



Actes de la
Conférence internationale de Géomatique et Analyse Spatiale
2012

*Proceedings of
Spatial Analysis and GEOmatics 2012*

Université de Liège, 6 au 9 novembre 2012

Éditeurs

R. BILLEN

M. BINARD

P. HALLOT

JP. DONNAY

Actes - SAGEO 2012 – *Proceedings*

©Unité de Géomatique, Université de Liège, Liège, novembre 2012.

ISBN à paraître

DOI <http://hdl.handle.net/2268/133126>

« Sageo est la cordelette qui noue les domaines, les approches, les disciplines et les acteurs qui font la géomatique. »

SAGEO a pour objectifs de :

- présenter les recherches récentes des domaines de la Géomatique, de l'Analyse Spatiale et de l'Information Géographique;
- réunir des communautés de chercheurs venant de disciplines diverses;
- promouvoir les relations entre les spécialistes développant et valorisant des recherches dans les organismes publics, dans les entreprises et dans les collectivités territoriales;
- fournir un lieu d'information et d'échange sur les recherches et les développements de l'information géographique en général, à un niveau national et international.

SAGEO 2012 est organisé à l'Université de Liège, du mercredi 7 au vendredi 9 novembre 2012. Les actes reprennent l'ensemble des communications orales et les textes de présentations des affiches soumises et acceptées par le comité après relecture « en double-aveugle ».

"Sageo is the rope which ties the different fields, approaches, disciplines and actors, promoting the geoinformatics."

SAGEO aims to :

- *present recent research in the fields of geoinformatics, spatial analysis and Geographic Information;*
- *gather communities of researchers from various disciplines;*
- *foster relations between specialists developing and enhancing research in government agencies, businesses and local authorities;*
- *provide an information and exchange place on research and development of geographic information in general, at national and international.*

SAGEO 2012 is held at the University of Liège, Wednesday 7 to Friday, 9 November 2012. Acts cover all the oral and poster submitted and accepted by a peer group after "double-blind" review.

Comité scientifique / *Scientific committee*

Roland BILLEN, Université de Liège (co-président)

Christophe CLARAMUNT, École Navale, Brest

Jean-Paul DONNAY, Université de Liège (co-président)

Thierry JOLIVEAU, Université Jean Monnet, Saint-Étienne

Thérèse LIBOUREL, Université de Montpellier II

Esteban ZIMANYI, Université Libre de Bruxelles

Comité d'organisation / *Steering committee* (Université de Liège)

Roland BILLEN*, Marc BINARD*, Yves CORNET*, Gwenaél DEVILLET**, Jean-Paul DONNAY*, Pierre HALLOT*, Mathieu JASPARD**, Jean-Paul KASPRZYK*, Luc LECLERCQ*, Christine PARTOUNE***, Anna POLETTO*, Nadia PONCELET*, Daniel SMETS*, Marie TROTTA*, Gilles WAUTELET*.

* Unité de géomatique

* Service d'études en géographie économique fondamentale et appliquée (SEGEFA)

* Laboratoire de méthodologie de la géographie

Orateurs invités / *Guest speakers*

Eliseo CLEMENTINI, Department of Industrial and Information Engineering and Economics, University of L'Aquila (Italy)

Thierry BADARD, Département des sciences géomatiques de l'Université Laval à Québec (Canada)

Frank CANTERS, Department of Geography, Vrije Universiteit Brussel (Belgium)

Comité de lecture / *Reviewers*

Thierry BADARD
Arnaud BANOS
Roland BILLEN
Marc BINARD
Alain BOUJU
Frank CANTERS
Nicholas CHRISMAN
Christophe CLARAMUNT
Eliseo CLEMENTINI
Claude COLLET
Yves CORNET
Helen COUCLELIS
Paule-Annick DAVOINE
Sophie DERUFFRAY
Gwenaël DEVILLET
Thomas DEVOGELE
Jean-Paul DONNAY
Johnny DOUVINET
Pierre DUMOLARD
Mauro GAIO
Pierre GANCARSKI
Jerome GENSEL
Jean-Marie HALLEUX
Pierre HALLOT
Eric HALLOT

Stéphane JOOST
Didier JOSSELIN
Jean-Paul KASPRZYK
Sylvie LARDON
Florence LEBER
Thérèse LIBOUREL
Michel MAINGUENAUD
Hélène MATHIAN
Sébastien MUSTIERE
Michel PASSOUANT
François PINET
Anne PUISSANT
Anne RUAS
Christian SALLABERRY
Serge SCHMITZ
Sylvie SERVIGNE
Cécile TANNIER
Marie TROTTA
Nico VANDEWEGHE
Christelle VANGENOT
Christiane WEBER
Éléonore WOLFF
Karine ZEITOUNI
Esteban ZIMANYI

Sponsors

- Groupe de Recherche MAGIS
- Institut Géographique National – Nationaal Geografisch Instituut, Bruxelles - Brussel
- Service Public Fédéral Finances, Administration Générale de la Documentation Patrimoniale – Federale Overheidsdienst Financien, Algemene Administratie van de Patrimoniumdocumentatie, Bruxelles - Brussel
- Service Public de Wallonie – SG Département Géomatique – Direction de la Géométrologie, Namur
- SIGGIS S.A. (Spatial Intelligence Genuine & Generic IT Solutions), Zaventem
- STAR-APIC, Liège
- Université de Liège : Faculté des Sciences, Fondation J.A. SPORCK, Service des Relations internationales

Avant-propos

La conférence SAGEO a pour objectif de réunir la communauté des chercheurs géographes et informaticiens travaillant dans le domaine de la géomatique. Elle est portée par le GdR MAGIS. Après avoir été accueilli à Toulouse (2010) et Paris (2011), SAGEO se tient en 2012 pour la première fois hors de France, à Liège.

Au terme de la sélection, 17 communications scientifiques ont été retenues, auxquelles s'ajoutent 4 communications techniques et 14 affiches scientifiques. Toutes les soumissions ont été soumises à une relecture en double-aveugle. Nous profitons de cet avant-propos pour remercier tous les collègues qui ont soumis une proposition et les membres du comité de lecture qui ont accepté la tâche ingrate mais indispensable à la qualité scientifique, de la relecture critique des soumissions.

Les communications scientifiques ont été rassemblées en 4 sessions thématiques : l'analyse régionale (3 communications), l'analyse urbaine (4 communications), la télédétection (5 communications) et la cartographie et les systèmes de référence (5 communications). Nous sommes particulièrement heureux du succès de la télédétection, thème quelque peu sous-représenté lors des dernières sessions SAGEO. D'autre part, les 4 communications de la session technique sont dédiées à des applications innovantes. Il s'agit de projets en cours ou de retours d'expérience, présentées par des géomaticiens travaillant dans le secteur privé de la géomatique. Enfin, à l'instar de ce qui s'est fait dans les sessions précédentes de SAGEO, les auteurs des affiches scientifiques sont invités à présenter oralement leurs travaux en séance plénière. Cette procédure originale offre plus de visibilité aux recherches réalisées dans la plupart des cas par de jeunes collègues, et elle facilite une discussion ultérieure face aux affiches, avec le public intéressé.

Pour introduire les trois journées du colloque, nous avons pu compter sur d'éminents spécialistes internationaux. E. CLEMETINI discute de l'application de la topologie à l'identification d'objets complexes. T. BADARD présente les concepts fondamentaux de la « GeoBI » (*Geospatial Business Intelligence*) et des SOLAP (*Spatial On Line Analytical Processing*), tandis que F. CANTERS démontre les potentialités de la télédétection des surfaces imperméables pour le suivi et la modélisation urbaine.

La communauté scientifique belge a bien accueilli l'organisation sur son territoire du colloque SAGEO. Le caractère international de la conférence s'en trouve ainsi renforcé, et gageons que ce soit l'occasion de nouveaux et fructueux contacts. Grâce au soutien logistique de nos sponsors – qu'ils en soient ici remerciés – nous avons tâché d'offrir à la conférence un cadre

digne de sa réputation et de sa qualité scientifique. Nous espérons ainsi que, pour une première « hors Hexagone », la session SAGEO de 2012 restera pour chacun une expérience agréable et enrichissante, et qu'elle suscitera auprès des responsables du GdR MAGIS l'envie de tenter l'aventure ailleurs.

Au nom du Comité scientifique, nous remercions toutes les personnes qui ont permis que puisse se tenir le colloque SAGEO 2012 dans les meilleures conditions, et en particulier tous les membres locaux du Comité d'organisation, les collègues qui ont accepté de présider les séances et les différentes autorités de l'Université de Liège.

Pour le Comité scientifique
Jean-Paul DONNAY & Roland BILLEN, co-présidents

Forewords

The SAGEO conference aims at gathering geographers and computer scientists working in the domain of Geomatics. SAGEO is supported by the GdR MAGIS. After being held in Toulouse (2012) and Paris (2011), SAGEO is organised in 2012, for the first time out of France, in Liege.

After completion of the double-blind review process, 17 scientific communications, 4 technical communications and 14 posters have been selected. We wish to thank all the researchers who have submitted their works and the members of the scientific committee who have accepted the thankless but necessary task of critical submission reviewing.

The scientific contributions are organized in 4 thematic sessions; regional analysis (3 communications), urban analysis (4 communications), remote sensing (5 communications) and cartography and reference systems (5 communications). We are especially happy with the success of the remote sensing theme which was a bit underrepresented in the last SAGEO meetings. On the other hand, the 4 technical communications are dedicated to innovative applications. It is about on-going projects or experiences feedback, presented by geomaticians working in the geomatics industry. Finally, following SAGEO habit, poster's authors are invited to make a brief oral presentation of their works during a plenary session; it brings more visibility to these contributions mainly done by early stage researchers and it facilitates further discussions in front of the posters.

Three leading international experts will introduced the three conference days; E. CLEMETINI will discuss the application of topology to complex object identification, T. BADARD will present fundamental concepts of Geospatial Business Intelligence (GeoBI) and Spatial Online Analytical Processing (SOLAP) and F. Canters will demonstrate the potential of remote sensing of sealed surfaces for monitoring and modeling of urban dynamics.

The Belgian scientific community is pleased to welcome the organisation of SAGEO on its « territory ». It strengthens the international aura of SAGEO and we are sure that it will generate new fruitful contacts. Thanks to our sponsors, we have made our best to welcome the participants in a venue worthy of SAGEO reputation. We hope that, for this first edition out of the "hexagon", SAGEO 2012 will remain for all of you a pleasant and rich experience and that it will encourage GdR MAGIS leaders to continue to "export" their conference.

In the name of the scientific committee, we wish to thank all the people who have contributed to the realization of SAGEO 2012, and especially all the members of local organizing committee, the sponsors, the colleagues who have accepted to chair sessions and finally the University of Liege authorities.

*For the scientific committee
Jean-Paul DONNAY & Roland BILLEN, co-chairs*

Table of Contents

Une approche ontologique d'aide à la classification des objets d'une image satellitaire	1
<i>Samuel Andrés, Damien Arvor and Christelle Pierkot</i>	
Contribution du raisonnement logique à la cartographie des flux	13
<i>Francoise Bahoken</i>	
Des variables didactiques pour éduquer l'analyse visuelle	30
<i>Raffaella Balzarini, Paule-Annick Davoine and Muriel Ney</i>	
Comprendre les pressions sur l'environnement à travers des modèles de potentiels anisotropes ?	35
<i>Joël Boulier</i>	
Simuler les évolutions urbaines à l'aide de données géographiques urbaines 3D	53
<i>Mickaël Brasebin, Sébastien Mustière, Julien Perret and Christiane Weber</i>	
Une fiche descriptive pour l'analyse des relations sémantiques entre textes et cartes	70
<i>Geoffrey Brun</i>	
A geomatic approach to Wildland-Urban Interface detection	73
<i>Alexandre Champendal, Carmen Vega Orozco, Raphaël Ceré, Mikhail Kanevski and Marj Tonini</i>	
Incident Citoyens capteurs. Retour d'expérience sur l'application G!v et Praj déployée au Danemark	77
<i>Anne-Sophie Collignon and Emmanuel Mondon</i>	
Géoréférencement et appariement de données issues des cartes de Cassini - Intégration dans un référentiel topographique actuel	96
<i>Benoit Costes, Éric Grosso and Christine Plumejeaud</i>	
Analyse des structures de haies et des linéaires pérennes dans deux paysages agricoles contrastés	113
<i>Sébastien Da Silva, Claire Lavigne and Florence Le Ber</i>	
Constructions cartographiques pour la caractérisation de l'aléa volcanique du Piton de la Fournaise : Cartographie de séries de données chrono-spatiales	129
<i>Paule-Annick Davoine, Cecile Saint-Marc, Andrea Dimuro, Thomas Staudacher, Marlène Villanova-Oliver and Patrice Boissier</i>	
Towards an internationally more accessible data model for the 2D large-scale topographic inventory of Flanders	145
<i>Ine De Cuyper and Jos Van Orshoven</i>	
Observatoire virtuel pour diffuser et réutiliser l'expertise en télédétection	158
<i>Laurent Demagistri, Yuan Lin and Thérèse Libourel</i>	
Can satellites help organic crop certification?	178
<i>Antoine Denis, Baudouin Desclee, Silke Migdall, Herbert Hansen, Heike Bach, Pierre Ott and Bernard Tychon</i>	

Comparison of discrete multi-criteria decision making methods for selection of afforestation sites	182
<i>René Estrella Maldonado, Ward Delabastita, Anja Wijffels and Jos Van Orshoven</i>	
Pertinence cognitive des représentations dans les cartes animées - Elaboration d'un protocole expérimental	198
<i>Emmanuel Ahr, Camille Guillemin, Nadine Mandran, Claire Cuntz and Elisabeth Chesneau</i>	
Vers une méthode de conception de SIG 3D collaboratif	203
<i>Rafika Hajji and Roland Billen</i>	
Student satisfaction towards software-specific and software-independent tutorials of GI ST 223	223
<i>Marianne Hubeau, Annelies De Meyer, Thérèse Steenberghen and Jos Van Orshoven</i>	
Analyse préalable à une infrastructure de données spatiales au Maroc	238
<i>Fatiha Ibannain and Jean-Paul Donnay</i>	
Map of economic risk of decrease in farmland costs in Stavropol Krai, caused by geological hazards as the territory-management tool	258
<i>Oleg Kiselev and Vera Pyrchenko</i>	
Le franchissement des obstacles en navigation intérieure	263
<i>Luc Leclercq</i>	
Une approche basée objet combinée avec les classificateurs avancés (SVM, RF, RN) pour la détection des changements du bâti	270
<i>Elmansouri Loubna</i>	
Vers une géospatialisation de l'éclairage public au Gabon	288
<i>Jean Sylvestre Makak, Daniel Thomas, Florence Palla, Etienne Emame and Hélène Elody Marcelle Ogouet Mboyo</i>	
Cours d'eau non navigables par intégration cartographique	312
<i>Catherine Marlier, Vincent Bombardier, Abdel Mokadem and Pierre Defourny</i>	
Les paysages numériques en 3D dans les démarches participatives. Vers une approche par les usages	323
<i>Matthieu Noucher, Sébastien Nageleisen, Laurent Couderchet, Julie Pierson, Maxime Demade, Irmi Obermaier, Nathalie Pinède and Véronique Lespinet-Najib</i>	
Apports de la technologie LIDAR dans l'objectivation écologique d'un territoire en amont d'un projet d'aménagement : aide à la caractérisation de l'habitat de l'Outarde canepetière dans la ZPS des Costières de Nîmes	329
<i>Pierre-André Pissard, Sylvie Vanpeene, Sylvie Durrieu and Laurent Albrecht</i>	
GéoPeuple: Analyse de l'évolution des territoires et de leur démographie sur 200 ans	338
<i>Anne Ruas, Christine Plumejeaud, Eric Grosso, Benoit Costes, Marie-Christine Vouloir, Claude Motte, Hervé Le Bras, Matthieu Cord, Nicolas Thome and Jonathan Guyomard</i>	
Une méthode de détermination du temps de parcours : application de recherches itinéraires dans un contexte d'urgence	344
<i>Mohamed Ayet Allah Bilel Soussi, Jean-Michel Follin, Guillaume Moreau, Alain Bouju and Laurent Polidori</i>	

Earth Observation for Reducing Emissions from Deforestation and forest Degradation	358
<i>Lauriane Cayet, Nathalie Stephenne and Mathieu Rahm</i>	
Une nouvelle méthode de calcul disovist en 2 et 3 dimensions	366
<i>Wassim Suleiman, Thierry Joliveau and Favier Eric</i>	
Do combined atmospheric and topographic correction methods improve land cover classification accuracy in mountain areas?	387
<i>Steven Vanonckelen, Stefaan Lhermitte and Anton Van Rompaey</i>	
Multifractal portrayal of the distribution of the Swiss population	392
<i>Carmen Vega Orozco, Jean Golay and Mikhail Kanevski</i>	
Formalisation des imperfections des données du système d'information agro-environnemental Observox - Gestion des imperfections	408
<i>Karima Zayrit and Eric Desjardin</i>	

Keyword Index

administration	77
afforestation	182
Afrique centrale	358
agricultural practices	408
agro-environmental indicator	408
aide à la décision	288
algorithme de recherche ditinéraires	344
analyse de documents	70
analyse spatio-temporelle	338
analyse visuelle	30
Android	77
animated map	198
anisotropie	35
appariement	96
belief functions	408
biodiversité	329
biophysical indicators	178
box-counting method	392
buildings' change detection	270
capteur	77
carte ancienne	96
carte animée	198
cartographic data integration	312
cartographie	13, 70, 129
cartographie de susceptibilité	30
Cassini	96
catalogage	238
CCSS4G	223
changement d'échelle	35
chaîne de traitements	158
citoyen	77
city clustering algorithms	73
cityGML	145
classification accuracy	387
co-production	203
collaboration	203
collaborative 3D GIS	203
colours	198
conceptual exercises	223
contextes spatiaux	35
corn	178
couleurs	198

crues	198
data merging	408
data modeling	145
densité	323
design	203
didactical variables.	30
données chrono-spatiales	129
données historiques	338
données imparfaites	408
données spatiales de référence	238
données spatio-temporelles	129
démarche participative	323
démographie	338
environmental information system	408
environnement	35
environnement de Workflow	158
espace visible	366
estimation du temps de parcours	344
evolutions urbaines	53
flux	13
fonctions de croyances	408
forest fires	73
forêt	358
FOSS4G	223
fouille de données spatiales	113
fusion de données	
fuzzy logic	
generalized entropy	
geographical database	258
geospatial	77
GIS	73, 258
GIS software	263
GIS T	223
	408
GRB	145
géoréférencement	96
géovisualisation 3D	323
hyperspectral	178
illumination	387
image satellite	158
images satellitaires	1
imperfect data	

imperfection	96
indicateur agroenvironnemental	408
indices	13
indoor navigation	263
infrastructure de données spatiales	238
infrastructure de mutualisation	158
integrated correction model	387
interaction spatiale	35
iPhone	77
isovist	366
landsat 5 TM image	387
lidar	329
lifelong learning	223
linéaires	113
logical-geometrical model	263
logique	13
logique floue	
map of economic risk	258
matrice	13
memorization	198
modélisation	35
modélisation conceptuelle	238
modélisation de paysage	113
morisita index	408
mountain areas	387
multi-criteria decision methods	182
multi-temporal image classification	73
multifractal	
mémorisation	198
natural language	263
neural Network	270
observation de la terre	1
obstacles crossing	263
oculomètre	198
ontologie	70, 238, 338
ontologies	1
organic agriculture	178
organic crop certification	178
PBL situation	30
perception	198
photométrie	366
population	392
potentiel	35

pratiques agricoles	
prevention of losses from hazardous processes	258
random Forest	270
REDD	358
reengineering	203
remote sensing	178, 258
RGE	288
rise of water level	198
risk from natural hazards	258
risques naturels	129
règles durbanisme.	53
réseau déclairage public	288
sandbox-counting method	
schema matching	145
SIG	77, 288
SIG 3D	53, 366
simulation	53
singularity spectrum	392
site location	182
situation-problème	30
structure de haies	113
student satisfaction	223
suitability mapping	30
supervised self-study	223
SVM	270
système dinformation environnemental	
sémiologie graphique	70, 129
temps	198
territoire	329
time	198
traitement dimage	338
trajectoire	344
télédétection	1, 358
usage	323
variables didactiques	30
VHSR image	270
visual analysis	30
voisinage	113
volcanologie	129
véhicules durgence	344
Walloon region	312
watercourses	312

wheat	178
wildland-urban interface	73

Une approche ontologique d'aide à la classification des objets d'une image satellitaire

Samuel Andrés — Damien Arvor — Christelle Pierkot

UMR 228 ESPACE-DEV (IRD - UM2 - UAG - UR)
Maison de la Télédétection
500 rue Jean-François Breton-34093 Montpellier Cedex 5
{samuel.andres, damien.arvor, christelle.pierkot}@ird.fr

RÉSUMÉ. Dans les systèmes d'information environnementaux, les observations effectuées à partir de capteurs de natures diverses constituent le socle de l'information. Les avancées technologiques et la mise à disposition des images satellitaires permettent aujourd'hui de traiter une grande quantité d'information à des échelles très différentes. Cependant, la connaissance humaine nécessaire à leur exploitation n'est pas issue du monde binaire, dépend d'une expérience, d'un contexte, d'un domaine et comporte des ambiguïtés. Dans le contexte de la télédétection, les recherches les plus récentes s'appuient sur une approche dite « orientée objet ». En informatique, la formalisation de la connaissance proposée par le génie ontologique permet l'expression de la connaissance thématique. Rapprocher ces deux domaines est l'un des principaux objectifs de projets scientifiques comme CARTAM-SAT et BIO-SOS.

ABSTRACT. In environmental information systems, the observations from sensors of various types are the foundation of information. Technological advances and the provision of satellite images now allow to process a large amount of information at very different scales. However, human knowledge necessary for their exploitation is not end of binary world, depends on experience, context, domain and contains ambiguities. In the context of Remote Sensing, the latest researches rely on an "object-based" called approach. In computing sciences, the knowledge formalization proposed by the ontological engineering allows the expression of thematic knowledge. Bringing these two domains closer is one of the main objectives of scientific projects as CARTAM-SAT and BIO-SOS.

MOTS-CLÉS : Ontologies, télédétection, observation de la terre, images satellitaires.

KEYWORDS: Ontologies, remote sensing, earth observation, satellite images.

2 SAGEO 2012 à Liège les 7, 8 et 9 novembre 2012

1. Introduction

L'accélération de la production de données, quel que soit le domaine d'application (environnemental, cartographie, ...), rend nécessaire l'évolution des outils et méthodes pour les exploiter. En télédétection, le développement des instruments de prise de vue satellitaire avec de meilleures résolutions spatiales, temporelles et spectrales (SPOT, Worldview, ...) donne accès à de plus en plus de détails dans l'image, élargissant de ce fait, le spectre des applications potentielles.

Pour analyser les images satellitaires à des fins d'exploitation dans un contexte particulier (e.g. cartographie de l'occupation du sol), l'expert en télédétection doit utiliser son expérience et sa connaissance personnelle, qui sont rarement formalisées et ne sont donc pas réutilisables en l'état par un processus automatique. Cela revient à constater l'existence d'un **fossé sémantique** que (Smeulders *et al.*, 2000) définit comme le « manque de concordance entre l'information que l'on peut extraire des données visuelles et l'interprétation que l'on peut en faire dans une situation donnée ». Pour réduire ce fossé, une première piste consiste à calibrer les images de manière à transformer la valeur des pixels en mesures physiques (réflectance, température...) qui sont déjà porteuses de sens. Cependant, cette première approche est encore très éloignée des concepts manipulés par les thématiciens.

La formalisation de la connaissance de l'expert apparaît donc comme un préalable indispensable à l'interprétation sémantique des images par un outil informatique. Cette piste de recherche est notamment évoquée par les projets CARTAM-SAT et BIO-SOS et nous pensons que l'utilisation des ontologies comme « spécifications formelles de conceptualisations » (Gruber, 1995), pourrait contribuer à améliorer l'exploitation des images satellitaires par les diverses communautés thématiques. De ce fait, une *ontologie de domaine* permettrait de passer d'un raisonnement assisté par ordinateur à un début d'automatisation.

Dans nos travaux, nous examinons le rôle que les ontologies peuvent jouer dans l'interprétation des images satellitaires. Notre réflexion porte sur l'explicitation de la connaissance experte et l'exploration des potentialités des logiques de descriptions, pour aider à la classification des objets des images satellitaires. Dans un premier temps, nous nous focalisons sur la description des concepts relatifs à l'image (e.g. bande spectrale...) et aux traitements (e.g. texture...) en télédétection. Cette représentation conceptualisée est ensuite utilisée pour construire une ontologie de la connaissance experte en télédétection explicitant la définition des concepts potentiellement représentés. À terme, nous souhaitons aussi intégrer, via des ontologies, de la connaissance spatio-temporelle et de terrain, afin d'affiner notre résultat et de tendre vers une interprétation thématique de l'image (cartographie, biodiversité...).

Après avoir examiné les travaux autour de l'utilisation des ontologies pour l'interprétation des images (2), nous présentons les ontologies de l'image et de la télédétection que nous avons produites au cours de notre approche, ainsi que la méthodologie associée (3). Nous validons ensuite notre démarche par une mise en œuvre expérimentale (4). Enfin, nous concluons et donnons quelques perspectives à ce travail (5).

2. État de l'art

L'interprétation des images a longtemps été basée sur des méthodes de classification par pixels (Lillesand *et al.*, 2003). Cependant, depuis quelques années, le développement d'une nouvelle approche basée sur le partitionnement de l'image en objets a fait son apparition dans la communauté (Hay *et al.*, 2008). Cette approche appelée *Geographic Object-Based Image Analysis* (GEOBIA) se déroule en trois étapes successives :

- 1) la segmentation délimite des régions de l'image ayant des caractéristiques communes. Idéalement, ces régions constituent des objets de l'image (Castilla *et al.*, 2008).
- 2) l'extraction définit les caractéristiques des objets, telles que la forme, la texture ou encore la réponse spectrale (Liu *et al.*, 2008).
- 3) la classification affecte une catégorie aux objets segmentés en fonction des attributs calculés en phase d'extraction. Cette dernière étape a pour but d'enrichir les objets de l'image afin de leur attribuer une sémantique significative pour l'expert (e.g. classe d'occupation du sol...).

Ces techniques ont montré des résultats satisfaisants mais se basent sur la connaissance experte du télédéacteur qui est subjective et non utilisable directement par un processus automatique.

Les ontologies sont une spécification explicite d'une conceptualisation, qui est une vision abstraite et simplifiée du monde (Gruber, 1995). En particulier, les *ontologies de domaine* permettent une représentation formelle des concepts et des relations d'un domaine scientifique. Pour développer ces représentations formelles, la communauté du génie ontologique produit et perfectionne des standards qui sont principalement le fruit de recommandations du World Wide Web Consortium (W3C). Ils représentent un intérêt certain si l'on souhaite bénéficier des réflexions de la communauté et d'une certaine interopérabilité. Le standard *Ressource Description Framework* (RDF) est destiné à la description des ressources relatives à un domaine d'étude. OWL (*Web Ontology Language*) enrichit le potentiel sémantique de RDF en intégrant les logiques de description. En s'appuyant sur ces règles logiques, un raisonneur¹ automatique peut alors inférer une nouvelle connaissance à partir de la connaissance explicitée par les ontologies et vérifier sa cohérence logique.

Les ontologies ont été appliquées dans de nombreux domaines scientifiques tels que la biologie, l'écologie ou l'hydrogéologie (Madin *et al.*, 2008 ; Tripathi et Babbaie, 2008). En particulier, l'application des ontologies pour l'interprétation d'images numériques intervient dans des domaines variés, des photographies de paysages à la télédétection, en passant par le biomédical. Chaque domaine présente des caractéristiques particulières (Schober *et al.*, 2004), (Durand *et al.*, 2007), (Hudelot *et al.*, 2008) et (Hassan *et al.*, 2009).

1. Un raisonneur est un logiciel spécifique permettant d'inférer de la connaissance en se basant sur les propriétés logiques qui caractérisent les concepts, les individus et les propriétés qui les relient.

4 SAGEO 2012 à Liège les 7, 8 et 9 novembre 2012

Certains travaux proposent d'utiliser les ontologies dès l'étape de segmentation. Dans (Hassan *et al.*, 2009), les ontologies proposées contiennent les paramètres de l'algorithme de segmentation et les étiquettes potentielles des régions. La description des concepts de l'ontologie se base principalement sur des caractéristiques géométriques. Après une segmentation initiale, les segments sont ajustés de manière à se rapprocher de leur description dans l'ontologie. Dans l'étape de classification, (Hudelot *et al.*, 2008) ont conceptualisé des relations spatiales floues dans des ontologies afin d'identifier des organes dans des images issues du domaine médical. Cette utilisation souligne l'importance de l'information spatiale dans ce domaine, car un organe est en grande partie identifiable par sa forme et sa position. Toutefois, dans les images satellitaires, les relations spatiales sont insuffisantes car on ne connaît pas *a priori* la disposition des objets géographiques dans les images (Durand *et al.*, 2007). L'information prise en compte prioritairement est alors la radiométrie² des objets. Enfin, il a été montré dans (Schober *et al.*, 2004) l'intérêt du raisonnement appliqué sur des attributs radiométriques pour l'interprétation automatique de photographies de paysage. L'approche de (Falomir *et al.*, n.d.) implique également les logiques de description dans la description qualitative d'images.

3. De l'image à la connaissance

Le fil directeur de notre approche consiste à **considérer le génie ontologique comme un terrain intermédiaire** de représentation entre le monde de l'image et le monde de la connaissance experte. La figure 1 illustre la manière dont l'information de l'image pourrait prendre une forme basée sur les ontologies (à gauche), de même que la connaissance de l'expert (à droite). Ces deux informations partageant désormais la même forme pourraient être utilisées conjointement.

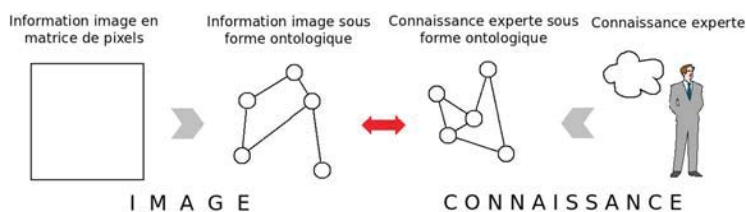


Figure 1. Les ontologies, un terrain intermédiaire entre l'image et la connaissance.

2. La radiométrie désigne la mesure du rayonnement électromagnétique, en particulier dans le domaine de la lumière visible (photographies de paysages) mais également dans d'autres longueurs d'onde (images satellitaires).

3.1. Conceptualisation de l'image satellitaire : ontologie de l'image satellitaire

Dans la perspective de l'exploitation automatique de l'image, il apparaît essentiel de commencer par automatiser sa description. Nous présentons dans la suite de l'article une description des caractéristiques de bas niveau exprimées à l'aide de concepts de l'image. D'autres niveaux de description peuvent être envisagés. Nous pensons par exemple à l'étiquetage des objets par un identifiant de classe issu d'une préclassification des pixels.

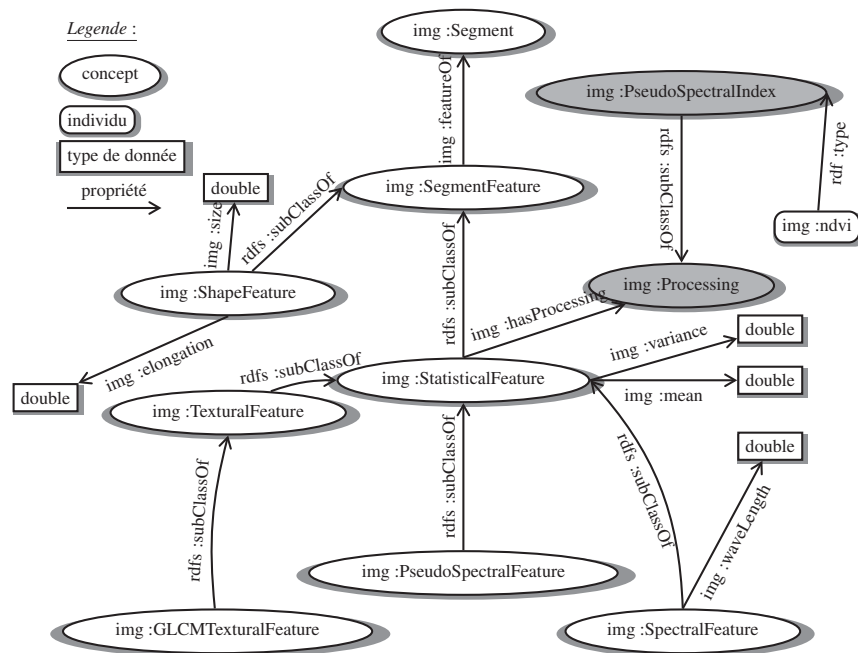


Figure 2. Les concepts de segment et de caractéristique de segment dans l'ontologie de l'image satellitaire. Les préfixes représentent les différents espaces de noms. Les concepts grisés concernent les traitements.

Dans un premier temps nous avons souhaité décrire l'image de manière objective, c'est-à-dire exprimer son contenu uniquement à l'aide d'un vocabulaire lié au monde de l'image. Pour cela, nous avons défini une ontologie de l'image satellitaire dont un extrait est donné en figure 2. Notre conceptualisation se rapproche des classes attributaires de (Durand *et al.*, 2007) tout en favorisant l'objectivité des concepts vis-à-vis des points de vue du capteur et du télédéacteur. En effet nous considérons que le concept de « caractéristique spectrale » est objectif car sa définition ne dépend ni du capteur ni du télédéacteur. En revanche, le concept de « caractéristique spectrale

6 SAGEO 2012 à Liège les 7, 8 et 9 novembre 2012

dans le rouge » serait subjectif à cause de la notion de « rouge », plus qualitative, assez vague et déjà sujette à interprétation. Nous pensons donc que la sémantique subjective doit être séparée de la sémantique objective et construite comme une connaissance experte de plus haut niveau sémantique dans une ontologie de la télédétection.

Deux familles de concepts émergent de notre proposition de conceptualisation de l'image satellitaire illustrée par la figure 2. La première concerne les **traitements** tels que la classification, la segmentation ou le calcul d'indices et de textures. Les concepts de *Processing* et de *PseudoSpectralIndex* font partie de ce groupe et permettent de s'y référer. L'indice *ndvi* souvent utilisé en télédétection pour discriminer la végétation est une instantiation de *PseudoSpectralIndex*. La seconde famille est centrée sur les **données**. Les concepts qui y sont attachés découlent principalement des concepts d'*image*, de *segment d'image* et de *pixel*. Le segment d'image (*Segment*) est au centre de notre réflexion. Chaque segment peut être associé à plusieurs caractéristiques (*SegmentFeature*) issues de l'analyse orientée-objet, selon la nécessité exprimée dans (Petridis *et al.*, 2006). Nous utilisons ici le mot *Feature* dans le sens de l'analyse d'images orientée objet qui associe à une image de segments (*Label image*), une image des caractéristiques (*Feature image*) que l'on souhaite leur attribuer (Lehmann, 2007). Cette notion se distingue donc de la définition faite dans les normes ISO 191xx dans lesquelles un *Feature* est défini comme une « abstraction du monde réel ».

La figure 2 présente le concept *ShapeFeature* qui permet de définir des propriétés relatives à la forme telles que la taille, l'élongation du segment... Les caractéristiques statistiques (*StatisticalFeature*) renvoient à des propriétés de type moyenne, variance etc. calculées sur la valeur des pixels du segment. Dans le cadre des images satellitaires, ces statistiques peuvent être relatives à une bande spectrale (*SpectralFeature*), un indice (*PseudoSpectralFeature*, e.g. image de NDVI) ou une texture (*TexturalFeature*, e.g. image d'énergie d'Haralick (Haralick *et al.*, 1973)). Les propriétés statistiques et de forme suivent la nomenclature définie dans (Lehmann, 2007).

3.2. Élicitation de la connaissance experte : ontologie de la télédétection

Sa connaissance experte permet au télédétecteur d'affecter une sémantique aux objets de l'image en fonction de leurs caractéristiques. Dans notre approche, une seconde ontologie explicite cette connaissance de manière à la numériser. Elle contient des *concepts définis* qui représentent les différentes sémantiques potentielles qu'un segment peut prendre pour un télédétecteur. Leur *définition* s'appuie sur les *concepts primitifs* issus de l'ontologie de l'image satellitaire et sur les constructeurs des logiques de description (Napoli, 1997). Par exemple, le concept de végétation dans l'image peut être rapproché d'un indice de végétation NDVI compris entre 0,3 et 0,8 (Derivaux *et al.*, 2008). Ainsi, nous pouvons définir ce concept comme suit :

$$\text{SegmentVegetal} \equiv \text{Segment} \sqcap \exists \text{hasFeature}(\text{SegmentFeature} \\ \sqcap \exists \text{hasProcessing}\{\text{ndvi}\} \sqcap \exists \text{Mean}\{\geq 0,3\} \sqcap \exists \text{Mean}\{\leq 0,8\})$$

où la partie de droite de l'axiome d'équivalence est constituée exclusivement de concepts et propriétés importés de l'ontologie de l'image satellitaire.

Il est important d'insister sur le fait que les concepts de l'ontologie de télédétection découlent des concepts de l'image : ce sont pour la plupart des spécialisations du concept de *segment d'image* et non pas des concepts de terrain. Il ne s'agit pas de donner une description de la *forêt comme environnement* mais bien du *morceau d'image qui représente de la forêt*. Il apparaît toutefois que les concepts de l'expert en télédétection sont en relation avec des concepts de terrain. Les ontologies pourraient également intervenir et contribuer à rapprocher ces deux niveaux (ou points de vue) en se basant par exemple sur des classifications existantes comme le *Land Cover Classification System*.

3.3. Approche méthodologique

L'approche que nous proposons suit les étapes de la méthodologie orientée objet en utilisant les ontologies précédemment définies pour la phase de classification. La figure 3 montre l'organisation de notre proposition.

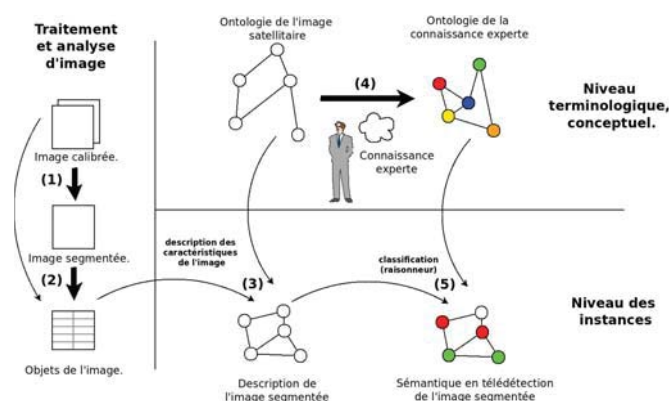


Figure 3. Expression de la sémantique des segments et inférence sur la description de l'image.

L'étape de segmentation (figure 3, (1)) détermine en grande partie la caractérisation et donc l'interprétation sémantique des objets. Il s'agit d'un domaine de recherche à part entière dans lequel il semble également intéressant de faire intervenir l'expertise (Derivaux *et al.*, 2010). En ce qui nous concerne, nous avons décidé de segmenter à partir d'une préclassification par pixel adaptée à la télédétection (Baraldi *et al.*, 2006) de manière à travailler sur des objets cohérents (Castilla *et al.*, 2008) et à nous affranchir au mieux des problèmes liés à la segmentation.

L'extraction des caractéristiques des segments est effectuée par une étape d'analyse orientée objet traditionnelle (figure 3, (2)). Les segments de l'image et leurs ca-

8 SAGEO 2012 à Liège les 7, 8 et 9 novembre 2012

ractéristiques de forme et statistiques (spectrale, pseudospectrale, texturale) peuplent l'ontologie de l'image satellitaire et constituent une base de connaissances relative à l'image étudiée (figure 3, (3)).

L'ontologie de la connaissance en télédétection est construite à partir de la connaissance experte, en utilisant les concepts de l'image satellitaire (figure 3, (4)). Combinée à la description ontologique de l'image satellitaire, elle permet de classer les segments. Cette dernière étape est réalisée en utilisant le potentiel des logiques de descriptions pour inférer une interprétation des segments (figure 3, (5)) via un raisonneur.

4. Mise en œuvre

Nous avons appliqué notre approche à un extrait d'image satellitaire avec l'exemple du concept de segment de végétation explicité ci-avant. Les illustrations que nous présentons ont été obtenues à partir d'une image Landsat 5 de la ville de Santarem (état du Para au Brésil) du 12 juillet 2009.

4.1. Outils et données

Nous avons commencé par choisir un outil de traitement des images satellitaires adapté à nos besoins. *Orfeo Toolbox* (OTB) est une bibliothèque logicielle développée par le Centre National d'Études Spatiales (CNES) français sous licence libre³ utilisée en télédétection. Ce choix permet de garder l'entière maîtrise du travail de recherche sans crainte de subir des modifications imposées par l'éditeur. De plus, le code source est écrit en C++, dont la popularité permet d'envisager la collaboration avec d'autres outils d'ingénierie des connaissances, même si dans la communauté de gestion des connaissances le langage Java jouit d'une popularité supérieure. OTB permet de réaliser des traitements sur des images, tels l'extraction de bandes spectrales, le calcul d'indices et de textures, des opérations de segmentation ou de classification.

Il s'agit ensuite de mettre sur le même plan (conceptuel) les connaissances issues de l'image et celles issues de l'expertise grâce aux différentes ontologies de domaine. Sur le plan pratique, ceci peut être mis en œuvre de plusieurs manières. L'interface de programmation *Soprano* qui utilise les vocabulaires standards du web sémantique offre la possibilité d'alimenter un graphe contenant la description conceptualisée de l'image, puis de sérialiser cette description au format RDF/XML. Cela permet ensuite de la communiquer à d'autres logiciels permettant de naviguer et de raisonner sur la description de l'image, comme *Protege*. Si l'on souhaite au contraire intégrer le raisonnement directement à la suite de l'analyse orientée-objet, il est possible d'utiliser le raisonneur FaCT++⁴ ; en revanche, FaCT++ ne permet pas directement la sérialisation des bases de connaissances dans un format standard.

3. http://smc.cnes.fr/PLEIADES/Fr/lien3_vm.htm

4. <http://owl.man.ac.uk/factplusplus/>

4.2. Procédé de description

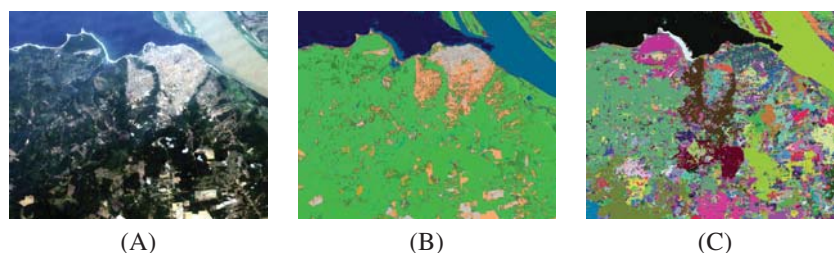


Figure 4. Classification par pixels (B) puis segmentation (C) de l'image initiale (A), prise dans les environs de Santarem (Brésil) sur une zone de 25×20 km.

L'étape de délimitation des segments cohérents se déroule en trois temps (voir figure 4). L'image initiale est calibrée en réflectance et température de manière à transformer les comptes numériques en valeurs physiques. On peut ensuite passer à une étape de classification par pixel en utilisant l'algorithme proposé dans (Baraldi *et al.*, 2006) (figure 4 (B)). Les pixels connexes de même classe sont ensuite regroupés pour former les segments (figure 4 (C)) à caractériser. On extrait ensuite les caractéristiques et on peuple l'ontologie de l'image satellitaire.

Le formalisme unique utilisé pour les différentes ontologies de domaine permet leur utilisation concertée. Par exemple, en utilisant le langage OWL, on peut utiliser pour cela le mécanisme d'importation d'ontologies. L'importation des concepts définis par le télédéacteur dans la base de connaissance issue de la description de l'image satellitaire rend possible de retrouver la sémantique experte des segments en inférant cette connaissance grâce à un raisonneur (La figure 5 concerne les segments de végétation).



Figure 5. Les segments correspondant au concept de végétation sont en blanc, les autres sont en noir.

Il n'est pas forcément évident de valider ce premier résultat. Les seuils utilisés dans l'expression de la connaissance experte ont été tirés de la littérature (Derivaux *et*

10 SAGEO 2012 à Liège les 7, 8 et 9 novembre 2012

al., 2008). Un risque est d'introduire un biais en faisant appel à une expertise basée sur une connaissance différente...un autre de « valider » simplement le raisonneur. De manière à évaluer notre approche, nous avons donc opté pour une comparaison à *une autre méthode* d'analyse de *la même image*, avec *la même connaissance*. La référence est un masque réalisé par seuillage des pixels sur leur valeur de l'indice ndvi (figure 6). Les deux images présentent une cohérence quant à la répartition des zones végétalisées. La figure 6 propose également une validation statistique représentant la matrice de confusion basée la sémantique des pixels dans les deux images (figure 5 et 6). Ces deux résultats se ressemblent... alors pourquoi chercher à remplacer des méthodes traditionnelles par une nouvelle approche ?

Notre objectif est ici d'explorer les possibilités offertes par les ontologies dans le cadre de l'exploitation des images satellitaires. Dans notre approche, la connaissance est formalisée et numérisée. La classification des objets est ainsi effectuée grâce à un *raisonnement réalisé par la machine* et basé sur la *connaissance experte explicitée*.

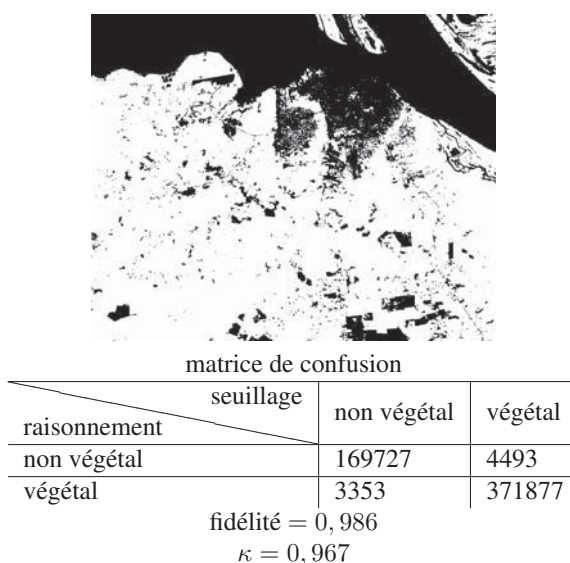


Figure 6. Les pixels dont le NDVI est compris entre 0,3 et 0,8 sont en blanc, les autres sont en noir. La matrice de confusion compare la sémantique des pixels de cette image avec celle de la figure 5. Cette classification permet d'évaluer qualitativement et quantitativement la pertinence du résultat produit par la méthode ontologique.

5. Conclusion et perspectives

Dans cet article, nous avons présenté une approche basée sur l'utilisation des ontologies pour rapprocher l'information de l'image de la connaissance de l'expert, ceci dans un but de classification sémantique des images satellites. Nous avons formalisé

la connaissance issue de l'image et de l'expertise du télédéacteur dans deux ontologies que nous utilisons conjointement pour effectuer un premier raisonnement sur l'image satellite. Dans cette tâche, il est possible de mettre à profit les standards du web sémantique. Nous avons vérifié notre méthodologie sur une image de Santarem en Amazonie brésilienne. Les premiers essais de raisonnement automatique se sont révélés encourageants et démontrent la faisabilité d'une telle approche.

Nous souhaitons maintenant consolider et développer notre méthode en affinant le résultat de la classification par des considérations relatives à la forme des objets, aux relations spatiales qu'ils entretiennent ou encore grâce à leur définition selon le point de vue du thématique sur le terrain. Nous pensons que les ontologies peuvent offrir la possibilité de combiner la connaissance de terrain et l'expertise en télédétection à des apports relatifs au caractère spatio-temporel des objets. Pour cela, nous envisageons de poser des passerelles entre ces différents niveaux de connaissance (image, terrain, spatio-temporel) afin d'exploiter au mieux la connaissance explicitée dans chacune des ontologies. Dans cette hypothèse, le rattachement de nos définitions de concepts à des entités définies par une ontologie de plus haut niveau permettrait de les homogénéiser et faciliterait ainsi l'établissement de ces passerelles. À cet égard, l'ontologie SWEET⁵ (*Semantic Web for Earth and Environmental Terminology*) qui propose un grand nombre de concepts sur lesquels il est possible de bâtir des concepts propres à un domaine de connaissance (image ou thématique dans notre contexte) représente une perspective intéressante. Ce n'est toutefois pas la seule piste, DOLCE⁶ étant par exemple utilisée dans la communauté des applications multimedia (Petridis *et al.*, 2006). Une ontologie cadre basée sur les normes géographiques, définie par notre équipe, spécifierait quant à elle les concepts relatifs au domaine spatio-temporel.

6. Remerciements

Les axes de recherche présentés sont développés au sein de l'UMR 228 ESPACE-DEV dans le cadre du projet CARTAM-SAT (programme PO FEDER 2007-2013) et du *European Union Seventh Framework Program* FP7/2007-2013 sous la convention de subvention n°263435 pour le projet BIOSOS.

7. Bibliographie

Baraldi A., Puzzolo V., Blonda P., Bruzzone L., « Automatic spectral rule-based Preliminary Mapping of calibrated Landsat TM and ETM+ images », *IEEE Transactions on geoscience and remote sensing*, vol. 44, n° 9, p. 2563 - 2586, september, 2006.

Castilla G., Hay G. J., *Object-based image analysis : spatial concepts for knowledge-driven remote sensing applications*, Springer, chapter Image objects and geographic objects, p. 91-110, 2008.

5. <http://sweet.jpl.nasa.gov/ontology/>

6. <http://www.loa.istc.cnr.it/DOLCE.html>

12 SAGEO 2012 à Liège les 7, 8 et 9 novembre 2012

- Derivaux S., Forestier G., Wemmert C., Lefevre S., « Extraction de détecteurs d'objets urbains à partir d'une ontologie », *Atelier Extraction de Connaissance à partir d'Images (ECOI), Journées Francophones Extraction et Gestion des Connaissances (EGC 2008)*, Sophia Antipolis, France, p. 71-81, Jan, 2008.
- Derivaux S., Forestier G., Wemmert C., Lefèvre S., « Supervised image segmentation using watershed transform, fuzzy classification and evolutionary computation », *Pattern Recognition Letters*, vol. 31, p. 2364-2374, 2010.
- Durand N., Derivaux S., Forestier G., Wemmert C., Gançarski P., Boussaïd O., Puissant A., « Ontology-based object recognition for remote sensing image interpretation », *19th IEEE International Conference on Tools with Artificial Intelligence*, vol. 19, p. 472 - 479, October, 2007.
- Falomir Z., Jiménez-Ruiz E., Escrig M. T., Museros L., « Describing images using qualitative models and description logics », n.d., University Jaume I, Castellon, Spain.
- Gruber T. R., « Toward principles for the design of ontologies used for knowledge sharing », *International Journal of human Computer Studies*, vol. 43, p. 907-928, 1995.
- Haralick R. M., Shanmugam K., Dinstein I., « Textural features for image classification », *IEEE Transactions on Systems, Man and Cybernetics*, vol. SMC-3, n° 6, p. 610-621, November, 1973.
- Hassan S., Hétry F., Palombi O., « Segmentation de maillages guidée par une ontologie », *22èmes journées de l'association francophone d'informatique graphique*, Arles, 2009.
- Hay G. J., Castilla G., *Object-based image analysis : spatial concepts for knowledge-driven remote sensing applications*, Springer, chapter Geographic Object-Based Image Analysis (GEOBIA) : A new name for a new discipline, p. 75-89, 2008.
- Hudelot C., Atif J., Bloch I., « Fuzzy spatial relation ontology for image interpretation », 2008.
- Lehmann G., « Label object representation and manipulation with ITK », July - December, 2007.
- Lillesand T. M., Kiefer R. W., Chipman J. W., *Remote Sensing and Image Interpretation*, Wiley, October, 2003.
- Liu Y., Guo Q., Kelly M., « A framework of region-based spatial relations for non-overlapping features and its applications in object based image analysis », *ISPRS Journal of Photogrammetry & Remote Sensing*, vol. 63, p. 461-475, 2008.
- Napoli A., Une introduction aux logiques de descriptions, Research Report n° RR-3314, INRIA, 1997.
- Petridis K., Bloehdorn S., Saathoff C., Simou N., Dasiopoulou S., Tzouvaras V., Handschuh S., Avrithis Y., Kompatsiaris Y., Staab S., « Knowledge representation and semantic annotation of multimedia content », *Vision, Image and Signal Processing, IEE Proceedings*, vol. 153, p. 255 - 262, 2006.
- Schober J.-P., Hermes T., Herzog O., « Content-based image retrieval by ontology-based object recognition », in *KI-2004 Workshop on Applications of Description Logics*, p. 61-67, 2004.
- Smeulders A. W. N., Worring M., Santini S., Gupta A., Jain R., « Content-Based Image Retrieval at the End of the Early Years », *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, vol. 22, n° 12, p. 1349-1380, December, 2000.

Contribution du raisonnement logique à la cartographie des flux.

Françoise Bahoken

Université Paris-Est / IFSTTAR / UR Splott
UMR Géographie-Cités / Equipe Paris/ Groupe Cartomouv'
Le Descartes II, 2, rue de la Butte Verte, 93166 Noisy-le-Grand
(frcse_bhk@yahoo.fr)

RÉSUMÉ. Le défi de la cartographie des flux consiste à réduire « l'effet-spaghetti » lié au chevauchement parfois trop important des figurés présents sur la carte. La solution classique vise à « sélectionner » les flux à représenter avec « parcimonie ». Plusieurs approches de la sélection sont envisageables selon que le raisonnement est mené sur les aspects graphiques, cartographiques ou lors du traitement préalable des données. Les analyses statistiques et mathématiques –appliquées à la cartographie– caractérisent une démarche qui appréhende la « sélection » notamment comme une « réduction » de l'univers des données par la production d'indices. Nous proposons ici une extension de l'approche algébrique des matrices de flux, mise en œuvre par Tobler (1979, 1982), vers le raisonnement logique. Trois opérations issues de la combinaison des opérateurs classiques des ensembles sont proposées comme variante des indices classiques utilisés en cartographie des flux. Leur efficacité est présentée lors d'une mise en œuvre sur une matrice des migrations internes.

ABSTRACT. The challenge of flow mapping is to reduce the "spaghetti-effect" related to the overlapping of sometimes too many graphic ties. The conventional solution aims to "select" them with "parsimony." Several approaches are possible depending on the way the selection is processed: on the graphical aspects, the cartographical one or during the pretreatment data step. Statistical and algebraic analysis-applied to mapping- consider "selection" as a way to "reduce" the universe of data especially by the production of indexes. We propose an extension of the flow matrix algebra approach, implemented by Tobler (1979, 1982), toward a logical reasoning. Three operations from the combination of logical operators are proposed as alternative commonly indices used in flow mapping. Their effectiveness is presented through an implementation on a matrix of internal migration.

MOTS-CLÉS : matrice Origine-Destination, cartographie, flux, indices, raisonnement logique.

KEYWORDS: Origine-Destination matrix, flow mapping, indexes, logical reasoning.

1. Introduction

Les cartes de flux Origine-Destination (OD) ont pour objet de représenter la structure spatiale des échanges qui se sont produits, sur une période, entre des couples de lieux d'origine (i) et de destination (j). La lisibilité des cartes dépend du nombre de lieux (N) considérés : le nombre de figurés est égal à $N*(N-1)$ quand il n'existe aucun flux d'un lieu vers lui-même ou bien que l'on ne considère que les flux interzones. En effet, lorsque N dépasse la dizaine (Tobler, 1987), la carte est caractérisée par un important chevauchement des liens préjudiciable à la lisibilité des échanges. C'est « l'effet-spaghetti » (Breukelman & al., 2009). La conséquence de cet effet est qu'il rend les cartes de flux inopérantes et par conséquent « inutiles » (Torricelli, 1997).

La solution classique au problème général de l'illisibilité des cartes de flux OD consiste à « simplifier » la figure (Bunge, 1969) c'est-à-dire à sélectionner avec « parcimonie » (*ibid.*) les éléments qui seront représentés ou, inversement, ceux qui seront éliminés de la carte. La notion de simplification appliquée aux données de flux OD prend différentes acceptations en fonction du cadre d'analyse dans lequel elle est réalisée. La carte des flux étant le résultat d'une série de traitements d'ordres cartographiques et liés au traitement des données (statistiques, mathématiques, ...), la sélection des éléments à représenter est envisageable à tous les stades de l'analyse.

Sur le plan graphique, la sélection des flux à représenter peut être assimilée à une réduction du nombre de figurés, c'est-à-dire au choix des figurés à représenter. Elle consiste généralement à :

- Choisir une focale de représentation des flux, du point de vue :
 - des liens : « *je représente les échanges entre les lieux (à l'aide de flèches)* » ;
 - des lieux : « *je représente la résultante des échanges au niveau des lieux (à l'aide de symboles ou d'aplats)* » ;
 - des liens et des lieux.
- Transformer éventuellement les flux en indices (si les données ont fait l'objet d'un traitement préalable) : « *je représente le volume des flux (entrants et sortants) des unités spatiales* ».
- Appliquer un critère graphique de sélection des flux fondé sur leur :
 - Valeur absolue : « *je représente les flux supérieurs à telle valeur* » ;
 - Valeur relative ou contribution au total des échanges : « *...les flux qui contribuent à x% du total des échanges* ».

Sur le plan statistique et mathématique, la sélection des flux à représenter vise à :

- Choisir une focale d'analyse de la matrice (F), c'est-à-dire à raisonner sur le tableau en axant l'analyse sur les éléments situés :
 - Au-dessus de la diagonale : « *j'analyse les flux émis ou sortants (Fij)* » ;
 - En-dessous de la diagonale : « *(...) entrants ou reçus (Fji)* » ;
 - Sur la diagonale : « *j'analyse les flux internes aux lieux (Fii)* » ;
 - Sur un vecteur-ligne¹ : « *(...) les flux en provenance de tel lieu (Oi)* » ;
 - Sur un vecteur-colonne : « *(...) les flux à destination de tel lieu (Dj)* » ;
- Analyser l'ensemble des flux :
 - bruts : « *(...) tous les flux (Fij, Fji)* » ;
 - Relatifs : « *(...) tous les flux en pourcentage du total des échanges* » ;
 - décomposés, réduits ou transformés en indices : « *j'analyse une matrice symétrisée* » ;
- Appliquer un critère de sélection graphique des échanges. La réduction-transformation des données de flux sous la forme d'indices n'a pas pour vocation première de procéder à l'élimination du nombre de liens. C'est plutôt l'application d'un critère graphique de sélection qui permet d'éliminer les liens (bruts ou transformés) qui ne contribuent pas à une part significative des échanges. Le critère graphique couramment utilisé est fondé sur une discrétisation de la série en quintiles : « *je représente les 5% (ou les 25%, les 50%, etc.) des flux (Fij, Fji) les plus importants* ».

En articulant les aspects graphiques, cartographiques et liés au traitement des données, nous pouvons considérer cinq manières différentes de sélectionner les flux. Celles-ci diffèrent selon que le raisonnement est mené sur :

- Les lieux (Oi, Dj) ;
- Les liaisons entre les lieux (Fij, Fji), ou conjointement sur les lieux et les liens (Fij, Oi, Dj, Fji) ;
- La valeur ou l'intensité des liaisons entre les lieux (Fij, Fji) ;
- l'emprise spatiale des flux appréhendée, par exemple, en termes de termes de distance entre les lieux (Dij).
- La sélection peut également être réalisée dans le cadre d'un raisonnement mixte qui prend en compte l'intensité des flux et la distance parcourue (Fij, Dij).

Notre propos intéresse la question de la sélection des flux à représenter réalisée lors du traitement des données et dans le cadre d'un raisonnement mené sur la valeur ou l'intensité des liaisons, si celles-ci ont été transformées en indicateurs. Nous

¹ Le vecteur-ligne (ou colonne) peut correspondre à un lieu particulier ou à une somme marginale.

rappelons, dans un premier temps, les solutions apportées par les approches classiquement mises en œuvre que sont l'approche probabiliste et l'approche statistique de transformation des flux en indices. L'approche algébrique, telle qu'elle est mise en œuvre par Tobler, est présentée dans un second temps. Elle consiste à décomposer la matrice en deux composantes symétriques et antisymétriques. Nous proposons d'étendre cette approche au raisonnement logique. Pour ce faire, nous présentons, dans un troisième temps, l'utilisation des opérateurs MIN et MAX pour réduire et transformer les flux au préalable de leur cartographie. Nous appliquons ensuite les indices proposés en soulignant les améliorations qu'ils permettent sur le plan cartographique et conceptuel, par rapport à un indice plus classique : le bilan net ou solde des flux.

2. Les approches couramment utilisées dans le traitement des flux

Les approches classiques du traitement des données de flux, au préalable de leur cartographie, consistent –dans leur phase descriptive– à transformer les données en indicateurs.

2.1. L'approche probabiliste

L'approche probabiliste est la plus couramment mise en œuvre pour préparer les données de flux à cartographier. Elle s'appuie sur certaines propriétés des tableaux de contingence considérant à juste titre que les matrices OD peuvent y être assimilées. La méthode consiste à transformer la matrice en tableau de fréquences pour pouvoir comparer entre elles des sous-populations de taille variable. Les valeurs brutes des flux sont alors transformées en pourcentages (%) en ligne ou en colonne qui représentent des distributions conditionnelles. Différents indices sont ensuite élaborés, tels que ceux présentés par le tableau 1.

Indice	Formulation
part des flux ayant pour origine i et destination j	$PO_{ij} = (F_{ij}/O_i) * 100$
part des flux ayant pour origine j et destination i	$PO_{ji} = (F_{ji}/D_j) * 100$
part des flux ayant pour destination j et origine i	$PD_{ij} = (F_{ij}/D_j) * 100$
part des flux ayant pour destination i et origine j	$PD_{ji} = (F_{ji}/O_i) * 100$

Tableau 1. Exemples d'indices de probabilité d'origine et de destination des flux

La transformation des valeurs brutes de la matrice en indices fait intervenir les sommes marginales en lignes (et en colonnes). Elles correspondent respectivement au total des flux émis $O_i = \sum_j (F_{ij})$ et reçus $D_j = \sum_i (F_{ji})$, dans le calcul des probabilités d'émission et de réception.

Le premier avantage de ces indices est qu'ils intègrent les valeurs des flux et leur effet marginal aux lieux (O_i , D_j) en les pondérant. Le second avantage, issu du

premier, est qu'ils permettent de mettre en évidence les parts d'échanges qui s'éloignent de la distribution marginale.

Sur le plan cartographique, la représentation de ces indices nécessite deux à trois composantes : des liens proportionnels aux échanges et une coloration discrétisée de leurs probabilités. Un critère graphique de sélection est éventuellement appliqué afin de limiter le nombre de figurés qui seront placés sur la carte.

2.2. *L'approche statistique*

L'approche communément qualifiée de statistique, également répandue, consiste à appliquer les méthodes d'analyse des données propres aux tableaux de contingence aux matrices de flux. C'est le cas, par exemple, de l'analyse des correspondances mise en œuvre par Piron (2010). La représentation graphique est celle du plan factoriel qui illustre l'importance des flux entre deux zones par leur proximité dans le plan.

Les matrices peuvent également être transformées en indices qui permettent de focaliser l'analyse sur les lieux et/ou sur les couples de lieux.

Pumain et Saint-Julien (2001 : 18) suggèrent, dans ce cadre, la production d'un certain nombre d'indices. L'indice de préférence illustre, par exemple, la présence de relations préférentielles entre des lieux, c'est-à-dire de valeurs de flux observées qui sont supérieures aux valeurs de flux estimées à l'aide d'un modèle², compte tenu des capacités d'émission et de réception des lieux (O_i , D_j). Ces indices sont élaborés dans le cadre de l'analyse des résidus issus des modèles et peuvent faire l'objet de plusieurs formulations (Pumain, Saint-Julien, 1990).

Dans un registre d'analyse descriptif, Courgeau (1980 : 63) préconise notamment le recours à un indice d'« efficacité des zones » dans le système d'échanges qui résulte de la part du solde (ou migration nette) dans le total des flux d'une zone ou à celui « de compensation » qui correspond à l'asymétrie des échanges en traduisant le rapport des migrations nettes (le solde) sur les migrations totales (le volume) (Courgeau, 1980 : 141).

Dans l'ensemble, la particularité des indices élaborés dans un objectif de cartographie des flux permettent de comparer la valeur ou l'intensité des échanges entre les zones. Les plus courants sont le volume, le solde et l'asymétrie :

- le volume : $(F_{ij} + F_{ji})$;
- le solde : $(F_{ji} - F_{ij})$;
- l'asymétrie : $(F_{ji} - F_{ij}) / (F_{ij} + F_{ji})$

² Les hypothèses mises en œuvre dans le cadre des modèles gravitaires permettent de modéliser mathématiquement les interactions, ce qui peut être appréhendé comme une manière de sélectionner les flux à représenter selon leur valeur : les cartes réalisées à partir des résidus positifs des modèles d'interaction illustrent, par exemple, des relations préférentielles entre les lieux.

Les résultats cartographiques obtenus associent généralement les flux allers et les flux retours, ce qui constitue une gêne pour la compréhension de la structure spatiale des échanges. Cette fusion ou association est due au mélange d'informations de natures différentes qui mériteraient d'être analysées individuellement. La solution à ce second problème de la cartographie des flux est à rechercher dans les travaux de Tobler. Ce dernier a, en effet, suggéré de décomposer les flux en deux sous-parties distinctes en adoptant une démarche issue de l'algèbre des matrices.

3. L'approche algébrique

L'approche algébrique du traitement des matrices de flux, au préalable de leur cartographie a été mise en œuvre par W. Tobler (1979, 1982), lors de la mise en place de la version additive du modèle gravitaire qu'il propose. Elle se distingue de l'approche statistique classique car elle conduit à la réalisation de deux cartes complémentaires d'un même système d'échanges issus d'indices qui contribuent en partie à résoudre l'effet de fusion.

D'après W. Tobler (1979, 1982), les matrices de flux asymétriques (si $i \neq j$, $F_{ij} \neq F_{ji}$) peuvent être décomposées en une « partie symétrique » notée (F+) et une « partie antisymétrique » notée (F-). (F+) et (F-) concourent toutes deux à la définition de la matrice complète puisque $F = (F+) + (F-)$. L'auteur définit la partie symétrique (F+) de la matrice comme celle qui relève de la stabilité du système et la partie asymétrique comme celle qui relève de sa dynamique (F-). La décomposition des matrices de flux en composantes positive (F+) et négative (F-) proposée par Tobler est fructueuse. Elle présente l'avantage d'améliorer la compréhension des processus sous-jacents. D'après Cauvin et Reymond (1991), le (F+) exprime le « fond vital du système », sa stabilité tandis que le (F-) exprime la « direction nette des échanges » (Cauvin C., Reymond H, 1991 : 468).

La solution apportée par Tobler est menée dans le cadre d'une approche algébrique de la réduction des matrices. Elle s'appuie sur les principes de trigonalisation des matrices de vecteurs utilisés en algèbre linéaire. Le (F+) proposé par Tobler correspond à l'opération de symétrisation des éléments situés de part et d'autre de la diagonale et le (F-) à une anti-symétrisation ainsi définies :

$$(F+) = (F_{ij} + F_{ji})/2 \quad [1]$$

$$(F-) = (F_{ji} - F_{ij})/2 \quad [2]$$

Tobler propose ainsi deux variantes des indicateurs classiques de volume ($F_{ij} + F_{ji}$) et de solde ($F_{ji} - F_{ij}$) classiquement utilisés en cartographie des flux, mais sous la forme de moyennes arithmétiques du volume (F+) et du solde (F-). La représentation cartographique du (F-) est particulièrement fructueuse car elle peut

être appréhendée comme une forme d'attractivité des lieux lorsque sont représentés sur une même figure : les liens (F-) et l'effet marginal des flux aux lieux, en termes de volume, par exemple.

Dans le cadre d'un raisonnement qui porte sur la valeur des échanges, la question générale de la sélection des flux à représenter consiste à déterminer un critère graphique α en deçà (ou au-dessus) duquel les flux seront cartographiés ; que ces derniers aient été préalablement transformés en indices ou pas. Si les approches statistiques et algébriques sont fructueuses, leur extension dans le cadre d'un raisonnement logique permet d'augmenter la qualité des cartes réalisées.

4. Vers l'approche logique

Le problème de la quantité de flux à sélectionner (ou à éliminer) au préalable de leur cartographie est difficile à résoudre. L'efficacité d'un indicateur est généralement liée à la capacité de son auteur à analyser les résultats produits, d'où la multiplicité d'indices ou de dénominations pour une même opération : les termes de « solde », de « bilan net » ou encore de « migration nette » résultent tous de la différence entre les flux entrants et sortants. De la même façon, le choix du critère de sélection graphique des flux ne fait pas l'objet d'un consensus. On peut même considérer que ce critère revêt un caractère imprécis.

Plusieurs raisons peuvent être avancées pour expliquer la complexité du problème de la sélection des flux à cartographier.

La première raison tient au nombre de figurés à représenter qui peut être très important. Cette caractéristique visuelle de l'effet-spaghetti des cartes de flux contraint à appliquer un critère de sélection des flux fondé sur la valeur ou sur l'intensité³ des $N*(N-1)$ liaisons.

La seconde raison tient à la manière d'appréhender les flux. Si la représentation territoriale⁴ (ou discrète) des flux est la plus courante, sa version continue⁵ permet d'étendre tout en les simplifiant, les approches proposés antérieurement grâce aux possibilités combinatoires offertes par le raisonnement logique.

4.1. De l'algèbre des matrices à la logique matricielle

Dans le cadre d'une vision spatiale des échanges, la matrice des flux est appréhendée comme un ensemble complexe formé par une infinité de valeurs définies dans le temps –la période d'observation des données– et dans l'espace, la zone d'étude. Ces flux sont caractéristiques d'une distribution nécessairement

³ Des critères de distance entre les couples de lieux, ainsi que les sommes marginales en ligne (O_i) et en colonne (D_j) peuvent également être pris en compte pour fixer ce seuil.

⁴ L'emprise des flux est représentée sur un maillage formé d'unités spatiales délimitées, contigües ou pas, souvent hétérogènes.

⁵ L'emprise des flux est assimilée à une surface, ensemble complexe formé par une infinité de valeurs.

définie sur un intervalle de valeurs borné par un minimum et par un maximum qui sont absolus. Des bornes intermédiaires, non absolues mais relatives, peuvent également être définies afin de partitionner la distribution : c'est l'objet des discrétisations utilisées en cartographie. Nous pouvons dès lors considérer que les flux sont définis précisément en leurs bornes et de manière imprécise ailleurs⁶.

En adoptant un tel raisonnement qui postule l'imprécision des valeurs des flux, nous considérons que l'ensemble des flux (F) bilatéraux d'une zone d'étude est défini sur l'intervalle borné par leur valeur minimum [3] et par leur valeur maximum [4] de la manière suivante :

$$F_{\text{Min}} = \text{Min}(F_{ij}, F_{ji}) \quad [3]$$

$$F_{\text{Max}} = \text{Max}(F_{ij}, F_{ji}) \quad [4]$$

L'intérêt de l'argument est d'autoriser l'application de la théorie des sous-ensembles flous. Pour ce faire, nous posons que les flux (F) sont constitués par un premier ensemble formé par les flux aller (F_{ij}) et par un second ensemble qui correspond aux flux retour (F_{ji}) ; les deux ensembles forment le tout défini sur $[\text{Min}(F_{ij}, F_{ji}) ; \text{Max}(F_{ij}, F_{ji})]$.

NOTE. — Nous aurions pu également considérer deux sous-ensembles à définir, composés d'une part imprécise de flux aller et de flux retours, tel que cela est classiquement utilisé en logique floue : cette approche constitue le stade suivant de l'analyse qui n'est pas présenté ici.

La question littérale de la sélection des flux peut ainsi être formulée de l'une des manières suivantes, prises à titre d'exemples :

- « *je représente les flux aller ET les flux retour ?* » ;
- « *je représente les flux aller OU les flux retour ?* » ;
- « *je représente les flux aller ET NON les flux retour (et inversement) ?* ».

Les réponses, en guise de solutions, peuvent être obtenues par l'application de règles d'inférence qui utilisent les opérateurs logiques des ensembles que sont « ET, OU, NON ». La question de la sélection des flux à cartographier peut ainsi être étendue à l'aide des opérateurs MIN et MAX tels que formalisés par [3] et [4] qui sont généralement combinés aux opérateurs logiques « ET, OU, NON » formalisés dans un raisonnement flou.

⁶ Les flux agrégés résultent, en effet, de l'agrégation de déplacements individuels liés à l'application de filtres spatiaux (taille du maillage) et temporels (pas de temps) qui conduisent à discrétiser les flux sur une base spatio-temporelle. Appréhender les flux de manière continue revient à faire fi des filtres.

4.2. Les opérateurs MIN et MAX

Le principe général de la logique floue est d'intégrer la part d'incertitude liée à l'appartenance des valeurs x d'un ensemble X à deux classes distinctes (Dubois, Prade, 1980). Elle diffère du raisonnement scientifique classiquement adopté qui est fondé sur la logique booléenne pour laquelle un objet appartient strictement à une classe (VRAI=1) et non à une autre (FAUX=0).

Dans le cas de la combinaison de deux ensembles A et B qui est celui qui peut être appliqué à la cartographie des flux, l'extension des opérateurs classiques d'Union (\cup) ou « OU », d'Intersection (\cap) ou « ET », de Négation (ou inverse ou complémentaire) (\neg) ou « NON » est étendue à l'aide des formulations floues proposées par L. A. Zadeh (1965, cité par Dubois et Prade, 1980) de la manière suivante (Dubois, Prade, 1980 : 11) :

$$[(A \cup B)(x)] = \text{MAX} [fA(x), fB(x)] \quad (\text{i})$$

$$[(A \cap B)(x)] = \text{MIN} [fA(x), fB(x)] \quad (\text{ii})$$

avec : $fA(x)$ et $fB(x)$ comme fonctions d'appartenances respectives de $(A \cup B)$ et de $(A \cap B)$ pour des valeurs réelles comprises entre 0 et 1. L'intervalle $[0,1]$ est considéré comme le pseudo intervalle de distribution des valeurs où les MIN-MAX ont les rôles respectifs de bornes inférieure et supérieure. Appliqués aux matrices de flux, les opérateurs d'union et d'intersection peuvent être ainsi formalisés :

$$F_{ji} \cap F_{ji} = \text{MIN} (F_{ij}, F_{ji}) \quad [3]$$

$$F_{ij} \cup F_{ji} = \text{MAX} (F_{ij}, F_{ji}) \quad [4]$$

NOTE. — Certaines approches privilégient la formation du « ET » par le produit $(A.B)$ et celle du « OU » par la moyenne $(+)$ appréhendée comme « opérateur somme », là où nous avons choisi les MIN et MAX.

C'est ainsi à partir des opérateurs MIN et MAX que nous étendons les opérateurs logiques d'union et d'intersection à la question de la sélection des flux [3] et [4]. Sans aller plus loin dans le raisonnement flou⁷, il est intéressant d'analyser les résultats cartographiques obtenus par l'application des [3] et [4].

⁷ Nous pensons que le raisonnement flou peut être appliqué de deux manières en cartographie de flux : la première voie est celle de la production d'indices à laquelle contribue en partie cet article et la seconde est celle de la définition de sous-ensembles flous, comme une forme de discrétisation ; toutes deux contribuent *in fine* à sélectionner et réduire les éléments à cartographier.

4.3. Application

L'utilisation des opérateurs MIN-MAX en cartographie des flux peut être effectuée de deux manières : de façon individuelle en réalisant deux cartes : l'une sur les $\text{MIN}(\text{Fij}, \text{Fji})$ et l'autre $\text{MAX}(\text{Fij}, \text{Fji})$; ou de manière combinée sous la forme d'indices (Bahoken, 2012).

NOTE. — Les opérateurs MIN-MAX peuvent également servir à la réduction de la distribution des flux à un intervalle⁸ aux bornes imprécises afin de sélectionner des plages de valeurs de flux à représenter. Cette procédure qui s'apparente à la cartographie d'une des classes d'une distribution obtenue après discrétisation des valeurs n'est pas présentée ici.

Dans le registre d'application des opérateurs MIN et MAX aux flux asymétriques à cartographier, l'opérateur logique « ET » consiste à considérer le $\text{Min}(\text{Fij}, \text{Fji})$, soit l'intersection des flux aller et retour et l'opérateur « OU » le $\text{Max}(\text{Fij}, \text{Fji})$, soit l'union des flux aller et retour. Ils conduisent donc d'abord à une réduction de la matrice complète en deux matrices diagonales par leur transformation en indices.

Nous avons choisi d'illustrer la présentation des problèmes qui rendent les cartes de flux illisibles et les solutions que nous proposons avec une matrice des migrations internes.

4.3.1. Présentation des données

Les données sont celles des flux bilatéraux (Fij, Fji) et asymétriques (si $i \neq j$, $\text{Fij} \neq \text{Fji}$) qui se sont produits entre 1989 et 1990 entre les 114 districts qui composent l'ex-Tchécoslovaquie. Il s'agit d'une matrice carrée des migrations résidentielles qui renseigne sur le nombre de déplacements (événements de migrations) interdistricts. Elles émanent de l'Annuaire Démographique Tchécoslovaque (Pohyb Obyvatelsva) et ont été préparées par Grasland (Grasland, 1994 : 439). Les données sont disponibles à l'échelle des 114 districts ainsi qu'à différents échelons administratifs tel que celui des dix régions.

La réalisation d'une carte des flux de l'ex-Tchécoslovaquie consiste à représenter les 114×113 liaisons interdistricts théoriques soit 18882 figurés, de tailles variables selon la valeur du flux. L'ensemble des flux n'est pas toujours disponible : dans le cas de l'ex-Tchécoslovaquie, 84% d'entre eux ont été collectés

⁸ La logique floue permet, en effet, de définir des ensembles (ou classes) de valeurs de flux dont les limites (bornes Min et Max) ne sont pas clairement définies. En pratique, le critère α de sélection des flux est difficile à estimer avec précision puisque l'allure de la carte, en termes de quantité de flux représentés, ne change pas fondamentalement dans un certain intervalle. Une voie intéressante consiste donc à discrétiser les valeurs des flux suivant un raisonnement flou.

soit 10556 flux. Compte tenu du nombre de liaisons effectives interdistricts, la représentation des flux de l'ex-Tchécoslovaquie pose problème, comme l'illustre la Figure 1. C'est « l'effet-spaghetti » (Breukelman & al., 2009).

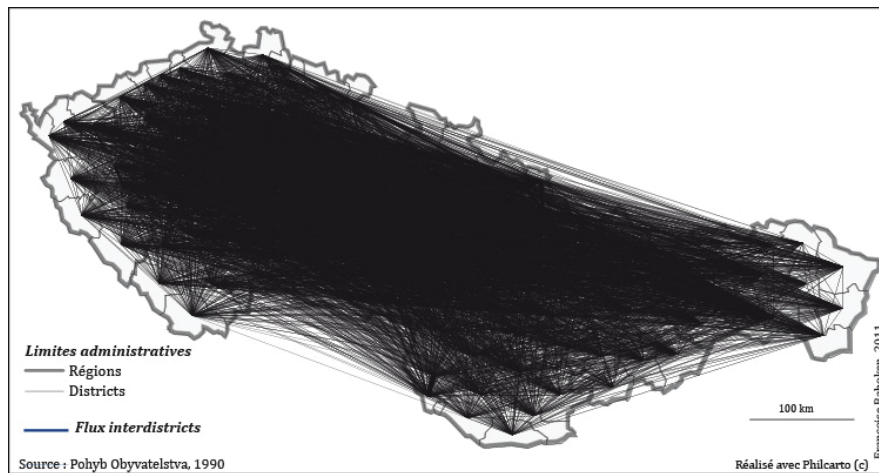


Figure 1. Illustration de l'effet-spaghetti sur les flux de l'ex-Tchécoslovaquie

Pour pallier à cet effet-spaghetti, nous présentons les résultats obtenus suite au calcul des indices fondés sur les opérateurs MIN-MAX.

4.3.2. Exemples de cartes

L'intérêt cartographique du recours aux opérateurs MIN et MAX pour représenter les flux est indéniable. Les figures 2 et 3 ne représentent, en effet, que les couples de lieux qui présentent des flux bilatéraux : les relations unilatérales sont donc éliminées. En effet, sur les 84% de flux disponibles, seuls 4526 liens, soit 36% des échanges sont représentés pour l'opérateur MAX et 5832 liens, soit 47% pour l'opérateur MIN. La différence du nombre de liaisons tient à l'asymétrie de la matrice (F) où le nombre de (Fij) diffère de celui des (Fji). Le recours aux MIN-MAX opère donc une vraie réduction de l'univers des données en éliminant une quantité importante des flux à représenter tout en donnant un vrai sens à la carte qui est analysé dans la section suivante.

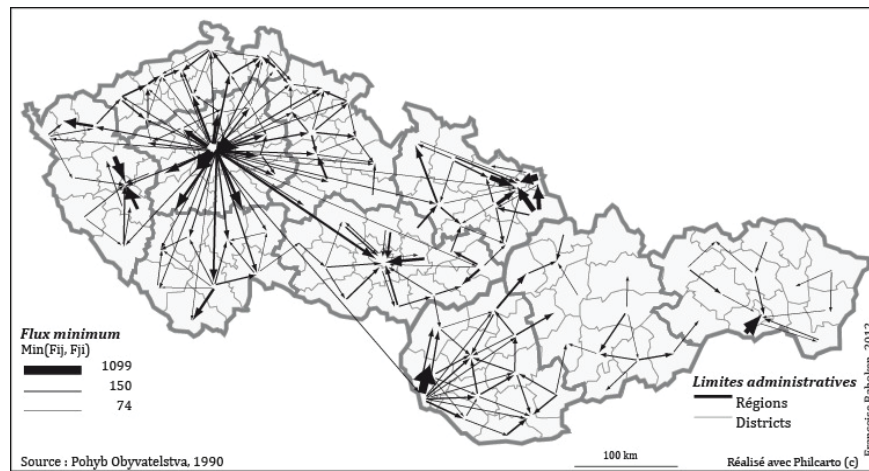


Figure 2. Exemple de cartographie du $\text{Min}(F_{ij}, F_{ji})$

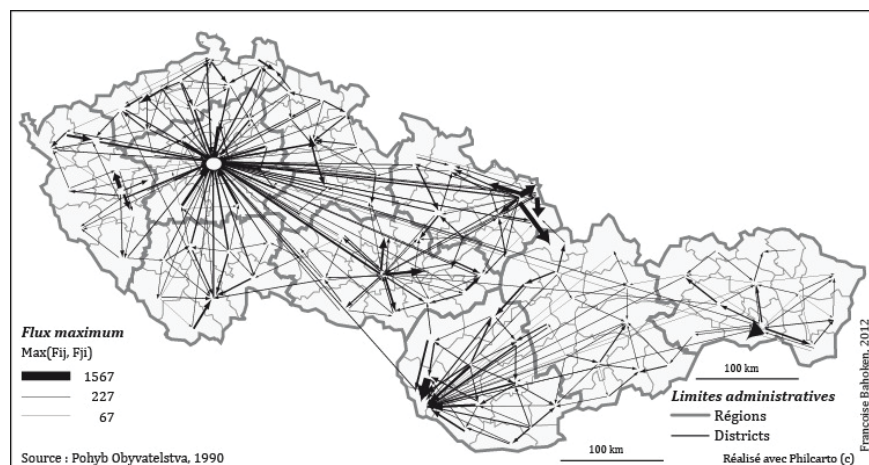


Figure 3. Exemple de cartographie du $\text{Max}(F_{ij}, F_{ji})$

Les opérateurs MIN et MAX ont des propriétés logiques et des possibilités combinatoires très intéressantes pour l'élaboration de plusieurs indices dans un objectif de cartographie des flux (Bahoken, 2012). Dans le cadre du raisonnement logique que nous proposons, nous considérons les MIN-MAX comme des indices qui traduisent des composantes symétriques complémentaires des échanges. Nous pouvons les considérer comme des variantes du volume classique ($F_{ij} + F_{ji}$) des échanges puisqu'elles sont fondées sur les opérateurs logiques d'Union et d'Intersection qui privilégient la formation du « ET » par la somme. En suivant le même raisonnement, il est possible de définir une variante du solde classique des échanges ainsi défini par [5] et représenté sur la figure 4.

$$F_{\text{Solde}} : [\text{Max}(F_{ij}, F_{ji}) - \text{Min}(F_{ij}, F_{ji})] \quad [5]$$

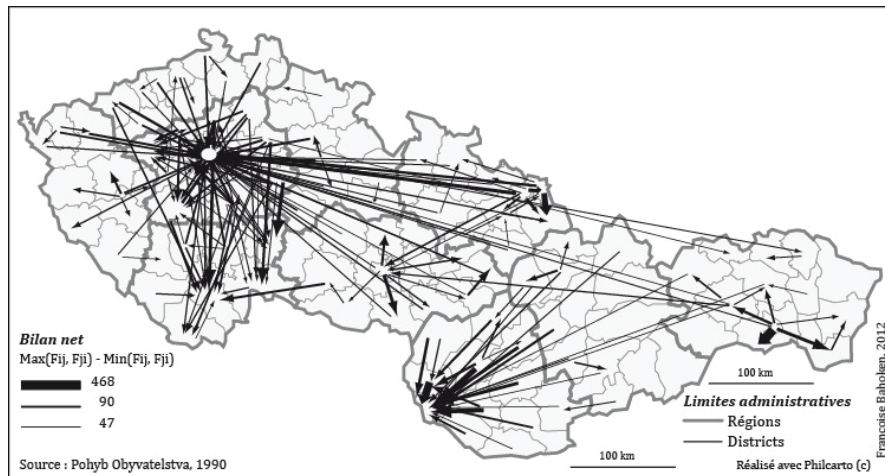


Figure 4. Exemple de cartographie du $[\text{Max}(F_{ij}, F_{ji}) - \text{Min}(F_{ij}, F_{ji})]$

L'intérêt du recours à une formulation logique du bilan net des échanges $[\text{Max}(F_{ij}, F_{ji}) - \text{Min}(F_{ij}, F_{ji})]$ est de pouvoir se départir, si on le souhaite, du dessin des têtes de flèches : la valeur de l'indice est, en effet, toujours positive à la différence de celle du solde ou du (F-). Elle évite aussi de réaliser deux cartes : l'une pour les valeurs positives et l'autre pour les valeurs négatives ou de calculer un indice intermédiaire telle que celui de l'asymétrie. Sur la figure 4, nous avons toutefois fait le choix de représenter le bilan net et de faire figurer l'orientation des échanges afin qu'elle puisse être comparée aux figures 2 et 3. Ici, l'orientation des flèches a pour objet de traduire également un processus de compétition entre les lieux.

4.3.3. Intérêts du raisonnement logique

L'intérêt du recours aux opérateurs MIN et MAX dans un objectif de cartographie des flux est, outre la réduction importante qu'ils permettent, un avantage conceptuel qui va au-delà de la seule combinaison logique qui a servi à leur élaboration. Deux intérêts principaux méritent qu'on s'y attarde.

Sur le plan conceptuel, le premier intérêt de la cartographie des MIN-MAX tient à la simplification du raisonnement apporté par la logique et à la précision des concepts de symétrie et d'antisymétrie des échanges appliqués à la cartographie des flux, tels qu'ils ont été proposés notamment par Tobler.

La différence conceptuelle entre le (F+) proposé par W. Tobler en [1] comme partie symétrique des échanges avec la représentation des opérateurs [3] et [4] de la même composante symétrique des échanges est sensible. Au sens mathématique, la composante symétrique d'une matrice (F) est sa transposée notée (F^T). La représenter consiste donc à permuter deux à deux les éléments situés de part et d'autre de la diagonale. La figure 5 illustre, à l'aide d'un exemple fictif, les modifications de la cartographie des flux liées à l'acceptation du concept de symétrie. La composante symétrique présente un réel intérêt en cartographie lorsqu'elle est appliquée aux matrices asymétriques où (F) diffère de (F^T), c'est-à-dire lorsque $F_{ij} \neq F_{ji}$; la cartographie de la composante symétrique des matrices triangulaires ($F_{ij} = F_{ji}$) est, en effet, moins intéressante⁹ à notre sens. Ainsi le (F+) tel qu'il est présenté en [1] correspond donc non pas à la composante symétrique de la matrice mais, comme nous l'avons indiqué précédemment, à la moyenne des flux bilatéraux (voir figure 5). La cartographie de la « vraie » composante symétrique des flux est par conséquent illustrée par la représentation des opérateurs MIN et MAX (voir figures 5, 2 et 3).

Valeurs observées :

$F_{ij} = 100$ et $F_{ji} = 65$

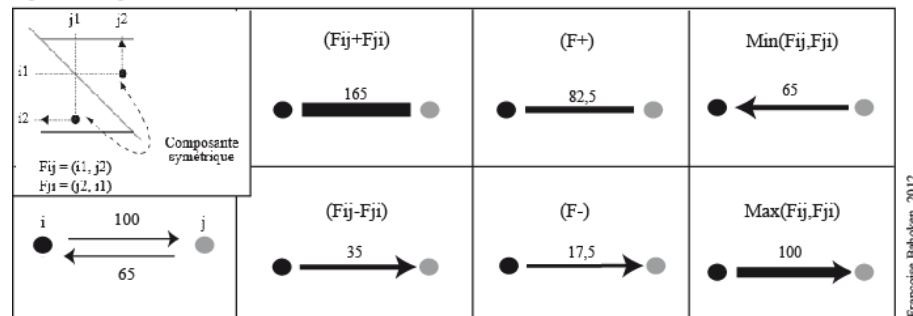


Figure 5. Illustration de l'application des MIN et MAX

⁹ Une matrice « symétrique » est une matrice carrée où $F = F^T$. Sa composante symétrique est alors formulée par : $(F + F^T)/2$.

Le second intérêt conceptuel du recours aux opérateurs MIN-MAX tient à l'effet de leur mise en œuvre sur la qualification des mobilités représentées. En d'autres termes, l'application des opérateurs logiques permet de cartographier différents types de mobilités distinguées, notamment, selon un critère de distance sans que celui-ci soit explicitement intégré l'analyse. Dans le cas de l'ex-Tchécoslovaquie, la carte du $\text{Min}(\text{Fij}, \text{Fji})$ ne représente que des mobilités locales qui se manifestent à courte distance, alors que celle du $\text{Max}(\text{Fij}, \text{Fji})$ illustre des mobilités à plus longue distance. Toutes deux figurent un type de mobilités particulier et non pas, comme on aurait pu le croire, les seuls flux minimaux ou maximaux. Les mobilités locales se manifestent sur la figure 2, majoritairement à une échelle infra régionale suggérant l'existence d'un effet-frontière de type barrière et/ou le rôle de la distance, processus qui ont pu être mis en évidence et estimés par Cattán et Grasland (1994) et Grasland (1994). A l'inverse, les mobilités sur longues distances illustrées par le $\text{Max}(\text{Fij}, \text{Fji})$ révèlent des flux interrégionaux qui caractérisent une perméabilité de la frontière. Il s'agit, d'une part, de flux plus coûteux en termes de déplacement (en raison de l'effort lié au franchissement de la frontière) et qui ont dépassé les éventuels freins au déplacement compte tenu de la distance qu'ils ont parcouru. De l'autre, des flux qui se dirigent vers des lieux fortement attractifs.

Nous pouvons dès lors considérer que la cartographie du $\text{Min}(\text{Fij}, \text{Fji})$ met en évidence un processus de coopération entre les lieux en illustrant les flux bilatéraux intenses qui se traduisent sur la carte par des mobilités locales, tandis que celle des $\text{Max}(\text{Fij}, \text{Fji})$ présente les phénomènes de compétition entre des lieux fortement attractifs.

5. Conclusion

Les cartes des flux réalisées à partir des flux transformés en indices, dans le cadre d'une approche logique, sont complémentaires à celles qui peuvent être mises en œuvre dans le cadre d'approches plus classiques. Les variantes proposées de la représentation des parties symétriques de la matrice présentent des intérêts multiples. Outre l'amélioration qualitative des cartes réalisées, elles permettent d'énoncer différentes hypothèses sur l'attractivité des lieux, sur la hiérarchisation des échanges mais également sur l'existence d'effets-frontières particuliers. Des possibilités de combinaisons des opérateurs MIN et MAX sont également envisageables pour la production d'indices variés (Bahoken, 2012).

Il convient de noter que la transformation des flux en indices ne doit pas faire perdre de vue que les flux ainsi transformés nécessitent ensuite d'être convenablement sélectionnés (Tobler, 1981) afin d'illustrer une part significative du système d'échanges. Cette sélection intervient davantage dans le cadre d'approches graphiques et cartographiques mises en œuvre après l'étape de traitement de données. Les procédures successives de réduction des données, de transformation en

indices puis de sélection des échanges sont, en effet, articulées pour une résolution du problème général d'illisibilité des cartes de flux à partir d'un raisonnement sur la valeur ou l'intensité des échanges. Les solutions que nous avons présentées concourent donc à la réduction de deux problèmes de la cartographie des flux. Le premier est celui de l'effet-spaghetti et le second est un effet de fusion lié au mélange d'informations de nature différentes. La solution à ce second problème réside dans la décomposition des flux bilatéraux puis à leur transformation dans le cadre d'approches algébriques et logiques.

Cependant, les résultats cartographiques proposés sont biaisés par l'hétérogénéité du maillage qui sert de fond de carte. Celle-ci est liée aux différentes tailles de lieux qui conduit à favoriser les zones les plus importants en termes de surface et/ou les couples de lieux dont l'éloignement est suffisant pour permettre une représentation aisée du figuré qui illustre le flux. Réduire puis sélectionner les échanges sans tenir compte de la taille des lieux et/ou d'une distance qui les sépare conduit à l'absorption visuelle des flux par les zones de grande taille et/ou des couples de lieux qui sont situés à une distance « raisonnable ». Cet effet d'absorption visuelle est le troisième problème d'une cartographie des flux qu'il convient de résoudre en introduisant des critères liés à la taille et/ou à la distance qui sépare les lieux dans la procédure de sélection des données au préalable de leur représentation.

6. Bibliographie

Bahoken F., Cartographie des flux et production d'indices sur l'intensité des échanges, *Actes du Colloque Les Interfaces : Enjeux de Nature, de Science et de Société (LIENS)*, Maison Méditerranéenne des Sciences et de la Société (MMSHS), Aix-en-Provence, 15-16 mai 2012.

Breukelman J., Brink G., De Jong T., Floor H., Selecting Flow Data, *Manual. Flowmap software v.3.2*, Faculty of Geographical Sciences, Utrecht University, the Netherlands, 2009, p. 42-44.

Bunge W., *Theoretical geography*, (Second edition), Lund Studies in Geography, Serie C: General and Mathematical Geography, No. 1., Lund, Sweden, Gleerup, 1966.

Cauvin C., Reymond H., Interaction spatiale et cartographie : les solutions de W. Tobler. *Espace, Populations, Sociétés*, n° 3, 199, p. 467-485.

Cattan N., Grasland C., Migrations et effets de barrières en Tchécoslovaquie, in Rey, V.(dir.), *La Tchécoslovaquie en 1992 : transition, fragmentation, recomposition*, Presses de l'ENS Fontenay-Saint-Cloud, Coll. Société, Espaces, Temps, 1994, p. 97-120.

Courgeau D., *Analyse quantitative des migrations humaines*, Masson, Paris, 1980. 225 p.

Dubois D, Prade H., M, *Fuzzy sets and systems: theory and applications*, Academic Press, INC, Mathematics in Sciences and Engineering, London, Vol. 144, 1980, 389 p.

Grasland C., Limites politiques et barrières migratoires : l'exemple de l'ex-Tchécoslovaquie (1965-1989), in Galluser W. (ed.), *Political Boundaries and Coexistence*, IGU Symposium, Basle, 1994, p. 425-440.

Pumain D., Saint-Julien T., *L'analyse spatiale. Les interactions spatiales*, Armand Colin, Coll. Cursus Géographie, 2001, 170 p.

Pumain D., Saint-Julien T., Les migrations interurbaines. *Mappemonde*, n°90, vol 1, 2 p.

Tobler W., A Model of Geographical Movement, *Geographical Analysis*, n°13, vol.1, 1981, p. 1-20.

Tobler W., Cartographic study of movement tables, *Statistical Graphics Mapping national Computer Graphics Association*, Anaheim, California, 1982, 4 p.

Tobler W., Experiments In Migration Mapping, *The American Cartographer*, n°14, vol. 2, 1987, p. 155-163

Torricelli G., Navetteurs en Suisse et en Italie du nord : images des flux et de leurs changements, *Mappemonde*, 1997, p. 9-13.

Françoise Bahoken est ingénieur d'études, recherche et formation en cartographie (traitement et représentation de données à représentation spatiale). Elle est actuellement accueillie en détachement à l'Institut Français des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux (IFSTTAR). Elle effectue par ailleurs une recherche doctorale sur les méthodes de cartographie des flux à l'UMR 8504 Géographie-Cités (équipe PARIS).

Des variables didactiques pour éduquer l'analyse visuelle

Raffaella Balzarini *, **- Paule-Annick Davoine *- Muriel Ney *

* LIG 681 rue de la Passerelle 38400 Saint Martin d'Hères

** ESRI France 64 chemin des Mouilles 69134 Ecully

{Raffaella.Balzarini, Paule-Annick.Davoine, Muriel.Ney}@imag.fr

RÉSUMÉ : Dans le cadre d'une recherche en cours sur les difficultés d'apprentissage des SIG dans l'enseignement supérieur et plus particulièrement auprès d'étudiants non géographes, nous nous intéressons au rôle des variables didactiques dans la résolution d'un problème spatial lorsqu'il s'agit de stimuler des opérations d'analyse visuelle sur une carte de susceptibilité du sol.

ABSTRACT: This poster shows an ongoing research on GIS learning difficulties for graduate students in environment domains. The poster aims to highlight the role of didactical variables in suitability mapping problems when visual analysis tasks need to be boosted.

MOTS-CLÉS: analyse visuelle, cartographie de susceptibilité, problème, variables didactiques

KEYWORDS: visual analysis, suitability mapping, PBL situation, didactical variables.

Lors de la résolution d'un problème spatial à travers une réalisation cartographique, on ne peut pas séparer le traitement de la visualisation des données (Dumolard, 2006). La carte intervient régulièrement lors des étapes intermédiaires du processus de résolution du problème : elle fait l'objet d'analyses visuelles systématiques. Force est de constater que ces étapes d'analyse visuelle de la carte échappent régulièrement aux étudiants, notamment lorsque ces derniers ne sont pas sensibilisés à l'information géographique. Cette étude tend à répondre à la question : comment amener des étudiants non cartographes à effectuer des analyses visuelles efficaces sur leurs productions cartographiques? En nous appuyant sur les outils théoriques issus de la didactique des mathématiques et des sciences, nous montrons en quoi la modélisation d'une « situation-problème » et la spécification de variables didactiques favorisent les stratégies d'analyse visuelle.

1. Des outils de la Didactique Des Sciences pour inciter le contrôle visuel

La Didactique des Sciences nous offre des outils pour structurer et réfléchir à la conception du problème. Plus particulièrement, la Théorie Des Situations (TDS) de G. Brousseau (1975), fondée sur une conception socioconstructiviste de l'apprentissage, précise que dans la situation didactique l'enseignant conçoit une **situation-problème** qui doit permettre aux élèves d'acquérir une connaissance nouvelle. Selon M. Fabre (Fabre, 1999) et JP. Astolfi (Astolfi, 1992), on distingue cinq caractéristiques principales à un problème : 1) il doit être d'une relative complexité, 2) mettre en jeu plusieurs compétences, 3) être tel que la solution n'est pas immédiatement disponible, 4) exiger de la part de l'étudiant mobilisation et initiative et 5) se fonder sur une difficulté objective concernant le savoir à construire. Par ailleurs, les apprenants doivent avoir un moyen de contrôler eux-mêmes leurs résultats par des interactions avec le milieu didactique (logiciel SIG et données). Les **variables** sont des paramètres de la situation - problème (Brousseau, 1975), elles sont **didactiques** dans la mesure où en agissant sur elles, on pourra provoquer des adaptations et des régulations (Piaget, 1937) : des apprentissages. Celles-ci peuvent prendre plusieurs « valeurs » dont la variation est susceptible de modifier le processus de résolution que les apprenants vont adopter.

2. Poser le bon problème

Nous nous plaçons dans le contexte d'un problème de recherche de site résolu avec des méthodes d'analyse spatiale multicritère ASMC (Malczewski, 2006) pour la création de cartes d'aptitude. Il s'agit de cartes de synthèse qui superposent plusieurs couches d'information géographique, correspondant chacune à critère. Chaque couche fait l'objet d'une classification, et se voit attribuée un poids différent selon son importance. Une fonction de ne somme pondérée agrège les différentes couches.

Le problème posé aux étudiants consiste à réaliser une carte d'aptitude permettant d'identifier des sites potentiels pour l'implantation d'une station de ski. Dans ce contexte, les variables didactiques prédominantes sont le type de milieu, les contraintes environnementales et géographiques (type de structure des pistes et des remontées mécaniques) et les facteurs socio-économiques (type de public visé). D'autres variables, liées aux méthodes et outils de résolution, tels que la fonction du SIG pour effectuer les calculs de superposition, le format des données en entrée, le but de l'analyse ont été définies. L'ensemble des variables agissent sur la résolution du problème à différents niveaux : la définition des critères, la méthode de calcul, l'analyse visuelle pour choisir le site.

Nous avons fixé les valeurs de trois variables prédominantes pour créer une première situation (A) qui s'avère ne pas stimuler les stratégies d'analyse visuelle chez les étudiants. Nous avons ensuite créé une deuxième situation (B) avec des valeurs plus ajustées à la spécificité géomorphologique de la région (zone alpine,

massifs cristallins très escarpés et hauts plateaux en altitude) afin d'inciter une lecture plus sélective.

2a. Situation A : les valeurs des variables didactiques sont pour le type de milieu : un domaine skiable en haute altitude ; pour le type de structure : des pistes larges et très faciles ; pour le type de public : familial.

Exemple de résolution d'étudiants

Critères définis :

Altitude	> 2000 m
Inclinaison de pentes	15°-25°
Végétation	clairsemée
Hydrographie	pas de ruissellement

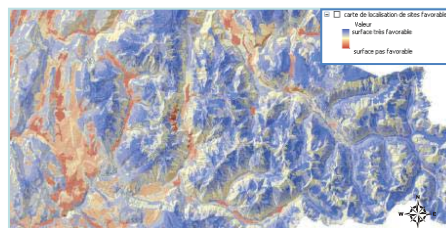


Figure1. Exemple de carte d'aptitude

Nous avons analysé les stratégies de contrôle visuel des étudiants lors de la réalisation de cette carte (Balzarini, *et al.* 2011). En nous appuyant sur la littérature relative aux variables visuelles (Bertin, 1967/83) et aux concepts de fonctionnement perceptuel (Lloyd, 1997 in Fabrikant, 2009) ainsi qu'aux descripteurs des tâches visuelles (MacEachren et M.Kraak, 2001) nous avons identifié trois moments clés de contrôle:

1. un contrôle « structurel » : il s'agit de reconnaître, identifier des objets et des formes, estimer des quantités. Or si la variable visuelle « couleur » (teinte et contraste) est unanimement utilisée par les étudiants, les attributs géométriques - forme et taille de l'objet géographique, correspondant ici à la zone n'ont pas été considérés. Dans cet exemple, les zones les plus favorables, en bleu foncé, sont très étendues : en se limitant à la prédominance de la couleur les étudiants ont cru pouvoir localiser plusieurs sites et répondre au problème.
2. un contrôle « topographique » où il s'agit de reconnaître la nature du terrain (sommets, falaises, vallées...). Les étudiants ont identifié les objets topographiques.
3. un contrôle « analytique » : il s'agit de donner du sens aux zones bleues foncé, d'identifier des corrélations ou des conflits. A' ce stade, les surfaces favorables sont tellement étendues que les étudiants n'ont pas réussi à récapituler et à mettre en relation les zones favorables intégrant les combinaisons des critères et la topographie: ils n'arrivent plus à retrouver, dans le résultat visuel, tout ce qu'ils ont calculé via les techniques d'ASMC. La carte n'a leur pas été pas utile pour décider et justifier leur choix.

2b. Situation B : les valeurs des trois variables didactiques sont : un domaine skiable à moyenne altitude, avec des pistes courtes et raides pour un public de clubs de ski alpin.

Exemple de résolution d'étudiants

Critères définis :

Variable	Valeur
Altitude	< 1500 et 2000 >
Inclinaison des pentes	40° - 50°
Végétation	forêt
Hydrographie	présence de cuvettes

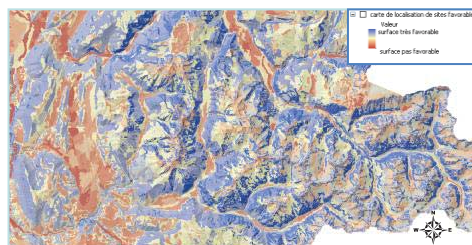


Figure2. Exemple de carte d'aptitude

Du fait de la nature de la région, le résultat de l'ASMC a donné des cartes (Fig.2) où les zones les plus favorables ont des petites surfaces, avec des limites beaucoup plus segmentées et irrégulières. Le contrôle structurel sur cette carte a été plus difficile pour les étudiants, mais ils ont mobilisé plus de stratégies : ils ont recourus aux attributs taille et forme des zones de couler favorable pour identifier des surfaces continues aptes à répondre au problème. Ceci a eu pour conséquence de faciliter le contrôle sémantique : une fois les patterns identifiés, leur connexion au monde réel est confirmée. Les patterns représentant la combinaison des critères sont plus réduits et il a été cognitivement plus facile de découvrir des corrélations ou des conflits (par ex. surfaces favorables sur des versants mal exposés). Cette carte leur a permis de prendre des décisions argumentées.

3. Conclusions

L'objectif de ce poster est d'ouvrir la discussion autour de la pertinence des problèmes posés et des leurs variables didactiques, dans le but de stimuler des stratégies des étudiants, dans un contexte de stratégies d'analyse visuelle. Appliqués à l'analyse cartographique, ces stimuli sont d'autant plus importants lorsque les étudiants sont sans background géographique, mais sont sensés appréhender dans leurs futurs métiers toute la complexité d'un SIG.

Bibliographie

- [AST 92] Astolfi J.-P., *L'école pour apprendre*, ESF, Paris, 1992.
- [BAL 11] Balzarini R., Davoine P.-A., Ney M., « GIS in Geosciences Training: an experience in pedagogical-engineering », *In Proceedings of International Cartographic Conference*, Paris, 2011.
- [BER 67] Bertin J., *Sémiologie graphique*, Paris, Mouton/Gauthier-Villars, 1967.
- [BRO 75] Brousseau G., recueil des travaux dans « *Théorie de situations didactiques* », La Pensée Sauvage, Grenoble, 1998.
- [DUM] Dumolard P., « Les SIG au cœur de la géographie ? La géographie au cœur des SIG ? », *Géographes associés* n° spécial. Savoir penser et partager l'information géographique : les SIG, 30, p. 151-167, 2006.
- [FAB 99] Fabre M., *Situations-problèmes et savoir scolaire*, PUF, 1999.
- [FAB 05] Fabrikant S.I., Goldsberry K., Thematic relevance and perceptual salience of dynamic geovisualization displays. *In Proceedings of 22th International Cartographic Conference*, Coruna Spain, 2005.
- [MAC 11] MacEachren A.M., Kraak M.-J., Research Challenges in Geovisualization. *Cartography and Geographic Information Science*, 28, p. 3-12, 2001.
- [MAL 06] Malczewski J., « GIS-based multi criteria decision analysis: a survey of the literature », *IJGIS*, Vol. 20, N.7, 703-726, 2006.

Comprendre les pressions sur l'environnement à travers des modèles de potentiels anisotropes ?

Joël Boulier*

UMR 8504 Géographie-cités
13 rue Du Four
75006 Paris
joel.boulier@univ-paris1.fr

RESUME : Si la méthode des potentiels en géographie n'est pas tout à fait nouvelle, l'originalité de ce travail repose sur son adaptation pour la prise en compte de l'anisotropie spatiale. Alors que les premiers auteurs s'attachaient à la question de la stabilité des représentations cartographiques lors de changements d'échelle, nous avons tout d'abord adapté la méthode générale pour mettre en évidence des relations amont-aval entre éléments explicatifs (occupation du sol) et importance de dégâts tempétueux en forêt. Dans un second temps, pour modéliser la pression anthropique sur la biodiversité, nous avons adapté la méthode aux espaces anisotropes, tels que celui des déplacements routiers. Nous proposons dans cet article de retracer cette construction méthodologique et de l'illustrer par les résultats thématiques principaux.

ABSTRACT : If potential method is not new in geography, the originality of this work lies in its adaptation to take into account spatial anisotropy. Whereas the first authors were attached to the question of stability of multiscalar cartographic representations, we adapted the general method to highlight relationships between upstream and downstream explanatory factors (land use) and importance of forest damages. In a second step, we tried to model the potential anthropogenic pressure on biodiversity. Then, we adapted the method and tool to anisotropic spaces, such as road transports one. We propose in this paper to trace this methodological way and to draw principal results.

MOTS CLES: interaction spatiale, modélisation, potentiel, anisotropie, environnement, contextes spatiaux, changement d'échelle

KEYWORDS : spatial interaction, modeling, potential, anisotropy, environment, spatial context, multiscalar

Introduction

Tout cartographe s'est heurté tôt ou tard à la question du maillage, de la collecte des données et de l'instabilité des représentations que les changements d'échelle induisent. En partant de ce constat, dès 1991, C. Grasland avait exposé les principes d'une méthode permettant de s'affranchir des maillages et problèmes de frontières grâce l'emploi de fonctions d'interactions spatiales continues, qu'il appliquait ensuite à d'autres données sociodémographiques. [GRA 1991] [GRA 1998]. Par la suite, avec H. Mathian et J.M. Vincent, l'auteur explicitait plus formellement la méthode. [GRA 2000]. Cette méthode générale consiste à décrire un lieu non pas par les caractéristiques explicites qui lui sont propres (population de la commune, nombre d'écoles,...), mais par des descripteurs inhérents à l'environnement spatial qui lui est rattaché (nombre d'écoles dans un rayon de 5 kilomètres,...), sous la forme d'une masse cumulée pondérée par l'inverse de la distance.

A l'instar des données sociodémographiques, la sémantique de l'information géographique environnementale est largement tributaire du maillage de récolte [GEN 2007] [BOU 2007]. Par ailleurs, les données environnementales sont très souvent partielles (échantillonnage souvent incontournable). Enfin, les acteurs agissent à des échelles très diverses et les prises de décision nécessitent des approches multiscalaire et dynamiques. Aussi, la famille des méthodes de *potentiels* nous semble-elle pertinente pour aborder des problèmes d'interactions spatiales dans nombre de questions environnementales. Cependant, il nous paraît également évident que le lissage par une fonction continue indépendante des propriétés directionnelles de l'espace n'est plus très adaptée : l'orientation des relations de type amont-aval ou de type accessibilité routière est une caractéristique fondamentale des phénomènes étudiés. Aussi, l'angle, sous lequel nous avons abordé le problème, est celui de l'anisotropie de l'espace, (telle que définie par Thérèse Saint-Julien [JUL 2009])

Dans un première temps, nous avons conçu des fonctions d'interaction dans un espace où les relations ne seraient prise en compte que selon une orientation privilégiée commune à tous les lieux (calculs de potentiels « *par secteurs* »). Appliquant cette idée aux dégâts tempétueux en forêts¹, nous considérons les fonctions continues d'interaction spatiale, utilisées pour le calcul des *potentiels de stocks*, comme traduction des relations amont-aval qui semblent expliquer la grande diversité des impacts.. Cette prise en compte d'interactions « sectorielles » décrit une première étape vers l'anisotropie [BOU 2003]. A travers un modèle prédictif plus abouti des dégâts potentiels en forêt, nous illustrerons ici cette étape importante qui a posé les véritables jalons de l'innovante méthode des potentiels « anisotropes ».

¹ Les tempêtes de 1999 ont été à l'origine de ce travail

Le second volet, plus original, tend à généraliser cette démarche en proposant une forme très approchée de l'anisotropie de l'espace. L'objectif y est double : construire une nouvelle famille d'indicateurs environnementaux spatialement continus dans un espace anisotrope et ouvrir de nouvelles voies de modélisation des phénomènes environnementaux anisotropes. Le développement de cette méthode est illustré par une première esquisse du rapport entretenu entre pression anthropique et biodiversité en zone d'urbanisation rapide.

Pour une mise en perspective claire de l'enrichissement successif de la méthode des potentiels, nous retracerons rapidement ici les principes initiaux qui ont concouru à l'élaboration de la méthode générale. Puis, nous présenterons la phase intermédiaire des potentiels « par secteurs », jusqu'ici peu valorisée, en nous appuyant sur le cas particulier des dégâts en forêt. Enfin, nous présenterons l'intérêt des calculs de potentiels dans les espaces anisotropes appliqués à la pression anthropique sur la biodiversité en Seine et Marne.

1. Principes et contexte de la méthode générale « des potentiels de stocks » : échapper aux mailles...

La méthode des potentiels de stocks n'est pas nouvelle. En 1991, C. Grasland et H. Mathian jetaient les bases d'une méthode de lissage par potentiels [GRA 1991]. Reprise ensuite de manière formelle par le groupe Hypercarte [GRA 2000], cette méthode initiale s'attachait à décrire de manière continue la distribution spatiale d'un phénomène démographique discret, obtenu à partir de l'information statistique échantillonnée à la maille d'observation.

L'application retenue alors visait à reconstituer une carte de densité à partir d'un ensemble d'individus statistiques dont on connaît la population dans la maille d'observation. La morphologie de la maille dans laquelle les informations individuelles sont collectées et agrégées influencent notablement la perception induite de l'information. Une cartographie des densités à partir des communes puis des zones iris, montre bien l'instabilité cartographique engendrée par la différence de mailles. L'atlas interactif des régions européennes² fournit une belle illustrations de cette instabilité.

Aussi, les auteurs ont-ils proposé de rendre compte pour tout point, non plus les informations agrégées de la maille qu'il représente, mais du *potentiel* de cette information dans son entourage. Cette méthode repose sur la probabilité d'interaction entre deux lieux : un lieu étudié prend alors comme valeur la somme des stocks observés à proximité pondéré par une fonction inverse de la distance avec les points de mesure. « *Tout interagit avec tout, mais deux objets proches ont plus de chance de le faire que deux objets éloignés* » [TOB 1970]. *In fine*, il s'agit de

² <http://aire.ums-riate.fr/fr#content/index.fr.php>

décrire un lieu par les caractéristiques de son environnement spatial plus ou moins proche.

Cette approche nécessite la construction d'un semis de points représentatif des individus étudiés : d'un côté, les individus connus possèdent des valeurs de stocks ; de l'autre, les individus inconnus sont seulement localisés et recevront les valeurs de potentiels de stocks liés à leur proximité avec les points connus.

La contribution d'un point *connu*, *valué*, au potentiel de chaque point de mesure (ou *inconnu*) dépend de plusieurs facteurs :

- de leur éloignement respectif, et donc de la métrique,
- de la portée³ d'étude : elle est considérée comme la « *distance moyenne d'action d'une masse sur son voisinage* » et « *...interprétée comme l'échelle spatiale de représentation choisie* » [PLU 2008-1]. Une portée courte accentue la décroissance des fonctions.
- de la forme de la fonction d'interaction spatiale : elle traduit les différences d'hypothèses concernant les lois de diffusion dans l'espace pour le phénomène étudié. Trois fonctions d'interaction ont été proposées dans ces premiers travaux. La *fonction rectangulaire*, courante dans les logiciels SIG, renvoie la valeur constante unitaire sur l'ensemble de la portée puis devient nulle au-delà. Cette fonction exprime une probabilité d'interaction très forte et très cadrée (telle qu'on pourrait la définir par exemple pour des zones à usage restrictif et mesures contraignantes : protection des captages,...). La *fonction dite « parétienne modifiée »* offre une décroissance très rapide et répond plutôt à des hypothèses de type interactions locales. Enfin, la *fonction dite « exponentielle modifiée »* décroît progressivement et correspond à des configurations spatiales où la distance représente un moindre frein.

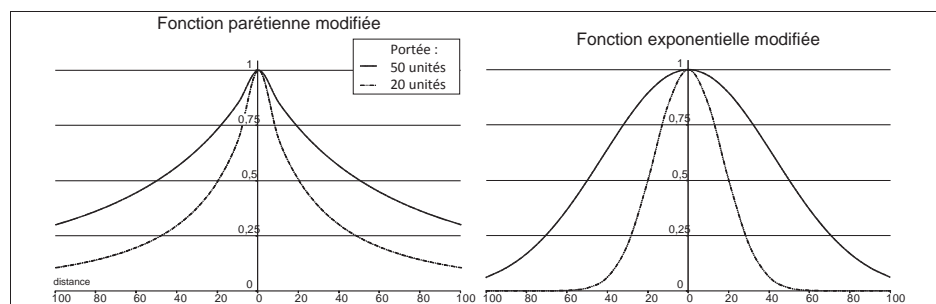


Figure 1. Influence des types de fonction d'interaction spatiale et portées sur la probabilité d'interaction

³ distance à partir de laquelle la valeur de la fonction d'interaction spatiale passe en dessous de 0,5.

Comme le montre la figure 1, le niveau d'interaction spatiale induit est très sensible à la combinaison type de fonction et portée.

Cette méthode générale permet des analyses multi-scalaires dynamiques intéressantes puisque pouvant opérer sur des mailles de collecte morphologiquement très différentes. L'exemple du découpage communal transfrontalier franco-belge est un exemple bien connu maintenant de l'application des lissages cartographiques par potentiels. [RUF 2011].

2. Modéliser la sensibilité des forêts au risque tempétueux : prise en compte des potentiels du contexte spatial sectoriel.

2.1. Un contexte thématique favorable.

Initialement, cette méthode de lissage répondait à une problématique cartographique. En la détournant de son objectif initial, nous avons cherché à construire une information spatialisée qui puisse à la fois traduire le concept de « contexte spatial » et être confrontée à d'autres indicateurs environnementaux issus de mailles et d'échelles hétérogènes. La thématique des tempêtes en forêt nous a offert un support de réflexion très favorable.

En effet, en 1999, les interrogations induites par les tempêtes qui ont ravagé les forêts de l'ouest européen [DRO 2000] n'ont guère trouvé de réponses satisfaisantes dans les nombreuses études du niveau local (système racinaire, station forestière,...). Comme alors répondre à la question restée en suspens : « *comment se préparer aux tempêtes ?* » [DRO 2000]. Comment comprendre la diversité des dégâts ?

Pour nous, il s'agit donc de construire un modèle prédictif mobilisant les dimensions spatiales du phénomène, dimensions souvent absentes des études. Mais, puisque l'échelon local seul ne suffit pas pour exprimer la sensibilité différenciée des peuplements, il devient indispensable de s'affranchir à la fois de l'échelle d'observation des dommages et des différentes mailles forestières (unités de peuplement, parcelles forestières,...).

Nous cherchons à modéliser l'importance des dégâts en fonction de critères spatiaux. Or, l'espace étudié est homogène tant du point de vue de la gestion et que de celui des caractères forestiers (essences, topographie,...). Cependant, la réaction des peuplements face à la tempête est très différenciée, et ne s'explique guère par les caractéristiques locales du lieu. Il faut donc en chercher une interprétation dans le contexte spatial de chacun de ces lieux et dans les interactions qui s'y nichent. La méthode des potentiels offre une perspective évidemment intéressante : traduire les interactions spatiales et conjuguer des informations issues de mailles très hétérogènes.

Cependant, le vent, lors de ces épisodes tempétueux, s'est montré très orienté et d'une puissance avec laquelle peu de peuplements forestiers devraient être à l'abri.

Or, la diversité des dégâts est très importante au sein même de peuplements homogènes. Ce constat conduit à une question triviale : quelles perturbations du flux d'air ont pu œuvrer avant chaque zone de dégâts et créer des conditions favorables à cette hétérogénéité de réactions ?

Pour tenter d'y répondre, nous supposons alors que des interactions spatiales s'expriment à travers une relation amont-aval forte entre les lieux de dégâts et l'espace parcouru par le flux d'air, et que les effets sont cumulatifs le long du parcours. Les potentiels de stock permettraient de mesurer l'importance de cette relation, continue et décroissante avec la distance :

- la nature de l'occupation du sol influencerait l'impact du vent pendant son trajet
- et les lieux proches du dégât seraient sans doute plus actifs que les lieux plus éloignés.

Or, si la méthode générale des potentiels, isotrope, répercute bien la nature continue et décroissante de l'interaction entre les lieux, en revanche, elle ne peut pas, en l'état, traduire la relation amont-aval, hypothèse primordiale ici. D'où l'émergence d'une méthode « *pseudo-anisotrope* » [BOU 2003], développée autour d'une étude sur le massif forestier du Huttier (Normandie, France) ; et dont nous présenterons ici les résultats les plus aboutis.

2.2 Principes des potentiels « par secteurs » : un premier niveau d'anisotropie ?

A travers cette méthode, il s'agit de décrire tout point de l'espace grâce à une nouvelle famille d'indicateurs potentiels particuliers calculés en fonction d'un secteur situé dans en amont de lui-même : le secteur autour de la direction du vent dominant. (figure 2).

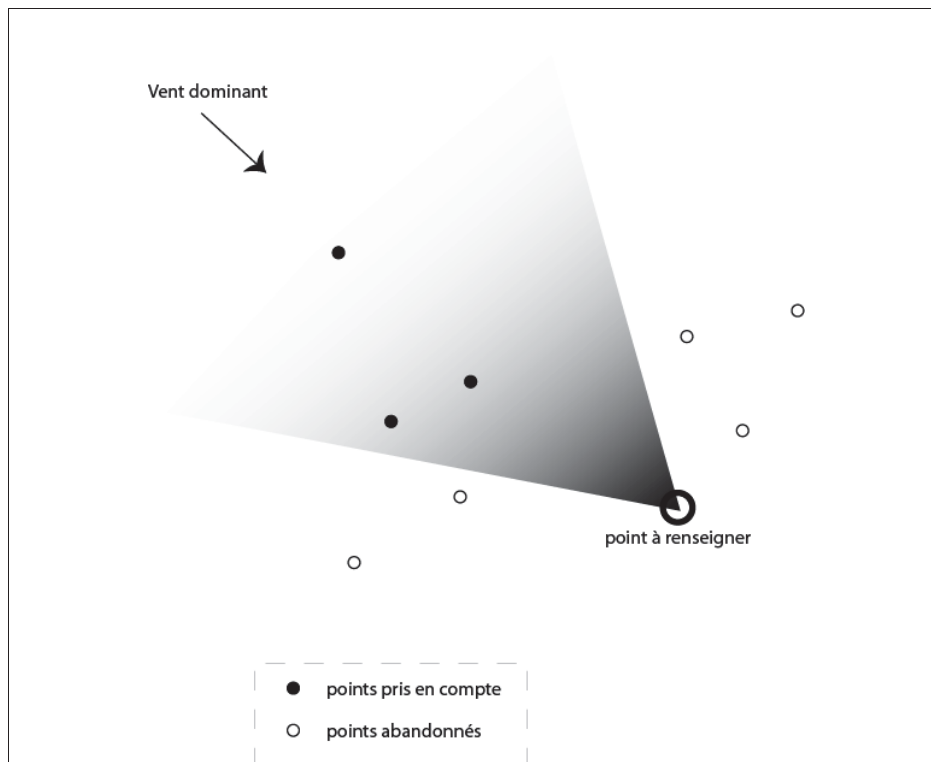


Figure 2. *Adaptation du principe général des potentiels à une orientation privilégiée d'interactions.*

Nous avons posé deux hypothèses importantes dans la mise au point de cette méthode :

- les dégâts sont liés au sens de circulation du vent dominant : nous bâtissons un modèle prédictif en ne considérant que les éléments explicatifs situés en amont de chaque point à expliquer ; les valeurs explicatives contribuant au potentiel de chaque point de dégât seront uniquement celles situées dans le secteur nord-ouest de chacun d'eux,
- la perturbation du flux d'air dominant par la nature des espaces parcourus explique en partie la diversité des dégâts constatés. Les espaces traversés n'ont pas tous la même faculté à écrêter le flux. Or, certaines occupations du sol peuvent être considérées comme étanches et favorisant le risque de turbulences en amont [DRO 2000]⁴. Pour mettre en relation cette notion de « *densité/résistance face au vent* » et

⁴ voir notamment l'article de J. Trouvilliez, pp 304-306

la densité des dégâts constatés, nous construisons plusieurs indicateurs de *densités potentielles* à partir des données forestières disponibles⁵ (densité potentielle de pins, de taillis sous futaie,...). Ces densités potentielles correspondent au ratio de deux potentiels de stocks (potentiel de pins / potentiels de surface par exemple).

2.3. Application aux dégâts dans le massif forestier du Huttier

Plusieurs tests du modèle ont été réalisés et confirment les hypothèses exploratoires initiales : la densité potentielle de dégâts en un lieu est liée aux densités potentielles sectorielles de grandeurs forestières. Le tableau de la figure 3 présente l'intensité des relations entre dégâts et grandeurs forestières « orientées ». La relation entre impact et ouvertures dans les peuplements ou types de peuplements est particulièrement forte. Nous sommes donc en face d'événements fonctionnant sur un espace que nous pourrions qualifier de *quasi-anisotrope*.

	Variables explicatives	Densité potentielle de dégâts
Densités potentielles	Futaie feuillue (toutes directions)	-0,362
	Futaie résineuse (toutes directions)	0,656
	Futaie feuillue (secteur NW)	-0,298
	Lisières (NW)	-0,497
	Taillis sous futaies (NW)	-0,407
	Futaies résineuses (secteur sud est - SE)	0,632
	Futaies feuillues (SE)	-0,218
	Taillis sous futaies (SE)	-0,417
Stocks potentiels	Lisières (total longueurs toutes dir.)	-0,201
	Carrefours (nombre, NW)	0,511

Figure 3. *Corrélations entre densité potentielle de dégâts et variables explicatives*

Par ailleurs, il ressort nettement que la fonction d'interaction spatiale permettant de construire les facteurs explicatifs les plus pertinents soit la fonction de type

⁵ Il s'agit pour l'essentiel, des images aériennes à basse altitude, des cartes de l'I.G.N. et les données de l'Inventaire Forestier National

Pareto. Cette dernière privilégie le poids les éléments les plus proches par rapport à ceux plus lointains ; d'autant plus que la portée est courte (voir figure 1).

L'hypothèse du rôle des perturbations du flux dominant semble confirmée, même si nous nous pensions qu'une fonction d'interaction spatiale plus nuancée serait plus pertinente (fonction exponentielle par exemple).

Au final, l'intensité potentielle des dégâts en un lieu s'explique à 85 % par une combinaison de facteurs explicatifs contextuels. L'importance de la position relative des lieux impactés par rapport aux types d'occupation du sol semble également confirmée. Les cartes de la figure 4 montrent que les écarts d'estimation du modèle. Si faibles soient ils, ces résidus sont très concentrés et dessinent des formes de répartition particulières : le modèle méritera d'être affiné en y introduisant, notamment, des éléments de météorologie locale⁶ ou d'autres informations infra-parcellaires.

⁶ *Notamment en ce qui concerne la mesure des vitesses de vents verticaux ou de directions à l'échelle très fine*

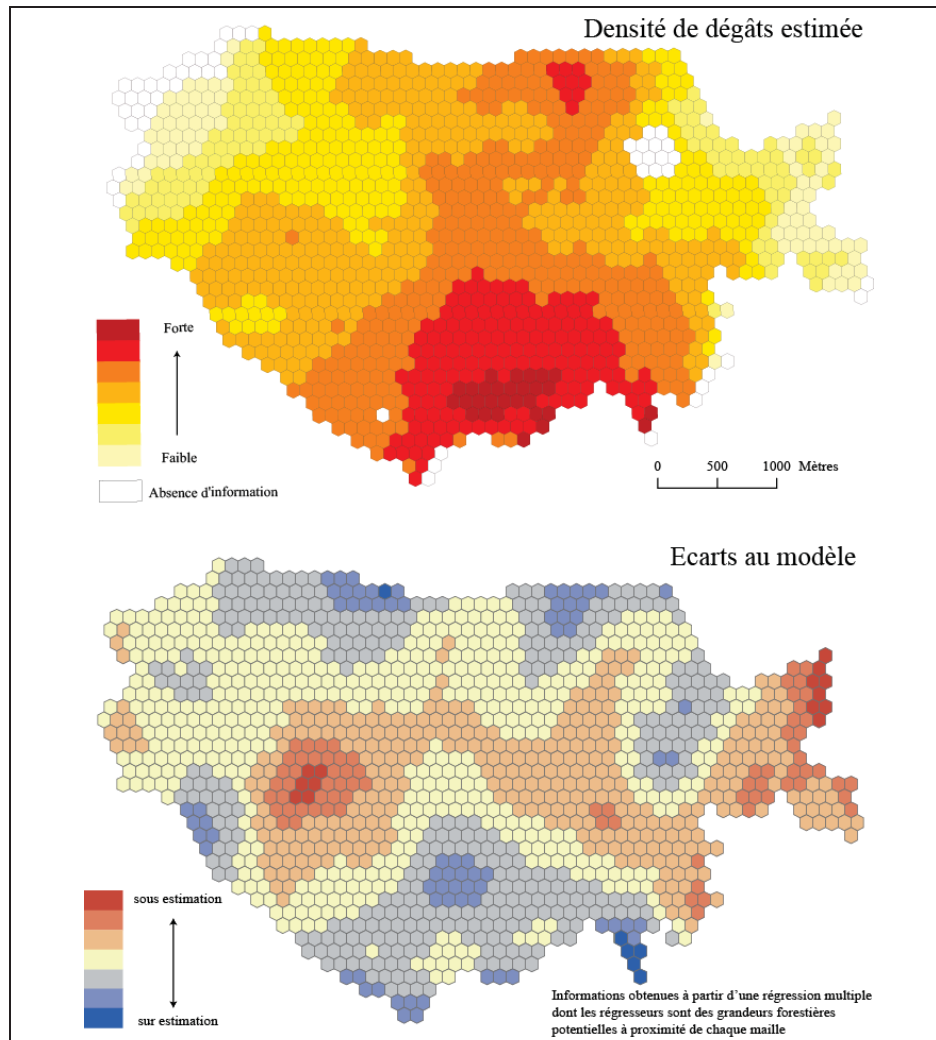


Figure 4. Estimation de la densité de dégâts dans le massif du Huttier (Eure, France)

Ces résultats sont prometteurs. Cependant, ils reposent sur un cadre particulier en dehors duquel la méthode paraît difficilement extrapolable. Si le contexte tempétueux, avec un vent dominant très marqué et stable à l'échelle régionale, autorise une approximation des phénomènes à travers une sectorisation des interactions, il est peu vraisemblable que des résultats aussi probants puissent être obtenus dans des domaines où, soit une échelle plus fine est requise, soit

l'anisotropie réelle de l'espace est un paramètre incontournable : risque hydrique et érosion des sols, pression anthropique sur la biodiversité,...

3. Prise en compte de d'un niveau d'anisotropie élevé de l'espace dans la méthode des potentiels : quelle pression sur la biodiversité ?

La voie vers l'anisotropie ouverte par les calculs de potentiels « sectoriels trouve assez naturellement sa place dans les domaines environnementaux où les notions de cumul de masses et d'interactions spatiales non isotropes ont du sens : circulation des flux dans des bassins versants, propagation des feux de forêts, pression anthropique potentielle sur la biodiversité...C'est en particulier une opportunité de recherche sur ce dernier thème qui nous a incité à reprendre les principes généraux des potentiels et à dépasser la rigidité de la version « *calculs sectoriels* ».

3.1. Accessibilité à la biodiversité et pression potentielle.

La question initiale est de mettre en évidence les enjeux et risques relatifs à la biodiversité, dans un espace à très forte progression urbaine : la Seine et Marne (France). Un certain nombre de décisions publiques visent à protéger l'environnement et atténuer les effets de l'expansion de la ville et des pressions qu'elle engendre. Mais, derrière cette panoplie de dispositions, se pose toujours la question de l'adéquation entre protection et risque réellement encouru par le site estimé comme sensible. La difficulté est donc d'évaluer ce risque. Nous ne trancherons pas cette question ici, cependant, notre démarche cherche à apporter un éclairage méthodologique novateur.

En s'attachant au risque encouru du fait d'une sur-fréquentation des lieux, nous nous interrogeons sur la probabilité de fréquentation potentielle de la population de l'espace seine et marnais. Dans cette première approche, il s'agira simplement de mettre en relation, pour l'ensemble du département, les densités potentielles de population en tout point de l'espace⁷ et les zones labélisées avec mesure de protection (ZNIEFF⁸, etc.).

Traduire sous forme de densités potentielles les déplacements possibles de populations résidentes paraît assez sensé : il s'agit bien d'interactions spatiales entre lieux de résidence et zones de biodiversité et la taille des mailles de collecte relativise les masses en jeu. Par ailleurs, l'attractivité de la « nature » peut être forte mais elle est contrainte par les temps de déplacement qui peuvent représenter un frein important pour les habitants (particulièrement en Île-de-France). Mais plus

⁷ Nous approximons la continuité de l'espace par la construction d'un semis régulier de points

⁸ Voir en particulier le site de la DRIEE : <http://www.driee.ile-de-france.developpement-durable.gouv.fr/les-inventaires-des-znieff-r372.html>

encore, l'éloignement croissant des habitants rend la probabilité de déplacement de moins en moins forte.

3.2. Construction des potentiels anisotropes et réseau de transport

Face à ces considérations générales, les méthodes de calculs de potentiels semblent pouvoir offrir une réponse adaptée pour la construction d'un indicateur de « pression anthropique ». Cependant, l'espace des transports est fortement anisotrope : *« le système de transport produit et renforce l'anisotropie de l'espace ; il permet d'atténuer le frein de la distance (rugosité de l'espace). Donc les autres directions deviennent relativement plus rugueuses que les nouvelles lignes de transport. »* (G. Pini)⁹.

Aussi, contrairement à l'espace forestier précédemment décrit, il serait très difficile de résumer ses propriétés directionnelles par une orientation sectorielle unique. Nous sommes placés ici dans un cadre où les interactions spatiales sont très contingentées par la présence et l'efficacité des réseaux de transport. Aussi bien la durée des trajets que leur orientation forcément particulière rendent la « sectorisation » inadaptée ; comme l'exprime si clairement D. Bariotti : *« l'anisotropie, quant à elle, exprime l'idée que la portée du voisinage d'une unité spatiale n'est pas forcément régulière, ni dans sa taille ni dans sa forme, puisqu'elle dépend de l'accessibilité de cette unité aux autres unités environnantes. »* [BAR 2006]. Il faut donc favoriser la pleine dimension anisotrope dans ces méthodes.

C'est pourquoi, nous avons redéveloppé la méthode et le logiciel Digitaline32¹⁰, afin de décrire chaque relation entre les lieux de manière non régulière, tant pour sa direction que pour son « coût » (métrique).

Dans l'application présentée ici, la relation entre zones de peuplement et zones écologiques est estimée grâce à la modélisation du réseau de transport. Le coût d'accès est exprimé en durée de déplacement. Dans les calculs de potentiels alors mis en œuvre, ce coût de transport, cette durée, sert également à fixer le seuil au delà duquel l'interaction entre deux points est nulle et à fixer la portée.

Concrètement, nous mettons en relation, dans un périmètre de 50 kilomètres autour de la Seine et Marne¹¹ :

– un semis régulier de points *inconnus*, où chaque point représente le centre d'un hexagone¹² et se situe à une distance de 1 km de tout autre point.

⁹ Propos cité dans un résumé de son cours en ligne sur la Géographie des transports : <http://www.la-zone.ch/wp-content/uploads/GdT-I-r%C3%A9sum%C3%A9-du-cours-2011-Olivier-Brand.pdf>

¹⁰ Digitaline32 est un logiciel développé par l'auteur et fait partie des rares outils à permettre des calculs de potentiels dans le domaine de la géographie

11. 50 kilomètres afin de s'affranchir des effets de bordure.

12. Forme favorable à l'équidistance entre voisins les plus proches, dans six directions

- un semis irrégulier des points connus : ce sont les centroïdes des zones Iris (INSEE) pour lesquelles nous disposons d'effectifs de population 2007 (6532 zones représentées par leur centroïde).

Par ailleurs, pour quantifier les flux potentiels de population vers les lieux protégés, nous arrêtons quelques règles simples :

- l'interaction entre semis de points et zones iris traduit plutôt un déplacement de loisir vers les lieux protégés de la biodiversité : nous privilégions donc le transport automobile,

- pour le coût de transport associé au réseau, distance-temps, nous nous inspirons des travaux de M. Appert [APP 2002] et de D. Robert [ROB 2001].

- nous retenons deux scénarii contrastés de circulation : celui d'une circulation très fluide (dimanche matin,...) et celui d'une circulation très dense et encombrée (fin de journée en semaine par exemple).

- s'agissant de déplacements de proximité et de type « loisirs », nous ne regardons que les trajets de moins d'une heure, posant comme préalable qu'au delà, il ne s'agit plus de déplacement ordinaire vers un service banal : ce sera le seuil d'interaction dans notre maquette. Dans le même ordre d'idée, nous fixerons, à titre exploratoire, la portée à 20 minutes, en écho à un déplacement moyen plausible d'une vingtaine de minutes pour accéder à des lieux riches en biodiversité.

Cette démarche globale, décrite par la figure 5, attribue ainsi à chaque point inconnu de l'espace une valeur de densité potentielle de population, selon les paramètres du scénario choisi.

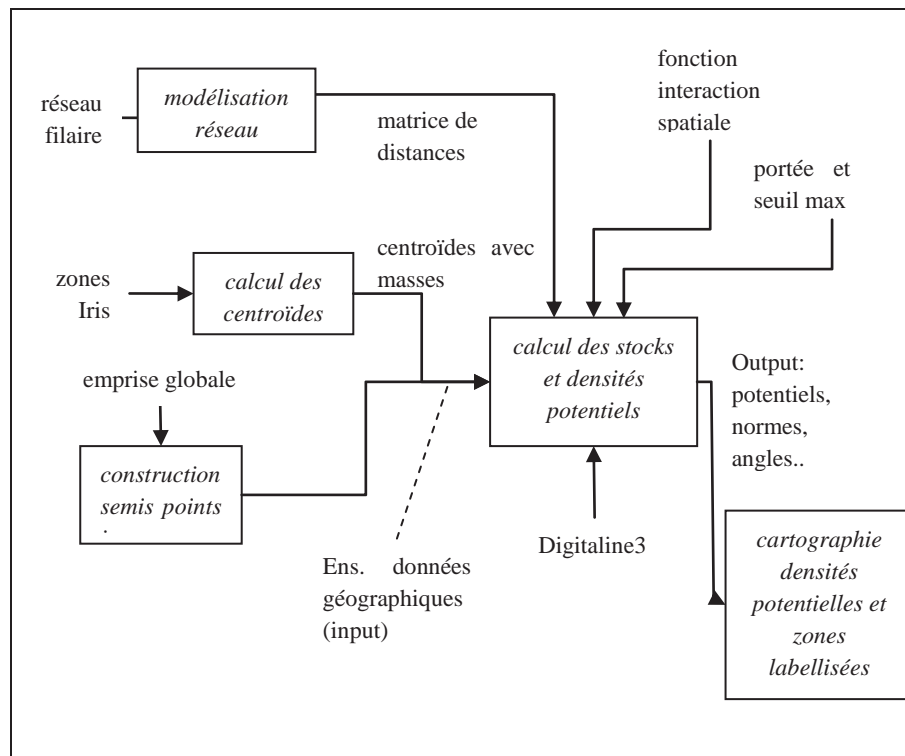


Figure 5. Décomposition fonctionnelle des calculs de densités potentielles de population en fonction du réseau routier.

3.3. Une première vision de la pression anthropique sur les zones protégées.

A partir des scénarii décrits précédemment, les deux premiers blocs-diagramme obtenus (figure 6) représentent les pics et creux de densités potentielles pour l'ensemble du département. Y sont superposées les zones de labellisation¹³.

¹³. et leur niveau de labellisation : certaines zones possèdent jusqu'à 7 mesures de protection simultanées

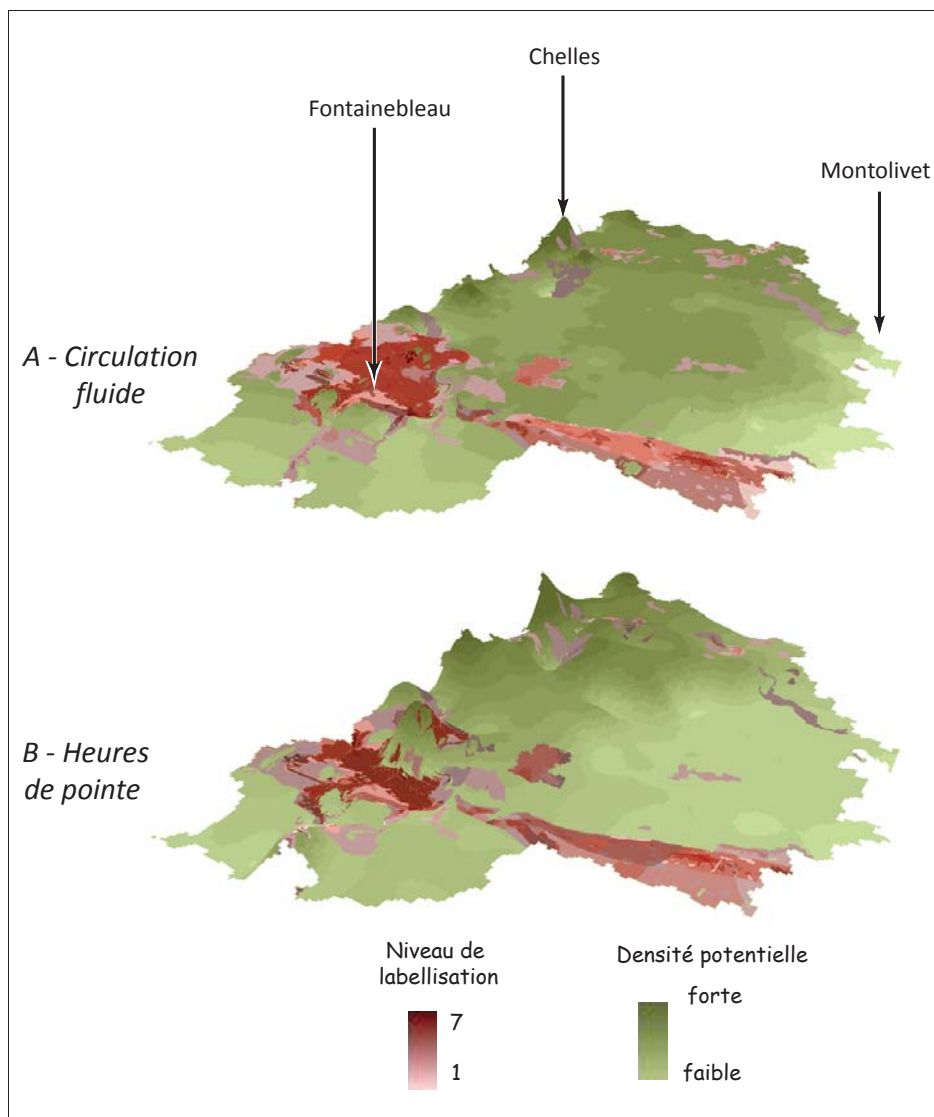


Figure 6 : Protection environnementale face aux densités potentielles de population en Seine et Marne (France) : deux scénarios de circulation...

La difficulté de circulation en Île-de-France aux heures de pointe explique en grande partie la très nette différence de densités potentielles. Ces cartes montrent bien que cette application souligne la légitimité de l'interrogation initiale : les zones protégées, telles que l'emblématique forêt de Fontainebleau, ne se situent pas là où la probabilité de pression anthropique est la plus forte ; quel que soit le scénario de

circulation routière envisagé. De même, on peut s'interroger sur le niveau de labellisation. Dans les deux contextes de circulation, les plus fortes pressions se situent au nord-ouest du département, vers la commune de Chelles, sans surprise...non loin de Paris. Peut-on dire si on a protégé les zones précieuses ou les zones sensibles ? Si un approfondissement de la modélisation générale (mesures de flux réels, enquêtes de comportement, études de fréquentations,...) est indispensable pour répondre sereinement, cette première application des potentiels anisotropes mérite de nourrir le débat ; débat qui reste ouvert...

Conclusion

Les travaux menés pour comprendre la sensibilité des peuplements forestiers tout comme ceux permettant de modéliser une forme de pression anthropique sur la biodiversité ont montré l'intérêt indéniable de la prise en compte des contextes spatiaux sous la forme de « *potentiels* ». En partant d'une méthode générale, nous avons ainsi enrichi la famille de ces méthodes par deux spécialisations qui aboutissent à une anisotropie plus fine et plus complète de l'espace pris en compte.

Les avantages de chacune des méthodes déclinées peuvent se conjuguer au sein d'un système d'information d'aide à la décision car faisant émerger des caractéristiques différentes de l'espace étudié et des interactions qui s'y jouent :

- la méthode initiale, que l'on peut qualifier de générale, fait ressortir les grands traits structurants de l'espace indépendamment de toute anisotropie de l'espace, et s'accorde bien avec les échelles plutôt petites, c'est à dire échelles pour lesquelles les niveaux de décision et de conception demandent souvent une forte généralisation,
- les deux autres méthodes permettent de spécifier des propriétés directionnelles privilégiées dont la signification et l'interprétation sont sensiblement différentes. La méthode *par secteurs* semble plutôt bien adaptée à des phénomènes orientés assez massifs (tempêtes, courants marins,...) alors que la méthode finale, anisotrope, s'accorde beaucoup mieux avec les échelles d'étude plus fines et les phénomènes interactifs plus locaux à forte variabilité directionnelle.

Cette méthode de potentiels anisotropes, balbutiante, ouvre des pistes de réflexion intéressantes et conduit à de nombreuses interrogations : métriques adaptées aux phénomènes à décrire, type des fonctions d'interactions, échantillonnage spatial, niveau d'anisotropie... Les choix conjugués des paramètres du modèle conduisent très vite à une explosion combinatoire. Le rôle du thématique reste primordial : la thématique permet-elle de fixer métrique, valeur des portées et seuils choisis ou type de fonction utilisée ? Une certaine complétude de cette réflexion sera atteinte si chaque individu peut être au final muni de ses propres règles d'interaction avec le monde qui l'entoure, idée chère à M. Takeyama [TAK 1997]. Mais, outre la difficulté pour modéliser ces règles, la question de la puissance de calcul se pose pour la construction des matrices de distances et l'appariement des données, surtout aux échelles fines...

Bibliographie

- [APP 2002] Appert M. & Chapelon L., Planification des transports régionaux en Languedoc-Roussillon et Nord-Pas-de-Calais : évaluation de la concurrence rail-route, analyse comparée des chaînes de transport à dominante routière et ferroviaire, programme de recherche INRETS – Groupement Régional Nord-Pas-de-Calais., 84p., 2002
- [BAR 2006] Badariotti D. et al., « Conception d'un automate cellulaire non stationnaire à base de graphe pour modéliser la structure spatiale urbaine: le modèle Remus », *Cybergeo : European Journal of Geography*, Dossiers, Sélection des meilleurs articles de SAGEO 2006, article 403, mis en ligne le 03 octobre 2007, consulté le 13 septembre 2012
- [BOU 2003] Boulier J., Grasland C., La forêt face au risque tempête : vers d'autres apports de l'analyse spatiale, Colloque « Les rencontres ThéoQuant », Besançon, 2003
- [BOU 2007] Boulier J. et al., " Hand Book for Data Collection " - Datanavigator 2 Part 1, ESPON 3.2 "Spatial scenarios and orientations in relation to the ESDP and EU cohesion policy (2004-2006)" , Luxembourg, 2007
- [BOU 2007] Boulier J. et al., " Experiments on database integration " - Datanavigator 2 Part 2, ESPON 3.2 "Spatial scenarios and orientations in relation to the ESDP and EU cohesion policy (2004-2006)" , Luxembourg, 2007
- [DRO 2000] Drouineau S. et al., « Expertise collective sur les tempêtes, la sensibilité des forêts et sur leur reconstitution », *Le courrier de l'environnement* n°40 INRA-ME&S, Paris, 336 p., 2003
- [GEN 2007] Gensel J., Boulier J. et al., " LTDB: a long term database application", ESPON 3.2 "Spatial scenarios and orientations in relation to the ESDP and EU cohesion policy (2004-2006)" , Luxembourg, 2007
- [GRA 1991] Grasland C., Potentiel de population, interaction spatiale et frontière: des deux Allemagne à l'unification, *L'Espace Géographique*, 3, pp. 243-254, 1991
- [GRA 1998] Grasland C., La composante d'échelle dans l'analyse des distributions spatiales. Application à la fécondité européenne en 1980 et 1988, *Revue Belge de Géographie*, 122ème année, fasc. 4, pp. 435-457, 1998
- [GRA 2000] Grasland C., Mathian H., Vincent J.-M., « Multiscalar Analysis and map generalisation of discrete social phenomena : Statistical problems and political consequences » s, 2000, *Statistical Journal of the United Nations ECE*, 17, pp. 157-188
- [GRA 2001] Grasland C., Lissage cartographique et animation spatio-temporelle : quelques réflexions méthodologiques, *Travaux de l'Institut de Géographie de Reims*, 12 p., 2001
- [JUL 2009] Sant-Julien Th., Isotropie, Hypergeo, Concepts, article 463 mis en ligne le 24 octobre 2009.
- [PIN 2001] Pini, G. 2001, La géographie des transports in *Les concepts de la géographie humaine*. A.Bailly et Al. Paris, Armand Colin

- [PLU 2008-1] Plumejeaud C. et al. , « HyperSmooth : calcul et visualisation de cartes de potentiel interactives » , Colloque International de Géomatique et d'Analyse Spatiale. Numéro spécial de la Revue Nouvelle des Technologies de l'Information. RNTI-E-13, pp 19-42, ed. Cépaduès, 2008
- [PLU 2008-2] Plumejeaud C. et al. , « HyperSmooth, a system for Interactive Spatial Analysis via Potential Maps » , *8th international Symposium on Web and Wireless Geographical Information Systems 2008 - W2GIS 2008*, Shanghai, December 11-12, China, 2008
- [ROB 2001] Robert D., « L'utilisation de la méthode du potentiel pour s'affranchir des contraintes imposées par les découpages géographiques dans l'analyse transport » , PREDIT Recherches stratégiques, Programme « Système d'information » , Paris, 2001
- [RUF 2011] de Ruffray S. et al., « Enjeux des territoires frontaliers à l'échelle nationale - Vers la détermination et la délimitation de pôles transfrontaliers », DATAR, Rapport sur les territoires transfrontaliers. http://www.ums-riate.fr/documents/rapportDATAR_mai2011.pdf
- [TAK 1997] Takeyama M. et Couclelis H., Map dynamics: integrating cellular automata and GIS through Geo-Algebra. *International Journal of Geographical, 1997 Information Science*.
- [TOB 1970] Tobler W., « A computer movie simulating urban growth in the Detroit region ». *Economic Geography* , 46(2): 234-240, 1970

Simuler les évolutions urbaines à l'aide de données géographiques urbaines 3D

Mickaël Brasebin* — Sébastien Mustière* — Julien Perret* —
Christiane Weber**

* IGN, COGIT, Université Paris Est
73 avenue de Paris
94 160 Saint Mandé
France

{prenom.nom}@ign.fr

** Laboratoire LIVE

3 rue de l'Argonne
67083 Strasbourg
France

{prenom.nom}@live-cnrs.unistra.fr

RÉSUMÉ. Comprendre et étudier les évolutions urbaines est important vis-à-vis des conséquences qu'elles peuvent entraîner. Pour appréhender ces phénomènes, la simulation urbaine basée agent est une des méthodes utilisées. Dans ce cadre, la disponibilité croissante des données 3D peut être mise au profit de la simulation afin d'avoir une connaissance plus fine des évolutions de la ville. Nous proposons une approche originale intégrant les données 3D et leur analyse combinées à l'usage des règles d'urbanisme lors des différentes étapes de simulation : la constitution de bases de données spatio-temporelles, la qualification de tissus urbains et la simulation du bâti. Une utilisation potentielle du simulateur pour l'aide à la détermination des paramètres de règles d'urbanisme est proposée.

ABSTRACT. Understanding and studying urban evolutions is important regarding their potential consequences. In order to apprehend such phenomena, agent based urban simulations can be used. In this context, the increasing availability of 3D urban data can be put to use for simulations so as to gain a finer understanding of urban evolutions. We propose a novel approach integrating 3D data and their analysis combined with urban regulation rules during the different steps of the simulation: the creation of spatio-temporal databases, the qualification of the urban fabric and buildings simulation. A potential application of such a simulator for the support of urban rules parameters choice is proposed.

MOTS-CLÉS : simulation, évolutions urbaines, SIG 3D, analyse spatiale, règles d'urbanisme.

KEYWORDS: simulation, urban growth, 3D GIS, spatial analysis, urban regulation

2 SAGEO 2012 Liège.

1. Introduction

La ville est un système dynamique dont les évolutions sont au centre de multiples préoccupations. [BAI 08] estime que chaque année 1% du parc immobilier est construit (environ 300 000 bâtiments) contre 0.1% détruit. Les préoccupations qui accompagnent les évolutions urbaines peuvent notamment être d'ordre sociales (accessibilité à l'emploi et aux équipements), économiques (attractivité du territoire ou du coût des logements) ou environnementales (qualité de l'air, réduction des émissions de CO₂, etc.). Il est important de comprendre ces dynamiques pour pouvoir évaluer leur impact futur afin de potentiellement le maîtriser. Dès lors, de nombreux travaux sont menés pour analyser les évolutions urbaines à différentes temporalités et différentes échelles en fonction des thématiques étudiées.

La simulation urbaine basée agents est une des méthodes employées pour reproduire les mécanismes d'évolution afin de développer ou vérifier la bonne connaissance de ces derniers ou d'en déceler d'autres. Lorsque les résultats de la simulation deviennent satisfaisants, il est possible d'extrapoler ces connaissances pour mettre en œuvre une démarche prospective afin d'étudier l'impact de différents scénarios (réglementaires, aménagements etc.) sur les évolutions de la ville et notamment sur ses évolutions morphologiques. Afin de pouvoir comparer la configuration des villes avant/après simulation, entre différentes temporalités ou entre les différents scénarios, il est nécessaire de définir une métrique permettant de les qualifier notamment à partir d'indicateurs urbains.

Les améliorations et la multiplication des techniques d'acquisitions de données permettent l'obtention de données de granularité de plus en plus fine et notamment des données 3D. Ces données offrent de nouvelles possibilités d'analyses à des échelles différentes de la zone urbaine au quartier. Les acteurs de la ville exploitent divers indicateurs à travers de nombreuses thématiques (thermiciens, aménageurs etc.).

L'objectif de cet article est de présenter une réflexion, menée par ailleurs dans le cadre d'une thèse de doctorat, sur les apports de l'utilisation des règles d'urbanisme dans le cadre de la simulation urbain. Afin de prendre pleinement en compte ces règles, l'utilisation de données 3D est indispensable. Par exemple, pour contraindre la visibilité vis-à-vis d'un monument historique ou d'un espace libre, l'usage de données 2D n'est pas toujours suffisant.

Dans un premier temps, une revue de littérature (*cf.* section 2) sur les différentes approches permettant de simuler l'évolution de la morphologie urbaine (*cf.* section 2.1) et sur les indicateurs morphologiques 3D (*cf.* section 2.2) sera l'occasion de discuter du positionnement de nos travaux (*cf.* section 2.3). Une proposition de méthodologie (*cf.* section 3), s'appuyant sur les apports potentiels des données 3D, sera ensuite présentée à travers les différents aspects de la simulation : constitution de bases de données 3D spatio-temporelles (*cf.* section 3.1), qualification du tissu urbain (*cf.* section 3.2) et simulation de l'évolution du tissu urbain (*cf.* section 3.3). Cette méthodologie sera ensuite discutée à travers les différentes problématiques rencontrées pour sa mise en œuvre (*cf.* section 3.5).

2. Revue de littérature et positionnement

2.1. *Les simulations des évolutions urbaines*

L'intérêt que suscite dans le domaine de la recherche la simulation des évolutions urbaines a permis l'émergence d'une multitude de systèmes notamment basés agents qui offrent des approches méthodologiques très variées. On peut noter tout d'abord différents points de vue dans la définition des agents. [BOU 99] a proposé une distinction entre les approches spatiales (qui simulent les changements d'occupation des sols [WHI 93]) et acteurs (qui simulent l'expansion urbaine du point de vue des acteurs du marché de l'immobilier [DEV 06]). Une manière simple de classer ces différentes approches est la manière dont l'espace est modélisé. On peut par exemple distinguer les modèles cellulaires et les modèles vectoriels.

Les modèles cellulaires, basés sur le principe du jeu de la vie ou du modèle de ségrégation de [SCH 71], consistent en une grille couvrant la zone géographique étudiée, un ensemble d'états caractérisant les cellules de la grille, un ensemble de voisinages d'une cellule, un ensemble de transitions qui détermine le changement d'état d'une cellule en fonction des cellules voisines et une séquence de pas temporels discrets lors desquels les états des cellules se mettent à jour [WHI 00]. Comme ces modèles cellulaires peuvent s'associer à des données de type raster, ils sont souvent utilisés pour modéliser les changements d'occupation des sols, pour en citer quelques uns : [WHI 00] au niveau des Pays-Bas, [DUB 03] avec le modèle SpaCelle au niveau de l'agglomération Rouennaise, [JAN 10] dans le cadre du projet Gigapolis comme application du modèle Sleuth [CLA 97, CLA 98] sur la baie de Chesapeake, etc. Les modèles cellulaires sont très utilisés du fait de la simplicité de leur mise en œuvre. Cependant, [HAM 03] note que les systèmes urbains étant particulièrement complexes, la représentation de ceux-ci dans des modèles cellulaires pour simuler les dynamiques urbaines est limitée et que les systèmes vecteurs agents sont une manière plus réaliste de modéliser le monde réel et ses processus.

Plus récents, les modèles vecteurs pour modéliser les évolutions urbaines sont aussi moins courants. Ils se basent sur les mêmes principes que les modèles cellulaires mais offrent plus de liberté au niveau de la modélisation de l'espace et une plus grande variété de relations de voisinage entre objets (Voronoi, réseaux etc.) [HAM 03, BAD 07]. Parmi les travaux de ce domaine, certains ont été appliqués pour le domaine vidéoludique. Par exemple, [WEB 09] a conçu un simulateur géométrique urbain interactif de villes pour peupler des environnements 3D divers en générant parcelles, rues, utilisation du sol, empreintes et volumes de bâtiments. Le système simule les évolutions en prenant en compte un modèle numérique de terrain, différents centres urbains, une simulation des prix de l'immobilier et du trafic. Des grammaires de formes [MUL 06] et des règles basiques d'urbanisme (hauteur maximale et recul à la route) sont utilisées pour générer le bâti. Cependant, les résultats ne sont pas croisés avec des données réelles, on parle alors de simulation géométrique. [SIL 10] propose une simulation du bâti et la génération d'équipements par rapport à des seuils de population. Le peuplement de l'espace urbain lié à l'influence de la proximité d'équipements et de

4 SAGEO 2012 Liège.

routes. La disposition du bâti est basée sur des lois morphologiques qui tendent à préserver des relations géométriques entre groupes de bâtiments basés sur [PLA 08]. Les bâtiments sont générés à partir d'inférences statistiques sur les propriétés des objets présents (orientation, surface, élongation, etc.). Plus récemment, le projet GeOpenSim [RUA 11] propose une simulation multi-échelle des évolutions urbaines en se concentrant sur les aspects morphologiques celle-ci se fait en fonction de la densité et du type de tissu des îlots (voir [PUI 10] pour la classification). Les îlots sont peuplés, en fonction d'une densité cible de bâti et d'un type de tissu [CUR 10].

Cependant, la distinction entre approche cellulaire et vectoriel n'est pas exclusive, il est possible de bénéficier des avantages des deux approches grâce au couplage de modèles. [ANT 08] fait par exemple cohabiter trois modèles dont un modèle multi-agents pour simuler la localisation et le déplacement de ménages et un modèle cellulaire pour l'occupation du sol.

Par ailleurs, d'après [CRO 10], la troisième dimension pour simuler les évolutions urbaines est surtout utilisée pour l'efficacité de ses capacités immersives afin que le public puisse mieux s'approprier l'espace simulé. Ainsi, on peut noter des travaux dont le but est de proposer une visualisation 3D des résultats de simulations notamment à partir de modèles cellulaires d'occupation du sol [LAM 05] ou de ville [CAN 06] mais également d'empreintes vectorielles de bâtiments [WEB 09]. Ce type de résultat est principalement utilisé dans le cadre de concertations entre planificateurs et grand public pour discuter de différents scénarios d'aménagements et établir de nouvelles propositions [CAN 11, ISA 11].

Concernant les plateformes pour mettre en place ce type de processus, on peut noter l'extension de Repast¹ GeoGraph3D [DIB 04]. Si, ces outils servent avant tout à la représentation, excepté le relief, très peu d'utilisations sont réalisées dans le processus de simulation de données 3D ou d'outils d'analyse spatiale.

2.2. Indicateurs 3D morphologiques

De nombreuses méthodes se basant sur les géométries 3D de bâtiments ont été proposées pour caractériser l'espace urbain. Certains indicateurs concernent parfois les bâtiments eux mêmes, il peut s'agir de mesures de dimensions (volume, aire des surfaces extérieures, hauteur), de modélisation (complexité [GAN 10]) ou caractérisant la forme (compacité [CAR 10], déviation [GRA 08]). Si ces indicateurs concernent un bâtiment seul, il peut être possible de les agréger pour caractériser une zone choisie. Le choix de la zone permettant l'agrégation influe beaucoup sur la signification de l'indicateur calculé, comme on peut le constater avec les mesures de densité du tissu [IAU 05a] à travers ses nombreuses expressions [FOU 97, ADA 10]. Cette zone est nécessaire pour définir certains indicateurs, [ADO 01] a proposé des indicateurs environnementaux basés sur les géométries de modèles urbains simples pour mesurer l'impact du tissu sur la circulation du vent (rugosité et porosité), la transmission de

1. <http://repast.sourceforge.net/>

chaleur (contigüité entre bâti) ou l'évapotranspiration (minéralisation). D'autres indicateurs impliquent les relations entre objets comme la surface des murs adjacents avec d'autres bâtiments, les différences de hauteurs pour une parcelle ou l'angle sous lequel est vu un bâtiment depuis les bordures opposées de la voirie attenante. Une dernière catégorie d'indicateurs caractérise les espaces vides à partir de différents points de l'espace urbain pour mesurer l'ensoleillement [CAR 10], la visibilité d'un objet [FRA 00, RAM 03] ou l'ouverture de ciel [FIS 05] et les méthodes pour analyser les formes de ciel [CAR 10]. Ce dernier indicateur est utilisé aussi bien pour caractériser le ressenti [TEL 01] que pour quantifier l'effet d'îlot de chaleur urbain [SVE 04, UNG 09].

2.3. Positionnement de nos travaux

À notre connaissance aucun simulateur d'évolutions urbaines n'intègre la 3^e dimension dans les phases de simulation excepté lors de la visualisation (à l'exception de [SIL 10] déjà mentionné). Pourtant, l'intégration de la 3^e dimension aux différentes étapes de simulation permet d'étudier des phénomènes, en particulier environnementaux, pour lesquels la 3^e dimension doit être prise en compte. Dans le domaine de l'environnement, la simulation est effectuée manuellement en intégrant de futurs aménagements [BOU 09, ADA 10]. Cette approche est très efficace localement pour des sites choisis dont on connaît les futures configurations potentielles, mais ne peut s'appliquer à l'échelle de la ville.

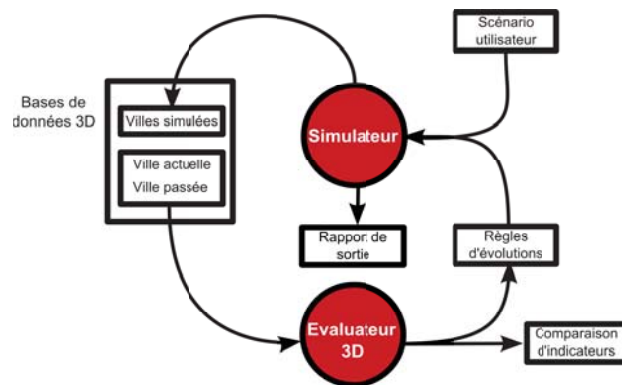


Figure 1 – Schéma de simulation proposé

Nos travaux s'inscrivent dans la conception d'un système de simulation spatial vectoriel 3D basé sur la morphologie du bâti dans lequel la composante 3D est exploitée à travers les différentes étapes de simulation. Nous considérons un système de simulation type (voir figure 1) composé d'une base de données urbaine (2D/3D) à différentes temporalités, qu'il faudra construire (*cf.* section 3.1). Un évaluateur doit permettre de comparer les valeurs d'indicateurs 2D/3D entre des bases de données à

6 SAGEO 2012 Liège.

différentes temporalités pour extraire des règles d'évolution ou entre différentes bases simulées ou réelles (*cf.* section 3.2). Ces bases simulées sont produites à partir de règles d'évolutions et d'un scénario utilisateur (*cf.* section 3.3) qui peuvent renvoyer en sortie des informations relatives à la simulation (*cf.* section 3.4).

Dans le cadre de nos travaux, une série de choix a été effectuée afin de mettre œuvre cette simulation 3D. Tout d'abord, nous avons choisi de nous baser sur la méthodologie de GeOpenSim et d'étudier les évolutions à partir de données topographiques 3D en centrant notre étude sur le bâti. Ce choix est justifié par le fait que les bases de données 3D ne contiennent généralement que le bâti et parce que leur verticalité influe de manière importante sur les configurations spatiales. Le niveau de détail choisi est le LOD2 (suivant la classification de CityGML@[KOL 03]), c'est le niveau de détail le plus courant dans les bases de données 3D. Concernant les règles d'évolution, nous les considérons en deux étapes. **La première étape** se déroule au niveau méso (l'îlot dans notre cas). Lors de celle-ci chaque objet choisit un état cible en fonction du contexte local et de méthodes de décision vers lequel la simulation les fait tendre. Cet état est décrit à travers un type de tissu et une densité bâtie. **La seconde étape** consiste à peupler les objets méso par des objets micro (par exemple, bâtiments dans notre cas) à partir de méthodes de peuplement. Ces méthodes ont pour but de reproduire l'organisation du bâti lors de la densification à partir des contraintes induites par l'état cible de l'objet méso. Dans le cadre de notre travail, nous utiliserons les règles d'urbanisme, notamment issues de PLU (ou de POS pour les dates les plus anciennes), pour réguler le peuplement. Le schéma figure 2 illustre les différentes étapes de la simulation de densification d'un îlot.

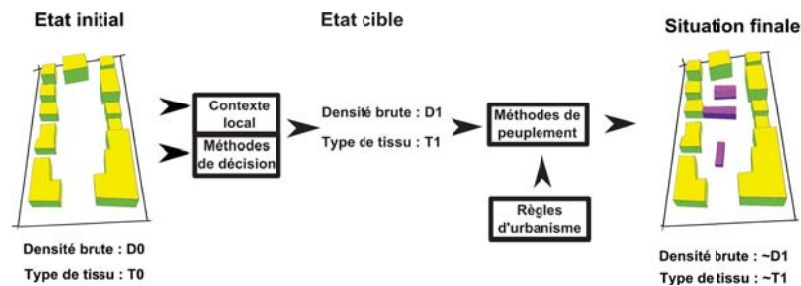


Figure 2 – Illustration de l'utilisation des méthodes de décision et de peuplement lors d'une simulation

Deux cas d'utilisation du simulateurs sont considérés :

– *Proposition de scénarios* : l'utilisateur fournit un scénario composé d'un ensemble de règles d'urbanisme et souhaite étudier les variations de valeurs d'un indicateur. Par exemple, il pourrait souhaiter mesurer la variation de la densité dans un quartier donné en fonction des règles de prospect ou de recul à la route. Ce type d'utilisation est l'occasion de simuler le rôle régulateur des règles dans le développement du tissu,

– *Génération de scénarios* : comme dans le premier type d'utilisation un ensemble presque infini de scénarios peut être défini, on propose ici d'inverser le problème et de générer un ensemble de valeurs paramètres de règles pour atteindre des valeurs souhaitées d'indicateurs (*cf.* section 3.4).

3. Propositions pour étudier les évolutions urbaines en 3D

3.1. Bases de données spatio-temporelles 3D

Les technologies classiques permettant de saisir en 3D la géométrie de bâtiments étant très récentes, il n'existe presque aucune base de données topographique 3D complète antérieure aux années 2000 (on pourra par exemple penser aux données TRAPU®[BOL 01]). On peut noter cependant de nombreux travaux dans le domaine de l'archéologie où l'on tente de modéliser des bâtiments en 3D à partir de diverses sources (descriptions, peintures, photographies anciennes, voir par exemple [ELH 01, STE 10]) mais cela ne permet généralement pas de couvrir qu'une petite zone géographique. Une phase de constitution d'une base de données 3D à des dates antérieures est donc nécessaire pour étudier les temporalités passées du point de vue de la 3^e dimension. Cependant, il est beaucoup plus aisé de constituer une base de données 2D en saisissant l'emprise de bâtiments à partir d'orthophotos ou de cartographies [RAV 08, RUA 11], qui sont des documents couvrant la totalité du territoire depuis au moins un siècle. Nous supposons qu'il est possible de s'appuyer sur de telles connaissances pour produire une base de données spatio-temporelle 3D. De telles bases de données spatio-temporelles sont en particulier intéressantes lorsqu'il s'agit de comprendre les comportements de construction à différentes périodes vis à vis des règles d'urbanisme. On pourra s'attacher à étudier leur respect, mais aussi la stratégie de construction, par exemple pour déterminer s'il s'agit en général d'une optimisation du volume (ou de la surface de plancher) construit par rapport aux contraintes réglementaires. À l'inverse, sans de telles bases, nous pourrions tester différentes hypothèses de simulation quant à cette stratégie de construction. Par ailleurs, nous n'avons aucune certitude à l'heure actuelle sur la continuité de telles stratégies. En particulier, il est possible que la sensibilisation aux enjeux environnementaux et au développement durable marque une rupture dans les habitudes de construction.

Notre méthode consiste à apparier l'emprise au sol d'une base de données 3D actuelle avec les emprises des bâtiments en 2D en essayant pour chaque bâtiment d'apparier des dates les plus récentes vers les dates les plus anciennes. L'opération est répétée jusqu'à ce que l'emprise n'existe plus ou qu'elle soit très modifiée. La difficulté de cette méthode est de réussir l'appariement 2D/3D malgré les différences de spécifications concernant notamment l'agrégation de bâtiments et la précision planimétrique. Cette méthode a deux inconvénients principaux : elle ne permet pas de détecter des modifications verticales (modification du toit ou ajout ou suppression d'étages) et ne permet pas de connaître la morphologie des bâtiments détruits sans l'apport d'informations complémentaires.

8 SAGEO 2012 Liège.

La figure 3 illustre des résultats sur un îlot de petite taille.

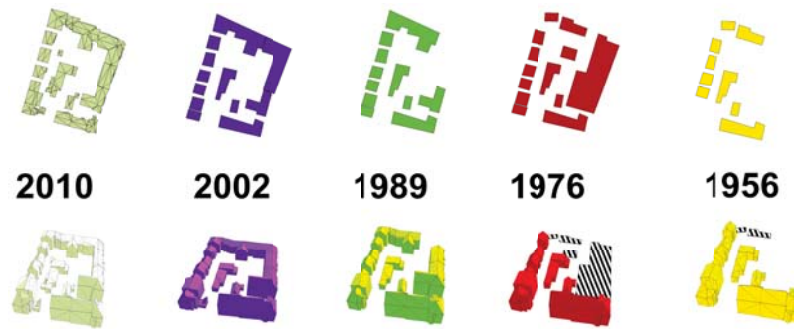


Figure 3 – Base de données temporelles 3D en partant de l’année 2010. Les bâtiments détruits avant cette date apparaissent sous la forme d’une empreinte hachurée.

Différentes méthodes peuvent être utilisées afin de reconstruire les bâtiments manquants :

- *Règles d’urbanisme* : dans ce cas, on génère la 3D en utilisant les informations contenues dans les POS et les PLU, en supposant que les règles sont bien respectées. On se retrouve dans la configuration où l’on souhaiterait peupler une zone en bâtiments avec une contrainte sur leurs emprises (cf. section 3.3),

- *Inférence statistique* : en mesurant des indicateurs morphologiques liés au bâtiment et grâce à des informations contextuelles (type de tissu urbain et période de construction), il s’agira d’inférer la morphologie 3D décrite à travers les indicateurs portant sur un bâtiment notamment la hauteur et la forme des toits. L’avantage de cette méthode est de ne nécessiter aucune connaissance a priori,

- *Données aériennes anciennes* : la reconstruction des bâtiments peut aussi être effectuée à partir de données aériennes du type BD ORTHO@Historique pour lesquelles la pose est estimée et des algorithmes de reconstruction 3D peuvent être utilisés [HAA 10],

- *Connaissances d’expert* : permettre à partir de connaissances sur des bâtiments particuliers (à partir de photos, témoignages, relevés topographiques, archives, permis de construire, base de données externe -comme la Base d’Informations Économiques Notariales-, etc.) de générer la troisième dimension.

3.2. Évaluer le tissu urbain le tissu urbain et ses évolutions

L’évaluation du tissu urbain intervient dans deux phases distinctes de notre processus : lors de l’extraction des méthodes de décision et lors de l’évaluation de bases de données simulées ou non. Cependant, ces deux phases sont semblables car il faut pouvoir comparer des états de bases de données différentes.

La phase d'extraction des méthodes de décision consiste à étudier les densité et les types de tissus à différentes dates et à extraire les changements de densités ou de type de tissu pour une zone donnée. Si la méthodologie d'étude des changements reste la même qu'en 2D, on peut s'attendre à ce que la modélisation 3D permette d'améliorer les résultats en termes de qualification du tissu. Par exemple, au niveau de la densité, la meilleure connaissance de la hauteur d'un bâtiment ou de la forme de son toit permette de mieux déterminer les surfaces de plancher. De plus, il sera intéressant de voir comment la classification des tissus proposée par [PUI 10] est modifiée ou améliorée en intégrant les indicateurs relatifs au 3D. Intuitivement, nous pouvons penser que la forme des toits pourra avoir un impact sur les résultats de classification. De plus, les calculs de densité effectués dans ce travail sont surfaciques, effectuer une classification à l'aide d'une densité volumique, mesure utilisable pour catégoriser les tissus urbains [IAU 05b], pourrait améliorer la classification.

Quant à la comparaison entre différentes bases de données simulées, elle sera l'occasion de faire intervenir des indicateurs et notamment ceux liés à l'environnement pour étudier comment ils évoluent à travers différents scénarios. Afin de couvrir un ensemble important d'indicateurs possibles et leurs agrégations, un modèle générique permettra d'intégrer des indicateurs en fonction des éléments qu'ils considèrent. Pour adapter la portée des indicateurs, des fonctions d'agrégations et d'échantillonnages sont proposées afin de produire un indicateur d'une portée globale et plus pertinente pour servir d'élément de comparaison. Pour les indicateurs concernant un bâtiment seul, les fonctions d'agrégation seront plutôt simples à mettre en œuvre car elles consisteront à prendre en compte les valeurs de l'indicateur sur la zone donnée (îlot, parcelle ou un polygone quelconque). Cependant, pour les indicateurs concernant un point de l'espace, comme par exemple l'indicateur d'ouverture de ciel, il est également nécessaire de définir des fonctions d'échantillonnage (sous forme d'un semis de points) pour construire des indicateurs plus globaux sur une zone donnée, comme par exemple l'ouverture de ciel moyenne sur un îlot.

3.3. *Peupler un espace en 3D*

Le processus de peuplement que nous considérons est présenté dans la figure 4. Une densité et un tissu cible sont affectés à l'îlot d'après les méthodes de décision. Si le tissu cible est très différent du tissu actuel, les bâtiments ne correspondant pas au tissu cible sont supprimés lors d'une phase préliminaire. Le type de tissu urbain évoluant, certains bâtiments caractéristiques d'un type de tissu particulier ne peuvent assurer de fonction adéquate dans le nouveau type de tissu. Un bâtiment industriel risque d'être détruit lorsque la zone dans laquelle se voit ouverte à une opération d'implantation d'habitats collectifs. Si la densité cible est inférieure à la densité actuelle, le peuplement de l'îlot par de nouveaux bâtiments est déclenché pour tendre vers la densité cible. Dans une première phase, il est nécessaire de sélectionner la parcelle qui accueillera le nouveau bâtiment. Pour cela, les volumes réglementaires en accord avec les règles du POS ou du PLU sont calculés. La prise en compte des

10 SAGEO 2012 Liège.

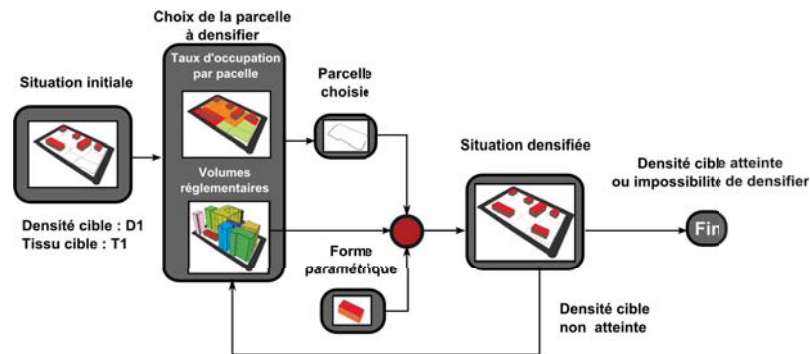


Figure 4 – Schématisation de la méthode de peuplement.

règles d'urbanisme pour générer des solides réglementaires est un sujet très étudié [ELM 87, FAU 01, SHE 10]. La parcelle à densifier est sélectionnée en fonction de critères permettant d'évaluer son attractivité. Outre des critères 2D courants (accessibilité au réseau de transport, distance au centre ville, aménités, etc.), de tels critères peuvent concerner le volume constructible disponible et la densité du bâti existant (les parcelles non ou peu bâties sont privilégiées par rapport aux autres). La seconde phase consiste à générer et placer le bâtiment sur cette parcelle. Le bâtiment proposé est généré à partir d'une bibliothèque de formes paramétriques 3D. Cette bibliothèque paramétrique est créée à partir de statistiques sur les bâtiments construits lors de la période considérée ou déterminée par l'utilisateur. Les paramètres que l'on compte considérer pour générer un bâtiment sont : les emprises des corps du bâtiment, les hauteurs maximales des différents corps et les formes des toits pour les différents corps. Les valeurs de ces paramètres sont contraintes de manière à générer un volume respectant les règles d'urbanisme. Différents travaux ont proposé d'utiliser les volumes réglementaires pour générer des bâtiments en 3D à l'aide d'un système multi-agents [BEL 10] ou à partir d'emprises 2D ou d'une forme 3D [BRA 11]. On peut noter que cette dernière méthode permet de générer des volumes à partir de l'empreinte de bâtiments et ainsi s'appliquer aux bâtiments détruits lors de la constitution de la base de données spatio-temporelle. Enfin, le positionnement du bâtiment sera fonction du respect du volume réglementaire, si il est impossible de placer le bâtiment généré du fait des règles d'urbanisme, celui-ci sera modifié (en jouant sur les différentes variables du modèle paramétré : hauteur, etc.) pour pouvoir être implanté ou une autre parcelle sera sélectionnée pour être densifiée. Le processus continue jusqu'à ce que la densité cible soit atteinte ou qu'il ne soit plus possible de densifier.

3.4. Proposition de règles

Cette étape intervient lorsqu'il s'agit de déterminer les paramètres d'un jeu de règles pour atteindre des valeurs souhaitées d'indicateurs ou en optimiser d'autres.

N'importe quelle métaheuristique peut être utilisée pour résoudre ce problème. Parmi les métaheurstiques utilisables, [KAM 10] propose un algorithme génétique dont le but est d'optimiser le volume du bâti tout en diminuant les pertes d'énergies cumulées dues au contact des surfaces des bâtiments avec l'extérieur. L'aspect génétique intervient lors de la génération des bâtiments, en effet, chaque bâtiment est porteur de gènes qui correspondent à sa hauteur et à la forme de son toit. Nous souhaitons reprendre cette idée de modéliser par un génome la configuration d'une ville, mais en faisant porter les gènes au niveau des règles d'urbanisme dont les valeurs des paramètres seraient ainsi des allèles (différentes versions d'un même gène). Cette technique permettra d'explorer l'espace des paramètres possibles des règles d'urbanisme. Pour mesurer l'adéquation d'une configuration conforme aux souhaits de l'utilisateur, une fonction de satisfaction prendra en compte des indicateurs à optimiser, des indicateurs dont on souhaite atteindre un certain intervalle de valeurs, des règles d'urbanismes fixées et des règles d'urbanisme dont les paramètres sont laissés libres. En retour, le système proposera un ensemble de valeurs pour les paramètres de règles laissés libres qui permettent de générer grâce au simulateur une configuration spatiale permettant d'atteindre les valeurs souhaitées pour les indicateurs définis en entrée. L'outil de simulation est utilisé pour générer des villes qui sont évaluées par la fonction de satisfaction. La figure 5 illustre ce processus à l'aide d'un exemple. L'utilisateur souhaite optimiser la surface habitable d'une zone tout en maintenant un indicateur d'ouverture de ciel élevé. Pour cela, il propose de jouer sur les paramètres d'angle de toit et de prospect tout en prenant en compte une hauteur maximale et un recul par rapport à la route.

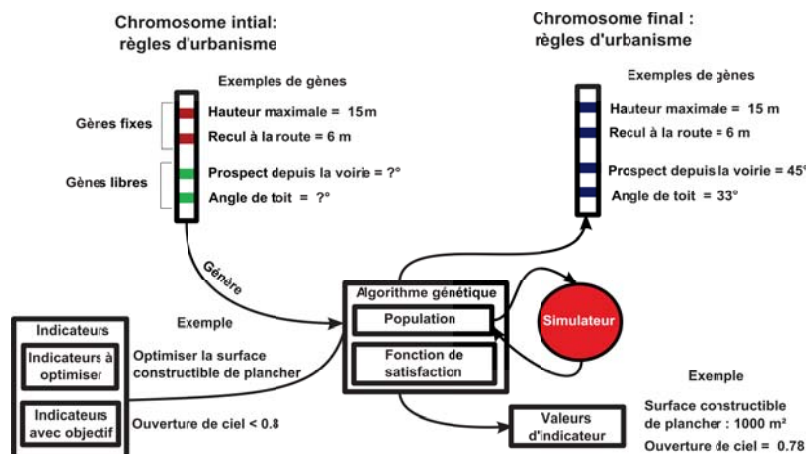


Figure 5 – Utilisation de l'algorithme génétique pour générer des propositions de règles.

12 SAGEO 2012 Liège.

3.5. *Discussions et mise en œuvre*

Nous avons proposé les principes d'un système de simulation d'évolutions urbaines intégrant la 3D. L'implémentation de ce système sera réalisée sur la plate-forme SIG GeOxylene et son module 3D [BRA 09] pour tout ce qui a trait à la modélisation, l'analyse spatiale et la visualisation de données 3D. Le moteur de simulation GeOpenSim sera utilisé pour la modélisation et la classification de données spatio-temporelles ainsi que pour les mécanismes de simulation. Les zones étudiées se situent à Strasbourg où des données spatio-temporelles 2D ont été saisies lors du projet GeOpenSim à cinq temporalités différentes (1956, 1966, 1976, 1989 et 2002). Plusieurs questionnements s'ajoutent autour de cette mise en œuvre. Le premier concerne le niveau de détail des données 3D. Produire des données 3D nécessite un coût beaucoup plus important que pour des données 2D. La conséquence est qu'actuellement seules quelques agglomérations (parmi les plus importantes) sont couvertes par ce genre de bases de données au niveau de détail LOD2 ou LOD3 (en se basant sur la définition de [KOL 03], déjà citée). Quant au reste du territoire français, il est couvert par la BD TOPO® qui fournit une modélisation du bâti au niveau de la gouttière, c'est à dire un niveau de détail LOD1 (2D5). Ainsi, l'influence de la modélisation sur les résultats sera soigneusement étudiée afin d'évaluer la répétabilité et la pertinence de ce processus sur d'autres zones. Des études sur ce sujet sont actuellement en cours. De plus, il est important de fixer le niveau de détail des bâtiments générés par la simulation afin qu'ils correspondent à l'indicateur qui leur sera appliqué. Une autre question est la proposition d'un indice de fiabilité des résultats. En effet, lors des cas d'utilisation proposés, une multiplication d'incertitudes s'inscrit dans le résultat final (incertitudes dues à la modélisation, aux échantillonnages lors de l'agrégation d'indicateurs, erreurs de calcul, incertitudes temporelles, non unicité des solutions de peuplement...), il sera donc nécessaire de fournir des éléments sur la confiance à accorder aux résultats et d'étudier les variations des calculs d'indicateur en resimulant plusieurs fois le même scénario (on parle aussi de sensibilité des résultats aux données et aux modèles).

3.6. *Conclusion*

Dans cet article, nous avons présenté une revue de littérature sur la simulation des évolutions urbaines afin de positionner nos travaux relativement à l'existant. Nous avons proposé une approche pour étudier ces évolutions à l'aide de données 3D en évoquant les différentes possibilités de reconstruction de bases de données spatio-temporelles 3D et l'usage d'indicateurs 3D urbains pour qualifier le tissu. Nous avons également introduit une méthode basée sur l'usage des règles d'urbanisme pour simuler l'implantation de bâtiments et montré comment ce simulateur peut être utilisé pour déterminer l'impact de règles d'urbanisme.

Ce travail de thèse a pour but d'entrevoir de nouvelles possibilités en termes de conception de systèmes basés agents de simulation des évolutions urbaines en 3D. Il s'inscrit dans une démarche globale de réflexion sur l'utilisation des bases de données

spatiales 3D. Il permettra d'ouvrir l'analyse des évolutions urbaines à de nouveaux outils : les indicateurs 3D, en proposant une première utilisation de simulation 3D à travers la régulation par les règles d'urbanisme. Cet article est l'occasion de présenter les réalisations qui seront mises en œuvre lors de cette thèse à travers quatre problématiques : la constitution de bases de données, l'évaluation du tissu urbain, le peuplement en bâtiments d'un espace et la proposition de règles d'urbanisme. L'utilité du système final sera double : étude de l'influence de règles d'urbanisme et aide à la détermination de ces règles à partir d'indicateur. Si, à ce stade de notre réflexion, l'intérêt de la modélisation 3D dans les différentes phases de simulation n'est que théorique, une nécessaire comparaison avec le modèle 2D GeopenSim permettra de faire émerger des avantages à ce type de simulation, on peut notamment penser à une meilleure classification du tissu et une meilleure compréhension des configurations générées. Le système sera l'occasion de se questionner sur le contenu actuel des PLU. Au niveau d'une utilisation appliquée, dans un premier temps, un travail d'échanges avec la communauté urbaine de Strasbourg aura lieu pour échanger sur les possibilités d'utilisation pour la planification. Dans un second temps, on pourrait imaginer l'utilisation d'applications métiers 3D dans les phases d'évaluation du tissu pour permettre aux analystes de la ville de bénéficier de villes 3D vectorielles simulées.

3.7. Remerciements

Nous remercions la Communauté Urbaine de Strasbourg pour la mise à disposition de ses données dans le cadre de la convention ZAEU (Zone Atelier Environnementale Urbaine).

4. Bibliographie

- [ADA 10] ADAM M., « Densité : étude transversale de l'évolution de la forme urbaine d'un quartier de grands ensembles entre arguments environnementaux et perceptions habitantes. », Master's thesis, École Nationale Supérieure d'Architecture de Nantes, École Centrale de Nantes, École des Mines de Nantes et Université de Nantes, août 2010.
- [ADO 01] ADOLPHE L., « A simplified model of urban morphology : application to an analysis of the environmental performance of cities », *Environment and Planning B : Planning and Design*, vol. 28, n° 2, 2001, p. 183–200, Pion Ltd.
- [ANT 08] ANTONI J.-P., FRANKHAUSER P., « Modéliser différentes dynamiques à différentes échelles pour simuler la complexité des mobilités urbaines. L'exemple de MobiSim. », *XLVe colloque de l'ASRDLF*, août 2008.
- [BAD 07] BADARIOTTI D., BANOS A., MORENO D., « Influence of network metrics in urban simulation : introducing accessibility in graph-cellular automata. », *15TH European Colloquium on Theoretical and Quantitative Geography*, 2007.
- [BAI 08] BAIN P., MAUJEAN S., THEYS J., « Agora 2020, Vivre, habiter, se déplacer en 2020 : quelles priorités de recherche ? », rapport, 2008, MEEDDAT/DRAST.

14 SAGEO 2012 Liège.

- [BEL 10] BELKAID A., BEN SACI A., « Système multi-agents pour la fabrication de la forme urbaine », *JFSMA'10*, octobre 2010.
- [BOL 01] BOLDO D., « Fabrication conjointe de Modèles Numériques de Surface et d'ortho-images pour la visualisation perspective de scènes urbaines », PhD thesis, Université René Descartes – Paris V, Paris, France, juin 2001, Interne.
- [BOU 99] BOUSQUET F., GAUTIER D., « Comparaison de deux approches de modélisation des dynamiques spatiales par simulation multi-agents : les approches "spatiale" et "acteurs" », *Cybergeog : European Journal of Geography*, vol. [En ligne], n° 89, 1999.
- [BOU 09] BOUYER J., « Modélisation et simulation des microclimats urbains Etude de l'impact de l'aménagement urbain sur les consommations énergétiques des bâtiments », PhD thesis, Université de Nantes, 2009.
- [BRA 09] BRASEBIN M., « GeOxylene : An Open 3D Framework for the Development of Geographic Applications », *12th International Conference on Geographic Information Science (AGILE)*, juin 2009.
- [BRA 11] BRASEBIN M., PERRET J., HAËCK C., « Un système d'information géographique 3D pour l'exploration des règles d'urbanisme : Application à la constructibilité des bâtiments », *Revue Internationale de Géomatique*, vol. 21, n° 4, 2011, p. 533–556.
- [CAN 06] CANEPARO L., GUERRA F., MASALA E., « UrbanLab - Generative platform for urban and regional design », *proceedings of the 8th International DDSS Conference*, Eindhoven, NL, juillet 2006, p. 235–253.
- [CAN 11] CANEPARO L., « Cities Personal Foresights : A case study in Turin metropolitan area », *My Ideal City. Scenarios for the European City of the 3rd Millennium*, 2011, p. 46–63.
- [CAR 10] CARNEIRO C., MORELLO E., DESTHIEUX G., GOLAY F., « Urban environment quality indicators : application to solar radiation and morphological analysis on built area », *Proceedings of the 3rd WSEAS international conference on Visualization, imaging and simulation*, VIS '10, Stevens Point, Wisconsin, USA, 2010, World Scientific and Engineering Academy and Society (WSEAS), p. 141–148.
- [CLA 97] CLARKE K. C., HOPPEN S., GAYDOS L., « A self-modifying cellular automaton model of historical urbanization in the San Francisco Bay area », *Environment and Planning B : Planning and Design*, vol. 24, n° 2, 1997, p. 247–261, Pion Ltd.
- [CLA 98] CLARKE K. C., GAYDOS L. J., « Loose-coupling a cellular automaton model and GIS : long-term urban growth prediction for San Francisco and Washington/Baltimore », *International Journal of Geographical Information Science*, vol. 12, n° 7, 1998, p. 699–714, Taylor & Francis.
- [CRO 10] CROOKS A., HUDSON-SMITH A., PATEL A., « Building 3D Agent-Based Models for Urban Systems », *CASA Working paper 161*, novembre 2010.
- [CUR 10] CURIE F., PERRET J., RUAS A., « Simulation of urban blocks densification », *13th AGILE International Conference on Geographic Information Science*, mai 2010.
- [DEV 06] DEVISCH, TIMMERMANS H. J. P., ARENTZE T. A., BORGENS A. W. J., « Modeling Residential Search and Location Choice - Framework and Numerical Experiments », *proceedings of the 8th International DDSS Conference*, Eindhoven, NL, juillet 2006, p. 185–200.
- [DIB 04] DIBBLE C., FELDMAN P., « The GeoGraph 3D Computational Laboratory : Network and Terrain Landscapes for RePast », *Journal of Artificial Societies and Social Simulation*,

- vol. 7, n° 1, 2004.
- [DUB 03] DUBOS-PAILLARD E., GUERMOND Y., LANGLOIS P., « Analyse de l'évolution urbaine par automate cellulaire : le modèle SPACELLE », *L'Espace géographique*, vol. 32, 2003, p. 357–378.
- [ELH 01] EL-HAKIM S. F., « A flexible approach to 3D reconstruction from single images », *SIGGRAPH '01 Sketches and Applications*, 2001.
- [ELM 87] EL MAKCHOUNI M., « Un système graphique intelligent d'aide à la conception des plans d'occupation des sols : SYGRIPOS », *12th Urban Data Management Symposium*, 1987.
- [FAU 01] FAUCHER D., « UrbanLab modélisation déclarative des enveloppes urbaines réglementaires », Thèse de doctorat, Université de Nantes, 2001.
- [FIS 05] FISHER-GEWIRTZMAN D., SHACH PINSKY D., WAGNER I. A., BURT M., « View-oriented three-dimensional visual analysis models for the urban environment », *Urban Design International*, vol. 10, n° 1, 2005, p. 23–37, Palgrave Macmillan.
- [FOU 97] FOUCHIER V., *Les densités urbaines et le développement durable : le cas de l'Île-de-France et des villes nouvelles*, Édition du SGVN, Paris, décembre 1997.
- [FRA 00] FRANKLIN W. R., « Applications of Analytical Cartography », *Cartography and Geographic Information Systems*, vol. 27, n° 3, 2000, p. 225–237.
- [GAN 10] GANITSEVA J., COORS V., « Automatic landmark detection for 3D urban models », *5th International Conference on 3D GeoInformation*, novembre 2010.
- [GRA 08] GRABLER F., AGRAWALA M., SUMNER R. W., PAULY M., « Automatic generation of tourist maps », *SIGGRAPH '08 : ACM SIGGRAPH 2008 papers*, New York, NY, USA, 2008, ACM, p. 1–11.
- [HAA 10] HAALA N., KADA M., « An update on automatic 3D building reconstruction », *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 65, n° 6, 2010, p. 570 - 580, ISPRS Centenary Celebration Issue.
- [HAM 03] HAMMAM Y., MOORE A., WHIGHAM P. A., FREEMAN C., « A Vector-Agent Paradigm for Dynamic Urban Modelling », p. 99–104, Citeseer, 2003.
- [IAU 05a] IAURIF, « Appréhender la densité : 2. Les indicateurs de densité », *Note Rapide sur l'occupation du sol*, vol. 383, 2005.
- [IAU 05b] IAURIF, « Appréhender la densité 3. Formes urbaines et densités », *Note Rapide sur l'occupation du sol*, vol. 384, 2005.
- [ISA 11] ISAACS J. P., GILMOUR D. J., BLACKWOOD D. J., FALCONER R. E., « Immersive and non immersive 3D virtual city : decision support tool for urban sustainability », *Journal of Information Technology in Construction*, vol. 16, 2011, p. 151–161.
- [JAN 10] JANTZ C. A., GOETZ S. J., DONATO D., CLAGGETT P., « Designing and implementing a regional urban modeling system using the SLEUTH cellular urban model », *Computers, Environment and Urban Systems*, vol. 34, n° 1, 2010, p. 1–16.
- [KAM 10] KAMPF J. H., MONTAVON M., BUNYESC J., BOLLIGER R., ROBINSON D., « Optimisation of buildings' solar irradiation availability », *Solar Energy*, vol. 84, n° 4, 2010, p. 596–603.
- [KOL 03] KOLBE T. H., GRÖGER G., « Towards unified 3D city models », *Challenges in Geospatial Analysis, Integration and Visualization II. Proc. of Joint ISPRS Workshop*, Stuttgart, 2003.

16 SAGEO 2012 Liège.

- [LAM 05] LAMMEREN R. J. A., MOMOT A., OLDE LOOHUIS R. J. W., HOOGERWERF T. C., « *Trends in real-time landscape visualization and participation* », chapitre 3D visualizations and 2D scenarios, Heidelberg : Wichmann, 2005.
- [MUL 06] MULLER P., WONKA P., HAEGLER S., ULMER A., VAN GOOL L., « Procedural modeling of buildings », *ACM Trans. Graph.*, vol. 25, n° 3, 2006, p. 614–623, ACM.
- [PLA 08] PLAZANET C., « Morphogenèse de la Métropole », *Projet de recherche interdisciplinaire EPFL. Rapport de recherche.*, EPFL, 2008.
- [PUI 10] PUISSANT A., LACHICHE N., BRAUD A., SKUPINSKI G., PERRET J., « Classification des tissus urbains à partir de données vectorielles - application à Strasbourg », *SAGEO 2010*, novembre 2010, p. 198–211.
- [RAM 03] RAMOS F., « Modélisation et Validation d'un Système d'Information Géographique 3D opérationnel », PhD thesis, Université Paris-Est, 2003.
- [RAV 08] RAVEAUX R., BURIE J.-C., OGIER J.-M., « Object Extraction from Colour Cadastral Maps », *DAS 08 Proceedings of the 2008 The Eighth IAPR International Workshop on Document Analysis Systems*, vol. 0, 2008, p. 506—514, IEEE Computer Society.
- [RUA 11] RUAS A., PERRET J., CURIE F., MAS A., PUISSANT A., SKUPINSKI G., BADARIOTTI D., WEBER C., GANCARSKI P., LACHICHE N., LESBEGUERIES J., BRAUD A., « Conception of a GIS-Platform to simulate urban densification based on the analysis of topographic data », *Advances in Cartography and GIScience. Volume 1*, Lecture Notes in Geoinformation and Cartography, p. 413–430, Springer Berlin Heidelberg, 2011.
- [SCH 71] SCHELLING T. C., « Dynamic models of segregation », *Journal of Mathematical Sociology*, vol. 1, n° May, 1971, p. 143–186, Routledge.
- [SHE 10] SHEN Z., KAWAKAMI M., « An online visualization tool for Internet-based local townscape design », *Computers, Environment and Urban Systems*, vol. 34, n° 2, 2010, p. 104–116.
- [SIL 10] SILVA V., « Conception et évaluation d'un prototype de simulation de la morphogenèse urbaine par agents vecteurs multi-échelles », PhD thesis, Ecole Polytechnique Fédérale de Lausanne, octobre 2010.
- [STE 10] STEFANI C., « Maquettes numériques spatio-temporelles d'édifices patrimoniaux : modélisation de la dimension temporelle et multi-restitutions d'édifices », PhD thesis, Arts et Métiers ParisTech, septembre 2010.
- [SVE 04] SVENSSON M. K., « Sky view factor analysis - implications for urban air temperature differences », *Meteorological Applications*, vol. 11, n° 3, 2004, p. 201–211, Cambridge Univ Press.
- [TEL 01] TELLER J., « La régulation morphologique dans le cadre du projet urbain. Spécification d'instruments informatiques destinés à supporter les modes de régulation performantiels », PhD thesis, Université de Liège, avril 2001.
- [UNG 09] UNGER J., « Connection between urban heat island and sky view factor approximated by a software tool on a 3D urban database », *International Journal of Environment and Pollution*, vol. 36, n° 1-2-3, 2009, p. 59–80.
- [WEB 09] WEBER B., MUELLER P., WONKA P., GROSS M., « Interactive Geometric Simulation of 4D Cities », *Computer Graphics Forum*, vol. 28, 2009, p. 481–492.
- [WHI 93] WHITE R., ENGELN G., « Cellular automata and fractal urban form : a cellular modelling approach to the evolution of urban land-use patterns », *Environment and Planning A*, vol. 25, n° 8, 1993, p. 1175–1199, Pion Ltd.

- [WHI 00] WHITE R., « High-resolution integrated modelling of the spatial dynamics of urban and regional systems », *Computers Environment and Urban Systems*, vol. 24, n^o 5, 2000, p. 383–400.

Une fiche descriptive pour l'analyse des relations sémantiques entre textes et cartes

Geoffrey Brun

IGN, COGIT, 73 avenue de Paris, 94160 Saint-Mandé, France
UMR 8504 Géographie-Cités, Université Paris-Diderot Paris 7, Université
Paris 1 Panthéon-Sorbonne, CNRS
geoffrey.brun@ign.fr

RÉSUMÉ. De nombreuses cartes thématiques transmettent de manière inexacte le message que souhaite faire passer leur auteur. Le sujet de thèse présenté ici propose une méthode permettant de créer une carte à partir d'un texte thématique donné. Préalablement à l'élaboration de cette méthode, il est nécessaire d'étudier les relations sémantiques existant entre les textes et les cartes. L'utilisation d'une fiche standardisée permet de caractériser des couples de textes et de cartes regroupés au sein de deux corpus thématiques. Cette fiche s'intéresse à trois éléments : les textes, les images cartographiques et l'habillage de ces images. Cette fiche descriptive permet d'identifier les éléments importants et communs aux textes et aux cartes. Ceux-ci seront ensuite réutilisés lors de l'élaboration de la méthode permettant de passer d'un texte à une représentation cartographique.

ABSTRACT. Many thematic maps provide a misguided message in comparison to the one that their author would like to communicate. The thesis presented in this paper suggests a method allowing the authors to create a map from a given thematic text. Prior to the elaboration of this method, the analysis of semantic links between texts and maps is required. A normalized description sheet will be used for characterizing several "texts and maps" couples gathered into two thematic corpuses. This sheet analyzes three elements: texts, maps and maps wraps. With the help of this sheet, we will find the essential and shared elements between texts and maps. These elements will be then employed for outlining the method that would convert a text into a cartographic representation.

MOTS-CLÉS: cartographie – sémiologie graphique – analyse de documents – ontologie.

KEYWORDS: cartography – graphic semiotics – text-mining – ontology.

1. Introduction

Dans le cadre du service *Cartes à la carte* de l'IGN, j'effectue une thèse visant à fournir un outil d'aide à la conception cartographique à partir de textes. L'objectif de ce travail de recherche est de poser les bases d'une éventuelle méthode permettant de créer une représentation cartographique à partir d'un texte transmis par un commanditaire. L'enjeu est de pouvoir créer à la demande de ce dernier une légende de carte transmettant le mieux possible le message contenu dans son texte afin de s'assurer que la carte qu'il publiera sera conforme à ses attentes.

Pour cela, il est nécessaire de trouver dans le texte des informations thématiques ayant une dimension spatiale, celle-ci garantissant la possible représentation cartographique de ces informations. Les thématiques et sous-thématiques du texte, ainsi que les concepts impliquant la spatialisation de ces thématiques, peuvent être extraits grâce à des outils de Traitement Automatique du Langage [TAL]. Les informations extraites doivent ensuite être utilisées via une ontologie et un moteur de raisonnement pour créer une légende cartographique (Dominguès *et al.*, 2012) à partir de différentes règles issues principalement de la sémiologie graphique (Bertin, 1967). Afin de hiérarchiser les différentes informations à cartographier, l'analyse des informations textuelles reflétant l'intention de l'auteur du texte doit également être effectuée.

2. Une fiche formalisée permettant de décrire un corpus {(textes, cartes)}

Afin d'identifier les informations pertinentes et cartographiables pouvant être extraites d'un texte, il est nécessaire d'étudier préalablement les relations sémantiques existant entre des textes et des cartes publiés conjointement, c'est-à-dire les informations communes présentes dans chaque élément (lieux, thématiques, etc.).

2.1. Constitution du corpus {(textes, cartes)}

La constitution d'un corpus {(texte, cartes)} permet de regrouper des exemples pertinents de cartes et de textes possédant une relation sémantique. Les exemples de couples présentant une faible relation sémantique, voire aucune, ne seront pas sélectionnés. Afin de restreindre l'ampleur de ce travail, deux corpus thématiques seront constitués : l'un aura pour thème l'*énergie*, l'autre portera sur celui des *conflits*.

Le choix des exemples considérés comme pertinents sera guidé principalement par des publications signalées par plusieurs cartographes ayant participé à une série d'entretiens menés au cours des mois de mars et d'avril 2012. Ces entretiens ont permis, entre autres, d'identifier un certain nombre de publications présentant des exemples de couples de cartes et de textes qui ont été pensés comme un tout au

moment de leur conception. Deux autres caractéristiques seront à prendre en compte lors de la constitution du corpus :

- la vocation de la carte jointe au texte : localisation, illustration, synthèse, etc. ;
- la nature de la publication : article de journal, atlas, livre, etc.

2.2. Vocation de la fiche descriptive

Une fois ce corpus constitué, chaque couple (textes, cartes) sera analysé grâce à une fiche descriptive formalisée. Cette fiche a pour vocation de décrire précisément et de manière standardisée les relations sémantiques existant entre cartes et textes en identifiant les informations qui sont présentes à la fois dans l'un et l'autre de ces éléments.

La fiche descriptive s'organise en trois rubriques. La première sert à décrire le texte (ou les textes) : elle permet notamment d'identifier les phrases-clefs et les Entités Nommées Spatiales [ENS] présentes dans le texte (Loustau *et al.*, 2007). Une seconde rubrique formalise la description de l'image cartographique : sont entre autres renseignés l'emprise géographique et le maillage de la carte. Une dernière rubrique s'attache à analyser l'habillage cartographique seul, en particulier le titre et la légende. Il s'agit notamment de caractériser les relations sémantiques (au sens de Bertin) existant entre les différents éléments qui composent la légende.

Une fois renseignées, ces fiches seront analysées afin d'identifier les éléments communs aux différents couples (textes, cartes). Ils seront ensuite utilisés lors de la conception d'une méthode de génération automatique de légende cartographique à partir d'un texte.

3. Bibliographie

- Bertin J., *Sémiologie graphique : les diagrammes, les réseaux, les cartes*, Paris, Editions de l'EHESS (3e éd.), 1967.
- Dominguès C., Christophe S., Jolivet L., « Connaissances opérationnelles pour la conception automatique de légendes de cartes », *Actes IC 2009*, 25-29 mai 2009, Hammamet, p. 253-264.
- Dominguès C., Corby O., Soualah-Alila F., « Raisonner sur une ontologie cartographique pour concevoir des légendes de cartes », *Actes EGC 2012*, 31 janvier - 3février 2012, Bordeaux, p. 1-6.
- Loustau P., Gaio M., Nodenot T., « Des déplacements à l'itinéraire, du syntagme au discours », *Actes SAGEO 2007*, 18-20 juin 2007, Clermont-Ferrand, p. 1-14.
- Mel'čuk I. A., Clas A., Polguère A., *Introduction à la lexicologie explicative et combinatoire*, Louvain-la-Neuve, Éditions Duculot, 1995.

A geomatic approach to Wildland-Urban Interface detection

Champendal A. – Vega Orozco C. – Céré R. – Kanevski M. – Tonini M.

*Institute of Geomatics and Risk Analysis, Amphipôle building
University of Lausanne
1015 Lausanne
Switzerland*

Alexandre.Champendal@unil.ch

CarmenDelia.VegaOrozco@unil.ch

Raphael.Cere@unil.ch

Mikhail.Kanevski@unil.ch

Marj.Tonini@unil.ch

ABSTRACT: *The recent tendency in GIS is to incorporate quantitative methods into integrated systems allowing simultaneously visualising, quantifying and modelling spatio-temporal data. This study aims at integrating different geocomputational methods (i.e. GIS capabilities, Remote Sensing image classification and geospatial advanced algorithms) to analyse the spatio-temporal evolution of the Wildland-Urban Interface (WUI). The study area concerns the alpine valleys of the Canton Ticino (Switzerland) which represents the more prone fire zone in Switzerland. This new approach allowed the definition, parameterisation and analysis of the evolution of WUI in the last three decades (1990, 2000 and 2010).*

RÉSUMÉ: *Récemment la tendance des SIG est d'incorporer des méthodes quantitatives avec des systèmes intégrés permettant simultanément la visualisation, quantification et modélisation des données spatiotemporelles. Cette étude vise à intégrer différentes méthodes de la géomatique (SIG, télédétection et algorithmes avancés d'analyse géospatiale) pour l'analyse de l'évolution spatiotemporelle des Interfaces Forêt/Habitat (IFH). La zone d'étude concerne les vallées alpines du canton Tessin (Suisse) qui représentent la zone suisse la plus sensible aux feux de forêts. Cette nouvelle approche a permis la définition, le paramétrage et l'analyse de l'évolution des interfaces forêt/habitat dans les trois dernières décennies (1990, 2000 et 2010).*

KEY WORDS: *wildland-urban interface, forest fires, city clustering algorithms, multi-temporal image classification, GIS.*

MOTS-CLÉS: *interfaces forêt/habitat, feux de forêts, algorithme des clusters urbains, classification d'image multi-temporelle, SIG.*

1. Introduction

The recent tendency in GIS is to incorporate quantitative methods into integrated systems allowing simultaneously exploring, visualising, quantifying, modelling and analysing spatio-temporal data. In this scenario the term geocomputation has recently been coined. This represents a powerful tool for understanding the structure of environmental phenomena and to investigate complex systems characterised by the interaction of various agents and showing a complex spatio-temporal dynamic.

The objective of the present study is to integrate different geomatic methods (e.g. GIS/Remote Sensing/geocomputational analyses) to analyse the spatio-temporal evolution of a complex zone, namely the Wildland-Urban Interface (WUI). The proposed method allows identifying the local conditions of the population/forest/fires distribution, their interactions and evolution. The study area concerns the alpine valleys of the Canton Ticino that is the more prone fire zone in Switzerland.

2. Wildland-Urban Interface

WUI is a broadly used term in the context of wild and forest fires to indicate areas where houses and other human infrastructures meet or intermingle with wildland vegetation, rural areas or forests. The most relevant problem related to the WUI is the risk of wildland fires induced by anthropogenic actions (Lampin-Maillet *et al.*, 2010). In fact, the increase of urban areas results also in urban sprawl observed in or near forest (Radeloff *et al.*, 2005) and this coexistence enhances both the anthropogenic ignition sources and flammable fuels.

Many researches for WUI detection have been developed; however, no attempts exist for the Alpine environment. The WUI in this complex region exhibits more distinct patterns than those observed in other environments (i.e. Mediterranean area). This is mainly due to the different related fuel types, fire behaviour, socio-economic and environmental conditions. In this respect, the WUI characterization must be assessed through the analysis of the local conditions.

3. Data and methods

Remote sensing images, forest fires, census and cadastral data of the canton Ticino were used for the detection of the WUI in the Swiss alpine environment. The forest fire database of canton Ticino consisted on 2,401 geo-referenced ignition-points in a 40-years period from 1969 to 2008. This information has been collected by the Swiss Forest Service. The census data was provided by the Swiss Federal Statistical Office (FSO) and is organised on a hectometric grid. Multi-temporal Landsat TM imagery (30 m resolution) was considered for land-use classification at different times (1990, 2000 and 2010 years). Methodologically, a supervised

classification method was applied for each image. Subsequently, a detection of the land-cover changes was carried out for mapping and monitoring the evolution of these changes in the last three decades. The main results of this phase are maps portraying the evolution of the forest/low vegetation land and urban space for the considered period.

The analysis and quantification of the WUI highly depend on the definition of “urban zone”. Therefore, the authors suggested a new concept of urban space, elaborated using recently proposed urban/city clustering algorithm U/CCA (Rozenfeld *et al.*, 2008). This algorithm allows the delimitation of urban clusters based on the population distribution. An urban cluster is defined as a continuous area connected by nonzero population cells where the only parameter to be defined is the grid size. The demographic data is normally required to be distributed in a regular grid (100x100 m in our case). In order to overcome this limitation, GIS operations were applied to elaborate a U/CCA based on a non-regular grid.

4. Results and conclusions

The image classification allowed detecting the forest/low vegetation land while the urban area’s detection is affected by confusion with rock-soil. The U/CCA method was then applied to define the “urban zone” in an elegant way. The WUI was defined as the result from the enhanced overlapping spatial zone between the urban area (defined by the U/CCA), the forest-fire ignition points and (detected from the Landsat imagery classification). A buffer of a characteristic value (resulting from the distance analyses between human caused ignition points and urban infrastructures) was also considered. This new approach allowed the definition, parameterisation and analysis of the evolution of WUI in the last three decades (1990, 2000 and 2010 years).

5. Bibliography

- Lampin-Maillet C., Jappiot M., Long M., Bouillon Ch., Morge D., Ferrier J.P., “Mapping wildland-urban interfaces at large scales integrating housing density and vegetation aggregation for fire prevention in the South of France”, *Journal of Environmental Management*, vol. 91 n° 3, 2010, p. 732-741.
- Radeloff V., Hammer R., Stewart S., Fried J., Holcomb S., McKeefry J., “The wildland urban interface in the United States”, *Ecological Applications*, vol. 15 n° 3, 2005, p. 799-805.
- Rozenfeld H., Rybski D., Andrade Jr. J., Batty M., Stanley H., Makse H., “Laws of population growth”, *Proceedings of the National Academy of Sciences of the USA (PNAS)*, vol. 105 n° 48, 2008, p. 18702-18707.

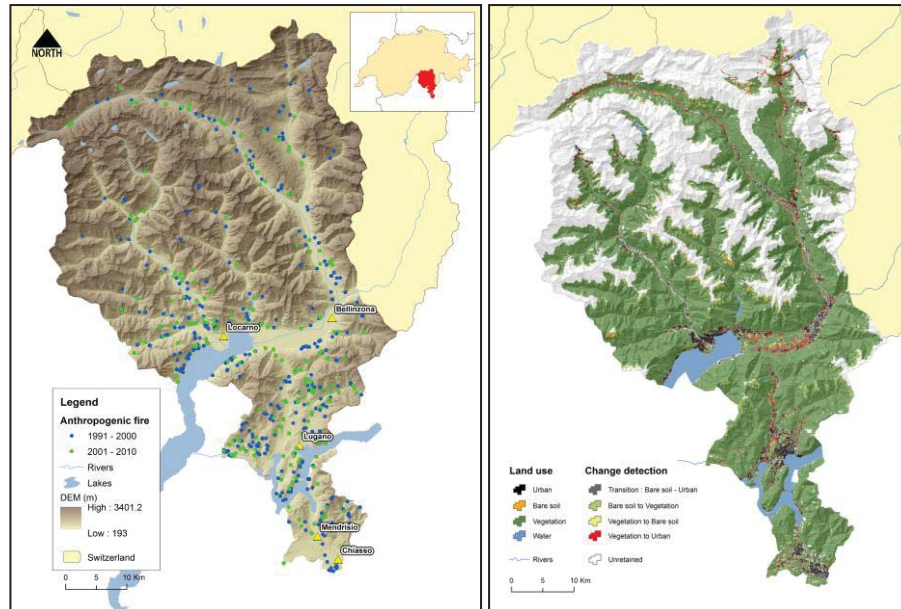


Fig.1 The forest fires ignition points in Ticino (Switzerland) (on the right) and the result of the land-use classification and the change detection (on the left).

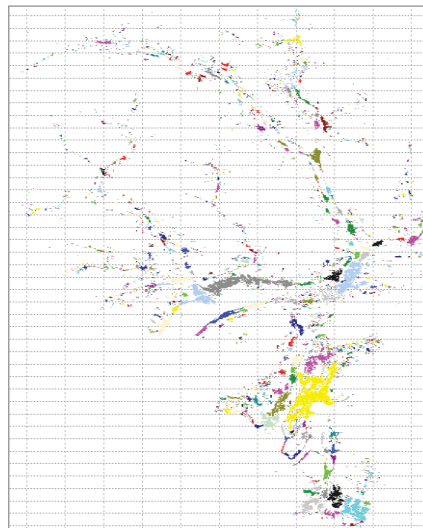


Fig.2 The U/CCA results. The input data are represented from the Swiss population census (hectometric grid). The different colors represent the individual detected urban clusters.

« !ncident » Citoyens capteurs

Retour d'expérience sur l'application « G!v et Praj » déployée au Danemark

Anne-Sophie Collignon - Emmanuel Mondon

Intergraph,
Quai Timmermans, 14/01
4000 Liège
Belgique
<http://www.intergraph.com>

RÉSUMÉ. « !ncident » en français, « Send a hint » en anglais ou « G!v et Praj » pour les danois est une application pour iPhone & Smartphone développée par Intergraph et déployée pour la première fois au Danemark (cf. <http://givetpraj.nu/>), en mars 2011, permettant à tout un chacun (équipé d'un mobile adéquat) d'être un citoyen capteur et ainsi de faire remonter des informations géolocalisées utiles aux collectivités locales.

ABSTRACT. "G!v and Praj" ("Send a hint") is an application for iPhone & Smartphones developed by Intergraph and deployed in Denmark (see <http://givetpraj.nu>) that allows anyone (that has an adequate Smartphone) to be a "sensor citizen" by providing useful geo-localized based information valuable for local administration.

MOTS-CLÉS : iPhone, Android, citoyen, geospatial, SIG, capteur, administration.

KEYWORDS: iPhone, Smartphone, geospatial, GIS, citizen, administration, sensor

1. Introduction

Qui n'a pas un jour été contraint de changer la roue de son véhicule car elle s'était abîmée dans un énorme nid de poule au milieu d'une route? Vous souvenez-vous ensuite du parcours du combattant pour obtenir un point de contact en vue de signaler votre mécontentement sur l'état de nos voiries... pour en fin de compte devoir toujours régulièrement rouler en étant attentif plutôt à l'asphalte qu'aux autres véhicules... Ceci n'est qu'un exemple général parmi d'autres, qui illustre parfaitement la raison d'être de l'application détaillée dans ces lignes.

« *Incident* » est la version française inspirée du danois « *G!v et Praj* », il s'agit d'une application pour *Smartphone* et *iPhone*, de type “*crowd sourcing*” qui permet aux citoyens de rapporter un incident directement à une autorité locale responsable de le résoudre. La procédure est simple car le citoyen qui a installé gratuitement l'application sur son mobile, peut simplement prendre une photographie du problème qu'il constate, par exemple un nid de poule, un graffiti, un panneau signalétique défectueux, un éclairage public en panne, *etc.* à laquelle il peut joindre un commentaire. Il choisit la catégorie de l'incident et il peut éventuellement fournir un complément d'information relative à sa position. Son constat est ensuite encodé dans une base de données et rapporté par e-mail à l'autorité compétente la plus appropriée pour le résoudre. Un rapport en ligne est également automatiquement accessible. L'utilisateur n'a donc pas besoin d'être connecté à un ordinateur portable, et il ne doit pas non plus se soucier du point de contact responsable pour solutionner le problème qu'il a identifié. Les informations envoyées sont automatiquement intégrées à la base de données, rapportées à l'autorité responsable ayant souscrit au système, laquelle est automatiquement identifiée et contactée.

L'application utilise des technologies avancées dans le domaine du géospatial afin de s'assurer que l'incident rapporté pourra être facilement localisé. Le rapport génère automatiquement la position géographique, grâce aux technologies A-GPS embarquées dans le mobile, permettant une précision de positionnement jusqu'à 1 mètre. Néanmoins, si le GPS n'est pas suffisamment précis (meilleur que 25 mètres en ville, meilleur que 50 mètres en dehors des villes), l'utilisateur sera alors invité à préciser sa position sur une carte ou à sélectionner l'adresse la plus proche manuellement.

L'application peut être gratuitement téléchargée sur *Apple App store* pour l'utilisation d'un *iPhone*, ou sur *Google Android-based* pour *Smartphones*.

L'application a été développée et mise en œuvre au Danemark en 2011, où elle compte actuellement plus de 10 000 utilisateurs et plusieurs milliers d'incidents ont été rapportés aux autorités responsables, permettant ainsi aux administrations locales de compter sur des milliers d'yeux sur le terrain afin d'identifier de façon efficace des problèmes nécessitant une intervention.

Plus de 35 administrations locales (plus de 33% de l'ensemble des administrations locales au Danemark) ont souscrit à la mise en œuvre des services « *G!v et Praj* » et reçoivent au moins une fois par jour les incidents rapportés. Plusieurs administrations locales utilisent les technologies SIG afin d'analyser et répartir les incidents identifiés aux responsables de département afin de s'assurer que les problèmes sont résolus aussi rapidement que possible.

Depuis que l'application a été mise en œuvre au Danemark, où elle y a déjà fait ses preuves, d'autres pays tels que la République Tchèque, la Suède, l'Espagne, la France et la Belgique ont également marqué leur intérêt. C'est pourquoi cet article vous donnera peut-être à vous aussi l'envie de tenter l'expérience tant du point de vue utilisateur que du point de vue client...

2. Le fonctionnement de l'application du point de vue de l'utilisateur

« *Incident* » est une application du type « *crowd sourcing* » dont la traduction littérale française serait « externalisation ouverte » et qui selon Wikipédia¹ consiste au « (...) fait d'utiliser la créativité, l'intelligence et le savoir-faire d'un grand nombre de personnes (...) pour réaliser certaines tâches traditionnellement effectuées par un employé ou un entrepreneur (...) ». Ces quelques mots résument parfaitement l'objectif premier de l'application qui permet à toute personne équipée d'un *Smartphone* ou d'un *iPhone* de rapporter un incident, qui ne serait peut-être jamais identifié ou très tardivement par un employé de l'administration compétente en vue de planifier sa réparation ou son entretien.

¹ <http://fr.wikipedia.org/wiki/Crowdsourcing>



Figure 1. *Incident*

Ainsi, prenons un exemple concret, Monsieur Dupond est équipé d'un *iPhone*. Il télécharge gratuitement sur *iTunes*, l'application « *Incident* » qu'il installe ensuite. En quittant sa maison, dans sa rue il remarque un important et dangereux nid de poule. Il démarre l'application, aussitôt et de façon transparente pour lui, avec l'aide du système GPS embarqué dans son *iPhone*, l'application identifie en quelques secondes la précision de sa position, pendant ce temps, l'utilisateur prend une photographie, puis détermine à quelle catégorie il appartient (il a le choix entre douze), dans son cas, il s'agira de la quatrième, pour information, voici le détail des catégories :

1. Animal mort	7. Trottoir dégradé
2. Luminaire défectueux	8. Signalisation incorrecte
3. Vandalisme	9. Plantations invasives
4. Chaussée dégradée	10. Signalisation endommagée
5. Graffiti/Affiche	11. Déchets sur la voie publique
6. Poubelle pleine	12. Autre

Tableau 1. *Détail des catégories*

Il peut si nécessaire ajouter une description ou un commentaire tel que: « *ce trou est présent depuis de nombreux mois et il s'agrandit dangereusement avec le gel que nous avons connu ces derniers jours* ». Ensuite, si sa position est connue avec suffisamment de précision, Monsieur Dupont pourra simplement envoyer l'incident qu'il a identifié, sinon, il lui sera demandé d'apporter manuellement plus d'information. L'application lui proposera soit de pointer l'incident sur une carte (*Open Street Map*) déjà zoomée et cadrée sur la zone d'intérêt, soit alors, d'encoder l'adresse la plus proche du trou, sur base d'information pré-encodée (*via Lifas OIS service* (www.lifaois.dk)). La philosophie de l'application est de contraindre le moins possible l'utilisateur de fournir des informations de positionnement. Ainsi, celles-ci ne sont demandées que lorsque cela est strictement nécessaire (si la précision est supérieure à 25 mètres en zone urbaine ou supérieure à 50 mètres en dehors d'une zone urbaine). Dans ce cas, l'information de positionnement sera stockée dans la base de données sous un code attributaire « -2 » pour différencier les cas où la précision du mobile était suffisante. Le travail de l'utilisateur sur le terrain s'arrête là.

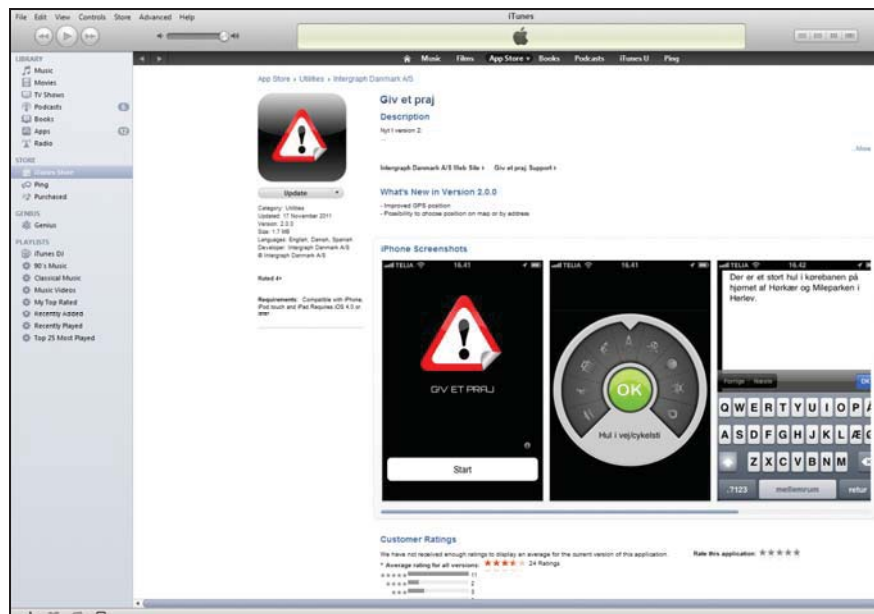


Figure 2. Exemple Danois du téléchargement de l'application sur iTunes



Figure 3. Démarrer l'application



Figure 4. Prendre une photo de l'incident

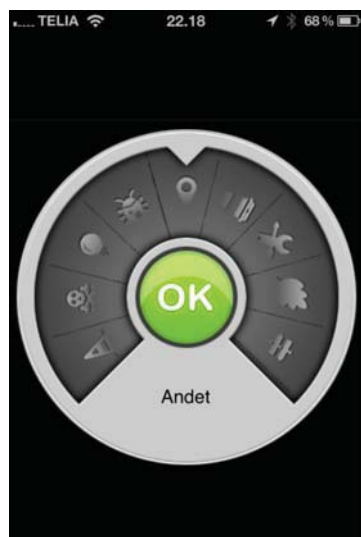


Figure 5. Choisir la catégorie de l'incident



Figure 6. Apporter un commentaire optionnel



Figure 7. Uniquement, si le positionnement GPS est imprécis, choisir entre l'encodage sur carte ou par adresse



Figure 8. Pointer la localisation de l'incident sur une carte



Figure 9. Choisir l'adresse la plus proche parmi une pré-sélection



Figure 10. *Quitter application*

3. Le fonctionnement de l'application du point de vue du client

Le client est l'administration locale ayant souscrit au service « *Incident* ». Lorsque l'incident est enregistré dans la base de données, en fonction de son type et de sa position, un rapport sera envoyé automatiquement à l'administration compétente et éventuellement la plus proche (tout dépend du ressort de celle-ci). Le client peut choisir de souscrire à un des trois niveaux de services disponibles : bronze, argent ou or. Le choix du niveau lui octroiera des capacités complémentaires facilitant la gestion des incidents rapportés.

Ainsi un client de type « bronze » aura un seul point d'accès e-mail, sur lequel il recevra les informations encodées par l'utilisateur telles que la photo, la catégorie de l'incident, les observations et commentaires des utilisateurs, la position et la précision de l'incident venant du GPS. Un client de type bronze a généralement déjà un système interne de gestion des incidents, il leur suffit donc d'encoder les e-mails reçus directement dans leur propre base de données.

Un client de type « argent » aura en plus, 12 comptes e-mails possibles pour filtrer les incidents reçus par catégorie. Il pourra également obtenir les informations relatives au positionnement par l'encodage d'adresse ou par l'identification sur une carte si la position du GPS n'était pas assez précise.

Enfin le client « or » pourra en plus, accéder au service de notification par SMS, à la console d'administration web et enfin il disposera d'accès aux services Web OGC tels que WMS (*WebMapService*) et WFS (*WebFeatureService*), ce dernier supportant les sorties en KML (format *GoogleEarth*) et RSS (pour le suivi des mises à jour).

	Or	Argent	Bronze
Simple Service e-mail <ul style="list-style-type: none"> - Catégorie - Commentaires - Position et précision GPS - Photographie - Une seule adresse e-mail 	-	-	+
Service e-mail avancé <ul style="list-style-type: none"> - Cartes de positionnement - Adresse la plus proche - 12 adresses e-mails 	+	+	-
OGC Service Web <ul style="list-style-type: none"> - WMS, WFS (+KML, +RSS) 	+	-	-
Service SMS	+	-	-
Administration web	+	-	-

Tableau 2. *Récapitulatif des trois niveaux de service*

La console d'administration web, disponible uniquement pour les clients du type « or », est un point d'entrée unique qui facilite la gestion des incidents et qui permet d'afficher l'ensemble des informations transmises par l'utilisateur. Les illustrations ci-dessous sont extraites de la page d'administration danoise, la figure 11 représente la liste des incidents représentés par un identifiant unique avec comme information supplémentaire pour chacun : une catégorie, une adresse, un statut qui détermine l'état actuel de l'incident (reçu, en cours de contrôle, traité ou impossible à traiter) ainsi que la date de soumission du problème.

IncidentID	Category	Address	Status	by	Date	Settings
5578	Hul i vej/cykeltsti	Havnegade 23, 1058 Indre by, Danmark	Received		17-02-2012 15:58:37	
5551	Affald på gaden	Bushøjvej 19, 2450 København, Danmark	Received		16-02-2012 07:38:20	
5548	Mangende/ødelagt betyning	Sjælar Boulevard & Hørdumsgade, 2450	Received		15-02-2012 17:47:00	
5547	Affald på gaden	Bushøjvej 15, 2450 København, Danmark	Received		15-02-2012 17:41:23	
5512	Fortov ujevnt/ødelagt	Hammerichsgade, 1611 Indre by, Danmark	Received		14-02-2012 21:01:29	
5487	Hul i vej/cykeltsti	Kronprinsessegade 92, 1306 Indre by, Danmark	Received		13-02-2012 20:34:04	
5475	Skadedyr/døde dyr	Hedemannsgade, 2100 København, Danmark	Received		12-02-2012 15:13:53	
5470	Hul i vej/cykeltsti	Sjælar Boulevard 55, 2450 Vesterbro/Kongens	Received		11-02-2012 13:43:31	
5469	Hul i vej/cykeltsti	Sjælar Boulevard 51, 2450 Vesterbro/Kongens	Received		11-02-2012 13:41:53	

Figure 11. *Liste des incidents via la console d'administration web*

La figure 12 correspond au résultat obtenu lorsque l'opérateur sélectionne un incident listé ci-dessus. Il obtient ainsi plus de détails sur celui-ci car il peut visualiser la photographie envoyée par l'utilisateur, la position du problème sur une carte, le commentaire éventuel fourni par l'utilisateur, l'adresse, les coordonnées de positionnement en longitude/latitude ainsi que la précision de la position.

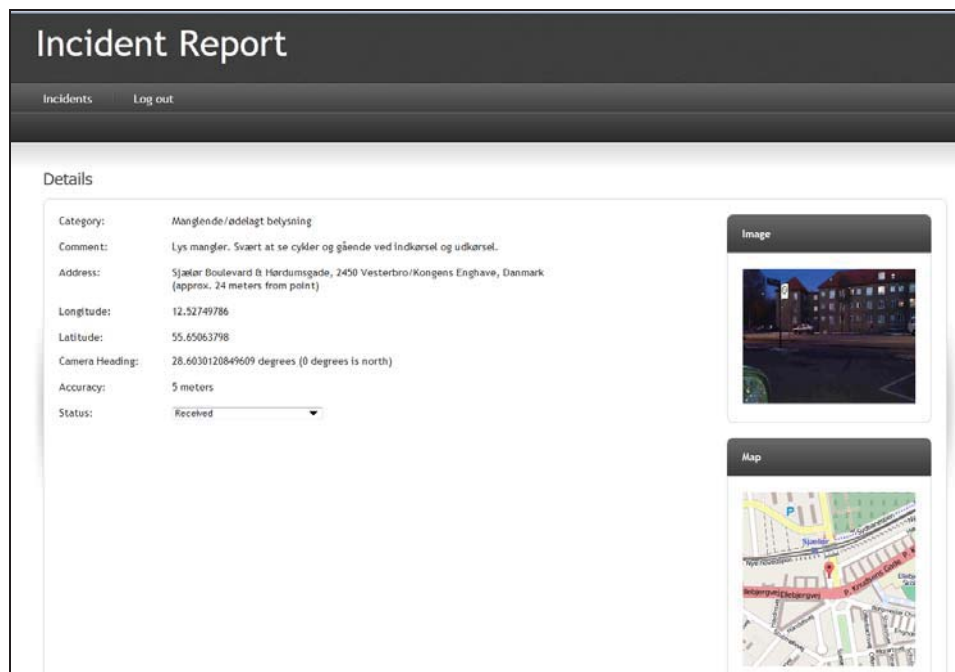


Figure 12. Détails d'un incident rapporté dans la console d'administration web

Les clients de type «or» disposent d'accès aux services web OGC (<http://www.opengeospatial.org/>) sous forme de *WebMapService* (WMS) et *WebFeatureService* (WFS). Ils permettent en utilisant une application cliente supportant ces spécifications OGC de visualiser la répartition des incidents sur une carte, et ils peuvent grâce aux informations attributaires disponibles au travers du service WFS, appliquer des filtres afin de réaliser des statistiques sur les incidents rapportés.

Par exemple, si nous chargeons le WMS correspondant aux incidents danois dans *Intergraph SDI Portal*, nous obtenons la répartition géographique de ceux-ci (figure 13). Avec l'outil d'information, en cliquant sur un des points dessinés sur la carte, il est possible d'afficher les informations attributaires, détaillant l'incident (figure 14).

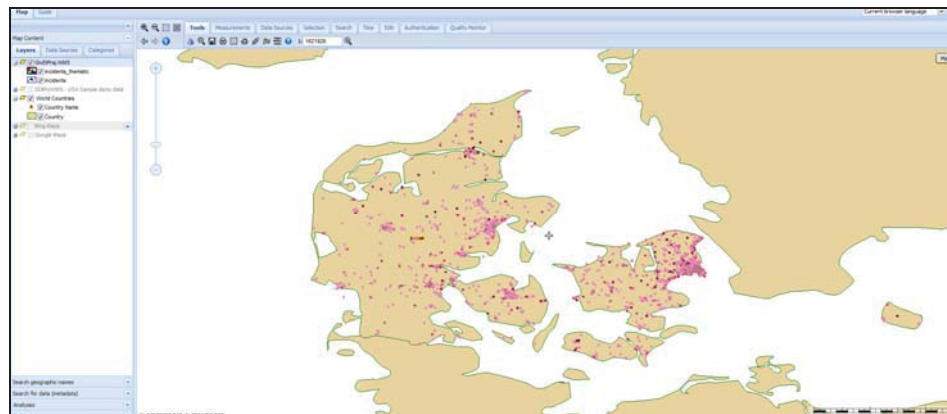


Figure 13. Visualisation de la répartition des incidents danois par WMS dans Intergraph SDI Portal

Feature info	
incidents_thematic incidents	
	Country Name Country
Name	Value
incidentGeomID	2780
type	Graffiti/plakater
comment	Upassende tekst på vejen, skrevet med hvid spraymaling.
deviceVersion	2.3.3
devicePlatform	Android
longitude	9.1501562
latitude	56.29318015
altitude	-1
heading	389.5
speed	-1
accuracy	53
altitudeAccuracy	-1
imageUrl	http://cdn.givetpraj.nu/images/5f88f5d9-338e-4eb6-a1a6-ab98c9e8dc5b.jpg
date	9/7/2011 9:55:10 AM
categoryID	8
municipalityCode	DK0791
address	Lærkevej 28, 7470 Viborg, Danmark
distanceToAddress	4.63775299317982
incidentID	3054
status	1
deviceUUID	

Figure 14. Détails de l'information attributaire d'un incident



Figure 15. Contenu de l'Image URL disponible via l'information attributaire

Lorsque le client dispose d'un outil SIG (du type *desktop*) plus complet qu'un simple client web, il peut alors réaliser plusieurs cartographies des incidents sur base de filtres et requêtes sur le service WFS. Par exemple, en utilisant l'application *GeoMedia* d'*Intergraph*, il est possible d'effectuer rapidement quelques requêtes attributaires. La figure 16 représente simplement le rendu, en temps réel, de l'ensemble des incidents rapportés sur le Danemark.

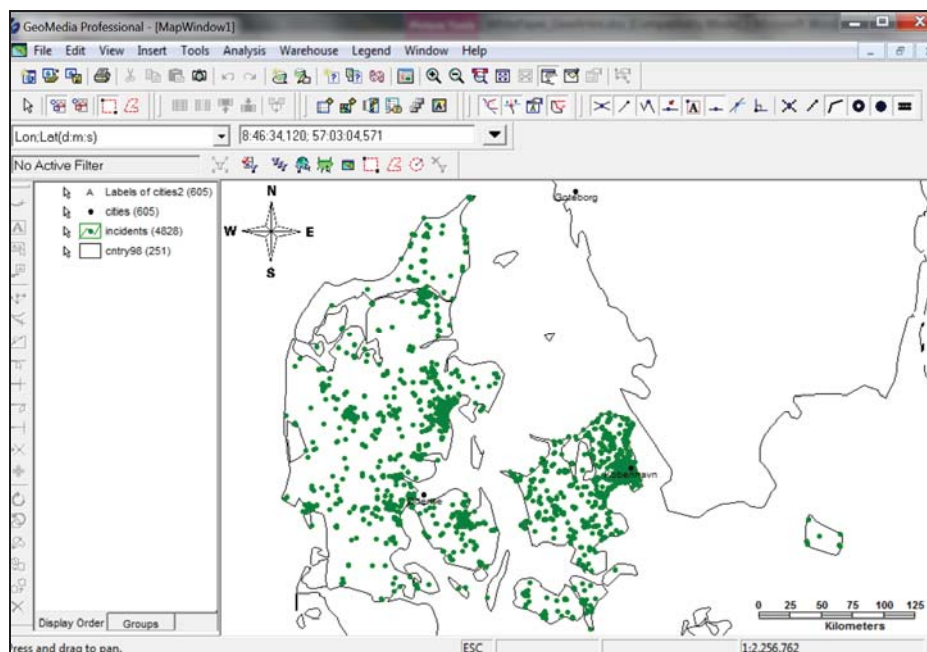


Figure 16. Rendu dans Intergraph Geomedia de l'ensemble des incidents via WFS

Ensuite, il est possible de définir une légende spécifique par type de catégorie, ceci mettant en évidence les incidents les plus souvent rapportés, par exemple le type « chaussée dégradée » apparaît clairement comme étant le plus important. A nouveau, l'opérateur peut afficher le contenu attributaire de chaque incident.

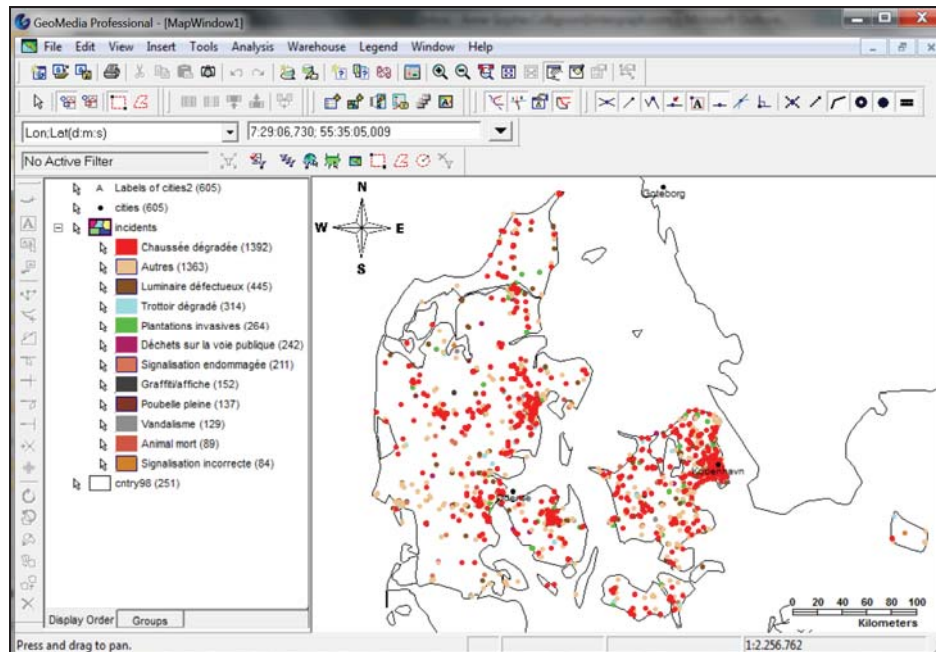


Figure 17. Rendu par catégorie d'incident

The screenshot shows the 'Incidents Properties' dialog box with the 'Attributes' tab selected. It displays a table of attributes for a specific incident.

Name	Value
gml_descriptor	
gml_name	
incidentGeom	3003
type	Hul i vej/cykelsti
comment	Hul i vej/ent
deviceVersion	4.3.3
devicePlatform	iPhone
longitude	9.748079
latitude	56.647436
altitude	0
heading	351.024231
speed	0
accuracy	50
altitudeAccuracy	95
imageUri	http://cdn.givnetprej.nu/images/d17b150d-4668-407d-bc6d-22de06bfe6ee.jpg
date	16/09/2011 15:08:10
categoryID	9
municipalityCode	DK0846
address	Plantagevej 25, 9500 Mariager, Denmark
distanceToAdc	15.6690287318626
incidentID	3275
status	1
deviceUUID	
fid	incidents 3003

Figure 18. Fiche attributaire dans Intergraph Geomedia



Figure 19. Contenu de l'Image URL disponible via la fiche attributaire

Sur base des informations attributaires contenues pour chaque incident, il est ainsi possible de réaliser plusieurs cartographies afin d'identifier les zones géographiques plus souvent touchées par un ou plusieurs types d'incidents. Par exemple, pour l'application administrée au sein d'une ville, elle lui permet de planifier des interventions de nettoyage plus fréquentes dans certains quartiers où un plus grand nombre d'incidents sont rapportés pouvant laisser penser qu'ils induisent plus de nuisances pour leurs habitants.

D'autres cartographies permettent par exemple, de visualiser si l'application est plus utilisée par des adeptes de l'*iPhone* ou du *SmartPhone*. On peut également imaginer une cartographie liée à la précision du système GPS et peut-être la coupler à celle de l'utilisation d'un type ou l'autre d'appareil. Enfin, il est aussi possible d'étudier les périodes durant lesquelles certains types d'incidents sont plus rapportés que d'autres, et ainsi tenter de les corréler avec des paramètres extérieurs (par exemple, plus de reports de trous dans les voiries à la fin de l'hiver à cause du gel). L'application peut donc être vue comme un outil de *reporting* facilitant la planification à court et moyen termes d'interventions sur le terrain.

4. Le fonctionnement de l'application du point de vue du hardware

La mise en œuvre de l'application pour le client final est plutôt immédiate, car il lui suffit simplement de souscrire à un des niveaux du service, il ne doit donc pas investir dans du coûteux matériel car l'ensemble de l'application est déployée sur un *Cloud Computing*. Selon Wikipedia², « le cloud computing est un concept qui consiste à déporter sur des serveurs distants des stockages et des traitements informatiques traditionnellement localisés sur des serveurs locaux ou sur le poste de l'utilisateur (...) le cloud computing est l'accès via le réseau, à la demande et en libre-service à des ressources informatiques virtualisées et mutualisées (...) ».

L'architecture de l'application est illustrée ci-dessous. L'objectif est de proposer une solution mondiale et performante, c'est pourquoi l'infrastructure se base sur un

² http://fr.wikipedia.org/wiki/Cloud_computing

cloud de type *Amazon EC2+S3*. Deux serveurs sont déployés, un pour la base de données où est installé *MS SQL Server Spatial 2008*, un autre pour le serveur d'applications. Les deux serveurs tournent sur *Windows Server 2008*, 64bit avec 7,5 Giga bytes de RAM. *Intergraph GeoMedia WebMap Professional* est installé sur le serveur d'application, afin de servir les services web OGC (WMS et WFS), le service d'e-mail et l'application web d'administration.

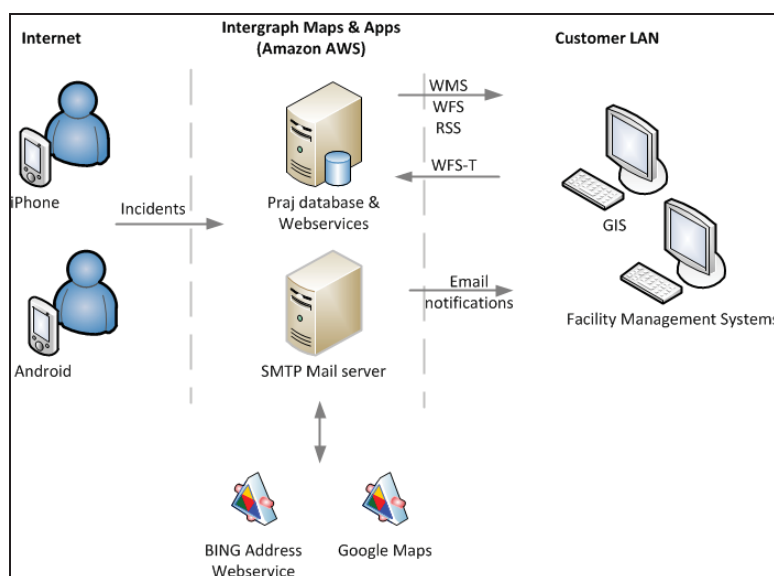


Figure 20. Architecture de l'application

L'avantage du *cloud* est qu'il est extensible et donc, adaptable aux besoins sans devoir migrer physiquement les serveurs. Actuellement, l'ensemble de l'infrastructure est centralisée et gérée par une seule équipe au Danemark et cela même pour les services déployées pour d'autres pays. Dès qu'un client de type « or » souscrit au service, il reçoit un lien sécurisé avec la console d'administration qui lui renvoie uniquement les incidents le concernant. Il reçoit aussi les URLs des services WMS et WFS qui sont eux, par contre, uniques par pays, d'où l'intérêt de disposer d'un client (*desktop* ou *web*) permettant de lire des services web OGC, afin de filtrer les résultats sur sa zone d'exploitation.

5. Quelques retours d'expérience

Au Danemark, l'application a été mise en œuvre en 2011. Actuellement, 35 administrations locales ont souscrit au service, ce qui représente 33% des administrations danoises. Plus de 10 000 utilisateurs ont téléchargé et installé l'application sur leur mobile, plusieurs milliers d'incidents (environ 7 600) ont été

rapportés jusqu'à ce jour. L'application a connu un succès médiatique inattendu, plus d'une centaine d'articles sont parus et trois interviews pour des radios nationales ont eu lieu.



Figure 21. Exemple d'article presse paru dans DagBladet Køge, le 15 octobre 2011

Les retours d'utilisateurs sont dans l'ensemble très positifs, car il s'agit d'une application à usage national, gratuite et facile d'utilisation. Elle a d'ailleurs été numéro 1 des téléchargements *iTunes* au Danemark, catégorie *Utilities*.

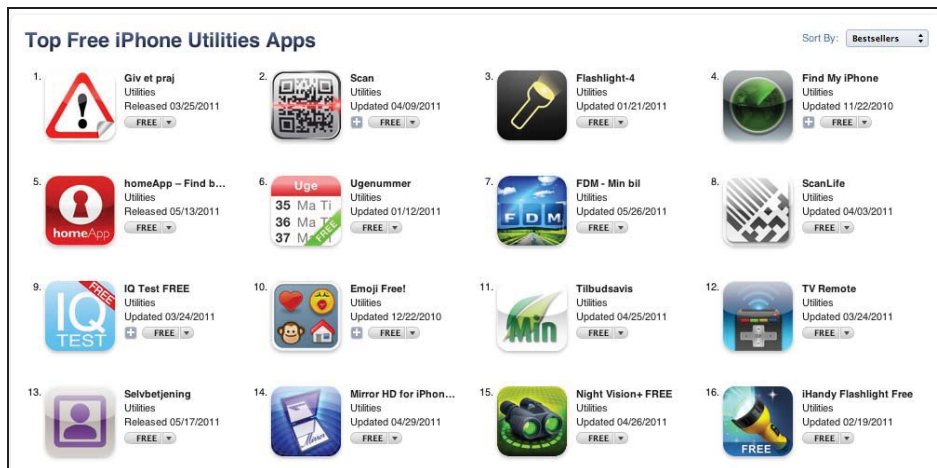


Figure 22. Numéro 1 des téléchargements au Danemark

Voici un extrait des commentaires reçus par des utilisateurs postés sur *iTunes* (traduis du Danois):

« Très bon software, particulièrement impressionnant, assez puissant pour rapporter des problèmes, on pourrait imaginer étendre l'application pour rapporter d'autres types d'incidents... et pourquoi pas en lien avec les services de secours... »
 « Utilisé trois fois à Copenhague, et les trois fois, les problèmes ont été fixés dans les 24 à 48 heures. On ne peut pas mieux faire !! »

Les administrations, c'est-à-dire les clients finaux sont aussi très satisfaits de l'application et de l'infrastructure proposée, d'autant plus qu'elle est gratuite pour les clients « bronze » et très bon marché pour les clients « argent/or ». Un exemple de commentaire, extrait et traduit de l'article en ligne : <http://sydkysten.dk/Borgere-i-Solroed-skal-give-et-praj/Sydkysten/artikel/158488>, provenant du maire de la ville de Solrød qui a souscrit au programme et qui dit : « cela permet de recueillir des informations rapidement et facilement, c'est simple et accessible (...), j'espère que l'ensemble des citoyens de ma commune profiteront de l'occasion de partager leurs observations », il voit également la possibilité d'étendre ces concepts de citoyens capteurs vers d'autres choses en vue d'améliorer la qualité de vie des citoyens de sa ville. Tout en sachant que l'ensemble des incidents ne seront pas nécessairement solutionnés car il établit un plan et une pondération pour prioriser la résolution des incidents rapportés.

L'application au Danemark aura ouvert également des perspectives *BtoB* comme la possibilité pour une société telle que *Daluso*, spécialisée dans la maintenance de panneaux signalétiques, d'optimiser leur procédure de commande en s'appuyant sur l'application « *Giv et Praj* ».

Pour terminer, pointons quelques limitations identifiées de l'application. Seuls les *Smartphones* équipés du système *Android* avec une résolution d'écran de 340 par 480 minimum sont supportés, le support de *BlackBerry* n'est pas prévu. Rappelons néanmoins que les mobiles de type *iPhone*, *iPad* et *iPodTouch* sont déjà supportés.

Il n'est pas possible de prendre une photo d'un incident avec un appareil qui n'est pas connecté au réseau wifi ou 3G et de rapporter l'incident à posteriori.

Il n'est pas possible de revenir sur un incident communiqué ou de suivre son état car la publication est anonyme et que les clients ayant souscrit au service ne sont pas tenus de répondre aux demandes sous certains délais, ni de rendre état d'une décision prise sur l'incident. Par ailleurs, chaque client est libre de diffuser de l'information sur les incidents qu'il gère par ses propres moyens.

5. Conclusions

Au Danemark, l'application vient d'introduire 12 nouvelles catégories pour répondre à de nouvelles demandes des sociétés de services de distribution et de gestion du gaz, de l'électricité, de l'eau, des télécommunications. Ainsi sont apparues des catégories telles que câbles cassés, terrains inondés, taques d'égout manquantes, *etc.* Il est également prévu de supporter d'autres plateformes pour mobiles comme *Windows Phone 7*.

Au vu du succès rencontré par « *G !v et Praj* » au Danemark, des déploiements nationaux ont eu lieu en République Tchèque, en Espagne et en Suède. Le déploiement en France et en Belgique a été lancé au second trimestre 2012.

Enfin le concept de citoyens capteurs n'est pas neuf. Il existe pour d'autres domaines tels que pour les relevés météorologiques et atmosphériques, le recensement d'oiseaux ou de papillons dans les jardins, *etc.* C'est pourquoi nous ne sommes pas surpris du succès rencontré par l'application car toute personne sensible au civisme sera enthousiaste pour rapporter tout dysfonctionnement en vue de rendre meilleur son cadre de vie et dans la foulée de redorer l'image de marque de sa ville, de sa région et de son pays. Néanmoins, l'application n'a pas de vocation à la délation et reste un outil pensé pour aider les administrations locales à améliorer le cadre de vie des citoyens et l'entretien de leurs infrastructures.

Géoréférencement et appariement de données issues des cartes de Cassini

Intégration dans un référentiel topographique actuel

Benoit Costes — Eric Grosso — Christine Plumejeaud

*Institut Géographique National, COGIT
73, avenue de Paris
F-94160 Saint-Mandé
prenom.nom@ign.fr*

RÉSUMÉ. L'intégration de l'information contenue dans les cartes topographiques anciennes dans un référentiel topographique historique (une base de donnée spatio-temporelle) est difficile du fait de l'imperfection inhérente à ces données. Cet article se focalise sur le géoréférencement et l'appariement des données ponctuelles extraites de la carte de Cassini, version noir et blanc, avec la BDTOPO, base de données topographique vectorielle du Référentiel à Grande Échelle de l'IGN, version 2011. Le géoréférencement proposé est réversible et appliqué indépendamment pour chaque feuille. Il permet de positionner des objets géographiques historiques vectoriels et ainsi d'étudier les imprécisions de leur localisation ainsi que les incohérences cartographiques des cartes de Cassini. Une amélioration du processus d'appariement suivant une approche multi-critères est ensuite proposée, afin d'automatiser partiellement le paramétrage et de faciliter l'analyse des résultats.

ABSTRACT. The integration of information contained in old topographic maps into a historical topographic framework (spatio-temporal database) is a complex task due to the inherent imperfection of data. This article focuses on the georeferencing and data matching process of vector point data extracted from the black and white version of Cassini's maps with the vector topographic database of the French mapping agency's large scale framework (BDTOPO, version 2011). Each Cassini sheet is georeferenced using a unique and reversible transformation enabling provision of information such as the location of historical geographical objects, thus allowing to study the imprecision of these locations and the cartographic inconsistencies of Cassini's maps. Then an improvement of the data matching process following a multi-criteria analysis approach is proposed in order to partially automate the data matching parameterisation and to ease results analysis.

MOTS-CLÉS : Cassini, carte ancienne, appariement, géoréférencement, imperfection

KEYWORDS: Cassini, old maps, data matching, georeferencing, imperfection

1. Introduction

Depuis 250 ans notre territoire s'est considérablement équipé, densifié au fil de l'accroissement démographique et des progrès technologiques, modifiant ainsi la topographie du paysage français. Par ailleurs, l'industrialisation a généré un important exode rural. L'étude des relations entre l'évolution du paysage français et l'évolution de la répartition de la population sur le territoire est pertinente, aussi bien du point de vue historique, que du point de vue de l'aménagement du territoire : si nous comprenons mieux notre passé, il est possible de mieux anticiper les besoins futurs. Or ce passé topographique nous a été légué sous la forme des cartes anciennes (Cassini – XVIII^{ème} siècle – et État-Major – XIX^{ème} siècle – par exemple), qui en proposent une trace codifiée selon les spécifications cartographiques propres à leur époque. Le projet de recherche GéoPeuple¹ se propose de d'analyser puis d'extraire par vectorisation le contenu des cartes de Cassini (échelle 1 : 86 400) puis d'État-Major (échelle 1 : 80 000), pour constituer une base de données géo-historique spatio-temporelle, et étudier ainsi la densification des territoires. Ce projet, financé par l'Agence Nationale de la Recherche, regroupe différents partenaires : le laboratoire COGIT de l'Institut Géographique National (IGN), le Laboratoire de Démographie Historique (LaDéHis) de l'École des Hautes Études en Sciences Sociales (EHESS) et l'équipe MALIRE du Laboratoire d'Informatique de Paris 6 (LIP6).

La constitution d'une telle base de données nécessite la numérisation, l'identification des concepts représentés avec leur organisation dans une ontologie appropriée, la vectorisation (manuelle ou automatique) [LEY 05], le géoréférencement de l'information [DUP 07, NOI 11], puis l'appariement des objets extraits avec ceux connus aux autres époques [DEV 97, OLT 08]. Or, pour cette tâche d'intégration, un certain nombre de travaux ont été effectués en vue de faciliter ces opérations entre données issues de bases de données contemporaines, non compatibles pour cause d'échelles différentes, ou d'objectifs de représentation différents. Avec les données anciennes, il s'agit également de tenir compte de l'imperfection de l'information.

Cet article se concentre sur les données extraites de la carte de Cassini, carte qui constitue la première tentative de créer un référentiel national de données topographiques géoréférencées [PEL 90]. Cependant, cette carte est imparfaite : les localités sont parfois placées à une lieue (soit 4 km) de leur véritable position [BER 98] (tome 1, page 173), elle est incomplète, et comporte des erreurs de positionnement relatif. En dépit de ces erreurs, il s'agit de proposer des méthodes qui puissent permettre d'identifier les objets qui ont perduré, qui ont disparu ou sont apparus, grâce, d'une part, à un géoréférencement assez précis des objets, et, d'autre part, à l'appariement des objets avec l'information topographique actuelle. L'étude se limite au ponctuel, et à trois natures d'objets (les moulins et industries, le bâti religieux ou non-religieux) pour lesquelles le biais sélectif est suffisamment faible. Aujourd'hui comme hier, ces concepts étaient choisis pour être représentés, même sous des formes (signes et appellations) différentes, contrairement aux dessertes ferroviaires nées au XIX^{ème} siècle.

1. <http://geopeuple.ign.fr>

Cet article est organisé comme suit : la première section introduit les données traitées, la deuxième section est consacrée au problème du géoréférencement de ces données, la troisième à celui de l'appariement de ces données et la dernière délivre les conclusions de cette expérience et annonce les perspectives de ce travail.

2. Données utilisées

Nous exploitons la carte de Cassini, levée au 86400^{ème}, puis éditée en noir et blanc, qui se présente sous la forme d'une collection de 181 feuilles. L'espace de travail a été restreint à quatre feuilles de Cassini, couvrant chacune des paysages différents : Saint-Malo (côte maritime, feuille 127 publiée en 1759), Reims (plaine, feuille 79 publiée en 1759), Agen (zone dense, feuille 73 publiée en 1778) et de Grenoble (haute montagne, feuille 119 publiée en 1779). Une cinquième zone, Châlons (feuille 80 publiée en 1757), située au sud de la feuille de Reims, a été ajoutée afin de pouvoir étudier les problèmes de raccord entre deux feuilles voisines, comme développé ultérieurement en section 3.3.

A partir de la numérisation des feuilles en haute définition (300 DPI) par l'IGN, le protocole conçu par le projet permet de vectoriser manuellement l'information en coordonnées image, quelque soit sa modélisation géométrique (ponctuelle, linéaire ou surfacique), et de lui associer des informations sur sa nature, (classe et sous-type), ainsi que sur sa désignation (via un attribut "toponyme"). Après géoréférencement, les objets ainsi extraits peuvent être confrontés à l'information topographique actuelle que contient la BDTOPO version 2011, base de données vectorielle de précision métrique composante du Référentiel à Grande Échelle constitué par l'IGN, qui utilise la projection Lambert93.

3. Géoréférencement

Avant cette étude, plusieurs géoréférencements de la carte de Cassini avaient été réalisés, dont l'un dans le cadre du projet "Des villages de Cassini aux communes d'aujourd'hui"² conduit par l'EHESS, et un autre par l'IGN dans le cadre de la dématérialisation du fonds de la Cartothèque où un scannage patrimonial enrichi d'un géoréférencement et d'une rectification des cartes anciennes scannées a été effectué. Dans ces deux cas, les méthodes de géoréférencement employées sont similaires et se décomposent en deux étapes : 1. géoréférencement de chaque feuille Cassini dans le tableau d'assemblage des cartes de Cassini afin de reconstituer la carte à l'échelle de la France entière en assurant au mieux leur juxtaposition ; 2. reprojection des feuilles géoréférencées dans le tableau d'assemblage dans une projection actuelle. Seule la méthode de reprojection diffère : calcul et utilisation de paramètres approchés de *Datum Shift* dans le premier cas, utilisation de l'ellipsoïde Clarke 80 comme support de la projection Cassini dans le second.

2. <http://cassini.ehess.fr>

Cependant, ces méthodes sont critiquables car il est historiquement difficile de connaître quelle fut la projection utilisée lors de la conception de la carte de Cassini, en particulier au niveau de la modélisation de la forme de la Terre (sphère ou ellipsoïde ?) [BER 98] (tome 1, page 44). De plus, si ce type de géoréférencement permet de répondre à une problématique de visualisation d'ensemble de la carte, il n'a pas la précision nécessaire pour l'appariement des données vectorisées avec l'information topographique actuelle. En effet, une transformation unique appliquée sur l'ensemble du territoire ne peut prendre en compte l'ensemble des déformations locales spécifiques à chaque feuille de la carte de Cassini.

3.1. Méthode employée

Nous proposons donc de traiter chaque feuille séparément, en lui appliquant une transformation de Helmert (à 4 paramètres) propre. Ce choix présente les trois avantages suivants : (i) il permet d'étudier les erreurs contenues dans les cartes de Cassini ; (ii) la transformation s'applique de manière indifférenciée sur des données raster et vecteur ; (iii) la transformation est inversible et permet de travailler dans l'espace de prédilection (soit l'espace Cassini originel, soit la projection liée à la BDTOPO). Ce procédé autorise ainsi l'emploi de la BDTOPO comme guide pour la vectorisation automatique des données. En effet, l'hypothèse qu'un moulin à eau est proche d'un cours d'eau permet de limiter l'espace de recherche ou *a minima* d'améliorer l'étape d'évaluation automatique du processus.

La transformation de Helmert, qui est une combinaison de deux translations, d'une rotation et d'un changement d'échelle, est valide dans notre contexte d'étude, car les repères, des cartes de Cassini comme des données actuelles, sont orthonormés. Mathématiquement, elle est définie comme suit :

$$\begin{bmatrix} x_c \\ y_c \end{bmatrix} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \cdot \begin{bmatrix} x_s \\ y_s \end{bmatrix} + \begin{bmatrix} c \\ d \end{bmatrix} \quad (1)$$

où : a et b sont les paramètres qui déterminent le changement d'échelle et la rotation ; c et d sont les paramètres de translation ; x_c et y_c sont les coordonnées d'un point clairement identifié (un amer, comme peut l'être par exemple un clocher d'église) dans le référentiel cible ; x_s et y_s sont les coordonnées de ce même point dans le référentiel source. L'inverse de cette transformation est définie comme suit, par résolution inverse du système linéaire défini par l'équation (1) :

$$\begin{bmatrix} x_s \\ y_s \end{bmatrix} = \begin{bmatrix} a_{inv} & b_{inv} \\ -b_{inv} & a_{inv} \end{bmatrix} \cdot \begin{bmatrix} x_c \\ y_c \end{bmatrix} + \begin{bmatrix} c_{inv} \\ d_{inv} \end{bmatrix} \quad (2)$$

$$\text{où : } a_{inv} = \frac{a}{a^2+b^2}, \quad b_{inv} = -\frac{b}{a^2+b^2}, \quad c_{inv} = \frac{b \cdot d - a \cdot c}{a^2+b^2} \quad \text{et} \quad d_{inv} = -\frac{a \cdot d + b \cdot c}{a^2+b^2}$$

L'existence de cette transformation inverse est garantie par le fait que $a^2 + b^2 > 0$. Sinon, $a = 0$ et $b = 0$, impliquant que l'ensemble des coordonnées cible soit réduit à l'unique point de coordonnées (c, d) , ce qui est impossible en pratique.

Deux amers seulement sont requis pour déterminer les paramètres de la transformation de Helmert. Toutefois, un nombre plus important d'amers est généralement utilisé. Ainsi, si n amers sont définis, l'équation (1) peut être réorganisée comme suit :

$$\underbrace{\begin{bmatrix} x_{s1} & y_{s1} & 1 & 0 \\ y_{s1} & -x_{s1} & 0 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ x_{sn} & y_{sn} & 1 & 0 \\ y_{sn} & -x_{sn} & 0 & 1 \end{bmatrix}}_A \cdot \underbrace{\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}}_X \approx \underbrace{\begin{bmatrix} x_{c1} \\ y_{c1} \\ \vdots \\ x_{cn} \\ y_{cn} \end{bmatrix}}_Y \quad (3)$$

L'estimation des paramètres de transformation relatifs au système linéaire $AX = Y$ défini dans l'équation (3) est réalisée par une procédure d'ajustement s'appuyant sur la méthode des moindres carrés. L'ensemble de ce processus de géoréférencement a été implémenté dans un outil qui permet le passage des données vectorisées dans Cassini vers le référentiel de la BDTOPPO ou l'inverse.

3.2. Résultats obtenus et évaluation du processus de géoréférencement

Afin de contrôler la précision de ce géoréférencement, des amers et des points d'évaluation sont sélectionnés conjointement sur les feuilles Cassini en coordonnées image (points source) et sur les données BDTOPPO en projection Lambert 93 (points cible). L'évaluation des géoréférencements réalisés pour chaque feuille est basée sur différents éléments, soit visuels, soit numériques : (i) la visualisation des erreurs résiduelles (niveau local) ; (ii) la visualisation des isolignes des erreurs résiduelles interpolées et liées aux amers, espacées de 10 mètres (niveau global), (iii) le calcul des erreurs en moyenne quadratique ou RMS (*Root Mean Square*), les erreurs résiduelles liées aux amers et celles liées aux points d'évaluation.

Zone considérée	Nombre d'amers	Nb points d'évaluation	RMS	Ecart-type	Erreur min	Erreur max
Agen	15	5	114.25	51.48	10.51	173.06
Grenoble	13	4	318.65	176.72	52.69	750.65
Reims	15	6	103.21	54.91	6.40	174.35
Saint-Malo	17	4	127.62	47.91	61.31	200.71
Châlons	13	5	118.24	59.53	14.38	203.95

Tableau 1. Principaux éléments d'évaluation numérique des géoréférencements.

Ces éléments d'évaluation sont employés pour chacune des 5 feuilles étudiées. Ces évaluations sont décrites de manière détaillée dans [GRO 11] dont les principaux éléments sont résumés dans le tableau 1, montrant des erreurs en moyenne quadratique égales à 110 ± 50 m, à l'exception de la feuille de Grenoble qui se situe en zone montagneuse (320 ± 175 m).

3.3. Étude des écarts entre feuilles de Cassini adjacentes

L'usage d'une transformation propre à chaque feuille pourrait entraîner des décalages importants entre feuilles voisines. C'est pourquoi nous avons évalué les écarts entre feuilles adjacentes avant et après géoréférencement, via une procédure illustrée ici à travers l'exemple des feuilles de Reims et celle de Châlons, au sud de Reims.

Écarts avant géoréférencement La mise bord à bord de deux feuilles adjacentes avant géoréférencement montre que le raccord entre deux feuilles est globalement cohérent. Néanmoins, de nombreuses incohérences cartographiques demeurent, comme l'illustre la figure 1. Le constat porte d'abord sur l'absence d'élément en correspondance sur une feuille voisine tel qu'illustré figure 1.1 (non continuité de deux routes) et figure 1.4 (arrêt de la limite paroissiale, en pointillés). Également, une erreur de relevés topographiques ou une erreur cartographique est constatée au niveau du raccord, l'erreur se propageant de manière régulière (respectivement irrégulière) sur l'ensemble d'une zone, impliquant ainsi un décalage constant (respectivement variable), tel qu'illustré figure 1.2 (discontinuité de la route empierrée bordée d'arbres mais continuité assurée pour les autres éléments). Enfin, un changement d'orientation d'un même élément de part et d'autre du raccord est observé, tel qu'illustré figure 1.3.

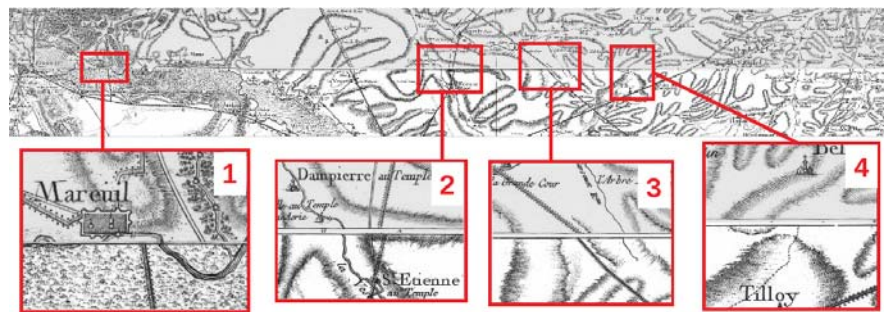


Figure 1. Exemples d'incohérences entre les feuilles Cassini de Reims et de Châlons.

Ajoutons que les incohérences de raccord entre les feuilles de Reims et de Châlons ne sont *a priori* pas relatives à des problèmes temporels, ces deux feuilles ayant toutes deux été réalisées lors de la première phase de la cartographie Cassini, entre 1750 et 1760.

Les erreurs topographiques ou/et cartographiques constatées sont de l'ordre d'une à quatre fois la largeur d'un symbole utilisé dans la légende Cassini. Dans l'exemple étudié, la largeur utile est généralement celle en relation avec le symbole d'une route. Cette dernière est comprise en moyenne entre 0.4mm pour les routes ordinaires (35 mètres dans la réalité) et 0.5mm pour les routes empierrées bordées d'arbres (43 mètres dans la réalité). L'erreur ainsi commise est comprise entre 35 mètres *a minima* et 172 mètres *a maxima*.

Écarts après géoréférencement Les écarts après géoréférencement sont maintenant étudiés. Pour ce faire, un repère relatif aux deux feuilles voisines considérées est défini. Il a pour point d'origine le point occidental commun aux deux feuilles, *i.e.* le point correspondant au coin sud-ouest de la feuille utile de Reims et le point correspondant au nord-ouest de la feuille utile de Châlons. L'axe horizontal est défini par la direction ouest-est et l'axe vertical par la direction sud-nord.

Des marqueurs verticaux placés le long de l'axe horizontal sont créés, espacés entre eux de 200 mètres, et ont une longueur suffisante pour figurer sur les deux feuilles. Ils sont ensuite géoréférencés de deux manières : une première fois en utilisant la transformation relative à la feuille de Reims et une seconde fois en utilisant la transformation relative à la feuille de Châlons. L'écart après géoréférencement entre ces marqueurs est ensuite comparé afin d'analyser l'impact dû au géoréférencement. Ces écarts varient en valeur absolue entre 0 et 179 mètres et évoluent de manière linéaire, ce qui est logique puisque c'est une transformation de type linéaire qui est utilisée pour le géoréférencement.

Discussion L'étude des écarts avant et après géoréférencement montre sur plusieurs exemples dont celui étudié ici que la méthode de géoréférencement employée a pour conséquence des décalages du même ordre de grandeur que les erreurs topographiques et/ou cartographiques liées aux feuilles de Cassini telles qu'initialement conçues. De fait, le raccord "exact" entre deux zones voisines pourra être obtenu à l'aide de méthodes d'intégration qui seront basées essentiellement sur l'alignement des géométries dans les zones tampons (fixes ou mobiles) définies autour des raccords de feuilles adjacentes. Ces méthodes seront *a priori* efficaces pour la majorité des feuilles traitées, à l'exception peut-être des feuilles situées dans des zones montagneuses, au vu des écarts constatés sur la feuille de Grenoble.

4. Appariement des données issues des cartes de Cassini avec la BDTOPO

Cette section détaille la méthode utilisée pour mettre en relation les données issues de la vectorisation des feuilles de Cassini avec celles de la BDTOPO de l'IGN, afin de créer des liens entre passé et présent. Nous développons notre étude sur les objets de nature "religieux" (église, prieuré, chapelle, etc.), "non religieux" (écart, hameau, etc.) et "moulin-activité-industrielle" (moulin à eau, à vent, forge, etc.), qui ont tous une représentation géométrique de type ponctuel.

L'évaluation du géoréférencement des cartes de Cassini a mis en évidence des décalages de position entre ces données, variant de 10 à 750 mètres, conséquence de l'imprécision des fonds de cartes. Cette imprécision peut s'expliquer par des approximations de localisation des objets lors du levé des cartes à l'aide de méthodes de triangulation, mais aussi par l'usage de taille de symboles relativement gros par rapport à l'échelle de la carte. Enfin, les positions relatives des objets ne sont pas toujours certaines, comme illustré par exemple figure 2, où un village indiqué au sud d'un hameau

8 SAGEO. Volume X - n° SAGEO/2012

par le passé se trouve aujourd'hui au nord de celui-ci. De fait, il apparaît que les techniques d'appariement privilégiant les relations géométriques entre objets ne sont pas nécessairement adaptées.

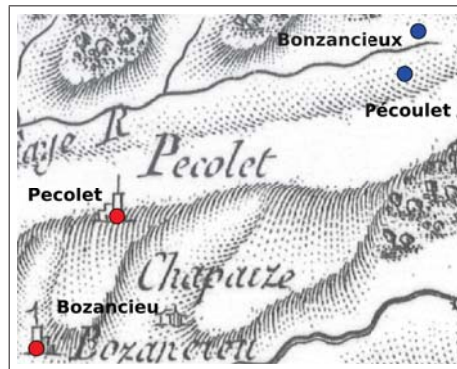


Figure 2. Modification de la position relative entre les villages de Bozancieux et de Pécoulet : Bozancieux est au nord de Pécoulet dans la BDTPO (en bleu) mais au sud sur la feuille Cassini de Grenoble (en rouge).

4.1. Travaux relatifs à l'appariement d'objets ponctuels

La plupart des processus d'appariement [WAL 99, SUI 04, BLA 04, ZHA 05, HAU 05, VOL 06, Lüs 07] ont une approche fondée uniquement sur un critère spatial. Après recalage et simplification des données, les objets à apparier sont sélectionnés de façon aléatoire, puis la définition d'une zone tampon (un *buffer*) autour de ces objets permet de filtrer les possibles candidats à l'appariement. Ce sont des mesures de distance, la distance euclidienne ou d'autres distances géométriques ([DEV 97, BEE 04, VOL 06, MUS 08]), qui interviennent ensuite pour décider de l'appariement effectif des candidats aux objets. Dans ces approches, le processus d'appariement est évalué par des mesures de confiance [CLO 07, MUS 08] ou en terme de précision et de rappel [BEE 04, SAF 06]. La *précision* des liens d'appariement (respectivement des objets non appariés) est le rapport entre le nombre de vrais positifs (respectivement de vrais négatifs) et la somme des vrais positifs et des faux positifs (respectivement des vrais négatifs et des faux négatifs). Le *rappel* des liens d'appariement (respectivement des objets non appariés) est le rapport entre le nombre de vrais positifs (respectivement de vrais négatifs) et le nombre de liens attendus, c'est-à-dire établis manuellement. L'objectif d'évaluation recherché est donc de déterminer le nombre de vrais positifs (liens d'appariement correctement établis), de vrais négatifs (objets correctement non appariés), de faux positifs (liens d'appariement établis par erreur) et de faux négatifs (objets non appariés par erreur).

Enfin, il existe peu de méthodes permettant un appariement d'objets à géométrie ponctuelle en dehors de celles proposées par [BEE 04] et [OLT 08]. L'approche basée

sur la théorie de l'évidence que propose [OLT 08] a l'avantage de tenir compte de l'imperfection des données et de combiner trois critères : géométrique, sémantique et toponymique. De façon classique, *le critère géométrique* évalue la proximité de deux objets avec la distance euclidienne. *Le critère sémantique*, qui mesure la similarité de nature entre les objets, utilise la distance de Wu-Palmer [WU 94] et s'appuie sur la taxonomie générale des objets géographiques de la BDTPOPO proposée par [ABA 08]. *Le critère toponymique* analyse la ressemblance lexicale entre deux toponymes, via l'usage de la distance de Levenshtein [LEV 65].

L'approche proposée par [OLT 08] repose sur cinq étapes (cf. figure 3.a) : sélection des candidats, initialisation des masses de croyance, fusion des critères, fusion des candidats et prise de décision. Un des problèmes majeur posé par cette approche réside dans l'étape d'initialisation des masses de croyance car elle nécessite un paramétrage délicat. En effet, la théorie des fonctions de croyance utilise les masses de croyance comme un filtre qui s'applique de façon particulière à chaque critère, filtre qui peut atténuer plus ou moins l'influence du critère et qui exprime le degré de croyance en l'appariement entre l'objet étudié et son candidat pour le critère mesuré. Ce filtre est généralement représenté par une fonction floue, qui doit préalablement être définie. [OLT 08] utilise à cette fin des fonctions linéaires mais précise qu'il serait intéressant d'étudier d'autres modélisations de fonctions. Par ailleurs, l'algorithme est très coûteux en temps de calcul.

4.2. Propositions d'améliorations pour l'appariement multi-critères basé sur la théorie des fonctions de croyance

Dans notre contexte d'intégration de données anciennes imparfaites, la méthode d'appariement multi-critères basée sur la théorie des croyances proposée par [OLT 08] apparaît comme étant la plus adaptée. Cependant sa complexité algorithmique est un inconvénient majeur. Nous proposons une approche simplifiée qui en reprend les principales étapes tout en s'affranchissant de la théorie de l'évidence. Notre algorithme apparie ainsi 35 fois plus vite à configuration matérielle équivalente (estimation réalisée à partir de 350 objets). Nous utilisons par la suite le terme de "poids" plutôt que celui de "masse de croyance" et l'étape de pondération des critères remplace celle d'initialisation des masses de croyance. Nous proposons également de nouvelles améliorations à cette approche – représentées en vert figure 3 – destinées à faciliter son paramétrage et l'analyse des résultats. Afin de vérifier les résultats, un appariement manuel est au préalable effectué. Cent points sont extraits aléatoirement sur trois zones pour chacune des classes étudiées, puis appariés manuellement, soit au total 900 objets Cassini, représentant un peu plus du dixième des données totales.

4.2.1. Sélection des candidats et évaluation des critères

Sélection des candidats Autour de chaque objet Cassini, une zone tampon, ou *buffer*, est définie afin de déterminer quels sont les objets de la BDTPOPO les plus probables de lui être appariés. La taille de ce *buffer* dépend de la zone étudiée, en

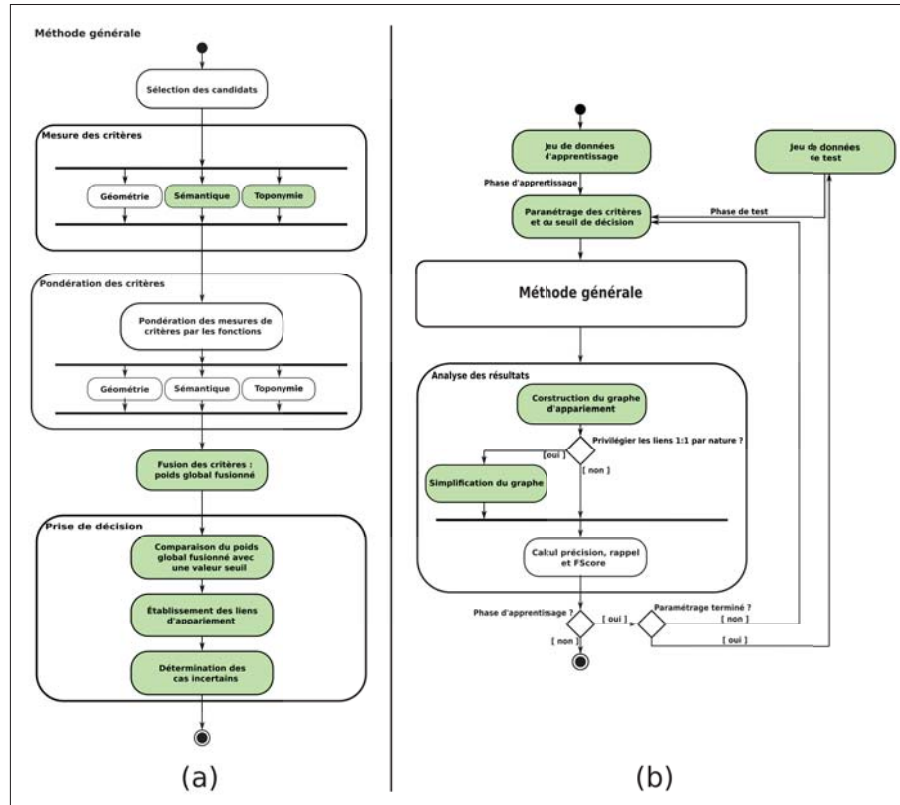


Figure 3. Méthode s'appuyant sur l'approche proposée par [OLT 08] (a) et méthode de paramétrisation automatique (b).

raison des différences de précision de chaque feuille Cassini, soulignées par les résultats du géoréférencement. Ces seuils ont été fixés empiriquement pour chaque zone en fonction de l'erreur maximale déterminée lors de l'appariement manuel, avec une marge d'erreur de 10% : 1400m pour Reims, 1800m pour Saint-Malo et 2600m pour Grenoble.

Modification de la distance sémantique De nombreux concepts saisis dans la base de données des objets Cassini n'existent soit pas du tout ("prieuré", "gentilhomme", etc.), soit de manière incomplète dans la taxonomie de [ABA 08] (pas de "moulin à vent en bois" mais seulement des "moulins à vent"). En l'absence d'une ontologie de domaine spatio-temporelle, couvrant les époques ici traitées, nous modifions le calcul de la distance sémantique utilisée par [OLT 08] de la manière suivante. D'une part, après lemmatisation des noms de nature d'objets, nous vérifions si chacun des lemmes ne sont pas présents dans la taxonomie, avant de calculer une distance

de Wu-Palmer minorée (en fonction de la proportion de lemmes absents de la taxonomie) ; ceci permet ainsi d'associer tous les moulins (à vent, à eau, à vent en bois, à vent en pierre) aux objets de nature "moulin" de la BDTPO. D'autre part, les toponymes des objets eux-mêmes sont analysés pour vérifier s'ils ne contiennent pas une nature d'objet présente dans la taxonomie. Des liens d'appariement sont ainsi établis entre un "moulin à eau" de Cassini et un lieu-dit de la BDTPO de nature "lieu-dit habité" et de toponyme "le moulin du clos", "moulin" étant un concept de la taxonomie.

Modification de la distance toponymique La toponymie joue un rôle primordial dans l'appariement des données anciennes mais de grandes différences orthographiques existent entre les toponymes actuels et ceux de la carte de Cassini. Comme leur comparaison exige de la souplesse dans la mesure du critère toponymique, la distance de Damerau-Levenshtein, extension plus tolérante de la distance de Levenshtein, est utilisée. De plus nous associons à la distance toponymique une valeur faible fixée à environ la moitié de la valeur maximale prise par le critère toponymique lorsque certains cas particuliers sont en jeu : (i) concaténation de termes, *e.g.* "Pont Aven" devenant "Pontaven" ; (ii) interversion de termes, *e.g.* "La Chapelle Felcourt" devenant "Felcourt la Chapelle" ; (iii) ajout ou suppression de termes, *e.g.* "Saint-Souplet" devenant "Saint-Souplet sur Py" ou "Hauteville" devenant "Hauteville-sur-Mer".

4.2.2. Modification de l'étape de prise de décision

Après que les critères aient été combinés, l'algorithme de [OLT 08] fusionne les candidats avant de choisir celui qui maximise une fonction de probabilité appelée probabilité pignistique. Nous préférons adopter un modèle de décision plus simple, en comparant la moyenne des poids des critères, appelée poids global fusionné, à un seuil défini préalablement, moins coûteux en terme de calculs.

4.2.3. Vers une paramétrisation adaptée de l'appariement

Comme les mesures de précision et de rappel sont aussi importantes l'une que l'autre, le protocole de paramétrisation automatique développé se base sur l'utilisation d'une fonction qui les combine équitablement, appelée *F-score* et définie comme suit :

$$F\text{-score} = 2 * \frac{\text{precision} * \text{rappel}}{\text{precision} + \text{rappel}} \quad (4)$$

Modélisation des fonctions L'influence du choix des fonctions de pondération, notées f_i , sur les résultats de l'appariement est maintenant étudié. Pour chaque critère, quatre mêmes familles de fonctions sont testées. Soit min et max les valeurs minimale et maximale de la fonction étudiée, et soit S le seuil pour lequel la fonction prend sa valeur maximale, les fonctions de pondération, notées respectivement f_1 , f_2 , f_3 et f_4 , vérifient $\forall i \in \llbracket 1; 4 \rrbracket, f_i(0) = min$ et $\forall i \in \llbracket 1; 4 \rrbracket, \forall x \geq S, f_i(x) = max$, et ont pour équations :

$$f_1(x) = \min + (\max - \min) \frac{1 - e^{-\frac{\lambda x}{S}}}{1 - e^{-\lambda}} \text{ si } x < S, \forall \lambda \in \mathbb{R} \quad (5)$$

$$f_2(x) = \min + (\max - \min) \frac{\ln(\frac{\lambda x}{S} + 1)}{\ln(\lambda + 1)} \text{ si } x < S, \forall \lambda \in \mathbb{R} \quad (6)$$

$$f_3(x) = \frac{\max - \min}{S} x + \min \text{ si } x < S \quad (7)$$

$$f_4(x) = \min + (\max - \min) \sqrt{\frac{x}{S}} \text{ si } x < S \quad (8)$$

La figure 4 illustre les courbes représentatives de ces fonctions de pondération :

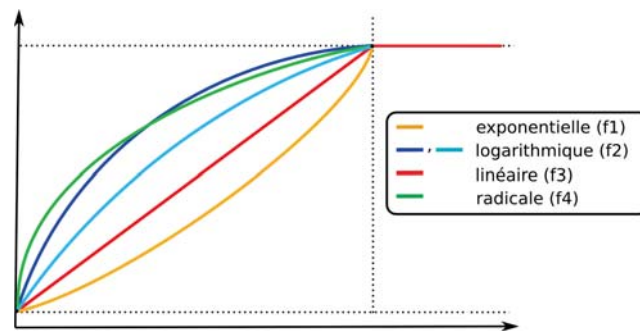


Figure 4. Courbes représentatives de quelques fonctions de pondération.

Le tableau 2 présente pour différentes combinaisons de fonctions et pour une série de paramètres fixés arbitrairement, les résultats du processus d'appariement sur les données appariées manuellement de la classe "non religieux" de la feuille de Reims.

Courbes critères [géométrique, sémantique, toponymique]	Précision LAP	Rappel LAP	Précision NAP	Rappel NAP	F-score moyen
[EXP($\lambda=1$), LOG($\lambda=8$), LINEAIRE]	97%	97%	96%	100%	97%
[EXP($\lambda=1$), LOG($\lambda=8$), EXP($\lambda=1$)]	95%	91%	91%	100%	94%
[EXP($\lambda=1$), EXP($\lambda=8$), EXP($\lambda=1$)]	87%	93%	94%	96%	92%
[LINEAIRE, LINEAIRE, LINEAIRE]	70%	97%	97%	76%	83%
[RADICALE, RADICALE, RADICALE]	100%	72%	78%	100%	86%
[LOG($\lambda=8$), LOG($\lambda=8$), LOG($\lambda=8$)]	100%	72%	78%	100%	86%

Tableau 2. Résultats d'appariement (classe "non religieux", feuille de Reims) pour différentes combinaisons de fonctions : précisions et rappel des liens d'appariement (LAP), précision et rappel des objets non appariés (NAP), F-score moyen. Paramètres utilisés pour les obtenir : geoMin=0.2, geoMax=0.5, geoS=1000, semMin=0.0, semMax=0.6, semS=0.3, topMin=0.1, topMax=0.7, topS=0.9, seuil=0.41.

Nous constatons que la qualité de la procédure d'appariement dépend du choix des fonctions de pondération. De plus, dans l'exemple ci-contre, les meilleurs résultats sont obtenus par l'utilisation d'une courbe différente par critère : une exponentielle pour la géométrie, une logarithmique pour la sémantique et une linéaire pour la toponymie. Pour un critère donné, le type de courbe déterminé par la paramétrisation dépend peu de la zone considérée, mais surtout de la classe étudiée. Les paramètres du processus d'appariement à définir sont donc le seuil global de prise de décision et, pour chaque critère, le type et les valeurs *min*, *max* et *S* de sa fonction de pondération. Pour trois critères étudiés, il y a donc 13 paramètres à déterminer ($1 + 4 * 3$).

Paramétrisation automatique des fonctions et détermination du seuil Dans la logique d'évaluation d'un appariement de données géographiques, il convient de s'intéresser autant aux liens d'appariement qu'aux objets non appariés, notamment dans un contexte d'appariement de données anciennes. Le processus de paramétrisation automatique repose sur un apprentissage supervisé. La base d'exemples des appariements manuels est scindée en deux : une base d'apprentissage et une base de tests (respectivement 2/3 et 1/3 des données). Le processus est itératif et prend en entrée un jeu de paramètres à tester. La moyenne des *F-scores* pour les liens d'appariement et pour les objets non appariés, que nous appellerons par la suite *F-score* moyen, est recalculée pour chaque nouveau paramètre. Les paramètres d'appariement recherchés sont ceux qui maximisent le *F-score* moyen dans la phase d'apprentissage. La base de tests sert à valider la paramétrisation issue de l'apprentissage.

Méthode d'analyse des résultats La notion de cardinalité des liens d'appariement est délicate à traiter. Il existe des liens *1:1* – un objet Cassini apparié avec un objet de la BDTOP –, *1:n* – un objet Cassini apparié avec plusieurs objets de la BDTOP – et *n:m* – plusieurs objets Cassini appariés avec plusieurs objets de la BDTOP –. Nous appelons lien *1:1* (ou *1:n* ou *n:m*) par nature un lien *1:1* (ou *1:n*, *n:m*) entre objets de même nature. Nous proposons une méthode d'étude locale des liens d'appariement par nature d'un objet en construisant le graphe d'appariement de l'objet relatif à une nature *N* donnée de candidat de proche en proche, tel qu'illustré figure 5.

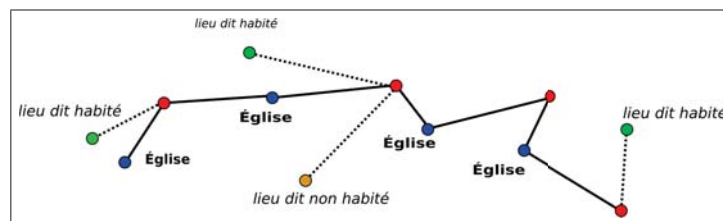


Figure 5. Exemple de graphe d'appariement (en trait plein) pour une nature "Église". Les objets Cassini sont en rouge.

Ce *graphe d'appariement* correspond au graphe formé par l'ensemble des liens d'appariement (les arcs), des objets et des candidats de nature N (les nœuds) tel qu'il existe un chemin entre l'objet initial et les autres objets ou candidats. Ce graphe permet d'étudier le contexte local d'un objet étudié en visualisant les objets et candidats en conflit.

Privilégier les liens $1:1$ par nature peut s'avérer nécessaire. Il est par exemple courant qu'une église Cassini soit appariée avec plusieurs lieux-dits différents au terme de l'étape de prise de décision. Nous proposons de simplifier ces graphes d'appariement conflictuels en choisissant comme meilleur candidat d'un objet Cassini de référence, le plus proche voisin du ponctuel de la BDTOPO portant un toponyme proche de celui de l'objet étudié (cf. figure 6.c). Dans les cas où la toponymie ne peut être utilisée, les arêtes de chaque graphe sont pondérées par la valeur du poids global fusionné associée à ce lien. Nous ne conservons alors que les n_{min} liens dont la somme des pondérations est minimale, où $n_{min} = \min(|Objets|, |Candidats|)$ avec $|E|$ représentant le cardinal de l'ensemble E .

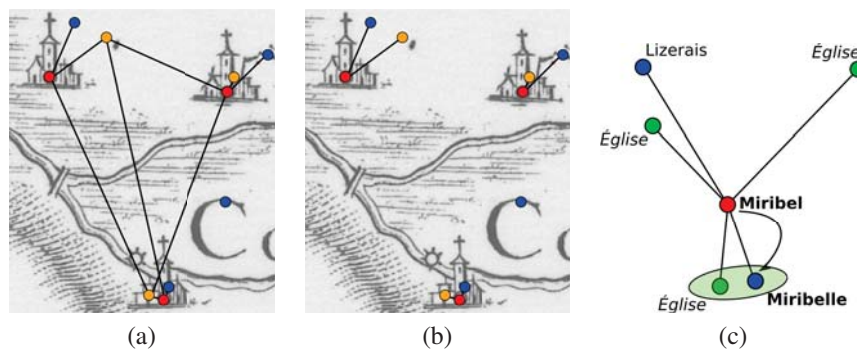


Figure 6. Graphe d'appariement pour la nature "Église" avant (a) et après (b) simplification. Les églises Cassini sont en rouge, les églises de la BDTOPO en orange, les lieux-dits de la BDTOPO en bleu. La figure (c) montre la prédilection des candidats proches du meilleur candidat au sens de la distance toponymique.

4.2.4. Résultats et discussions

Le tableau 3 montre les résultats obtenus pour chaque classe de chaque zone étudiée. Les résultats sont globalement bons, le F -score moyen variant entre 91% et 100% sur les données de validation. Sur les trois zones, nous ne constatons pas de fort différentiel de précision et de rappel entre classes, malgré des proportions d'objets aujourd'hui disparus très variées : faible pour la classe "religieux" (entre 10% et 30%), moyenne pour la classe non "religieux" (entre 35% et 55%) et importante pour la classe "moulins" (entre 65% et 80%).

Nous avons également créé, pour chaque classe, une base contenant l'union des objets des trois zones, et exécuté l'algorithme de paramétrisation sur ces données.

Zone	Classe	Apprentissage					Validation				
		Préc. LAP	Rapp. LAP	Préc. NAP	Rapp. NAP	F-score moyen	Préc. LAP	Rapp. LAP	Préc. NAP	Rapp. NAP	F-score moyen
Reims	Religieux	100%	97%	94%	100%	98%	100%	97%	100%	100%	99%
	Non religieux	100%	95%	95%	100%	97%	91%	100%	100%	100%	98%
	Moulins	83%	83%	98%	98%	91%	83%	100%	100%	96%	95%
St-Malo	Religieux	99%	97%	75%	100%	92%	100%	98%	100%	99%	99%
	Non religieux	100%	92%	83%	100%	93%	100%	88%	75%	100%	90%
	Moulins	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Grenoble	Religieux	99%	95%	82%	100%	93%	97%	99%	100%	67%	90%
	Non religieux	98%	100%	93%	96%	97%	100%	100%	100%	100%	100%
	Moulins	100%	93%	98%	100%	98%	100%	100%	100%	100%	100%

Tableau 3. *Résumé des résultats d'appariement.*

Les bons résultats obtenus (cf. tableau 4) nous permettent d'envisager la définition d'une paramétrisation unique par classe, qui serait hypothétiquement applicable à l'ensemble du territoire français, au seuil de prise de décision près. Nous proposons une méthode alternative à l'apprentissage permettant de déterminer automatiquement la valeur de ce seuil, basée sur l'étude de *ruptures* dans la classification des natures des candidats appariés. Pour cela, nous analysons les valeurs du seuil pour lesquelles une nouvelle nature (valeur de l'attribut "nature") de candidat est mise en correspondance avec un objet Cassini. La valeur de seuil pour laquelle une nature d'objet non désirée apparaît (*i.e.* une nature correspondant à une erreur d'appariement, par exemple un moulin apparié avec un donjon) peut être utilisée pour l'appariement, en tant que valeur limite maximale.

Classe	Apprentissage					Validation				
	Préc. LAP	Rapp. LAP	Préc. NAP	Rapp. NAP	F-score mean	Préc. LAP	Rapp. LAP	Préc. NAP	Rapp. NAP	F-score mean
Religieux	97%	97%	94%	97%	96%	97%	96%	88%	82%	91%
Non religieux	96%	95%	91%	97%	95%	93%	90%	87%	92%	91%
Moulins	93%	98%	99%	99%	97%	95%	90%	99%	100%	96%

Tableau 4. *Résultats d'appariement par classe pour la paramétrisation globale sur les trois zones.*

5. Conclusion et perspectives

L'intégration de données anciennes comme les cartes de Cassini, présentant des écarts d'actualité de plus de 250 ans avec le monde d'aujourd'hui, est un sujet difficile, du fait des fortes évolutions sémantique et toponymique des objets, et de l'imprécision de localisation des cartes utilisées. Pour répondre à cette problématique, que nous avons détaillé au long de l'article, nous avons proposé une méthode de géoréférencement adaptée, qui tient compte des déformations propres à chaque feuille de la carte. Nous avons par ailleurs démontré que si la transformation de Helmert utilisée peut induire de légers décalages, ceux-ci sont tout à fait comparables aux inexactitudes

observées dans la carte. Ensuite, à partir de l'approche proposée par [OLT 08], nous avons proposé un processus d'appariement multi-critères, dont le paramétrage est réalisé par apprentissage à partir de données manuellement appariées, et qui permet l'analyse visuelle des liens de correspondance entre objets étudiés et leurs candidats.

Concernant l'appariement, nos perspectives à court terme concernent l'élaboration d'une ontologie de domaine spatio-temporelle qui devrait permettre d'améliorer l'évaluation de la distance sémantique, en tenant compte des époques des objets à appairer. Par ailleurs, certains toponymes d'objets homologues présentent d'importantes modifications orthographiques ("Le Ménéil L'Epinoy" devenant "Ménéil-Lepinois"). Afin de réduire les erreurs dues à la toponymie, nous envisageons d'exploiter une base de données recensant les évolutions des toponymes des communes de France [MOT 03].

A plus long terme, l'étude des correspondances avec des données intermédiaires comme les minutes d'État-Major ou les fonds de carte de 1960 de l'IGN pourra permettre de valider ou d'infirmer les résultats obtenus par notre approche, mais aussi de détecter des erreurs de représentation et de contenu sur les cartes anciennes entraînant des incohérences lors de l'appariement.

6. Bibliographie

- [ABA 08] ABADIE N., MUSTIÈRE S., « Création d'une taxonomie géographique à partir des spécifications de bases de données », *Actes de SAGEO 2008*, Montpellier, 2008.
- [BEE 04] BEERI C., KANZA Y., SAFRA E., SAGIV Y., « Object Fusion in Geographic Information Systems », *Proceedings of the 30th VLDB Conference*, Toronto, 2004.
- [BER 98] BERTHAUT H.-M.-A., *La carte de France 1750-1898, étude historique*, Impression du Service géographique, 1898.
- [BLA 04] BLASBY D., DAVIS M., KIM D., RAMSEY P., « GIS conflation using open source tools », Rapport technique, 2004.
- [CLO 07] CLODOVEU A. D., FONSECA F., « Assessing the Certainty of Locations Produced by an Address Geocoding System », *GeoInformatica*, vol. 11, n° 1, 2007, p. 103-129.
- [DEV 97] DEVOGELE T., « Processus d'intégration et d'appariement de bases de données géographiques. Application à une base de données routières multi-échelles », Thèse de doctorat, Université de Versailles, 1997.
- [DUP 07] DUPOUEY J.-L., BACHACOU J., COSSERAT R., ABERDAM S., VALLAURI D., CHAPPART G., CORVISIER DE VILLÈLE M.-A., « Vers la réalisation d'une carte géoréférencée des forêts anciennes de France », *Revue du Comité Français de Cartographie (CFC)*, vol. 191, 2007, p. 85-98.
- [GRO 11] GROSSO E., « Géoréférencement des feuilles Cassini (L2.1-1) », Rapport technique, 2011, Projet ANR GéoPeuple, Institut Géographique National, Laboratoire COGIT.
- [HAU 05] HAUNERT J., « Link based Conflation of Geographic Datasets », *In Proceedings of the 8th ICA Workshop on Generalisation and Multiple Representation*, la Corogne, 7-8 juillet 2005.

- [LEV 65] LEVENSHTAIN V., « Binary codes capable of correcting deletions, insertions and reversals », *Doklady Akademii Nauk SSSR*, vol. 163, n° 4, 1965, p. 845–848.
- [LEY 05] LEYK S., « Computing the Past - Utilizing Historical Data Sources for Map-Based Retrospective Landscape Research », Thèse de doctorat, Universität Zürich, 2005.
- [Lüs 07] LÜSCHER P., BURGHARDT D., WEIBEL R., « Matching road data of scales with an order of magnitude difference », *In Proceedings of the XXIII International Cartographic Conference*, Moscou, 4-10 août 2007.
- [MOT 03] MOTTE C., SÉGUY I., THÉRÉ C., TIXIER-BASSE D., *Communes d'hier, communes d'aujourd'hui. Les communes de la France métropolitaine, 1801-2001. Dictionnaire d'histoire administrative*, Institut National d'Études Démographiques, Paris, 2003.
- [MUS 08] MUSTIÈRE S., DEVOGELE T., « Matching networks with different levels of detail », *GeoInformatica*, vol. 12, n° 4, 2008, p. 435–453.
- [NOI 11] NOIZET H., GROSSO E., « The ALPAGE project : Paris and its suburban area at the intersection of history and geography (9th-19th century) », *25th International Cartographic Conference (ICC'11)*, Paris (France), 3-8 juillet 2011.
- [OLT 08] OLTEANU A.-M., « Fusion de connaissances imparfaites pour l'appariement de données géographiques. Proposition d'une approche s'appuyant sur la théorie des fonctions de croyance », Thèse de doctorat, Université Paris-Est, 2008.
- [PEL 90] PELLETIER M., *La carte de Cassini, l'extraordinaire aventure de la carte de France.*, Presses de l'école nationale des Ponts et Chaussées, 263 p., 1990.
- [SAF 06] SAFRA E., KANZA Y., SAGIV Y., DOYTSHER Y., « Efficient Integration of Road Maps », PRESS A., Ed., *In Proceedings of the 14th annual ACM international symposium on Advances in geographic information systems*, Arlington (Etats-Unis), 10-11 novembre 2006, p. 59–66.
- [SUI 04] SUI H., LI D., GONG J., « Automatic feature-level change detection (FLCD) for road network », *In Proceedings of the 20th ISPRS Congress*, Istanbul, 12-23 juillet 2004.
- [VOL 06] VOLTZ S., « An Iterative Approach for Matching Multiple Representations of Street Data », *In Proceedings of ISPRS Workshop, Multiple representation and interoperability of spatial data*, Hanovre (Allemagne), 22-24 février 2006, p. 101–110.
- [WAL 99] WALTER V., FRITSCH D., « Matching Spatial Data Sets : Statistical Approach », *International Journal of Geographical Information Science*, vol. 13, n° 5, 1999, p. 445–473.
- [WU 94] WU Z., PALMER M., « Verb Semantics and Lexical Selection », *In Proceedings of the 32nd Annual Meetings of the Associations for Computational Linguistics*, 1994, p. 133–138.
- [ZHA 05] ZHANG M., SHI W., MENG L., « A generic matching algorithm for line networks of different resolutions », *In Proceedings of ICA Workshop on Generalisation and Multiple Representation*, La Corogne, 7-8 juillet 2005.

Analyse des structures des haies et des linéaires pérennes dans deux paysages agricoles contrastés

S. Da Silva^{*,**} — C. Lavigne^{**} — F. Le Ber^{*,***}

^{*} LORIA UMR 7503, 54500 Vandœuvre-lès-Nancy
sebastien.da-silva@loria.fr

^{**} INRA, UR 1115 Plantes et Systèmes de culture Horticoles, 84000 Avignon

^{***} LHyGeS UMR 7517 - Université de Strasbourg/ENGES - CNRS, 67000
Strasbourg

RÉSUMÉ. Dans cet article, nous nous intéressons aux linéaires structurant les paysages agricoles (routes, canaux et haies). Nous étudions différentes caractérisations possibles des structures des haies, leur interaction avec les linéaires pérennes et présentons des résultats pour deux paysages contrastés. En particulier nous étudions comment et à quelle échelle la distribution spatiale des haies est structurée par des linéaires plus pérennes. Nos premiers résultats montrent que les structures des haies sont dépendantes des linéaires pérennes à courte distance et que le voisinage des haies devient uniforme au-delà de 150 m. De plus nous mettons en évidence différentes structures de voisinage selon les orientations des haies et le paysage considéré.

ABSTRACT. In this paper we focus on lines structuring the agricultural landscape (roads, channels and hedgerows). We investigate ways to characterize hedgerow structures, their interaction with perennial linear and show results for two contrasting agricultural landscapes. In particular, we assess whether, and under what circumstances, the spatial distribution of hedgerows is structured by the more perennial linear landscape elements and the scale of such structures. First results show that hedgerow structures were dependent on perennial lines at short distance and that the neighbourhood of hedgerows is uniform beyond 150 meters. Furthermore different structures of neighbourhood were highlighted according to the main orientations of hedgerows and the landscape considered.

MOTS-CLÉS : Modélisation de paysage, Analyse spatiale, Structure de haies, Voisinage, Segment

KEYWORDS: Landscape modelling, Spatial analysis, Structure of hedgerows, Neighbourhood, Segment

1. Introduction

La modélisation et la simulation de paysages connaissent de nombreux domaines d'application, comme la géographie, l'écologie, ou l'agronomie. La simulation est utilisée pour étudier l'influence des motifs paysagers sur les phénomènes écologiques [TUR 91] ou les liens entre les activités humaines, la configuration du paysage et les problèmes environnementaux [CAR 02]. La simulation de paysage peut être basée sur des données réelles ou sur des données aléatoires. Celles-ci sont très utilisées pour trois raisons majeures. Premièrement, les données réelles ne sont pas forcément disponibles sur le territoire visé ou alors trop spécifiques pour obtenir un modèle de simulation généraliste. Deuxièmement, dans le cas des paysages agricoles, il est nécessaire d'envisager de nouvelles configurations dans le but d'en déterminer les effets ou de trouver la meilleure disposition par rapport à un processus agro-écologique [GRO 10]. Troisièmement, les paysages simulés peuvent être utilisés pour tester la sensibilité des modèles de processus à la variabilité du paysage agricole.

Les modèles de paysages aléatoires peuvent porter sur différents objets à différents niveaux. Les modèles aléatoires utilisés en écologie du paysage sont principalement appuyés sur des approches raster [SAU 00]. Plus récemment, une approche vecteur utilisant un pavage géométrique a été proposée pour simuler un paysage agricole [GAU 06, Le 09]. Dans cette dernière approche, le paysage réel est caractérisé par la distribution de centroïdes qui sont utilisés pour construire le pavage. Cette approche modélise les motifs paysagers (parcelles vues comme des polygones) en contrôlant leurs caractéristiques (taille moyenne, périmètre moyen,...). Mais, le paysage n'est pas seulement constitué d'éléments polygonaux, mais également de linéaires (haies, bandes enherbées,...) qui ne sont pas toujours connectés en un réseau. Alors que l'analyse et la simulation de réseaux sont des questions majeures dans l'analyse des données spatiales [DES 07], peu d'études s'intéressent aux seuls linéaires et à leur organisation spatiale. La plupart de ces travaux concernent les relations entre haies, fossés et distribution d'animaux [ARN 83] ou de plantes [CAM 09].

Nous nous intéressons dans cette étude aux relations qu'il peut exister entre les haies et leur voisinage et plus particulièrement avec les éléments pérennes du paysage, tels que les routes et les canaux ou fossés. Nous cherchons à déterminer s'il existe un effet structurant de ces linéaires pérennes sur les haies et si cette influence dépend des caractéristiques des haies considérées. Nous chercherons également à définir les distances de cette influence, si elle existe. Les caractéristiques des haies considérées *a priori* sont l'orientation des haies et leur longueur car elles jouent un rôle important vis-à-vis des ravageurs, des maladies et des flux de pollens entre parcelles [PAS 88]. Ces analyses seront menées sur deux paysages contrastés et eux-mêmes hétérogènes en terme de densité d'éléments linéaires.

La démarche adoptée consiste alors à caractériser les haies et leur voisinage, selon les longueurs, les orientations et les distances entre ces différents linéaires. Pour cela nous les considérons comme des segments. Nous nous démarquons ainsi de la majorité des travaux existants, qui portent sur l'extraction de réseaux (routes ou canaux) à partir de photos satellitaires [STO 01, MON 05, LAC 05]. Dans ces approches, les réseaux sont extraits en utilisant des modèles aléatoires qui sont affinés en fonction des données auxquelles ils s'appliquent : sont pris en compte localement les valeurs des pixels, l'orientation et la largeur du réseau (cf. le modèle Candy pour [STO 01]). Des travaux plus récents ont utilisé des approches similaires pour extraire des haies à partir de photos satellitaires à haute résolution [VAN 08, VAN 10]. Dans ce cas, les critères pris en compte sont la composition des pixels, leur arrangement spatial ainsi que la morphologie des objets. Nous travaillons sur des données issues de systèmes d'informations géographique et donc le problème d'extraction ne se pose pas.

Le plan de l'article est le suivant. Dans la section 2 nous présentons les deux jeux de données utilisés et les prétraitements que nous leur avons appliqués. La section suivante s'attache à la caractérisation globale et locale des linéaires. Dans la section 4, nous proposons une méthodologie pour caractériser des voisinages des haies. Les résultats obtenus sur les deux jeux de données sont présentés et discutés. Dans la conclusion nous évoquons quelques perspectives concernant la simulation des linéaires, appuyée sur les résultats présentés ici.

2. Données et prétraitements

Nous travaillons sur deux jeux de données, issus de deux territoires français distincts. Le premier, (données A), est fourni par l'équipe Écologie de la Production Intégrée (EPI) de l'unité PSH (Plantes et Systèmes de culture Horticoles) à l'INRA PACA. Ces données sont à notre disposition dans le cadre des études menées sur le rôle des haies sur les populations de ravageurs [RIC 09, RIC 11]. Le second (données B) nous est fourni par l'UMR COSTEL de Rennes, il est choisi au regard de sa différence marquée avec le premier tant sur le plan géographique qu'agricole.

2.1. Description du milieu concerné

Les zones d'où proviennent les jeux de données sont très différentes. Les données A fournies par l'unité PSH représentent un paysage dans le Sud-Est de la France, au nord du département des Bouches du Rhône (basse vallée de la Durance : coordonnées en WGS84 : de 43°46'27" N à 43°51'23" N et de 4°51'12" E à 4°57'34" E). Il s'étend sur près de 70 km² et comprend majoritairement des vergers (pommiers et poiriers) pour 70% de la SAU, ainsi que du maraîchage. Cette région est caractérisée par une forte densité de haies brise-vent, essentiellement composées de cyprès ou peuplier. Les autres haies présentes dans le

4 SAGEO 2012. Volume /2012

paysage sont composées de platanes en bord de route, ou sont composites. Les données contiennent également les informations relatives aux réseaux routiers et hydriques fournies par l'IGN.

En Bretagne, les données B ont été collectées sur la Zone Atelier Armorique et la zone étudiée couvre environ 120 km^2 dans la région de Pleine Fougère, au nord-est de Rennes (de $48^\circ 25' 32'' \text{N}$ à $48^\circ 34' 06'' \text{N}$ et de $1^\circ 31' 39'' \text{O}$ à $1^\circ 39' 07'' \text{O}$). Nous avons à notre disposition le fichier contenant les coordonnées et caractéristiques des linéaires de haies et, comme pour l'autre région, ceux concernant les réseaux routiers et hydriques produits également par l'IGN. Le sud de la zone est constitué de bocage historique contrairement au nord remembré durant les dernières années. Les haies sont réparties en quatre catégories, à savoir : végétation éparse, haie arbustive, haie arborée discontinue et haie arborée continue.

2.2. Production des données

En basse vallée de la Durance (données A), les haies ont été numérisées manuellement à partir d'une orthophoto aérienne (source IGN) dans un système d'informations géographiques. Elles représentent donc l'implantation réelle des haies, avec prise en compte de leurs discontinuités. Le fichier vectoriel est constitué de polygones. Pour le territoire Breton, les limites administratives des parcelles ont été dessinées à partir du cadastre numérique puis chacune d'elle a été notée soit comme vide, soit comme appartenant à une des quatre catégories de haies définies précédemment grâce à une photo prise lors d'un vol ULM (2006). Le fichier vectoriel ainsi formé comporte également des polygones.

2.3. Prétraitements

2.3.1. Découpage en segments

Nous avons segmenté de la même façon les polygones des haies et des réseaux routiers et hydriques afin de garantir l'homogénéité de la forme des données. Chaque segment est attaché à une catégorie (haie, route ou canal).

Le découpage en segments a eu des effets différents selon les régions : il y a seulement 56 segments de plus que de polygones en Basse vallée de la Durance alors qu'à Pleine-fougères, le nombre de segments des haies est le double du nombre de polygones (Tableau 1).

En ce qui concerne les routes et les canaux, moins nombreux que les haies, l'effet du découpage en segments est encore plus marqué, le ratio nombre de segments sur nombre de polygones allant de 5 (routes de A) à 28 (canaux de B).

Tableau 1. *Effet du découpage des polygones en segments sur le nombre d'éléments par catégorie*

Paysage	A		B	
Type de Données	Polygones	Segments	Polygones	Segments
Haies	11501	11557	7561	14819
Canaux	183	2014	328	9152
Routes	796	3737	2516	43416

2.3.2. Découpage en cellules

Nous avons choisi de découper chaque zone d'étude en cellules pour deux raisons principales : (i) le découpage permet de mettre en évidence l'hétérogénéité de chacun des paysages étudiés ; (ii) cela permet de réduire la taille des données et donc le temps de calcul. Nous avons découpé la fenêtre initiale de chaque paysage (carré circonscrit aux données) en cellules carrées mais en créant un chevauchement de toutes les cellules voisines égal à 10% de leur taille. Il est nécessaire pour la suite du travail de bien choisir la taille des cellules, pour cela, nous avons, sur les données A, procédé au découpage pour des tailles de cellule variant entre 200 et 1300 m de côté et, nous avons déterminé, pour chaque valeur, la distribution du nombre de haies par cellule.

Nous avons ensuite tracé les histogrammes représentant le nombre de cellules en fonction du nombre de haies à l'intérieur. Nous avons sélectionné le découpage en cellules de taille 1100 m car il présentait un nombre de haies par cellule suffisant et permettait de recouvrir la zone avec un nombre de cellules suffisamment grand pour montrer la diversité des configurations dans les paysages traités. L'objectif est en effet également de faire un juste équilibre entre le nombre de haies par cellule et le nombre de cellules par paysage. La même taille de cellule a été choisie pour les données B.

3. Caractérisation des données

3.1. A l'échelle des paysages

L'objectif à cette échelle est de construire des classes de haies pertinentes à partir des paramètres considérés, ici la longueur et l'orientation. La longueur est une variable simple à obtenir une fois les données produites ; l'orientation d'un segment est définie par rapport à l'axe de référence nord-sud.

Nous nous intéressons principalement au paramètre orientation, la variable longueur n'ayant pas permis d'obtenir des résultats pertinents. Sur la figure 1(a), nous présentons l'histogramme circulaire des angles pour les données A. Les angles étant calculés par rapport à la verticale, nous avons effectué une rotation de $\frac{\pi}{2}$ afin d'avoir une représentation des angles équivalente à leur disposition

sur le terrain. Nous constatons que la distribution n'est pas uniforme et qu'il se dégage 3 classes d'angles différentes. Tout d'abord, une première classe proche de l'horizontale, ayant pour direction l'axe (ouest-nord-ouest – est-sud-est). Celle-ci peut être considérée comme la classe des haies brise-vent (notée HV). La seconde classe (notée HP) est, semble-t-il, perpendiculaire à la première, et s'oriente selon l'axe (nord-nord-est – sud-sud-ouest). La dernière classe est composée du reste des valeurs. Nous avons donc défini les trois classes d'angle suivantes $[0, \frac{\pi}{9}] \cup [\frac{8\pi}{9}, \pi]; [\frac{\pi}{3}, \frac{5\pi}{9}];]\frac{\pi}{9}, \frac{\pi}{3}[\cup]\frac{5\pi}{9}, \frac{8\pi}{9}[(\text{modulo } \pi)$.

Nous avons effectué les mêmes études pour les données B. Bien qu'il n'existe pas un vent fort dominant dans cette région de Bretagne, l'orientation des haies présente elle aussi des classes (figure 1(b)). Tout d'abord, il existe une première classe d'angles autour de l'axe nord-sud avec un nombre élevé de haies. La seconde classe se situe autour de l'axe est-ouest. La troisième classe d'angle contient le reste. En accord avec les données A, nous notons les haies de la première classe (HP) et celles de la seconde (HV). Nous avons donc, pour les données B, les trois classes d'angles suivantes $[0, \frac{\pi}{9}] \cup [\frac{8\pi}{9}, \pi]; [\frac{7\pi}{18}, \frac{11\pi}{18}];]\frac{\pi}{9}, \frac{7\pi}{18}[\cup]\frac{11\pi}{18}, \frac{8\pi}{9}[(\text{modulo } \pi)$.

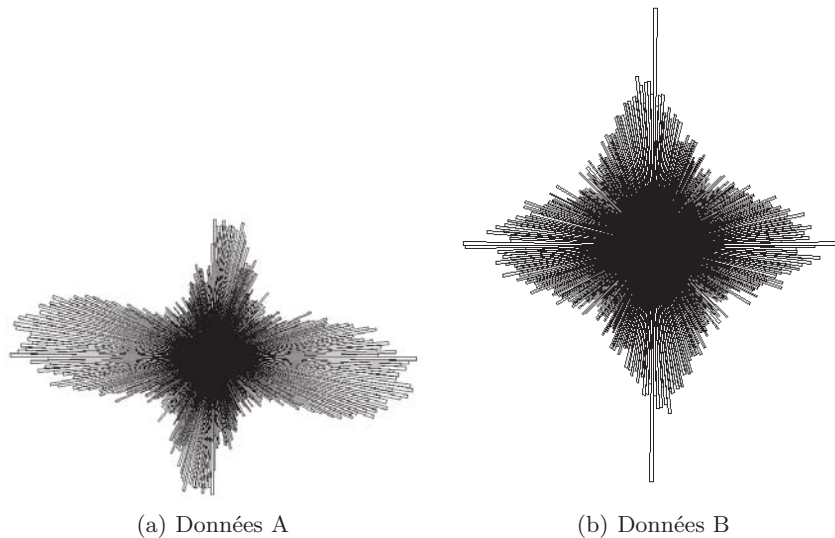


Figure 1. *Histogrammes circulaires des angles pour les deux jeux de données*

3.2. A l'échelle locale

Pour caractériser les haies à l'échelle locale, nous avons travaillé sur un sous-ensemble de cellules de chaque paysage. Par la suite, nous appellerons ces cellules, des cellules cibles.

3.2.1. *Choix des cellules cibles*

Pour choisir les cellules cibles, nous avons représenté les paysages A et B découpés en cellules de 1100 m de côté avec, pour chacune d'elle, son nombre d'éléments total (NB), de haies (H), de canaux (C) et de routes (R). Nous avons prévu d'utiliser des cellules ayant différentes densités de haies et de routes (faible, moyenne, forte), mais les cellules ayant trop peu d'éléments n'étaient pas utilisables : elles-mêmes, ou leurs voisinages, se trouvaient en bordure de paysage et donc, des effets de bord étaient trop forts. De plus, le manque d'éléments rendait la détection d'une structure impossible.

Nous avons donc choisi pour chacun des paysages, des cellules ayant des densités de haies différentes (moyenne ou forte) et des densités d'éléments pérennes quelconques. Nous avons vérifié également que chaque zone cible (constituées par la cellule cible et ses 8 cellules attenantes) soit suffisamment pourvue en éléments (densité moyenne ou forte) et que les cellules cibles ne se situent pas au bord du domaine de nos données afin d'éliminer les possibles effets d'absence de voisins. Sous ces hypothèses et de manière empirique, nous avons choisi 8 cellules, décrites dans le tableau 2.

Tableau 2. *Récapitulatif du nombre d'éléments par cellule cible et zone cible*

Cellule	Elts dans la cellule cible			Elts dans la zone cible		
	Haies	Routes	Canaux	Haies	Routes	Canaux
A1_9	148	19	69	1241	263	531
A2_2	202	27	60	1949	186	415
A8_3	208	33	25	1457	222	257
A8_9	130	19	39	1711	158	492
B2_7	179	28	388	1282	220	3924
B6_3	137	81	433	1486	498	3199
B6_7	299	96	394	2207	992	4027
B8_3	318	103	639	2789	616	4806

Deux exemples des cellules choisies sont donnés sur la figure 2 (paysage A) et la figure 3 (paysage B). La cellule cible est représentée au cœur de sa zone, incluant ses huit voisins.

4. Etude des voisinages de haies

Nous nous intéressons maintenant aux relations entre haies et éléments pérennes dans le but de caractériser des structures à différentes distances.

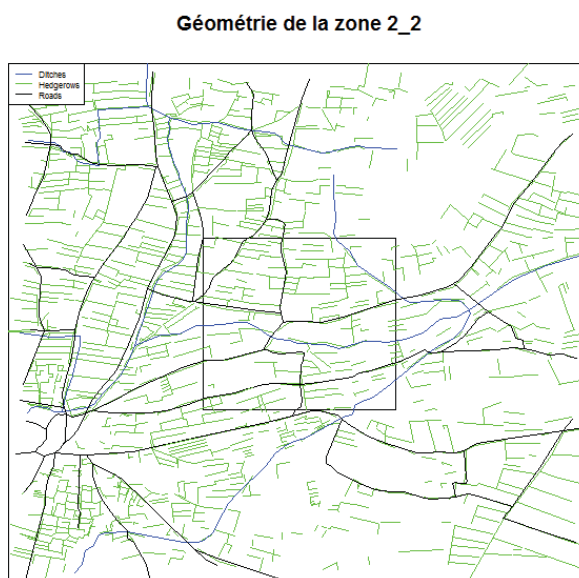


Figure 2. *Géométrie de la zone cible centrée sur la cellule 2_2 du paysage A*

4.1. *Méthodologie*

4.1.1. *Distance entre segments*

Les haies sont considérées comme des segments, donc, comme des ensembles compacts de \mathbb{R}^2 . La distance usuelle entre ensembles dans \mathbb{R}^2 est la distance de Hausdorff, définie comme la distance maximum entre les points des deux ensembles. Or nous cherchons à mettre en évidence des proximités entre haies (et entre linéaires pérennes) et de ce fait, nous n'utiliserons pas cette distance. Nous pourrions alors calculer une distance entre les barycentres des segments mais elle ne transcrit pas la notion de proximité géographique, deux haies se touchant pouvant avoir des barycentres très éloignés dès qu'elles sont un peu longues.

Nous avons donc décidé d'utiliser une pseudo-distance correspondant à la distance minimale entre les points des deux segments et dont la définition est donnée ci-dessous.

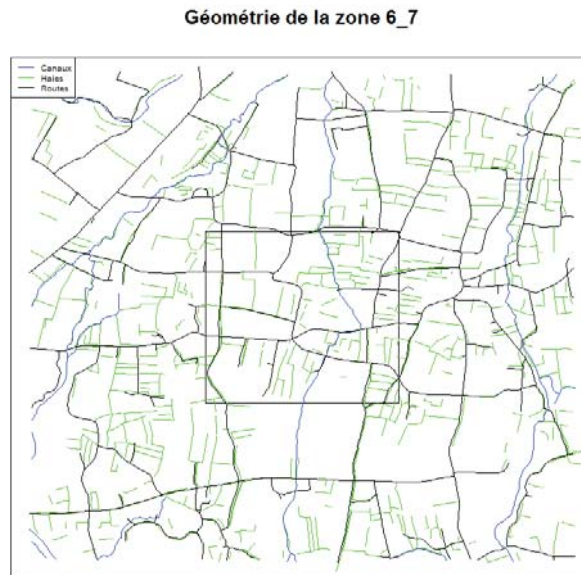


Figure 3. Géométrie de la zone cible centrée sur la cellule 6_7 du paysage *B*

Définition Soit S et S' deux segments de \mathbb{R}^2 , soit d la distance euclidienne entre deux points du plan, nous déterminons la distance entre deux segments S et S' , notée $diSt(S, S')$, par

$$diSt(S, S') = \min\{d(x, y), \forall x \in S, \forall y \in S'\}$$

4.1.2. Angle entre segments

Soit θ , l'angle entre un segment haie et un segment voisin quelconque. Nous avons séparé les valeurs de θ en trois classes, ci-après dénommées « perpendiculaire », « parallèle », et « autre » (cf. tableau 3). La classe « perpendiculaire » correspond aux angles de l'intervalle $[\frac{7\pi}{18}; \frac{11\pi}{18}]$; la classe « parallèle » aux angles de l'ensemble $[0; \frac{\pi}{9}] \cup [\frac{8\pi}{9}; \pi]$; enfin, la classe « autre » recouvre toutes les autres valeurs.

4.1.3. Densité de segments dans le voisinage d'un élément

La densité relative d'un type d'élément dans le voisinage d'une haie est calculée comme le rapport entre le nombre d'éléments de ce type situés à une distance $\leq b$ de la haie et le nombre d'éléments attendus dans ce voisinage,

Tableau 3. *Récapitulatif des classes d'angle entre les deux segments*

Intervalle de valeur des angles	$[0; \frac{2\pi}{18}[$	$[\frac{2\pi}{18}; \frac{7\pi}{18}[$	$[\frac{7\pi}{18}; \frac{11\pi}{18}[$	$[\frac{11\pi}{18}; \frac{16\pi}{18}[$	$[\frac{16\pi}{18}; \pi]$
Nature de la liaison entre segments	Parallèle	Autre	Perpendiculaire	Autre	Parallèle

compte tenu de la densité de ces éléments sur la zone totale. Le nombre de voisins d'un segment S à une distance b est

$$N_r(S, b) = |\{S' | dist(S, S') \leq b\}|$$

où S' est un segment quelconque. Nous définissons le nombre de voisins, de type route (respectivement haie, canal), d'une haie H à une distance $\leq b$ de façon similaire,

$$N_r^R(H, b) = |\{R' | dist(H, R') \leq b\}|$$

où R' est un segment route.

Le nombre d'éléments attendus dans un voisinage est défini pour un segment et une distance donnés de la façon suivante. Soit A la surface de la zone cible, T_A le nombre de segments de routes (respectivement haies, canaux) dans cette zone, $\frac{T_A}{A}$ est la densité moyenne de routes dans la zone. La surface du voisinage d'un segment de longueur L à la distance b est déterminée par $B(L, b) = \pi \cdot b^2 + 2 \cdot L \cdot b$. Le nombre attendu de routes dans ce voisinage est alors

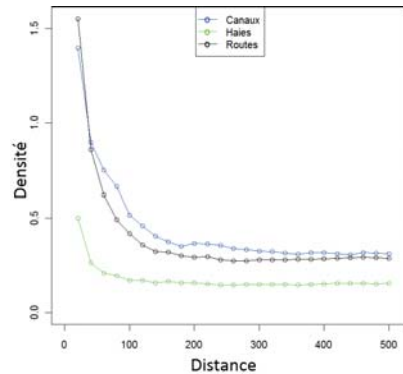
$$N_t^R(L, b) = \frac{B(L, b) \cdot T_A}{A}$$

La densité relative de routes dans le voisinage d'un segment H , notée $D_r^R(H, b)$, est alors donnée par

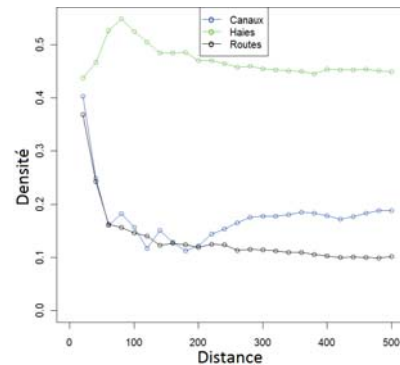
$$D_r^R(H, b) = \frac{N_r^R(H, b)}{N_t^R(H, b)}$$

4.2. Résultats

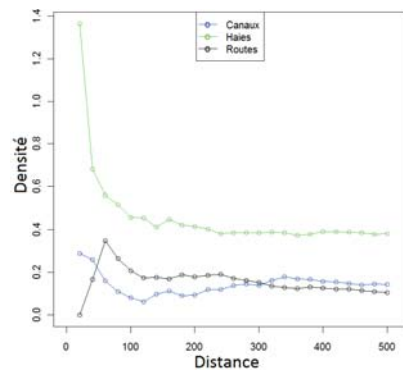
Nous avons mené l'étude sur les différents types de haies, HV, HP ou autre, et leurs différents types de voisins (haies, routes ou canaux), dans leurs différentes liaisons (parallèle, perpendiculaire ou autre), sur les huit cellules sélectionnées. Nous avons tracé les courbes de densité de voisins pour les distances allant de 0 à 500 m. Quelques unes des courbes de densité obtenues pour la cellule 8_9 du jeu de données A sont présentées sur la figure 4. Nous ne présentons pas ici les résultats pour les haies d'orientation *autre*.



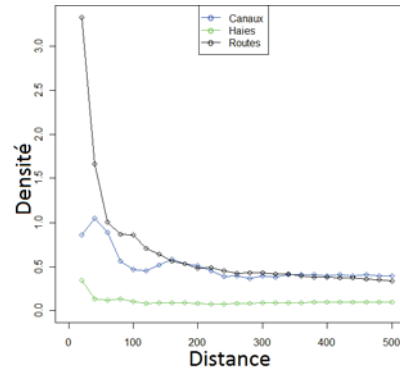
(a) Cellule 8_9 (Données A), courbe de densité pour le voisinage perpendiculaire aux haies (tous types)



(b) Cellule 8_9 (Données A), courbe de densité pour le voisinage parallèle aux haies HV



(c) Cellule 8_9 (Données A), courbe de densité pour le voisinage perpendiculaire aux haies HP



(d) Cellule 8_9 (Données A), courbe de densité pour le voisinage parallèle aux haies HP

Figure 4. Exemple de courbes de densité pour la cellule 8_9 (Données A)

Tout d'abord, nous remarquons que pour l'ensemble des haies (figure 4(a)), la distance joue un rôle quant à la densité d'éléments dans le voisinage d'une haie : en effet, quelle que soit sa valeur de départ, la densité de linéaires de toutes les catégories baisse rapidement et devient constante à partir de 150 m jusqu'à 500 m.

Lorsque nous considérons les voisins parallèles aux haies HV (figure 4(b)), nous voyons que la densité de routes ou de canaux est très faible (inférieure à 0,2) et relativement constante à partir d'une distance faible. Ainsi, dans le voisinage immédiat des haies HV (inférieur à 60 m) nous trouvons quelques routes et canaux mais ce ne sont pas les linéaires les plus présents. En effet,

la densité de haies parallèles aux haies HV est plus importante que toutes les autres densités à toute distance (entre 0 et 500 m) avec un maximum relatif apparaissant pour une petite distance (inférieur à 80 m).

Concernant les voisins perpendiculaires aux haies HP (figure 4(c)), le voisinage immédiat (inférieur à 20 m) comporte une densité de haies forte (supérieure à 1) et encore élevée (supérieure à 0,6) pour les distances plus grandes (inférieures à 40 m). Enfin, si nous étudions le voisinage parallèle aux haies HP (figure 4(d)), nous nous apercevons que celui-ci est composé avant tout de linéaires pérennes et que la densité de ceux-ci à faible distance est très supérieure à celle à grande distance.

Ces résultats se généralisent sur l'ensemble des cellules étudiées comme nous le montrons ci-dessous (figures 5, 6, 7). Pour comparer les résultats obtenus sur chaque cellule, nous avons divisé chaque valeur de densité par la valeur à 500 m. Nous présentons les voisinages inférieurs à 20 m et ceux inférieurs à 100 m, afin d'illustrer les effets des voisinages immédiat et proche dans les différentes directions. Les huit cellules sont regroupées selon leur région d'origine (A ou B) et leur nombre de haies (moyen ou élevé).

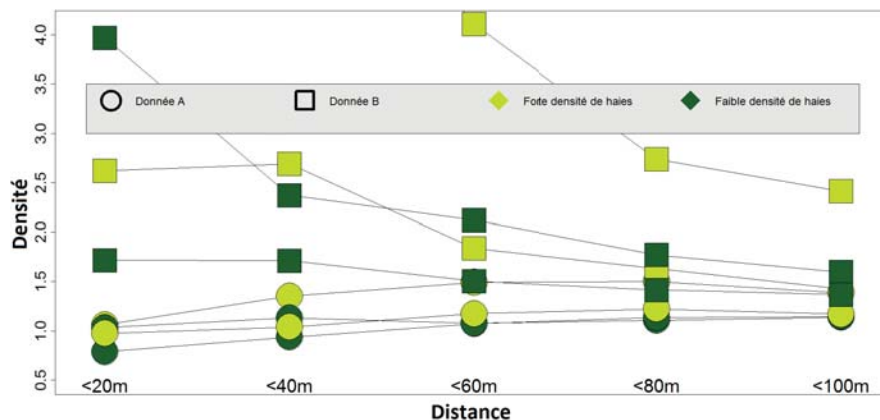


Figure 5. Densité relative des haies dans le voisinage des haies HV divisée par la densité relative à 500 m. Les valeurs sont données pour les distances de voisinages inférieure à 20 m, à 40 m, à 60 m, à 80 m et à 100 m et pour les haies parallèles aux haies HV. Chaque point correspond à une cellule. Une cellule des données B n'est pas représentée pour les voisinages inférieure à 20 m et inférieure à 40 m car la valeur sont trop haute (8.19 et 5.25)

Le premier graphique (figure 5) présente la densité relative de haies parallèles aux haies HV pour l'ensemble des cellules des deux paysages. Nous constatons qu'il n'y a pas de comportement spécifique lié au nombre de haies dans les cellules, mais une différence entre régions. Les cellules issues du paysage B affichent des densités relatives supérieures à celles du paysage A. Deux

comportements distincts sont visibles dans ce graphe. Le premier montre une diminution de la densité lors de l'augmentation de la taille du voisinage, le second montre une élévation de la densité jusqu'à une certaine distance. Ceci indique que les voisinages immédiats des haies sont plutôt dépourvus de haies parallèles dans le paysage A mais pas dans le B.

Le second graphique (figure 6) présente la densité relative des routes dans le voisinage des haies HV et dans les trois directions parallèle, perpendiculaire et autre. D'une part, nous pouvons dire que les effets de l'orientation des routes dans le voisinage des haies HV disparaît avant 100 m, quelle que soit cette orientation. D'autre part, les haies HV des deux paysages semblent s'organiser différemment suivant l'orientation des routes. Ainsi, sur les données A, les routes dans le voisinage proche des haies HV sont plutôt perpendiculaires aux haies alors que pour les données B, les haies HV présenteraient un voisinage un peu plus mixte mais avec une préférence pour les routes parallèles.

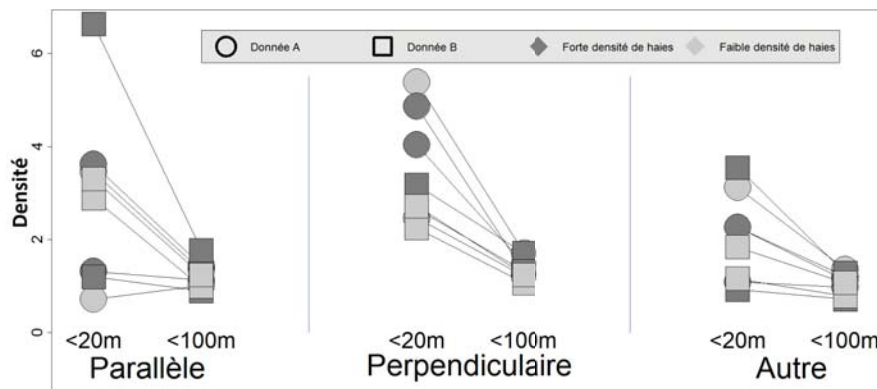


Figure 6. Densité relative des routes dans le voisinage des haies HV divisée par la densité relative à 500 m. Les valeurs sont données pour les distances de voisinages inférieures à 20 m et inférieures à 100 m et pour les routes parallèles, perpendiculaire ou sans orientation particulière par rapport aux haies HV. Chaque point correspond à une cellule.

Le troisième graphique (figure 7) présente la densité relative des routes dans le voisinage des haies HP. Comme pour les graphes précédents, les effets de voisinages disparaissent au delà de 100 m. La densité de routes d'orientation quelconque par rapport aux haies HP est la même sur l'ensemble des cellules et ce, quelle que soit la distance. En revanche, nous observons une forte densité de routes parallèles dans le voisinage proche des haies HP et ce d'autant plus que les cellules comportent un grand nombre de haies.

14 SAGEO 2012. Volume /2012

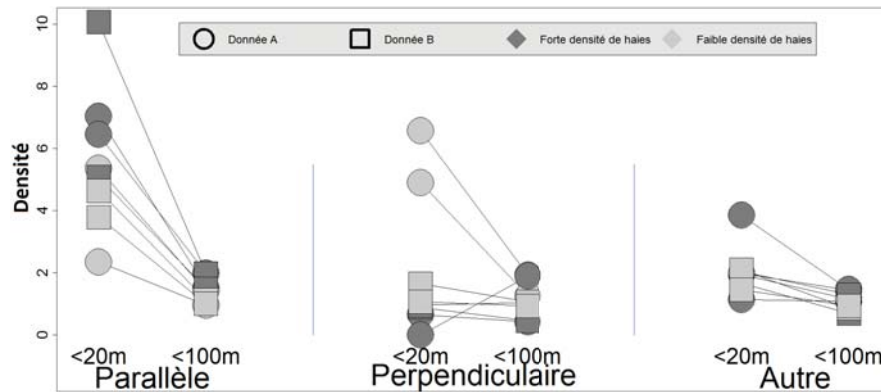


Figure 7. Densité relative des routes dans le voisinage des haies HP divisée par la densité relative à 500 m. Les valeurs sont données pour les distances de voisinages inférieures à 20 m et inférieures à 100 m et pour les routes parallèles, perpendiculaires ou sans orientation particulière par rapport aux haies HP. Chaque point correspond à une cellule.

5. Conclusion

Cet article présente les premiers résultats de notre travail pour la modélisation de la structure des haies dans les paysages agricoles. Nous nous sommes appuyés sur deux paysages très contrastés. L'hypothèse de dépendance entre les haies et les éléments pérennes (routes, canaux) est confirmée. De plus, nous avons montré que la distance jouait un rôle dans la structuration des haies, et qu'il n'y avait pas d'effet au delà de 100-150 m. Cela induit que la modélisation pourrait être pensée à cette échelle. Enfin, nous avons examiné le rôle de la direction des haies, séparées selon les directions ouest-nord-ouest – est-sud-est (HV) et nord-nord-est – sud-sud-ouest (HP), et prouvé que la modélisation indépendante de ces deux types de haies était pertinente pour les données A mais apparemment pas pour les données B. Les méthodes de simulation envisagées avec ces résultats s'appuieraient sur l'utilisation des processus ponctuels [STO 01, Van 00]. Il s'agirait de simuler les barycentres des haies et d'utiliser leurs relations de voisinage avec des linéaires pérennes afin de marquer ces points avec les valeurs d'angle les plus probables.

6. Bibliographie

- [ARN 83] ARNOLD G., « The influence of ditch and hedgerow structure, length of hedgerows, and area of woodland and garden on bird numbers on farmland », *Journal of applied ecology*, vol. 20, n° 3, 1983, p. 731-750.

- [CAM 09] CAMPAGNE P., BUISSON E., VAROUCAS G., ROCHE P., BAUMEL A., TATONI T., « Modeling landscape structure constraints on species dispersal with a cellular automaton : Are there convergences with empirical data ? », *Ecological Complexity*, vol. 6, n° 2, 2009, p. 183–190.
- [CAR 02] CARSIJENS G. J., VAN DER KNAAP W., « Strategic land-use allocation : dealing with spatial relationships and fragmentation of agriculture », *Landscape and Urban Planning*, vol. 58, n° 2–4, 2002, p. 171–179.
- [DES 07] DE SMITH M. J., GOODCHILD M. F., LONGLEY P. A., Eds., *Geospatial analysis : a Comprehensive Guide to Principles, Techniques and Software Tools*, vol. 1, Troubador, July 2007.
- [GAU 06] GAUCHEREL C., FLEURY D., AUCLAIR D., DREYFUS P., « Neutral models for patchy landscapes », *Ecological Modelling*, vol. 197, n° 1–2, 2006, p. 159–170.
- [GRO 10] GROOT J. C. J., JELLEMA A., ROSSING W. A. H., « Designing a hedgerow network in a multifunctional agricultural landscape : Balancing trade-offs among ecological quality, landscape character and implementation costs », *European Journal of Agronomy*, vol. 32, n° 1, 2010, p. 112–119.
- [LAC 05] LACOSTE C., DESCOMBES X., ZERUBIA J., « Point processes for unsupervised line network extraction in remote sensing », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, n° 10, 2005, p. 1568–1579.
- [Le 09] LE BER F., LAVIGNE C., ADAMCZYK K., ANGEVIN F., COLBACH N., MARI J.-F., MONOD H., « Neutral modelling of agricultural landscapes by tessellation methods - Application for gene flow simulation », *Ecological Modelling*, vol. 220, n° 24, 2009, p. 3536–3545.
- [MON 05] MONESTIEZ P., BAILLY J.-S., LAGACHERIE P., VOLTZ M., « Geostatistical modelling of spatial processes on directed trees : Application to fluvisol extent », *Geoderma*, vol. 128, n° 3–4, 2005, p. 179–191.
- [PAS 88] PASEK J. E., « 30. Influence of wind and windbreaks on local dispersal of insects », *Agriculture, Ecosystems & Environment*, vol. 22–23, 1988, p. 539–554.
- [RIC 09] RICCI B., FRANCK P., TOUBON J.-F., BOUVIER J.-C., SAUPHANOR B., LAVIGNE C., « The influence of landscape on insect pest dynamics : a case study in southeastern France », *Landscape Ecology*, vol. 24, 2009, p. 337–349, Springer Netherlands, 10.1007/s10980-008-9308-6.
- [RIC 11] RICCI B., FRANCK P., BOUVIER J.-C., CASADO D., LAVIGNE C., « Effects of hedgerow characteristics on intra-orchard distribution of larval codling moth », *Agriculture, Ecosystems & Environment*, vol. 140, n° 3–4, 2011, p. 395–400.
- [SAU 00] SAURA S., MARTÍNEZ-MILLÁN J., « Landscape patterns simulation with a modified random clusters method », *Landscape Ecology*, vol. 15, 2000, p. 661–678, Springer Netherlands.
- [STO 01] STOICA R., « Processus ponctuels pour l'extraction de réseaux linéiques dans les images satellitaires et aériennes », PhD thesis, Université de Nice-Sophia Antipolis, Nice, France, 2001.
- [TUR 91] TURNER M. G., GARDNER R. H., « Quantitative methods in landscape ecology : an introduction », TURNER M. G., GARDNER R. H., Eds., *Ecological Studies Analysis and Synthesis : Quantitative Methods in Landscape Ecology : The Analysis and Interpretation of Landscape Heterogeneity*, chapitre 1, p. 3–14,

16 SAGEO 2012. Volume /2012

Springer-Verlag, New York, 1991.

[Van 00] VAN LIESHOUT M., *Markov point processes and their applications*, Imperial College Press, 2000.

[VAN 08] VANNIER C., HUBERT-MOY L., « Detection of Wooded Hedgerows in High Resolution Satellite Images using an Object-Oriented Method », *2008 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, vol. 4, july 2008, p. 731–734.

[VAN 10] VANNIER C., HUBERT-MOY L., « Wooded hedgerows characterization in rural landscape using very high spatial resolution satellite images », *2010 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, july 2010, p. 347–350.

Constructions cartographiques pour la caractérisation de l'aléa volcanologique du Piton de la Fournaise

Cartographie de séries de données chrono-spatiales

Paule-Annick Davoine* — **Cécile Saint-Marc*** — **Andrea Dimuro**** — **Thomas Staudacher**** — **Marlène Villanova-Oliver*** — **Patrice Boissier****

* *Laboratoire d'Informatique de Grenoble, Université Joseph Fourier*

681 Rue de la Passerelle

38402 Saint-Martin-d'Hères

{paule-annick.davoine, cecile.saint-marc, marlene.villanova-oliver}@imag.fr

** *Observatoire Volcanologique du Piton de la Fournaise / Institut de Physique du Globe de Paris*

{dimuro, staud}@ipgp.fr, boissier@univ-reunion.fr

RÉSUMÉ. La caractérisation de l'aléa éruptif d'un volcan effusif tel que le Piton de la Fournaise s'appuie sur l'élaboration d'un certain nombre de cartes représentant la répartition spatiale des phénomènes observés dans le temps lors des différentes éruptions, à savoir fissures, fractures, coulées, cônes, séismes... Le caractère temporel des phénomènes étudiés en volcanologie renvoie à des questions conceptuelles, aussi bien en termes de structuration des données que de choix sémiologique pour leur représentation. Cet article s'intéresse à la cartographie de séries de données chrono-spatiales relatives à la volcanologie, c'est-à-dire des matrices d'information géographique dans lesquelles les objets géographiques sont identifiés à partir de leur date d'occurrence et organisés chronologiquement. Il traite des problèmes rencontrés et des réflexions méthodologiques menées lors de l'élaboration de cartographies de ce type de données dans un contexte d'analyses volcanologiques.

ABSTRACT. To characterize volcanic hazard on the Piton de la Fournaise volcano, it is essential to have a deep knowledge of past eruptions events: faults, cinder cones, lava flows, seisms... The first step is to map these past events. However, the temporal aspect of these data raises mapping issues. This article deals with the choice of suited cartographic symbology to visualize chrono-spatial data, which are GIS data chronologically ordered.

MOTS-CLÉS : cartographie, données spatio-temporelles, risques naturels, sémiologie, volcanologie

KEYWORDS: Cartography, natural hazards, spatio-temporal data, semiology, volcanology

1. Contexte et problématique

L'île de la Réunion est soumise à de nombreux risques naturels. Parmi eux, le risque volcanique lié à l'activité intense du Piton de la Fournaise, pour lequel on dénombre plus d'une trentaine d'éruptions entre 1998 et 2011 [BES 11] et dont certaines (1977, 1986, 1998, 2007) ont touché des habitations et des infrastructures de communication [PAY 07], [PEL 09]. Ceci amène donc les autorités à s'intéresser à l'évaluation du risque volcanique avec comme première étape, l'élaboration d'une cartographie de l'aléa éruptif [DIM 12].

Le rôle de la carte dans l'évaluation des aléas naturels n'est plus à démontrer [LEO 06] [ARN 09a]. La cartographie de l'aléa nécessite d'élaborer, au préalable, une succession de cartes, permettant de spatialiser les événements passés afin de mieux appréhender le phénomène étudié. Dans le cas de l'étude de l'aléa éruptif, il s'agit d'élaborer un ensemble de cartes représentant la répartition spatiale de chaque phénomène lié aux éruptions en fonction de la chronologie des éruptions : fractures, fissures, cônes, coulées de laves, séismes, L'objectif est d'identifier des variations ou des régularités spatio-temporelles de ces phénomènes, de les combiner, afin de comprendre le fonctionnement du volcan. Toutefois, le caractère temporel des données géographiques utilisées complexifie les représentations cartographiques. En effet, les phénomènes observés sont organisés sous la forme de matrices d'informations géographiques organisées chronologiquement, dans lesquelles chaque ligne correspond à une entité géographique et chaque entité est identifiée par sa date d'occurrence. Il s'avère que les phénomènes éruptifs observés ont des localisations géographiques proches au cours du temps, voire identiques, des formes similaires et peuvent se superposer en partie ou en totalité. Ceci réduit donc le nombre d'éléments visibles sur la carte représentant une certaine période temporelle, et altère sa lisibilité. Ce problème est d'autant plus marqué que le nombre d'éléments à cartographier est important et que le support cartographique est statique.

L'objet de cet article est de présenter les réflexions méthodologiques menées lors de l'élaboration de cartes « temporelles » nécessaires à la caractérisation de l'aléa volcanologique¹. Les spécificités des données volcanologiques, présentées dans la section 2, nous ont conduit à nous interroger sur les différentes façons de cartographier le temps sur un support statique et à proposer des représentations adaptées (section 3).

¹ Ce travail a été effectué en collaboration avec les chercheurs de l'Observatoire Volcanologique du Piton de la Fournaise, et s'inscrit dans une étude financée par la DEAL de l'Île de la Réunion, liée à l'évaluation du risque volcanologique.

2. Spécificités des données volcanologiques pour la caractérisation de l'aléa

2.1. Des données multiples pour un aléa multiple

Le Piton de la Fournaise est l'un des volcans les plus actifs et aussi l'un des plus surveillés. L'Observatoire Volcanologique du Piton de la Fournaise, créé en 1979, a pour missions la surveillance du volcan, l'étude de son fonctionnement et la diffusion scientifique des données acquises relatives à l'activité éruptive. Différents types de données sont relevés : celles relatives aux signes précurseurs de l'éruption, et celles relatives aux manifestations mêmes de l'éruption à la surface de la terre. Dans le premier cas, il s'agit de données issues de stations de mesure ou de capteurs tels que les GPS, extensomètres, inclinomètres, permettant de mesurer la déformation ou les sismomètres qui enregistrent l'activité sismique du massif volcanique. Dans le deuxième cas, il s'agit plutôt de données d'observation de surface correspondant aux éléments géomorphologiques qui apparaissent lors de l'éruption. Dans le cas d'un volcan effusif tel que le Piton de la Fournaise, ces éléments appelés aussi phénomènes éruptifs sont les coulées de lave, les cônes, les fissures, les fractures et les téphras (blocs, cendres, cheveux de pélé ...).

L'évaluation de l'aléa éruptif s'appuie sur la connaissance historique de l'activité du volcan et de ses manifestations phénoménologiques : coulées, fissures, cônes, téphras, séismes précurseurs. Celle-ci doit permettre de comprendre le fonctionnement du volcan, de dresser une cartographie de la localisation des phénomènes éruptifs passés. Pour chaque phénomène un aléa spécifique est évalué :

- l'aléa « Coulées » caractérise les zones susceptibles d'être touchées par des coulées de lave ;
- l'aléa « Ouverture de bouche éruptive » définit les zones les plus probables d'apparition d'un phénomène éruptif de surface, en prenant en compte la localisation des fissures et cônes ;
- l'aléa « Fractures » s'intéresse aux fracturations du sol en surface, susceptibles d'entraîner des effondrements ;
- l'aléa « Téphras » caractérise les zones susceptibles d'être touchées par des projections de cendres, de bombes ou de blocs volcaniques au cours d'une éruption.

L'analyse combinée de ces différents aléas permet d'élaborer une carte des zones menacées ou carte de l'aléa volcanologique. Quant aux séismes, signes précurseurs à une éruption, ils interviennent dans la compréhension du fonctionnement du volcan et donc de l'aléa et font aussi l'objet de représentations cartographiques.

2.2. Les données volcanologiques : des données chrono-spatiales

La description et l'analyse des phénomènes éruptifs s'appuient sur différentes catégories d'information :

- une information géographique relative à la localisation du phénomène et à sa description géométrique, qui est de type surfacique pour les coulées et les cônes, linéaire pour les fissures et fractures, et ponctuelle pour les séismes mais aussi les cônes lorsque l'on s'intéresse plutôt à leur répartition spatiale ;
- une information temporelle correspondant à la date d'occurrence du phénomène, c'est-à-dire le moment d'apparition du phénomène, généralement la date de l'éruption à laquelle il est rattaché : dans le cas de deux coulées de lave apparues le même jour, on considère une seule entité géographique multipartie.
- une information thématique, sous forme attributaire, décrivant les caractéristiques du phénomène observé.

Nous disposons donc d'un ensemble de jeux de données référencées, dans le temps et dans l'espace, relatifs aux fissures, aux fractures, aux cônes, aux coulées et aux séismes, qui correspondent en fait à un ensemble d'entités géographiques de même nature, ayant une géométrie propre, relevées dans le temps et organisées chronologiquement (notons que les blocs et les cendres ne sont pas caractérisés par une date). Ainsi on peut donc parler de série de données chrono-spatiales :

- Les objets géographiques qu'elles contiennent sont superposables puisque chaque enregistrement correspond à l'apparition d'un nouvel objet (nouvelle fissure, fracture, coulée...). Cette superposition peut s'effectuer en totalité ou en partie.
- La cadence d'enregistrement des entités, c'est-à-dire le temps qui s'écoule entre deux enregistrements ou deux observations et qui correspond aux moments d'occurrence des phénomènes éruptifs, est irrégulière. Il s'agit en effet de phénomènes aléatoires.
- Le niveau de granularité temporelle, c'est-à-dire le pas de temps qui décrit le phénomène observé, est variable selon les différents jeux de données, mais aussi au sein d'un même jeu de données. Le niveau de granularité commun reste l'année, mais plus les phénomènes sont récents plus la granularité est détaillée, allant jusqu'à l'heure d'occurrence du phénomène en passant par le jour et le mois. Inversement, plus les phénomènes sont anciens, plus l'information temporelle est vague et arbitraire.

Dans le contexte de notre étude, la cartographie doit permettre de localiser les phénomènes observés, pour une date ou une période donnée. Il s'agit de :

- répondre à la question de la localisation spatiale et temporelle des objets (où ?, quand ? quoi ?) et de pouvoir visualiser leurs caractéristiques spatiales et phénoménologiques.

- repérer l'ordre d'apparition des entités, leur séquençement, leur fréquence d'apparition, dans le contexte d'une même éruption mais aussi plus largement à des fins de compréhension des différents mécanismes éruptifs. Il s'agit de répondre aux questions du type « dans quel ordre ? », « à quelle fréquence ? » et d'identifier des structures spatiales et temporelles. La question de l'ordre se rapporte aux notions de succession et de simultanéité temporelle. Elle renvoie aux relations temporelles entre les objets : « avant », « après », « en même temps » et donc aux relations d'Allen [ALL 83]. D'un point de vue cartographique, ces relations impliquent de conserver la trace des phénomènes passés, et de représenter les empreintes spatiales de tous les phénomènes recensés, même si au cours du temps et en fonction des événements éruptifs, certains d'entre eux ont pu être recouverts en partie ou en totalité, par des phénomènes plus récents.

3. Propositions cartographiques pour la représentation des données volcanologiques

Afin de répondre aux objectifs décrits dans la section précédente et compte tenu des éléments de sémiologie graphique proposés pour représenter le temps dans les cartes, nous avons été amenés à construire, de façon expérimentale, un nombre important de cartes représentant les caractéristiques phénoménologiques de l'activité éruptive du Piton de la Fournaise [DIM12]..

A partir d'un échantillon de ces cartes, considérées comme étant les plus représentatives tant d'un point de vue des aspects phénoménologiques pour la caractérisation de l'aléa volcanique, que des spécificités des données volcanologiques, nous présentons les difficultés rencontrées et les réflexions menées. Ainsi, nous traitons ici des exemples de prise en compte du temps dans la cartographie de phénomènes géographiques datés qu'ils soient surfaciques tels que les coulées, linéaires tels que les fissures ou encore ponctuels tels que les séismes.

3.1 *Quelle sémiologie pour une représentation du temps ?*

C'est à partir des années 70, que les géographes ont commencé à s'intéresser aux interactions temps/espace. Mais les réflexions sur l'expression cartographique du temps datent plutôt des années 90 [CHE 94], [THI 99], même si la nécessité d'introduire le temps dans les cartes est ancienne [PAL 00], [CAU 08].

L'expression cartographique du temps est un problème récurrent et constitue un enjeu fondamental pour la représentation des changements et la compréhension de la dynamique des processus spatiaux [CHE 07], [CAU 08], [CUN 12]. Les nouvelles technologies de l'information et de la communication rendent possible son introduction en cartographie, via l'usage des techniques d'animations, du

multimédia et de l'interactivité [KRA 97], [KRA 11], [CAU 08], [ARN 09 a, b], même si les problèmes conceptuels et sémiologiques sont loin d'être résolus [CUN 12]. Ces approches sont utilisées dans un contexte de géovisualisation associé parfois à une démarche d'analyse exploratoire [AND 10], mais ne permettent pas de résoudre la question de la représentation du temps, qu'il s'agisse du mouvement ou du changement sur un support cartographique statique. Or l'impression papier est souvent le support de communication final : les cartes sont imprimées, affichées, diffusées à travers les rapports papiers. C'est en tout cas ce qui était attendu dans notre contexte. La carte statique traditionnelle reste donc le support de communication privilégié.

La nécessité d'introduire le temps dans les cartes statiques n'est pas récente, malgré les difficultés rencontrées, [BER 67], [CAU 08] :

Cartographier les taux de variations des attributs quantitatifs d'une même entité géographique entre deux dates constitue une façon classique de cartographier les dynamiques temporelles (comme par exemple les variations d'indices ou de taux ...).

La collection de cartes est un processus permettant de représenter des variations temporelles. Elle peut aussi bien représenter les changements liés à la composante attributaire, qu'à la géométrie des objets. [CAU 08]. Toutefois ce procédé trouve vite ses limites dès lors que le nombre de séquences de temps à représenter, et donc que le nombre de cartes, est important.

Vasilev, cité dans [CAU 08], propose une sémiologie en fonction des différentes catégories du temps et de la géométrie de l'objet. Les catégories de temps considérées sont le moment et la durée entre deux événements. Il suggère d'appliquer sur la carte, en implantation ponctuelle, une étiquette représentant la date d'occurrence de l'objet, pour représenter le moment. Ce procédé, bien qu'il n'implique pas une cartographie spécifique, est souvent utilisé pour représenter des historiques ou l'apparition de nouveaux phénomènes. Ce procédé est souvent utilisé pour la cartographie des risques naturels : historique des inondations, des avalanches, ... [ARN 09a]. Il permet, éventuellement, une lecture de la carte au niveau détaillé (ie identification de la modalité « date » pour chaque unité géographique), mais, rend difficile l'identification de structures spatiales et ne rend pas compte des dynamiques ou des variations spatio-temporelles des phénomènes. Par ailleurs, lorsque que le nombre d'éléments à représenter est important, se pose alors le problème de la surcharge cognitive de la carte.

La notion de mouvement implique aussi une relation espace-temps. Cette notion implique le déplacement d'un objet ou d'un individu d'un lieu au temps T1 vers un autre lieu au temps T2 [CAU 09]. L'« aquarium spatio-temporel » ou « cube spatio-temporel », qui s'appuie sur une géométrie 3D, (chaque axe représentant une dimension Temps - Espace - Attributs), constitue l'expression cartographique associée à ce type de temporalité [AND 03], [CAU 08]. Il est essentiellement utilisé

pour l'étude de problématiques liées à la mobilité des individus (visualisation des trajectoires) ou à des séries chronologiques d'évènements ponctuels tels que les séismes [AND 03].

Une façon classique de cartographier des données géographiques datées est d'effectuer une analyse thématique sur la variable « date ». Deux solutions peuvent être envisagées selon les règles de la sémiologie de Bertin [BER 67] :

- Soit la date est identifiée comme une variable qualitative (c'est-à-dire une variable discrète, caractérisant l'objet en lui-même), et dans ce cas on peut appliquer la variable visuelle couleur (précisément ici la teinte, selon le modèle TSL [SMI 78]), qui permet de mettre en avant les ressemblances et les différences [LEF 07], en associant à chaque date une teinte spécifique. Se posent alors plusieurs problèmes, notamment, celui de la lisibilité de la carte lorsque le nombre de modalités à représenter est particulièrement important, ainsi que celui du choix des couleurs et des teintes qui peut perturber la lecture et la compréhension de la carte [CHR 11].
- Soit la date est perçue comme une variable quantitative ou qualitative ordonnée, et c'est alors la variable visuelle valeur (appelée aussi intensité [SMI 78]) qui peut être utilisée pour identifier des relations d'ordre. Dans ce cas, se pose le problème de la construction des classes (choix du type de classification, et des seuils de classes), en plus du problème de l'usage de la couleur. Afin d'améliorer le problème de l'harmonie des couleurs dans une carte [CHR 11], il peut être possible de s'appuyer sur les travaux proposés par Brewer [BRE 03], notamment si le nombre d'événement à cartographier reste limité.

3.2. Cartographie de phénomènes surfaciques datés : les coulées

3.2.1. La date comme variable quantitative pour faciliter la lecture

Pour cartographier des données chrono-spatiales liées à la volcanologie, nous avons considéré la date comme une donnée quantitative continue. Le premier choix à faire est de définir la granularité à laquelle représenter l'information (heure, jour, mois, année, décennie, etc.). Quelle que soit cette granularité, le nombre de classes peut rapidement être important compte tenu des périodes de temps étudiées en géologie. Or le nombre maximal de classes interprétables par l'œil humain sur une carte est de 6 ou 7 [ROB 78]. Parallèlement, la granularité peut être imposée par l'objectif d'analyse de la carte.

Prenons l'exemple des coulées de lave sur le massif du Piton de la Fournaise. L'observatoire volcanologique du Piton de la Fournaise souhaitait une carte montrant les coulées de lave récentes, de 1970 à nos jours, dans le but d'analyser leur répartition spatiale et d'identifier des variations spatio-temporelles. Sur cette période, 29 années ont connu une ou plusieurs éruptions, pour un total de 129 coulées de lave.

Il a été convenu qu'une classification des coulées par année (soit 29 classes) représenterait une précision d'analyse suffisante. Pour faire face au grand nombre de classes à distinguer sur la carte, nous avons choisi d'utiliser deux variables visuelles : la teinte et l'intensité, comme le montre la figure 1.

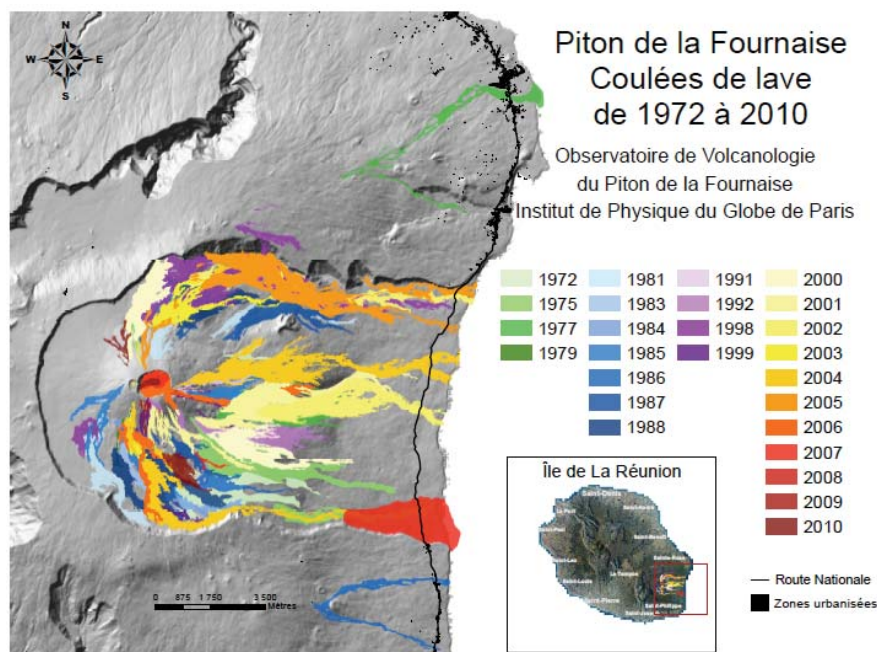


Figure 1. Utilisation conjointe des variables visuelles « teinte » et « intensité » pour représenter un nombre important de données datées

L'utilisation conjointe de ces deux variables permet de diviser l'interprétation visuelle en deux niveaux : à l'échelle de la décennie puis à l'échelle de l'année. La variable teinte s'applique à l'échelle de la décennie et permet d'identifier l'appartenance d'une coulée à un grand ensemble temporel. La variable intensité s'applique à l'intérieur de ce grand ensemble pour distinguer plus précisément la date de la coulée. La seule variable date est donc considérée à la fois comme une variable qualitative (teinte) et quantitative (intensité).

La lecture de la carte est facilitée par ces deux niveaux d'interprétation visuelle. Malgré tout, le nombre important d'objets sur la carte empêche de visualiser la totalité de l'information : certaines coulées de lave sont partiellement, voire totalement masquées par d'autres plus récentes. Un moyen simple de contourner ce problème est de réaliser des cartes par période plus réduite, par exemple une carte de 1972 à 1989 et une de 1990 à 2010. Cependant, des morceaux de coulées resteraient

toujours masqués à la vue du lecteur. Nous avons proposé un nouveau mode de représentation pour pallier ce problème.

3.2.2. Visualiser les superpositions temporelles

Les phénomènes naturels surfaciques que sont les coulées, sont donc susceptibles de survenir plusieurs fois aux mêmes endroits au cours du temps. La visualisation conjointe de ces événements superposés spatialement est alors problématique, notamment si les superpositions sont nombreuses. Si la transparence est un moyen classique de visualiser des phénomènes superposés, elle n'est pas adaptée dès lors que la couleur est utilisée comme variable visuelle. En effet, visualiser des couleurs différentes par transparence dénature ces couleurs et altère la légende. Cette solution ne s'applique donc pas à notre mode de représentation. Cependant, en se basant, sur cette idée, nous avons finalement choisi de tracer le contour des entités pour produire un « effet de transparence » tout en conservant les couleurs d'origine. Comme nous pouvons le voir sur la figure 2, ce contour permet de visualiser à la fois les tracés des coulées les plus récentes et le chemin des coulées plus anciennes.

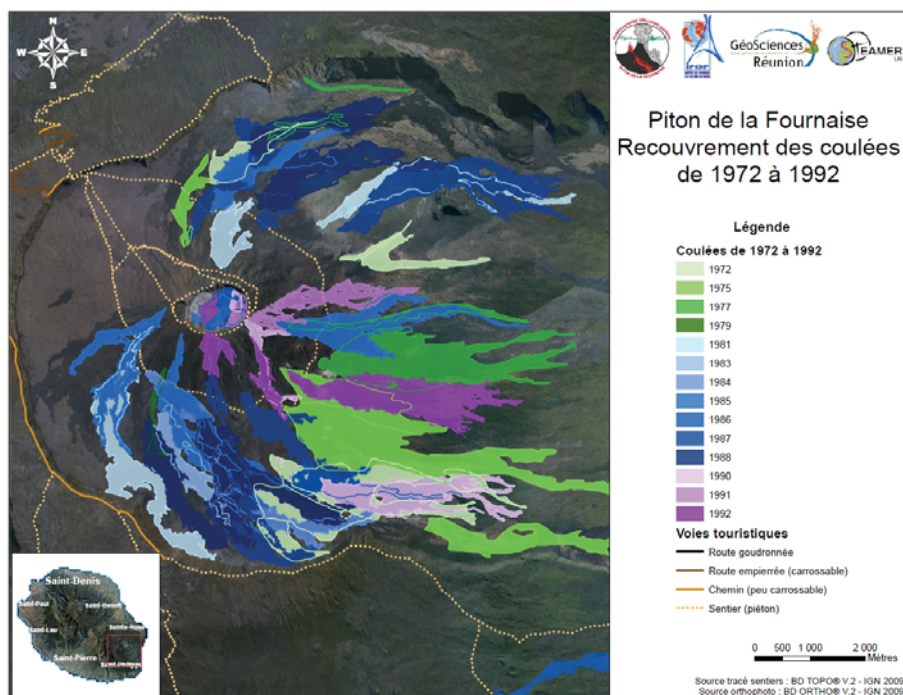


Figure 2. Illustration du tracé du contour des entités permettant de conserver la variable « teinte » tout en visualisant les superpositions

3.3. Cartographie de phénomènes linéaires datés : les fissures

La sélectivité de la variable « teinte » est moins grande pour les figurés linéaires et ponctuels [LEF 07]. Elle nécessite donc de choisir des couleurs fortement différenciées pour produire une carte facile à interpréter. Par conséquent, la représentation à deux niveaux choisie pour les coulées de lave ne peut pas être appliquée à des phénomènes linéaires comme les fissures éruptives.

Les fissures éruptives sont des phénomènes fortement liés aux coulées de lave puisqu'une coulée naît à partir d'une fissure dans le sol. Par conséquent, nous avons choisi d'utiliser encore une fois cette variable visuelle, dans les mêmes teintes, pour faire écho aux cartes précédentes. La variable « intensité », peu discriminante pour ce type de données vectorielles, a été abandonnée. La représentation des fissures, visible en figure 3, n'a donc pu se faire qu'à la granularité de la décennie.

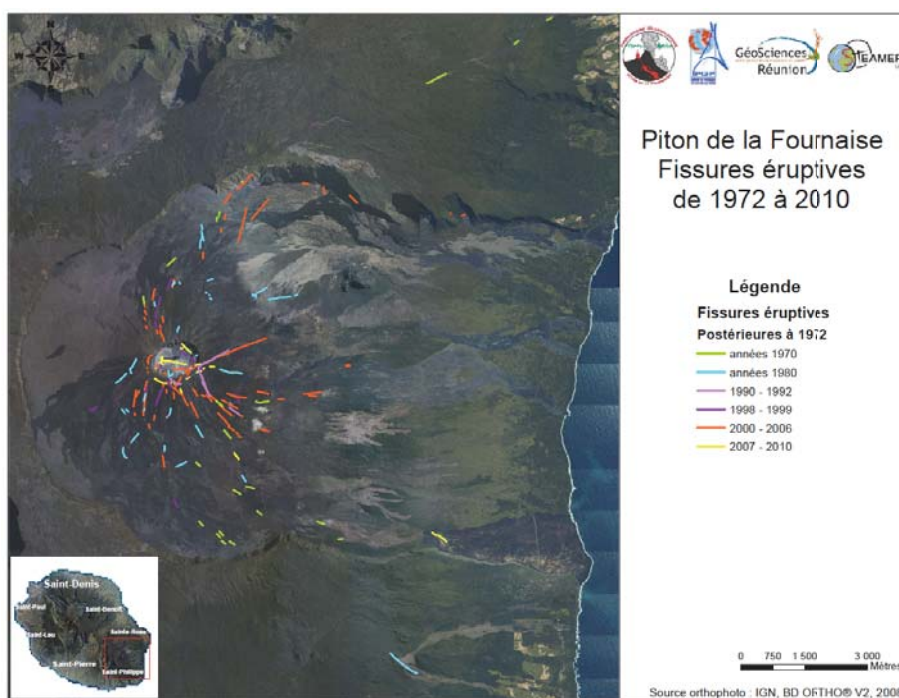


Figure 3. Cartographie d'objets temporels linéaires : Utilisation de la variable « teinte » sur des données agrégées au niveau des décennies.

Deux problèmes apparaissent sur cette carte. D'une part, il est difficile de distinguer des motifs spatiaux entre les différents groupes de fissures : elles sont nombreuses dans chaque classe et leur éparpillement dans l'espace entraîne une

absence de continuité d'une zone de la carte à l'autre, ce qui freine un balayage fluide de l'image par le regard. D'autre part, aucune analyse n'est réalisable au pas de temps de l'éruption (échelle mensuelle).

Nous nous sommes alors demandé comment descendre à une résolution temporelle plus fine tout en facilitant la lecture des motifs spatiaux formés par les événements. La solution que nous avons envisagée consiste à généraliser l'information disponible pour aboutir à la construction de polygones, représentant non pas les fissures elles mêmes, mais les zones de fissures.

Sur la figure 4, chaque polygone convexe englobe l'emprise des fissures pour une période temporelle donnée. En repassant à une représentation sous forme de surfaces, on peut alors réintroduire la variable visuelle intensité en plus de la teinte et représenter les données à l'échelle de l'année.

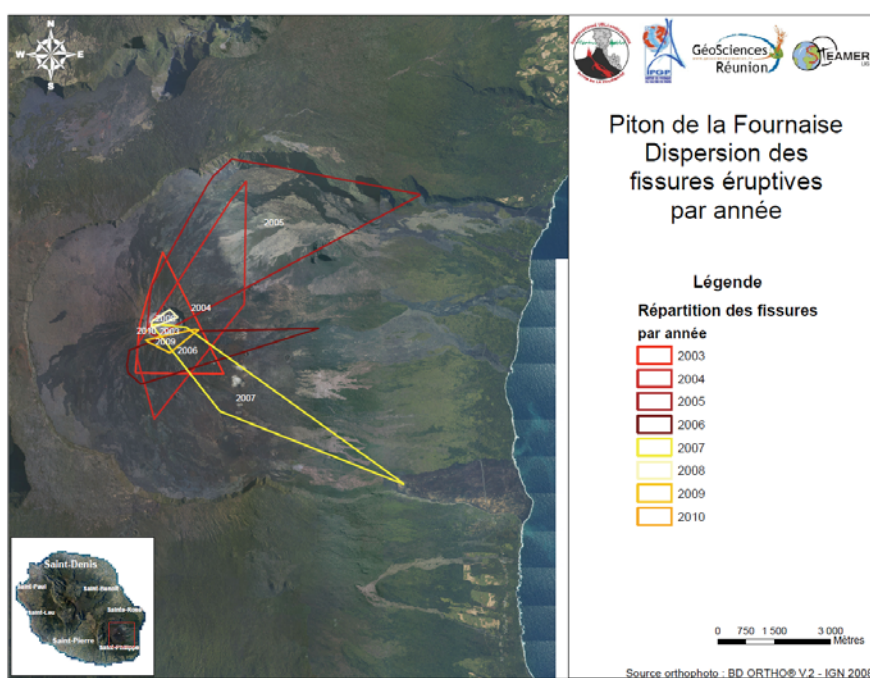


Figure 4. Surfaces représentant l'emprise des fissures par année. Cette proposition facilite l'identification de motifs spatiaux et permet de conserver la variable intensité.

3.4. Cartographie de phénomènes ponctuels datés : les séismes

3.4.1. Des phénomènes ponctuels caractérisés par leur densité

Partant de l'hypothèse qu'il existe un lien entre la localisation des hypocentres des séismes et la montée du magma, l'analyse de la répartition spatio-temporelle des séismes intervient dans l'évaluation de l'aléa éruptif. Dans ce contexte, il ne s'agit pas forcément de s'intéresser à l'intensité des séismes mais plutôt à l'évolution de leur localisation en fonction du temps qui reflète le déplacement souterrain de la lave, principalement sur la période qui précède l'éruption. La durée de cette période est très variable selon les éruptions, allant de quelques heures à plusieurs mois.

Les séismes correspondent à des événements ponctuels. Deux contraintes sont rencontrées dans la représentation de ce type de données. Tout d'abord, la quantité de données enregistrées et localisées étant particulièrement importante, l'information est difficilement perceptible sur une carte. Par ailleurs, à petites échelles, les symboles de représentation ponctuels doivent avoir une taille suffisante pour être visibles [CAU 08] (le seuil de visibilité d'une entité surfacique est de 0,12 mm² [RUA 99]). Deux événements aux localisations proches peuvent donc apparaître superposés, ce qui rend leur dénombrement impossible.

Comme il est difficile de comparer le nombre de points présents dans chaque zone d'une carte, qu'ils soient ou non masqués par d'autres, on utilise communément des cartes de densité. Cette représentation, présente l'inconvénient de ne pas conserver la chronologie des événements. Nous avons donc cherché un moyen plus adapté de représenter l'information temporelle pour ces phénomènes ponctuels.

3.4.2. Cartographier l'évolution temporelle de phénomènes ponctuels

Habituellement, la variable « teinte » est utilisée par les sismologues pour représenter la répartition spatio-temporelle des séismes préalables aux éruptions. Il semble donc naturel de l'utiliser aussi pour nos cartes, comme pour les coulées et les fissures.

Pour analyser la dynamique des séismes à l'échelle de plusieurs éruptions, nous avons produit la figure 5. Si celle-ci permet de visualiser une tendance à la variation spatiale des séismes au fil des mois, les points sont trop nombreux pour que l'interprétation soit fiable.

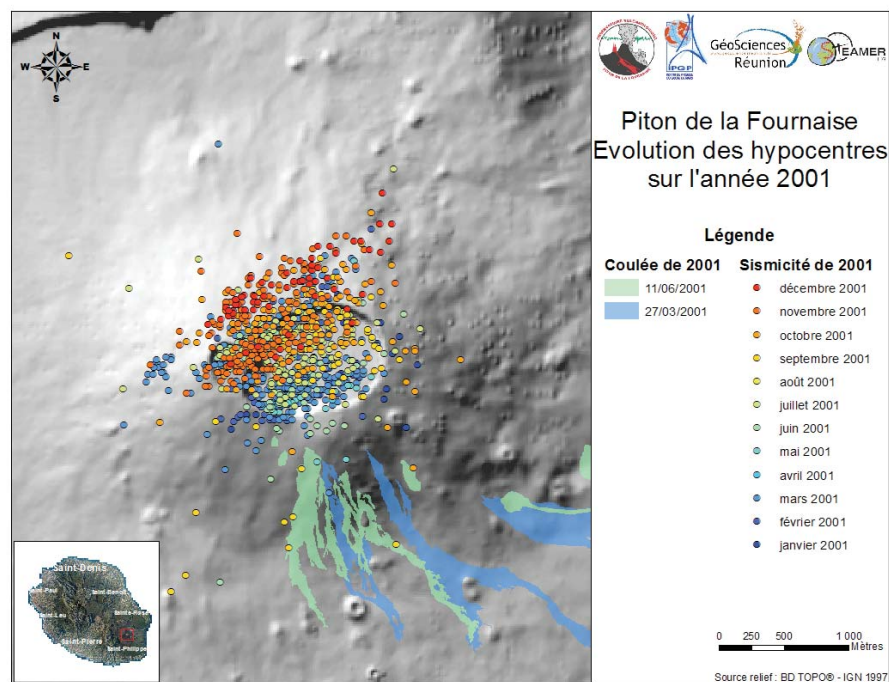


Figure 5. Le nombre élevés de points fait obstruction à la lecture de la variable « teinte » et à l'analyse du temps

Pour pallier ce problème, nous avons représenté les points d'un même mois sous forme d'un seul point situé en leur barycentre (figure 6). Il est possible de conserver l'information du nombre de séismes en utilisant la variable « taille ». On conserve ainsi la date et la quantité sous forme lisible et il est possible d'analyser les déplacements des hypocentres des séismes au cours du temps. Ceux-ci sont représentés sur la figure 7 par des lignes reliant les points.

Pour ne pas réduire un nuage de points à un point unique, une technique d'agrégation a récemment été expérimentée par [JAN 12], qui consiste à agréger les points d'un jeu de données par zones, de façon à obtenir des cercles proportionnels aux nombres de points présents dans chaque zone, tout en évitant les recouvrements entre cercles. Cette méthode permet de co-visualiser sur une carte de façon assez lisible jusqu'à quatre jeux de données ponctuels. Néanmoins, dans notre étude des séismes où un jeu de données correspondrait à un mois, le nombre de périodes à co-visualiser dépasse largement quatre et cette technique ne pourrait pas être adoptée.

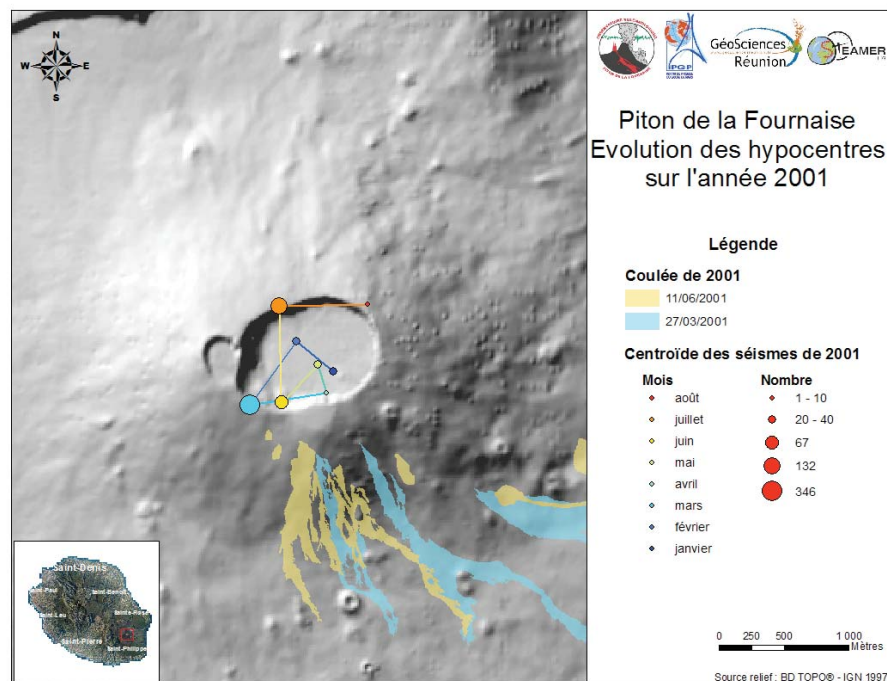


Figure 6. L'usage de la variable « taille » associée à celle de la « teinte » permet de conserver les attributs « date » et « nombre » pour visualiser les variations spatio-temporelles des séismes

3. Conclusion

A travers l'exemple de la cartographie de l'aléa volcanique, nous avons voulu montrer les difficultés rencontrées, lors de la cartographie d'entités géographiques datées, qu'elles soient surfaciques, linéaires ou ponctuels et organisées chronologiquement sous forme de séries de données chrono-spatiales. Nous avons donc élaboré des propositions pour la représentation cartographique temporelle, en considérant par exemple la date comme une variable quantitative ou qualitative en fonction de la nature de l'objet géographique et de ses caractéristiques temporelles, ou en utilisant la technique du contour apparenté à la transparence pour mémoriser l'empreinte spatiale des phénomènes passés qui se superposent, ou encore en proposant la technique des polygones englobant auxquels on associe l'étiquette de la date. Si ces propositions ont contribué à analyser le processus éruptif du Piton de la Fournaise, il n'en reste pas moins qu'elles sont très liées à ce contexte. Une évaluation de leurs apports sur d'autres cas d'étude liés aux risques naturels, (domaine où les séries de données chrono-spatiales sont fréquemment utilisées) est en cours d'investigation

5. Bibliographie

- [ALL 83] Allen J.-F., "Maintaining knowledge about temporal intervals", *Communications of the ACM*, volume 26, n°11, November 1983, ACM Press, p. 832-843.
- [AND 03], Andrienko N., Andrienko G., Gatalsky P., Exploratory spatio-temporal visualization, : an analytical review, *Journal of Visual Languages and Computing*, 14 2033,pp 503-541
- [AND 10] Andrienko G., Andrienko N., Demsar U., Dransch D., Dykes J., Fabrikant S., Jern M., Kraak M-J, Schumann H., Tominski Ch., "Space, time and visual analytics", *International Journal of Geographical Information Science*, 24:10, 2010, p. 1577-1600.
- [ARN 09a] Arnaud A., « Valorisation de l'information dédiée aux événements de territoires à risque. Une application sur la couronne grenobloise ». Thèse de doctorat de Géographie, Université Joseph Fourier, Grenoble 1, 500p. 2009
- [ARN 09b] Arnaud A., Davoine P-A., Temporal geovisualization in risk aera, International Cartography Conference, November 15-21, 2009, Santiago du Chili
- [BER 67] Bertin J., *La sémiologie graphique*. Paris, Gauthiers-Villars, 1967, 431 p.
- [BES 11] Bès de Berc S., Gehl P., Di Muro A., Villeneuve N. (2011). « Risque volcanique et gestion de crise à La Réunion », In : Les volcans actifs de l'outre-mer français, des édifices à haut risque, Boudon G., *Geosciences*, vol. 14, p. 78.
- [BRE 03] Brewer C., A transition in improving maps : the Colorbrewer example, *Cartography and Geographic information Science*, 2003, 30(2°, p.159-162.
- [CAU 08] Cauvin C., Escobar F., Serradji A., *Cartographie thématique*, tome 1 et 5, Traite IGAT, Hermès, 2008
- [CHE 07] Cheylan J-P., « Les processus spatio-temporels : quelques notions et concepts préalables à leur représentation ». *Les dossiers de Mappemonde « Cartographie animée »*, n°87, 2007, 21 p.
- [CHR 11] Christophe S. Zanin C., Roussaffa H., 2011, Colours harmony in cartography, *25th International Cartographic Conference (ICC'11)*, 3-8 July, Paris (France)
- [CUN 12] Cuntly C., Mathian H., Premiers pas vers une proposition de sémiologie graphique animée. Un outil d'expérimentation, *Cartes et Géomatique, Revue du Comité Français de Cartographie*, mars 2012.
- [Di 12] Di Muro A., Bachelery P., Boissier P., Davoine P-A., Fadda P., Favalli M., Ferrazzini V., Finizola A., Leroi G., Mairine P., Manta F., Michon L., Morandi A., Nave R., Peltier A., Principe C., Ricci T., Roult G., Saint-Marc C., Staudacher Th., Villeneuve N. Évaluation de l'aléa volcanique à la Réunion, Rapport d'Etude, Ministère de l'Environnement, DEAL île de la Réunion, Avril 2012.
- [JAN 12] Jänicke S., Heine C., Stockmann R., Scheuermann G., "Comparative Visualization of Geospatial-Temporal Data". *Proceedings of the 3rd International Conference on Information Visualization Theory and Applications, IVAPP 2012*, Rome, Italy, 2012, p. 613-625.

- [KRA 97] Kraak M-J., Edsall R., MacEachren A-M., "Cartographic Animation and Legends for Temporal Maps: Exploration and or Interaction". *Proceedings of the 18th International Cartographic Conference*, Stockholm, Sweden, 1997, p. 253-260.
- [KRA 03] Kraak M-J., "Geovisualization illustrated", *ISPRS Journal of Photogrammetry and Remote Sensing*, 2003, 57(6-5), p. 390-399.
- [KRA 11] Kraak M.J., Ormeling F. and F-J., *Cartography: visualization of spatial data*. New York, Guildford Press, 2011.
- [LEF 07] Le Fur A., *Pratiques de la cartographie*, 2e édition, Paris, Editions Armand-Colin, 2007.
- [LEO 06] Léone F., Leales T., « Des cartes pour gérer le risque volcanique à la Martinique (Antilles françaises) », *Revue Internationale de Géomatique*, N° special, volume 16 – n°3-4/2006, 291-546, pp 341-358.
- [PAL00], Palsky G. Le temps des cartographes, in Les cartes de la connaissance, sous la direction de J-P Laborde et P-R Baduel,, KARTHALA, -URBAMA, 2000
- [PAY 07] Payet G., « Les, Réunionnais et leur Volcan, Mission Victimologie, Eruption du Piton de la Fournaise », Antenne Réunion de l'Institut de Victimologie, Avril 2007.
- [PEL 09] Peltier, A., Bachèlery P., Staudacher T., "Magma transfer and storage at Piton de La Fournaise (La Réunion Island) between 1972 and 2007: a review of geophysical and geochemical data", *Journal of Volcanology and Geothermal Research*, 184(1-2), 2009, p. 93-108.
- [ROB 78] Robinson A.H., Sale R., Morrison J., *Elements of Cartography*, New York, Jon Wiley & Sons, 1978.
- [RUA 99] Ruas A., « Modèle de généralisation de données géographiques à base de contraintes et d'autonomie », Thèse de doctorat, Université de Marne la Vallée, 2009.
- [SMI 78] Smith A. R., "Color Gamut Transform Pairs", *Computer Graphics*, volume 12 - n°3, *SIGGRAPH 78 Conference Proceedings*, 12-19 August 1978.
- [THI 99] Theriault M., Claramut C., « La représentation du temps et des processus dans les SIG : une nécessité pour la recherche interdisciplinaire », *Revue Internationale de Géomatique*, volume 9, n°1, 1999, p. 67-103, Hermès, Paris.

Towards an internationally more accessible data model for the 2D large-scale topographic inventory of Flanders

Ine De Cubber*, — Jos Van Orshoven****

** KaHo Sint-Lieven*

Gebroeders Desmetstraat 1, 9000 Gent, Belgium

Ine.decubber@kahosl.be

*** KU Leuven*

Celestijnenlaan 200E - box 2411, 3001 Heverlee, Belgium

Jos.vanorshoven@ees.kuleuven.be

ABSTRACT: The large-scale topographic inventory of Flanders (GRB) is based on a vernacular data model. In this paper, a schema matching is presented based on which the similarities and dissimilarities between the data model of the GRB and CityGML, an OGC standard, are identified, with a view to standardise the GRB data model, to make the information flow more consistent and to make the GRB data model internationally more accessible. By this, the potential of CityGML is shown to model and store 2D data from existing LSTI, like the GRB in a more standardized environment.

KEY WORDS: Data modelling, CityGML, GRB, Schema matching

1. Introduction

1.1. Data modelling

A definition for data modelling is “abstracting a portion of the real world” (The Open Geospatial Consortium, 2012). Three abstraction levels are distinguished, whereby modelling actions can be defined for each level, as shown in table 1 (De Cubber *et al.*, 2011). An overview of the three abstraction levels is given in figure 1 (Longley *et al.*, 2007). At the conceptual level, object classes and high-level relationships are defined. At the logical level, attributes and relationships, with primary keys and operations are added. At the physical level, tables are created and stored in files or databases. Where it regards the geometry, the actions at conceptual level are the choice of view, the geometry representation and the high-level primitives. At logical level, the spatial data structure is determined and at physical level, spatial tables are created and stored in spatial files or databases.

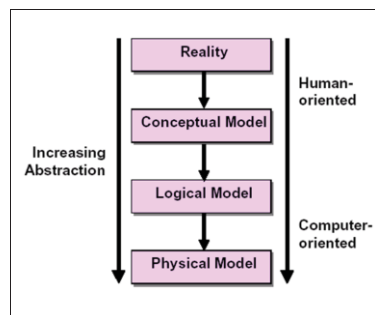


Figure 1: The three abstraction levels for data modelling (Longley *et al.*, 2007)

1.2. Standards for semantic and spatial data modelling

UML (ISO/IEC 19501:2005) and ERM (Chen, 1976) are popular approaches for non-spatial data modelling.

For spatial data modelling, the European ISO standards for gridded and vector geometry representation like the ISO 19125-1:2004 Simple Features standard and the ISO19136:2007 Geography Markup Language are grouped into the ISO 19100 series of Geographic Information Standards (Kresse *et al.*, 2004). The major de facto standards for spatial data modelling are the OGC Keyhole Markup Language (KML) and the industrial standards ESRI Shapefile (SHP).

Whereas most of the existing standard modelling approaches focus either on the semantics or on the geometry, the de facto OGC standard CityGML combines both the semantics and the geometry (The Open Geospatial Consortium – CityGML, 2012).

Abstraction levels	Semantic data modelling actions	Spatial data modelling actions
Reality	Requirements analysis	
Conceptual	Definition of object classes	Choice of view: - Discrete object/ Continuous field Choice of geometry representation: - Gridded / Vector
	Definition of relationships at high-level	Choice of primitives at high-level: - Geometric / Topologic / Gridded
Logical	Definition of attributes and relationships	Choice of spatial data structure: - Geometric / Topologic / Gridded
	Definition of primary keys and operations	
Physical	Creation of tables	Creation of spatial tables, managed as spatial files or using DBMS

Table 1: Abstraction levels for modelling geographic reality and specific actions for semantic data modelling and spatial data modelling (De Cubber *et al.*, 2011)

1.3. Large-scale topographic inventories and their data models

In Belgium, large-scale topographic inventories (LSTI) are produced and maintained by each of the three political regions separately. Three distinct databases result: Grootschalig Referentie Bestand in Flanders (GRB), the Projet Informatique de Cartographie Continue in Wallonia (PICC) and Urbis (Brussels region) (De Cubber *et al.*, 2009). Among the many other LSTI-initiatives in Europe are the OS Master Map in the UK (Ordnance Survey, 2012), and the GBKN, BGT and IMGEO (respectively the current LSTI, the future LSTI and the extended LSTI) in the Netherlands (Geonovum, 2012). Each of these LSTI has its own vernacular data models encompassing specific modelling actions.

2. Problem statement and objectives

The starting point of this paper is that although international standards for data modelling are available and although the distinction between modelling at conceptual, logical and physical level is widely accepted, many LSTI are based on their own vernacular data models encompassing non standardized modelling actions.

A detailed examination of the data models of the Flemish LSTI (GRB) and their documentation revealed a lack of consistency in the information flow through the GRB modelling process and the absence of a number of expected modelling actions at each abstraction level (De Cubber *et al.*, 2011).

Against this background, the objectives of this paper are to search for similarities and dissimilarities between the data model of the GRB and the CityGML standard, with a view to standardise the GRB data model, to make the information flow more consistent, and to make the GRB data model internationally more accessible. Moreover the potential of the 3D CityGML standard to accommodate 2D data from vernacular LSTI like the GRB is tested.

3. Methodology

Schema matching is used to identify similar elements in the different abstraction levels of the data models and to identify the dissimilarities. (Rahm *et al.*, 2001) present a taxonomy of the existing schema matching approaches, with a detailed description of these approaches.

For the comparison of the GRB data model and the CityGML data model, a linguistic, element-based schema matching approach is used, based on names and textual descriptions of schema elements, like object classes and attributes.

4. Materials

4.1. The GRB, large-scale topographic inventory of Flanders

The GRB is a database holding large-scale/high resolution topographic object structured data according to a specific data model and data format. The GRB data model makes provision for buildings and building infrastructures, parcel separations and borders, roads and road infrastructure, railways and water infrastructures, ... The GRB is expected to cover the region of Flanders in 2014 (AGIV, 2012). The GRB dataset is used by cities, municipalities, cable and pipeline administrators, educational institutions and many other users, as a topographic reference database which can be enriched with more domain-specific information.



Figure 2: View of an extract of the GRB database (left) and virtual 3D model of Ettenheim, using CityGML (Research Center Karlsruhe, 2012)(right)

4.2. *The OGC CityGML standard*

CityGML is an encoding standard for the representation, storage and exchange of virtual 3D city and landscape models (The Open Geospatial Consortium - CityGML, 2012) based on OGC's standard GML3 and covering the geometrical, topological, and semantic aspects of 3D city models. CityGML combines both spatial and semantic properties. It distinguishes object classes which are similar to those in the GRB (like buildings and other man-made artifacts, vegetated objects, water bodies, and transportation facilities like streets and railways). CityGML provides a multiscale model with 5 well-defined hierarchical levels of detail (LoD), from a regional model (LoD0), through a city model (LoD1 – LoD2) to an architectural exterior model (LoD3) and even an architectural interior model (LoD4).

5. Results

The GRB and CityGML data models were compared at conceptual, logical and physical level. As CityGML provides 5 levels of detail, the comparison of the data models had to be done at each of these five levels. For each GRB object class, the corresponding CityGML object class was identified and for each CityGML object class, the most relevant levels of detail are indicated between square brackets []. In what follows, the sign \cong should be read as “matches with”.

5.1. Conceptual level

At this level, the real-world entities were translated into object classes and high level relationships were defined, together with the choice of view, the geometry and the high-level primitives.

Certain entities like the WaterSurface and Railways are described in the same way in both the GRB and the CityGML data model, resulting in identical object classes (GRB-WTZ \cong CityGML-Waterbody[0-1], GRB-SBN \cong CityGML-Railway[1-2-3-4]).

Thanks to the multiple levels of detail in CityGML, certain entities, like Buildings, can be modelled in much more detail within CityGML, resulting in a complex comparison between the two data models (GRB-GBG \cong CityGML-Building[1], GRB-GBG \cong CityGML-GroundSurface[2-3-4], GRB-GVL \cong CityGML-WallSurface[2-3-4]). At the lower LoDs, CityGML provides extra object classes, giving the possibility to store information about roof surfaces, wall openings like doors and windows, and even about building furniture and building installations.

Some entities are represented in less detail within CityGML, like parcels (GRB-ADP \cong CityGML-Landuse[0-1-2-3-4]), where the border points of parcels in the

GRB data model cannot be matched with any CityGML object class (GRB-PRP \cong ?).

Entity classes, like manholes and façade points, are part of the GRB data model, but are not present at all within CityGML (GRB-WRI \cong ?, GRB-GVP \cong ?). But some other entity classes, like relief and vegetation, are modelled within CityGML, but have no counterpart in GRB(? \cong CityGML-relief, ? \cong CityGML-vegetation).

Comparison of entities...	GRB - conceptual data model: object classes	CityGML - conceptual data model: object classes
Identical object classes: Water, Railway	<div>watergang (wtz)</div> <div>spoorbaan (sbn)</div>	<div><<Feature>> WaterBody</div> <div><<Feature>> Railway</div>
With more detail in CityGML: Building	<div>gebouw aan de grond (gbg)</div> <div>gevellijn (gvl)</div> <div>/</div> <div>/</div>	<div><<Feature>> Building</div> <div><<Feature>> WallSurface</div> <div><<Feature>> Opening</div> <div><<Feature>> RoofSurface</div>
With less detail in CityGML: Parcels	<div>administratief perceel (adp)</div> <div>perceels-reconstructiepunt (prp)</div>	<div><<Feature>> LandUse</div>
Not in CityGML: Manholes, Façade Points	<div>gevelpunt (gvp)</div> <div>putdeksels (wri)</div>	<div>/</div> <div>/</div>
Not in GRB: Relief, vegetation	<div>/</div> <div>/</div>	<div><<Feature>> ReliefFeature</div> <div><<Feature>> _VegetationObject</div>

Table 2: Examples of entity classes and their representation as object classes in the conceptual data model of GRB and CityGML

Within the GRB conceptual data model, only three high-level relationships are defined: (i) GRB-Façade line bounds GRB-Building, (ii) GRB-Façade point bounds GRB-Façade line and (iii) GRB-building has GRB-façade points. Within CityGML, the same high-level relationships are defined, but sometimes with different syntax: CityGML-Abstract building is bounded by CityGML-boundary surface. As CityGML does not have an object class for the façade points, relationships in which façade points occur cannot be translated.

Within CityGML, other relationships are defined, like “CityGML-AbstractBuilding consists of CityGML-BuildingPart”. Although the GRB has also the object classes GRB-building and GRB-building part, such relationship is not foreseen in the conceptual GRB-data model.

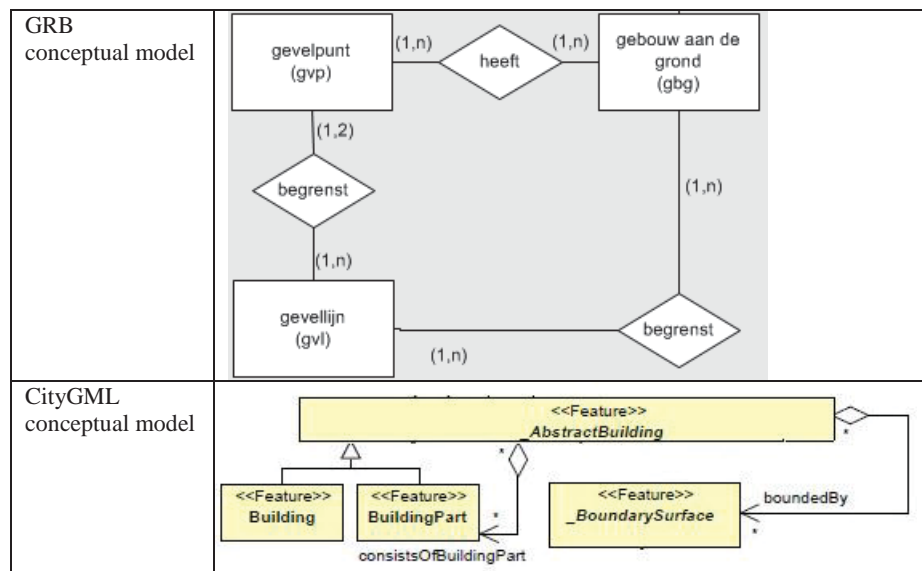


Figure 3: Comparison of the relationships present at conceptual level in GRB and CityGML

The chosen view, geometry and high-level primitives are not explicitly mentioned in the GRB data model. The spatial modelling actions at conceptual level for both CityGML and GRB are the same, namely: discrete objects are represented by vectors built upon geometric primitives. As CityGML uses the GML3 standard, the possibility exists to store also continuous fields (represented as geometry), and to store the topology in an explicit way.

5.2. Logical level

At the logical level, the attributes, primary keys and operations are added to the conceptual model, the high-level relationships are translated into explicit types of relationships like aggregation and composition, and the spatial data structure is chosen.

The comparison of all the attributes between GRB and CityGML, showed that each object class has its own specific attributes. Some of these attributes from the GRB, can be accommodated by CityGML, others can't and vice versa.

The high-level relationships, found at conceptual level, disappear in the GRB logical model but are translated into explicit types of relationships at the logical level of the CityGML data model. Each object class in the GRB data model has a primary key, named IDN. With CityGML, primary keys are not defined. (Figure 4)

The GRB uses the Simple features data structure (ISO 19125-1:2004) for storing points, lines and polygons. CityGML uses the GML 3.1.1 spatial data structure (The Open Geospatial Consortium-GML, 2012). GML 3.1.1 is also an ISO standard (ISO 19136:2007) and provides possibilities to store points, linestrings and polygons.

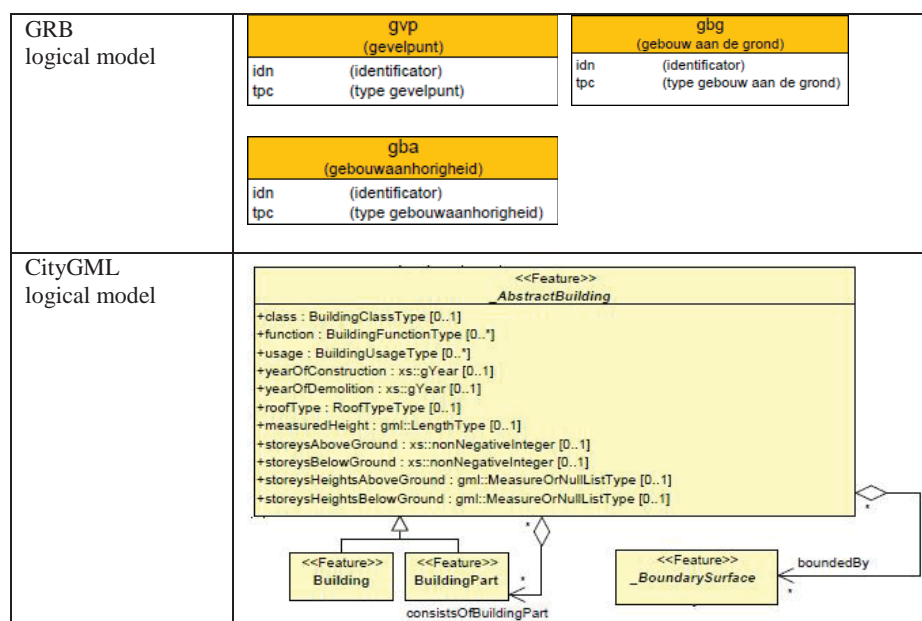


Figure 4: Comparison of the attributes, relationships and primary keys in the GRB logical data model and the CityGML logical data model.

5.3. *Physical level*

At physical level, the non-spatial and spatial tables are created, according to the specification in the logical data model. The physical model is software dependent, based on the file format or the database format in which the data will be stored. For the GRB, the geometric data is stored in shapefile-format and additional semantic tables are stored in DBF format. For CityGML, XML schema definitions are available, describing how a CityGML file should be stored, resulting in a text-file, which can be read by specific CityGML viewers, like LandXplorer CityGML viewer (Autodesk, 2012). The logical model of CityGML can also be translated into a spatial database, like the 3D CityDB (<http://www.3dcitydb.net/>).

6. Discussion

6.1. *Schema matching for a new data modelling process versus schema matching between existing data models*

Within a data modelling process, starting from the user requirements, it is extremely interesting to start with a conceptual model, which can be fine-tuned in a next phase into a logical model, and transformed into databases or files, according to the physical model. For a new data modelling process, schema matching can already be done at conceptual level, to see whether the newly designed conceptual model fits with other existing conceptual data models, how existing data can be easily integrated in the new model. In the next phase, when fine-tuning, a schema matching can be done again between the new logical model and existing logical models, to find similarities and the same can be done at physical level. Schema matching at the three levels, ensures that the newly designed data models will match as closely as possible the reference data models.

In the case of standardising the GRB data model, where the data modelling process for both the GRB and CityGML was already done, resulting in data models at the three abstraction levels, there is no real advantage in doing the schema matching at all of the three levels. As the logical level is an extension of the conceptual level, a matching at conceptual level only results in a comparison of the object classes, ignoring the importance of the attributes. Therefore a schema matching at logical level is sufficient to demonstrate the similarities and dissimilarities between the two data models. Also the physical data model, was already chosen for both GRB and CityGML. A fundamental comparison between the OGC CityGML format and ESRI-shapefile format (ESRI, 2012) can be done, but a schema matching at physical level in this specific case gives no added value for creating a standardised data model for the GRB, as the ESRI shapefile is already a de facto standard for vector spatial data.

6.2. Matching cardinalities

Certain object classes of the GRB cannot be easily matched with one object class within CityGML, but have to be matched with different independent object classes. (Rahm *et al.*, 2001) describe this as “the matching cardinality”, where one element or object class of the first data model may match with more elements or object classes of the other data model and vice versa. For the matching of multiple object classes at a time, complex expressions will be needed to specify how the object classes and attributes are related. A simple textual element based matching, as done before, will not be satisfactory for this situation. Two types of matching cardinalities can be distinguished, a matching into independent object classes and into parent-child object classes.

Independent object classes: Within the GRB, different types of objects can be stored in one object class. In table 3, the different types of the object class “civil construction” are shown. When trying to match this object class to CityGML, some of the types correspond with the object class CityGML-BuildingFurniture, (with specific BuildingFurniture FunctionType) like tanks (type 9) and water mills (type 11), others are stored as CityGML-CityFurniture, (with specific CityFurnitureFunctionType) like columns (type 5) and chimneys (type 7) and again others cannot be stored at all within CityGML, like breakwaters (type 22) (Table 4).

1: overbrugging	9: silo, opslagtank
2: waterbouwkundige constructie	10: cabine
3: cultuurhistorisch monument	11: watertoren
4: hoogspanningsmast / openbare TV-mast	12: tunnelmond
5: pijler	13: chemische installatie
6: rooster	22: golfbreker
7: schoorsteen	23: havenpier
8: koeltoren	24: staketsel

Table 3: Different types of objects of the object class “Civil construction”, found in the textual documentation of the conceptual GRB.

GRB logical model	<table><tr><th colspan="2">knw (kunstwerk)</th></tr><tr><td>idn</td><td>(identifier)</td></tr><tr><td>tpc</td><td>(type kunstwerk)</td></tr><tr><td>mtc</td><td>(vorm)</td></tr></table>	knw (kunstwerk)		idn	(identifier)	tpc	(type kunstwerk)	mtc	(vorm)
knw (kunstwerk)									
idn	(identifier)								
tpc	(type kunstwerk)								
mtc	(vorm)								
CityGML logical model	<table><tr><th><<Feature>> CityFurniture</th><th><<Feature>> BuildingFurniture</th></tr><tr><td>+class : CityFurnitureClassType [0..1] +function : CityFurnitureFunctionType [0..*]</td><td>+class : BuildingFurnitureClassType [0..1] +function : BuildingFurnitureFunctionType [0..*] +usage : BuildingFurnitureUsageType [0..*]</td></tr></table>	<<Feature>> CityFurniture	<<Feature>> BuildingFurniture	+class : CityFurnitureClassType [0..1] +function : CityFurnitureFunctionType [0..*]	+class : BuildingFurnitureClassType [0..1] +function : BuildingFurnitureFunctionType [0..*] +usage : BuildingFurnitureUsageType [0..*]				
<<Feature>> CityFurniture	<<Feature>> BuildingFurniture								
+class : CityFurnitureClassType [0..1] +function : CityFurnitureFunctionType [0..*]	+class : BuildingFurnitureClassType [0..1] +function : BuildingFurnitureFunctionType [0..*] +usage : BuildingFurnitureUsageType [0..*]								

Table 4: Matching of one GRB object class “civil construction” with multiple CityGML object classes “CityFurniture” and “BuildingFurniture”

Parent-child object classes: In the GRB, one object class may encompass different types of objects. The information on the specific type of object is always stored as an attribute (TPC) in the GRB data model. Within CityGML, different types are sometimes modelled as an attribute, like for CityFurniture and BuildingFurniture, but sometimes result in the creation of new object classes, children of the parent, which inherit all the characteristics of the parent. In this example, the object class GRB-road, has an attribute TPC, indicating the difference between squares and roads. Within CityGML, the object class transportation complex, is parent of the object classes track, railway, square and road. The difference in how road types are modelled (object classes with type attributes versus parent-child object classes) is not necessarily a problem when matching the data models. The advantage of a parent-child structure is that each of the characteristics (attributes) of the parent is inherited by the child, but each child can have its own specific extra characteristics. In the example above, each object class “road” (child) inherits the function and the usage of the object class “TransportationComplex” (parent), but can have extra attributes, like for example width and material. When using the “type” attribute, adding additional information to one specific type of object, becomes difficult.

GRB logical model object class	<table border="1"> <tr> <td colspan="2">wb (wegbaan)</td></tr> <tr> <td>idn</td><td>(identifier)</td></tr> <tr> <td>tpc</td><td>(type wegbaanelement)</td></tr> </table>	wb (wegbaan)		idn	(identifier)	tpc	(type wegbaanelement)
wb (wegbaan)							
idn	(identifier)						
tpc	(type wegbaanelement)						
CityGML logical model object classes	<pre> classDiagram class TransportationComplex["<<Feature>>\nTransportationComplex"] { +function : TransportationComplexFunctionType [0..*] +usage : TransportationComplexUsageType [0..*] } class Track["<<Feature>>\nTrack"] class Railway["<<Feature>>\nRailway"] class Road["<<Feature>>\nRoad"] class Square["<<Feature>>\nSquare"] TransportationComplex < -- Track TransportationComplex < -- Railway Track < -- Road Track < -- Square </pre>						

Table 5: Comparison of the entity “road” at logical level: the type attribute, tpc, of the GRB matching with parent-child object classes in CityGML

6.3. CityGML for modelling 2D reality?

the fact that the GRB data is only representing two dimensions, and CityGML is a standard for modelling city environments in three dimensions, was not taken into account in this schema matching exercise. The linguistic, element-based schema matching only focussed on what was modelled (semantic schema matching), but not on how these entities are represented geographically (spatial schema matching). This spatial schema matching was not the focus of this paper, but definitely provides inspiration for future upgrading of the 2D GRB into a 3D database.

7. Conclusion and outlook

The comparison of the 2D GRB data model and the CityGML data model has shown that most of the GRB object classes with their attributes can be remodelled according to the CityGML data model specifications. Such remodelling would make the GRB data model more standardised and internationally more accessible. The approach for remodelling could be tested and fine-tuned in a case study in which a GRB extract is transformed into a CityGML dataset, based on the GRB-CityGML schema matching results of this paper,

For those object classes and attributes of the GRB, that do not fit in the CityGML data model, the type and amount of information lost must be quantified. If a requirements analysis among the GRB-users would reveal that this non-accommodated information is essential, a solution should be found by extending CityGML with Application Domain Extensions (ADE) or by making use of generic objects and attributes. Moreover the CityGML standard makes provision for 3D, and gives thus inspiration for upgrading the 2D GRB to 3D.

8. Bibliography/References

- AGIV, *Agentschap voor geografische informatie Vlaanderen*. Available from: <www.agiv.be> [February 2012].
- Autodesk, *Landexplorer CityGML viewer*. Available from: <<http://usa.autodesk.com/adsk/servlet/item?siteID=123112&id=14184672>> [February 2012].
- Chen P., "The Entity-Relationship Model: Toward a Unified View of Data", *ACM Transactions on Database Systems*, vol. 1 no 1, 1976, p. 9-36.
- De Cubber I., Kips G., Van Orshoven J., "Upgrading 2D large scale topographic inventories for 3D-applications in Belgium", *Proceedings of the 4th international workshop on 3D-geo-information, International Workshop on 3D-geo-information*, ed. P. De Maeyer, T. Neutens, M. De Ryck, 2009, p.69-76.

- De Cubber I., Van Orshoven J., “Unstandardized terminology complicates the communication about 2D and 3D spatial data models”, *Proceedings of the IAMG conference 2011*, Salzburg, Austria, 2011, p.1452-1463.
- ESRI, *ESRI Shapefile Technical Description*. Available from: <<http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>>[February 2012].
- Geonovum, *BGT/Imgeo*. Available from: <<http://www.geonovum.nl/dossiers/bgtimgeo>>[February 2012].
- ISO 19125-1:2004 - *Geographic information -- Simple feature access*. Available from: <http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=40114>[February 2012].
- ISO 19136:2007 *Geographic information -- Geography Markup Language (GML)*. Available from: <http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=32554>[February 2012].
- ISO/IEC 19501:2005 *Information technology -- Open Distributed Processing -- Unified Modelling Language (UML) Version 1.4.2*. Available from: <http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=32620>[February 2012].
- Kolbe T., Gröger G., Plümer L., “CityGML – Interoperable Access to 3D City Models”, *Proceedings of the International Symposium on Geo-information for Disaster Management*, 2005, ed. P. Van Oosterom, S. Zlatanova, E. Fendel, Springer Verlag, p.21-23.
- Kresse W., Fadaie K., *ISO Standards for Geographic Information*, Springer, Berlin, 2004.
- Longley P. A., Goodchild M. F., Maguire D. J., & Rhind D. W., *Geographic Information Systems and Science*, 2nd ed. John Wiley and sons, 2007.
- Ordnance survey, *OS mastermap*. Available from: <<http://www.ordnancesurvey.co.uk/oswebsite/products/osmastermap/>> [February 2012].
- Research Center Karlsruhe, Available from: <<http://www.iai.fzk.de/www-extern/index.php?id=222&L=1>>[February 2012].
- Rahm E., Bernstein P.A., “A survey of approaches to automatic schema matching”, *The VLDB Journal*, Vol. 10 Berlin / Heidelberg: Springer , 2001, p. 334-350.
- The Open Geospatial Consortium. Available from: <<http://www.opengeospatial.org/>> [February 2012].
- The Open Geospatial Consortium, *CityGML*. Available from: <<http://www.opengeospatial.org/standards/citygml>>[February 2012].
- The Open Geospatial Consortium, *GML*. Available from: <<http://www.opengeospatial.org/standards/gml>>[February 2012].

Observatoire virtuel pour diffuser et réutiliser l'expertise en télédétection

Laurent Demagistri* — Yuan Lin* — Thérèse Libourel*

* UMR ESPACE-DEV (IRD-UM2-UAG-ULR)
500 Avenue JF Breton
F-34000 Montpellier - nom.prenom@ird.fr

RESUME. L'observation de la Terre et de l'Espace à l'aide d'images satellitaires donne lieu à la mise en place de plusieurs organisations : "réseaux" de stations de réception. Leur principale vocation est de diffuser les produits bruts ou élaborés (les images). Cependant pour augmenter les potentialités offertes aux chercheurs, il serait souhaitable que ceux-ci puissent facilement décrire et exécuter leurs propres expérimentations en concentrant ainsi leur effort sur les traitements qu'ils souhaitent effectuer sur les données. L'objectif que nous nous sommes fixé est de faciliter l'expression et l'exécution de telles expérimentations par la mise en place d'observatoires virtuels conçus à partir des réseaux de stations de réception existants. Notre proposition sera détaillée sur un exemple traité à partir du réseau SEASnet.

ABSTRACT. Earth and Space observation by means of satellite images gives rise to the implementation of several organizations: "networks" of receiving stations. Their main vocation is to disseminate raw or elaborated products (the images) to users. However to increase the potentialities that they are given, researchers should also be offered the ability to easily describe and execute their own experiments on the data. They could thus concentrate their effort on the processings that they are designing. We settled the objective of facilitating the definition and the execution of such experiments by the implementation of virtual observatories designed from the existing networks of receiving stations. Our proposal will be detailed on an example treated from the SEAS network (SEASnet).

MOTS-CLÉS : Image satellitaire, Chaîne de traitements, Infrastructure de mutualisation, Environnement de Workflow

KEYWORDS: Satellite image, Processing line, Infrastructure, workflow environment

1. Introduction

L'imagerie satellitaire constitue une source d'information précieuse pour la connaissance et la gestion des observations de phénomènes terrestres. En effet, l'image satellite donne aux chercheurs la possibilité de croiser une vision globale d'une zone avec l'observation plus locale de terrain. Les données brutes recueillies nécessitent un ensemble de traitements avant d'être utilisables. De plus leur pérennisation demande des supports autorisant la sauvegarde et la diffusion aisée. De véritables infrastructures sont donc nécessaires et celles-ci doivent respecter les standards actuels à des fins d'interopérabilité. La mise en place de ce type d'infrastructure autorisant la mutualisation de ces données ainsi que celle des traitements d'analyse adéquats a constitué une des priorités que ce sont fixées l'IRD (Institut de Recherche pour le Développement), et plus précisément l'unité mixte de recherche Espace-Dev, pour répondre aux enjeux actuels liés à l'observation de la Terre et de l'Espace.

La profusion des images actuellement accessibles fait qu'au delà des traitements standards qui doivent être appliqués aux images brutes, les défis concernent les chaînes de traitements plus élaborés qui doivent être établies pour construire des informations utiles ou indicateurs. Les systèmes dédiés aux observations à l'aide d'images satellitaires (que nous désignons observatoires) se multiplient et sont disponibles sur des sites distribués, ce qui complique le travail du scientifique qui souhaite résoudre des problèmes mettant en œuvre divers traitements et données dispersés sur ces divers observatoires. Le challenge que nous souhaitons relever est de concevoir une réponse pragmatique, prenant en compte ces différents aspects. C'est ce que nous désignons par *observatoire virtuel*. Un observatoire virtuel doit permettre aux chercheurs d'accéder aux données et aux traitements (quelle que soit leur localisation), et leur offrir les outils pour mettre au point leurs diverses expérimentations en explicitant et modélisant leur raisonnement.

L'article déclinera, en premier lieu, dans la section 2 la description générale de ce que l'on désigne par infrastructure d'acquisition ou réseau de stations dédiées à l'imagerie satellitaire. La section 3 présentera la notion d'observatoire virtuel, celle de contexte de travail et d'architecture de chaîne de traitements afférants établis à partir d'un bref état de l'art autour de travaux qui ont inspiré cette proposition. La section 4 présentera la déclinaison de notre proposition sur un exemple. La section 5 précisera les perspectives ouvertes.

2. Infrastructure de données satellitaires et traitements (Observatoire)

Notre équipe a eu la chance de mener à bien la mise en place de plusieurs projets d'installation de systèmes d'acquisition d'images satellites durant les vingt dernières années : acquisition d'images des satellites météorologiques à large champ (basse résolution) mais également d'images "haute résolution" de satellites optiques ou radar.

2.1. Les principes généraux

Quel que soit le type de système d'acquisition d'images satellites, trois grands sous-systèmes sont généralement présents :

- le sous-système "Antenne" (l'infrastructure physique de télécommunication et son pilotage),
- le sous-système de production des données images "brutes",
- le sous-système d'exploitation de ces données brutes.

Ces différentes parties sont détaillées sur la figure 1. Nous les décrivons succinctement dans les paragraphes qui suivent.

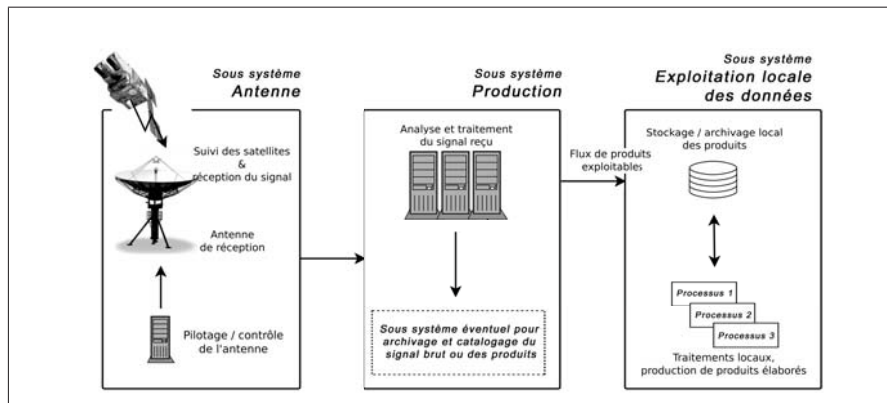


Figure 1 – Infrastructure locale d'acquisition de données satellites (pied d'antenne).

2.1.1. Réception du signal

Les satellites d'observation de la Terre les plus courants sont soit géostationnaires (apparemment immobiles par rapport au point au sol qu'ils survolent) soit à orbites polaires (ils survolent les pôles terrestres à chaque révolution, la rotation de la Terre entraînant un décalage de leur orbite par rapport aux zones survolées). Ils peuvent enregistrer à bord le signal correspondant aux mesures qu'ils effectuent en vol puis le restituer lorsqu'ils survolent une zone de réception ou bien émettre en permanence le signal mesuré. Une antenne motorisée programmée de manière adéquate pour suivre leur trajectoire peut acquérir ces données (fig. 1, sous-système antenne).

2.1.2. Production des images

Dans ce sous-système on trouve tous les équipements permettant de transformer en donnée exploitable pour les utilisateurs le signal brut issu du sous système "antenne". La plupart du temps il s'agit d'équipements "propriétaires" avec des logiciels spécifiques sur lesquels il est en principe impossible d'intervenir. Certains opérateurs

privés incluent à ce stade un système de catalogage des données acquises disponibles et des données produites (fig. 1, sous-système de production).

2.1.3. L'exploitation des données

Cette partie héberge l'ensemble des équipements mis en place pour exploiter, en fonction des besoins recensés, les données issues du système d'acquisition (fig. 1, sous-système d'exploitation des données).

Au minimum il s'agit d'un système de stockage suffisamment bien dimensionné : en effet, en fonction des caractéristiques des données acquises, le flux quotidien issu de l'antenne peut être conséquent.

Malgré une diffusion "grand public" de plus en plus répandue, les données satellitaires restent la plupart du temps largement sous-exploitées. Ce constat impose de mettre en place tous les outils permettant d'exploiter au mieux leur potentiel avec, en particulier, un effort particulier mis sur les capacités de diffusion : faciliter donc l'accès direct aux données (accès aux mesures "télédétectées") mais également publication sur internet via un catalogage normalisé de leurs métadonnées. Les utilisateurs potentiels ne disposant pas tous nécessairement d'outils de télédétection ou de traitement d'images suffisamment évolués, un objectif serait également de faciliter le traitement de haut niveau des données, via la mise en place d'infrastructures de traitement des données à distance.

Dans ces conditions, tout utilisateur potentiel, pourrait avoir accès non seulement à des données de télédétection mais également aux traitements standards de ce type d'images : prétraitements, corrections radiométriques, corrections géométriques, calcul d'indices, etc. De plus, il devrait aussi accéder aux produits dérivés mis en place par les producteurs à partir des données brutes. Ces produits dérivés, parfois complexes, constituent pour les utilisateurs des "boîtes noires" spécifiques et de ce fait ne sont pas facilement réutilisables. Pour aller au delà, il serait judicieux d'ouvrir ces "boîtes noires" afin d'offrir des traitements simples, réutilisables et combinables, à loisir, par les scientifiques. L'infrastructure d'exploitation doit donc comporter des serveurs de traitement puissants, afin d'offrir aux utilisateurs le loisir de concevoir leurs expérimentations, en connaissance de cause, à partir des différents traitements disponibles (simples / complexes).

2.2. Expérience de notre équipe

Dès 1991, dans le cadre de ses propres programmes de recherche et développement, l'Institut de Recherche pour le Développement (IRD) a développé et exploité plusieurs stations de réception d'images des satellites météorologiques à orbite polaire (NOAA/AVHRR et ORBview/SeaWiFS). Les capteurs à bord de ces satellites permettent d'acquérir en continu des données multispectrales. Ils ont une résolution spatiale à la verticale du satellite, de l'ordre du kilomètre et lors d'un passage au dessus d'une région, la fauchée (largeur de la zone au sol imagée) est de l'ordre de

2900 km. Ces caractéristiques font de ces systèmes d'acquisition de très bons outils de suivi à grande échelle de l'environnement. En 1999, les différentes stations se sont regroupées pour donner naissance au réseau SEASnet [PET 02]. La figure 2 présente la localisation des différentes antennes d'acquisition du réseau SEASnet avec une schématisation de leur couverture respective.

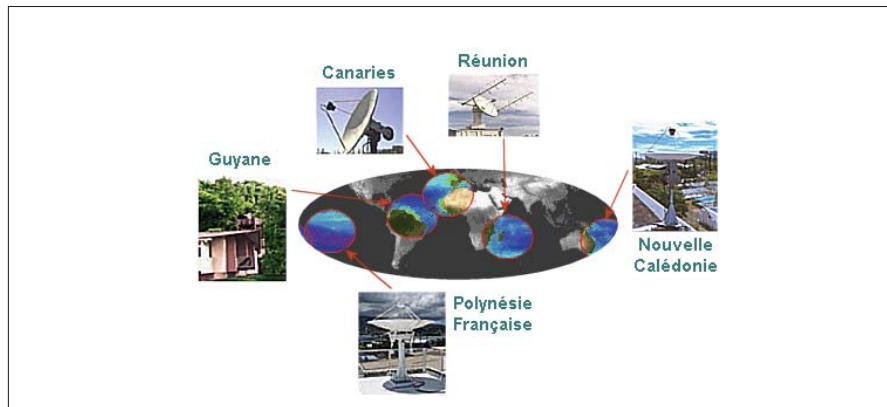


Figure 2 – Localisation des infrastructures d'acquisition du réseau SEASnet.

Cette initiative a été le point de départ de la réflexion que nous menons actuellement : comment mettre à la disposition des chercheurs un *observatoire virtuel* leur permettant la mise au point de diverses expérimentations à partir de ces infrastructures d'acquisition de données réparties. Un tel observatoire autorise la conception et la réalisation de chaînages entre divers services dans l'objectif de construction d'indicateurs pertinents pour l'observation terrestre.

Pour la diffusion de ces données et traitements, plusieurs choix étaient possibles :

- mettre en place un observatoire complet à chaque pied d'antenne,
- centraliser l'accès à ces divers observatoires sur un site dédié (situé à Montpellier)

Dans un premier temps, c'est le système centralisé qui a été préféré (sans créer de réels observatoires à chaque pied d'antenne) ; la figure 3 schématise cette configuration.

Les produits élaborés calculés en pied d'antenne de chacune des implantations locales sont transférés (via le protocole FTP) sur un serveur unique. Différents protocoles d'accès ont été mis en place sur ce dernier. Tout d'abord OPENDAP ¹, très répandu dans la communauté scientifique des océanographes et des météorologues mais également les protocoles de l'OGC ², tout particulièrement le WMS et le WCS (Web Map Service et Web Coverage Service).

1. Open-source Project for a Network Data Access Protocol : <http://opendap.org/>

2. Open Geospatial Consortium : <http://www.opengeospatial.org/standards/is>

6 Revue. Volume /

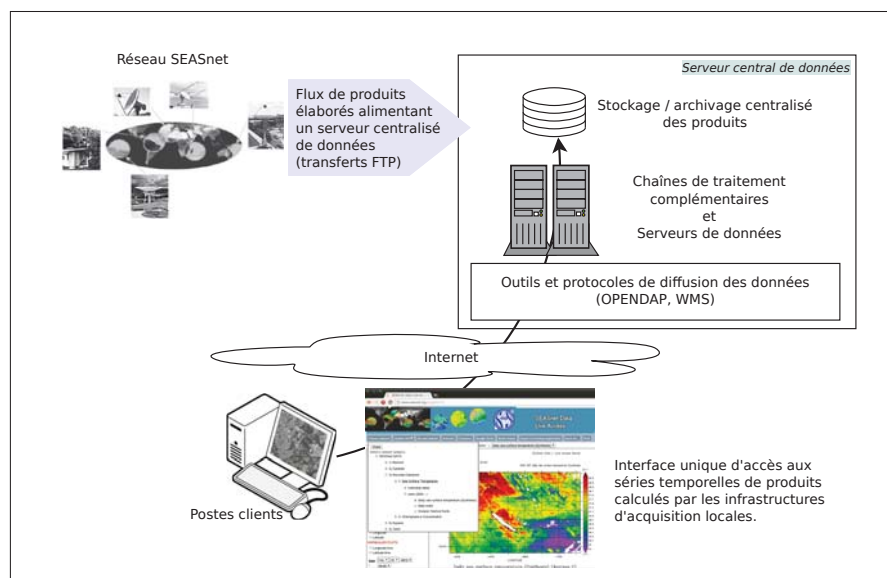


Figure 3 – Observatoire centralisé.

Des bases de métadonnées ont été élaborées afin d’annoter le flux de fichiers calculés quotidiennement. Elles respectent les normes courantes relatives à la gestion des métadonnées d’images d’observation de la Terre (normes ISO19115³ et ISO19115-2⁴). Les images sont indexées géographiquement et temporellement afin de constituer un outil parfaitement adapté à la gestion de séries temporelles de données cartographiques.

Par la suite, en 2006, une infrastructure d’acquisition de données à haute résolution spatiale a été implantée en Guyane, au centre IRD de Cayenne (réception de données optiques SPOT et radar ENVISAT). Le même type de plateforme a été installé à la Réunion en 2011 et, en 2013, le projet SEAS-Gabon⁵ permettra également la mise en place d’une telle infrastructure à Libreville au Gabon. Cette fois, une approche centralisée est difficilement envisageable en raison du poids des données qu’il faudrait transférer quotidiennement (plus de 500 Mo pour une image SPOT5 panchromatique et des performances de réseau pas toujours optimales).

Sur l’ensemble de ces installations l’objectif est de mutualiser les développements. Nous travaillons donc à l’harmonisation des systèmes d’exploitation et de diffusion

3. ISO 19115 :2003 : Information géographique – Métadonnées, http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020

4. ISO 19115-2 :2009 : Information géographique – Métadonnées – Partie 2 : Extensions pour les images et les matrices, http://www.iso.org/iso/fr/catalogue_detail.htm?csnumber=39229

5. <http://seas-gabon.espace-dev.fr/seasgabon/> et <http://www.ageos.org/ageos>

des données grâce à l'utilisation d'infrastructures de données spatiales (SDI) normalisées.

La démarche se veut la plus générique possible afin de tendre vers l'interopérabilité des systèmes.

3. Observatoire virtuel

Pour rendre l'architecture décrite précédemment encore plus générique, et se mettre réellement dans une configuration d'*observatoire virtuel* tel que nous l'imaginons, l'infrastructure de la figure 3 devra évoluer, au fur et à mesure des démarches de standardisation que nous menons, vers une architecture entièrement normalisée schématisée par la figure 4. L'accès aux données et traitements se fera directement sur les infrastructures informatiques de chaque site de réception d'images ; et ce de manière complètement transparente et virtuelle pour l'utilisateur.

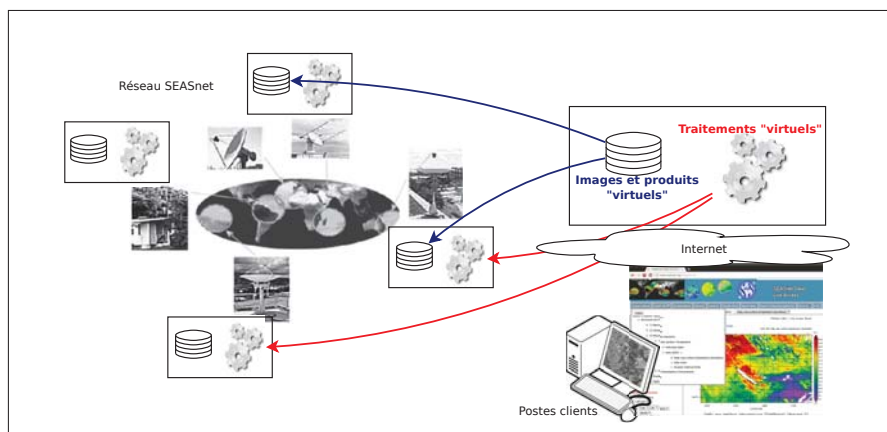


Figure 4 – Vers un observatoire virtuel SEASnet : accès direct et combinaison des données et traitements de chaque site.

Notre idée est que les chercheurs sont amenés à construire leurs propres expérimentations pour vérifier et valider leurs hypothèses. Pour construire ces expérimentations complexes, ils ont souvent besoin de rechercher les ressources adéquates et de les organiser ou réorganiser.

Un observatoire virtuel doit mettre à leur disposition ces données et les traitements existants (quelle que soit leur localisation) ; il doit en outre leur offrir un environnement pour concevoir et réaliser les expérimentations correspondant aux raisonnements qu'ils envisagent de mettre en œuvre.

Chaque plan d'expérimentation doit pouvoir être conçu, sauvegardé afin de pouvoir être réexécuté plusieurs fois soit dans diverses configurations, soit sur diverses données de test.

Ces deux observations nous amènent tout naturellement aux notions de workflows et de contextes de travail associés.

3.1. *État de l'art*

Nous établissons ici un rapide état de l'art autour des propositions existantes en termes de protocole d'accès aux données et de workflow scientifique.

3.1.1. *Diffusion de données*

Beaucoup de produits et images de télédétection sont actuellement accessibles sur internet. Plusieurs organisations mettent en ligne des sites de diffusion de données. Trois exemples nous paraissent particulièrement pertinents :

- Le site du "Physical Oceanography Distributed Active Archive Center" (PO.DAAC⁶) fournit gratuitement des produits d'océanographie physique (courants, température, vents, salinité, etc.).
- Le site du "Global Land Cover Facility" (GLCF⁷) de l'Université du Maryland met à disposition des utilisateurs des produits et des données pour l'étude de la Terre et de l'environnement, particulièrement les images des satellites LANDSAT⁸.
- Du côté européen, l'Organisation européenne pour l'exploitation des satellites météorologiques (EUMETSAT⁹) fournit sur son site des données, images et produits satellitaires dédiés à la météorologie et la climatologie.

La plupart du temps ces sites proposent de récupérer leurs fichiers via des protocoles standards afin que les utilisateurs puissent les exploiter localement sur leur poste. En fonction du type de données, ces fichiers peuvent être lourds et nécessitent un certain degré de performances en termes de connexion réseau ou de capacité de stockage. Le protocole le plus couramment utilisé est le transfert de fichiers "FTP", certes simple d'utilisation mais ne permettant pas de mettre en place des outils élaborés de recherche de données.

Le protocole **OPeNDAP**, cité précédemment, s'est beaucoup développé dans les communautés scientifiques de l'océanographie et de la météorologie, donnant naissance à différents outils, quasiment "clé en main", permettant une diffusion de données sur un serveur WEB à l'aide de formulaires de requête. Nous citons ici ceux qui ont été utilisés dans le cadre du réseau SEASnet : le *middleware* **THREDDS** (Thematic Realtime Environmental Distributed Data Services¹⁰) permettant d'activer un

6. Le PO.DAAC est un service du "Jet Propulsion Laboratory", un laboratoire dépendant de la NASA - <http://podaac.jpl.nasa.gov/dataaccess>

7. <http://glcf.umiacs.umd.edu/data/landsat/>

8. <http://landsat.gsfc.nasa.gov/>

9. <http://www.eumetsat.int/Home/Main/DataAccess/index.htm>

10. <http://www.unidata.ucar.edu/projects/THREDDS/>

service OPeNDAP sur un serveur d'applications et le **LAS** (Live Access Server¹¹) encapsulant les opérations THREDDS de base et offrant des possibilités plus élaborées de recherche, d'exploitation voire de traitement en ligne des données.

Les protocoles de l'*OGC*, également cités précédemment, sont de plus en plus utilisés, particulièrement en lien avec l'effort de normalisation des données géographiques mené conjointement avec l'*ISO*.

3.1.2. Environnements de workflow scientifique

La notion de workflow scientifique existe depuis longtemps ; en traitement d'images, des outils comme Khoros [KON 94] ont montré le chemin en insistant sur la notion d'outil dédié à la conception de chaînes de traitements et à leur paramétrisation au travers d'expérimentations itératives.

Une rapide analyse, issue des travaux de [LIN 27] nous a permis de sélectionner un ensemble de projets et de démarches existants. Chacun des projets de workflow (scientifique ou généraliste) retenus propose un environnement de travail reposant sur une interface graphique et des formalismes divers, grâce auquel les utilisateurs construisent leurs plans de travail.

Kepler [LUD 06] est un environnement complet pour le workflow scientifique construit sur la plateforme Ptolemy II de l'Université de Berkeley. Dans cet environnement, les *acteurs* correspondent aux différents traitements et opérations envisageables, et ils sont dotés de *ports* qui représentent leurs paramètres d'entrée / sortie. Les compositions entre traitements sont faites interactivement par les scientifiques en reliant les *ports* d'*acteur* par des *chanel*s. Le contrôle et l'orchestration de modèle de workflow est confié à des *directeurs*. Les adaptations nécessaires sont réalisées par des programmes intermédiaires (*senders* et *receivers*), qui assurent la compatibilité des données transférées sur un *chanel*.

Taverna [HUL 06] est un projet de workflow initialisé par l'équipe *myGrid* en Angleterre, utilisé principalement dans les domaines biologiques. Les traitements dans cet environnement sont essentiellement des services web (auxquels peuvent s'adjoindre des bibliothèques locales, des scripts manuscrits, etc.). Lors de la composition de traitements, l'utilisateur apparie manuellement les paramètres d'entrée / sortie des services web, ou invoque des *shim services*, adaptateurs spécifiques conçus antérieurement à partir d'appariements établis dans des expérimentations déjà construits et testés.

NetBeans est un environnement IDE généraliste, dont un des modules intégrés permet de construire un workflow par composition de services web en utilisant le langage BPEL (Business Process Execution Language) [AND 03]. Une bonne connaissance du standard BPEL est un pré-requis. La composition se fait par appariement ou transformation manuelle entre éléments de messages échangés, ces règles d'apparie-

11. LAS (<http://ferret.pmel.noaa.gov/LAS>) est un logiciel Open-Source développé par un laboratoire de la NOAA (National Oceanic and Atmospheric Administration)

10 Revue. Volume /

ment sont ensuite traduites à l'aide du langage XSLT (eXtensible Stylesheet Language Transformations) [KAY 07].

Weka [CUN 93] est une application du domaine apprentissage et fouille de données, réalisée par l'Université Waikato de Nouvelle Zélande. Elle comporte un composant **Weka KnowledgeFlow**, qui permet un chaînage de traitements relatifs aux expérimentations de fouille de données. Le modèle général de KnowledgeFlow obéit à la chaîne *Sélectionner les données* → *Filtrer* → *Classifier* → *évaluer* → *Visualiser*. Grâce à l'interface graphique, les scientifiques concrétisent interactivement leurs expérimentations, et choisissent les convertisseurs pré-établis pour assurer la compatibilité de leurs workflows. L'environnement repose sur des catégories de données et d'algorithmes relatifs aux différents traitements précédemment construites.

Tous ces projets disposent d'un environnement plus ou moins élaboré, et nous proposons de formaliser celui-ci sous la forme de ce que nous désignons par *contexte de travail*.

3.2. Contexte de travail

Le contexte de travail¹² (fig.5) proposé est composé de trois grandes organisations de descriptions de ressources, dénommées :

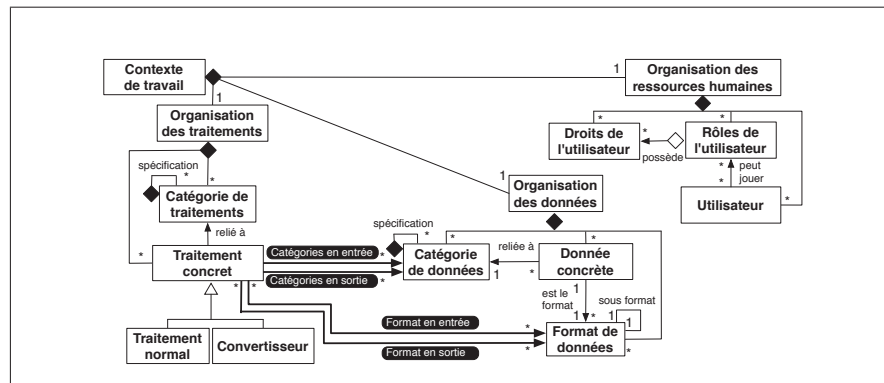


Figure 5 – Contexte de travail

- *Organisation des ressources humaines*, qui gère la description des utilisateurs de la plateforme ainsi que celles de leurs différents rôles et droits d'accès associés,
- *Organisation des données*, qui gère la description des catégories de données, des données concrètes et des divers formats de données associés¹³.

12. Pour simplifier la figure, on a noté traitements, données et ressources humaines, alors qu'il s'agit de descriptions de ces ressources.

13. Remarque : il est à noter que plusieurs catégories de données peuvent partager le même format.

12 Revue. Volume /

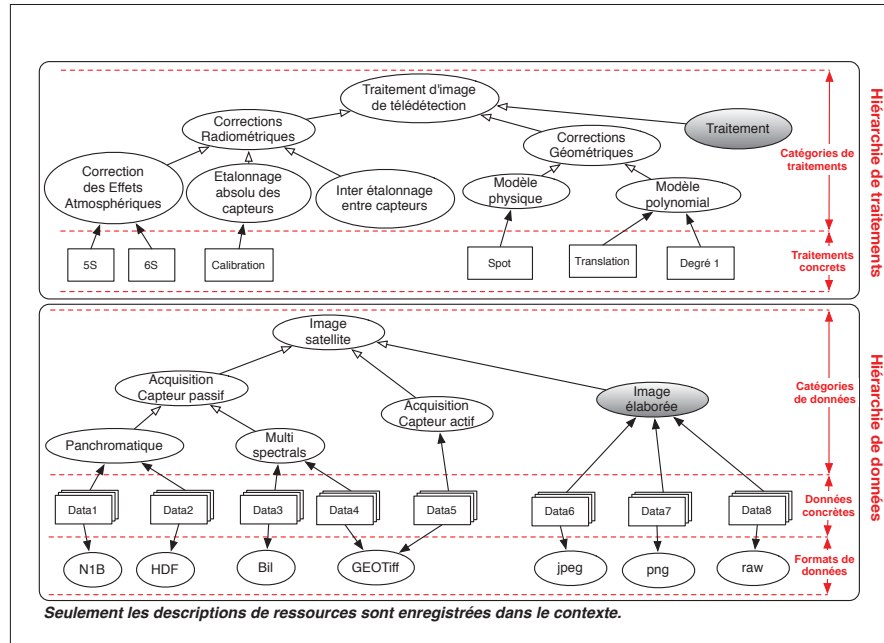


Figure 7 – Un exemple de contexte de travail

Pour une meilleure compréhension de la figure 7, nous explicitons ci-dessous les éléments qui la constituent.

Dans le domaine de l'imagerie satellitale [CAL 01], il est d'usage de distinguer les capteurs passifs des capteurs actifs. Les capteurs actifs, émettent un signal électromagnétique en direction de la cible (le sol Terrestre) et enregistrent la quantité de signal rétrodiffusé par celle-ci. Les capteurs passifs quant à eux, enregistrent le signal électromagnétique émis (cas des capteurs thermiques) ou réfléchi par le sol (réflexion du rayonnement solaire). En raison des processus physiques impliqués, les deux technologies induisent des types d'images différents tant par leur contenu que par leur analyse et les traitements pour les exploiter. Dans la catégorie des capteurs passifs, on peut imaginer faire la distinction entre les capteurs *panchromatiques* et les *multispectraux* :

- Les capteurs panchromatiques mesurent le signal émis par chaque portion élémentaire du terrain, les pixels, sur une large fenêtre spectrale (par exemple tout ou partie du spectre de la lumière visible) ; ils produisent une image comportant une seule "bande".
- Les capteurs multispectraux font cette mesure sur plusieurs fenêtres spectrales de largeurs variables (par exemple les bandes spectrales du rouge, du vert, du proche infrarouge, etc.) ; ils produisent une image comportant plusieurs canaux.

La résolution des capteurs étant dépendante des performances spectrales, les capteurs panchromatiques présentent un meilleur pouvoir séparateur (la capacité à distinguer deux éléments distincts du sol espacés d'une certaine distance). Quoi qu'il en soit, ce paramètre sera bien évidemment un attribut de la classe "Image satellite".

De la même manière que pour les données, on peut essayer de bâtir une classification des traitements habituellement réalisés à partir d'une image satellite. Quel que soit le mode d'acquisition, on se retrouve confronté à des problèmes de corrections géométrique et radiométrique des images : les prétraitements [CAL 01].

Les corrections géométriques visent à corriger les déformations liées à l'angle d'acquisition, à la courbure de la Terre, etc., afin de mettre l'image en conformité géographique (projection de l'image) ou pour la rendre superposable à une source de référence (par exemple une autre image). On distingue les modèles physiques qui modélisent précisément la géométrie d'acquisition du capteur et les modèles polynomiaux plus génériques.

Les corrections radiométriques ont pour objectif la calibration des valeurs numériques contenues dans les images afin d'homogénéiser les données acquises à différentes dates ou par différents capteurs [BRO 85]. On range dans cette catégorie la correction des effets atmosphériques, la calibration "absolue" permettant de convertir les "comptes" numériques (ou niveaux de gris) d'une image en valeurs géophysiques (albedo, températures de brillance, coefficient de rétrodiffusion, etc.), l'égalisation radiométrique entre plusieurs images, etc.

Les images pré-traitées subissent des "Traitements" pour constituer des "Images élaborées" (en grisé sur la fig.7).

Un ensemble de formats de données (*NIB, HDF, Bil, GEOTiff, jpeg, png, raw*) est aussi présenté. Lors de l'enchaînement des traitements, prétraitements radiométriques et géométriques d'une part puis traitements élaborés d'autre part, des images intermédiaires sont produites dans ces formats d'images standards.

3.3. Chaîne de traitements (workflow)

La réflexion menée a été guidée par le point de vue "métier" (expérimentateurs). Réaliser un protocole expérimental correspond au modèle général comportant trois étapes : 1) **Définition** : définition abstraite d'une chaîne de traitements correspondant à une expérimentation (planification d'expériences), 2) **Instanciation** : définition plus spécifique après identification des divers éléments de la chaîne (données / traitements), 3) **Exécution** : exécution personnalisée (selon des stratégies correspondant aux besoins).

3.4. Architecture

C'est à partir de ce cycle de vie expérimental et inspirés par les styles architecturaux proposés par l'Object Management Group [OMG 06], que nous avons proposé pour le composant workflow la vision architecturale à 3-niveaux (fig.8) :

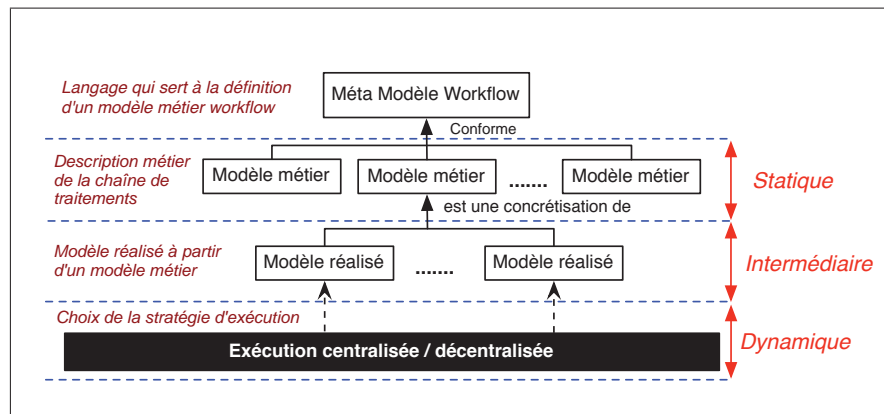


Figure 8 – Architecture du composant workflow

1) Le niveau **statique** concerne la phase de conception. Il s'agit de construire des modèles de traitement métier (abstraits) à partir d'un langage simple.

2) Le niveau **intermédiaire** traduit une phase d'instanciation et de pré-contrôle. à partir du modèle du traitement métier, l'utilisateur construira la chaîne de traitements réelle en déterminant et localisant les traitements et les données les plus adéquates à l'expérimentation visée. Cette phase sera réalisée à l'aide de moteur de recherche à partir des descriptions de ressources nécessaires (fig.6).

3) Le niveau **dynamique** concerne la phase d'exécution proprement dite. Celle-ci peut alors se dérouler à partir de différentes stratégies définies à la fois par l'expérimentateur et par les configurations opérationnelles.

Le niveau *statique* a été approfondi et étudié dans [LIN 09, LIB 10].

Il se base sur un langage défini par un méta modèle dont les éléments abstraits *tâches* ou *traitements* sont reliés par les liens unidirectionnels, et par l'intermédiaire de *port*. Pour faciliter la manipulation des chaînes de traitements abstraits, un langage graphique correspondant a été proposé (fig.9) ; il sera utilisé dans les exemples ultérieurs.

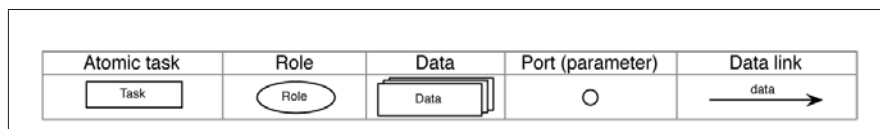


Figure 9 – Extrait des éléments essentiels de notre langage graphique

4. Illustration

4.1. Données et traitements de départ

Pour illustrer l'intérêt de l'architecture proposée ici, nous détaillons en exemple la chaîne de traitements aboutissant à l'obtention de cartes de température de surface de la mer à partir des données issues de notre réseau de stations d'acquisition d'images basse-résolution. Cette chaîne de traitement a été déployée dans chacune de nos implantations "pied d'antenne" basse-résolution.

Les données sont acquises via la réception des satellites américains NOAA¹⁵ (capteurs AVHRR-2 et AVHRR-3). Trois sont actuellement considérés comme opérationnels. Les mesures satellitaires sont réalisées par des radiomètres multicanaux capables d'enregistrer le rayonnement électromagnétique de la Terre selon six longueurs d'onde.

Canal	Résolution au Nadir (km)	Longueur d'onde (μm)
1	1.09	0.58 - 0.68
2	1.09	0.725 - 1.00
3A	1.09	1.58 - 1.64
3B	1.09	3.55 - 3.93
4	1.09	10.30 - 11.30
5	1.09	11.50 - 12.50

Tableau 1 – Résolution et longueurs d'onde des canaux AVHRR-3.

Ces satellites étant *héliosynchrones à orbites polaires*, ils permettent un survol plusieurs fois par jour d'une même région. Ils ont la particularité d'émettre leur signal enregistré à bord en continu. De fait, lorsque la zone d'émission d'un satellite intersecte le cercle de réception d'une antenne d'acquisition, une image correspondant à la trace au sol du satellite est enregistrée par les infrastructures en pied d'antenne. Dans des conditions optimales, il est possible d'acquérir plus de 10 images NOAA quotidiennement (tous satellites confondus). Les infrastructures d'acquisition du réseau

15. NOAA KLM USER'S GUIDE : <http://www2.ncdc.noaa.gov/docs/klm/>
 NOAA Satellite Information System : <http://noaasis.noaa.gov/NOAASIS/>

16 Revue. Volume /

SEASnet ont permis la réception des données de la série de satellites météorologiques américains NOAA : de 1991 avec NOAA-12, jusqu'à maintenant avec NOAA-15, 18 et 19, les satellites actuellement encore exploitables.

Cette multiplicité d'images quotidiennes est intéressante dans un contexte inter-tropical à fort couvert nuageux car des processus de synthèse de données peuvent être envisagés. C'est cette chaîne de traitements précisément qui est utilisée en guise d'exemple dans ce qui suit et qui pourrait donner lieu, dans le cadre d'un observatoire virtuel tel que nous l'envisageons, à diverses adaptations ou expérimentations.

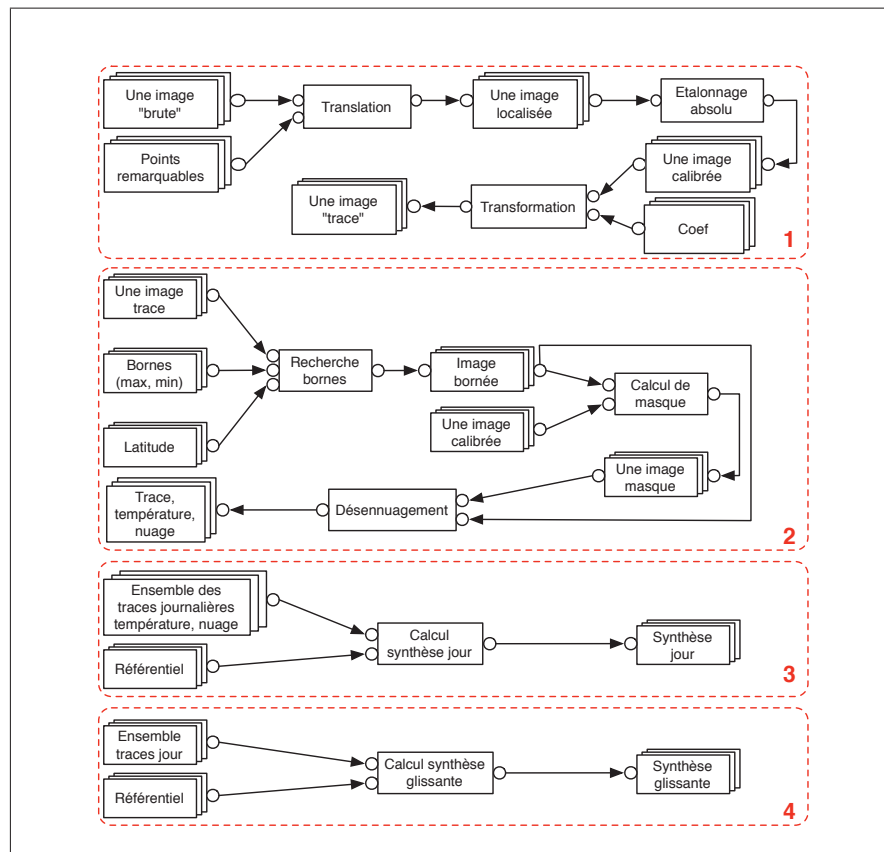


Figure 10 – Chaînes de traitements abstraites

La description complète de l'ensemble des traitements des quatre blocs (fig.10) est la suivante (seul le bloc 1 sera détaillé aux deux niveaux abstrait et concret conformément à la figure 11¹⁶).

16. Certains maillons de la chaîne peuvent être court-circuités comme le suggère la partie ombrée de la figure.

– Bloc 1 : Avant tout, il faut préciser que la localisation géographique des données NOAA n'est pas précise car il n'y a pas de GPS à bord du satellite pour affiner le positionnement orbitographique des images. Ces dernières sont donc générées à partir du modèle orbitographique prévisionnel utilisé pour le suivi du satellite par l'antenne d'acquisition. Une image NOAA est constituée d'un entête contenant toutes les informations relatives aux conditions d'acquisitions et aux paramètres de calibration du capteur. Viennent ensuite l'ensemble des enregistrements associés à chaque ligne de l'image. Pour chaque ligne, une information de localisation des pixels qui la compose est disponible en début d'enregistrement. Cette donnée n'est fournie que pour un sous échantillonnage de pixels (tous les 40 points). Une première lecture complète de l'image permet de récupérer ces informations afin de constituer une grille de localisation des pixels. La localisation de cette grille peut être ajustée par simple translation. Une interface utilisateur a été développée afin de visualiser l'image et d'afficher suivant cette grille de localisation la position du trait de côte. L'ajustement de ce dernier sur l'image, à condition bien évidemment qu'il soit visible dans l'image, permet d'ajuster le positionnement global de l'image. La correction utilisée pour ajuster le trait de côte est répercutée à l'ensemble des points de la grille de localisation et sauvegardée dans le fichier. Nous modélisons cette étape concrète de *localisation* de l'image par le traitement abstrait *Translation*. Le traitement prend en entrée une image NOAA au format N1B et un ou plusieurs points de localisation. Nous obtenons en sortie une image au format N1B dont le géoréférencement a été ajusté.

Le calcul d'une valeur de température sur un pixel donné de l'image est réalisé selon une procédure standard bien documentée dans la littérature : la "Multi Channel Sea Surface Temperature" (MCSST). Il s'agit d'appliquer un algorithme de *split window* prenant en compte les températures de brillance des canaux thermiques 4 et 5 des images NOAA ainsi que l'angle de prise de vue du capteur par rapport au sol [MCC 85]. Pour pouvoir travailler avec des températures de brillance des canaux 4 et 5, il est nécessaire de passer par une étape de calibration des comptes numériques de l'image (les niveaux de gris). Cette étape de traitement, notée *Etalonnage absolu* au niveau abstrait (bloc 1 de la fig.10 et fig.11) prend au niveau concret (bas de la fig.11) une image N1B et donne en sortie une image dans un format RAW. Le calcul de MCSST sur l'image calibrée avec un jeu de coefficients donne une image de température en sortie au format PNG (l'image "trace").

– Bloc 2 : Dans le contexte intertropical de notre réseau de stations de réception, la couverture nuageuse importante est souvent problématique pour les applications de cartographie. Dans les algorithmes de calcul de cartes de température de l'eau, les nuages parasitent les résultats car ils apparaissent beaucoup plus froids que la mer. Le calcul d'un masque des nuages est nécessaire si l'on veut pouvoir réaliser des synthèses entre différentes images. L'étape 1 est modélisée par le traitement *Recherche de bornes* dans l'image (fig.10). Les entrées sont : l'image de température et les paramètres permettant de filtrer ces valeurs en fonction de la latitude. On obtient en sortie une image bornée. L'étape 2 est modélisée par le traitement *Calcul de masque*. Il prend en entrée une image de température et une image NOAA calibrée (afin de pouvoir accéder aux températures de brillance). On obtient en sortie une image de

18 Revue. Volume /

température et un masque que l'on croise avec le traitement désenuagement. En sortie de ce bloc on obtient des images de températures dans lesquels ont été masqués les pixels correspondant à des zones nuageuses.

– Bloc 3 : Nous l'avons vu précédemment, les satellites NOAA survolent plusieurs fois par jour la même région. Théoriquement 2 passages pendant la journée et 2 passages pendant la nuit. Durant la journée, l'ensoleillement provoque un réchauffement de la surface de la mer. Ainsi, pour calculer à partir des images, des valeurs de température plus proches de la réalité, il conviendrait de ne considérer que les images acquises durant la nuit et donc ne pas exploiter 50% des données acquises en pied d'antenne. Pour utiliser la totalité des images, nous avons mis au point une procédure de correction des images acquises durant la journée en prenant comme référence les températures calculées sur les images de nuit. Le croisement de toutes ces données réalise une carte quotidienne des températures : les images sont projetées dans un référentiel cartographique et la valeur de température effectivement retenue est la valeur maximale sur l'ensemble des images de la journée.

– Bloc 4 : Au terme de la synthèse décrite ci-dessus, calculée quotidiennement (au jour J), la couverture nuageuse reste importante. Pour réduire son étendue, un second processus de synthèse est envisagé et prend en compte 5 jours de données. Cette étape est modélisée par le traitement *Calcul synthèse glissante* (synthèse des 5 images quotidiennes en se centrant sur le jour J et en recherchant une valeur désenuagée à J, J-1, J+1, J-2 ou J+2 successivement).

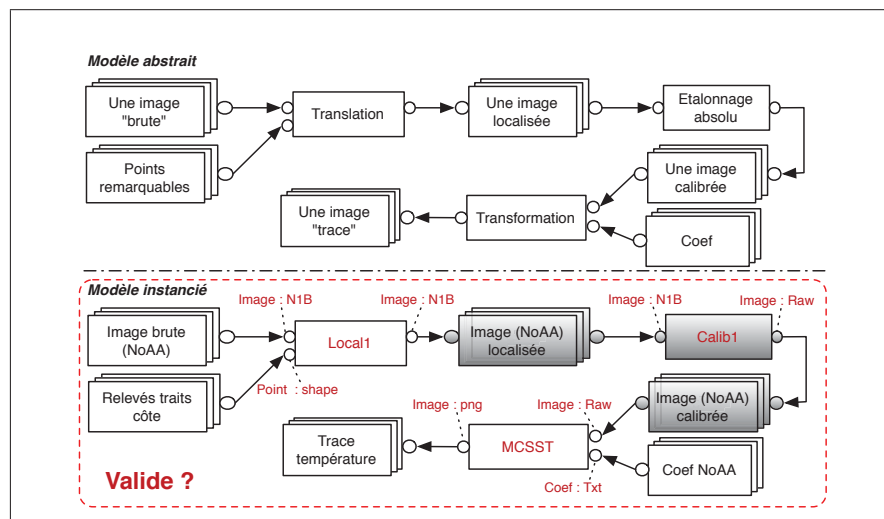


Figure 11 – Instanciation

Après la description des traitements, se pose la question de la validité de cette chaîne concrète (bas de la fig.11). Pour prendre en compte celle-ci, nous transcrivons la chaîne concrète formellement comme une suite de compositions de signatures de traitements. Une signature sera exprimée par :

Nom (liste de Param_E) : (liste de Param_S), avec

- **Nom** correspond à la désignation du traitement ou de l'opération concrète.
- **Param_E**, **Param_S** correspondent respectivement à un paramètre d'entrée et à un paramètre de sortie. Chaque paramètre **p** est formalisé (**dc :fo**), avec **dc** et **fo** relatifs aux catégories et formats de données utilisés.

Une composition $p1 \rightarrow p2$ est le lien reliant un paramètre **p1** de sortie d'un traitement **T1** à un paramètre **p2** d'entrée d'un traitement **T2**.

Signature et composition nous permettent d'introduire la notion de compatibilité, sur laquelle reposera la validation de la chaîne concrète. L'objectif est d'assurer l'exécution sans faute de la chaîne concrète.

5. Conclusion et perspectives

La mise en place d'observatoire virtuel reposant sur des infrastructures distribuées devient, à nos yeux, incontournable. De nombreux domaines scientifiques, adoptent déjà cette démarche (astronomie, biologie, etc.). Les travaux déjà effectués, tant sur le plan de la réalisation des infrastructures, de l'accessibilité aux données et aux traitements (via des protocoles standardisés) que de la réflexion menée autour du contexte de travail, nous semble autant de maillons nécessaires à cette mise en place.

L'architecture workflow proposée est simple mais nous semble judicieuse car elle a pu être confrontée au domaine des images satellitaires mais aussi à d'autres domaines scientifiques (hydrologie, bioinformatique, etc. [LIN 27]). Les recherches ultérieures envisagées portent sur la validation des chaînes concrètes, c'est-à-dire sur les définitions relatives aux divers cas de *compatibilités*. L'étape ultime consistera à exécuter les chaînes concrètes valides au sein de l'observatoire virtuel (prises en compte de la virtualité et de la distribution).

6. Bibliographie

- [AND 03] ANDREWS T., CURBERA F., DHOLAKIA H., GOLAND Y., KLEIN J., LEYMAN F., LIU K., ROLLER D., SMITH D., THATTE S., TRICKOVIC I., WEERAWARANA S., Business Process Execution Language for Web Services, Version 1.1 , , 5 May 2003.
- [BRO 85] BROWN O., BROWN J., EVANS R., Calibration of Advanced Very High Resolution Radiometer Infrared Observations , *Journal of Geophysical Research*, , 1985.
- [CAL 01] CALOZ R., COLLET C., *Précis de Télédétection, Traitements numériques d'images de télédétection*, vol. 3, Presses de l'Université de Québec, Agence Universitaire de la Francophonie, 2001.
- [CUN 93] CUNNINGHAM S., DENIZE P., A Tool for Model Generation and Knowledge Acquisition , *Proc International Workshop on Artificial Intelligence and Statistics*, Fort Lauderdale, Florida, USA, 1993, p. 213-222.

20 Revue. Volume /

- [HUL 06] HULL D., WOLSTENCROFT K., STEVENS R., GOBLE C., POCOCK M., LI P., OINN T., Taverna : a tool for building and running workflows of services , , 2006.
- [KAY 07] KAY M., XSL Transformations (XSLT) Version 2.0 , , 23 January 2007.
- [KON 94] KONSTANTINIDES K., RASURE J., The Khoros software development environment for image and signal processing , *Image Processing, IEEE Transactions on*, vol. 3, n° 3, 1994.
- [LIB 10] LIBOUREL T., LIN Y., MOUGENOT I., PIERKOT C., DESCONNETS J.-C., A Platform Dedicated to Share and Mutualize Environmental Applications , *ICEIS (1)*, 2010, p. 50-57.
- [LIN 27] LIN Y., Méthodologie et composants pour la mise en œuvre de workflows scientifiques , PhD thesis, Université Montpellier II. École doctorale I2S, 2011.12.7.
- [LIN 09] LIN Y., LIBOUREL T., MOUGENOT I., A Workflow Language for the Experimental Sciences , *ICEIS (3)*, 2009, p. 372-375.
- [LUD 06] LUDÄSCHER B., ALTINTAS I., BERKLEY C., HIGGINS D., JAEGER E., JONES M. B., LEE E. A., TAO J., ZHAO Y., Scientific workflow management and the Kepler system , *Concurrency and Computation : Practice and Experience*, vol. 18, n° 10, 2006, p. 1039-1065.
- [MCC 85] MCCLAIN E., PICHEL W., WALTON C., Comparative performance of AVHRR-based multichannel sea surface temperatures , *Journal of Geophysical Research*, , 1985.
- [OMG 06] OMG, *Meta Object Facility (MOF) Core Specification, Version 2.0, formal/06-01-01*, January 2006.
- [PET 02] PETIT M., RAMOS A., From receiving stations to concept : Survey of the Environment Assisted by Satellite (SEASnet), a net for research and development , *Proceedings of ESA Workshop GODAE GHRSSST, ESRIN, 02-04*, 2002.

Can satellites help organic crop certification?

Les satellites peuvent-ils aider la certification des cultures biologiques ?

**Antoine DENIS * — Baudouin DESCLEE ** — Silke MIGDALL
*** — Herbert HANSEN ** — Heike BACH ** — Pierre
OTT**** — Bernard TYCHON***

** Water, Environment and Development Unit, Arlon Campus Environment,
University of Liege, Belgium,
185 Avenue de Longwy, 6700 Arlon, Belgium
Antoine.Denis@ulg.ac.be*

*** KEYOBS SA
CAP Business Center, 31 Rue d'Abhooz, 4040 Herstal, Belgium
info@keyobs.com*

**** VISTA GmbH
51, Gabelsbergerstraße, D – 80333 München, Germany
mail@vista-geo.de*

***** ECOCERT SA
BP 47, Lieu dit Lamothe, 32600 L'Isle Jourdain, France
pierre.ott@ecocert.com*

ABSTRACT Organic agriculture, while producing healthy food and contributing to protect the environment, needs to be certified in order to meet the consumers confidence. The objective of this study was to investigate how earth observation techniques could enhance the crop certification process and in particular the possibility to discriminate organic and conventional fields. These different crop management methods results in crop biophysical differences which are supposed to be observable by earth observation techniques. A set of satellites with varying spatial and spectral resolution was used to compute discriminant indicators. Under the best conditions the results show a complete discrimination between organic and conventional crops.

RÉSUMÉ. L'agriculture biologique, alors qu'elle fournit une nourriture saine et contribue à protéger l'environnement, se doit d'être certifiée afin de rencontrer la confiance des consommateurs. L'objectif de cette étude était d'investiguer comment les techniques d'observation de la terre pourraient améliorer le processus de certification des cultures et en particulier la possibilité de discriminer les champs cultivés en mode biologique et

conventionnel. Ces différents modes de gestion des cultures résultent en des différences biophysiques supposées être observables par les techniques d'observation de la terre. Un ensemble de satellites de diverses résolutions spatiales et spectrales a été utilisé pour calculer des indicateurs discriminants. Les résultats montrent, dans les meilleurs cas, une discrimination complète entre les cultures biologiques et conventionnelles.

KEYWORDS: *organic crop certification, remote sensing, wheat, corn, hyperspectral, biophysical indicators, organic agriculture.*

MOTS-CLÉS: *certification biologique des cultures, télédétection, blé, maïs, hyperspectral, indicateurs biophysiques, agriculture biologique.*

1. Introduction

Organic agriculture provides healthy food and protects the environment by practising methods that avoid the widespread dissemination of chemicals. Crops can only be labelled as “organic” if they are produced according to a set of standards, undergo an evaluation and pass a yearly inspection. Since organic and conventional crops are treated differently (fertilizer, pesticide), their biophysical characteristics are also different and can be studied by earth observation (EO) techniques.

2. Objectives

The purpose of this study was to test the capacity of EO satellites to spot differences between organic and conventional crops in order to support the organic crops certification process.

3. Methodology

The concept was tested on winter wheat and corn grown in relatively large and flat fields in eastern Germany.

Multi- and hyperspectral satellite imagery of very high to medium spatial resolution (Table 1) was used to derive several indicators aimed to differentiate conventional and organic practises. Images from five different EO sensors were acquired during the study (SPOT4, Kompsat-2, Landsat 5 TM, Chris-PROBA, Worldview-2).

Both spatial and spectral content of the images were exploited to derive indicators based on biophysical justification and crop management practises. Three groups of indicators were investigated:

- **Spectral indicators:** organic crop management techniques use less fertilizer and pesticide, do not use growth regulators, their sowing density is often smaller and their crop variety different from conventional ones. These differences were studied

by spectral indicators related to crop nitrogen and chlorophyll content or field biomass.

Satellite sensors	CHRIS - PROBA	WorldView2	Kompsta2	SPOT4-5	Landstat5
Spectral bands	64	8 (XS) 1 (Pan)	4 (XS) 1 (Pan)	4 (XS) 1 (Pan)	7
Spectral range	VIS-NIR	VIS-NIR	VIS-NIR	VIS-NIR-SWIR	VIS-NIR-SWIR
Spatial resolution	34 m	2 m (XS) 0.5 m (Pan)	4 m (XS) 1 m (Pan)	20 m (XS) 10 m (Pan)	30 m (b6: 120 m)

Table 1. Main features of the satellite sensors used in this study. VIS: visible; NIR: near infrared; SWIR: short wave infrared; XS: multispectral; Pan: panchromatic.

- **Spatial heterogeneity indicators:** the use of less fertilizer and pesticide in organic management is supposed to result in higher spatial heterogeneity in organic fields due to soil fertility's spatial variability and crop disease spread. Diverse indicators of spatial heterogeneity of field biomass were computed, among which object based spatial heterogeneity using the properties of homogeneous spectral object obtained by segmentation.

- **Tractor traces:** conventional crops need numerous treatments for application of fertilizers and pesticides resulting in numerous tractor passages in the fields leaving visible traces, whereas they are difficult to identify in organic fields. Photointerpretation and object oriented automatic tractor traces recognition and characterisation were used to study the traces.

4. Results

In the limited context of this study (wheat and corn in eastern Germany), satellite images proved to be an efficient tool to discriminate organic from conventional crops with separability rates of 80% to 100%. The separability rate is defined as the global accuracy of the classification of all fields in organic and conventional classes.

In particular, regarding the three groups of indicators:

- **Spectral indicators** enabled complete separability for wheat and corn. In particular the Green NDVI, related to the plant nitrogen content, and the ratio of the two spectral bands 931 nm / 653 nm, related to the plant chlorophyll content, both computed from the hyperspectral sensor CHRIS-PROBA, gave 100 % separability.

- **Spatial heterogeneity indicators** enabled separability between 90% and 100% for wheat and corn. On very high spatial resolution images, object based spatial heterogeneity was however often more important in conventional fields where the impact of tractor traces is more pronounced.

- **Tractor traces** enabled complete separability for wheat fields while results were between 60% and 80% for corn.

5. Perspectives

These first results are encouraging but call for more research to refine and validate the method for other areas and climatic conditions in order to turn this concept into an operational service that is reliable and affordable for organic certification companies. More tests are actually being carried out.

Comparison of discrete multi-criteria decision making methods for selection of afforestation sites

René Estrella Maldonado* — **Ward Delabastita*** — **Anja Wijffels****
— **Jos Van Orshoven***

* *Department of Earth and Environmental Sciences, KU Leuven, Leuven, Belgium*

** *Universidad de Cuenca, Cuenca, Ecuador*

ABSTRACT. *This paper reports on a comparison of six Multi-Criteria Decision making Methods (MCDM) when used to rank sites for afforestation. ELECTRE III, PROMETHEE II, Analytic Hierarchy Process (AHP), Compromise Programming (CP), Stochastic Multi-criteria Acceptability Analysis (SMAA-2) and Iterative Ideal Point Thresholding (IIPT) were applied to a rasterized database containing a number of land units, with each land unit representing a group of spatially scattered pixels with identical land performance values. The 20 land units in the database were used as decision alternatives during the comparison. The decision criteria were five attributes expressing the biophysical and socio-economical performance of these land units 30 years after being covered by *Pinus patula*. The specific question addressed with these MCDM was: “Which land units should be afforested with *Pinus patula* to achieve an optimal land performance 30 years after planting the trees?” The output of each method was a ranking of the land units, ordered in accordance to their multi-criteria performance.*

A certain degree of consistency was observed in the rankings produced by all MCDM. Specifically, all methods designated the same four land unit as the most suitable alternatives, whereas only IIPT showed minimal differences regarding the least suitable land unit. Besides this, for all MCDM with the exception of IIPT, consistency was also observed in some intermediate positions in the ranking, although for other positions the results produced by the MCDM did not exactly coincide. This inconsistency can be attributed to inherent particularities of the methods and to the specific values assigned to the parameters used in each of them.

KEYWORDS: *multi-criteria decision methods, site location, afforestation*

2 SAGEO 2012.

1. Introduction

Due to unsustainable land use, land degradation has become a crucial issue in the mountainous Andes of southern Ecuador. Most pertinent aspects of land degradation are soil fertility loss and sedimentation of water reservoirs due to soil erosion. Also floods and drought occur more frequently due to loss of water storage capacity of the degraded land ([SOU 92], [VAN 03]). Afforestation of land sensitive to degradation has been proposed as a promising measure to improve overall land performance ([MOR 95], [WIJ 09], [WIJ 10]). It is clear however that decisions regarding afforestation (where and how ?) must take account of all important aspects of the land degradation problem, either ecological or socio-economical as perceived by all relevant stakeholders. In this decision context with multiple criteria, geographic information systems (GIS) and GIS-based decision support systems (DSS) are expected to be beneficial ([VAN 11]).

For the region of interest, i.e. the southern Andes of Ecuador, a prototype of spatial DSS (sDSS) for afforestation, called ForAndesT was recently developed ([WIJ 10]). This sDSS was designed to answer ‘Where ?’, ‘How ?’ and ‘How long ?’ questions related to afforestation planning. ForAndesT was a further development of its predecessor AFFOREST-sDSS ([GIL 05b]). Both sDSS are extensions of a GIS to which a multi-criteria decision making method (MCDM) encompassing an algorithm, previously called Interval Goal Programming (IGP) ([GIL 05b]), was added. We renamed IGP as Iterative Ideal Point Thresholding (IIPT) to avoid confusion with other Goal Programming methods, e.g. [CHA 55] and [CHA 72].

In addition to IIPT, in this article we explore, compare and discuss the performance of five other MCDM. These MCDM are ELECTRE III ([ROY 71], [FIG 05]), PROMETHEE II ([BRA 85]), Analytic Hierarchy Process (AHP) ([SAA 77]), Compromise Programming (CP) ([YU 73]) and Stochastic Multi-criteria Acceptability Analysis (SMAA-2) ([LAH 98], [LAH 01]). Among these methods, only SMAA-2 allows to deal directly with data uncertainty.

2. Materials and Methods

2.1. Study area

The study area is the hydrographical catchment of the river Tabacay, which is situated in the southern Andes of Ecuador. This catchment is located in the province of Cañar and its area is 66.52 km^2 . Its altitude ranges between 2400 and 3760 m asl. The climate, which is typical for the Andes-region, is characterised by relatively low temperatures although strong daily fluctuations are frequent. For the sake of illustration, a digital elevation model of the study area is visualised in Figure 1.

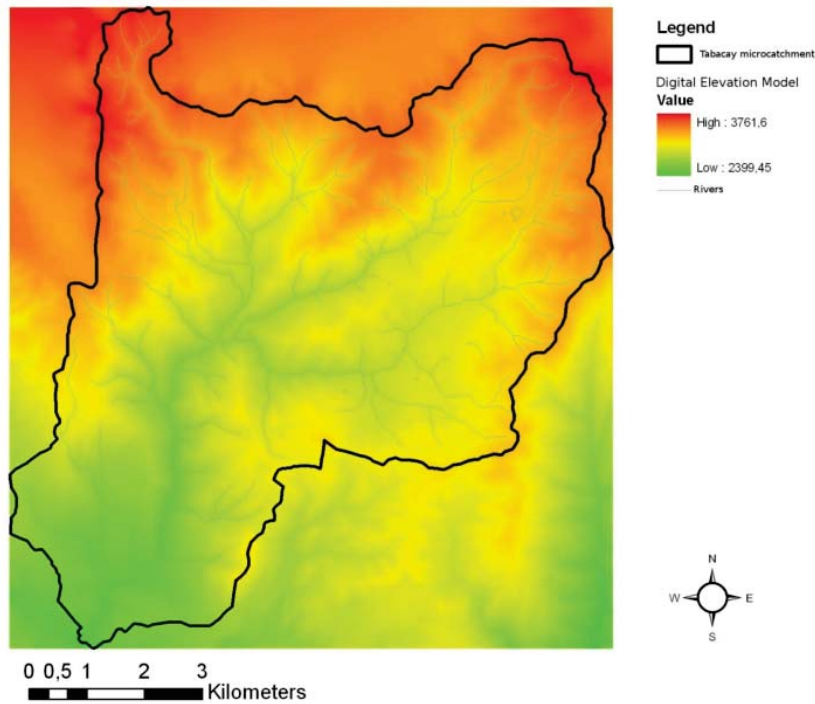


Figure 1. *Digital elevation model of the Tabacay microcatchment*

2.2. ForAndesT-sDSS

ForAndesT is a spatial DSS designed for providing support to forest planners and decision makers regarding questions like where, with which tree species, and for how long an afforestation project should be implemented ([WIJ 10]). The system comprises four modules: (i) a spatial database describing the territory of interest in terms of land units and holding all performance attribute values (to be used as decision criteria) for these land units, (ii) a MCDM module, and (iii) a user interface that includes (iv) tools for geospatial visualization and further analysis of results. The MCDM module basically consists of an implementation of the Iterative Ideal Point Thresholding (IIPT) method.

2.3. Database

The database used in ForAndesT contains the necessary data about the Tabacay catchment. This catchment is represented by a rasterized database with a spatial resolution of 30 m by 30 m. In this database pixels with the same performance attribute

4 SAGEO 2012.

Attribute	Short form	Unit	Optimisation
Water run-off at surface	Run-off	mm ha ⁻¹ yr ⁻¹	Minimisation
Sediment production	Sediment	Mg ha ⁻¹ yr ⁻¹	Minimisation
Organic carbon in soil	SOC	Mg ha ⁻¹	Maximisation
Organic carbon in biomass	BOC	Mg ha ⁻¹	Maximisation
Income generated	Income	\$ ha ⁻¹ yr ⁻¹	Maximisation

Table 1. Land performance attributes considered in ForAndesT-sDSS

values were grouped into land units. The values of these five performance attributes for each land unit were computed from the following land unit properties characteristics: altitude, slope gradient, slope curvature (convex, concave, straight), precipitation, soil type, lithology and land use type ([WIJ 09], [WIJ 10]). All performance attributes are continuous and were used as the criteria in the decision process. These performance attributes are listed in Table 1.

The performance attribute values corresponding to each land unit were computed off-line for the following time lapses: before afforestation, 10 years after afforestation, and 30 years after afforestation. The tree species considered for afforestation were *Pinus patula* and *Eucalyptus globulus*. This resulted in a total of six values for each combination of land unit and performance attribute, corresponding to the three time steps and two tree species mentioned above.

The stored performance attribute values are the basis for ForAndesT to address the following questions, which were considered relevant for supporting afforestation planning:

- Where to afforest ?, which refers to the land units on which a given afforestation project should be implemented;
- How to afforest ?, which refers to the tree species that should be used for afforesting a given land unit;
- How long to afforest ?, which refers to the period of time that a given type of forest should be kept in place.

The answers that ForAndesT provides for these questions are always targeted to achieve an optimal land performance, expressed in terms of one or more of the five performance attributes under consideration, cumulated over the considered time lapse. IIPT is used as the optimization module in ForAndesT.

2.4. Multi-criteria decision methods

The Tabacay database described above was used to study the performance of a number of different MCDM when answering the following question: “Where to afforest with *Pinus patula* to obtain an optimal cumulated land performance 30 years after establishing the forest ?” The decision alternatives were chosen to be the 20 land units

Land unit	Runoff mm ha ⁻¹ yr ⁻¹	Sediment Mg ha ⁻¹ yr ⁻¹	SOC Mg ha ⁻¹	BOC Mg ha ⁻¹	Income \$ ha ⁻¹ yr ⁻¹
1	19.50	1.59	153.64	232.66	2728.21
2	19.50	1.59	181.41	232.66	2728.21
3	30.00	1.59	153.64	232.66	2728.21
4	30.00	1.59	153.64	67.80	-129.79
5	30.00	1.59	181.41	232.66	2728.21
6	30.00	1.59	181.41	67.80	-129.79
7	30.00	2.77	153.64	163.30	1280.68
8	30.00	2.77	181.41	163.30	1280.68
9	45.24	1.59	89.63	163.30	1964.55
10	45.24	1.59	89.63	67.80	-858.56
11	45.24	1.59	89.63	163.30	678.88
12	45.24	1.59	89.63	57.02	-1951.72
13	45.24	1.59	153.64	67.80	-129.79
14	45.24	1.59	181.41	67.80	-129.79
15	45.24	1.59	329.41	67.80	-858.56
16	45.24	2.77	89.63	67.80	377.98
17	45.24	2.77	89.63	67.80	-129.79
18	45.24	2.77	153.64	163.30	1280.68
19	45.24	2.77	181.41	163.30	1280.68
20	45.24	2.77	329.41	67.80	377.98

Table 2. Performance attribute values for the 20 land units used as decision alternatives. Values correspond to 30 years after afforestation with pine

held by the ForAndesT database. Table 2 shows the performance attribute values for each of these 20 land units. The output of all MCDM is a ranking of these 20 land units from most to least suitable.

Since a multitude of MCDM suitable to address the stated question have been developed and described in literature, a preliminary exploration was conducted to select five methods, in addition to IIPT, for further study. This exploration was based on seven recent review articles concerning multi-criteria decision analysis. Three of these articles deal specifically with forestry related problems ([KAN 05], [Dia 08], [ANA 09]), while the other four refer to the application of MCDM in other domains ([ZOP 02], [POH 04], [WAN 09], [HO 10]).

Five of the methods to be discussed, namely ELECTRE III, PROMETHEE II, AHP, CP and SMAA-2, were selected on the basis of their frequency of appearance in the aforementioned articles. ELECTRE III, PROMETHEE II and AHP belong to the family of pairwise comparison methods, while CP is an example of an ideal point method, in which a decision alternative is considered better, or more suitable, if it is closer to the optimal, normally hypothetical alternative. The SMAA-2 method, as all other methods that belong to the SMAA family, is suitable for problems in which

6 SAGEO 2012.

Threshold	Explanation
Preference	$< (a_i - b_i)$ a is preferred over b $> (a_i - b_i)$ no preference
Indifference	$< a_i - b_i $ a is different from b $> a_i - b_i $ a and b are not different
Veto	$< (a_i - b_i)$ a is preferred over b, no matter the values of the other criteria $> (a_i - b_i)$ preference of a over b depends on values of the other criteria

Table 3. *Explanation of thresholds used in ELECTRE III*

Performance attribute	Preference	Indifference	Veto
Run-off	5	2	7
Sediment	0.3	0.1	0.5
SOC	15	5	20
BOC	15	5	20
Income	50	20	60

Table 4. *Threshold values used in ELECTRE III*

uncertainty is recognized in the data. The way in which each of these methods was applied is briefly described below.

2.4.1. ELECTRE III ([ROY 71], [FIG 05])

Before executing ELECTRE III, values for three parameters need to be defined for each performance attribute, namely the preference, the indifference and the veto thresholds. The meaning of these thresholds is explained in Table 3, in which a_i and b_i represent the values that performance attribute i takes for alternatives a and b .

Moreover a single value for an extra parameter, $s(\lambda)$, needs to be chosen to indicate the validity of comparisons that are made to build the resulting ranking.

The different values used for the preference, indifference and veto thresholds are listed in Table 4. They are expressed in the same units as their corresponding criteria. For $s(\lambda)$ a value of 0.15 was chosen in analogy to [RAY 03].

2.4.2. PROMETHEE II ([BRA 85])

PROMETHEE II converts differences between criteria values into a preference for one of the alternatives. To this end, a preference function has to be chosen among a group of six types: Usual, Quasi-Criterion, Linear, Level, Linear with indifference and Gaussian. These are all functions of the absolute value of the difference between criteria values for two given alternatives. In our case, a Gaussian preference function type was applied since it was considered that a function that gradually varies from

0 to 1 better represents preference variability in a real decision making context. For this type of function, one parameter value for each criterion had to be defined. This parameter, called sigma, allows to tune the width of the Gaussian curve corresponding to the preference function. Values for this parameter are expressed in the same units as the corresponding criteria. In this case the following values were used: 7.62 for runoff, 0.59 for sediment, 45.89 for SOC, 47.75 for BOC and 1096 for income.

2.4.3. AHP ([SAA 77])

The Analytic Hierarchy Process (AHP) formulates a decision problem as a three level hierarchy. The first level corresponds to the purpose of the problem, e.g. “select the land units in which a given tree species should be planted in order to optimize certain criteria”. The second level is composed of the criteria under consideration, for example runoff or sediment production. The third level comprises the decision alternatives.

Once the hierarchical structure of the problem has been sketched, the decision maker must state the corresponding preferences for the second and third levels of the hierarchy. This is carried out by means of a pairwise comparison in levels 2 and 3, using a scale to designate the relative importance of criteria and alternatives. This scale ranges from 1 (equally important criteria or alternative) to 9 (one criterion or alternative is extremely more important than the other). The relative importance values assigned during the pairwise comparison are used to ultimately compute a score for each alternative, which is in turn used to build the final ranking.

Since no expert knowledge was available for the application of AHP during this study, the pairwise comparison and assignment of relative importance values were automated using a MATLAB script. More specifically, a relative importance value ranging from 1 to 9 corresponding to the pair of alternatives (a_i, a_j) was computed with the Equation 1:

$$\left\lceil 8 \frac{a_{ik} - a_{jk}}{a_k^* - a_{*k}} \right\rceil \quad [1]$$

where:

- a_{ik}, a_{jk} are the values of alternatives a_i and a_j , respectively, for criterion k ;
- a_k^* is the optimal value for criterion k ;
- a_{*k} is the anti-optimal value for criterion k .

The fraction in Equation 1 is multiplied by 8, rounded to the next integer, and added to 1 in order to obtain a value between 1 and 9 as AHP requires.

2.4.4. Compromise Programming ([YU 73])

The first step that is performed when applying Compromise Programming (CP) is to compute the ideal point. The ideal point is a vector whose coordinates are given

8 SAGEO 2012.

Performance attribute	Ideal Value	Unit
Water runoff at surface	19.5	mm ha ⁻¹ yr ⁻¹
Sediment production	1.59	Mg ha ⁻¹ yr ⁻¹
Organic carbon in soil	329.406	Mg ha ⁻¹
Organic carbon in biomass	232.662	Mg ha ⁻¹
Income generated	2728.21	USD ha ⁻¹ yr ⁻¹

Table 5. Ideal point corresponding to the Tabacay catchment. Values correspond to 30 years after afforestation with pine

by the optimal values of the performance attributes, when every attribute is optimized independently of each other while taking into account the constraints of the problem. The ideal point is normally unfeasible, since multi-criteria decision problems involve conflicting objectives. The ideal point for this study is shown in Table 5.

The execution of CP results in what is known as the compromise solution. It corresponds to the feasible solution (vector of performance attribute values for a land unit) that is closest to the ideal point. The definition of ‘closeness’ requires the formulation of a distance function. This is a crucial step in CP, since the selection of distance function will certainly determine the resulting compromise solution. A general form of a distance function is expressed in the Equation 2:

$$L_p = \left[\sum_{i=1}^n w_i^p \left[\frac{f_i^* - f_j(x)}{f_j^* - f_{*j}} \right]^p \right]^{1/p} \quad [2]$$

where:

- n is the number of criteria under consideration;
- w_i is the relative importance (weight) assigned to performance attribute i ;
- p is a parameter that determines the type of distance function;
- f_i^* is the optimal value for environmental performance i ;
- $f_j(x)$ is the value of the i th performance attribute expressed as a function of the decision variables x ;
- f_{*j} is the anti-ideal corresponding to the i th attribute, that is the “worst” value for this attribute.

Note that the division by $f_j^* - f_{*j}$ is necessary because normally performance attributes are expressed in different units. Therefore, for aggregating them in a sum, this normalization procedure is used to obtain unit-less values in the range [0, 1].

As mentioned above, the parameter p defines the type of distance function. For example, when $p = 2$ the Equation 2 becomes the Euclidean distance function. In this study a value of $p = 1$ was considered convenient. When $p = 1$ is applied in Equation 2, it becomes Equation 3.

$$L_1 = \sum_{i=1}^n w_i \frac{f_i^* - f_j(x)}{f_j^* - f_{*j}} \quad [3]$$

It is important to note that the solution corresponding to L_1 amounts to maximizing the weighted sum of the achievement of all criteria. Note however that, in some cases, this solution might be quite unbalanced, that is some attributes can be much closer to their ideals than others. Note as well that the higher the value of p , the more balanced the resulting compromise solution will be. Nevertheless, a value of 1 was chosen in this study to maximize the overall achievement of the solution.

2.4.5. Stochastic Multi-criteria Acceptability Analysis ([LAH 98], [LAH 01])

Stochastic multi-objective acceptability analysis (SMAA) ([LAH 98]) is a multi-criteria decision support technique that is applied when multiple decision makers participate in the decision process. This technique is very suitable for problems with a high degree of uncertainty on the data, which can be present on the criterion values. In SMAA information about the preferences of the decision makers is not necessary at all; instead this technique assesses the weight values that would make each alternative the preferred one. Additionally, this method produces indicators regarding the support for an alternative to be chosen as well as about whether the accuracy of the input data is enough for making an informed decision.

SMAA-2 ([LAH 01]) computes two main measures in order to assign a rank to each alternative. The first measure is called rank acceptability and it is produced for each possible combination of alternative-rank. It expresses the probability that a given alternative is assigned the corresponding rank. The second measure provided by the SMAA-2 method is called the central weight vector. The central weight vector expresses, for each alternative, a combination of criteria weights that would make this alternative optimal.

2.4.6. Iterative Ideal Point Thresholding ([GIL 05b])

Finally, the Iterative Ideal Point Thresholding (IIPT) method was implemented and executed. Note that [GIL 05b] used the term Interval Goal Programming to refer to this method. As mentioned before, in this study we prefer to rename it as IIPT to avoid confusion with other well-known multi-criteria decision support methods introduced in [CHA 72]. The workflow of IIPT is outlined below.

Step 1: Besides the relative importance of each criterion expressed as weights, the user of IIPT is required to specify a maximum number of iterations. The value assigned to this parameter is used to compute the interval by which each coordinate of the threshold is relaxed, as explained below.

Step 2: The first threshold corresponds to what is called “ideal point” in Compromise Programming terminology, that is, each coordinate of the first threshold is given by the optimal value for each criterion when considered separately from the other cri-

10 SAGEO 2012.

teria. The database is then queried to determine whether there exist land units that satisfy this threshold.

Step 3: When no land units are found that satisfy the first threshold, each coordinate of the latter is relaxed according to Equation 4.

$$t_{ij} = t_i^* \pm j \frac{w_{max}}{w_i} \frac{|t_i^* - f_{*i}|}{n} \quad [4]$$

where:

- t_{ij} is the value that the coordinate of the threshold corresponding to criterion i takes in iteration j ;
- t_i^* is the coordinate of the first threshold (ideal point) corresponding to criterion i ;
- j is the current iteration;
- w_{max} is the maximum value among the weights assigned to all criteria;
- w_i is the weight assigned to criterion i ;
- f_{*i} is the anti-ideal value for criterion i ;
- n is the maximum number of iterations.

Note that, given the way in which thresholds are relaxed during the execution of IIPT, the whole range of values for the criterion with the maximum weight will be completely processed only after all iterations have been executed. Ranges of criteria with lower weights will be processed in fewer iterations. In other words, the higher the weight for a criterion, the smaller the “relaxation” of its corresponding threshold coordinate. This means that the “exploration” for criteria with higher weights is more fine-grained than for criteria with lower weights.

Once a new threshold has been produced by relaxing all coordinates of the previous threshold, the database is queried again to determine whether any alternatives satisfy the new threshold.

The procedure of relaxing and querying the database is repeated until the maximum number of iterations is reached. The rank assigned to each alternative depends on which iteration step it satisfied the threshold. This ensures that, after executing the maximum number of iterations, every alternative has been assigned a position in the final ranking.

Note that, when large intervals are used to relax the threshold in each iteration, the probability of more than one land unit being assigned the same ranking position is increased.

When setting the maximum number iterations a sensible compromise should be made between the detail level in which the “exploration” is conducted and, on the other hand, the time required for IIPT to produce results. This compromise also depends on

the criteria values. In our specific case, a maximum number of iterations of 20 was considered convenient.

All the MCDM mentioned above, with the exception of SMAA-2, were implemented using MATLAB. For executing SMAA-2 a software tool called JSMAA ([TER 12]) was used.

2.5. Criteria weights

The MCDM described above except SMAA-2 require that the decision makers define a weight for each decision criterion. These weights are values that indicate the relative importance of the different criteria ([WAN 09], [MAL 99]). To this end, several professionals and researchers in the area of land use planning who work in the study region were consulted by means of interviews. Specifically, they were asked to perform a pairwise comparison among the five criteria and assess their relative importance. This procedure resembles the steps carried out in AHP in the criteria level, as described in [GIB 07]. The resulting weights for each criteria were: 0.235 for runoff, 0.192 for sediment, 0.131 for SOC, 0.164 for BOC, and 0.278 for income.

2.6. Uncertainty on the database

In addition to comparing the performance of different MCDM, an additional goal of this study was to assess the behaviour of one of these methods, namely SMAA-2, under conditions of data uncertainty. For this purpose a partial inventory of performance attributes was conducted in the field. Based on the original and re-assessed values, interval limits for each attribute were computed. It is expected that the real value of a given performance attribute is somewhere between these interval limits. This amounts to say that the certainty regarding validity of data is increased when ranges, instead of deterministic values, are considered. SMAA-2 was then applied using the defined ranges as its input.

3. Results and Discussion

Table 6 shows the ranking of the 20 land units under consideration produced by each of the six studied MCDM. The first column lists the numeric, sequential identifiers assigned to the land units, as indicated in Table 2. The numbers in the remaining columns indicate the rank assigned by each MCDM to the corresponding land unit.

Note in Table 6 that the ranking produced by IIPT is very coarse. It distributes the 20 land units into only 6 different rank categories. In other words, the only positions of the ranking that were assigned to at least one alternative were 1, 2, 5, 7, 9 and 20. It is therefore evident that in some iteration steps several land units satisfied the current threshold, while in other iteration steps no alternatives were found to satisfy it. The

12 SAGEO 2012.

Land unit	Electre III	Promethee II	AHP	CP	SMAA-2	IGP
1	2	2	2	2	2	2
2	1	1	1	1	1	1
3	4	4	4	4	4	2
4	10	9	10	7	8	5
5	3	3	3	3	3	2
6	5	8	9	6	6	5
7	14	6	8	9	15	7
8	6	5	6	8	7	7
9	9	7	5	5	5	9
10	18	17	17	17	17	9
11	11	10	12	10	10	9
12	19	19	18	18	18	20
13	16	15	16	13	16	9
14	7	14	15	12	14	9
15	13	13	7	11	9	9
16	17	18	19	19	19	9
17	20	20	20	20	20	9
18	15	12	14	15	12	9
19	8	11	13	14	11	9
20	13	16	11	16	13	9

Table 6. *Rankings produced by the studied MCDM*

coarse output of IIPT greatly handicaps the comparison with the rankings produced by the other MCDM.

Several interesting patterns are observed when considering the results produced by all MCDM except IIPT. For instance, the same land units are designated as the 4 most suitable and the least suitable alternatives by all five methods. It can be said as well that, for the case of intermediate positions in the ranking, a certain degree of consistency can be observed. Note for instance the ranks assigned to land units 10, 11, 12 and 16.

This consistency coincides with findings reported in a related study ([GIL 05a]). [GIL 05a] describes a comparison among six MCDM, namely PROMETHEE II types 1, 5 and 6, ELECTRE III, AHP and IIPT (referred to as IGP in that publication) when selecting tree species among three alternatives (beech, spruce and pine) to be planted on a given set of land units. However, [GIL 05a] also reports different results regarding other topics on the application of the above MCDM. For instance, they point out that, in many cases, ELECTRE III produces what they term as “plural solutions”, i.e. solutions in which more than one alternative is recommended for a given land unit. This behaviour was not observed in our case, since our implementation of ELECTRE III produced no plural solutions, as shown in Table 6. On the other hand, [GIL 05a] do not report this behaviour for IIPT, while in our particular case it certainly produced

plural solutions. Nevertheless, the existence of plural solutions in the results of both ELECTRE III and IIPT can be considered a common situation given the characteristics of these methods. The discordances just described can be attributed to differences both in the underlying criteria data and in the tuning of the MCDM, i.e. the specific values assigned to their parameters.

Besides the apparent consistency observed in the rankings produced by five of the studied MCDM, it is also clear from Table 6 that, for some land units, there is no correspondence among the ranking position assigned by the different MCDM. These differences might be explained by particularities both in the procedures applied by each method, and in the set of parameters and their values that are used in each specific case. In general, given the different nature of each method, it is not realistic to expect that they produce identical rankings. Additionally, in the specific case of SMAA-2, the criteria data entered to the method was different from the data used by the other MCDM. As was explained above, interval limits for each criterion, instead of deterministic values, were used in SMAA-2. Intuitively, it can be claimed that this fact could have contributed to the inconsistencies between the ranking produced by SMAA-2 and the output of the other MCDM.

The coarse nature of the ranking produced by IIPT was presumably caused by three facts: the characteristics of the underlying criteria data, the particularities of the internal working of this method, and the parameter values used when this method was executed. It can be seen in Table 2 that the criteria values for all land units are distributed into few specific values. This fact causes that some land units show identical values in certain criteria, which in turn increases the similarity among land units. This similarity added to the chosen size of the relaxation intervals caused that, in some iterations, more than one land unit meet the relaxed threshold. This is the reason why, in most cases, the same ranking position is assigned to more than one alternative. Nonetheless, note that the most suitable alternative suggested by IIPT corresponds to the results of the other MCDM, as well as, at certain level, the least suitable land unit (least suitable alternative found by IIPT, i.e. land unit 12, was assigned positions 18 or 19 by the other methods). Using smaller intervals would have probably produced a slightly more “fine grained” ranking, although we can intuitively claim that the impact in this regard would have been minor, given the similarity among the performance attribute values.

Note the particular case of alternative 17, which is designated as the least suitable alternative by all methods except IIPT. IIPT assigns it position 8 out of 20 in the ranking, which can be somewhat misleading at first sight. Nevertheless, a more careful look at Table 6 allows to see that the 8th position is in fact the penultimate place in the coarse ranking produced by IIPT. As a matter of fact, IIPT assigns the 8th position to 11 alternatives. Therefore we can conclude that, for alternative 17, there is also a certain degree of consistency between IIPT and the other methods.

Note that ELECTRE III and PROMETHEE II are methods that require the specification of a relatively large number of parameters. In the case of ELECTRE III, the values for three parameters per criterion plus one global parameter must be set.

14 SAGEO 2012.

PROMETHEE II requires the user to provide values for one parameter per criterion. CP and IIPT require a single parameter to be defined. All these methods need that, in addition to the aforementioned parameters, one weight for each criterion is specified. On the other hand, SMAA-2 does not require any weights nor any specific parameters. Therefore, if we use the term complexity to refer to the number of parameters that a certain method or model needs, ELECTRE III is clearly the most complex method among the ones that were studied, while SMAA-2 is the least complex in this sense. Note that, although indirectly related, this definition of complexity does not refer to ease of use, which is a rather subjective concept.

4. Conclusions

In this article a comparison among six MCDM for locating afforestation sites has been presented and discussed. The different methods were applied with the aim of ranking a set of 20 land units based on four environmental and one economic criteria representing the cumulated performance of the land units 30 years after afforestation with *Pinus patula*.

In general, a major degree of consistency was found among the rankings produced by the six studied methods. This fact allows us to claim that the selected MCDM is not a decisive factor for the ranking, which gives the user a certain extent of freedom to select the method considering practical issues like, for example, ease of use. In the particular case of IIPT, although consistency is also observed for the most and least suitable land units, due to the criteria values, inherent characteristics and parameter settings of this method, the intermediate positions in the rankings are assigned to groups of alternatives, instead of producing a full ranking as was the case for the other MCDM. This behaviour handicapped the comparison between the output of IIPT and the results of the other methods.

Although uncertainty can be dealt with in ELECTRE III and PROMETHEE II, SMAA-2 is the MCDM that most naturally incorporates it both on the data and on the criteria weights. On the other hand, CP and AHP apparently leave small room for considering uncertainty.

The term complexity has been used here to refer to the number of parameters required by each MCDM. According to this definition, ELECTRE III is the most complex method, while SMAA-2 can be designated as the least complex MCDM, since it does not require the specification of any parameter at all. This notion of complexity may be a relevant factor when choosing the most suitable MCDM for solving a particular problem instance.

Since it can be claimed, intuitively, that the ranking produced by every method (except SMAA-2) depend on the values that its parameters take, it would be informative to undertake a sensitivity analysis with regard to the parameter values for each of these MCDM. This sensitivity analysis would consist in fixing the values for all except one parameters, while varying the value of the remaining parameter within a sensible pre-

defined range. Such a process would be repeated for every of the parameters involved in each method. This technique would allow to formally determine the real impact that tuning the parameters in one or another way has on the output of each method. It is clear as well that sensitivity analysis would ultimately result in more appropriate values for the required parameters.

5. References

- [ANA 09] ANANDA J., HERATH G., “A critical review of multi-criteria decision making methods with special reference to forest management and planning”, *Ecological Economics*, vol. 68, num. 10, 2009, p. 2535–2548.
- [BRA 85] BRANS J. P., VINCKE P., “A Preference Ranking Organisation Method: (The PROMETHEE Method for Multiple Criteria Decision-Making)”, *Management Science*, vol. 31, num. 6, 1985, p. 647–656.
- [CHA 55] CHARNES A., COOPER W. W., FERGUSON R. O., “Optimal Estimation of Executive Compensation by Linear Programming”, *Management Science*, vol. 1, num. 2, 1955, p. 138–151, ArticleType: research-article / Full publication date: Jan., 1955 / Copyright 1955 INFORMS.
- [CHA 72] CHARNES A., COLLOMB B., “Optimal economic stabilization policy: Linear goal-interval programming models”, *Socio-Economic Planning Sciences*, vol. 6, num. 4, 1972, p. 431–435.
- [Dia 08] DIAZ-BALTEIRO L., ROMERO C., “Making forestry decisions with multiple criteria: A review and an assessment”, *Forest Ecology and Management*, vol. 255, num. 8, 2008, p. 3222–3241.
- [FIG 05] FIGUEIRA J., GRECO S., EHRGOTT M., *Multiple criteria decision analysis: state of the art surveys*, Springer, 2005.
- [GIB 07] GIBNEY R., SHANG J., “Decision making in academia: A case of the dean selection process”, *Mathematical and Computer Modelling*, vol. 46, num. 7, 2007, p. 1030–1040.
- [GIL 05a] GILLIAMS S., RAYMAEKERS D., MUYS B., ORSHOVEN J. V., “Comparing multiple criteria decision methods to extend a geographical information system on afforestation”, *Computers and Electronics in Agriculture*, vol. 49, num. 1, 2005, p. 142–158.
- [GIL 05b] GILLIAMS S., VAN ORSHOVEN J., MUYS B., KROS H., HEIL G., VAN DEURSEN W., “AFFOREST sDSS: a metamodel based spatial decision support system for afforestation of agricultural land”, *New Forests*, vol. 30, num. 1, 2005, p. 33–53.
- [HO 10] HO W., XU X., DEY P. K., “Multi-criteria decision making approaches for supplier evaluation and selection: A literature review”, *European Journal of Operational Research*, vol. 202, num. 1, 2010, p. 16–24.
- [KAN 05] KANGAS J., KANGAS A., “Multiple criteria decision support in forest management: the approach, methods applied, and experiences gained”, *Forest Ecology and Management*, vol. 207, num. 1, 2005, p. 133–143.
- [LAH 98] LAHDELMA R., HOKKANEN J., SALMINEN P., “SMAA - Stochastic multiobjective acceptability analysis”, *European Journal of Operational Research*, vol. 106, num. 1, 1998, p. 137–143.

16 SAGEO 2012.

- [LAH 01] LAHDELMA R., SALMINEN P., “SMAA-2: Stochastic Multicriteria Acceptability Analysis for Group Decision Making”, *Operations Research*, vol. 49, num. 3, 2001, p. 444–454.
- [MAL 99] MALCZEWSKI J., *GIS and multicriteria decision analysis*, John Wiley and Sons, April 1999.
- [MOR 95] MORGAN R. P. C., “Soil erosion and conservation.”, , num. 2, 1995.
- [POH 04] POHEKAR S., RAMACHANDRAN M., “Application of multi-criteria decision making to sustainable energy planning: A review”, *Renewable and Sustainable Energy Reviews*, vol. 8, num. 4, 2004, p. 365–381.
- [RAY 03] RAYMAEKERS D., “Meervoudige doelplanning bij de bebossing van landbouwgronden: naar een ruimtelijk beslissingsondersteunend systeem”, Master’s thesis, Faculteit Landbouwkundige en Toegepaste Biologische Wetenschappen, KU Leuven, 2003.
- [ROY 71] ROY B., “Problems and methods with multiple objective functions”, *Mathematical Programming*, vol. 1, num. 1, 1971, p. 239–266.
- [SAA 77] SAATY T., “A scaling method for priorities in hierarchical structures”, *Journal of Mathematical Psychology*, vol. 15, num. 3, 1977, p. 234–281.
- [SOU 92] SOUTHGATE D., WHITAKER M., “Promoting Resource Degradation in Latin America: Tropical Deforestation, Soil Erosion, and Coastal Ecosystem Disturbance in Ecuador”, *Economic Development and Cultural Change*, vol. 40, num. 4, 1992, p. 787–807.
- [TER 12] TERVONEN T., “JSMAA: open source software for SMAA computations”, *International Journal of Systems Science*, , 2012, p. 1–13.
- [VAN 03] VANACKER V., GOVERS G., BARROS S., POESEN J., DECKERS J., “The effect of short-term socio-economic and demographic change on landuse dynamics and its corresponding geomorphic response with relation to water erosion in a tropical mountainous catchment, Ecuador”, *Landscape Ecology*, vol. 18, num. 1, 2003, p. 1–15.
- [VAN 11] VAN ORSHOVEN J. M. F., KINT V., WIJFFELS A., ESTRELLA MALDONADO R., BENCSEK G., VANEGAS P., MUYS B., CATTRYSSSE D., DONDEYNE S., “Upgrading Geographic Information Systems to Spatio-Temporal Decision Support Systems”, *Mathematical and Computational Forestry & Natural-Resource Sciences (MCFNS)*, vol. 3, num. 1, 2011, p. 36–41 (6).
- [WAN 09] WANG J., JING Y., ZHANG C., ZHAO J., “Review on multi-criteria decision analysis aid in sustainable energy decision-making”, *Renewable and Sustainable Energy Reviews*, vol. 13, num. 9, 2009, p. 2263–2278.
- [WIJ 09] WIJFFELS A., VAN ORSHOVEN J., “Contribution of afforestation to the enhancement of physical and socio-economic land performance in the Southern Andes of Ecuador: Assessment, modelling and planning support”, report , 2009, Vlaamse Interuniversitaire Raad (VLIR).
- [WIJ 10] WIJFFELS A., VAN ORSHOVEN J., MUYS B., CATTRYSSSE D., “*Decision support systems in agriculture, food and the environment: trends, applications and advances*”, chapter GI-technology: Adaptive generator of spatio-temporal decision support systems for land use planning, Idea Group Inc (IGI), 2010.
- [YU 73] YU P. L., “A Class of Solutions for Group Decision Problems”, *Management Science*, vol. 19, num. 8, 1973, p. 936–946.
- [ZOP 02] ZOPOUNIDIS C., DOUMPOS M., “Multicriteria classification and sorting methods: A literature review”, *European Journal of Operational Research*, vol. 138, num. 2, 2002.

Pertinence cognitive des représentations dans les cartes animées

Élaboration d'un protocole expérimental

Emmanuel Ahr* & Camille Guillemin**

Sous la direction de Nadine Mandran* — Claire Cunty**** — Elisabeth Chesneau*******

* ENS de Lyon

emmanuel.ahr@ens-lyon.fr

** Université Lumière Lyon 2

camille.guillemin@univ-lyon2.fr

*** UMR 5217 - Laboratoire LIG - Maison Jean Kuntzmann

110 av. de la Chimie - Domaine Universitaire de Saint-Martin-d'Hères - BP 53 - 38041 Grenoble cedex 9 – France

**** Université Lyon 2. Faculté GHHAT – UMR5600 EVS - IRG

5 avenue Pierre Mendès-France - 69676 Bron Cedex

***** "EVS" (Environment, City, Society) UMR CNRS 5600, CRENAM laboratory

6 rue Basse des Rives, 42023 Saint-Etienne Cedex 2, France

RÉSUMÉ. Le but de cette étude est de créer un protocole expérimental réutilisable en se basant sur une méthodologie issue des Sciences cognitives et appliquée aux cartes animées de géographie. Cela afin de déterminer laquelle de deux représentations d'un élément de la carte est la plus pertinente du point de vue de la perception et de la mémorisation. Cette approche devrait permettre d'améliorer les cartes animées à l'aide d'arguments scientifiques. Elle a ici été testée sur des cartes illustrant les crues du Rhône et de ses affluents sur une période de 6 ans.

ABSTRACT. This experiment aimed at proposing a reusable protocol based on cognitive science methodologies. It was here applied to geographic animated maps in order to determine which one of two representations of a single element of the map was the most relevant as far as perception and memorization are concerned. More specifically, it was tested on maps illustrating rises of water of the Rhône and its tributaries over a six-year period. This study shows that the use of a scientific/cognitive approach can participate in improving animated maps.

MOTS-CLÉS: carte animée, perception, mémorisation, oculomètre, temps, couleurs, crues.

KEYWORDS: animated map, perception, memorization, time, colours, rise of water level.

1. Introduction

Cette étude s'inscrit dans le cadre du projet PerçuRisk regroupant des chercheurs et ingénieurs de l'UMR5600 EVS-IRG de Lyon-Saint-Etienne et du LIG de Grenoble. Il est financé par un cluster « Risques et Environnement » de la région Rhône-Alpes, portant sur la pertinence de l'animation pour les représentations temporelles dans les cartes liées aux risques naturels. La carte animée apparaît comme un moyen intéressant de communiquer des informations spatio-temporelles, mais qu'en est-il de son efficacité d'un point de vue cognitif (Valorge, 2010 ; Harrower *et al.*, 2008) ? Face à la brièveté des images qui défilent, que perçoit l'œil ? Face à l'accumulation des informations qui s'enchaînent, que mémorise-t-on ? Le premier objectif de cette étude est d'élaborer un protocole expérimental issu des sciences cognitives et des sciences humaines permettant de savoir laquelle de deux représentations est la mieux perçue et mémorisée par des utilisateurs non-experts de cartes animées. Le second objectif est de mettre ce protocole à l'épreuve en comparant effectivement différentes représentations simples de cartes animées. Pour ce faire, quatre cartes animées représentant les mêmes crues du Rhône et de ses affluents entre novembre 2002 et juin 2008 ont été conçues par des étudiants en cartomatique (Brunner *et al.*, 2012). Une carte utilise un dégradé de bleu pour marquer l'intensité d'une crue, une autre utilise une gamme polychromatique allant du vert au rouge. Nous avons fait l'hypothèse que la gamme polychromatique, par le biais d'une meilleure catégorisation sémantique, permettait une meilleure perception (Winawer *et al.*, 2007) et mémorisation (Imai *et al.*, 2011) de l'intensité des crues à un moment donné. Ces deux cartes ont été associées ou non à une bande sonore annonçant le mois et l'année à chaque changement de mois. Les quatre versions de la carte s'accompagnent d'une frise et d'une date (présentée sous la forme jj/mm/aaaa). Notre hypothèse est que la bande sonore favorisera la mémorisation des dates de crues, suivant la théorie de l'encodage multimodal (par l'intermédiaire de plusieurs sens, ici la vue et l'ouïe, Allport *et al.*, 1972).

2. Méthode

Un protocole expérimental spécifique a été conçu. La méthodologie respectée ici, classique aux sciences cognitives, ne consiste pas à étudier l'influence de tous les paramètres en jeu. Elle préconise une approche réductrice, mais quantitative et rigoureuse. Premièrement, des fonctions cognitives à tester sont choisies. De ce choix découle la sélection d'outils ou de tests expérimentaux appropriés, eux-mêmes associés à une mesure qui doit être définie. La deuxième étape consiste à choisir quel va être le paramètre étudié. L'expérience est alors reproduite deux fois à

l'identique, mis à part ce paramètre qui va varier entre les deux expériences. Par conséquent, si la mesure réalisée diffère entre les deux sessions expérimentales, cette différence est due au paramètre que l'expérimentateur a modifié. Un grand nombre de mesures doit être établi pour obtenir un résultat statistiquement significatif, c'est-à-dire qu'il y a moins de 5% de chance que ce résultat soit dû à une variation aléatoire. Si deux paramètres sont étudiés lors d'une même expérience, les tests statistiques choisis prennent alors en compte une possible interaction entre eux. Ainsi, la présente expérience a nécessité la participation de 44 bénévoles à raison d'une unique session d'une heure chacun, soit 11 sujets par version de la carte, suivant un protocole prédéfini. En premier lieu, les mouvements oculaires du sujet sont enregistrés pendant le visionnage de la carte à l'aide d'un oculomètre. Le pourcentage de temps de regard sur les différents éléments de la carte est calculé pour chaque modalité expérimentale. Ces données permettent de détecter des éléments de l'animation qui n'auraient pas été perçus ainsi que ceux qui ont été les plus attirants pour l'œil novice. À la suite du visionnage de la carte, un premier entretien permet de recueillir les remarques et impressions spontanées des sujets. Ne pas négliger de telles données qualitatives permet de faciliter l'interprétation des résultats quantitatifs. Cette étape sert également de distraction pour le sujet afin que les tests de mémorisation qui suivent mesurent la mémoire à long terme et non la mémoire à court terme (<1 minute). Suite à l'entretien, les sujets réalisent deux tâches de rappel indicé, une première appelée « Dangerosité » pour tester leur mémorisation de l'intensité des crues et une seconde nommée « Temps » pour tester leur mémorisation de la temporalité, soit des dates des crues présentées. Les tâches de rappel indicé, c'est-à-dire pour lesquelles un indice contextuel aide au rappel, sont des tests classiques utilisés par les cognitivistes pour mesurer la mémorisation. Pour le test Dangerosité, quinze images extraites de l'animation sont présentées au sujet. Pour chaque extrait, un tronçon de cours d'eau a été supprimé. Le sujet doit spécifier l'intensité de la crue du tronçon supprimé, à ce moment de l'animation. Pour le test Temps, on présente au sujet quinze courts extraits vidéo de la carte animée desquels la frise, la date, et le cas échéant la bande sonore sont retirées. On demande alors au sujet de rappeler le tiers-temps de l'animation auquel appartient l'extrait, puis de renseigner l'année, et enfin le mois auxquels l'évènement présenté a eu lieu. Un second entretien à visée exploratoire clôt l'heure de passation. On demande notamment au sujet quelle représentation aurait pu l'aider à mieux percevoir ou mémoriser l'intensité et le moment des crues.

3. Résultats

L'hypothèse selon laquelle la gamme polychromatique permet une meilleure mémorisation de l'intensité des crues par rapport au dégradé de bleu est vérifiée par le test Dangerosité. Une analyse de variance (ANOVA) effectuée sur les résultats issus des 44 sujets révèle un effet principal de la représentation de l'intensité de la crue. L'analyse des temps de regard par zone d'intérêt montre que cette différence de mémorisation n'est pas due à une différence de temps de regard. Notre seconde

hypothèse, selon laquelle la bande sonore aiderait à mémoriser le temps, n'est en revanche pas vérifiée par les tests statistiques. Au contraire, il apparaît que les sujets sont meilleurs pour mentionner l'année des crues lors du test Temps s'ils ont vu les cartes sans bande sonore. Les données oculométriques nous renseignent sur l'origine de ce résultat. Les sujets qui avaient la bande sonore regardent significativement moins la date et la frise temporelle que les sujets sans bande sonore. Lors des entretiens, les sujets qui ont bénéficié du son reconnaissent effectivement qu'ils n'ont pas prêté attention à la frise ni à la date, puisqu'une voix la leur dictait. En revanche, ils cessaient rapidement de faire attention à cette voix car elle n'était pas directement reliée aux événements de crue.

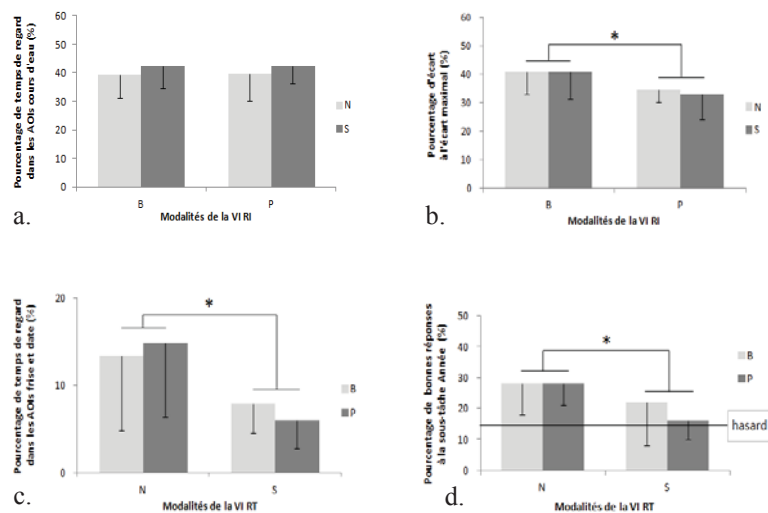


Figure 1. Résultats. a. Le type de carte ne modifie pas le temps de regard sur les fleuves ; b. La gamme polychromatique induit moins d'erreurs de rappel de l'intensité des crues ; c. La date et la frise sont moins regardées si une bande sonore annonce les changements de mois ; d. La date des crues est moins bien retenue si une bande sonore annonce les changements de mois. B : représentation de l'intensité des crues selon un dégradé monochromatique de Bleu ; P : représentation de l'intensité des crues selon une gamme Polychromatique allant du vert au rouge ; S : présence d'une bande Sonore annonçant le mois et l'année des changements de mois ; N : absence de bande sonore ; VI : Variable Indépendante, le paramètre étudié ; RT : Représentation du Temps ; RI : Représentation de l'Intensité des crues ; AOI : Area Of Interest, zones d'études ; * : différence significative à l'ANOVA ($p < 0.05$).

4. Conclusion

Cette étude démontre que les sciences cognitives sont aujourd'hui suffisamment mûres pour servir d'outil à l'amélioration scientifique des cartes de géographie. L'approche décrite ici nécessite seulement une vingtaine de sujets par paramètre étudié. Elle donne une réponse quantitative aux questions de la perception et de la mémorisation des informations contenues dans la carte animée, mais elle ne néglige pas pour autant l'aspect qualitatif et humain du visionnage, grâce aux deux entretiens effectués. Rigoureux et complet, ce protocole est adaptable à l'étude d'autres cartes et d'autres paramètres.

Remerciements

Nous tenons à remercier pour leur aide, quelle que soit la forme que celle-ci ait prise Rémy Versace, professeur à l'EA 3082, EMC, Lyon, France, toute l'équipe PerçuRisk ainsi que la région Rhône-Alpes pour avoir financé le projet par un cluster « Risques et environnement », Eric Liegho, Gwendolyn Brunner, Julien Chopy et Carole-Anne Mazur, pour la conception des cartes animées, enfin, tous les sujets qui sont venus passer bénévolement une heure d'expérience au laboratoire IRG de l'Université Lumière Lyon 2.

5. Bibliographie

- Allport D.A., Antonis B., Reynolds P., « On the division of attention : A disproof of the single channel hypothesis » *Quarterly Journal of Experimental Psychology*, volume 24, issue 2, 1972.
- Brunner G., Chopy C., Lieghio E., Mazur C-A., Réalisation de maquette animées. Représentation des inondations du Rhône et de ses affluents au cours du temps, rapport de stage, février 2012, Master STEP OSIG, Université de Saint Etienne.
- Harrower M., Fabrikant S., The role of map animation for geographic visualization in Dodge M., McDerby M., Turner M., *Geographic Visualization*, Wiley, 2008.
- Imai *et al.*, « Hybrid coding of color : How can we unify color discrimination, categorization and memory ? » *Journal of Vision*, septembre 2011.
- Valorge A., Perception visuelle de cartes animées dédiées aux risques naturels et technologiques, mémoire, juin 2010, Master TPE SIG, Université de Saint Etienne.
- Winawer *et al.*, « Russian blues reveal effect of language on color discrimination » *Proceeding of the National Academy of Sciences*, avril 2007.

Vers une méthode de conception de SIG 3D collaboratif

Rafika Hajji* — Roland Billen**

** Filière de Formation en Géomatique et Ingénierie Topographique*

Institut Agronomique et Vétérinaire Hassan II

BP 6202-Rabat Instituts Rabat, Maroc

r.hajji@iav.ac.ma

*** Unité de Géomatique, Université de Liège*

17 Allée du six Août (Bât. B5a) 4000 Liège, Belgique

r.billen@ulg.ac.be

RÉSUMÉ. Dans le contexte d'une nouvelle géomatique axée sur le partage et l'instauration d'une logique de travail participatif et collaboratif, les professionnels doivent repenser les méthodologies, les outils et les usages autour de l'information géographique. Certes, les formes partenariales autour de l'information géographique voient le jour mais leurs apports sont trop réduits et ponctuels. Les SIG 3D collaboratifs par leur prise en charge de la 3D et leur mise en œuvre autour de solutions conceptuelles et techniques collaboratives nous paraissent une solution évoluée et prometteuse pour partager les coûts de production et de mise à jour ainsi que pour répondre aux nouvelles exigences en interopérabilité. Leur conception méthodologique fait objet de plusieurs questions organisationnelles et techniques. L'objectif de ce papier consiste à mettre en perspective les potentialités d'un SIG 3D collaboratif en présentant les premières prémisses d'une réflexion conceptuelle sur leur mise en place.

ABSTRACT. In the context of a sharing-based geomatics and growth of participatory and collaborative approaches, professionals must rethink geographic information methodologies, tools and practices. Although partnership forms around geographic information are emerging, their contributions are, up to now, rather small and isolated. Collaborative 3D GIS by their support for 3D and their implementation based on collaborative design and techniques seem promising; they represent an advanced solution for sharing production and update costs and for meeting new interoperability requirements. Their design methodology is subject to several organizational and technical issues. The objective of this paper is to put into perspective the potential of a collaborative 3D GIS by presenting first premise of a conceptual reflection on their implementation.

MOTS-CLÉS: SIG 3D collaboratif, collaboration, coproduction, conception, reengineering.

KEYWORDS: Collaborative 3D GIS, collaboration, co-production, design, reengineering

1. Introduction

Le territoire constitue un terrain d'actions pouvant supporter plusieurs recompositions thématiques qui ont donné lieu à plusieurs schémas conceptuels du monde réel. L'implémentation et l'utilisation de ces schémas étaient réduites à un cadre organisationnel précis pour répondre à des besoins ponctuels de gestion. Ces solutions individuelles ont soulevé en plus des problèmes de discordances d'ordre technique, d'autres problèmes de divergences sémantiques. Avec les avancées technologiques et l'évolution du Géoweb, les SIG se sont démocratisés pour toucher une grande communauté d'utilisateurs avec des besoins différents. Les coûts de production et de mise à jour qui dépassent les budgets, la mutualisation des données et la gestion de leur qualité constituent les principales problématiques qui ont dynamisé les débats autour des données. La modélisation des données semble élargir le fossé entre les différentes composantes sociales. En effet, les chercheurs pensent déjà aux modèles de données 5D intégrés représentant la géométrie, le temps et l'échelle (Oosterom et al, 2010), les utilisateurs expriment un véritable engouement envers la 3D tandis que les producteurs potentiels des données de base sont handicapés par les limites budgétaires d'une part et les difficultés autour de la mutualisation des données, des traitements et des connaissances d'autre part. Les SIG 3D collaboratifs ne seraient-ils pas une solution potentielle à ces grands défis ?

C'est précisément sur cette problématique que se focalise le présent papier qui fait un tour d'horizon sur les problématiques actuelles autour des données géographiques et des SIG en mettant en valeur les solutions collaboratives et en présentant leur potentiel dans leur résolution. L'objectif étant d'expliquer et de démontrer l'utilité d'une solution SIG 3D collaborative, les enjeux et les contraintes de son adoption ainsi qu'une réflexion sur sa mise en œuvre conceptuelle et méthodologique. Pour ce faire, nous présentons préalablement les concepts des SIG organisationnels, des données et de la collaboration en focalisant sur leurs aspects sémantiques et en soulevant les grandes questions qui s'y rattachent. Nous faisons ensuite une mise au point sur le concept du SIG 3D collaboratif dans l'objectif de montrer son utilité, ses contraintes et ses enjeux. Enfin, nous présentons nos premières réflexions et propositions autour d'une méthode conceptuelle de sa mise en œuvre.

2. Le SIG 3D collaboratif : Un système multidimensionnel

2.1. *Le SIG : Un concept multi-facettes*

Bien que les SIG aient connu une longue période de gestation et de mûrissement au niveau de leur construction technique et conceptuelle, leur concept fait toujours objet de mutations. Les définitions du SIG convergent, implicitement ou explicitement, vers son caractère multidimensionnel qui est justifié par la nature même de l'information géographique qu'il gère. En effet, la conception d'un SIG peut être fondée sur trois types de constructions : une construction technique où le SIG est vu comme boîte à outils, comme environnement informatique et comme base de données ; une construction organisationnelle qui considère le SIG comme système d'une organisation et une construction sociale et politique qui dépasse la sphère organisationnelle (Joliveau, 2004).

Les premières préoccupations autour des SIG étaient d'ordre technique centrées sur l'automatisation des productions cartographiques et la performance dans la gestion et l'intégration des données pour répondre à un besoin organisationnel (Pornon, 2007). Les grandes innovations technologiques et la révolution des pratiques sociales qui ont bouleversé tout le paysage de la géomatique tendent à diluer l'aspect technique des SIG et à revaloriser leur dimension sociale. Dans ce contexte, le SIG est vu comme une superposition de couches reliées entre elles par des protocoles dans un contexte de coordination et de communication qui recentre les SIG sur leur contribution à la compréhension du territoire et à l'amélioration de la transversalité dans les organisations. Plusieurs auteurs constituent des jalons de ce point de vue (Joliveau, 2004 ; Pornon, 2007). Aujourd'hui, Le SIG est devenu un réseau complexe et dynamique qui transforme et se transforme par les révolutions technologiques, sociales et politiques et son avenir en est fortement tributaire.

2.2. *La donnée géographique : Un concept non encore maîtrisé*

Nous sommes aujourd'hui confrontés à une grande vague de concepts et néologismes qui ont changé les pratiques traditionnelles autour de l'information géographique. Les communautés, les usages et les outils ont connu un changement radical et de nouvelles pratiques se sont installées auprès des utilisateurs et se sont transformées en exigences sociales, notamment, l'accès et l'appropriation de l'information géographique (Joliveau, 2004). Un mouvement qui semble faire une rupture technologique et culturelle par rapport aux méthodes traditionnelles de penser et gérer l'information géographique (Turner, 2006). Les nouvelles pratiques du Géoweb ont fait un renversement de paradigme dans la conception des données spatiales et une rupture avec la logique descendante selon laquelle l'information

émanait seulement de grands producteurs d'informations institutionnels et privés (Mericskay et al, 2010). D'un autre côté, un véritable engouement pour la donnée 3D impose aux organismes détenteurs des données de base de répondre à cette nouvelle demande par la construction d'un fond de plan 3D.

La normalisation des données géographiques a franchi un grand pas mais elle est loin de maîtriser toutes les situations d'échange. Les données sont souvent ancrées dans une pratique métier pour répondre à un usage prédéfini et sont encapsulées dans un schéma conceptuel du monde réel. Ainsi décrites, ces données posent des problèmes de partage, de maîtrise des formats d'échange, des discordances et un appauvrissement sémantique. S'il est plus facile de standardiser les données de référence (cadastre, données topographiques, etc.), la standardisation des données thématiques n'est pas triviale (Pornon, 2007). La qualité des données est toujours un débat sur la table. Les nouvelles contraintes liées à l'échange et la mutualisation des données et la formalisation des partenariats inter organisations lui posent des défis supplémentaires. Le grand flux de données disponibles sur le web engage aussi de grandes questions sur la qualité.

Aujourd'hui, la modélisation 3D urbaine a franchi un grand pas avec l'adoption du standard CityGML par l'Open GIS Consortium (l'OGC) pour la normalisation de l'échange des modèles 3D urbains. Le CityGML définit un modèle géométrique, topologique et sémantique de l'espace selon différents modules thématiques dont le bâtiment. Il adopte la notion de niveau de détail ou Level of Detail (Lod) pour représenter la granularité de l'information 3D (voir Gröger et al, 2008). Depuis sa validation par l'OGC, le CityGML a apporté une solution à la standardisation et l'échange de villes numériques et a ouvert des perspectives à de nouveaux traitements de l'information 3D. Cependant, la question autour de la modélisation et l'analyse spatiale 3D n'est pas encore maîtrisée malgré le nombre de recherches dans ce domaine. Les systèmes disponibles sont toujours qualifiés de 2.5D et une véritable solution 3D n'est pas encore atteinte (Goodchild, 2010).

2.3. La collaboration : un processus complexe, dynamique et prometteur

La communauté scientifique a largement balisé le champ de la collaboration. Le concept a fait objet de plusieurs recompositions qui, malgré leurs diversités, gravitent autour de principes communs. Selon (Roche, 2009), « On peut parler de la collaboration lorsqu'un groupe de personnes se réunit, poussé par un intérêt commun, pour explorer de façon constructive de nouvelles possibilités et créer quelque chose qu'il n'aurait pas pu créer seul ». Serge K. Levan (cité par Roche) précise que la collaboration est vue comme un processus cyclique articulé autour de quatre phases: la co-analyse; la co-définition; la co-réalisation et la co-évaluation. Cette définition montre que l'action collaborative est présente dans toutes les phases autour de la réalisation d'un objectif commun, depuis la conception et l'analyse jusqu'à l'évaluation. La collaboration est vue comme une forme évoluée des autres

actions partenariales comme la coopération qui suppose que chacun pourrait seul réaliser la tâche et la coordination où on cherche à atteindre un objectif global en travaillant de manière indépendante. Plusieurs auteurs comme (McDougall et al, 2005; Dedekorkut, 2003; Pornon, 1998) ont bien cerné ces différences.

En géomatique, les enjeux de la collaboration sont multiples: décloisonnement des activités, une volonté de partager les coûts d'acquisition des référentiels géographiques, le catalogage pour le partage de métadonnées; l'augmentation de la fréquence des mises à jour et enfin la coproduction qui nécessite le plus fort degré d'action mais qui représente le principal carburant qui donne un potentiel collaboratif aux SIG (Noucher, 2009).

Si l'échange et la diffusion des données se heurtent toujours à des problèmes d'interprétations sémantiques, la coproduction permettra d'anticiper ces problèmes en tentant de confronter et d'homogénéiser des visions hétérogènes sur les données. Les données coproduites circuleront plus facilement dans les processus de gestion inter-organisationnels. C'est une façon de régler les discordances et les différences avant que la donnée ne soit mise sur le chemin de la diffusion. Aussi, la coproduction apporte des solutions potentielles pour l'appropriation collective de données 3D dont les coûts exorbitants d'acquisition et de mise à jour constituent le grand frein à l'engagement des organisations dans l'adoption de solutions 3D. Les données 3D coproduites doivent être définies sur base de spécifications géométriques et sémantiques communes assurant leur interopérabilité. Ainsi, un bâtiment représenté selon le modèle géométrique et sémantique du CityGML peut être facilement intégré dans plusieurs modèles urbains destinés à des usages différents.

Paradoxalement, les nouvelles perspectives ouvertes par le potentiel des données collaboratives soulèvent en contrepartie de nouveaux problèmes de qualité et de responsabilité. Ces problèmes ne sont pas que techniques, mais également déontologiques et juridiques (Mang, 2010). La collaboration pose des questions d'ordre juridique et institutionnel et nécessite des conventions pour partager les coûts et les responsabilités.

Si la collaboration apporte des solutions aux grandes préoccupations autour des données, elle engendre aussi des contraintes organisationnelles qu'il faut savoir maîtriser. S'intégrer dans une action collaborative est loin d'être trivial et constitue un grand défi pour l'organisation. Les objectifs sont rarement convergents et le processus est long et risque de porter atteinte à la chaîne de production. Les données coproduites pourront-elles être adoptées, dans le futur, comme données de base pour satisfaire les objectifs organisationnels ou l'organisation devra résoudre la question technique autour de la cohabitation de ses propres données avec celles coproduites?

La collaboration est loin d'être une décision ponctuelle statique mais elle représente un processus dynamique et évolutif qui doit être mené d'une manière pragmatique et bien réfléchie pour minimiser, sans prétendre éliminer, les risques

d'échec. Une clé de succès serait de l'entreprendre selon un cheminement parallèle qui sauvegarde les intérêts et minimise les impacts organisationnels.

La collaboration engendre inévitablement des coûts supplémentaires pour l'organisation afin de réaliser le projet commun : le coût de reengineering, celui de la formation et de la mise à niveau de toutes les organisations partenaires pour leur contribution dans le projet et l'installation du nouveau système. Cependant, les bénéfices doivent être évalués à long terme vu la nature du projet. L'évaluation ou plutôt la co-évaluation peut être quantitative ou qualitative et doit couvrir plusieurs aspects. Les critères en sont multiples : évaluation par rapport aux objectifs communs, par rapport aux systèmes existants ou par rapport aux attentes organisationnelles, etc. Les produits doivent être évalués en termes de qualité tandis que l'évaluation du processus porte sur son efficacité pour supporter le développement de projets collaboratifs.

3. Le SIG 3D collaboratif : Une forme de collaboration à contraintes et enjeux multiples

3.1. Limite des SIG intra-organisationnels

Les SIG et leurs applications métier se sont pour longtemps implantés dans les organisations dans un contexte figé pour être greffés sur les procédures et dispositifs organisationnels existants et pour résoudre des problèmes de l'organisation étroitement liés à son processus de production. Avec le besoin accentué d'échange des données qui se pose à toute organisation, le SIG intra organisationnel trouve sa limite. Un SIG conçu par et pour l'organisation n'a plus de place dans la nouvelle géomatique centrée sur le partage et la collaboration. Les SIG sont appelés à dépasser les limites intra organisationnelles pour répondre aux exigences de la collaboration et permettre à l'organisation de s'ouvrir et de s'épanouir au sein de formes collaboratives. Leur adaptation aux contraintes de collaboration nécessite de revoir les méthodes de leur conception.

3.2. Vers des SIG 3D collaboratifs

L'utilité de la 3D n'est plus à démontrer devant le nombre d'applications qui pourraient bénéficier d'une modélisation 3D de l'espace comme la planification et la gestion urbaine, le cadastre, la protection civile, l'architecture, l'archéologie, les sciences environnementales, etc. La représentation des objets par deux modèles séparés : les SIG 2D pour la gestion de la planimétrie et le modèle numérique de terrain pour la représentation du relief entraîne une redondance des données et un

manque de consistance (Abdul Rahman et al, 2008). Souvent, des transformations sont nécessaires pour convertir l'information dérivée d'un MNT (pente, hauteur, etc.) selon une couche SIG pouvant être superposée avec d'autres couches thématiques pour effectuer des analyses spatiales. L'intégration de l'information spatiale dans un modèle unique permet de faire l'analyse spatiale et la mise à jour de manière efficace sans avoir recours à des transformations ou conversions de formats. Par exemple, une requête du type : quelles sont les parcelles qui risquent de subir une inondation d'un mètre de haut? pourrait être résolue à travers un seul modèle intégré pouvant être implémenté dans un SIG 3D (Abdul Rahman et al, 2008).

Les SIG 3D sont vus comme des solutions étendant les fonctionnalités d'un SIG 2D ou 2.5D pour prendre en charge les structures volumiques et répondre à des requêtes élaborées dans un environnement 3D. Il est nécessaire pour un SIG 3D d'être en mesure d'afficher, de manipuler et d'interroger un modèle 2D et d'obtenir à nouveau une modélisation 3D dès que cela est nécessaire. Les requêtes peuvent être exécutées dans un environnement 2D tout en gardant l'exactitude de l'information 3D (De Cambray, 1993). A titre d'exemple, dans le cas de structures multi-niveaux, une requête d'intersection entre un pont et une rivière doit tenir compte de l'information 3D et renvoyer un résultat nul même si elle est exécutée dans un environnement 2D (De Cambray, 1993). Nous pouvons ainsi dire qu'un SIG 3D est une solution évoluée pouvant prendre en charge les fonctions d'un SIG 2D et aussi les étendre aux formes 3D, ce qui pourrait justifier l'investissement dans une telle solution.

L'intérêt porté aux SIG et à la collaboration par la communauté scientifique a donné naissance à plusieurs définitions et visions parfois hétérogènes autour de ces deux concepts. En contrepartie, le concept de SIG collaboratif, même si ayant émergé il y a longtemps dans la littérature, n'a pas fait l'objet d'une définition précise qui cerne le concept dans un contexte bien défini. Nous avons soulevé une grande divergence sémantique autour de l'usage du terme. En effet, le SIG collaboratif s'est vu attribué deux visions différentes : la première vision est trop réductrice et considère le SIG comme outil de collaboration dans le sens de fournir une infrastructure technique qui aide les organisations à travailler ensemble pour une gestion participative du territoire, et une deuxième vision plus ouverte qui tente de le considérer comme forme ou produit de mise en œuvre d'une action de collaboration. Cette diversité de visions a été aussi soulevée par (Orban-Ferauge, 2011) autour des SIG participatifs dans le sens où les chercheurs se classent selon ce qu'ils placent dans leurs priorités autour du SIG participatif : l'approche pratique du SIG au service de la participation (sigP), le développement de l'outil à des fins de participation (SIGp) ou dans la mise en œuvre de cet outil (SIGP).

Notre propre vision s'appuie sur l'idée de mettre la collaboration au service des SIG. En effet, les fonctions classiques d'un SIG ne peuvent pas être extrapolées à tout contexte d'échange, c'est tout le processus de sa mise en place conceptuelle et méthodologique qui doit être revu dans un contexte collaboratif. La collaboration ne doit pas être vue comme une finalité mais comme un processus offrant une solution aux principaux problèmes et défis qui s'imposent pour les SIG intra-organisationnels existants. Par SIG collaboratif, nous nous entendons une solution collaborative de conception et de mise en place d'un SIG qui fait appel à la collaboration dans toute la démarche méthodologique de sa conception à sa mise en œuvre. Dans cette situation, la collaboration est sollicitée pour confronter des visions différentes et construire une vision commune qui profitera aux différentes organisations adhérentes au projet par le partage des coûts d'acquisition et de mise à jour, des ressources et du savoir-faire. La conception collaborative apporte un potentiel majeur dans le sens où elle permettra de maîtriser toutes les divergences autour de l'information géographique qui naissent d'initiatives individuelles et de visions parfois contradictoires.

Les SIG 3D est une solution émergente qui suscite un grand intérêt auprès de la communauté des chercheurs mais qui, auprès des producteurs et utilisateurs, pose des contraintes de coût et de mise en œuvre. Le coût de production, de mise à jour et d'échange des données 3D impose le fait de travailler en coproduction pour définir une donnée 3D de référence qui soit interopérable. Un SIG 3D collaboratif permettrait de gérer cette information dans un cadre inter-organisationnel pour supporter les projets de prise de décision commune.

Contrairement à ce qu'on pourrait prétendre, on est loin d'ajouter un autre degré de complication par la prise en charge de la 3D, mais l'idée se trouve justifiée par le constat suivant : les SIG 2D qui ont pris le temps d'évoluer dans un contexte intra organisationnel peinent à répondre aux nouvelles exigences. Si on n'est pas encore

arrivé à maturité au niveau des solutions 3D, la collaboration poserait moins d'impacts sur les solutions existantes. Ça serait une façon d'anticiper les problèmes futurs d'échange des données 3D. Les SIG 3D collaboratifs se veulent comme solution potentielle qui permet de maîtriser le maximum de défis en tendant vers des solutions de conception et de réalisation collaborative.

3.3. Les SIG 3D collaboratifs : Une suite ou une révolution ?

Les SIG 3D collaboratifs ouvrent de nouvelles perspectives. Mais leur mise en œuvre et leur appropriation collective posent des problèmes méthodologiques. L'une des principales difficultés reste de structurer l'information dans un système cohérent pour pouvoir la partager, l'exploiter et la valoriser. Aussi, il faut savoir comment rapatrier les solutions existantes provenant des différentes parties dans le nouveau système. La conception collaborative est loin de répondre à ce besoin. Le problème ne se réduit pas à un simple partage de l'information pour lequel il existe des solutions techniques mais le grand défi est l'ajout d'une plus-value aux données existantes et leurs confrontations pour ressortir les plus pertinentes et leur intégration dans la nouvelle solution (Rousson et al, 2007). Si ces problématiques s'imposent à une étape du processus, la coproduction de données 3D selon de nouvelles spécifications permettra de les dépasser dans le futur. Aussi, l'enrichissement par de nouvelles acquisitions de données pour construire un fond de données 3D est une étape non négligeable. La question de l'interopérabilité et des grandes orientations régionales et internationales ajoute une problématique à résoudre. Le SIG 3D collaboratif revêt donc un caractère révolutionnaire.

4. Vers une méthode de conception de SIG 3D collaboratif

4.1. Vue du problème

D'une façon générale, la mise en place d'applications géomatiques suscite un cadre méthodologique organisationnel adéquat répondant à l'implémentation spécifique de SIG de par la nature des données manipulées, leurs complexités, leur volume et du nombre d'acteurs intervenant dans la mise en place du système. Des méthodes de conception existent pour la mise en place d'un SIG d'une organisation mais sont loin de répondre à tous les défis. Ces méthodes classiques sont aujourd'hui dépassées par l'avancée spectaculaire au niveau technique, social et politique et les grandes mutations autour de l'information géographique.

L'introduction de la 3D au niveau des méthodes de mise en place de SIG engendre des contraintes additionnelles. L'impact technique consiste en l'adaptation des outils et méthodes à un référentiel 3D. Le passage à un SIG 3D demande une investigation dans la modélisation et le choix de bonnes structures de stockage. Aussi, la 3D au niveau conceptuel aide à spatialiser les besoins et concevoir des méthodes de reengineering des données existantes. Dans ce contexte, nous orientons notre vision vers une méthode pour la mise en place d'un SIG 3D collaboratif qui aborde les questions techniques et conceptuelles autour de la 3D.

Il existe un bon nombre de formalismes et méthodes dédiés pour la modélisation des applications géomatiques. Nous citons sans exhaustivité les méthodes MODULR, MECOSIG, Pollen, OMEGA, T-OMEGA, MADS, GeoFrame,... Dans le contexte des méthodes orientées objets géomatiques, seules quelques-unes proposent une démarche méthodologique telles que MECOSIG, OMEGA et T-OMEGA (Ben Youssef, 2010). Notre choix a porté sur la méthode MECOSIG comme méthode sur base de laquelle nous pouvons mener notre réflexion sur le problème. Ce choix se justifie par le fait que la méthode est très bien documentée et supportée par une démarche méthodologique que nous proposons de mettre en examen en face de notre objectif de SIG 3D collaboratif. Nous renvoyons le lecteur vers l'ouvrage de (Pantazis et al, 1996) qui présente la méthode MECOSIG sur base de laquelle nous fondons notre analyse.

4.2. MECOSIG : Critiques et insuffisances

Notre objectif n'est pas de faire une étude d'évaluation sur MECOSIG dans tous ses aspects mais nous focalisons sur ce qu'il est possible de faire avec cette méthode pour répondre à notre besoin de conception et de « réalisation » d'un SIG 3D collaboratif. Ainsi, la méthode sera confrontée aux besoins d'interopérabilité des données et à la question conceptuelle et technique autour de la 3D dans un cadre collaboratif. Des propositions seront formulées pour répondre à ces problèmes.

MECOSIG propose une grande diversité de concepts qui rend parfois l'interface entre les étapes difficile à discerner dans la matrice de conduite du projet (ex : la conception dans le niveau descriptif et l'analyse dans le niveau conceptuel peuvent se chevaucher). Le concepteur/analyste peut tomber dans la redondance ou ne pas pouvoir cerner la question de chaque étape de la démarche. Telle qu'elle est conçue, la méthode présente un aspect conceptuel théorique, il faut des recherches pour la transformer en méthode de conception et de réalisation pour porter solution aux conflits conceptuels/techniques qui peuvent s'imposer dans la démarche. La démarche suppose une connaissance préalable des objectifs stratégiques, ce qui n'est pas le cas dans un processus de collaboration qui tente de rapprocher des visions divergées parfois hétérogènes pour aboutir à une vision commune. MECOSIG ne traite pas la question sur l'aspect de normalisation et d'interopérabilité des données dans les étapes conceptuelles. En résumé, la méthode telle qu'elle est conçue ne

répond pas aux aspects collaboratifs des SIG 3D mais nous allons l'exploiter comme ligne directrice pour guider nos réflexions que nous proposons de présenter dans la partie suivante.

4.3. Une méthode SIG 3D collaboratif : Principes et fondements théoriques

Les SIG 3D collaboratifs ont plusieurs spécificités liées à leur dimension collaborative qui impose des contraintes additionnelles aux projets de SIG classiques, notamment la multiplicité des acteurs impliqués et la tension entre l'individuel et le collectif. En outre, leur prise en charge de la 3D, un produit non encore mature, engendre des contraintes techniques d'appropriation, d'intégration et de gestion de la donnée 3D collaborative. Il en sort que la mise en œuvre d'un SIG 3D collaboratif doit faire appel à une méthode robuste qui offre aux collaborateurs un cadre méthodologique adéquat.

- La méthode doit fournir une démarche qui tient compte des aspects et des contraintes liées à un travail collaboratif.
- Elle doit proposer des orientations/des bases, voire des outils pour soutenir le travail collaboratif et faciliter la prise de décision.
- Elle doit proposer des directives de reengineering pour les partenaires de l'action collaborative.
- Elle doit s'aligner sur les normes et orientations internationales concernant les infrastructures de données spatiales.
- La méthode doit être ouverte et non contraignante en offrant un cadre méthodologique permettant au concepteur d'adapter les solutions conceptuelles aux contraintes existantes.
- La méthode doit fournir des moyens d'aides à l'analyse et à la conception du problème.
- La méthode doit proposer une démarche flexible pour atténuer la complexité du projet.

Une méthode de SIG 3D collaboratif par la diversité des aspects à gérer est basée sur le principe de la collaboration qui valorise le processus. Elle a la particularité de ne pas être dédiée à une organisation comme c'est le cas de MECOSIG mais la démarche qu'elle propose est par analogie centrée sur une organisation dite « abstraite » qui représente toutes les organisations partenaires. Toutes les solutions conceptuelles pour les classes d'analyse doivent être injectées dans les organisations pour un reengineering. L'unité de base pour le traitement est la classe (inversion de la matrice de conduite de projet dans MECOSIG) et la problématique de la 3D doit s'imposer au démarrage pour spatialiser les besoins en 3D.

4.4. *Quels outils pour une méthode de SIG 3D collaboratif ?*

4.4.1. Les moyens d'analyse

La méthode MECOSIG propose de conduire l'étude sur base d'un ensemble de classes de préoccupations (Donnée, Traitements, Flux de données, Organisation et Intégration) et de niveaux d'abstraction (niveau Descriptif, Conceptuel, Organisationnel et Logico-physique) (Pantazis et al, 1996). Le projet de SIG 3D collaboratif porte sur des organisations multiples qui adhèrent à une action de collaboration et qui sont ciblées avec le même niveau de préoccupation. Les données et traitements constituent des composantes principales de tout SIG, nous les retenons comme classes principales de préoccupations. Pour maîtriser le cheminement de l'analyse, nous proposons d'associer les flux et l'intégration à la classe « Donnée » que nous considérons comme classe basique qui étudie tous les aspects qui se rapportent aux données. Cette concaténation est loin de minimiser l'importance de ces composantes mais nous l'adoptons dans le sens de centrer notre étude sur tous les aspects qui relèvent des données et pour lesquels nous ne pouvons pas prévoir autant de classes. La question d'interopérabilité nous paraît primordiale et nous proposons de lui associer une classe dite : classe « Interopérabilité ».

L'interopérabilité porte aussi bien sur les données que sur les traitements, la classe interopérabilité est considérée comme une superclasse dont l'analyse doit être projetée sur les deux classes : donnée et traitement pour répondre à la question de la mutualisation des données et celle des traitements. Pour les niveaux d'abstraction, nous proposons une analyse croisée selon un niveau descriptif-analytique, un niveau conceptuel et un niveau technique. Par niveau technique nous faisons référence au niveau qui permet au concepteur de mettre en œuvre les solutions conceptuelles de natures diverses que soient-elles sans s'attacher uniquement au niveau d'implémentation de la base de données. Pour chaque niveau, nous retenons des sous niveaux d'analyse et de conception proposés par MECOSIG.

Le niveau organisationnel (proposé par MECOSIG) n'est pas pris en compte vu que notre objectif n'est pas focalisé sur la géomatisation de l'organisation. Dans une solution collaborative, les contraintes organisationnelles ne doivent pas trop influencer les choix conceptuels. C'est l'organisation qui doit s'adapter aux contraintes du projet. Nous visons à ce que la collaboration apporte des solutions à ces contraintes organisationnelles par le partage et l'échange des ressources. Cependant, le retour aux organisations est essentiel pour un reengineering et nous le prévoyons comme étape consistante de la démarche.

4.4.2. *La démarche*

Plusieurs variables autour de la mise en œuvre d'un SIG 3D collaboratif font qu'on ne peut pas prévoir une démarche universelle qui maîtrise toutes les questions. En effet, la nature des données et systèmes existants, le niveau de disponibilité de la 3D, la nature des relations entre partenaires sont des paramètres qui influencent les choix conceptuels et aussi les cheminements à adopter. Cependant, le concepteur/analyste doit disposer d'un outil pratique pour aborder la problématique. Nous visons une démarche qui traite en priorité des questions d'interopérabilité, de solutions conceptuelles 3D et de reengineering.

4.4.2.1 Des étapes à valoriser

A un premier niveau, la démarche vise des organisations potentielles dans la production des données qui centrent leur objectif autour des données de référence 3D. Les organisations sont généralement trop hésitantes sur l'adhésion à une action collaborative. Elles ont un regard très restrictif et douteux sur les apports de la collaboration dans la satisfaction de leurs propres besoins et peuvent montrer des réticences en cours de processus surtout lorsqu'elles voient leur chaîne de production étudiée et leurs données mise à l'examen, ce qui risque d'entraver le projet en cours de chemin. La phase de sensibilisation qui revêt déjà une importance dans les projets de géomatisation d'une organisation, a une plus grande plus-value dans le cas d'une action de collaboration. L'objectif est de démontrer l'utilité de la collaboration et inciter les partenaires à y adhérer, présenter de façon objective les avantages (partage de savoir-faire, des coûts, interopérabilité,...) et les inconvénients (processus longs, impact organisationnel, reengineering,...).

Une phase de pré-collaboration permet aux organisations d'instaurer une première base qui servira de plateforme pour la conduite du projet. Dans cette phase, les partenaires sont invités à se réunir pour se connaître, faire des choix stratégiques et définir un cadre (un référentiel) de travail. Sachant que les objectifs des partenaires sont souvent divergents et réduits dans un contexte intra-organisationnel, cette étape se veut une étape de concertation et de rapprochement des visions. Dans une action collaborative, il faut se mettre d'accord de façon claire et irrévocable sur des objectifs communs qui font dissoudre les identités organisationnelles. Des objectifs préliminaires peuvent toujours être révisés au cours de la démarche. L'idéal serait de démarrer de façon très peu formalisée au niveau des individus (Pornon, 1998). Ceux-ci serviront de locomotives par la suite. Il faut centrer les objectifs sur la définition de l'information de référence sans les réduire au seul contexte technique. Les objectifs communs doivent être classés en objectifs techniques immédiats (ex : établir un modèle 3D de bâtiment) et ceux plus stratégiques. D'autres étapes de la démarche sont illustrées par l'organigramme de la figure1.

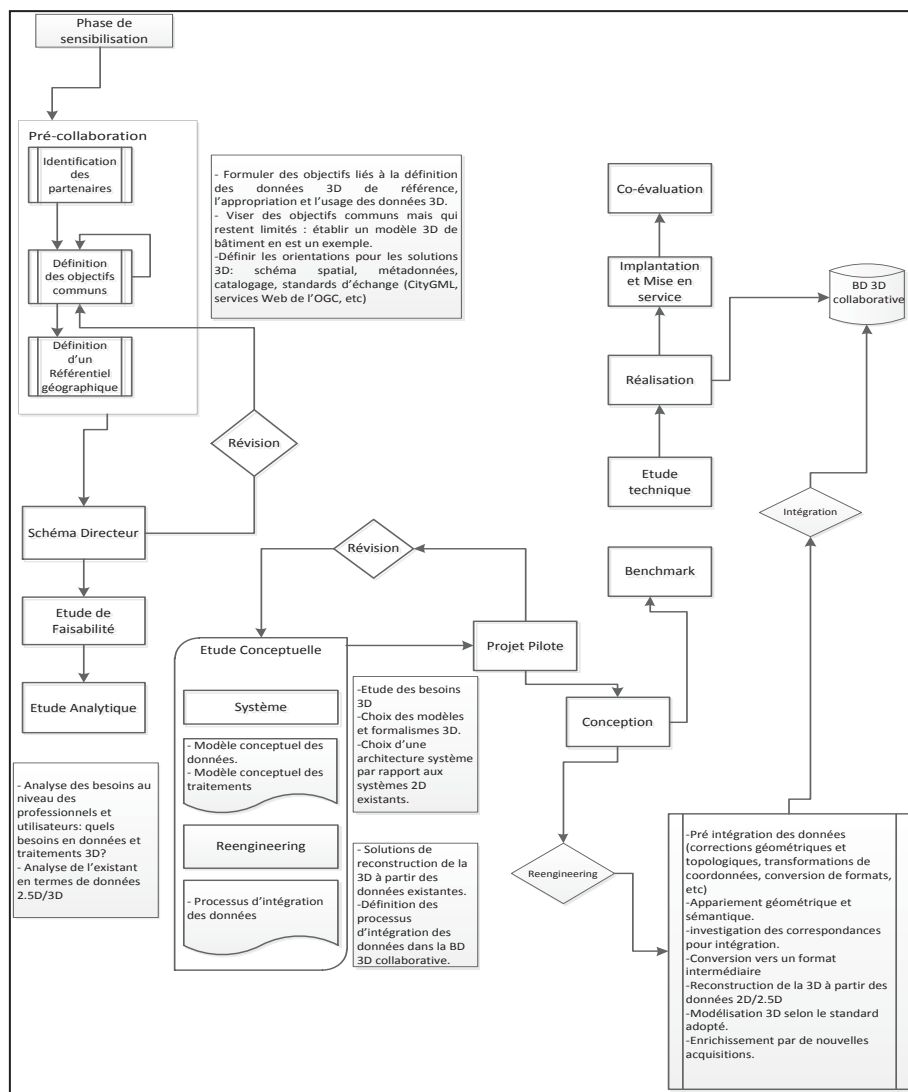


Figure 1. Organigramme de notre démarche

4.4.2.2. La question de l'interopérabilité des données

Si l'initiative de la collaboration est née du besoin de faire face aux contraintes imposées par les avancées que connaît l'information géographique, celle-ci ne peut se développer que sur base d'un référentiel géographique qui répond aux exigences de la nouvelle information géographique et qui constitue un cadre pour le

développement du projet. Dans cette phase, il faut mener des réflexions sur les défis autour de l'information géographique: les normes, les standards, la multi dimensionnalité des données et les grandes orientations internationales (notamment la norme Inspire). Cette étape doit définir une vision, des objectifs, des principes et orientations en matière de gestion de l'information ; préciser des obligations et exigences de conformité aux lois et règlements en matière de gestion de l'information géographique et définir les normes et standards à adopter. Le référentiel géographique offre une vision globale et des orientations à une échelle plus grande qui peut dépasser la dimension et les objectifs arrêtés pour le projet. Néanmoins, il servira de référence.

Un catalogue de données commun est une des premières étapes à valoriser pour l'interopérabilité des données. Dans son analyse des démarches de partenariat, (Pornon, 1998) considère que la réalisation de catalogues est un préalable nécessaire à la mise en œuvre d'échanges de données entre organismes. Au stade de la collaboration, les données deviennent des moyens et non des fins. La question technique est largement maîtrisée par les travaux de l'ISO (norme 19110). La difficulté est plutôt d'ordre sémantique.

4.4.2.3 La Modélisation conceptuelle 3D

Il est communément admis que la modélisation conceptuelle doit s'affranchir de toute considération technologique. Mais souvent, le technique et le conceptuel s'entremêlent et il se trouve que la conception de modèles de données soit influencée par des choix techniques préalables justifiés par des contraintes de coût ou de conformité entre les composantes du système. Par exemple, une solution combinant le choix du SGBD PostgreSQL avec sa cartouche spatiale Postgis et du standard CityGML poserait la contrainte d'intégration des données vu que jusqu'à maintenant, il n'y a pas d'intégrateur CityGML dans Postgis. A travers cet exemple, nous voulons exprimer le fait qu'une connaissance préalable pour les solutions techniques 3D existantes et leur interopérabilité est requise pour guider la modélisation conceptuelle.

La modélisation doit faire appel à une approche orientée modèle qui s'affranchit de toutes les questions de transfert ou de système. Dans ce contexte, la norme suisse Interlis offre une solution libre potentielle basée sur une approche de modélisation qui permet de décrire un modèle de données en utilisant un schéma de langage conceptuel (CSL) qui conserve les structures de données indépendamment de tout système ou de tout format de transfert. Les formats de transfert peuvent être dérivés automatiquement en utilisant des compilateurs appropriés. Interlis offre un langage textuel qui peut être utilisé comme complément d'UML. Des modèles hiérarchiques peuvent être dérivés d'un modèle de base en assurant une interopérabilité sémantique (Staub, 2007).

Dans la modélisation conceptuelle des bases de données 3D, il existe des formalismes étendus comme le PVL spatial 3D de Perceptory: langage UML enrichi

de « Plug-ins for Visual langage » (PVL) qui propose des pictogrammes pour la modélisation géométrique 3D (Larrivée et al, 2002). Le langage est supporté par un AGL qui utilise le logiciel Visio de Microsoft (la version Java est en cours de développement). Perceptory possède un générateur de codes pour Oracle et de rapports compatibles avec la norme ISO/TC211. L'utilisation de Perceptory offre l'avantage de pouvoir exprimer les besoins indépendamment de l'implantation de la base de données en permettant aussi le choix entre technologies 2D, 2.5D ou 3D pour chaque classe. Cette mise en évidence en termes de dimensionnalité de l'information permet de faciliter les choix techniques sur les logiciels 3D. (Larrivée et al, 2002), (Larrivée et al, 2005)

4.4.2.4 Le reengineering

Le reengineering est un processus global portant sur plusieurs actions autour des données pour leur attribuer une valeur ajoutée et les intégrer dans un nouveau système. Ces actions peuvent être classées en différentes fonctions dont les principales portent sur la préparation des données (correction géométrique, correction topologique, suppression des redondances,...), leur enrichissement par de nouvelles acquisitions et des fonctions de standardisation (Hardy, 2001). Les données sont souvent plaquées sur une culture métier ce qui rend leur reengineering difficile à accomplir. C'est une étape très longue et très fastidieuse qu'il faut concevoir de manière pragmatique pour minimiser les risques d'échec (Talla et al, 2011). Une fois les solutions conceptuelles du système mises en place, cette étape peut être menée en parallèle avec la suite du projet. Dans une démarche de conception d'un SIG 3D collaboratif, une étape de reengineering permet le retour aux organisations pour répondre aux solutions conceptuelles qui concernent l'interopérabilité des données. Une question déterminante par rapport au processus à adopter porte sur l'état des données existantes et les choix conceptuels adoptés pour la base de données collaborative. Les données existantes doivent subir un reengineering pour alimenter le catalogue commun et la base de métadonnées. Après la préparation des données, celles-ci doivent être intégrés dans le nouveau système. Un processus d'intégration doit tenir compte de plusieurs critères : la qualité, le niveau de détail, l'actualité, etc. Dans la figure 2, nous proposons un scénario d'intégration de données existantes dans une base de données collaborative 3D. Les données sont converties en un format intermédiaire qui répond à des spécifications prédéfinies avant qu'elles soient soumises à des contrôles de cohérence et de qualité pour être intégrées dans la base de données. Pour des fournisseurs « expérimentés » ou des données nouvellement acquises selon les spécifications, le passage au format intermédiaire peut être épargné.

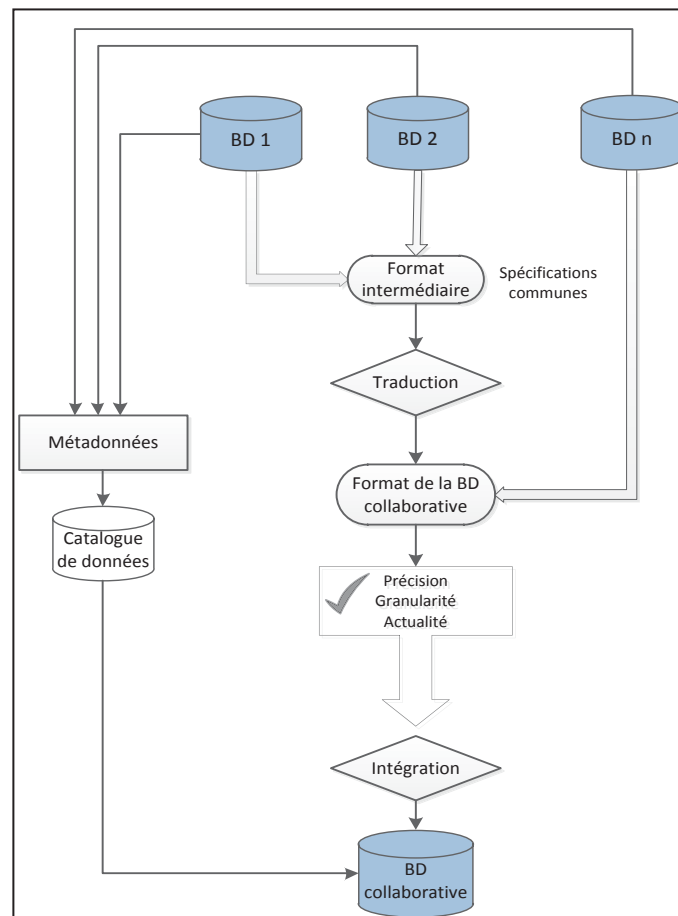


Figure 2. *Organigramme d'intégration des données*

5. Conclusion et perspectives

L'objectif de ce papier consistait, à travers une analyse et une mise au point sur les concepts existants et ceux émergents en géomatique à mettre en évidence les potentialités des SIG 3D collaboratifs dans la maîtrise d'une grande part des questions qui s'imposent aux organismes potentiels dans la production des données

de base, notamment le passage à la 3D, la minimisation des coûts d'acquisition des données, leur mise à jour et leur interopérabilité. Nous avons pu expliquer d'une manière objective le contexte dans lequel ces solutions peuvent évoluer en structurant les préoccupations et les questions à résoudre et en mettant en surface les contraintes qu'elles engendrent. Il en sort que le SIG 3D collaboratif s'avère une solution très prometteuse qui, malgré son caractère révolutionnaire par rapport aux systèmes existants et la difficulté de mise en œuvre, permet de maîtriser les diversités sémantiques, les objectifs divergents en vertu d'une conception et une mise en œuvre collective d'une solution bénéfique pour ses partenaires. Pour cerner une part de la question de faisabilité, nous avons traité l'aspect méthodologique et la modélisation conceptuelle d'un SIG 3D collaboratif en avançant nos premières réflexions sur une méthode de conception collaborative permettant d'atténuer la difficulté et d'offrir une base pour conduire une telle solution. Nos futures recherches se proposent d'aller vers une démarche élaborée qui s'inspire de cas concrets. Nos réflexions seront ainsi confrontées à une des réalités et peuvent évoluer et s'enrichir.

6. Bibliographie

- Abdul Rahman A., et Pilouk M., *Spatial Data Modelling for 3D GIS*, Springer-Verlag Berlin Heidelberg, ISBN 978-3-540-74166-4, 2008
- Ben Youssef M., Modélisation conceptuelle pour les applications géomatiques temps réel, Thèse de doctorat, l'Institut National des Sciences Appliquées de Lyon et l'Ecole Supérieure de Communications de Tunis, 2010.
- De la Cambay B., Three-Dimensional (3D) Modelling in a Geographical Database, Publié dans les Actes de la conférence: *Auto-Carto'11, Eleventh International Conference on Computer Assisted Cartography*, Minneapolis, USA, p. 338-347.1993
- Dedekorkut A., Determinants of Success in Interorganizational Collaboration for Natural Resource Management, thèse de doctorat, Université de l'Etat de Floride, 2003.
- Gröger G., Kolbe T.H., Czerwinski A., Nagel C., OpenGIS City Geography Markup Language (CityGML) Encoding Standard, Version 1.0.0, OGC Doc. No. 08-007r1, Open Geospatial Consortium, 2008
- Hardy P., Active objects and dynamic topology for spatial data re-engineering and rich data modeling, séminaire Dagstuhl, 2001
- Joliveau T., Géomatique et gestion environnementale du territoire. Recherches sur un usage géographique des SIG, mémoire d'Habilitation à Diriger des Recherches, Département de Géographie, Université de Rouen, vol. 1, 2004.

- Larrivée S., Bédard Y., Pouliot J., Fondement de la modélisation conceptuelle des bases de données géospatiales 3D, 2002, (en ligne), (<http://yvanbedard.scg.ulaval.ca/wp-content/documents/publications/412.pdf>)
- Larrivée S., Bédard Y., Pouliot J., How to Enrich the Semantics of Geospatial Databases by Properly Expressing 3D Objects in a Conceptual Model, R. Meersman et al. (Eds.): OTM Workshops 2005, LNCS 3762, pp. 999 – 1008, 2005. Springer-Verlag Berlin Heidelberg, 2005.
- Mang S., Contribution à la problématique de la qualité des données géographiques collaboratives Cas d'OpenStreetMap, Rapport GAIAGO, 2010
- McDougall K., Rajabifard A., Williamson I., Understanding the Motivations and Capacity for SDI Development from the Local Level, From Pharaohs to Geoinformatics, *FIG Working Week 2005 and GSIDI-8*, Cairo, Egypt April 16-21, 2005.
- Mericskay B., et Roche S. al, 2010, « Cartographie et SIG à l'ère du Web 2.0 Vers une nouvelle génération de SIG participatifs », *SAGEO 2010* Toulouse.
- Mericskay B., Roche S., « Cartographie numérique nouvelle génération: impacts de la néographie et de l'information géographique volontaire sur la gestion urbaine participative », *HyperUrbain.2*, Paris, Europa, 2009.
- Noucher M., « Quels leviers d'action pour rendre nos SIG plus collaboratifs? », *Bulletin e-geo.ch* 2009, p. 12-14
- Oosterom P., Stoter J., 5D Data Modeling: Full Integration of 2D/3D Space, Time and Scale Dimensions, *GIScience*, LNCS 6292, pp. 310–324, 2010. Springer-Verlag, Berlin Heidelberg, 2010.
- Orban-Ferauge F., *Systèmes d'information géographique participatifs et aménagement du territoire, Expériences philippines citoyennes de désenclavement*, Namur, Presses Universitaires de Namur, 2011.
- Pantazis D., Donnay JP., *La conception de SIG, méthode et formalisme*, Paris, Editions Hermès, 1996
- Pornon H., Noucher M., « Bilan et perspective de 20 années de géomatique : Vers des SIG plus collaboratifs, la Géocollaboration », *Géomatique Expert*, vol. 58, 2007, p. 56-60.
- Pornon H., Partenariats relatifs aux SIG : Proposition d'une typologie et d'une grille d'évaluation, Rapport CERTU, IETI Consultants, 1998.
- Roche S., « Géoweb et travail collaboratif: quelques éléments de cadrage », *Bulletin e-geo.ch*, 2009
- Rousson C., Roche S., « Vers l'élaboration d'une infrastructure géomatique collaborative- la gestion forestière comme modèle d'étude, 1^{ère} soumission *SAGEO*, 2007
- Staub P., « A model-driven Web Feature Service for enhanced Semantic interoperability » *OSGeo Journal*, vol .3, 2007, p. 38-43
- Talla M., Valverde R., Data oriented and process oriented strategies for legacy information systems Reengineering, *Joint Colloquium on Emerging Technologies in Computer Electrical and Mechanical*, 2011.

Turner Andrew J., Introduction to Neogeography, Londres, *O'Reilly Media*, 2006.

Student satisfaction towards software-specific and software-independent tutorials of GI S&T

Hubeau, M.* - De Meyer, A. - Steenberghen, T.* - Van Orshoven, J.*****

** R&D Division SADL (Spatial Application Division Leuven),
KU Leuven, Celestijnenlaan 200^E, 3001 Leuven, Belgium*

Marianne.hubeau@SADL.kuleuven.be

Therese.steenberghen@SADL.kuleuven.be

Jos.vanorshoven@SADL.kuleuven.be

*** Division Forest, Nature and Landscape,
Department of Earth and Environmental Sciences,
KU Leuven, Celestijnenlaan 200^E, 3001 Leuven, Belgium*

Annelies.demeyer@ees.kuleuven.be

ABSTRACT: Commercial and closed source desktop GIS-software and their free and open source counterparts (CCSS4G and FOSS4G) have reached a significant level of standardisation regarding interfaces, content of their toolboxes and supported data formats. Therefore, it has become envisagable to develop tutorials which are independent of a specific software, the so-called 'conceptual exercises'.

A web-based questionnaire survey addressing all students of the introductory GI S&T-practical was held during the 2010-2011 and 2011-2012 academic year at KU Leuven. This questionnaire survey allowed the comparison of the student satisfaction regarding different GI S&T educational methods. The concerned methods are (i) the traditional in-class approach whereby six collective sessions of two hours in a computer class were held under guidance of a teaching assistant and (ii) Supervised Self-Study (SSS) whereby students independently solve and report about a set of exercises while having the opportunity to receive advice and feedback. Moreover, we investigated the student satisfaction towards three different tutorial formats, namely (i) the software-dependent prescriptive tutorial-CCSS4G, (ii) the software-dependent prescriptive tutorial-FOSS4G, and (iii) the software-independent tutorial-conceptual.

The results of the web-based survey revealed a significant difference between bachelor and master students towards the SSS-approach. Master students were more comfortable with the SSS. Moreover, the results revealed that conceptual exercises were appreciated which stimulates to further develop, implement and investigate this tutorial format.

KEY WORDS: GI S&T, Supervised Self-study, Conceptual exercises, Life-long learning, Student satisfaction, CCSS4G and FOSS4G.

1. Introduction

Since the late 1980ies until today, GI S&T-students at KU Leuven have been exposed to specific commercial and closed source software for geomatics (CCSS4G) like pcArcInfo, MapInfo, IDRISI and ArcView-GIS. Although, since 1999, supervised self-study (SSS) is embodied in the educational vision underpinning all study programmes at KU Leuven, the adherence to CCSS4G prevented a true implementation of SSS, despite broad intra- and internet connectivity.

Since 2009, SSS in software-based GI S&T-practical became possible thanks to the availability of reliable free and open source software for geomatics (FOSS4G) (Schweik *et al.*, 2009, Van Orshoven *et al.*, 2010). SSS implies that students work independently to solve and report about a set of exercises while having the possibility to receive thorough advice and feedback. An advantage is that students acquire extra skills like solving problems related to installation, hardware and operating system compatibility and data interoperability. Since its introduction, the combination of SSS with FOSS4G has been well accepted by KU Leuven students. Moreover the learning effectiveness, access to learning materials and staff satisfaction is similar to what is noticed in more traditional GI S&T-courses (Van Orshoven *et al.*, 2010; Hubeau *et al.*, 2011a, Hubeau *et al.*, 2012).

With the rapid evolution of software systems (Steiniger *et al.*, 2009), even more rapid for FOSS4G than for CCSS4G, software-specific tutorials are rapidly outdated and important efforts are required to keep the exercises compatible with the latest software versions. Table 1 illustrates the rapid evolution of the FOSS4G QuantumGIS (QGIS) (<http://www.qgis.org>) and OpenJUMP (<http://www.openjump.org>). Eight new versions of both softwares were released over a period of 3 years (January 2009-January 2012). A negative consequence of this evolution is that many (if not all) screenshots and workflow descriptions need to be updated and replaced every time when a new version is released leading to a short-term validity of the learning materials. Also, the tutorials developed for a specific software typically encompass prescriptive exercises which usually describe exercises step-by-step. This allows students to follow the description of the exercises without acquiring in-depth knowledge of the GIS-concepts behind the software functionalities.

Table 1. *Consecutive versions of QGIS and OpenJUMP*

QGIS version	Release date	OpenJUMP version	Release date
1.0.0 Kore	January 5, 2009	OpenJUMP 1.3	April 19, 2009
1.1.0 Pan	May 12, 2009	OpenJUMP 1.3.1	December 6, 2009
1.2.0 Daphins	September 1, 2009	OpenJUMP 1.4	October 24, 2010
1.3.0 Mimas	September 20, 2009	OpenJUMP 1.4.0.2	November 8, 2010
1.4.0 Enceladus	January 10, 2010	OpenJUMP 1.4.0.3	December 1, 2010
1.5.0 Tethys	July 29, 2010	OpenJUMP 1.4.1	July 2, 2011
1.6.0 Capiapo	November 27, 2010	OpenJUMP 1.4.2	September 1, 2011
1.7.0 Wroclaw	June 19, 2011	OpenJUMP 1.5.0	January 2, 2012

Therefore, in 2011, we introduced SSS based on what we call ‘conceptual exercises’. Conceptual exercises are described in a software-independent way so that they can be solved with every GIS-software (CCSS4G or FOSS4G) that contains the necessary functionalities. This approach has become feasible since desktop GIS-software has become largely standardised with regard to interface, functionality, terminology, read-write of data formats etc.

This paper investigates the student satisfaction with two different educational methods (traditional approach (TA) and supervised self-study (SSS)) and with three different tutorial formats (software-dependent IDRISI-CCSS4G, software-dependent QGIS-FOSS4G and software-independent-conceptual). The results of a web-based questionnaire survey addressing all the students who followed one of these teaching approaches are reported and discussed. The learning effectiveness, learning experience and learning efficiency of the latter two approaches is presented in another paper (Hubeau *et al.*, 2012).

2. Materials and methods

2.1. Educational methods

During the fall semester of the academic year 2011-2012, three different educational methods were implemented for the practical part of the introductory GIS&T-course organised jointly for the Bachelor programmes of the Faculty of BioScience Engineering and Sciences of KU Leuven: (i) the traditional in-class approach (TA) with the software-dependent prescriptive tutorial using the CCSS4G IDRISI (<http://www.clarklabs.org/>), (ii) SSS with the software-dependent prescriptive tutorial using the FOSS4G QGIS and GRASS, and (iii) SSS with a software-independent conceptual tutorial. Also, in the fall semester of the academic year 2010-2011, the first two were implemented. All three methods aimed at allowing students to put the theoretical concepts of GIS&T into practice, to acquire hands-on skills in the use of the selected GIS-software and to become prepared for use of arbitrary GIS-tools in other circumstances (life-long learning).

Students of the TA followed the practical module in a computer classroom during six mandatory collective sessions of two hours guided by a teaching assistant. The GIS-software to solve the exercises was IDRISI, a commercial and closed source software for geomatics (CCSS4G). This traditional approach (TA) was mainly a top-down approach, whereby the teaching assistant gave the instructions and guided the students intensively. Consequently, students had little opportunity to achieve secondary objectives such as solving software, hardware and operating system compatibility problems, overcoming lack of data interoperability and consulting internet-based software user communities.

Supervised self-study (SSS) was taught through web-based learning. All the exercises, datasets and instructions were made available online through the e-

learning platform Toledo (<http://toledo.kuleuven.be>), a blackboard software application. This e-learning environment enabled to create an active learning environment by providing interactive learning modules, assignments and feedback to students (Dorner et al., 2011). Moreover, students were supposed to submit assignments, communicate with the tutor or peer students, and share documents (Valentine, 2002). SSS requires that students have independent access to the GIS-software. Therefore, free and open source software for geomatics (FOSS4G) was introduced. Two variants of the web-based SSS were organised. The first variant encompassed SSS with four non-compulsory feedback sessions and online interaction using step-by-step explained exercises for QGIS and its GRASS-plugin (<http://grass.fbk.eu/>). The second variant was of the SSS-type whereby the interaction between students and tutor was limited to online interaction using software-independent exercises which can be addressed with any appropriate GIS-software, CCSS4G or FOSS4G.

2.2. Conceptual exercises

Conceptual exercises are software-independent exercises which can be solved with any GIS-software that contains the necessary functionalities. The objectives of conceptual exercises are (i) to stimulate acquisition of in-depth knowledge about GIS-concepts by students by encouraging them to explore GIS-functionalities without step-by-step explained exercises (i.e. no screenshots, no illustration of buttons etc.), (ii) to force students to choose and select a GIS-software which encourages them to work with internet-based communities and to differentiate between different GIS-desktop software, and (iii) to develop long-living tutorials requiring limited maintenance due to software and software version independence. The structure of a conceptual exercise is rather standardised (Figure 1). Not any reference is made to a specific GIS-software. An example of a part of a conceptual exercise is presented in Figure 2. Firstly, all GIS-functionalities required to solve the exercise are listed in a checklist. This allows students to check whether the GIS-software they intend to use offers all the necessary functions to complete the exercise. Secondly, a short description represents the specific objectives of the exercise. Thirdly, the actual exercise is explained by a general workflow which allows students to solve the exercise based on the solution scheme. Finally, a demo-movie is available of every exercise. This movie shows the workflow of the exercise using an arbitrary GIS-software. In this way, the different GIS-concepts and functionalities are visualised and illustrated so that the students can correct or improve their own solution approach (Hubeau *et al.*, 2011b).

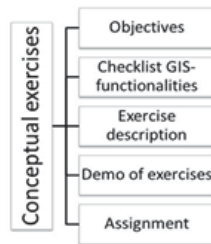


Figure 1. *Structure of a conceptual exercise*

Exercise: Spatial data analysis

Checklist GIS-functionalities

- Import vector-layers → Shapefiles
- Execute spatial queries/data analysis and more specifically:
 - Buffer analysis
 - Overlay analysis: Intersect operation
 - Dissolve
- Create a new column in attribute tables
- Calculate the area as a value to be displayed in a new column of the attribute table → a field-calculator need to be present

Assignment

Most GIS-software have some analytical capabilities, i.e. the possibility to process one or more geodatasets to answer particular questions.

The aim of this exercise is:

1. To discover some analytic capabilities of the GIS-software;
2. To process geodata and to perform spatial queries:
 - a. Buffer analysis
 - b. Overlay analysis: Intersection
 - c. Dissolve function

The purpose of the assignment is to determine which land cover or land use classes occur in 100m wide zones along the river system of the Tabacay watershed. You also need to calculate the area of the present land cover classes. The geodatasets we need for this analysis to describe the river system on the one hand and the land cover/land use distribution within the watershed on the other hand.

The necessary geodatasets are:

- *Rios.shp* ; *Coberturas.shp*

Solution scheme

- Create a new GIS-project.
- Load the shapefiles *rios.shp* and *coberturas.shp*.
- Make sure the map units are *meters*.
- Change the Coordinate Reference System to *WGS84/UTMzone17S*.

STEP 1: Buffer analysis

The first operation is a **buffer** operation. The process involves generating a buffer round an existing geographic object and then identifying or selecting objects based on whether they fall inside or outside the boundary of the buffer zone.

- The input vector is *rios.shp*.
- The buffer distance is **100 meters**.
- The new layer *rios_buffer.shp* needs to be displayed.
- Examine whether the expectations are met. How many polygons are in this geodataset? (TIP: Check the attribute table)

Demo movie:

<https://videolab.avnet.kuleuven.be/video/?id=5bec022e39d93760bd522b9c459806a2&autostart=false&height=388&width=640>

Figure 2. *Example of a conceptual exercise about spatial data analysis following the standardised structure*

2.3. Web-based survey

The student satisfaction was measured through a web-based survey hosted by Toledo. A questionnaire based on previous surveys (Abu el Nasr *et al.*, 2008, Van Orshoven *et al.*, 2010, Hubeau *et al.*, 2011a) was conducted in December 2010 and December 2011 immediately after the final assignment of the GI S&T-practical. Table 2 presents the questions proposed by the survey. These questions cover different subjects including the students' attitude towards the different teaching approaches and towards the tutorial format.

Table 2. Structure of the web-based questionnaire

Question	Possible answers
I. Express the level of agreement with the following statements: Ia. Questions concerning the students' attitude towards the teaching approach <ul style="list-style-type: none"> - Human support, advice and feedback were sufficiently available - The form of the course gave me the freedom to work independently - The form of the course was stimulating to deal active with the course materials - There was sufficient possibility for regular interaction with peer students Ib. Questions concerning the students' attitude towards the form of the tutorial <ul style="list-style-type: none"> - The form of the tutorial was time efficient - The software contained too many bugs - The theoretical background of GIS became more clear - Experience with the software prepared for professional life 	Opinion scale: (6) Strongly agree, (5) Agree, (4) Rather agree, (3) Rather disagree, (2) Disagree, (1) Strongly disagree

The survey was anonymous but participation was tracked. Student responses to the survey were analysed using SPSS 16.0 statistical analysis software. Since the opinion scale represents ordinal data, the non-parametric Kruskal-Wallis one-way analysis of variance by ranked data was used to compare the three methods and the Mann-Whitney U test was applied when 2 samples were compared. This study assumed that a statistically significant difference between rating distributions associated with two different teaching approaches would suggest a difference (positive or negative) in student satisfaction. On the other hand, distributions that did not differ significantly were assumed to suggest similar levels of satisfaction. The analyses compared the student satisfaction towards the teaching method and towards the tutorial format. The different analyses are presented in Figure 3:

- Analysis 1: The comparison of the students' attitude towards two different teaching approaches: Traditional approach (TA) and Supervised Self-Study (SSS).
- Analysis 2a: The comparison of the attitude of bachelor and master students towards the traditional approach;
- Analysis 2b: The comparison of the attitude of bachelor and master students towards the SSS;
- Analysis 3: The comparison of the students' attitude towards the tutorial formats: the software-dependent prescriptive exercises with IDRISI, the software-dependent prescriptive exercises with QGIS/GRASS, and the software-independent conceptual exercises.

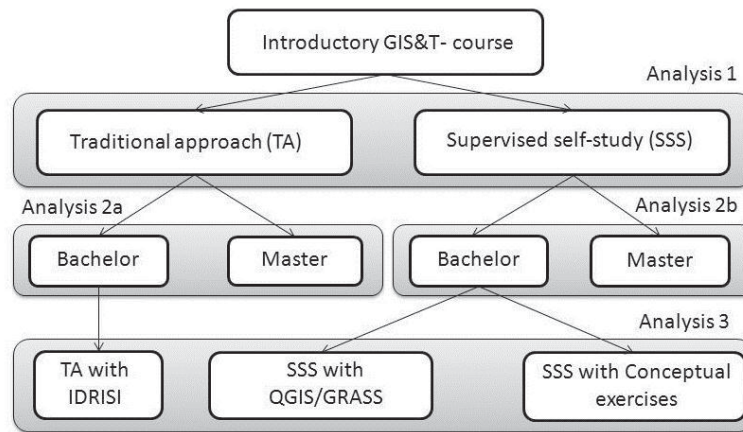


Figure 3. *Structure of the analyses*

3. Results

3.1. *Analysis 1: Traditional approach versus supervised self-study*

A total of 90 students following the traditional approach (TA) and 54 students following SSS completed the survey for the two considered academic years. The answers to the questions regarding the teaching approach are summarised in Table 3. The data indicate no significant difference in the satisfaction concerning the support and the possibility for interaction with peer students. Both the TA- and SSS-students agreed or rather agreed that support, advice and feedback were sufficiently available (82% of the TA and 79% of the SSS) and that regular interaction with peer students was sufficiently possible (82% of the TA and 82% of the SSS). On the other hand,

significant differences were observed regarding the freedom to work independently and the stimulation to deal actively with the learning materials. 73% of the SSS-students indicate that the SSS provides freedom to work independently as opposed to 49% of the TA-students ($p < 0.05$). Furthermore, SSS stimulated to deal more actively with the learning materials (52% of the SSS-students vs. 32% of the TA-students, $p < 0.01$).

Table 3. Results of the comparison of satisfaction with the traditional approach (TA) and with the supervised self-study (SSS)

1. Human support, advice and feedback were sufficiently available						
(%)	Agree	Rather agree	Rather disagree	Disagree	No response	p-value
TA	41	41	8	11		0.497
SSS	52	27	14	7		
2. The teaching approach gave me the freedom to work independently						
(%)	Agree	Rather agree	Rather disagree	Disagree	No response	p-value
TA	49	32	14	3		0.015*
SSS	73	24	2	2	1	
3. There was sufficient possibility for regular interaction with peer students						
(%)	Agree	Rather agree	Rather disagree	Disagree	No response	p-value
TA	54	28	12	6		0.674
SSS	55	27	11	7		
4. The teaching approach stimulated to deal actively with the course materials						
(%)	Agree	Rather agree	Rather disagree	Disagree	No response	p-value
TA	32	41	17	9	1	0.006*
SSS	52	38	5	5		

3.2. Analysis 2a and 2b: Bachelor versus master students

The bachelor and master students were not equally distributed over the three different tutorial formats. Therefore, the difference in satisfaction between bachelor and master students was compared in the second analysis. The two teaching approaches were considered separately to determine whether or not bachelor and master students have different experience.

Analysis 2a compared the bachelor and master students of the TA. 78 Bachelor students and 12 master students responded the questionnaire. The overall results (Table 4) revealed an equal satisfaction for the bachelor and master students. Only a significant difference in appreciation about support was observed. More master students agreed that support, advice and feedback were sufficiently available (75% master students vs. 33% bachelor students, $p < 0.01$). Although the answers of the other questions did not differ significantly, master students answer slightly more positively as opposed to bachelor students.

Table 4. Results of the comparison of satisfaction of bachelor and master students with the traditional approach

1. Human support, advice and feedback were sufficiently available						
(%)	Agree	Rather agree	Rather disagree	Disagree	No response	p-value
Bachelor	33	46	8	13	0	0.006*
Master	75	17	8	0	0	
2. The teaching approach gave me the freedom to work independently						
(%)	Agree	Rather agree	Rather disagree	Disagree	No response	p-value
Bachelor	46	33	17	4	0	0.274
Master	67	25	0	0	8	
3. There was sufficient possibility for regular interaction with peer students						
(%)	Agree	Rather agree	Rather disagree	Disagree	No response	p-value
Bachelor	53	31	10	6	0	0.484
Master	67	8	17	8	0	
4. The teaching approach stimulated to deal actively with the course materials						
(%)	Agree	Rather agree	Rather disagree	Disagree	No response	p-value
Bachelor	32	41	18	9	0	0.884
Master	33	42	17	8	0	
5. The format of the exercises was time efficient						
(%)	Agree	Rather agree	Rather disagree	Disagree	No response	p-value
Bachelor	40	41	14	4	1	0.487
Master	58	17	8	17	0	

6. The software contained too many bugs						
(%)	Agree	Rather agree	Rather disagree	Disagree	No response	p-value
Bachelor	26	24	28	22	0	0.833
Master	17	33	33	8	8	
7. Experience with the software prepared for professional life						
(%)	Agree	Rather agree	Rather disagree	Disagree	No response	p-value
Bachelor	10	45	26	18	1	0.469
Master	8	42	17	33	0	
8. The theoretical background became more clear						
(%)	Agree	Rather agree	Rather disagree	Disagree	No response	p-value
Bachelor	36	40	14	10	0	0.647
Master	42	33	17	8	0	

Analysis 2b considered the questionnaire answers of 32 bachelor and 22 master students who followed the SSS (Table 5). In contrast with the results of analysis 2a, more questions revealed a significant difference between bachelor and master students (5 out of 8 questions). In general, master students were more positive towards the SSS. 68% of the master students versus 41% of the bachelor students agreed that support, advice and feedback were sufficiently available ($p<0.01$). Also, 77% of the master students in comparison with 44% of the bachelor students indicated that sufficient possibility existed to interact with peer students ($p<0.05$). The difference in satisfaction between master and bachelor students even increased with the questions concerning the tutorial format. Two times more master than bachelor students endorsed the time efficiency of the exercises (64% vs. 31%, $p<0.01$) and disagreed that the software contained too many bugs (55% vs. 22%, $p<0.05$). Moreover, 82% of master students experienced that the exercises prepared for professional life as opposed to 50% of the bachelor students ($p<0.01$).

Table 5. Results of the comparison of the satisfaction of bachelor and master students with the Supervised Self-Study approach

1. Human support, advice and feedback were sufficiently available						
(%)	Agree	Rather agree	Rather disagree	Disagree	No response	p-value
Bachelor	41	31	19	9		0.009*
Master	68	23	9	0		

2. The teaching approach gave me the freedom to work independently						
(%)	Agree	Rather agree	Rather disagree	Disagree	No response	p-value
Bachelor	63	25	3	3	6	0.353
Master	77	14	0	0	9	
3. There was sufficient possibility for regular interaction with peer students						
(%)	Agree	Rather agree	Rather disagree	Disagree	No response	p-value
Bachelor	44	31	16	9		0.019*
Master	77	18	5	0		
4. The teaching approach stimulated to deal actively with the course materials						
(%)	Agree	Rather agree	Rather disagree	Disagree	No response	p-value
Bachelor	47	41	9	3		0.162
Master	59	36	0	5		
5. The form of the exercises was time efficient						
(%)	Agree	Rather agree	Rather disagree	Disagree	No response	p-value
Bachelor	31	25	19	25		0.002*
Master	64	32	5	0		
6. The software contained too many bugs						
(%)	Agree	Rather agree	Rather disagree	Disagree	No response	p-value
Bachelor	31	28	19	22		0.032*
Master	14	23	9	55		
7. Experience with the software prepared for professional life						
(%)	Agree	Rather agree	Rather disagree	Disagree	No response	p-value
Bachelor	50	38	3	9		0.006*
Master	82	14	5	0		
8. The theoretical background became more clear						
(%)	Agree	Rather agree	Rather disagree	Disagree	No response	p-value
Bachelor	47	38	3	3	9	0.757
Master	55	27	9	0	5	

3.3. Analysis 3: Comparison between the software-dependent CCSS4G, software-dependent FOSS4G and software-independent conceptual tutorials

Because of the significant difference in appreciation between bachelor and master students towards the SSS and the unequal numbers of bachelor and master students in the sample, the master students were excluded from the last analysis. So, the responses of 78 TA-bachelor students using IDRISI, 25 SSS-bachelor students using QGIS/GRASS and 10 SSS-bachelor students using conceptual exercises were compared concerning the tutorial format. The results are presented in Table 6 and Figure 4. The students using the software-dependent prescriptive IDRISI-tutorial confirmed the time-efficiency of the exercises as opposed to the students solving the QGIS-tutorial (40% vs. 32%, $p < 0.05$). In contrast, twice as much students solving the conceptual tutorial experienced that the exercises format prepared for professional life opposed to IDRISI-students (70% vs. 31%, $p < 0.05$). The results revealed no significant difference between the students using QGIS on the one hand or those selecting an own GIS-software on the other hand. Moreover, also no significant difference regarding the software appreciation and the theoretical background was observed. However, the students taking the software-independent conceptual exercises were slightly more positive. 60% of these students vs. 16% of QGIS-students and 22% of IDRISI-students disagreed that the software contained too many bugs and 60% of the students taking the conceptual exercises vs. 48% of the QGIS-students and 36% of the IDRISI-students agreed that the theoretical background became more clear. The low significance rate can be explained by the small number of students working with the conceptual tutorial.

Table 6. Results of the comparison between the different tutorial formats

	Q1	Q2	Q3	Q4
IDRISI-tutorial vs. QGIS-tutorial	0.041*	0.265	0.172	0.089
IDRISI-tutorial vs. Conceptual tutorial	0.826	0.124	0.019*	0.159
QGIS-tutorial vs. Conceptual tutorial	0.427	0.084	0.106	0.659

Q1 'The format of the tutorial was time efficient'

Q2 'The software contained too many bugs'

Q3 'Experience with the software prepared for professional life'

Q4 'The theoretical background became more clear'

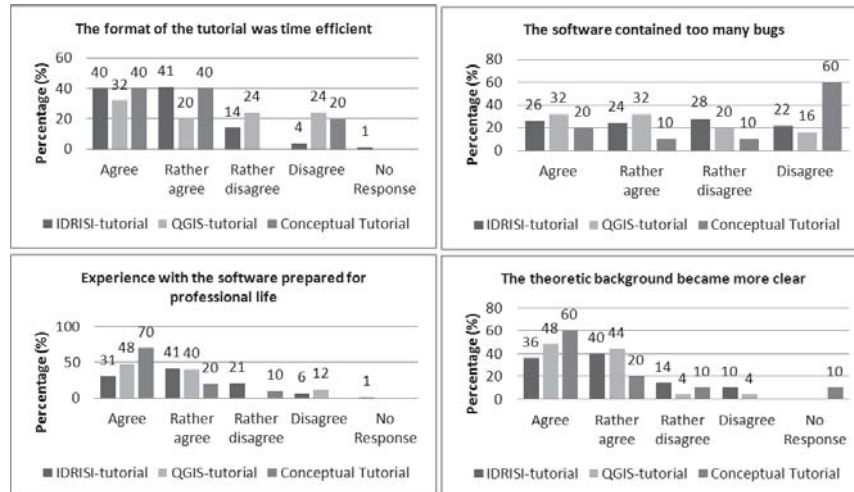


Figure 4. Results of the comparison between the different exercises formats

4. Discussion

The debate about the advantages and disadvantages of educational methods suffers from a lack of data about the student perception and appreciation. In this paper, we aimed to provide data to contribute to this debate by investigating the student satisfaction. The survey focussed simultaneously on two different educational methods (traditional in-class and supervised self-study) and three tutorial formats, i.e. software-dependent with CCSS4G, software-dependent with FOSS4G and software-independent.

The results of the first analysis indicated that both educational methods were equally appreciated by the students. However, the SSS approach was evaluated slightly more positively in comparison with the traditional approach, especially concerning the freedom to work independently and the stimulation to deal actively with the learning materials. Previous surveys (Van Orshoven *et al.*, 2009, Hubeau *et al.*, 2011a) revealed that SSS has the risk to provide insufficient support and to allow insufficient interaction with peer students. In the studied issues of the course, these possible drawbacks seemed to be overcome as no significant difference in satisfaction was observed with the more traditional in-class approach.

The second analysis revealed no clear difference between bachelor and master students regarding their appreciation of the traditional approach. However, the analysis strongly confirmed previous observations (Van Orshoven *et al.*, 2010) that

master students are more enthusiastic about supervised self-study. Master students have a higher maturity and are far more positive about different aspects of SSS such as the time efficiency of the practical work, the interaction with peer students and the human support, advice and feedback. Master students liked to work independently.

The final analysis investigated the difference in appreciation towards the tutorial format. Because the conceptual exercises had been implemented during only one academic year, the sample size was rather small and conclusions are indicative. All the students appreciated the different tutorial formats but the students who took the conceptual exercises were slightly more positive (however not significantly). The students taking the conceptual tutorial felt that they became well prepared for professional life. In addition, this result points out that these students had more confidence about their own competences. Furthermore, it can be assumed that the conceptual tutorial facilitates the acquisition of more in-depth knowledge about the GIS-concepts and the differentiation between different desktop GIS-software. Master students are more comfortable with SSS and both master and bachelor students are aware of professional challenges after studies.

5. Conclusion

FOSS4G and CCSS4G have reached an important level of standardisation. Therefore, conceptual software-dependent tutorials can be developed and implemented in GI S&T-courses. The satisfaction of bachelor students with this new type of tutorials in a SSS-teaching approach was at least as high as with a more traditional type of tutorial in both in-class and SSS-settings. Conceptual exercises fit perfectly within supervised self-study and may significantly enhance the validity duration of tutorial materials. In our opinion, these findings justify the further development, testing and investigation of the applicability, performance and appreciation of conceptual exercises.

6. Bibliography/References

- Abu el Nasr, A., Van Orshoven, J. *CSS or FSS for education in GIS?* Proceedings (L. Bernard, A. Friis-Christensen, H. Pundt and I. Compte, eds.) of the 11th AGILE International Conference on Geographic Information Science 2008, Girona, Spain. CD-ROM: 73_DOC.pdf, http://plone.itc.nl/agile_old/conference/2008-Girona/PDF/73_DOC.pdf.
- Dorner, W., Scheffer, J., Zink, R. 2011. *GvSIG in the academic education of heterogeneous target groups – experiences in lectures, exercises and eLearning*. Proceedings of Geoinformatics, Prague, Czech Republic. ISSN 1802-2669. <http://geoinformatics.fsv.cvut.cz/pdf/geoinformatics-fce-ctu-2011-07.pdf>: 5-10.

- Hubeau, M., De Meyer, A., Verbeeck, K., Heremans, S., Van Orshoven, J. 2011a. *Self-study or in-class-study of geographic information science and technology: the student's perspective*. Proceedings of EDULEARN11 2011, Barcelona, Spain. ISBN 978-84-615-0441-1. <http://library.iated.org/view/HUBEAU2011SEL>: 2872-2881.
- Hubeau, M., Steenberghen, T., Vileikis, O., Van Balen, K., Van Orshoven, J. 2011b. *Software-independent tutorials for supervised self-study of GIS&T*. Proceedings of the LeGIO-workshop 'GIS-education in a changing academic environment, Leuven, Belgium, 2011. ISBN 978-90-8826-217-3. 117-125.
- Hubeau, M., De Meyer, A., Steenberghen, T., Van Balen, K. Van Orshoven, J. 2012. *Software-independent tutorials and their effectiveness for supervised self-study of GIS&T*. Proceedings of EUGISES 'GIS-education: Where are the boundaries?', Leuven, Belgium 2012. ISBN 948-90-8826-258-6.
- Schweik, C., Fernandez, M., Hamel, M., Kashwan, P., Lewis, Q., Stapanoc, 2009. *A. Reflections of an online geographic information system course based on open source software*. Social Science Computer Review 27: 118-129.
- Steiniger, S., Bocher, E. 2009. *An overview on current free and open source desktop GIS developments*. International Journal of Geographical Information Systems 23: 1345-1370.
- Valentine, D. *Distance learning: Promises, problems and possibilities*. 2002. Online Journal of Distance Learning Administration 5 (3). State University of West Georgia, Distance Education Center
- Van Orshoven, J., Wawer, R. and Duytschaever, K. *Effectiveness of a train-the-trainer initiative dealing with free and open source software for geomatics*. Proceedings (J.-H. Haunert, B. Kieler and J. Milde, Eds.) of the 12th AGILE International Conference on Geographic information Science 2009, Hannover, Germany. IKG, Leibnitz Universität, ISBN 2073-8013. CD-ROM: 136.pdf and http://plone.itc.nl/agile_old/conference/2009-hannover/pdfs/136.pdf: 9p.
- Van Orshoven, J., Verbeeck, K., Heremans, S. *Attitude of Undergraduate and Graduate Students Towards Free And Open Source Software For Geomatics In A Supervised Self-Study Context*. Proceedings of the 13th AGILE International Conference on Geographic Information Science 2010, Guimarães, Portugal. ISBN 978-989-20-1953-6. http://agile2010.dsi.uminho.pt/pen/PosterAbstracts_PDF%5C90_DOC.pdf: 2p.

Analyse préalable à une infrastructure de données spatiales au Maroc

Fatiha Ibannain* - Jean-Paul Donnay**

** Agence Nationale de la Conservation Foncière, du Cadastre et de la Cartographie
Avenue Hassan II, Km 4, Rabat, Maroc*

f.ibannain@ancfcc.gov.ma

*** Unité de Géomatique, Université de Liège*

17 Allée du 6 Août (B5a) 4000 Liège, Belgique

jp.donnay@ulg.ac.be

RÉSUMÉ. Dans cet article, nous discutons les moyens de promouvoir les infrastructures de données spatiales (IDS). Outre des considérations d'ordre organisationnel réglementant ces infrastructures, le développement d'outils et de modèles tenant compte de la diversité des intervenants est inévitable pour atteindre les objectifs d'une IDS. Notre étude a ciblé les préalables techniques à la mise en place d'une IDS au Maroc. Une approche ontologique des données de référence a facilité la définition des types d'objets de référence lors du catalogage des entités, puis celle des concepts retenus lors de la modélisation conceptuelle. Un catalogue d'entités conforme à la norme ISO 19110 est élaboré selon les principes de cette ontologie illustrant l'indépendance et la complémentarité des types d'objets, leur adaptabilité à des domaines spécifiques et leur hiérarchisation. Il permet ainsi de mettre en place un modèle de référence, c'est-à-dire une structure de base pouvant être adaptée à une large gamme d'applications. Les développements de ce travail ont pour objectif de sensibiliser les décideurs quant à l'instauration d'une IDS au Maroc.

ABSTRACT. In this paper, we discuss ways of promoting a spatial data infrastructure (SDI). Further considerations regulating the organizational infrastructure, the development of tools and models taking into account the diversity of actors is inevitable to achieve the objectives of an SDI. Our study focused on the technical prerequisites for the establishment of an SDI in Morocco. An ontological approach to the reference spatial data has facilitated the definition of object types when cataloging reference entities, as well the concepts used in conceptual modeling. A feature catalog complying with ISO 19110 is developed using the principles of this ontology illustrating the independence and complementarity of object types, their adaptability to specific application domains and their inheritance. This approach allows establishing a reference model, that is to say a base structure which can be adapted to a wide range of applications. Developments of this work aim to sensitize policy makers about the introduction of an SDI in Morocco.

MOTS-CLÉS: IDS, données spatiales de référence, ontologie, catalogage, modélisation.

KEYWORDS: SDI, spatial reference data, ontology, cataloguing, spatial data design.

1. Introduction

Les infrastructures de données spatiales (IDS) constituent une réponse logique au besoin incessant d'exploiter et de partager efficacement l'Information Géographique. Les premières expériences ont ciblé principalement l'intégration de données qui réclame une bonne compréhension des besoins des divers utilisateurs. Ces pratiques nécessitent au préalable une coordination et une collaboration étroite entre les différents acteurs nationaux pour arriver à un consensus sur les objectifs et les priorités. Ensuite, des accords sur les normes et sur les autorisations nécessaires doivent être établis. Enfin, une attention particulière doit être donnée au développement de modèles et d'outils pour l'exploitation et la gestion de l'information géographique [MOL 98], [DON 02], [NOU 06], [PAR 06], [IBA 11]. Dans ce qui suit, nous reviendrons sur plusieurs de ces aspects tout en mettant l'accent sur les concepts les plus fondamentaux de l'information à référence spatiale [IBA 09]. Il s'agit d'approfondir l'analyse des diverses composantes d'une IDS, et particulièrement celle concernant la mutualisation de l'information qui constitue le point central autour duquel s'articule une panoplie d'idées, de concepts et de ressources. Cet élément incontournable d'une IDS devrait être bien documenté pour en faciliter l'usage et en optimiser l'exploitation. Sa compréhension doit permettre d'identifier les sources de conflits et de concilier les points de vue favorisant ainsi la mise en place d'un modèle de référence, soit une structure de base pouvant être adaptée à une large gamme d'applications. Cet article propose d'exposer la démarche poursuivie pour mettre en place un tel modèle au Maroc. Elle passe par l'établissement d'une ontologie qui s'inspire des pratiques courantes entre producteurs et utilisateurs des données à référence spatiale, tout en se focalisant sur les données de base que nous qualifions de « données de référence ».

2. IDS et politique de géoinformation ?

2.1. Les enjeux de l'information géographique

La représentation du monde réel tente de communiquer une vision qui reproduit autant que possible la réalité. Cette représentation reste liée à l'objectif du produit final. Selon qu'il s'agisse d'une simple cartographie ou de la constitution d'une base de données géographiques regroupant toutes les composantes environnementales du phénomène étudié, plusieurs facteurs sont à considérer. Dans le cas d'une représentation cartographique classique, malgré son utilité et sa richesse sémantique, son utilisation reste peu adaptée à une exploitation s'intéressant au contenu de l'information plutôt qu'à sa reproduction, elle-même entachée des opérations de symbolisation et de généralisation. Dans le même ordre d'idées, certains auteurs qualifient la carte *« d'une vue ou d'une représentation particulière d'un espace géographique adaptée à une utilisation précise, avec des informations différentes »*

[DEV 97, p. 15]. L'utilisateur se retrouve ainsi limité dans ses manœuvres par les spécifications même de la carte.

En d'autres termes, la représentation cartographique ne peut plus constituer le support unique de l'information géographique. Elle devient une application, parmi d'autres, issue d'une base de données géographiques. Il faut voir là une optimisation des potentialités de l'information géographique, dans ses dimensions tant spatiales que thématiques. Mais comment constituer la base de données géographiques en amont ? La démarche implique naturellement l'intégration de données se rapportant à diverses thématiques mettant en interaction autant de catégories d'utilisateurs. D'une part, on identifie des organisations responsables de la production des données géographiques de référence, et d'autre part celles qui se dotent de ces informations pour constituer leurs propres systèmes d'information géographique en y greffant leurs données métiers spécifiques. Ces dernières institutions se placent en tant qu'utilisateurs vis-à-vis du producteur de données géographiques de référence, appelées aussi « référentiels » [BAD 00]. Cette situation requiert la définition des responsabilités de chacune des parties.

D'une manière générale, l'information géographique est une ressource indispensable pour tout pays moderne. Cependant, plusieurs facteurs tant d'ordre technique que juridique forment une barrière à sa diffusion auprès des principaux acteurs économiques et, en conséquence, ne favorisent pas la concertation entre la communauté scientifique, les décideurs et la société civile [GAZ 06]. C'est la raison pour laquelle, partout dans le monde, des acteurs nationaux conscients de l'aspect stratégique que revêt l'information géographique, ont initié des démarches de rapprochement pour instaurer un cadre favorable à l'optimisation de l'exploitation de l'information à référence spatiale. Ces démarches ont pu profiter des avancées technologiques apparues dans le domaine de l'information géographique, au premier rang desquelles il faut mentionner le développement des normes, des formalismes et des modèles.

2.2. L'information géographique au Maroc

L'utilisation de l'information à référence spatiale au Maroc s'est étendue à des domaines de plus en plus variés. Après une aire de cartographie conventionnelle dont l'unique produit était la carte papier, les outils et les méthodes de représentation de l'espace géographique ont évolué.

Dans ce sens, la cartographie se trouve aujourd'hui confrontée à de nouveaux défis liés principalement aux avancées technologiques dans le domaine de la géomatique. À côté des logiciels et matériels devenus accessibles et des méthodes d'acquisition des données géographiques de plus en plus nombreuses, l'utilisation d'Internet comme vecteur de diffusion a contribué à répandre encore plus l'information géographique.

Une volonté de se conformer aux directives internationales en matière de géomatique a été exprimée au Maroc dès la fin des années 90 par l'informatisation de certains maillons de la chaîne de production cartographique qui relève des missions de l'agence nationale de la conservation foncière, du cadastre et de la cartographie (ANCFCC). Ces efforts se sont poursuivis par l'adoption de l'approche « tout numérique » pour l'optimisation du processus de production [ANC 04].

2.2.1. Contexte de production des données de base

Des procédures ont été mises en place pour structurer les données cartographiques selon les possibilités offertes par des logiciels de cartographie assistée par ordinateur. Il s'agit d'attribuer à chaque objet des valeurs d'attributs dans des fichiers au format DGN1. Ces valeurs correspondent au type d'implantation ponctuelle, linéaire ou zonale (type), au niveau (LV), à la couleur (CO), au style de trait (LC), et à l'épaisseur de trait (WT) attribués à ce détail. La représentation cartographique est confiée à des applications spécialisées qui exploitent ces fichiers. Ce changement de support a offert de nouvelles possibilités de réutilisation des données cartographiées. Mais il n'a pas réellement facilité la tâche des utilisateurs. Ils restent limités par les contraintes de production des données ayant servi à la réalisation des cartes, dont l'interprétation nécessite le recours systématique à la structuration en vigueur à l'entité de production afin d'associer à chaque combinaison (type, LV, CO, LC, WT) le détail qui lui correspond. Par ailleurs, l'avancée technologique s'est accompagnée du lancement de projets SIG par différents organismes nationaux. Après avoir constitué, pendant longtemps, un outil complexe que seuls les spécialistes pouvaient maîtriser, les nouvelles générations des logiciels SIG ont contribué à élargir son public. De nouveaux besoins sont apparus, notamment chez les utilisateurs sectoriels, qui, pour mettre en place leurs propres systèmes d'information géographique, réclament des données géographiques de référence. Cependant, l'exploitation de ces données dans un environnement SIG nécessite certains traitements ou reformatages (re-engineering de données) de la part du producteur, dont la justification demeure discutable selon l'ampleur et la finalité du projet considéré.

2.2.2. Besoins en information de base

Rares sont les enquêtes officielles portant sur les capacités d'un SIG au niveau national. Les SIG se retrouvent dans les grandes institutions et organisations, gouvernementales ou non, mais également à différents niveaux hiérarchiques, selon les découpages administratifs (SIG nationaux, régionaux et locaux). Ce phénomène entraîne une forte dispersion et une hétérogénéité des données géographiques et cartographiques [POI 07]. Dans le même ordre d'idées, certains utilisateurs mettent l'accent sur le manque de cartes de base récentes pouvant refléter le rythme accéléré des changements du tissu urbain [DSM 07], [DSN 07].

1 Format spécifique au logiciel DAO MicroStation de la compagnie Bentley Systems.

La multiplicité des besoins engendre une augmentation de la fréquence des mises à jour des données de base, mais surtout, elle réclame la documentation de l'information de référence (métadonnées) afin d'évaluer son degré de correspondance par rapport à la réalité. Les producteurs de données, tant aux niveaux national que régional, optent pour la définition et la mise en œuvre de spécifications de qualité pour satisfaire les exigences des utilisateurs (clients) et pour améliorer l'efficacité du système de gestion de la qualité [ISO 00], [IGN 01], [GIN 03], [IGN 03], [NSH 04], [SHO 04].

Parmi les incohérences mises en lumière par les analyses de qualité, on constate que de nombreuses d'entre-elles sont liées à la nature même des éléments cartographiés. C'est le cas, par exemple, du découpage administratif, délimitation abstraite constituant l'une des unités spatiales d'intégration (USI) les plus utilisées. Plusieurs acteurs économiques utilisateurs de cette donnée, relatent les difficultés que pose son utilisation, particulièrement en présence des changements de découpages nécessitant des mises à jour [DSM 07], [DSN 07].

D'une manière générale, ce sont les producteurs des données de référence qui sont tenus d'en livrer les mises à jour afin que les utilisateurs puissent disposer d'une vue de l'espace géographique qui reflète autant que possible la réalité du terrain. Le même souci se pose aussi aux producteurs de données dérivées, lors de la réalisation de leurs propres produits. La propagation des évolutions spatio-temporelles du référentiel dans les bases de données géographiques des utilisateurs ou dérivées, ne va cependant pas sans poser des risques importants de perte d'information ou conduire à l'obtention d'états incohérents dans les bases de données [BAD 00].

L'élaboration et la mise en œuvre des méthodologies de mise à jour régulière ou permanente de ces données de référence et dérivées constituent donc un préalable à leur exploitation par les différents utilisateurs. L'expérience de la division de statistique des Nations Unies (DSNU) a montré que le développement et la maintenance de bases de données géographiques sont coûteux et très consommateurs de moyens humains. Il est, par conséquent, nécessaire de minimiser les redondances en encourageant le partage de données de base [DSN 07]. Cependant, une telle action réclame le développement des outils nécessaires pour accompagner les producteurs et les utilisateurs dans leurs missions à l'échelle nationale.

2.2.3. *Peut-on envisager une IDS au Maroc ?*

L'analyse de la situation actuelle a permis d'identifier de nouveaux besoins en termes d'administration de l'information géographique. En d'autres termes, l'attention doit être orientée vers l'instauration des dispositions nécessaires pour une possible mise en commun de l'information géographique, susceptible de conduire à la mise en place d'une infrastructure nationale de données. Mais la question qui se pose est : dispose-t-on au Maroc des composantes principales d'une IDS telles qu'elles sont définies dans la littérature ? [GIN 02], [ECA 04], [SAL 06].

À l'exception du patrimoine national, d'énormes efforts sont encore à déployer quant à la documentation des informations géo-référencées. La notion de métadonnées n'a pas encore atteint son degré de maturité dans de nombreuses organisations. D'autre part, les mécanismes de découverte et d'exploration des données doivent être mis en place comme préalables à l'accès proprement dit aux données géographiques. À côté de ces volets qui restent à développer, on doit aussi noter le faible niveau d'avancement des accords techniques et organisationnels effectifs à l'échelle nationale.

Par ailleurs, la réussite de la mise en place d'une IDS est conditionnée par son développement, son utilisation et son entretien par un ensemble d'organismes responsables des ressources de base [GIN 02], [SAL 06]. Or au Maroc, les principaux producteurs de l'information géographique de base relèvent d'un même organisme (ANCFCC), ce qui constitue une assise prometteuse pour faciliter la mise en commun des données géographiques à l'échelle nationale. Il reste que l'administration des données reste le cœur du problème, nécessitant des préalables techniques pour soutenir une telle démarche.

Il s'agit tout d'abord de mettre au point un cadre pour accompagner les différentes catégories d'utilisateurs des données géo-spatiales à l'échelle nationale. Outre les aspects techniques, qui se traduisent principalement par l'adoption des normes, la prise en considération de l'aspect juridique s'avère indispensable dans la perspective d'arriver à un consensus sur les conditions de production et d'utilisation des données géographiques.

Néanmoins, face à la diversité des intervenants et afin de devancer les sources de conflits, la mutualisation des connaissances liées aux différentes composantes d'une IDS reste le principal défi. Comme le souligne [DAV 09], les insuffisances sémantiques constituent un facteur limitant l'interaction des divers acteurs. Leurs points de vue nécessitent des moyens plus efficaces pour passer au travers des ensembles de concepts, ou ontologies, caractérisant chaque communauté.

3. Modélisation et ontologies des données géographiques de référence

La modélisation est une représentation de la réalité selon des choix répondant à un objectif défini. On peut en conséquence parler d'une synthèse de l'information dans le sens où elle implique une simplification de la complexité de la réalité basée sur des hypothèses définies au départ de la modélisation. Il s'agit d'un ensemble de règles dictées par notre perception et guidées par nos connaissances qui constitueront les spécifications qui définissent un modèle donné. De nombreuses méthodes de conception adaptées aux données géographiques ont été proposées. D'autres travaux sur les modèles de représentation des données spatiales continuent à susciter l'intérêt et concernent notamment la normalisation et la gestion des métadonnées.

La représentation des éléments du monde réel dans une base de données géographiques nécessite une définition sémantique claire et précise. Même si la bonne communication repose avant tout sur des formalismes pouvant être aisément interprétés par les intéressés, l'espace géographique reste suffisamment complexe pour que sa formalisation en termes d'objets et d'interactions reste difficile. C'est dans ce sens qu'un dictionnaire de données s'avère indispensable pour enrichir la description des phénomènes du monde réel. Il constitue en fait une partie de l'ensemble des métadonnées et il a pour avantage d'être particulièrement flexible et accessible à l'utilisateur final [PAN 98].

L'élaboration d'un dictionnaire de données dans le cadre de ce travail a débuté par la définition du mot « référence » qui correspond à une information ou un élément servant de repère. Dans le domaine de l'information à référence spatiale, cette définition reste applicable. Le terme référentiel est a priori associé à l'infrastructure géodésique. Plusieurs mutations dans le monde de la géomatique ont fait évoluer son sens et on parle également de « référentiel géographique ». Il s'agit d'un ensemble de données géographiques de base qui constituent le centre d'intérêt d'une large communauté de producteurs et d'utilisateurs potentiels de l'information géographique. Ce sont ces données d'intérêt commun qui sont qualifiées de données de référence.

L'identification et la définition des données de référence au Maroc ont constitué le premier pas vers la modélisation envisagée dans cette recherche. Au travers d'une étude de l'existant et des besoins, il s'agissait tout d'abord, d'identifier les producteurs et leurs données géographiques, mais aussi d'imaginer une exploitation optimale de ces données de référence. Nous avons ainsi confronté les objectifs des producteurs et ceux des utilisateurs des données, et analysé en parallèle les configurations des données existantes. Ces analyses ont permis de mettre en avant les principaux problèmes liés à l'utilisation des données, en ce compris par le producteur initial lorsqu'il souhaite en dériver d'autres produits. Les incohérences sémantiques sont ainsi apparues comme l'un des obstacles majeurs à l'interopérabilité des données. L'absence de vocabulaire commun, les divergences voire l'insuffisante compréhension des concepts et de leurs relations, en un mot l'absence d'une ontologie de domaine et de métadonnées explicites, constituent un frein à la mise en commun de l'information et, d'une manière générale, à l'établissement d'une IDS [GES 05]. Pour apporter un premier élément de réponse à cette lacune, nous avons procédé au catalogage des données géographiques en se conformant à la norme ISO 19110 [ISO 05], [IBA 11].

3.1. Catalogage des entités ou ontologie des données de référence ?

L'élaboration d'un catalogue d'entités géographiques concrétise notre analyse dans le sens où elle permet dans un premier temps la description des données existantes de manière structurée. Il s'agit de définir les objets et les attributs que

devront posséder les données inventoriées. En tenant compte des producteurs de ces données et des exigences d'une large gamme d'utilisateurs de données de référence, nous avons essayé de définir les attributs d'un ensemble de classes génériques abstraites auxquelles ces objets appartiennent et dont les valeurs constituent des métadonnées sur ces objets. Nous avons ainsi défini onze types d'objets repris au tableau 1.

Tableau 1. *Descriptif des types d'objets de référence [IBA 09]*

Types d'objets	Description
REFERENTIEL GEODESIQUE	Abstraction de l'ensemble des éléments du référentiel géodésique.
LIMITES	Abstraction des limites politiques, administratives, de parcellaire, de zoning, et toute autre délimitation.
RELIEF	Abstraction de l'ensemble des objets qui permettent de décrire la configuration de la surface terrestre telle qu'elle est déterminée par les inégalités de cette surface en se basant sur les formes de relief.
HYDROGRAPHIE	Abstraction de l'ensemble des objets qui décrivent l'organisation du réseau hydrographique et des étendues d'eau commandée par l'agencement du relief, la disposition et la topographie des masses solides qui contiennent les eaux marines ainsi que tous les ouvrages hydrauliques liés à l'intervention humaine.
TRANSPORT TERRESTRE	Abstraction de l'ensemble des objets constitutifs des réseaux routier, de pistes et des lignes téléphériques.
RESEAU FERROVIAIRE	Abstraction de l'ensemble des objets constitutifs du réseau ferroviaire.
STRUCTURES	Abstraction des constructions et structures permanentes telles les bâtiments et les constructions en relation avec les divers réseaux et les ouvrages d'équipement.
INSTALLATIONS AERONAUTIQUES	Abstraction des objets conçus et réservés pour l'atterrissage et le décollage d'aéronefs, y compris les installations d'entreposage et d'entretien des aéronefs ainsi que pour les passagers et le fret.
INSTALLATIONS PORTUAIRES	Abstraction des eaux navigables et des ouvrages portuaires.
EQUIPEMENT	Abstraction des objets qui relèvent du domaine des équipements électrique, de communication, industriel, minier...
VEGETATION	Abstraction de toutes les entités qui composent le couvert végétal ainsi que les surfaces couvertes de végétation.

3.2. Principes d'une ontologie pour les données de référence

La démarche qui consiste à définir des types d'objets de référence lors du catalogage, en s'inspirant des pratiques courantes entre producteurs et utilisateurs des données à référence spatiale, peut dans une certaine mesure être assimilée à l'élaboration d'une ontologie [LAU 07]. Pour ce faire, nous avons développé une méthodologie s'inspirant des matrices de conduite de projet de SIG [PAN 96] où

l'adaptabilité à une application particulière, de toute entité que nous qualifions d'objet de référence, est tracée au travers d'un ensemble de critères de pertinence. Parmi les classes de préoccupation de cette matrice (en colonne – voir tableau 2), figure en première place l'organisation, commandant la définition des données de référence. L'objectif qui motive la création de ces données de référence, ainsi que l'utilisation qui leur est réservée sont ensuite décrits. Le type d'objet de référence qu'il est possible d'en inférer lors de la phase de catalogage et l'évaluation de la cohérence de la proposition forment les deux dernières classes de préoccupation. Ces classes de préoccupation sont analysées successivement au travers des critères énoncés par Gruber pour évaluer et mettre en évidence les bases d'une ontologie : clarté, extensibilité et engagement ontologique minimal [GRU 93]

Tableau 2. Mise en évidence des principes d'une ontologie des données de référence [IBA 09]

	Organisation	Objectif	Utilisation	Types d'objets de référence	Cohérence
Clarté	Identification des données de référence et des acteurs	Objectif et possibilités d'utilisation	Typologie des domaines d'applications	Premier dictionnaire	Principe 1 : Indépendance / complémentarité
Extensibilité	Principe 3 : Hiérarchisation	Principe 2 : Adaptabilité			Héritage multiple
Engagement minimal	Contraintes	Chevauchement de domaines			Responsabilités de production

3.2.1. Indépendance et complémentarité des types d'objets

L'analyse du tableau 2 permet d'exprimer les principes de cette ontologie en considérant les diverses classes de préoccupation proposées. Ainsi, la première classe, l'organisation, reflète le contexte général de production et d'utilisation de l'information de référence. À un niveau descriptif exprimant le critère de clarté, notre démarche est concrétisée par l'identification des données de référence au Maroc, ainsi que les divers acteurs impliqués. Il s'agit de mettre en évidence les ensembles de données qui sont d'un intérêt commun à un maximum d'utilisateurs pour cerner le premier champ d'investigations.

En combinant ce premier élément d'analyse aux deux classes suivantes, à savoir l'objectif et l'utilisation des données de référence, on est en mesure de dresser une première typologie des pratiques courantes des producteurs et des utilisateurs de ces données. D'une part, les objectifs qui ont motivé la production de ces données, ainsi que les possibilités d'utilisation, constituent un paramètre qui renseigne sur le contenu et le sens de l'information. D'autre part, le degré d'adaptabilité de ces données est partiellement documenté à travers les applications pour lesquelles ces données ont été employées jusqu'ici. Ces considérations sont à la base de la définition des types d'objets de référence qui formeront le catalogue des entités. Ce dernier est ainsi construit selon une démarche inductive procédant par synthèse [CHA

07] : la définition des types d'objet au niveau générique du catalogue s'est basée sur la prise en compte de l'ensemble de données qui peuvent relever de chaque type d'objet.

Quant aux contrôles de cohérence qui font l'objet de la dernière classe considérée, ils traduisent le premier principe ontologique des données de référence, à savoir l'indépendance et la complémentarité des types d'objets de référence. À un niveau global, les contrôles de cohérence permettent d'éviter les chevauchements entre les domaines couverts par les divers types d'objets. À un niveau plus détaillé, la cohérence peut concerner des entités qui relèvent de plusieurs types d'objets, en faisant appel à la notion d'héritage multiple qui souligne la complémentarité des types d'objets autorisant ainsi le chevauchement de leurs domaines.

3.2.2. *Adaptabilité des types d'objets à des domaines spécifiques*

L'analyse à un niveau descriptif traduisant le critère de clarté a permis d'accumuler un ensemble de connaissances sur les données de référence dont un premier résultat concret est le dictionnaire de données. La prise en compte de l'objectif initial de création des jeux de données, ainsi que leurs utilisations antérieures, ont permis d'identifier des propriétés que doit posséder un objet en vue de répondre aux besoins d'un domaine spécifique. La différenciation des lots de données peut dès lors être faite sans ambiguïté, en permettant la constitution des types d'objets de référence offrant une adaptabilité à un maximum d'applications. En d'autres termes, les objets qualifiés de référence sont définis d'une manière complète, indépendamment de leur contexte. Leur exploitation dans une application particulière doit se faire de manière fonctionnelle tout en assurant des extensions propres au domaine concerné.

3.2.3. *Hiérarchisation*

Comme nous venons de le présenter, le catalogage des données a été réalisé dans l'esprit d'assurer l'adaptabilité des objets de référence à divers domaines d'application. Les types d'objets peuvent ainsi être introduits dans autant de modèles conceptuels spécifiques. Le modèle conceptuel de référence consiste lui en une hiérarchisation qui se concrétise par des relations de spécialisation liant divers types d'objets spécifiques à chaque type d'objet de référence. À un niveau de détail plus fin, des relations de spécialisation peuvent aussi lier certaines classes d'entités à d'autres classes d'entités de référence.

Le tableau 3 illustre ces principes par l'exemple des structures bâties figurant dans la légende des cartes de base selon diverses modalités. Le catalogage des entités a permis d'identifier le type d'objet « Structures » qui concerne tout édifice conçu pour un usage donné. C'est ainsi qu'on a défini deux catégories de constructions de type d'objet « Structures » qui se retrouvent dans plusieurs modèles et ne se spécialisent qu'à un niveau de détail assez élevé exigeant la présence explicite de ces sous-classes lors du processus de mise en relation dans ces modèles.

Tableau 3. Exemple de mise en évidence des principes d'une ontologie des données de référence

	Organisation	Objectif	Utilisation	Types d'objets de référence	Cohérence
Clarté <i>Niveau descriptif</i>	Les constructions dans les cartes de base existantes: <ul style="list-style-type: none"> ▪ Groupées remplies, ▪ Isolées remplies ▪ Légères (Isolées non remplies) ▪ Bidonville ▪ Construction en dur (terre) ▪ Bâtiment remarquable 	Production des cartes de base	Généralisation cartographique Cartographie thématique	<ul style="list-style-type: none"> ▪ Type d'objets : Structures Définitions textuelles : <ul style="list-style-type: none"> ▪ classes d'entités, ▪ attributs ▪ relations 	Principe 1 : Indépendance / complémentarité <ul style="list-style-type: none"> ▪ Une classe d'entités est définie dans un seul type d'objet ▪ L'import d'une classe d'entités d'un modèle spécifique à l'autre assure la complémentarité des types d'objets
Extensibilité <i>Niveau conceptuel et organisationnel</i>	Principe 3 : Hiérarchisation	Principe 2 : Adaptabilité			Évolution de la classe d'entités BATIMENT (type d'objet Structures) à SURFACE_BÂTIE (type d'objet Limites) ⇔ Généralisation conceptuelle
	BATIMENT	Grande échelle	<ul style="list-style-type: none"> ▪ Cadastre ▪ Cartes urbaines 	BÂTIMENT est du type d'objet Structures	
	SURFACE BATIE	Petites échelles	<ul style="list-style-type: none"> ▪ Cartes de base 	SURFACE_BATIE est du type d'objet Limites	
Engagement minimal <i>Niveau logico-physique</i>	Contraintes et mécanismes d'évolution de la classe ou sa modélisation spatiale	Chevauchement de domaines à des niveaux de détail plus fin CONSTRUCTION_PARTICULIERE se spécialise en STRUCTURE_RESEAU qui se spécialise à son tour en OUVRAGE_DE_TRAVERSEE qui peut se spécialiser en PONT ou TUNNEL			Définir les responsabilités de production des données de référence pour faciliter les mises à jour.

D'une part, la classe d'entités BATIMENT désigne toute construction permanente, murée et couverte. En plus de sa description par l'attribut d'usage, qui renseigne sur l'activité qui lui est dévolue, la considération des caractéristiques géométriques et dimensionnelles de l'objet BATIMENT a permis de mettre en évidence la sous-classe d'entités BATIMENT_REMARQUABLE. D'autre part, on a défini la classe d'entités CONSTRUCTION_PARTICULIERE comme généralisation des constructions dépendantes des divers réseaux ou des ouvrages d'équipement.

En conclusion, sans prétendre avoir mené une analyse ontologique de données de référence, nous avons essayé d'en exprimer les principes à travers les critères qui ont conduit notre démarche de catalogage. Une réflexion plus approfondie sur l'ontologie des données de référence peut ensuite être envisagée en tenant compte des ontologies de domaine existantes.

3.3. Modèle de référence / Schéma intégré d'une IDS nationale

3.3.1. Modélisation conceptuelle des données de référence

L'élaboration de modèles conceptuels de données consiste à décrire l'information structurée selon un schéma cohérent en spécifiant les contraintes et les conditions à respecter. Le catalogage des données nous a permis de distinguer onze types d'objets. Cette classification s'est basée principalement sur la caractérisation des classes d'entités en termes de traitements ultérieurs possibles. Les relations qu'elles entretiennent entre elles ont permis de mettre en commun des sous-domaines d'interaction et de consolider les différents modèles conceptuels élaborés autour des types d'objets répertoriés. Le recours à l'expertise internationale en termes de modèles de référence existants [IGN 02], [IGN 03], [OS 06] a permis d'améliorer la perception et la définition de certaines classes d'entités. L'utilisation de Web2GIS [LAP 06], l'environnement de conception de bases de données spatiales sur Internet, a simplifié l'élaboration des différents modèles en important les classes d'entités du catalogue des objets de référence réalisé. Il suffit d'associer leur spatialité aux types d'objets et à les regrouper au sein de paquetages en veillant à compléter par de nouvelles associations en cas d'importation d'une classe d'entités à partir d'un autre modèle.

Nous présentons dans ce qui suit un exemple d'application qui reprend le modèle spécifique du type d'objet « Transport terrestre » qui comporte le paquetage des pistes. Le modèle de référence global est ensuite enrichi par (1) l'import des classes d'entités relatives aux parcelles cadastrales, aux sections de chemin de fer et à la végétation zonale et (2) l'ajout de nouvelles associations pour témoigner des relations qu'elles entretiennent avec la classe d'entités des sections de pistes. Pour ce faire, nous avons utilisé les options d'import et d'ajout de liens proposées par le module de modélisation Web2GIS (Figure 1).

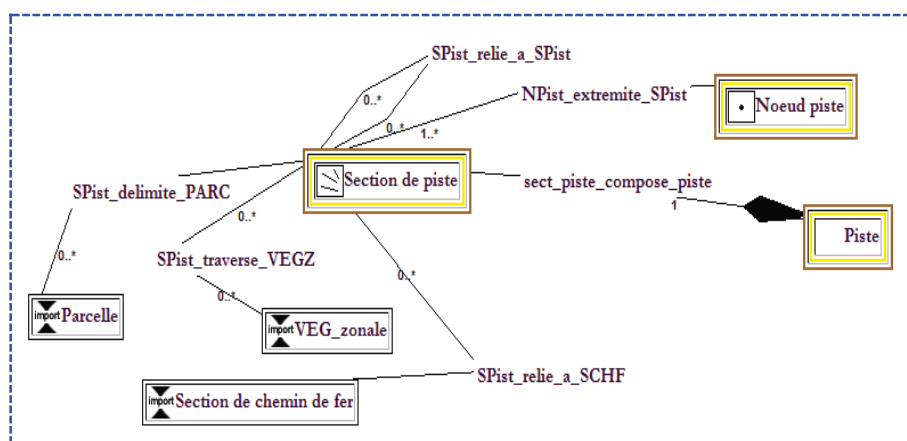


Figure 1. Extrait du modèle du type d'objet Transport terrestre enrichi [IBA 09]

Nous avons donc retenu les diverses classes d'entités qui relèvent du paquetage des pistes, appartenant au type d'objet « Transport terrestre », ainsi que les relations présentes entre ces classes d'entités (Figure 3). Comme la modélisation des pistes requiert la prise en compte des contraintes d'intégrité de réseau pour assurer la connectivité du réseau des pistes, nous avons défini la classe d'entités « Section piste » comme une division de l'axe de piste, délimitée par deux « nœuds de piste » dès qu'il y'a un changement de la valeur d'un attribut de la piste, ainsi qu'à l'intersection avec une autre piste ou un élément d'un autre réseau. La classe d'entités « Section piste » est reliée à la classe d'entités « Piste » par une relation de composition². Cependant, un enrichissement sémantique à travers des attributs permet de mieux caractériser la piste. Les classes d'entités du paquetage pistes sont reliées au type d'objet « Transport terrestre » par une relation de composition libellée « est de type d'objet ». Par ailleurs, l'importation des classes d'entités issues d'autres modèles spécifiques, doit prendre en compte l'implémentation des classes d'entités dont elles dépendent. Ainsi, la parcelle cadastrale importée du modèle « Limites » concrétise l'interaction entre le paquetage parcellaire et le réseau des pistes à travers la relation de délimitation des parcelles par des pistes. Comme la parcelle cadastrale entre dans la composition de la propriété foncière, qui, elle-même, est répertoriée sur une mappe cadastrale, c'est tout le paquetage parcellaire qui est considéré en faisant appel à la parcelle cadastrale. Il en est de même pour la classe d'entités des sections de chemin de fer, importée du modèle « Réseau ferroviaire » où elle forme, par composition, la classe d'entités « Chemin de fer ». La relation entre les sections de piste et les sections de chemin de fer invoque ainsi la totalité du paquetage du réseau ferroviaire. Par contre, la classe d'entités « Végétation zonale », importée du modèle « Végétation », n'a pas nécessité la prise en considération de classes d'entités additionnelles (Figure 2).

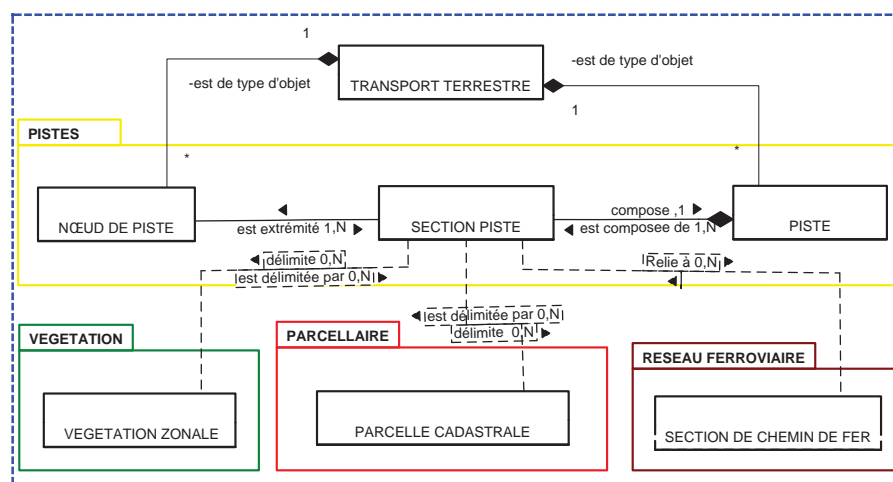


Figure 2. Exemple d'interactions entre les modèles spécifiques

² Les relations sont illustrées selon les notations du formalisme UML

Fiche de la classe d'entités SECTION_PISTE

Descriptif de la classe d'entités SECTION_PISTE du modèle Transport Terrestre

Classes d'entités	Définition	Attributs		Sous classes d'entités
SECTION_PISTE	Axe de la plus grande division d'une piste	désignation_piste	PISTE_CARROSSABLE, PISTE_CYCLABLE	
		largeur_piste		
		consistance_piste		

Descriptif des attributs de la classe d'entités SECTION_PISTE

Attribut	Définition	Type de données des valeurs	Unité de mesure des Valeurs	Type du Domaine des Valeurs	Domaine des valeurs
désignation_piste	Utilisation ou fonction assurée par la piste	Char(25)	-	Valeurs précodées	PISTE_CARROSSABLE, PISTE_CYCLABLE CHEMIN_EXPLOITATION LAIE_FORESTIERE, COUPE_FEU, RUELLE, SENTIER /LAYON, ESCALIERS
largeur_piste	Largeur de piste	Real	mètre	-	-
consistance_piste	Composition de la surface de la piste	Char(25)	-	Valeurs précodées	EMPIERREE EN_TERRE, EN_DUR, NON_AMENAGEE, INCONNU

transport_terrestre.section_de_piste

Field	Type	Length	Lengthvar	NotNull
largeur_de_piste	varchar	-1	14	f
consistance_de_piste	varchar	-1	29	f
designation_piste	varchar	-1	29	f
section_de_piste_id	int4	4	-1	t
the_geom	geometry	-1	-1	f
sect_piste_compose_piste_piste_id	int4	4	-1	t

transport_terrestre.piste

Field	Type	Length	Lengthvar	NotNull
piste_id	int4	4	-1	t
piste_est_de_type_d_objet	int4	4	-1	t
transport_terrestre_type_d_objet_de_reference_id	int4	4	-1	t

Lien entre les tables relatives aux pistes et aux sections de pistes

Figure 3. Fiche du catalogue normalisé relative à la classe des sections de piste

3.3.2. Exploitation du modèle de référence

L'exploitation d'un modèle unique des objets de référence constitue un atout considérable pour le partage des données en palliant aux problèmes d'hétérogénéité (sémantique, schématique et syntaxique) qui entravent l'intégration des bases de données géographiques issues de modèles et de disciplines différentes [BIS 98], [HAK 03]. Mais une autre source d'hétérogénéité peut apparaître entre différents niveaux de détail au sein d'un même domaine. C'est le cas, par exemple, des agences cartographiques nationales qui disposent de plusieurs catalogues d'entités correspondant chacun à une échelle donnée, engendrant ainsi de grands écarts entre le nombre de concepts retenus pour décrire l'espace géographique d'une échelle à une autre [SHE 07]. Pour résoudre ces problèmes d'hétérogénéité, certains ont proposé le recours à la cartographie pour la génération automatique d'une ontologie, en analysant le mécanisme qui, « cartographiquement », relie des éléments hétérogènes, et ce au départ du catalogue relatif à l'échelle 1/25 000^e [GOM 08]. À côté de cette démarche qui vise à la création d'un cadre d'intégration pour maintenir les diverses bases de données existantes comme alternative au manque d'outils de généralisation « complètement » automatique, d'autres auteurs proposent un modèle de bases de données topographiques intégrées, enrichi sémantiquement pour soutenir une généralisation « semi-automatique », en ajoutant des modèles de transition d'une échelle à une autre [STO 08].

Dans le cas de la présente recherche, le modèle de référence développé peut être exploité par différentes catégories d'utilisateurs et à différents niveaux de détail. Selon le point de vue de l'utilisateur, les modèles des types d'objets peuvent être implémentés sous leur forme actuelle ou étendus par la considération des ontologies de domaine existantes. À ce niveau, il convient de rappeler les objectifs de ce travail, en particulier les besoins constatés en matière d'outils de modélisation et de gestion de l'information géographique à l'échelle nationale. La modélisation des données de référence au Maroc doit être perçue comme un préalable à la mise en place d'une infrastructure nationale de données spatiales. En ce sens, le modèle de référence constitue un champ d'intérêt commun à diverses disciplines. Son adoption signifie la constitution de bases de données issues d'un même modèle conceptuel de données (MCD). Cette approche permet de pallier aux quatre sources d'hétérogénéité relevées par [HAK 03] et citées par [GOM 08] : (1) la modélisation conceptuelle, (2) la sémantique des objets, (3) la structure des bases de données et enfin (4) la modélisation spatiale des entités.

D'une part, les divers MCD développés sont issus d'un catalogue d'entités conforme à la norme ISO 19110. Ce dernier est élaboré selon les principes d'une ontologie de données de référence, motivés principalement par les points de vue des producteurs et utilisateurs des données, les configurations des données existantes et leurs domaines d'application antérieurs ; ce qui limite les incohérences sémantiques. D'autre part, le modèle de référence est conçu d'une manière globale dans le but de consolider les divers modèles spécifiques de types d'objets autour du modèle de type

d'objet de référence, permettant ainsi de constituer des bases de données qui partagent le même schéma.

En ce qui concerne les problèmes d'hétérogénéité susceptibles d'être engendrés par les diverses modélisations spatiales d'un même objet, des mécanismes ont été envisagés dès la modélisation conceptuelle. Ainsi, le principe d'adaptabilité, soutenu par diverses formes d'enrichissement du modèle de référence, rejoint les mécanismes de transitions prévus par [STO 08] pour une généralisation semi-automatique. Il s'agit de la notion d'héritage multiple entre les classes d'entités qui relèvent de types d'objets différents, ainsi que les possibilités d'évolution de certaines classes d'entités ou de leur modélisation spatiale.

4. Conclusions et perspectives

La modélisation des données de référence au Maroc est motivée principalement par les besoins incontestables des divers utilisateurs de l'information géographique à l'échelle nationale. L'objectif visé dans ce travail est avant tout de mettre en place un modèle de référence, considéré comme une structure de base pouvant être adaptée à un maximum d'applications. En ce sens, le catalogage des entités a constitué le premier pas vers cette modélisation, et a permis de mettre en avant les principes d'une ontologie que l'on peut qualifier d'« ontologie de données de référence ». En d'autres termes, la définition des types d'objets de référence – en autorisant un même concept à appartenir à plusieurs catégories différentes – amplifie la portée du modèle global de référence tout en assurant sa consolidation par les interactions entre les modèles spécifiques aux divers types d'objets de référence définis.

Dans le cadre des données de référence, un repositionnement des ontologies dans le processus général de développement d'un SIG permettrait d'enrichir la caractérisation des relations et des propriétés spatiales en vue d'une adaptation à diverses applications. En effet, pour une application couvrant un domaine donné, le développement d'un SIG se base sur le dictionnaire de données de l'organisme demandeur qui traduit l'ontologie du domaine étudié. Dans le cas des données de référence, l'élaboration d'un SIG débute à partir de la réflexion sur le dictionnaire de données, mais le contexte réclame la réintégration de l'ontologie au-delà du dictionnaire.

Une réflexion plus approfondie sur l'ontologie des données de référence peut être envisagée et permettrait d'étendre le modèle de référence en tenant compte des ontologies de domaine existantes pour les modèles des types d'objets spécifiques. Le modèle de référence élaboré dans le cadre de ce travail peut être exploité en implémentant l'ensemble des modèles des types d'objets développés. Par ailleurs, des extensions pourraient être apportées en considérant divers niveaux de détail pour concrétiser les possibilités d'héritage multiple de certaines classes d'entités. Dans ce sens, il conviendrait d'explicitier davantage la notion de « référence » pour décider du niveau d'exploitation des objets qualifiés précisément de référence.

Par ailleurs, les aspects techniques présentés ci-dessus devraient inciter les décideurs à prendre des mesures concrètes pour encadrer les efforts à l'échelle nationale. En ce qui concerne l'aspect organisationnel de la mise en place d'une IDS nationale, et en attendant l'instauration des réglementations nécessaires, les développements de ce travail pourraient jouer un rôle de sensibilisation des différents acteurs nationaux grâce au fait que :

- le catalogue d'entités conforme ISO constitue un grand pas vers la normalisation ;
- le dictionnaire de données élaboré dans le cadre du catalogage des entités est de nature à faciliter la communication entre les divers acteurs nationaux ;
- le modèle de référence permet la constitution de bases de données issues d'un même MCD, permettant ainsi de pallier aux problèmes d'hétérogénéité rencontrés généralement lors de l'intégration de données multi sources.

Enfin, l'adoption du modèle de référence peut venir en appui au développement des missions de normalisation et de standardisation de l'information géographique nationale. À cette fin, un Conseil National de l'Information Géographique (CNIG) devrait être mis en place et reconnu comme régulateur national légitime par l'ensemble des acteurs locaux, et constituer leur représentant dans les instances internationales traitant des IDS [IBA 2011].

Remerciements

Les auteurs tiennent à remercier la Coopération Technique Belge et les membres de l'unité de Géomatique de l'Université de Liège pour leur soutien durant les travaux de cette recherche.

5. Bibliographie

- [ANC 04] ANCFCC, "Cartographie : vision et stratégie", rapport interne, 2004, Direction de la Cartographie, ANCFCC, Rabat, Maroc.
- [BAD 00] Badard T., "Propagation des mises à jour dans les bases de données géographiques multi-représentations par analyse des changements géographiques", Thèse de doctorat, Université de Marne-la-Vallée, 2000.
- [BIS 98] Bishr, Y., "Overcoming the semantic and other barriers to GIS interoperability", *International Journal of Geographical Information Science*, Vol. 12, n° 4, 1998, p. 299-314.
- [CHA 07] Chaidron C., Billen R., Teller J., "Investigating a Bottom-up Approach for Extracting Ontologies from Urban Databases", *Studies in conceptual Intelligence (SCI)* n° 61, 2007, p. 131-141.

- [DAV 2009] Davis Jr., C.A., Fonseca, F.T., Câmara, G., “Beyond SDI: Integrating Science and Communities to Create Environmental Policies for the Sustainability of the Amazon”. *International Journal of Spatial Data Infrastructures Research*, Vol. 4, 2009, p. 156-174.
- [DEV 97] Devogele T., “Processus d’intégration et d’appariement de bases de données géographiques : Application à une base de données routières multi-échelles”, Thèse de doctorat, Université de Versailles, 1997.
- [DON 02] Donnay J.P., “Distributed GIS for sharing large scale data between public agencies. A case study in Belgium”, *Proceedings of GIS 2002 International Conference*, March 11-13 2002, Bahrain, p. 47-59.
- [DSM 07] Direction de la Statistique du Maroc, “Expérience marocaine en matière de cartographie et du SIG pour les recensements de la population”, *Atelier régional des Nations unies sur la cartographie et l’utilisation des SIG pour l’organisation des Recensements*, Rabat, 12-16 novembre 2007, 28 p.
- [DSN 07] Direction de la statistique des Nations Unies, “Rapport final, recommandations et conclusions”, *Atelier régional des Nations unies sur la cartographie et l’utilisation des SIG pour l’organisation des Recensements*, Rabat, 12-16 novembre 2007, 31 p.
- [ECA 04] Economic Commission for Africa, “SDI Africa : *An Implementation Guide*”, Geoinformation Team, ECA, 2004.
- [GAZ 06] Gazel H., Gadal S. et Lekavicut J. “Réalisation d’un système d’information géographique et d’aide à la décision pour optimiser la gestion des ressources forestières”, Colloque : Gestion concertée des ressources naturelles et de l’environnement, 28-30 juin 2006, Saint-Quentin.
- [GES 05] Gesbert N., “Étude de la formalisation des spécifications de bases de données géographiques en vue de leur intégration”, Thèse de doctorat, Université de Marne-la-Vallée, 2005.
- [GIN 02] GINIE, “Infrastructures de données spatiales (SDI) : Recommandations pour actions”, rapport du projet GINIE D5.3.2(a), *Geographic Information Network in Europe*, 6-8 Mai 2002, programme IST de l’Union Européenne, Ispra.
- [GIN 03] GINIE, “Survey of key GI players within Europe”, rapport du projet GINIE D2.1.1 *Geographic Information Network in Europe*, 11 novembre 2003, programme IST de l’Union Européenne, Ispra.
- [GOM 08] Gomez-Pérez A. Ramos J.A., Rodriguez-Pascual A., Vilches-Blazquez L.M., “The IGN-E Case : Integrating through a hidden ontology”, In : Ruas A. et Gold C. (Eds). *Headway in Spatial Data Handling*, 13th International Symposium on Spatial Data Handling. 2008, Berlin-Heidelberg, Springer-Verlag, p. 417- 435.
- [GRU 93] Gruber, T. R., “Toward Principles for the Design of Ontologies Used for Knowledge Sharing”, *International Journal of Human and Computer Studies*, n° 43, 1993, p. 907-928.
- [HAK 03] Hakimpour F., “Using ontology to resolve semantic heterogeneity for integrating Spatial database schemata”, Thèse de doctorat, Université de Zurich, 2003.

- [IBA 09] Ibannain F., “Modélisation des données géographiques de référence : Préalable à la mise en place d’une infrastructure de données spatiales au Maroc”, Thèse de doctorat, Université de Liège, 2009.
- [IBA 11] Ibannain F., “Mise en place d’une Base de Données Géographiques Nationale : Elaboration d’un catalogue d’entités conforme ISO et d’un MCD en adéquation avec la stratégie de l’ANCFCC”, Présentation à la direction de la cartographie, mars 2011, ANCFCC, Rabat.
- [IGN 01] IGN France, “Plan qualité du processus de fabrication de Cartign, Filière de production Cartign”, Rapport interne, 2001, Institut Géographique National, Paris.
- [IGN 02] IGN France, “BDTopo Pays / Agglomération: Spécifications de contenu”, Version 1.2. Service des Bases de Données Vecteurs”, rapport interne, 2002, Institut Géographique National, Paris.
- [IGN 03] IGN belge, “SGISR CDM v3.0 .Modèle conceptuel des données de références élaboré dans le cadre du projet SGISR Seamless Geographic Information System”, Rapport interne, 2003, Institut Géographique National, Bruxelles.
- [ISO 00] ISO, “ISO 9001 : Système de management de la qualité – Exigences”, 2000, Organisation internationale de normalisation, France.
- [ISO 05] ISO/TC 211. “ISO/TC 19110, Geographic information: Methodology for feature cataloguing”, 2005, Technical Committee ISO/TC 211, Geographic information/Geomatics, norme reprise intégralement comme norme marocaine (NM ISO 19110, 2007) par le Service de Normalisation Industrielle Marocaine (SNIMA), Rabat, Maroc.
- [LAP 06] Laplanche F., “Environnement de conception de bases de données spatiales sur Internet”, Thèse de doctorat, Université de Liège, 2006.
- [LAU 07] Laurini R., “Pre-consensus Ontologies and Urban Databases”, *Studies in computational Intelligence* (SCI), n° 61, 2007, p.27-36.
- [MOL 98] Molenaar M., “*An Introduction to the Theory of Spatial Object Modelling for GIS*”, Research Monographs in GIS Series, London: Taylor & Francis, 1998.
- [NOU 06] Noucher M., “Mutualisation de l’information géographique : Infrastructure de données ou communauté d’acteurs ?” *GEO-EVENEMENT* 2006, Paris, France, 10 p.
- [NOU 08] Noucher M., Golay F., Sède-Marceau M. H., Pornon H., “OPDE 2008 : les outils pour décider ensemble?”, OPDE 2008, Québec, 5-6 juin 2008, 23 p.
- [NSH 04] NSHC, “Report from the 26th Conference of the North Sea Hydrographic Commission”, 20-23 septembre 2004, Cardiff, 153 p.
- [OS 06] Ordnance Survey, “Implementing OS Mastermap. Information Sheet 1: Data models and management (v1.4 – April 2006)”, Ordnance Survey, Southampton.
- [PAN 96] Pantazis, D., Donnay, J.P., *La conception de SIG : méthode et formalisme*, Collection géomatique, Paris, Hermès, 1996.

- [PAN 98] Pantazis D., Donnay J-P., “Objets géographiques à limites indéterminées. Modélisation et intégration dans un modèle conceptuel de données”, *Revue Internationale de Géomatique*, Vol. 7, n° 2, p. 159–186.
- [PAR 06] Parent, C., Spaccapietra, S., Zymanyi, E., *Conceptual Modelling for Traditional and Spatio-Temporal Application. The MADS Approach*, Berlin-Heidelberg, Springer-Verlag.
- [POI 07] Poirel G., “Cadre d’optimisation de la cartographie pré-censitaire”, *Atelier régional des Nations Unies sur la cartographie et l’utilisation des SIG pour l’organisation des Recensements*, Rabat, 12-16 novembre 2007, 22 p.
- [SAL 06] Salgé F., “Patrimoine géographique et infrastructure de données géoréférencées”, Séminaire “*Information géographique et aide au développement durable*”, Paris, 4 mai 2006, CNIG.
- [SHE 05] Sheeren D., “Méthodologie d’évaluation de la cohérence inter représentations pour l’intégration de bases de données spatiales : une approche combinant l’utilisation de métadonnées et l’apprentissage automatique”, Thèse de doctorat, Université de Paris 6, 2005.
- [SHO 04] SHOM, “La lettre du SHOM”, n°21, Décembre 2004.
- [STO 08] Stoter J.E., Morales J.M., Lemmens R.L.G., Meijers B.M., Van Oosterom P.J.M., Quak C.W., Uitermark H.T., Van Den Brink L., “A data model for multi-scale topographical data, in Ruas A. et Gold C. (Eds). “*Headway in Spatial Data Handling, 13th International Symposium on Spatial Data Handling*”, 2008, Berlin-Heidelberg, Springer-Verlag, p 233-254.

NOTE BIOGRAPHIQUES

Fatiha Ibannain est ingénieur géomètre topographe à l’Agence Nationale de la Conservation Foncière, du Cadastre et de la Cartographie. Membre du conseil national de l’Ordre des Ingénieurs Géomètres Topographes (ONIGT, Maroc). Titulaire d’un doctorat en Sciences de l’université de Liège préparé au sein de l’unité de géomatique. Ses travaux portent sur la cartographie, la normalisation et les systèmes d’information géographique.

Jean-Paul Donnay est professeur ordinaire à l’Université de Liège, où il a créé en 1997 l’unité de géomatique au sein de la faculté des sciences. Titulaire d’un doctorat en sciences et d’une maîtrise en sciences appliquées, ses domaines de recherche et d’enseignement sont la cartographie, l’analyse spatiale et les systèmes d’information géographique.

Map of economic risk of decrease in farmland costs in Stavropol Krai, caused by geological hazards as the territory-management tool

Kiselev O., Pyrchenko V.

Sergeev Institute of Environmental Geoscience, Russian Academy of Sciences (IEG RAS)

Ulansky per. 13, bld. 2, P.O. Box 145, 101000 Moscow, Russia, e-mail:
vapo41@list.ru, okiselew@gmail.com

RÉSUMÉ. Une méthode est ici proposée pour cartographier les risques de perte économique en terme de valeur des terres agricoles sachant certains risques naturels. Le matériel d'étude utilisé par les auteurs se compose : de cartes topographiques, de notes d'arpentage, de données de télédétection, d'estimations cadastrales du prix des terres. Une carte a été créée en utilisant la méthode des auteurs afin d'évaluer la vulnérabilité et les risques causés par les catastrophes naturelles. La base de données géographiques contient différentes couches vectorielles et images. Le SIG et la carte finale ont été créés avec le logiciel ArcGIS d'ESRI. La carte des risques économiques liés aux terres agricoles pourra être utilisée pour aider à prendre des décisions administratives concernant la réduction et la prévention des pertes, ainsi que pour justifier des investissements visant à protéger les régions.

ABSTRACT. The method for mapping the risk of economic loss in value of farmlands from natural hazards is offered. Authors used following materials: topographic maps, survey notes of past years, remote sensing data, cadastral estimations of lands cost, etc. The map was created using the author's method for assessing vulnerability and risk from natural hazards. Geographical database contains different raster and vector layers. GIS and final map were made using ESRI ArcGIS software. Map of economic risk of decrease in farmland costs can be used for justification of administrative decisions on reduction and the prevention of losses from hazardous processes, as well as for justification of investments into projects to protect these lands.

MOTS-CLÉS: carte de risque économique, SIG, base des données géographique, imagerie satellitaire, risques des dangers naturels, prévention des pertes de processus dangereux.

KEYWORDS: map of economic risk, GIS, geographical database, remote sensing, risk from natural hazards, prevention of losses from hazardous processes.

Within the last decades rates of development of new territories, intensity of use of the already mastered many times over grew. A direct consequence of it is the increase in material losses and cases of death of people from hazardous natural processes. Damage and risk mapping from hazardous natural processes is the actual and profitable preventive measure allowing considerably to lower a damage from natural hazards.

Stavropol Krai – one of the largest agricultural regions of Russia. Farmlands occupy about 80 % of its territory. At the same time in the territory of the region more than dozen of hazardous geological processes take place. They damage farmlands, considerably reducing their efficiency and cost. We considered three processes, making the greatest impact on change of qualities of lands and their cost: landslides, gully erosion and deflation.

As a result of this work the Map of economic risk of decrease in farmland costs of Stavropol Krai, caused by geological hazards from the listed processes is made. Similar maps are the synthetic cartographical works uniting mapping of a natural and social and economic component, received as a result of analytical estimated work.

The work was implemented author's technique, which allows linking data on the activity of natural processes, quantitative characteristic of cost of land and the qualitative characteristics of the study area. Important elements of this technique are tables of vulnerability of the farmland from impact of hazardous processes. The quantitative assessment of loss of efficiency of the farmland is determined by them depending on intensity of hazardous process. Work was made in the following sequence: collecting and the analysis of necessary information, adaptation of basis surface chart, drawing up of intermediate thematic maps of categories of lands and hazardous natural processes, calculation and drawing up of the map of decrease in farmland costs from dangerous geological processes.

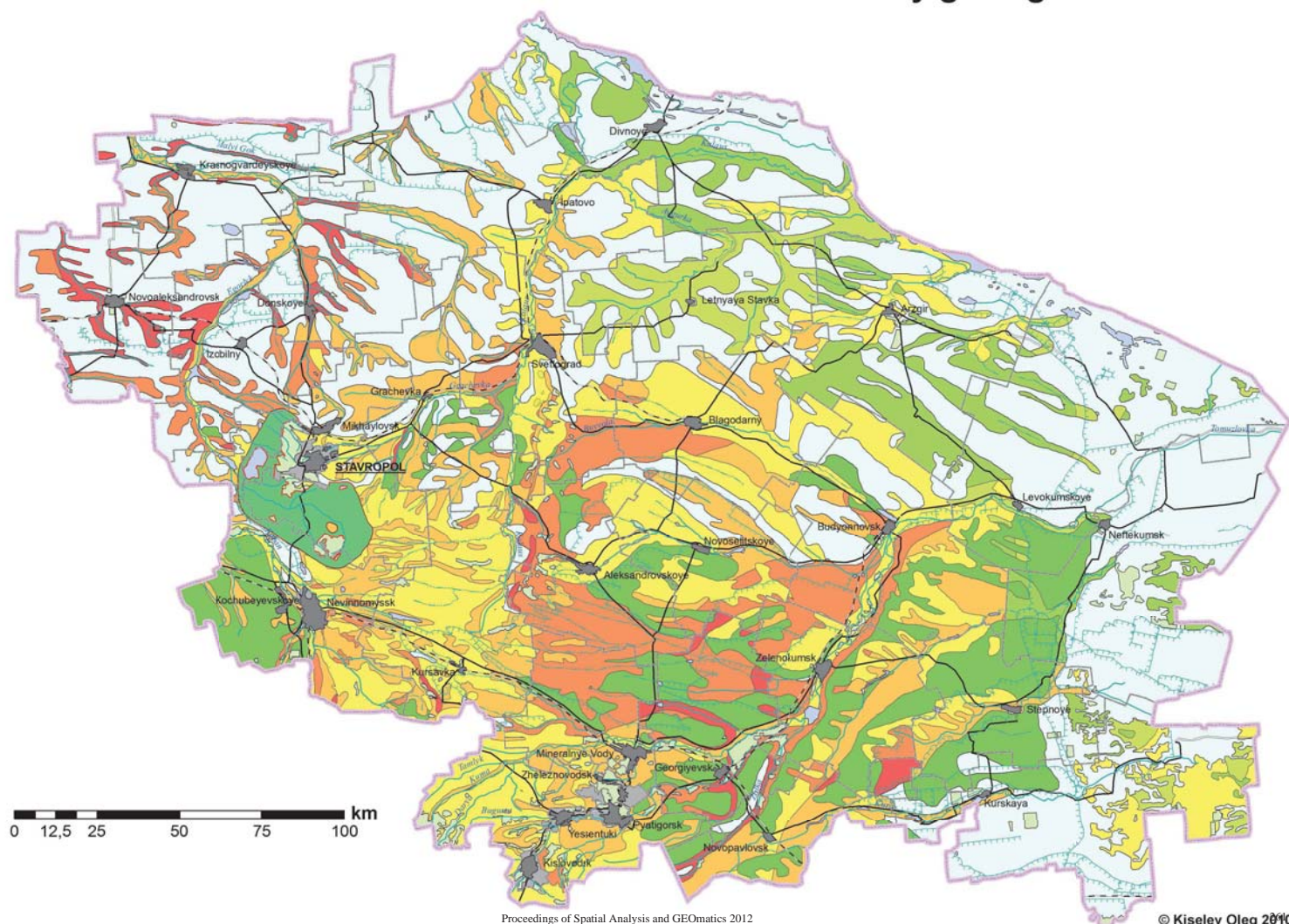
The authors have used different types of sources: the survey notes of past years, cartographic sources, remote sensing, literary (normative acts, territorial planning documents and decrees of the Russian and the Stavropol Krai government), statistical data on the cost of land, resources Internet (mostly statistics site of the Federal service for State registration, Cadastre, and Cartography (Rosreestr). Cartographic work was carried out in software ArcGis.

The main tool to create the final map was method of identical crossing (Identity Overlay). As a result, the boundaries were separate final (estimated) taxons and identified areas for which the calculation of risk is not produced. According to calculations, the range of values obtained specific annual economic risk has been divided into seven grades, which are reflected in the final map. It is a qualitative description of the Stavropol Krai in terms of the likely cost of loss of agricultural land, which takes into account the territorial differences in the cadastral value of land and the dynamics of hazardous geological processes.

Map of economic risk of decrease in farmland costs can be used, first of all, for justification of administrative decisions on reduction and the prevention of losses from hazardous processes, and also for justification of investments into projects to protect these lands. Similar maps serve as the basis for development of standard and regulatory documents and justification of methods of management by the region. Also, it can be carried out by the calculations of return on various types of agricultural activity on certain lands.

Economic risk of decrease in farmland costs (caused by geological hazards)

Actes de la Conférence internationale de Géomatique et Analyse Spatiale 2012











Proceedings of Spatial Analysis and GEOmatics 2012

© Kiselev Oleg 2010


Legend

Values of economic risk of decrease in farmland costs


	Thousands rubles per year/ha	Economic risk level due to its value
	0	Zero
	0-20	Ignorable
	20-30	Low
	30-100	Moderate
	100-200	High
	200-300	Very high
	> 300	Extreme
 Taxons' boundaries		


Unrated areas

Farmlands

 Farmlands without impact of geological hazards

Hydrography


 Lakes, reservoirs

 Channels

 Intermittent watercourses

 Watercourses

Vegetation

 Woods including arboretum

Protected areas

 National park


Settlements

 Towns

 City districts

Roads

 Railway

 Highway

 Unpaved roads

Boundaries

 Federal subject

 Districts

 STAVROPOL Administrative center

 Izobilny District center

Le franchissement d'obstacles en navigation intérieure

Luc Leclercq

*Unité de Géomatique
Université de Liège
17 Allée du 6-Août,
B-4000 Liège, Belgique
LucLeclercq@ulg.ac.be*

RÉSUMÉ. Dans le contexte de la navigation en espace fermé (navigation intérieure), nous nous sommes intéressés à la manière de franchir un obstacle et la façon de communiquer le nouveau trajet à un utilisateur. Pour ce faire, quelques bases de réflexion en ce qui concerne la perception et la modélisation de l'espace clos seront posées. Ensuite, des définitions et des règles formalisées seront établies afin de créer un modèle logico-géométrique répondant à notre problématique du franchissement d'obstacles. Sur ce modèle, un algorithme sera implémenté permettant de créer un nouveau trajet. Le moyen de communiquer ce trajet aux utilisateurs en langage naturel sera également abordé. Finalement, le modèle de franchissement pourra être validé en le soumettant à quelques scénarios dans un logiciel SIG.

ABSTRACT. In the context of navigation in closed space (indoor navigation), we looked at means of crossing obstacle and how to communicate the new path to a user. To do so, we expose some basic ideas about perception and modeling of confined space. Then, definitions and formalized rules will be established to create a logical-geometrical model corresponding to our problem of obstacle clearance. Based on this model an algorithm will be implemented to create a new path. The way to communicate this route to users in natural language will also be addressed. Finally, the crossing model can be validated by subjecting it to a few scenarios in GIS software.

MOTS-CLÉS : navigation intérieure, franchissement d'obstacles, modèle logico-géométrique, langage naturel, logiciel SIG.

KEYWORDS: indoor navigation, obstacles crossing, logical-geometrical model, natural language, GIS software.

1. Introduction

Les progrès techniques en matière de positionnement ont influencé ces dernières années les recherches en transport et en navigation, permettant, par exemple, à un utilisateur de s'orienter et de trouver le chemin le plus adéquat via les réseaux routiers. On parle dans ce cas de navigation extérieure dans un espace ouvert (« outdoor »). Leurs équivalents existent en espaces intérieurs « indoor », c'est-à-dire en milieu contraint fermé. Ces deux systèmes sont composés de trois modules : (i) localisation de « l'agent » à informer dans l'espace considéré, (ii) génération des informations de navigation (analyse spatiale), (iii) diffusion des informations de navigation (sous la forme la plus adéquate possible). Un espace contraint fermé (typiquement, un bâtiment) de par sa structuration spatiale complexe, pose généralement plus de problème de navigation qu'un espace ouvert. En effet, ce type d'espace est souvent composé de nombreuses cellules qui contiennent des objets de toute sorte (natures et structures différentes). Il peut également présenter une certaine dynamique, sa configuration peut changer au cours du temps (ex: un meuble déplacé, un accès bloqué, ...) et ainsi modifier considérablement l'information de navigation à transmettre. Ainsi, une aide à la navigation peut être envisagée spécialement si un obstacle (objet, trou,...) est présent sur le trajet. La manière de franchir cet obstacle et la façon de représenter et communiquer le trajet à suivre pour le contourner peut être envisagé de différentes manières.

La première partie sera consacrée à la description des différentes approches réalisées dans le domaine de la navigation intérieure pour des utilisateurs tant au point de vue de la perception de l'espace qu'au niveau des modèles conceptuels déjà réalisés. La deuxième partie comportera des développements terminologiques et théoriques permettant d'arriver à un modèle conceptuel adapté parfaitement à la problématique. La manière de franchir l'obstacle sera ensuite étudiée d'un point de vue algorithmique. Puis, sur base des travaux de cognition spatiale décrits, la façon de communiquer le trajet aux utilisateurs sera envisagée. Finalement, la troisième partie du travail portera sur la phase applicative. Le modèle ainsi créé sera implémenté dans un logiciel SIG.

2. État de l'art

Cette partie contient des bases de réflexion utiles au développement d'un modèle spécifique à la navigation intérieure. Nous ne présenterons ici qu'une sélection limitée qui nous a inspirée. Ceci ne constitue donc pas un inventaire exhaustif de ce qui a été étudié dans le domaine.

2.1. Localisation

En matière de localisation « indoor », nous pouvons distinguer deux types de systèmes. Le premier est dit passif, car il ne réclame pas d'action spécifique pour l'utilisateur, il doit juste disposer d'un récepteur. Un tel système peut reposer sur des

technologies comme des senseurs Wi-Fi utilisant la triangulation [7] et exploitant les compas et accéléromètres des Smartphones. L'utilisateur peut également agir activement sur le système, comme par exemple en pointant un élément spécifique (un QRcode ou une écriture particulière) [4]. Pour un utilisateur « robot », la localisation s'effectue par comparaison des mesures avec le modèle de représentation présentant une série de points d'intérêts (« landmarks »). FINALE et NEURONAV sont des systèmes fonctionnant sur ces principes [2]. Les recherches les plus récentes tendent à combiner différents modes de positionnement afin d'assurer de meilleure précision de localisation et couverture [5]. Ce domaine est actuellement en plein développement.

2.2. Analyse spatiale

Afin de pouvoir utiliser des outils d'analyse spatiale et ainsi générer une information de navigation, il faut pouvoir modéliser l'espace englobant. Un tel modèle de représentation doit prendre en compte la topologie ainsi que la géométrie des différentes entités [6] pour que des algorithmes de plus court chemin comme Dijkstra puissent s'y déployer. Ce type de modèle prend généralement la forme d'un graphe (nœuds et arcs [6]). En espaces contraints et fermés, la sémantique prend toute son importance afin de pouvoir définir chaque entité et arriver à des ontologies de déplacement [8], mais il existe actuellement peu d'implémentation d'un tel modèle enrichi de formalismes propres à un espace « indoor ». Ces thèmes ont émergé depuis peu notamment en sciences de l'information géographique.

2.3. Langage naturel

L'information doit ensuite être diffusée à l'utilisateur. Cette information est communiquée de manière vocale ou graphique le plus clairement possible (en langage naturel) [10]. Cette interface homme-machine nécessite donc des formalismes et des ontologies suffisamment puissants pour que l'information soit comprise par tout le monde. Des travaux de cognition spatiale comme la notion de scène élémentaire et d'espace conceptuel commun [1] [3] permettent de décrire quelques relations spatiales qui existent entre des différentes entités, mais elles sont forts basiques et limitées à deux dimensions.

3. Développements

3.1. Modèle conceptuel

Il est indispensable de définir les différentes entités qui feront partie du modèle adopté afin qu'il ne subsiste aucune ambiguïté ni aucun doute sur la nature et la fonction de ces entités (bâtiment, compartiment, couloir, local, liaison et obstacle). Sur base de ces définitions et de relations topologiques décrites par [8], un modèle sémantique sera tout d'abord créé (Figure 1).

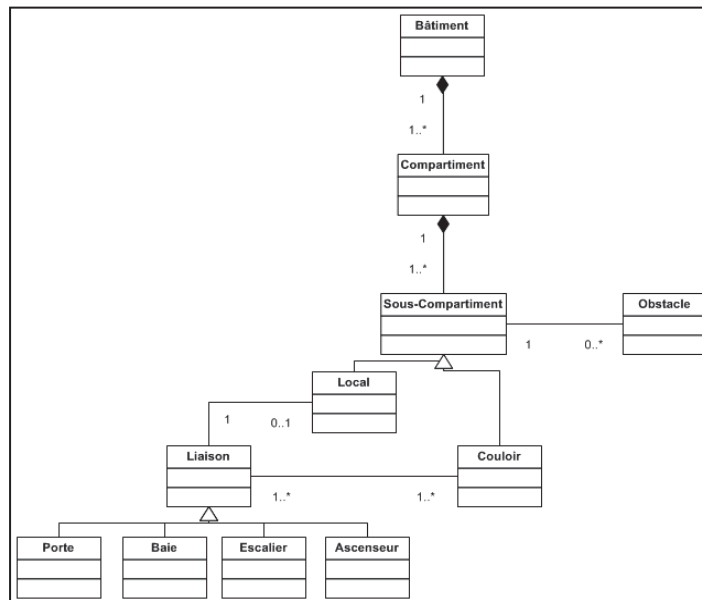


Figure 1. *Modèle sémantique formalisé en UML*

Ensuite, grâce à ce modèle sémantique, le modèle logico-géométrique est créé. La représentation choisie sera celle du graphe nœud-relation (ou nœud-arc) [6] tenant compte de la géométrie (logico-géométrique). Un nœud est un point relié à un autre nœud par un arc. Un arc est une relation entre deux nœuds et est représenté par une ligne. Un polygone dans une telle représentation est donc composé de lignes et de nœuds tels que le premier nœud correspond au dernier (fermeture). La représentation visuelle des entités devra également être formalisée. Par exemple, un local sera composé d'une seule liaison (nœud de type « liaison » en gris clair situé au centre de gravité de la porte par exemple) et pourra comporter des nœuds indiquant la présence de l'utilisateur (de type « local » en gris foncé) ou également d'un objet comme un extincteur. Un arc reliera se basant sur un algorithme de plus court chemin un nœud de type « local » au nœud de type « liaison » (Figure 2). De même, un couloir sera composé d'arcs et de nœuds. L'arc sera la représentation de l'axe médian du couloir [9] et des nœuds seront créés à des intersections, à un tournant (de type « couloir » en gris clair) et devant un local (de type « intersection » en gris foncé) obtenus en projetant le nœud de type « liaison » perpendiculairement sur l'axe médian (Figure 3). Un arc sera créé entre les nœuds de type « intersection » et « liaison ». Les régions non convexes seront découpées en sous-régions convexes et la même méthode sera appliquée [11]. Un obstacle sera lui un polygone et sera traité comme un objet individuel. Les liaisons peuvent être elles-mêmes des obstacles (porte fermée, ascenseur trop petit ou en panne, franchissement de

l'escalier impossible, passage trop étroit). Il sera symbolisé par son emprise sur le sol (polygone) et l'algorithme de franchissement créera des nœuds de type « obstacle » et les arcs les rejoignant.

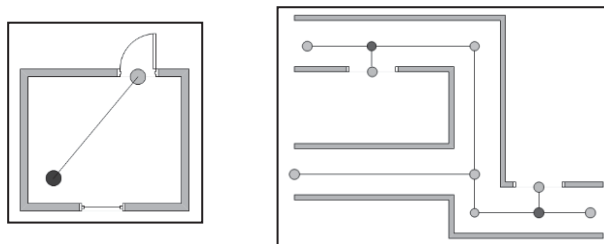


Figure 2 et 3. Exemple de représentation d'un local et d'un couloir

3.2. Algorithme de franchissement

La manière de franchir un obstacle va maintenant être abordée. Pour ce faire, un algorithme faisant appel à des opérateurs sera créé puis exploité. Bien sûr, pour que l'algorithme fonctionne, il faut que la géométrie de l'obstacle soit connue dans le modèle. La solution retenue a été celle de l'enveloppe convexe. L'enveloppe convexe (*Convex Hull*) d'un objet est le sous-ensemble convexe le plus petit le contenant. Un ensemble est dit convexe si et seulement si le segment joignant deux points quelconques de cet ensemble lui est inclus.

3.3. Nouveau chemin

Une fois le réseau mis à jour, le nouveau chemin doit passer par les nœuds et arcs ainsi créés. Ce nouveau chemin sera communiqué aux utilisateurs. C'est en langage naturel qu'il conviendra d'exprimer le nouveau trajet. C'est des instructions de franchissement comme « contourner » ou « longer » qui seront utilisées, et pour cela, il sera nécessaire de formaliser ces instructions afin que chacun les comprenne de la même façon. On se basera sur les travaux de [1] et [3] qui ont formalisé de manière conceptuelle (« contourner » implique un changement de direction) et graphique (« longer » est un trait droit et « contourner » une courbe) ces opérations (Figure 4).

(4) Action Localisée	Longer la poste. Passer entre la poste et la boulangerie.	
----------------------	---	--

Figure 4. Action localisée [10]

Sur base du coefficient angulaire de la droite joignant l'utilisateur à sa destination, nous pourrions déterminer si ces deux opérations se font par la gauche ou la droite.

4. Application

Le modèle ainsi décrit, les entités définies et l'algorithme de franchissement choisi, une implémentation au sein d'un logiciel de traitement des informations géographiques et géométriques efficace (logiciel dit « SIG ») peut s'effectuer. C'est le logiciel ArcGIS version 10 d'ESRI qui a été choisi car pouvant représenter les entités sous forme de réseau et la possibilité d'y chercher un chemin. Un schéma se basant sur le modèle sémantique (voir Figure 1) est créé. Ensuite, il est implémenté dans ArcGIS sous forme d'une « Personal Geodatabase » et des données y sont ajoutées. Nous allons maintenant franchir l'obstacle en créant des nouveaux nœuds et arcs grâce à l'algorithme de l'enveloppe convexe (*Convex Hull*) (Figure 5). Ensuite, un réseau est créé et l'extension « Network analyst » permet de calculer et de visualiser le plus court trajet entre deux points (Figure 6). Cette visualisation comprend une fenêtre qui indique étape par étape le trajet à effectuer.

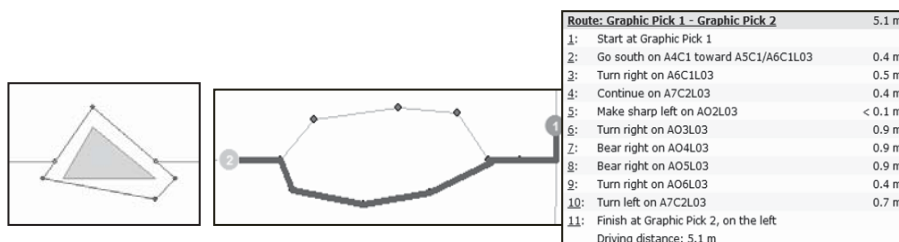


Figure 5 et 6. Résultat de l'algorithme et visualisation du trajet

5. Conclusions

Dans ce travail, nous nous sommes intéressés à la navigation intérieure au sein d'un bâtiment. Plus précisément, la manière de franchir un obstacle dans ce milieu contraignant a été étudiée. La première partie inventorie quelques modèles qui ont été développés et nous avons pu mettre en évidence certains manques de formalismes propres à la navigation « indoor ». Une série de réflexions et terminologies théoriques ont ensuite été faites pour élaborer un modèle logico-géométrique adapté à notre problématique du franchissement d'obstacles. En sélectionnant un algorithme et en regardant comment communiquer le nouveau chemin aux utilisateurs, la manière de franchir l'obstacle a été définie. Ce modèle simple a ensuite été implémenté dans le logiciel ArcGIS qui s'est montré assez performant. Bien que des améliorations pourraient être apportées (gestion de l'obstacle en temps réel, le meilleurs chemin n'est pas toujours le plus court, prise en compte de la troisième dimension), ce modèle de franchissement a pu être validé.

6. Bibliographie

- [1] Couclelis H., « Verbal directions for way-finding: Space, cognition, and language », The construction of cognitive maps, In J. Portugali (Ed.), 1996, p. 133-153.
- [2] Desouza G., Kak A.C., « Vision for Mobile Robot Navigation: A Survey », *Proceedings IEEE transactions on pattern analysis and machine intelligence*, vol. 24, n° 2, 2002, 31 p.
- [3] Gärdenfors P., « Conceptual spaces: The geometry of thought », MA: The MIT Press, Cambridge, 2000, 317 p.
- [4] Koltringer T. et al., « Understanding Spatial Mobile Service Discovery using Pointing and Touching », *Mobile Spatial Interaction Workshop Whitepaper, San Jose*, 2007, 4 p.
- [5] Korbinian F. et al., « Development and Evaluation of a Combined WLAN and Inertial Indoor Pedestrian Positioning System », *4th International Symposium on Location and Context Awareness, Georgia (USA)*, 2009, 9 p.
- [6] Lee J., « 3D Data Model for Representing Topological Relations of Urban Features », *21st Annual ESRI International User Conference, San Diego*, 2001, 16 p.
- [7] Lee W. et al., « Indoor Navigation System for Mobile Robot using Wireless Sensor Network », *Korea University, 2008*, 2 p.
- [8] Li Yuan, He Zizhang, « 3D Indoor Navigation: a Framework of Combining BIM with 3D GIS », *44th ISOCARP Congress, China*, 2008, 10 p.
- [9] Meijers M., Zlatanova S., Pfeifer N., « 3D Geo-information indoors: Structuring for evacuation », *Proceedings of Next Generation 3D City Models, Bonn*, 2005, 6 p.
- [10] Przytula-Machrouh E., Ligozat G., Denis M., « Des ontologies transmodales pour la description d'itinéraires », *Revue Internationale de Géomatique*, vol. 14, n° 2, 2004, p. 285-302.
- [11] Stoffel E.-P., Lorenz B., Ohlbach H. J., « Towards a Semantic Spatial Model for Pedestrian Indoor Navigation », *Advances in Conceptual Modeling – Foundations and Applications, Auckland (New Zealand)*, 2007, p. 328-337.

Une approche basée objet combinée avec les classificateurs avancés (SVM, RF, RN) pour la détection des changements du bâti.

Loubna Elmansouri

Unité de Géomatique, Université de Liège
Bât. B5A Unité de Géomatique, allée du 6 Août 17, 4000 Liège 1, Belgique
l.elmansouri@student.ulg.ac.be

RÉSUMÉ. L'objectif de cet article est de développer une méthodologie automatique pour la détection des changements du bâti en milieu urbain. L'approche est basée sur la comparaison entre un plan existant et une image à très haute résolution spatiale (THRS) récente. Elle s'enchaîne sur 4 étapes. Premièrement l'image est segmentée pour produire des primitives. La sélection des attributs est faite ensuite quantitativement par l'algorithme MrMR, basé sur l'information mutuelle. Des classificateurs par apprentissage supervisé avancés ont été utilisés dans l'étape de la classification (SVM, Ensembles d'arbres de décision et Réseaux de Neurones). La prédiction finale des objets est déduite par un système de vote à partir des 3 meilleurs classificateurs. Enfin, une étape de détection de changements permet d'identifier les segments qui représentent les changements du bâti. Les données utilisées concernent la ville de Rabat (Maroc). Une image QuickBird a été utilisée avec un plan existant à l'échelle 1:10,000. Indépendamment de la qualité de la forme des bâtiments détectés, la méthode offre des meilleurs taux en complétude (95%) et en bonne détection (83%).

ABSTRACT. The aim of this study is the development of an automatic methodology for change detection of buildings in urban environments. The approach is based on a comparison between an existing map and a newly registered very high spatial resolution satellite image (VHSR). The whole process is performed mainly on four steps. First, the image was segmented to give primitives. Then, features are selected quantitatively by a mutual information-based algorithm (MrMR). Thereafter, advanced supervised learning algorithms are used in a classification step (SVM, Ensembles of decision trees and Neural Network). The final objects' prediction is deducted thanks to a vote-based system from the three better classifiers. Last, a post classification change detection step allows to identify the segments which represent the changes of buildings. The data used in this research concern the city of Rabat (Morocco). A QuickBird image has been used with an old map at the scale of 1:10,000. Regardless of the quality of the detected buildings' shape, the method achieves good rates of completeness (95%) and correctness (83%).

MOTS-CLÉ S. Détection des changements du bâti, image à (THRS), SVM, Random Forest, Réseaux de Neurones.

KEYWORDS. Buildings' change detection, (VHSR) image, SVM, Random Forest, Neural Network.

1.Introduction

Dans tous les pays, qu'ils soient développés ou en voie de développement, un besoin pressant en matière d'informations précises sur le territoire est constaté. Les bases de données géographiques, comme source principale de ces informations, sont indispensables pour l'étude et la prise de décision dans des situations d'urgence ou pour des planifications à long terme [BOU 07]. L'efficacité de ces bases de données est directement liée à la fiabilité de leur contenu [PUI 03], qui ne peut être assurée avec des fréquences de mise à jour de l'ordre de plusieurs années [WAL 04]. Face à ce constat, notre réflexion s'est portée sur les méthodes pratiquées par les services compétents pour la mise à jour de leurs bases de données géographiques (BDG).

Deux stratégies de mise à jour cartographique ont été recensées: *la mise à jour systématique* et *la mise à jour continue*. On parle de *mise à jour systématique* lorsque la date prévue pour la révision de la base de données est fixée à l'avance au moment de la production de la base de données. Elle consiste en une reproduction cartographique complète du territoire, souvent réalisée par une restitution manuelle des photographies aériennes complétée par des levés sur terrain. En plus du coût élevé, l'inconvénient majeur de la méthode est que la base de données existante est complètement ignorée dans le processus de sa mise à jour [KNU 03]. *La mise à jour continue*, quand à elle, consiste à cartographier uniquement les zones de changements à partir de l'ancienne version de la base de données, ce qui garantit implicitement une réduction du coût de l'opération de mise à jour. Or, souvent la détection de changement se fait d'une manière interactive (*on- screen change detection*) qui nécessite, à son tour, un temps et des efforts considérables [BOU 07, HOL 06]. Par conséquent, sa faisabilité reste restreinte au milieu rural où le taux de changements est faible. L'automatisation de cette étape se fait ressentir par les producteurs de données et a fait l'objet de plusieurs études.

Depuis au moins une décennie, les recherches se sont dirigées vers l'automatisation de l'étape de détection de changements dans le processus de la mise à jour continue des BDG. Plusieurs agences de cartographie Européennes s'y mettent; les projets: *ATKIS* [WAL 04], *ATOMI* [BAL 04], *TOP10DK* [KNU 03] et *ORDONNANCE SURVEY* [HOL 06] en témoignent. Indépendamment de la méthode utilisée et des données exploitées (images satellites à très haute résolution spatiale (THRS) ou photographies aériennes), les résultats sont encourageants puisqu'ils garantissent une grande épargne de temps en comparaison aux méthodes classiques de détection de changement [HOL 06]. Le détail des procédures de détection automatique de changements utilisées dans ces projets et d'autres procédures testées entre 2004 et 2008, peut être trouvé en [MAY 08]. Parallèlement, l'organisation *EuroSDR* (*European Spatial Data Research*) a confirmé, dans son rapport de synthèse sur la détection automatique de changement de bâtis dans les bases de données topographiques en 2D, que ces méthodes automatiques sont de grand intérêt économique puisque plus de 80% des bâtiments dans la base de données n'ont pas à être révisés [CHA 09].

Deux approches typiques sont utilisées pour détecter automatiquement les changements dans la mise à jour des BDG: la comparaison *image-image* et la comparaison *image-plan*. La première méthode consiste à comparer les résultats de la classification de deux images diachroniques. C'est une méthode générale de détection de changement, pratiquée depuis quelques décennies pour le suivi des phénomènes naturels dans le temps; son utilisation pour la mise à jour des BDG exige la disponibilité des images à la même date de la production de la base de données à mettre à jour, ce qui a limité son utilisation [KNU 03]. La seconde méthode est basée sur la comparaison du résultat de l'extraction des données à partir d'une image récente avec la base de données existante. L'avantage principal de cette dernière méthode est que la base de données existante peut guider le processus de détection des changements. Elle peut intervenir dans le processus de l'extraction, soit pour contraindre la segmentation [CAR 06], soit pour entraîner la classification [WAL 04], ou aussi pour fournir de la connaissance a priori [BOU 07].

La thématique la plus dynamique et nécessitant une fréquence de mise à jour importante est le bâti, notamment en milieu urbain. Plusieurs études s'y sont intéressées. Une revue exhaustive est présentée dans [MAY 08].

Les tendances actuelles dans les méthodes de détection de changements du bâti s'orientent vers l'utilisation des images à THRS qui présentent plusieurs avantages (*moindres coûts, répétitivité à court intervalle*). Ce qui leur permet de rivaliser avec la photographie aérienne en urbain [PUI 03]. Les approches basées pixel trouvent leurs limites avec ce type d'imagerie. Ainsi, dans [KNU 02] les auteurs ont combiné deux classifications basées pixel, une supervisée et une autre non supervisée, pour détecter les changements sur une image à THRS. Les résultats n'étant pas satisfaisants, l'intégration d'un MNS a été nécessaire pour les améliorer [KNU 03]. Les approches basées objet conviennent aux mieux pour les images à THRS, l'entité de base n'est plus le pixel mais un groupe de pixels adjacents à propriétés similaires. Ce qui permet (1) d'éviter l'effet "salt and pepper" causé par l'hétérogénéité spectrale et la variance spatiale au sein de l'image (2) de décrire les objets par d'autres types d'attributs (géométrie, contexte, texture, association) en plus de leur réponse spectrale [BAA 00] et également (3) de s'approcher des objets tels qu'ils sont perçus par l'œil humain. Dans [WAL 04] les auteurs proposent une approche de mise à jour basée objet. Les limites des anciens objets sont tirées directement de l'ancienne base de données. Les objets sont caractérisés par différents attributs et classés ensuite par Maximum de Vraisemblance. Une approche basée objet en 3 étapes typiques pour localiser les nouveaux bâtiments sur les images à THRS a été présentée dans [HER 11]. Les primitives résultantes d'une première étape de segmentation ont été caractérisées par différents attributs et classées ensuite par arbre de décision. Dans le contexte de mise à jour de BDG, une étude comparative a été réalisée dans [HOL 08]. Trois types de classifications ont été testés: deux classifications basées pixels par SVM (Machines à supports vectoriels) et arbre de décision et une classification basée objet par l'algorithme des plus proches voisins. Les auteurs ont conclu sur l'efficacité de la classification basée objet et ont

recommandé de la combiner avec l'algorithme d'arbre de décision. Dans [LEI 10] les changements du bâti ont été détectés en comparant le résultat d'une classification basée objet d'une image à THRS avec une version simplifiée de la base de données existante. Les auteurs ont comparé différents algorithmes de classification par apprentissage supervisé. Un système de règles de connaissance tirées en partie à partir de la carte existante pour la reconnaissance du bâti sur une image à THRS récente a été proposé dans [BOU 07]. Les anciens objets de la carte ont servi pour (1) amorcer la segmentation, (2) entraîner l'apprentissage et (3) automatiser la détermination des paramètres des traitements.

Dans les approches basées objet, deux variantes peuvent être identifiées selon la méthode de classification adoptée. La première consiste en une classification par règles de connaissance [BOU 07]. Ces règles sont déduites des propriétés intrinsèques des objets ainsi que des relations contextuelles entre les différentes classes d'objets. Ce type de classification est spécialement critiqué pour la non portabilité de la base de connaissance construite [PUI 07, CHA 09]. Dans la seconde méthode de classification, différents algorithmes d'apprentissage, provenant de l'intelligence artificielle, sont exploités. Les résultats de ces classifications sont jugés encourageants [LEI 10, HOL 08].

Dans cet article nous examinons la détection automatique des changements du bâti pour la mise à jour des plans de ville au Maroc, où la mise à jour se fait habituellement en mode systématique. Notre objectif principal est de recenser les nouveaux bâtiments. La méthodologie proposée est basée sur une comparaison image-plan dans un processus basé objet. Différents algorithmes d'apprentissage automatique supervisé ont été utilisés.

L'article est structuré comme suit: la section 2 présente la zone d'étude et les données utilisées. La méthodologie complète est détaillée dans la section 3. Dans la section 4 nous présentons les résultats, leur validation et des discussions. La conclusion et les travaux en cours sont donnés dans la section 5.

2.Site d'étude, données et prétraitements

2.1. Site d'étude

Le site d'étude correspond à Rabat, capitale administrative du Maroc. Elle est définie géographiquement par sa latitude 33° 1' 31" et sa longitude 6° 53' 10". Son relief est plat avoisinant les 45 m. Rabat compte pour la 3^{ème} ville du Maroc avec une population de 1.6 million, selon le dernier recensement officiel de 2004 [BOU 07]. Les quartiers périphériques de la ville connaissent une forte urbanisation difficile à contrôler.

2.2. Données

Les données images utilisées dans ce travail consistent en une image Quickbird acquise en Aout 2004 avec une orientation du capteur de 88°. Cinq bandes sont disponibles: 1 bande panchromatique avec 0.6m de résolution spatiale et 4 bandes multispectrales avec 2.4m de résolution spatiale.

La donnée cartographique utilisée est le plan de ville de Rabat à l'échelle de 1/10000, il est produit en 1999 par restitution manuelle à partir des photographies aériennes. Depuis sa production, il n'a bénéficié d'aucune mise à jour. Ce plan est réparti en plusieurs couches d'informations couvrant les thèmes suivants: le couvert végétal, le relief, le réseau hydrographique, la voirie, le bâti et la toponymie.

2.3. Prétraitements

La fusion d'image permet d'exploiter en synergie la très haute résolution spatiale de la bande panchromatique et les 4 bandes multispectrales. Ce type de fusion a prouvé son intérêt dans la classification des images à THRS [PUI 03]. Notamment en milieu urbain où un grand nombre d'objets est présent sur des surfaces restreintes. La littérature est bien fournie pour les différents algorithmes permettant la fusion [PUI 03]. Dans notre travail, nous avons utilisé l'algorithme de *BROVEY* [HAL 83].

Notre approche de détection de changements est basée sur une comparaison image-plan. Dans ce cas, les deux données doivent partager un même système de référence spatial pour assurer la correspondance de leurs objets.

3. Méthodologie

La méthodologie proposée pour détecter les changements du bâti est basée sur une comparaison entre le résultat de l'extraction du bâti à partir d'une image à THRS récente et la couche du bâti existant sujet de mise à jour. La contribution principale de la méthodologie c'est qu'elle combine les avantages de l'analyse basée objet avec des méthodes de classifications avancées. La figure 1 présente les étapes de la méthodologie proposée. Chacune d'entre elles est développée en détail dans la suite de cette section.

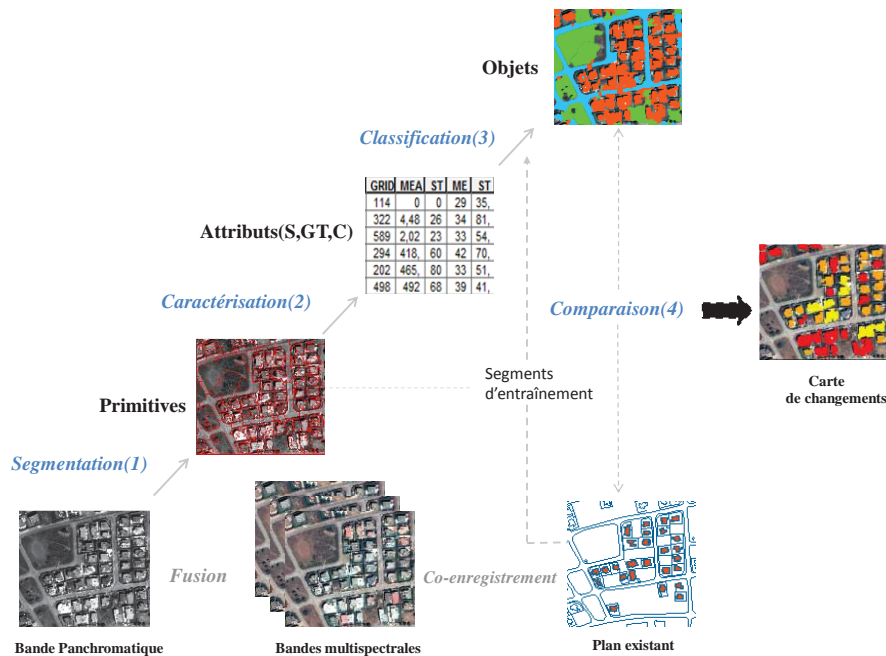


Figure 1. Méthodologie proposée pour la détection des changements du bâti.

3.1. Segmentation

Dans une classification basée objet, la segmentation de l'image est en amont du traitement. Elle permet de partitionner l'image en régions connexes et homogènes selon un critère donné. Les zones résultantes, dites primitives, sont qualifiées d'être plus intelligentes que les pixels dans le sens où elles peuvent être décrites par d'autres types d'attributs que la signature spectrale et aussi par leurs connaissances et gestion du voisinage [PUI 03].

Dans ce travail, une variante de la ligne de partage des eaux (LPE, Watershed en anglais) [BEU 90] a été développée pour segmenter la bande panchromatique. La LPE tient ses fondements de la morphologie mathématique. L'idée est de prendre l'image pour un relief topographique où les intensités des pixels sont considérées comme des altitudes. Les points à fortes élévations sont vus alors comme des crêtes qui séparent les bassins versants (les régions homogènes) créés à partir des minimas locaux de l'image.

Habituellement la LPE est calculée à partir du module de gradient, mais dans cet article, et comme le contexte est l'urbain avec des objets anthropogéniques à formes quasi régulières, nous avons pris la carte de distance comme support de calcul de la LPE.

Dans sa version originale, la LPE produit une myriade de segments dont les plus petits sont souvent dus aux variations locales non significatives du gradient. Ce problème s'accroît davantage avec la THRS. Remédier à cette sur-segmentation, revient à supprimer les petits bassins causés par les faux minima du bruit. Les vrais minima sont imposés, soit par des marqueurs, soit par un seuillage dynamique [BEU 90]. Dans notre travail, nous avons utilisé le seuillage dynamique. Le seuil est fixé en fonction du rang dynamique de la carte de distance. Les minima locaux inférieurs à ce seuil sont éliminés. Le résultat de la segmentation de l'image est présenté dans la figure 2.

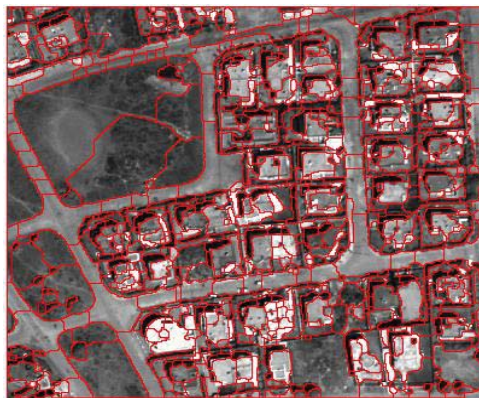


Figure 2. *Résultat de la segmentation de la bande panchromatique.*

3.2. Calcul et sélection des attributs des segments

3.2.1. Calcul des attributs des segments

Les primitives résultantes de l'étape de la segmentation sont décrites par des statistiques zonales calculées à partir d'une seule bande ou en combinant plusieurs bandes. En analysant les objets urbains sur le plan existant et les images disponibles, des indices spectraux et géométriques ont été calculés. Ces indices sont connus par leur pouvoir séparatif entre les différentes classes d'objets urbains. La texture est également utilisée. 8 indices d'Haralick, justifiés par leur grande utilisation dans la littérature, ont été considérés [HAR 73]. Les attributs d'un segment sont alors la moyenne et l'écart type des 4 bandes fusionnées et des différents indices calculés. Pour tenir compte du voisinage d'un segment, nous avons pris l'ensemble des attributs (spectraux, géométriques et texturaux) de ses voisins adjacents pour ses attributs contextuels. Le tableau 1 présente la liste des indices utilisés avec leurs formules de calcul.

Categorie	Indice	Formule
spectrale	NDVI	$(PIR-R)/(R+PIR)$
	IBS(Indice de brillance du sol)	$\sqrt{(R^2+PIR^2)}$
	IC(Indice de constructions)	$(B-PIR)/(B+PIR)$
	IO(Indice d'ombre)	$(R+V+B+PIR)/4$
Texturale	Homogénéité	Indices d'Haralick
	2nd moment	
	Corrélation	
	Entropie	
	Contraste	
	Variance	
	Moyenne	
	Dissimilarité	
Géométrique	Longueur / Largeur	
	Aire	
	Périmètre	
	Compacité	$\sqrt{(2\pi \cdot \text{Aire})}/\text{Périmètre}$

Tableau 1. Liste des indices calculés (bandes spectrales(R: rouge, V: verte, B: Bleu, PIR: proche infra rouge)).

3.2.2.Sélection quantitative des attributs des segments

Dans la plupart des méthodes de classification supervisées, choisir quantitativement les attributs pertinents pour une meilleure discrimination entre les classes, est laissé implicitement à la charge des algorithmes d'apprentissage en amont des méthodes de classification. Ces algorithmes utilisent des wrapper pour leur choix d'attributs. Les wrappers ne s'intéressent pas aux interactions entre les différents attributs [PEN 05]. Ils exploitent uniquement la contribution de l'attribut dans la prédiction de la classe de sortie pour évaluer sa pertinence. C'est ce qui rend le modèle d'apprentissage construit très collant à l'échantillon d'entrée (*apprentissage par cœur*) et ce qui justifie également son échec à la généralisation [PEN 05]. C'est pourquoi dans ce travail, nous avons utilisé un algorithme introduit sous le nom de *MrMR* : Minimum de Redondance et Maximum de Relevance pour la présélection des attributs à considérer dans l'étape de l'apprentissage. Cet algorithme sélectionne le sous-ensemble d'attributs les moins corrélés entre eux d'une part et les plus significatifs dans la détermination de la classe de sortie, d'autre part. Il offre plusieurs avantages: (1) comparé à d'autres algorithmes de sélection d'attribut, le *MrMR* maintient les attributs dans leur espace original (contrairement, par exemple, à l'analyse par composantes principales), (2) dans le *MrMR*, les calculs sont fondés sur la théorie de l'information mutuelle, connue pour sa meilleure discrimination des relations non linéaires entre les attributs et (3) il est indépendant

du classificateur choisi ensuite. Deux variantes de cet algorithme ont été utilisées: la combinaison multiplicative(MIQ) et la combinaison additive(MID). Les détails de cet algorithme et les codes utilisés sont présentés dans [PEN 05].

A noter que les deux variantes de l'algorithme ont sélectionnées en premier rang, l'indice NDVI en moyenne et en écart type, suivis de la moyenne des trois indices d'Haralick, calculés sur une fenêtre de (7*7): la variance, l'entropie et la moyenne. Dans les indices de forme, le rapport longueur/largeur est sélectionné en premier.

3.3. Classification

L'objectif de la classification est d'étiqueter chaque primitive résultante de la segmentation par un thème. De l'analyse des données disponibles, nous avons retenu cinq classes thématiques : bâtiment, route, sol avec couvert végétal, végétation urbaine, et ombre.

Une étape préalable à la classification par apprentissage supervisé a permis de détecter les segments appartenant aux classes végétation urbaine et ombre. L'indice NDVI a permis de classer binairement l'image en "végétation urbaine" et "non-végétation urbaine". Les zones d'ombre sont ensuite distinguées grâce à l'indice d'ombre (figure 3). Le reste des segments de l'image "non-ombre" sont ensuite examinés dans un processus de classification par apprentissage supervisé à partir des objets du plan existant.

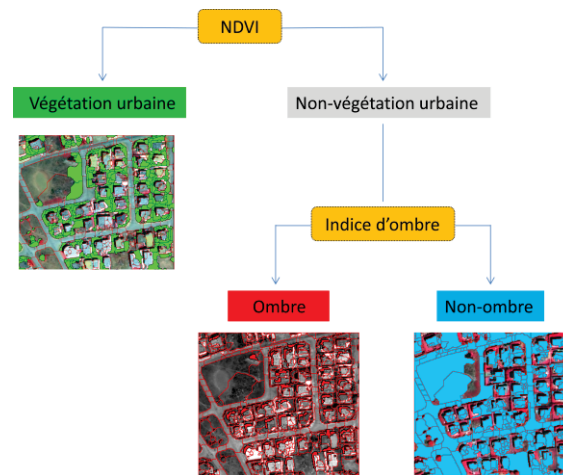


Figure 3. Détection des deux classes "végétation urbaine" et "ombre".

Les trois classes bâti, route et sol avec couvert végétal sont séparées entre elles en utilisant des classificateurs d'apprentissage supervisé automatique avancés(CASAA) où les zones d'entraînement sont sélectionnées automatiquement à partir du plan existant. En effet; pour chaque objet du plan existant, le segment (ou les segments) qui lui correspond spatialement est identifié. En règle générale, le segment est retenu pour l'apprentissage si ses propriétés géométriques (taille et forme) sont proches de celles de l'objet du plan. D'autres critères spécifiques à chaque type d'objets peuvent être ajoutés au moment de la sélection. Pour les segments d'apprentissage du bâti par exemple, l'adjacence du segment à un segment d'ombre confirme son appartenance à la classe du bâti et élimine l'hypothèse de la destruction de l'ancien bâtiment du plan.

Dans cet article, nous avons utilisé les *CASAA suivants*: les machines à supports vectoriels (*SVM*), les réseaux de neurones (*RN*) et les ensembles d'arbres de décision (*Bagging*, *Random Forest(RF)*, *Boosting*). L'avantage principal de ces classificateurs comparés aux classificateurs classiques (maximum de ressemblance par exemple) c'est qu'ils n'acquiescent aucune hypothèse sur la normalité de la distribution de données, le modèle construit sur la base de l'échantillon d'entrée est directement applicable sur tout nouveau jeu de données [MAR 05]. Les prédictions des meilleurs CASAA pour un segment seront agrégées selon un système de vote pour déduire sa classe finale.

Dans le contexte de la classification des images à THRS, Les auteurs dans [MAL 10], [FRA 05], [CAM 08] et [LEI 10] se sont servis respectivement des *SVM*, des *RN*, du *Boosting* et des *RF* pour séparer le bâti des autres classes d'objets. Dans une classification basée objet pour la détection des changements urbains, les auteurs dans [LEI 10] ont testé trois classificateurs : les plus proches voisins(*PPV*), les *SVM* et les *RF*. Ils ont confirmé que les scores des deux derniers classificateurs sont meilleurs que celui réalisé par la l'algorithme classique des *PPV*. Ce dernier souffre également d'un autre désavantage qui est la difficulté de traiter avec précision des données à grand nombre d'attributs. Une étude comparative dans [MAR 05] a fait ressortir la performance des méthodes d'ensembles d'arbres comparées à l'algorithme classique d'induction par arbre de décision, une amélioration de 16% en précision globale de la classification est atteinte par les méthodes d'ensemble d'arbres. Dans la suite de cette sous-section, nous introduisons brièvement les définitions, les paramètres et l'environnement d'implémentation des différents CASAA utilisés dans ce travail.

Les méthodes d'ensembles d'arbres consistent à améliorer l'algorithme classique d'arbre de décision en combinant la prédiction de plusieurs modèles [MAR 05]. Ces modèles sont construits sur la base des sous-échantillons générés aléatoirement à partir de l'échantillon complet d'entrée; c'est le cas du *Bagging*. La méthode du *Random forest* est une autre variante des méthodes d'ensembles d'arbres où les modèles sont construits de la même manière que le *Bagging*, cependant, durant la construction d'un arbre, uniquement un petit nombre k d'attributs est choisi aléatoirement parmi l'ensemble des attributs. Quand au *Boosting* et par opposition

aux deux méthodes précédentes, produit ses modèles de manière séquentielle et déterministe. Pour implémenter ces méthodes nous avons utilisé les codes développés par *Pierre Geurts* [GEU 02], la mesure de score utilisé pour évaluer les tests aux nœuds internes des arbres (*splitting criterion*) est l'indice de GINI . Le paramètre principal dans ces méthodes est T le nombre d'arbres développés.

L'idée de base dans la classification par *SVM* est de produire une frontière de décision linéaire, dite hyperplan pour chaque paire de classe. L'hyperplan optimal est défini de telle sorte que la distance qui le sépare des points les plus proches, qu'on appelle les vecteurs de support, est maximisée. Dans notre travail, nous avons utilisé les *SVM* à noyau radial de l'outil LibSVM3.11, implémenté sous Matlab [<http://www.csie.ntu.edu.tw/~cjlin/libsvm>]. Ici, les paramètres importants sont gamma de la fonction du noyau et le coût.

Le principe des *RN* est d'entraîner sur un échantillon d'apprentissage (entrée-sortie) des couches de neurones interconnectés, tout en ajustant itérativement les poids entre elles de manière à minimiser la différence entre la sortie prédite et la sortie donnée dans l'échantillon de l'apprentissage. L'outil *Neural Network Toolbox* [<http://www.mathworks.nl/products/neural-network/>] fournit par Matlab a été exploité pour implémenter ce classificateur. Trois paramètres sont à déterminer : le nombre de couches cachées, la fonction d'activation et le nombre d'époques.

A l'exception du paramètre *k* dans les *RF*, qui a été fixé à la racine carré du nombre total d'attributs, tous les autres paramètres des différents classificateurs ont été déterminés automatiquement par validation croisée sur l'ensemble de l'échantillon d'apprentissage en se référant au meilleur taux d'erreur.

3.4. Détection des changements du bâti

Cette étape est basée sur la comparaison automatique en mode vecteur du résultat de la classification du bâti et de la couche du bâti de 1999. La comparaison est basée sur une correspondance spatiale. Nous partons de l'hypothèse que la précision de positionnement entre le plan et l'image est suffisante dans le sens que la correspondance (*matching*) entre les polygones des deux données pourrait être accomplie sur la base d'un chevauchement spatial. Des méthodes similaires ont été utilisées dans autres études de détection de changements [MAY 08]. Dans les règles prévues pour la détection de changements de bâtis, des marges d'erreurs sont retenues pour tolérer certaines différences de localisation ou de forme qui sont dues principalement (1) au défaut de co-enregistrement entre l'image et le plan, (2) à la propagation des erreurs à partir des attributs et (3) également à la différence intrinsèque dans la nature des objets des deux données (image et plan). Les différences qui sont plus importantes sont traduites comme des changements ou des erreurs de détection nécessitant d'autres mesures dans la procédure de détection de changements.

En se basant sur le pourcentage de chevauchements en surface d'une paire de polygones, l'un appartient à la couche du bâti extraite et l'autre appartient à la couche du bâti existante, nous avons définis quatre classes de changement de bâtis:

- Nouveau bâtiment (*classe1*): un polygone de la couche du bâti détecté ne correspond à aucun polygone de la couche du bâti existante.
- Bâtiment démoli (*classe2*): un polygone de la couche existante du bâti ne correspond à aucun polygone de la couche du bâti détecté.
- Ancien bâtiment confirmé (*classe3*): un polygone de la couche du bâti détecté correspond à un polygone de la couche du bâti existante, avec la vérification des deux inégalités de surfaces suivantes [1] et [2]:

$$100 * \text{Aire_chevauchement} / \text{Aire_bâti-plan} > \text{Seuil} \quad [1]$$

et

$$100 * \text{Aire_chevauchement} / \text{Aire_bâti-détecté} > \text{Seuil} \quad [2]$$

- Ancien bâtiment élargi (*classe4*): un polygone de la couche du bâti détecté correspond à un polygone de la couche du bâti existante avec la vérification uniquement de l'inégalité de surfaces [1], l'inégalité [2] est non vérifiée. Dans ce cas, l'ancien bâtiment existe mais son correspondant dans le résultat de la détection de bâti est plus large.

$$100 * \text{Aire_chevauchement} / \text{Aire_bâti-plan} > \text{Seuil} \quad [1]$$

et

$$100 * \text{Aire_chevauchement} / \text{Aire_bâti-détecté} < \text{Seuil} \quad [3]$$

Dans notre cas, le seuil est fixé à 60 %. Dans [MAT 10] les auteurs proposent d'adopter le seuil de 80% pour les zones industrielles, 70% pour les zones résidentielles et 50% pour les zones à petites maisons familiales. Les petits seuils sont intéressants si l'objectif est de mesurer la performance de la méthode dans la détection des bâtiments sans trop se soucier de la qualité de leur taille.

4. Résultats, validation et discussions

4.1 Résultats, validation et discussions: la classification

Tout d'abord les différents classificateurs utilisés ont été évalués. Dans le cas d'une classification basée objet, les attributs utilisés pendant la classification sont calculés sur des segments; par conséquent, les entités de la validation doivent également être des segments et non pas des pixels. 99 segments sélectionnés par échantillonnage stratifié et annotés par interprétation visuelle à partir de l'image ont constitués l'échantillon de validation. La précision globale, le coefficient kappa et son écart type ont été calculés [CON 91]. Le tableau 2 présente le meilleur résultat obtenu par chaque classificateur sur la base de 47 attributs sélectionnés par MrMR.

Classificateur	Précision globale %	Kappa %	σ_{Kappa}
Bagging (T=50)	73.56	64.21	2.16
Boosting (mu=0.2, comp=5)	74.76	64.98	3.27
RF (T=200)	82.85	72.21	0.03
SVM (radial, c= 10, γ = 0.00001)	83.03	72.41	0.11
RN (30 couches cachées, Sgd)	82.86	72.27	0.15

Tableau 2. Meilleurs (Précision globale, kappa et σ_{Kappa}) des différents CASAA utilisés.

Les trois classificateurs: *SVM* (radial), *RN* et *RF* présentent de meilleurs scores avec une légère supériorité des *SVM*. En effet; ces derniers n'utilisent qu'une petite partie de l'échantillon d'entraînement pour construire leur modèle d'apprentissage ce qui pourrait justifier leur succès dans notre cas, où le nombre de segments d'entraînement est réduit. A partir d'un certain nombre d'arbres développés, Les *RF* assurent au mieux la relation biais/variance [GEU 06] ce qui explique la stabilité de leurs résultats (σ_{kappa} = 0.03 à partir de 200 arbres développés). Quand aux *RN*, ils sont théoriquement moins sensibles au bruit ce qui les rend adaptables à la classification des images à THRS.

Pour exploiter conjointement les avantages de ces classificateurs et vu la similarité de leur résultats, nous avons décidé de n'en pénaliser aucun et de combiner leurs résultats dans un système de vote. La prédiction qui résulte alors de l'agrégation est sûrement plus stable que la prédiction à partir d'un seul classificateur. Le système de vote adopte la prédiction la plus présente. Dans le cas de conflits entre les trois classificateurs, le système compare les probabilités des différentes prédictions et opte pour la plus élevée; dans le cas échéant où ces probabilités sont égales, la prédiction donnée par les SVM est privilégiée. Un autre

avantage d'un tel système est sa capacité à doter la prédiction de chaque segment par un degré de confiance. Un segment prédit dans la même classe par les trois classificateurs a un degré de confiance de classification de 100%. Ce facteur pourrait être utile pendant la supervision des résultats de détection de changements. La figure 5 montre les résultats de la classification.

La matrice de confusion calculée sur la base des prédictions du système de vote est présentée dans la figure 4. Elle montre qu'avec les attributs sélectionnés par *MrMR* une bonne distinction entre le bâti et les routes est assurée. La confusion entre les 2 classes bâtis et sol avec couvert végétal est à l'origine de la minimisation de la valeur de kappa. Ceci est dû principalement (1) aux défauts de la segmentation, certains segments sont composés d'une mixture de pixels qui présentent des attributs non significatifs conduisant par la suite à des classifications erronées et (2) à la date d'acquisition de l'image (Août) qui n'est pas convenable pour séparer le bâti et le sol avec couvert végétal au Maroc. Des règles post-classificatoires sont prévues dans l'étape de détection de changements pour réduire ces confusions.

		Référence								
		B	R	S	Total ligne	Précision Producteur	Omission %	Précision Globale %	Taux Erreur %	Kappa %
Classification	B	44	1	0	45	97,8	2,2	82,8	17,2	72,3
	R	3	21	0	24	87,5	12,5			
	S	12	1	17	30	56,7	43,3			
	Total Colonne	59	23	17	99	80,6				
	Précision Utilisateur %	74,6	91,3	100	88,6					
	Commission %	25,4	8,7	0						

Figure 4. Matrice de confusion de la classification par système de vote (B: Bâti, R: Route, S: Sol avec couvert végétal).



Figure 5. Résultats de la classification (■ Bâti ■ Route ■ Sol avec couvert végétal).

4.2 Résultats, validation et discussions: détection des changements du bâti

Appliquer les règles de détection de changements directement sur la couche du bâti résultante de la classification mène à une surestimation dans les changements du bâti. Des règles additionnelles ont été prises pour contraindre la sur-détection de changements du bâti. Les transitions possibles entre les classes d'objets et les relations contextuelles (proximité, adjacence) entre les objets de différentes classes ont été exploitées. Egaleme nt des opérations morphologiques (ouverture et fermeture de surface) ont permis d'enlever les trous dans les bâtiments détectés.

Pour valider les résultats de détection de changements de bâti, une carte de changements établie par interprétation visuelle de l'image a constitué une référence. Elle est comparée à la carte de changements de bâti produite par notre approche. La comparaison est faite de la même manière qu'à l'étape de la détection de changements. Tout chevauchement spatial supérieur à 60% entre, un bâtiment de la carte de référence et son correspondant dans les résultats de détection des changements et qui sont étiquetés de la même classe, conduit à considérer que la détection du bâtiment est correcte. Les figures 6a et 6b montrent respectivement les résultats de la détection de changements de bâtis par notre approche et la carte de changements du bâti de référence. Le Tableau 3 présente les résultats des changements de bâti détectés par notre approche comparé aux changements du bâti de la carte de référence.



Figure 6a. Carte de changements de bâti détectés par notre approche



Figure 6b. Carte de changements de bâti de référence

(■ Nouveau bâtiment ■ Ancien bâtiment confirmé ■ Bâtiment élargi)

		Référence				
		Nouveau	Elargi	Confirmé	Démoli	Total
détection de changement	Nouveau	19	0	0	0	19
	Elargi	3	1	5	0	9
	Confirmé	0	0	20	0	20
	Démoli	0	0	0	0	0
	Non détectés	2	0	0	0	2
	Total	24	1	25	0	50

Tableau 3. Résultats des changements de bâti détectés par notre approche comparé aux changements du bâti de la carte de référence.

Le tableau présente les résultats de la détection de changement. La comparaison des résultats avec la carte de référence montre que sur les 50 bâtiments de la carte de référence 48 bâtiments sont détectés, soit un taux de complétude (*completeness*) [HEP 97] de 95%. 40 bâtiments parmi eux sont déterminés avec une bonne identification de la nature de changement, soit 83% de bonne détection (*correctness*). Les 8 autres bâtiments sont détectés mais avec une fausse identification de la nature de changement; 5 anciens bâtiments confirmés et 3 nouveaux bâtiments dans la carte de référence ont été détectés comme des bâtiments élargis. En effet; cette confusion résulte de la classification des segments pourtour des bâtiments (entrées de garage, cours de maison, passages bitumés, murs séparateurs) comme du bâti, ce qui introduit un élargissement dans l'emprise des bâtiments détectés arrivant même à rassembler certains bâtiments(détachés entre eux dans la réalité) en un seul bloc

La classe la plus fréquente, celle des nouveaux bâtiments (48%) est correctement détectée à 87.5%. Ce taux est jugé suffisant pour répondre à notre objectif de recensement des nouveaux bâtiments. Les anciens bâtiments confirmés qui n'ont pas changé présentent 50% de la totalité des bâtiments, ils sont détectés correctement avec un taux de 80%, ce qui présente une garantie pour l'opérateur, lors de la mise à jour, s'il se restreint uniquement sur les zones de changements.

Par ailleurs, 2 bâtiments non pas été détectés, ce qui présente une erreur d'omission de 4%. Les segments composants ces bâtiments sont hétérogènes, contiennent des mixtures (ombre, bâti et végétation). Ce qui rend leurs attributs différents des attributs pris pour l'apprentissage des bâtiments. L'hétérogénéité des segments vient également de la présence de grandes antennes paraboliques sur les toits des bâtiments au Maroc.

Enfin, nous soulignons que l'absence de bâtiments démolis n'est pas choquant, puisqu'il s'agit d'une zone résidentielle en plein croissance.

5. Conclusion et travaux en cours

Nous avons présenté une approche pour la détection de changements de bâti. Les premiers tests sur des données de Rabat(Maroc) donnent des résultats prometteurs. Le taux de bonne détection est de 83%. Il est comparable à d'autres méthodes développées dans la littérature [BOU 07, CHA 09, LEI 10]. L'idée de base et les éléments de cette approche, notamment la classification par un système de vote, sont solides et à caractère générique. Des tests sur d'autres morphologies urbaines au Maroc sont envisagés.

L'erreur typique nécessitant d'autres mesures dans le traitement est la connexion entre les bâtiments qui ne sont pas attachés réellement entre eux. L'origine de cette erreur, est la non distinction entre les constructions et les surfaces bitumées qui les entourent, à l'intérieur de la classe du bâti. A présent, nous sommes concentrés sur la résolution de ce problème. Une révision du résultat de la classification s'avère indispensable.

6. Bibliographie

- [BAA 00] Baatz M., Schäpe A., Multiresolution Segmentation – An optimisation approach for high quality multi-scale image segmentation. 2000.
- [BAL 04] Baltsavias E., Object extraction and revision by image analysis using existing geodata and knowledge: Current status and steps towards operational systems. ISPRS Journal of photogrammetry and Remote Sensing 58, p.3-4, 129-15, 2004.
- [BEU 90] Beucher S., segmentation d'images et morphologie mathématique. Thèse de doctorat, Ecole des mines de Paris , 1990.
- [HAL 83] HALLADA W.A., Image sharpening for mixed spatial and spectral resolution satellite systems. Proc. 17th Int.Symp. Rem. Sens. Env, p.1023–1032, 1983.
- [BOU 07] Bouziani. M., Détection automatique des changements de bâti en milieu urbain sur des images à très haute résolution spatiale en utilisant des données cartographiques numériques, Thèse de doctorat, Université de Sherbrooke, 2007.
- [CAR 06] Carleer A., Wolff E., Urban land cover multi-level region-based classification of VHR data by selecting relevant features. International Journal of Remote Sensing 27 (6), p. 1035-1051, 2006.
- [CAM 08] Camps G., González A., Speech, Audio, Image and Biomedical Signal Processing using Neural Networks Studies in Computational Intelligence, Vol. 83/2008, 2008.
- [CHA 09] Champion N., Detection of unregistered buildings for updating 2D databases. EuroSDR Official Publication N°56. Asmterdam: Gopher, 2009.
- [CON 91] Congalton R, A review of assessing the accuracy of classifications of remotely sensed Data. Remote Sensing of Environment, 37(1), p. 35-46, 1991.

- [FRA 05] Frate F., Change detection in urban areas with quickbird imagery and neuralnetworks algorithms, 2005.
- [GEU 02] Geurts P., Contributions to decision tree induction. Bias/variance tradeoff and time series classification. Thèse de doctorat. Université de Liège, 2002.
- [HAR 73] Haralic M., Shanmugan K., and DisteinI., Textural features for image classification. IEEE Transactions System, Man and Cybernetics, vol. SMC-3, p. 610-621,1973.
- [HER 11] Hermozilla T.,Evaluation of Automatic Building Detection Approaches Combining High Resolution Images and LiDAR Data, Remote Sens., 3, p.1188-1210, 2011.
- [HOL 06] Holland D., Boyd D., Marshal P., Updating topographic mapping in GreatBritain using imagery from high-resolution satellite sensors. ISPRS Journal of Photogrammetry and Remote Sensing 60 (3), p. 212-223, 2006.
- [KNU 03] Knudsen T., Olsen B., Automated change detection for updates of digital map databases. Photogrammetric Engineering & Remote Sensing 69(11), p. 1289–1296, 2003.
- [LAN 94] Langley P., Selection of Relevant Features in Machine Learning, Proc. AAAI Fall Symp. Relevance, 1994.
- [LEI 10] LEIGNEL C., Detecting Man-made structure changes to assist geographic data producers in planning their update strategy. ISPRS vol. XXXVIII, Part 4-8-2-W9, Haifa, Israel, 2010.
- [MAL 10] Malpica J., Alonso M., Urban changes with satellite imagery and lidar data. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science, Vol. XXXVIII, Part 8, 2010.
- [MAR 05] MAREE R., Classification automatique d'images par arbres de décision, Thèse de doctorat, Université de Liège, 2005.
- [MAT 10] Matikainen L., Automatic Detection of Buildings and Changes in Buildings for Updating of Maps. *Remote Sensing*. 2, p. 1217-1248; 2010.
- [MAY 08] Mayer H., Objects extraction in photogrammetric computer vision. ISPRS Journal of Photogrammetry and Remote Sensing, 63(2), p. 213-222, 2008.
- [PEN 05] Pen H., Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-relevance, and Min-Redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, N° 8, 2005
- [PUI 03] Puissant A., Information géographique et image à très haute résolution: utilité et application en milieu urbain, Thèse de doctorat. Université de Strasbourg, 2003.
- [WAL 04] Walter V., Object-based classification of remote sensing data for change detection. ISPRS Journal of Photogrammetry and Remote Sensing 58, p. 3-4, 225-238, 2004.

Vers une géospatialisation de l'éclairage public au Gabon

Makak Jean Sylvestre.^{1*}, Thomas Daniel.^{2**}, Palla Florence.^{3***},
Emane Etienne.^{4****}, Ogowet Mboyo Hélène Elody Marcelle.^{5*****}

* Directeur Général de GEOCOM, BP 13392 Libreville-Gabon, jeanmakak@gmail.com

**Consultant GEOCOM, BP 13392 Libreville-Gabon, dthomas@geovalor.com

***Membre associée GEOCOM, BP 13392 Libreville-Gabon, pallaflorence@hotmail.com

****Directeur Technique du CNEE, BP 20498 Libreville-Gabon, etienne.emane@gmail.com

*****Responsable de la cellule SIG du CNEE, BP 20498 Libreville-Gabon, ogelody@hotmail.com

RÉSUMÉ.

*Les systèmes mobiles de cartographie sont une bonne opportunité pour répondre au besoin grandissant de collecte d'informations géographiques en Afrique subsaharienne. Ils sont constitués d'un dispositif mobile équipé de capteurs de localisation et de perception. Parmi eux, les systèmes terrestres à laser permettent d'effectuer des relevés laser mobiles en environnement urbain, à une vitesse beaucoup plus importante que les systèmes fixes. Dans cet article, nous en présentons les principes et les résultats préliminaires d'un projet en cours au Gabon, basé sur l'acquisition des données par **camera-gps** des réseaux d'éclairage public pour la mise en place d'un « système d'information géographique » (**sig**) et des outils d'aide à la décision.*

ABSTRACT.

*The **mobile systems of mapping** are a good opportunity to meet the growing need for collecting geographic information in sub-Saharan Africa. They consist of a mobile device equipped with sensors location and collection. Among them, the terrestrial laser systems to perform mobile laser surveys in urban environments, at a speed much greater than fixed systems. In this article, we present the principles and the preliminary results of a project underway in Gabon, based on the data acquisition by **camera-gps** of lighting public networks for the establishment of a « geographic information systems » (**gis**) and tools of decision support.*

MOTS-CLÉS : Gabon - Eclairage public- Camera-gps- SIG - Aide à la décision

KEY WORDS: Gabon - Lighting public - Camera-gps - GIS - Decision support

1. Introduction

Depuis quelques années, on observe un fort développement des « systèmes d'information géographique » (**sig**), popularisés auprès du grand public par les bases de données routières embarquées ou la navigation virtuelle dans des villes en **3d**, mais aussi tirés par d'autres usages [GOU09]. Rappelons qu'une base de données cartographiques collecte des informations structurées de façon géographique, et qu'un **sig** est un système informatique permettant de gérer ces informations : interrogation, visualisation, renseignement de la base, etc. De fait, ce type de bases de données apporte des avantages pour de nombreuses applications [GOU09].

Rappelons aussi que les gestionnaires et les décideurs urbains sont divers et leurs besoins en information sont par conséquent variés, car la décision en milieu urbain implique la prise en compte de plusieurs facteurs, notamment la configuration du site, le foncier, le logement, les équipements, les infrastructures et les problèmes sociaux ou économiques [ASS07]. L'échec de nombreuses politiques de planification urbaine en Afrique se justifie en grande partie par le manque d'une information de qualité, permettant d'alimenter les processus décisionnels [ASS07].

Le Gabon est un pays francophone d'Afrique centrale qui couvre officiellement une superficie de 267 670 km² à cheval sur l'équateur géographique et juste en dessous de l'équateur climatique. Avec 1,3 million d'habitants en 2005, la densité moyenne de population est de 5 habitants par km². Il existe une forte disparité entre le milieu rural, peu peuplé et enclavé, et les zones urbaines où la population est essentiellement regroupée (85% de la population totale) [MAD10], [BUR07]. Face à cette forte disparité, les défis en termes de maîtrise d'équipements d'éclairage et de fourniture d'eau à usage public dans les quartiers et le long des axes de circulation restent énormes.

En effet, L'éclairage public est l'ensemble des moyens d'éclairage mis en œuvre dans les espaces publics, à l'intérieur et à l'extérieur des villes, très généralement en bordures des voiries et places, nécessaires à la sécurité ou à l'agrément de l'homme [WEBWIKIPEDIA].

Le conseil national de l'eau et d'électricité (**cnee**) est un établissement public, créé en 1995 comme organe consultatif, juridictionnel et exécutif pour assurer au nom et pour le compte de l'Etat gabonais la création, la gestion des réseaux d'éclairage public et d'installations des collectivités locales alimentées en eau potable et en électricité.

Sur le plan opérationnel, le **cnee** fonctionne comme un département du Ministère chargé du Pétrole, de l'Energie et des Ressources Hydrauliques. A ce titre, il est notamment chargé de décider à la création des réseaux d'eaux et d'éclairage public, d'assurer les dépenses liées au fonctionnement de ces réseaux et de procéder aux arbitrages en matière d'extension des réseaux à l'intérieur des collectivités locales desservies par le service public. L'éclairage public fait donc partie de son environnement quotidien et la maintenance préventive, corrective et sécuritaire des

installations reste une préoccupation majeure pour les acteurs concernés. Le budget du **cnee** est évalué à plus de **fcfa 5 milliards** soit environ 8 millions d'euros.

Toutefois, des insuffisances organisationnelles significatives existent au **cnee** et entravent la réalisation efficace de ses missions : absence de procédures formalisées, insuffisance du personnel, contrôles clés non identifiés et non réalisés, absence d'éléments de reportings et d'outils de pilotage. A titre d'exemple, les informations recueillies dans les rapports techniques des sous-traitants privés en charge des opérations de maintenance et de réalisation infrastructures, indiquent seulement les noms des quartiers bénéficiaires des infrastructures réalisées ou qui ont font l'objet de maintenance. Les factures payées par le **cnee**, se font sur la base d'une estimation du nombre d'équipements. Par contre, aucune coordonnée géographique et aucun plan de situation des infrastructures ne sont fournis, qui permettraient au client de faire un suivi rigoureux des opérations de terrain.

C'est face à cet imbroglio qu'un contrat d'assistance technique entre le **cnee** et le cabinet conseils geospatial company (**geocom**) a été signé le 16 janvier 2012, relatif à la mise en place d'un outil cartographique opérationnel d'aide à la décision, en matière de gestion des équipements spatialisés à grande échelle des infrastructures d'eau et d'éclairage à usage public, dont le **cnee** a la charge, pour un meilleur pilotage des interventions dans ce secteur. A travers ce projet, nous cherchons à répondre aux questions suivantes :

- quels types d'outils et quelles données pour améliorer le suivi, la maintenance et faciliter la prise de décisions des gestionnaires publics et privés du secteur de l'eau et d'électricité en Afrique ?
- la géomatique peut-elle contribuer à une mise en œuvre durable de projets de planification des réseaux d'éclairage public et d'eau dans les villes et villages du Gabon ?

Ainsi les objectifs spécifiques sont formulés en s'appuyant sur la réalité de terrain pour proposer des éléments de réflexion prospectifs en particulier :

- proposer des alternatives aux problèmes existants à une structure opérationnelle de collecte des données ;
- mesurer l'impact des outils utilisés pour la mise en place d'un système d'information géographique pertinent et capable d'améliorer le suivi du système d'éclairage et d'eau au Gabon.

Dans la section 2, nous présentons notre méthode d'acquisition automatique des données et leur localisation dans un logiciel SIG. Nous montrons dans la section 3 comment les données localisées peuvent être exploitées pour orienter les opérations de maintenance et d'extension du réseau. Ensuite, nous proposons une discussion sur l'intérêt de ces systèmes dans la section 4. Enfin nous concluons et présentons des perspectives.

2. Méthodologie

2.1. Analyse et formalisation des besoins

Un outil informatique est destiné à apporter une assistance dans l'exécution de tâches spécifiques, observées, existantes ou à développer.

Dans cette perspective, l'analyse, le recensement des besoins existants ou à prévoir constituent un fondement incontournable de la mise en place d'un système d'information.

Cette phase commence par le recensement des demandes exprimées par les services et l'ensemble des acteurs. C'est aussi l'occasion de préciser l'utilité d'un **sig** et d'évaluer les besoins des utilisateurs. L'analyse des besoins peut également être effectuée à partir de la description des applications souhaitées par un agent mais également à partir d'entretiens auprès des utilisateurs [VAL05].

Au cours des entretiens que nous avons eus avec la direction technique du **cnee**, nous avons constaté que le groupe des services de cette institution estime important d'avoir une bonne connaissance de leur patrimoine (Candélabres, Armoires, bornes fontaines...) à l'échelle du pays, de suivre de manière régulière les activités de maintenance dudit équipement, d'étendre le réseau vers de nouvelles localités. Il désire utiliser le **sig** pour aider ses services techniques à suivre les recommandations contenues dans les différents documents officiels tels que la loi 14/95 ou celles des audits (voir **annexe 2**).

Pour cette première phase pilote, l'objectif assigné était de pouvoir démarrer une phase de localisation rapide des équipements à Libreville (capitale du Gabon), et de faire une démonstration d'analyses spatiales basée sur des requêtes simples et la production des cartes d'aide à la décision. Le défi était également de prévoir l'appropriation du projet par le **cnee**, en s'appuyant sur une équipe de deux personnes avec une expérience limitée en géomatique.

2.2. Logistique

- Matériel :
 - 4 ordinateurs portables et 3 ordinateurs fixes (permanents),
 - 3 ordinateurs (ponctuels),
 - 6 caméras contours gps,
 - 12 batteries pour camera-gps,
 - 6 Adaptateurs des cartes mémoires pour le transfert des données,
 - 3 disques durs (2 téra, 1 téra, 750 giga),
 - 6 cartes mémoires de 16 giga,
 - 6 cartes mémoires de 2 giga,
 - Un véhicule land cruiser immatriculation: **av-355-aa**,
 - Une connexion internet **ipi9 512 mega**,
 - 2 imprimantes,
 - 2 casques de sécurité...

- Logiciels :
 - dashware,
 - contour storyteller,
 - microsoft office (power point, excel)
 - qgis
 - arcgis
 - veremap.pro et veremap.admin

2.3. Démarche méthodologique

La démarche méthodologique utilisée répond à l'architecture de création d'une base de données « **urbasig** » dont les principales étapes sont les suivantes [MAK12] :

- acquisition des données par caméra-gps ;
- analyse du film et sélection de points d'intérêt thématiques (poi) en fonction des besoins ;
- localisation des poi sous arcgis (**poi**: candélabres, bornes fontaines, armoires/postes d'alimentation...) ;
- préparation d'une base de données cartographique.

La figure 1 présente l'architecture simplifiée des étapes utilisées.

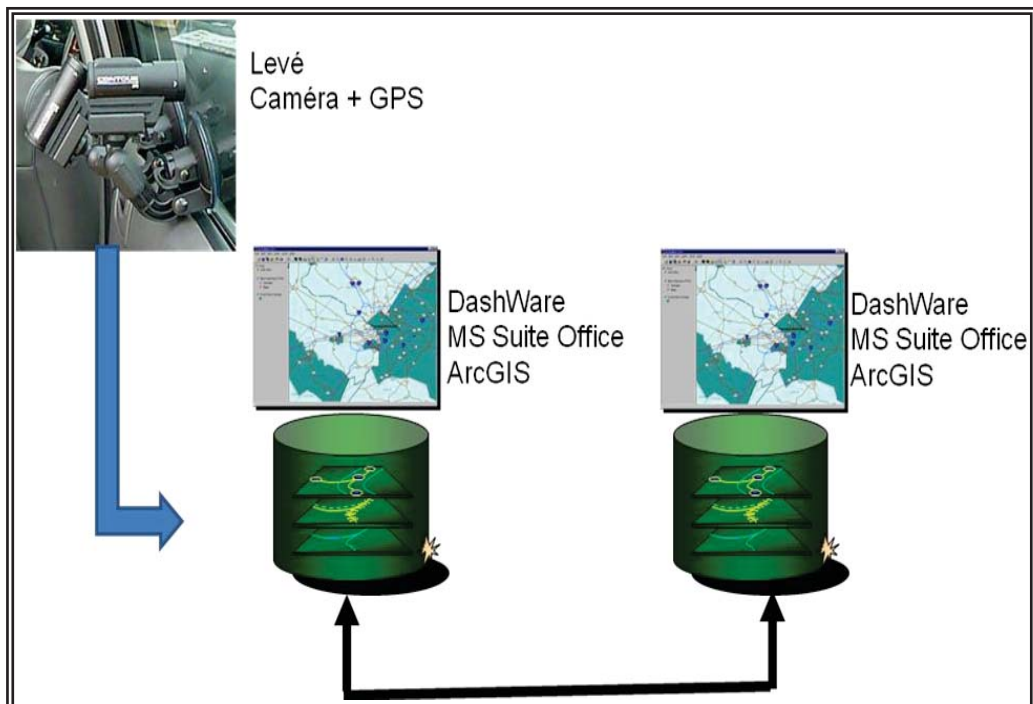


Figure 1. Architecture simplifiée des principales étapes utilisées

L'ensemble des activités réalisées intègre aussi bien le volet formation (initiation à la cartographie, conduite d'un projet **sig** et réalisation des levées sur le terrain, prise en main du webmapping et du logiciel **arcgis**) que la mise en place d'une Base de données cartographique pilote. Les lignes qui suivent présentent les principales tâches réalisées dans le cadre des relevés mobiles des données, l'analyse et la localisation des points d'intérêts (**poi**) et la construction d'une base de données cartographique pilote nécessaire pour la mise en place d'un **sig** pour le **cnee**.

2.3.1. *Relevés mobiles des données*

Le principe des scanners laser pour le relevé d'informations tridimensionnelles est connu depuis plusieurs années dans le monde de la recherche et de nombreux systèmes commerciaux sont déjà disponibles [GOU09]. En Afrique subsaharienne en général et dans la sous région Afrique centrale, il s'agit encore d'un système nouveau, d'un outil de luxe.

La phase préliminaire consiste évidemment à préparer le véhicule pour accueillir la batterie d'équipements nécessaires à l'acquisition des données et à leur enregistrement [WEBTOPCON]. Il existe plusieurs versions d'équipements :

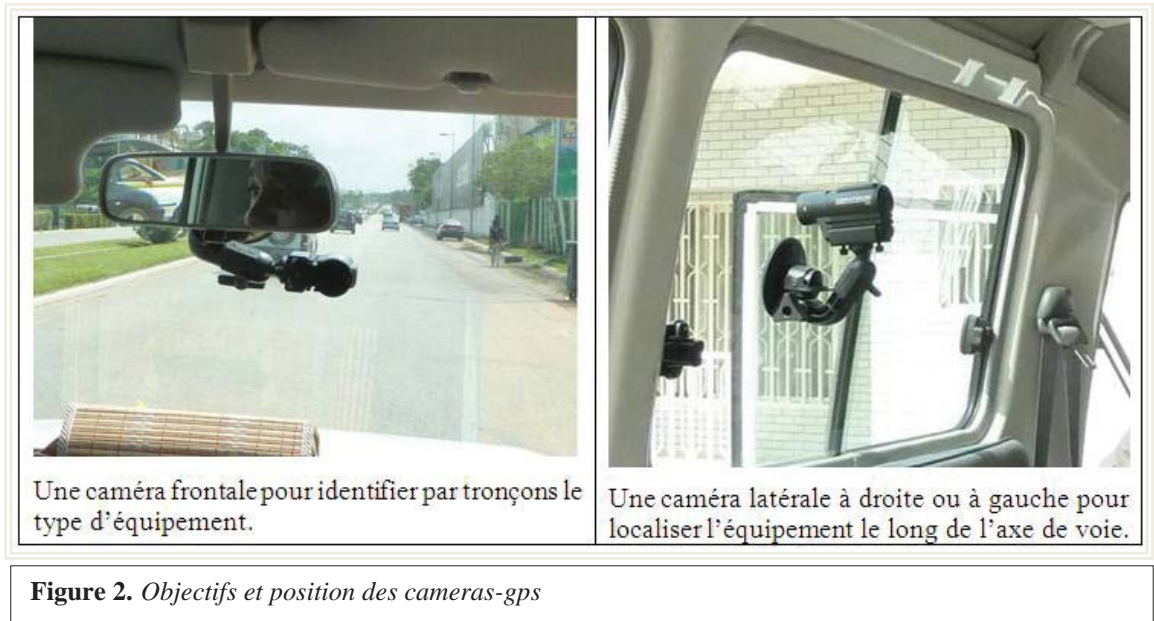
- **Topcon** par exemple possède deux versions de son produit dédié aux relevés mobiles : la station standard **ip-s2** pourvue d'un récepteur gps, d'une centrale inertielle, de trois faisceaux laser de télémétrie (d'une portée typique de trente mètres), d'une caméra panoramique « coccinelle » et d'un odomètre – ce dernier ne participe pas directement à la collecte des données, il sert uniquement à contrôler le déroulement de l'acquisition (on n'acquiert pas de données quand le véhicule stationne) ; la station haut de gamme **hd** est équipée d'un télémètre à 64 faisceaux qui collecte environ un million et demi de points par seconde, et de caméras améliorées [WEBTOPCON].
- D'autres systèmes moins complexes, moins coûteux et faciles d'utilisation existent sous forme de camera vidéo couleur qui contient un récepteur gps. Ils sont capables de collecter 30 points par seconde [MAD08], [ELL01]. Dans le cadre de ce travail nous avons utilisé un modèle qui rentre dans cette gamme, c'est la caméra **contourgps** équipée d'un gps intégré qui vous permet de retracer vos itinéraires, votre vitesse et votre altitude. Cet outil vous permet de collecter en fonction des options 30 à 60 images par seconde [WEBLARA], [WEBNOM10]. Le modèle utilisé est présenté en **annexe 3**.

Sur le plan opérationnel, nous avons effectué deux tâches la préparation du levé et sa mise en œuvre sur le terrain. Toutes les étapes ont été consignées dans un guide fourni au **cnee**.

- **Préparer un levé**

Dans la mesure du possible, il est important de préparer le levé, pour pouvoir recueillir dans les meilleures conditions, les informations qui répondent aux objectifs. Dans le travail réalisé, l'objectif était de localiser et caractériser les

équipements de l'éclairage public. Cet objectif s'est traduit par le choix du paramétrage et de la position des caméras. La figure 2 illustre la position retenue pour chacun des objectifs



C'est la position de la caméra latérale qui nécessite le plus grand soin. Si la durée du levé dépasse 2 heures, il faut penser à prendre une batterie de secours par caméra.

- **Effectuer un levé par caméra + gps embarquée dans un véhicule**

Pour démarrer une séquence, il est préférable que le véhicule soit arrêté. Le ou les opérateurs du levé doivent démarrer l'enregistrement des caméras de manière simultanée. Cela permet ensuite de pouvoir facilement coordonner l'analyse des séquences en se basant sur le temps affiché sur la vidéo.

Afin d'éviter de traiter des vidéos trop longues, il est recommandé d'arrêter une séquence au bout de 10 à 15 minutes, si possible à un carrefour, pour faciliter la logique de la séquence. De la même façon que pour le démarrage d'une séquence, le ou les opérateurs doivent arrêter l'enregistrement des caméras de manière simultanée.

Après test, nous avons proposé une méthode d'installation des caméras suivant les cas de figure sur le terrain. C'est ce qui ressort sur la figure 3.

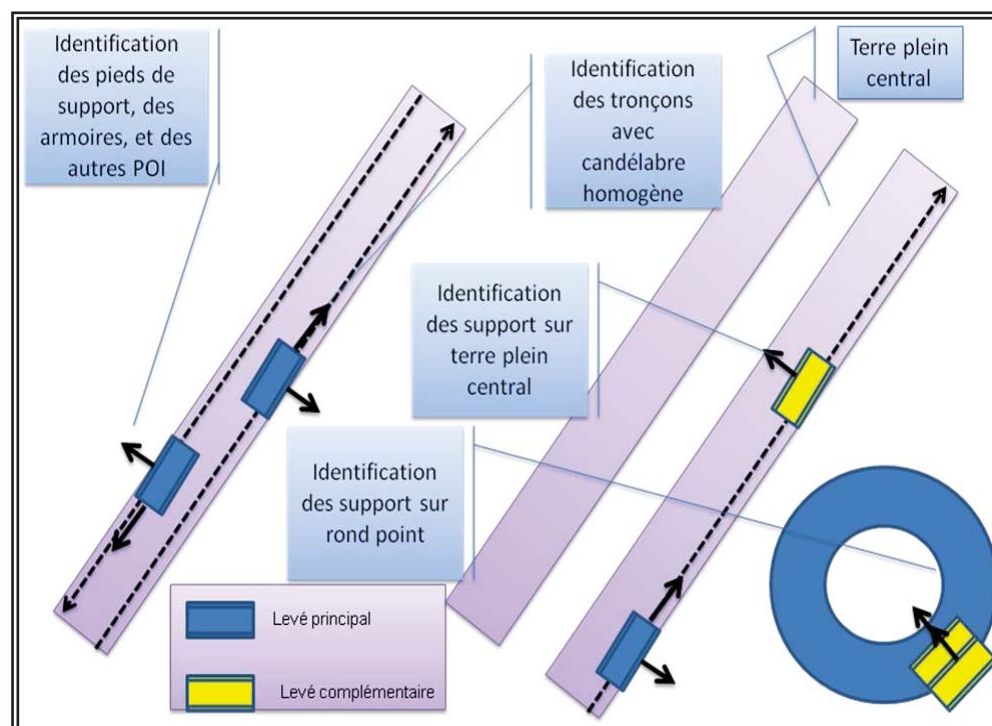


Figure 3. Proposition d'utilisation des cameras

2.3.2. Analyse d'une séquence

Chaque séquence est analysée en visualisant avec dashware son fichier vidéo et en sélectionnant des images correspondant à des points d'intérêt (**poi**).

– Préparation d'une analyse

Il faut créer un répertoire sur le disque qui permet une indépendance pour analyser la séquence. Ce répertoire contient le fichier vidéo à analyser ainsi que les fichiers contenant les résultats de l'analyse. Comme il peut y avoir plusieurs analyses d'une même séquence, l'analyse est dénommée en ajoutant au nom de la séquence un numéro d'analyse, par exemple **l00nc1_d_00nn_01**.

Créer un sous répertoire dans \02ana, dénommée du nom complet de la séquence en ajoutant 001 pour la première analyse. Par exemple le nom **l00nc1_d_00nn_01**, correspond à la première analyse du levé L00n, caméra C1, prise de vue latérale à droite, pour la séquence la séquence 00nn.

Copier les fichiers PPT et Excel du répertoire **02ana\template**, dans le répertoire de l'analyse, et renommer les fichiers du nom complet de l'analyse.

– Sélection des prises de vue

Il s'agit d'une séquence latérale permettant d'identifier directement des points d'intérêt sans autre aide. Par exemple des pieds de candélabres, sans ambiguïté. L'analyste utilise le logiciel Dashware pour visualiser le film vidéo de la séquence à étudier [THO12].

- A- L'analyste parcourt la vidéo et fait un arrêt sur image correspondant un point d'intérêt. Les commandes permettent de revenir ou d'avancer quelques images pour positionner la prise de vue correctement.
- B- L'analyste fait une copie de l'écran en utilisant la commande Impression Ecran de son clavier.
- C- L'analyste colle l'image copiée dans une diapositive de son fichier PPT de l'analyse en cours

Il faut répéter les opérations **a, b, c** pour chacun des **poi** de la séquence et enregistrer régulièrement le fichier ppt. C'est ce qu'illustre la figure 4.



Figure 4. Situation des prises de vue pour une séquence

Pour le cas d'une séquence frontale pour laquelle il est difficile d'identifier directement des points d'intérêt sans ambiguïté. Par exemple des pieds de poteaux portant ou non un luminaire. Il est alors nécessaire d'utiliser conjointement à la séquence latérale, la séquence frontale réalisée en même temps. Sur le poste principal de l'analyse, on affiche la vidéo latérale, alors que sur un poste de contrôle est affichée la vidéo frontale simultanée. On utilise alors les prises de vue de la vidéo frontale, pour contrôler le type de poi de la séquence latérale.

– Valorisation du temps

Lorsque la séquence a été analysée, il faut valoriser le temps dans le fichier Excel de la séquence. Pour cela, il faut afficher à l'écran en même temps le fichier PPT et le fichier Excel, en s'arrangeant pour faire apparaître le temps pour chaque diapositive du fichier PPT et les colonnes correspondant au temps du fichier Excel.

Il faut alors valoriser les champs avec les valeurs de temps affichées dans les diapositives de chaque fichier Excel. La formule utilisée pour calculer le temps est la suivante :

$$[(C2*60+D2)*1000+E2]*90]$$

C2= la partie minute du temps mesuré

D2= la partie secondes du temps mesuré

60= nombre de seconde/minute

1000= facteur multiplicatif pour passer de seconde en milliseconde

E2= la partie millisecondes du temps mesuré

90= facteur de calibrage pour être cohérent avec le temps gps (variable qtmediat_1)

La figure 5 présente clairement les éléments décrits précédemment dans leur contexte opérationnel.



Figure 5. Exemple d'un tableau de valorisation du temps d'une prise de vue

2.3.3. Localisation des POI du levé

La localisation des points d'intérêt passe d'abord par l'extraction des points gps à partir du **contour storyteller**, puis par leur transformation et leur localisation à l'aide du module de référencement linéaire dans la boîte à outil d'**arcgis**. Trois actions sont nécessaires avec ce module pour avoir les points géoréférencés dans un logiciel de cartographie : (i) créer des itinéraires, (ii) localiser les entités le long des itinéraires (iii) générer une couche d'événements des itinéraires. L'**annexe 4** présente les trois étapes en image dans **arcgis**.

A la suite de multiples tests de localisation, il ressort que l'ensemble des **poi** ne se positionnent que par rapport au tracé gps et non par rapport au référentiel cartographique de Libreville. Ce constat rejoint ceux montrés par d'autres auteurs [SOU09], [SOU07] et [WEBTOPCON].

Une phase de correction des données est donc envisagée de manière automatique sur la base des tronçons vectoriels du référentiel du Gabon. En fonction des cas, certains se retrouvent sur le terre-plein central, d'autres à gauche ou à droite. La figure 6 montre les différents cas de corrections proposées.

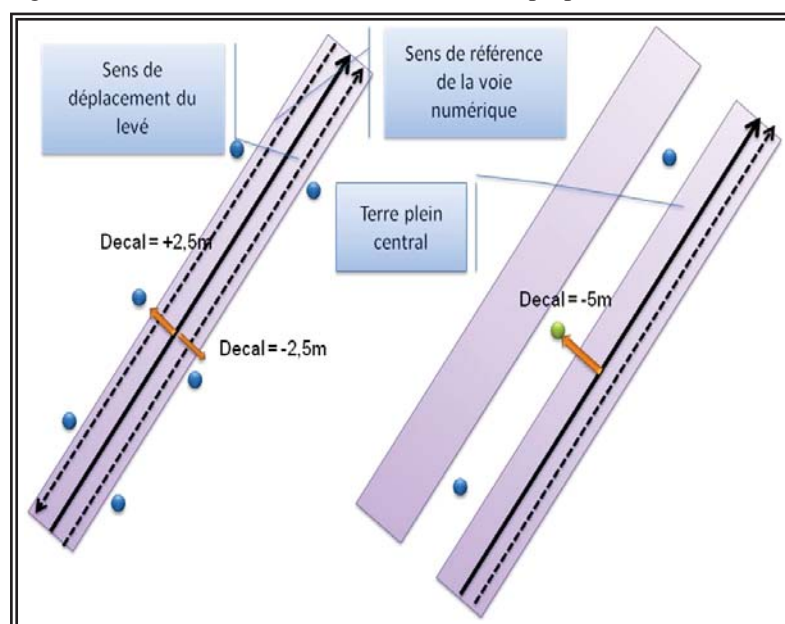


Figure 6. Différents cas d'analyse du positionnement des poi

2.3.4. Construction d'une base de données cartographique pilote

A ce stade du projet, il s'agit de la structuration des données produites et de leur archivage dans des dossiers. La figure 7 présente l'organisation de la base de données.

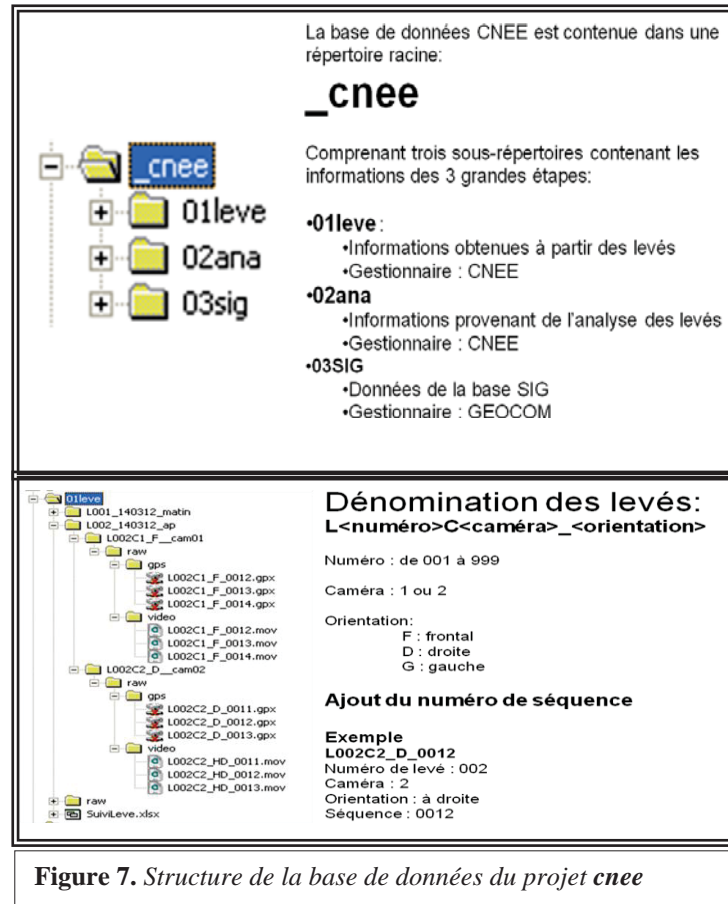


Figure 7. Structure de la base de données du projet *cnee*

3. Résultats préliminaires

Le **cnee** a conclu le 11 janvier 2012 avec **geocom**, un contrat d'assistance technique dans le cadre de la mise en place d'un laboratoire **sig** au sein du **cnee**. Ce contrat de collaboration a permis d'asseoir une méthodologie opérationnelle d'acquisition et d'analyse des données sur les réseaux d'éclairage public au Gabon et de construire une base de données cartographique comme outil pilote de production des informations d'aide à la décision au sein de l'institution. L'ensemble du processus s'est appuyé sur une équipe volontaire et dévouée du **cnee** avec des formations d'initiation en cartographie, conduite de projet, administration et utilisation du webmapping, et la prise en main du logiciel **arcgis**. Les acquis relevés permettent désormais de trouver au sein du **cnee** une cellule **sig** en gestation.

Dans cette section, nous allons présenter tour à tour la manière dont les tâches ont été réalisées, les résultats de l'exploitation thématique de la base de données cartographique et une petite discussion sur l'intérêt et la pertinence des outils utilisés.

3.1. Répartition des tâches entre Institutions

Dans un souci d'appropriation du système par le client, le projet prévoit au même moment une phase de formation sur certains aspects pratiques qui permettent déjà au client d'asseoir une stratégie opérationnelle de mise en œuvre du projet. Pour cela et grâce à ce projet, une cellule **sig** a été créée au sein du **cnee** et trois agents y ont été affectés avec une mission d'appropriation des résultats du projet. Le tableau 1 montre le premier niveau d'appropriation mis en place basé sur la répartition des tâches.

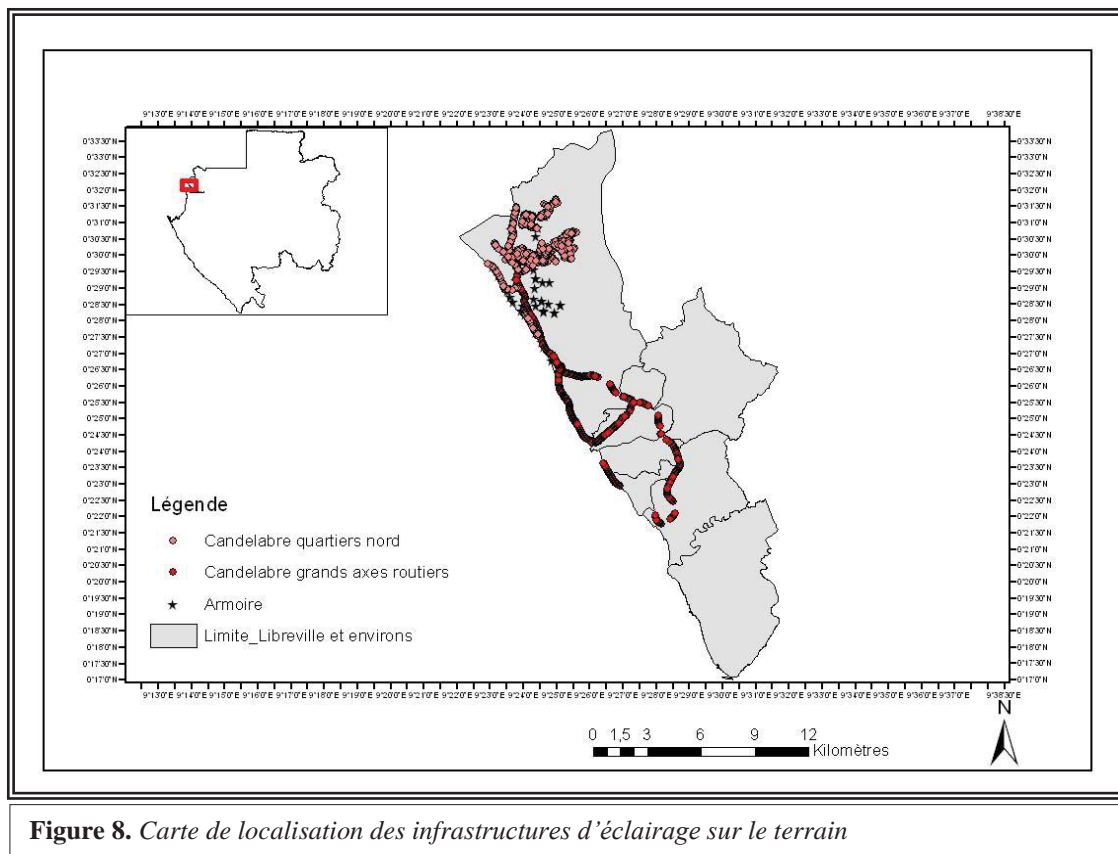
<i>Tâche</i>	<i>Compétences</i>	<i>Opérateur</i>	<i>Commentaire</i>
Préparation et gestion du référentiel géographique	Géomatique avancée	GEOCOM	Le référentiel géographique est fourni au CNEE
Réalisation du levé	Connaissance métier	CNEE	GEOCOM apporte une assistance technique
Analyse du levé: identifier et informer les POI	Connaissance métier	CNEE	GEOCOM apporte une assistance technique
Localisation des POI du levé	Géomatique avancée	GEOCOM	Les POI localisés sont fournis au CNEE
Exploitation des POI	Connaissance métier avec utilisation SIG	CNEE	Après formation à ArcGIS, accompagnement par GEOCOM sous forme d'assistance technique.

Tableau 1. Répartition des tâches entre institutions

3.2. Exploitation thématique de la base de données

Les données collectées par camera-gps et analysées nous ont permis d'élaborer une première analyse spatiale thématique, pouvant permettre au projet d'apporter un début de réponse aux défis des gestionnaires et des décideurs bénéficiaires du projet.

Le premier besoin était de pouvoir localiser les équipements sur le terrain et avoir à une meilleure estimation de leur nombre, afin de permettre une vision objective des budgets de maintenance et d'orienter clairement les sous-traitants de ce secteur. C'est ce qu'illustre la figure 8.



La figure 8 ci-dessus montre la quantité de données levées et traitées concernant les candélabres et les armoires/postes de livraison. A ce jour, nous avons **927 candélabres** et **75 armoires** localisés dans le groupement des quartiers nord de Libreville (**agondje et okala**), environ **1010 candélabres** le long des axes de circulation dans la ville. Ce qui donne un total de **2012 points de localisation** dont le traitement est déjà finalisé. Nous pouvons aussi évaluer le travail qui est déjà fait à Libreville en distinguant les zones couvertes de celles qui ne le sont pas.

Le deuxième point d'accord avec le **cnee** était d'avoir à une meilleure connaissance des zones qui doivent faire l'objet d'extension dans les quartiers. La figure 9 présente un cas d'école de distribution des équipements d'éclairage public dans le quartier **avorbam-agondje**.

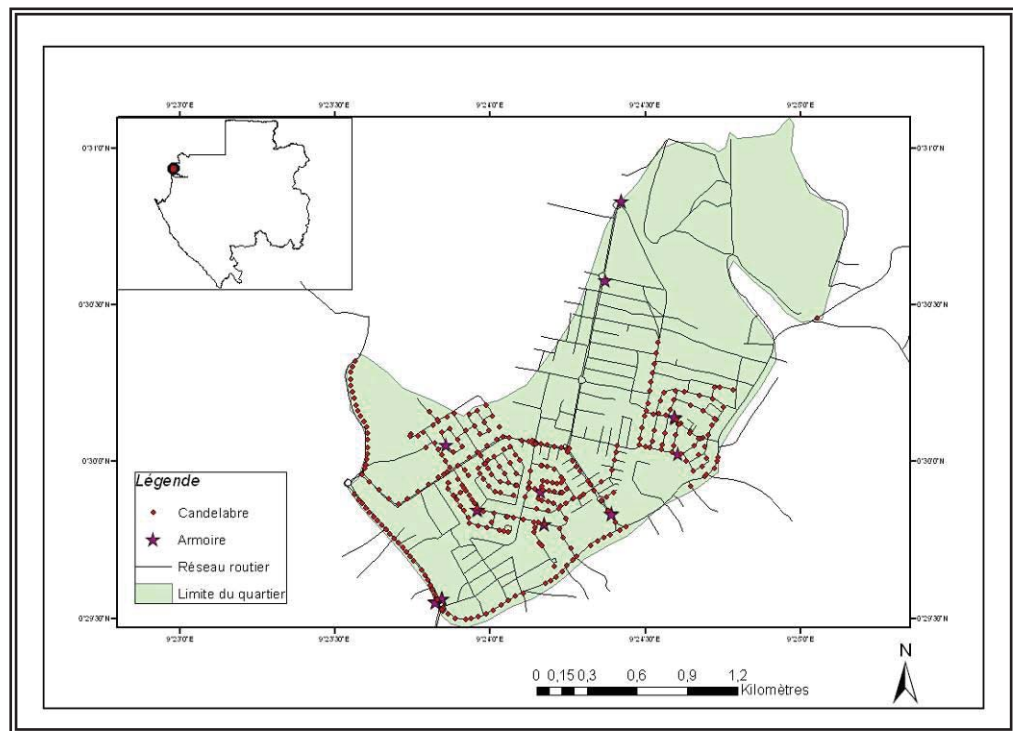
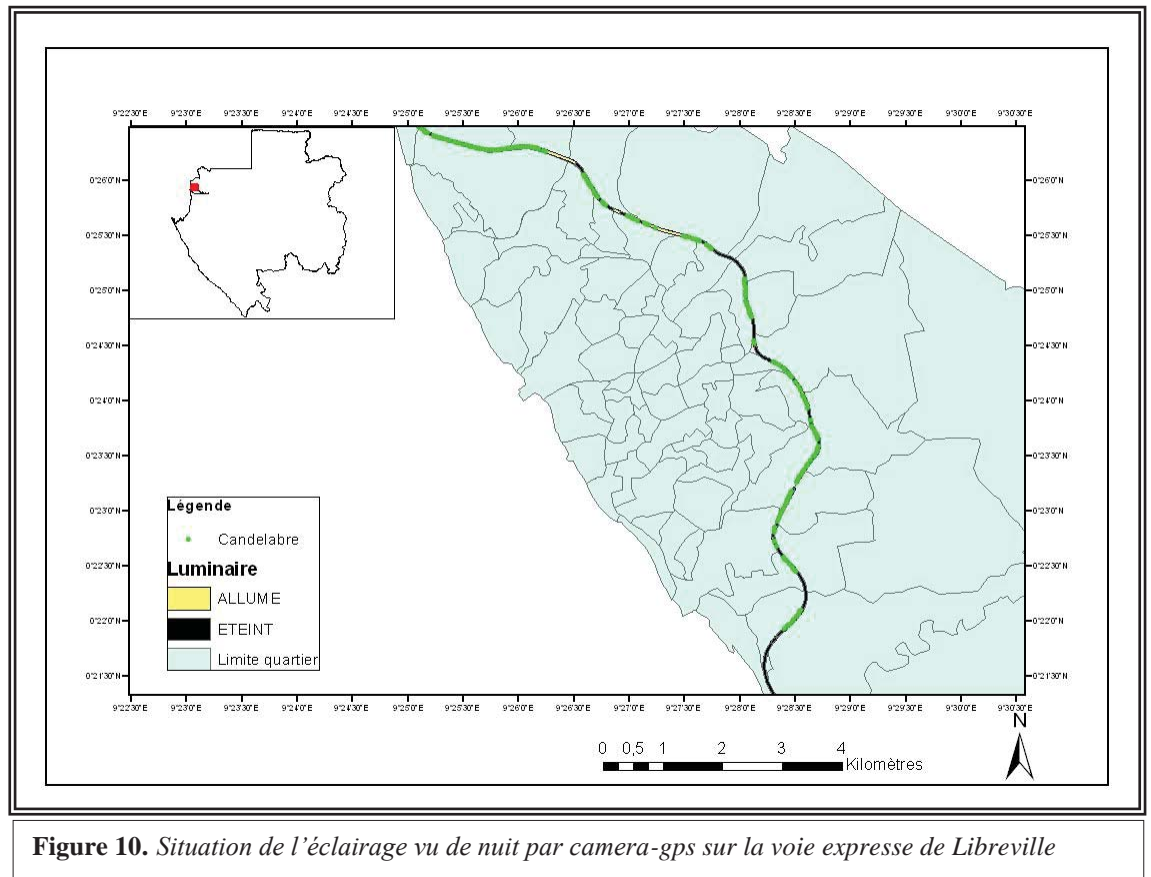


Figure 9. Situation de l'éclairage public dans le quartier Avorbam-Agondje à Libreville

Il ressort dans cette figure que l'éclairage public dans le quartier **avorbam-agondje** ne concerne qu'environ 50% de la zone habitée. Et pourtant, depuis le passage de la coupe d'Afrique des Nations (**can**) au Gabon au premier trimestre 2012, cette zone qui en faisait partie, est souvent considérée comme l'un des modèles réussis d'éclairage public par le **cnee**. On peut néanmoins observer en même temps la présence des armoires disponibles et évaluer le niveau d'investissements par quartier ou arrondissement en cas d'extension du réseau.

L'autre défi était également de pouvoir constituer un outil d'alerte pour orienter les équipes de maintenance sur les zones d'ombre ou d'éclairement, notamment sur les tronçons où les infrastructures existent déjà. À l'aide de la camera-gps nous avons pu effectuer un levé de nuit en 1 heure de temps pour répondre à ce besoin sur un tronçon test. La figure 10 présente les résultats de la situation réelle de l'éclairage vue de nuit par camera-gps.



Sur la figure 10, on voit clairement les segments de tronçon en noir où l'éclairage ne fonctionne plus, soit par manque de candélabres ou parce qu'ils sont en panne. Pour plus de compréhension, nous avons fait une superposition des points de candélabres en vert (levé pris un mois avant) sur le résultat du levé de nuit, ce confirme la fiabilité du travail effectué.

3.3. Discussion

Les caméras-gps utilisés suscitent actuellement un engouement important pour les levés effectués, en comparaison avec les systèmes fixes dont les gens ont l'habitude d'utiliser dans les pays africains. Pour autant, vont-ils remplacer les systèmes fixes ou simplement offrir une nouvelle modalité d'acquisition ? Pour éclaircir cette question, nous pouvons évoquer quelques critères de compréhension.

Le premier cas concerne la configuration des villes africaines. Les systèmes d'identification et de création des voies, constructions et parcelles n'ont tout simplement pas pu suivre le rythme de l'urbanisation, conduisant à une situation dans laquelle les villes de ces pays ont la moitié ou plus de quartiers d'accès difficiles, en particulier dans les quartiers les plus pauvres, et pourtant ces zones possèdent des réseaux d'eau et d'éclairage à usage public. Libreville et les autres villes du Gabon n'échappent pas à cette réalité [BURG07]. Dans ce cas, nous envisageons d'utiliser les systèmes fixes pour compléter les équipements de ces quartiers.

Le deuxième cas concerne le gain réalisé en matière de collecte des données sur le terrain. En 5 jours, il a été possible de lever plus de 2000 équipements, ce qui est exceptionnel dans le quotidien du client. Mais, il faut aussi dire que ce gain est fortement réduit avec le processus d'analyse et de transformation des coordonnées qui reste encore fastidieux en termes de volume de travail.

Le troisième point concerne l'accessibilité à l'outil. Les caméras-gps utilisées dans le cadre de ce projet ont coûté environ **700 euros ttc** l'unité, alors qu'un gps fixe de bonne précision coûterait environ **2000 euros ttc**.

Un dernier critère est la qualité du relevé laser. Plusieurs éléments sont à prendre en considération que l'on peut catégoriser ainsi : précision, résolution et homogénéité, complétude. La précision des points est une question qui se pose de façon spécifique dans les systèmes mobiles par rapport aux systèmes fixes car elle dépend du système de localisation (**pos**). Actuellement les **pos** ne garantissent pas la qualité de service en termes de précision. Il faut nécessairement une étape de correction des données par rapport au référentiel géographique connu. Il faut dire que ce problème n'est pas spécifique aux systèmes mobiles mais se pose différemment des systèmes fixes. La complétude concerne le problème des zones occultées, produites par les "ombres" d'objets indésirables au moment du relevé, comme les véhicules, piétons, arbres, etc. Ce problème n'est pas non plus spécifique aux relevés mobiles mais ils offrent probablement moins de liberté qu'avec les stations fixes pour choisir un point de vue adapté.

En conclusion, les systèmes mobiles ne répondent actuellement que partiellement aux besoins sur la qualité des relevés, en particulier sur la précision des données. L'analyse des critères de qualité permet toutefois de travailler sur la conception de ces systèmes pour mieux répondre aux besoins [GOU09].

4. Conclusion et perspectives

Les systèmes mobiles de cartographie ouvrent de nouvelles possibilités de relevés en environnement urbain. Cet article a présenté l'utilisation des cameras-gps dans le cadre d'un projet de géospatialisation des réseaux d'éclairage public à Libreville au Gabon.

En offrant une plus grande vitesse d'acquisition, les relevés mobiles rendent techniquement et économiquement opérationnels des chantiers jadis inaccessibles, notamment le secteur de l'éclairage public où le volume d'équipements à manipuler est complexe.

L'objet de cette initiative pilote était d'expérimenter une méthode géomatique à plusieurs articulations, et de faire une démonstration des résultats liés aux besoins opérationnels du centre national de l'eau et d'électricité (**cnee**). Il s'agit d'une expérience novatrice dans la sous-région qui pourra inspirer les pays d'Afrique centrale où les enjeux en termes d'équipements énergétiques et de gestion des infrastructures sont énormes.

Toutefois, des améliorations sont attendues sur la précision des points mesurés. De plus, les spécificités des systèmes fixes maintiendront un intérêt qui seront complétés avec ces nouvelles technologies.

A moyen terme, la collecte des données sera élargie à l'ensemble de la ville de Libreville, puis dans celles de Port-Gentil et de Franceville. La finalité étant d'élaborer un système de gestion de l'information et des opérations de la direction technique du **cnee**.

5. Remerciements

*Notre profonde gratitude va à l'endroit du Directeur général du **cnee** monsieur **Joël Hubert James** qui a bien voulu accorder ce projet à notre jeune compagnie. Nous tenons aussi à adresser nos sincères remerciements à toute l'équipe technique du **cnee**, en particulier ceux du service de la maintenance et de la cellule sig pour leurs appuis multiformes pendant les opérations de levés et d'analyses des séquences vidéo. Enfin, nous ne saurons oublier nos partenaires **d'esri france** et de **sambus company du ghana**, qui ont bien voulu nous mettre en contact avec le client via son directeur technique.*

6. Bibliographie

• Ouvrages, articles et rapports internes

- [ASS07] Assako Assako R. J., Ayissi Etame A., Courel M.F. et al, « *Observatoires urbains et environnementaux en Afrique* », ouvrage, Dianôia, 1ere édition, 2007, 223 p., Merignac, Chennevières-sur-Marne.
- [BAR08] Barber D., Mills J. and Smith-Voisey S., « *Geometric validation of a ground-based mobile laser scanning system* », ISPRS, Journal of Photogrammetry and Remote Sensing, n. 63, 2008, p.128-141.
- [BUR07] Burgel G., Ndong Mba J. C., « *Villes en parallèle* », ouvrage, Université PARIS X – NANTERRE, Laboratoire de Géographie urbaine, N°40-41, 2007, 241 p., Paris.
- [ELL01] Ellum C. M., El-Sheimy N., « *A mobile Mapping Systems for the Survey Community* », Proceeding of the 3rd International Symposium on Mobile Mapping Technology (MMS 2001), article, January 3-5, 2001, 12 p., Cairo, Egypt.
- [GOU09] Goulette F., « *Relevés laser urbains par Systèmes Mobiles de Cartographie* », article, n°119–2^e trimestre, 2009, 5 p., Revue XYZ, Paris.
- [MAD10] Madebe D., « *Hétérogénéités spatiales, réorganisation du territoire et développement du Gabon* », ouvrage, 2010, 141 p., CENAREST, Libreville.
- [MAK12] Makak J. S., Thomas D., Ogowet H. et al « *Système d'information géographique des réseaux d'eau et d'éclairage public : activités en cours à Libreville au Gabon* », rapport préliminaire du projet, 2012, 17 p., Libreville.
- [MAS08] Madeira S., Gonçalves J., Bastos L., « *Low cost mobile mapping system for urban survey* », LNEC, May 12-15, 2008, 12 p., Lisbon.
- [SOU09] Sourimant G. Colleu T., Jantet V., Morin L., « *Recalage GPS / SIG / Vidéo, et synthèse de textures de bâtiments* », article, 2009, 6 p., site www.irisa.fr.
- [SOU07] Sourimant G., Morin L., Bouatouch K., « *Vers une reconstruction 3D des zones urbaines : mise en correspondance de données Gps, Sig et Vidéo* », article, 2007, 4 p., site www.irisa.fr.
- [THO12] Thomas D., Makak J. S., « *Guide pour effectuer un levé URBASIG et l'analyser dans le cadre du SIG du CNEE* », guide d'assistance technique, 2012, 15 p., Libreville.
- [VAL05] Valkova S., « *Accompagnement du projet de renouvellement et de déploiement du SIG de la ville de Mâcon et étude comparative des solutions de publication cartographique sur le WEB* », rapport de stage, 2005, 69 p., Saint-Etienne.

• Sites Web

- [WEBNOM10]<http://www.nomadicsolutions.biz/upload/file/produits/WedJat%20100/manuel%20utilisateur%20logiciel%20Wedjat%2010.pdf>.
- [WEBTOPCON][http://www.topcon.fr/Documents/geo80_pp24-28_topo%20\(2\).pdf](http://www.topcon.fr/Documents/geo80_pp24-28_topo%20(2).pdf).
- [WEBLARA]http://www.lara.prd.fr/imara/positions/2012conduite_urbaine_autonome.
- [WEBWIKIPEDIA]http://fr.wikipedia.org/wiki/%C3%89clairage_public.

7. Biographie

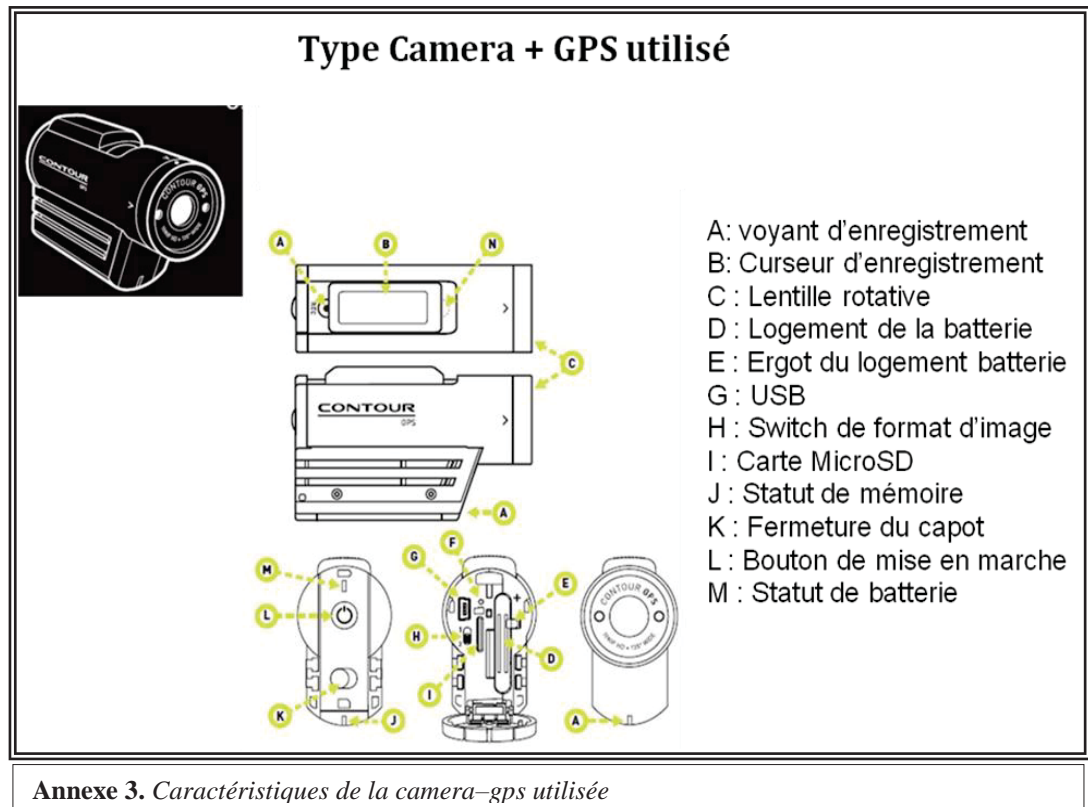
*Créé en 2008 au Gabon comme geospatial company, **geocom** est un cabinet conseil et d'assistance technique **privé**, spécialisé dans le développement des outils d'aide à la décision, sur la base des technologies géomatiques. La compagnie est également l'un des distributeurs agréés des produits **Esri** (arcgis), **Geoeye** (images satellitaires haute résolution) et Orbitica (GPS-GNSS).*

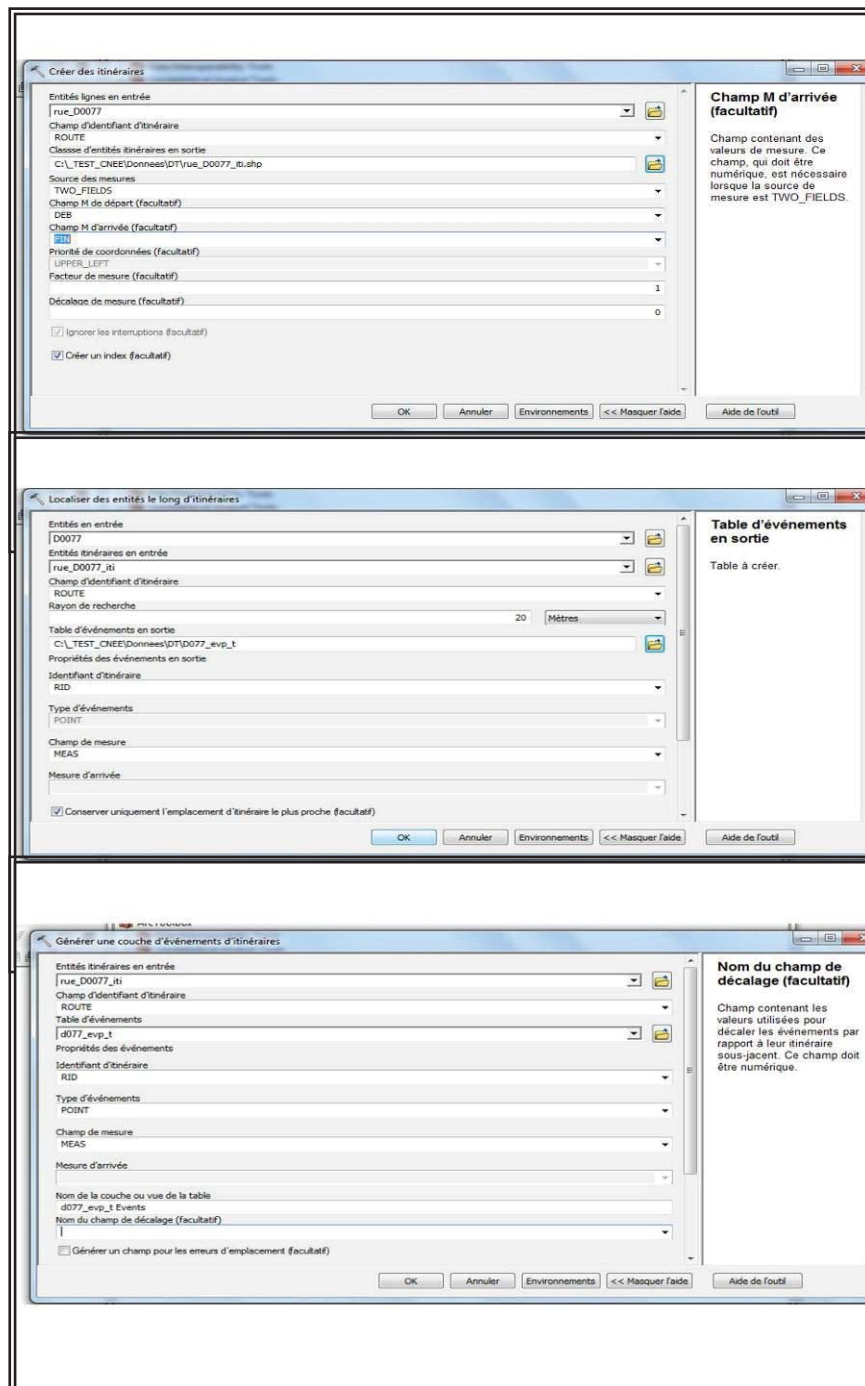
*Le conseil national de l'eau et d'électricité (**cnee**) par contre un établissement **public** créé en 1995, chargé de la mise œuvre de la politique de l'Etat en matière de création, gestion des réseaux d'éclairage public et d'installations des collectivités locales en eau potable et en électricité au Gabon.*

8. Annexes

Annexe 1: Toward a geospatialisation of public lighting in Gabon

Loi n°8/93 du 7 avril 1993 fixant le régime juridique de la production, du transport et de la distribution de l'eau potable et de l'énergie électrique
Loi n°9/93 du 7 avril 1993 portant création du fonds spécial de l'eau
Loi n°10/93 du 7 avril 1993 portant création du fonds spécial de l'électricité
Loi n°14/95 du 23 novembre 1995 portant création, composition et fixant les attributions et le fonctionnement du conseil national de l'eau et de l'électricité
Décret n°1142/PR/MERH du 10 août 1993 fixant les pouvoirs de contrôle de la concession du service public de l'eau potable et de l'énergie électrique délégués aux communes de Libreville et de Port-Gentil
Décret n°11143/PR/MERH du 10 août 1993 portant désignation du concessionnaire pour la production, le transport et la distribution de l'eau potable et de l'énergie électrique
Décret n°586/PR du 9 juin 1997 autorisant la cession des actions détenues par l'Etat au capital de la SEEG.
Décret 628/PR/MMEP du 18 juin 1997 portant désignation du concessionnaire du service public de la production, du transport et de la distribution de l'eau potable et de l'énergie électrique
Décret 629/PR/MMEP du 18 juin 1997 fixant les modalités d'application de certaines dispositions de la loi n°8/93 du 7 avril 1993
Décret n°769/PR/MMEPRH du 26 juin 2003 portant réglementation de la production indépendante d'énergie électrique en République gabonaise
Décret n°770/PR/MMEPRH du 26 juin 2003 fixant les modalités d'octroi de concession du service public de l'eau potable et de l'énergie électrique à un opérateur privé en République gabonaise
Décret 771/PR/MMEPRH du 26 juin 2003 portant modification de certaines dispositions du décret n°629/PR/MMEP du 18 juin 1997
Décret 772/PR/MMEPRH du 26 juin 2003 portant réglementation de l'utilisation de l'énergie électrique et des appareillages fonctionnant à l'énergie électrique en République gabonaise
Décret n°744/PR/MMEPRH du 22 septembre 2005 portant projet d'intérêt général et création de zone de préservation de la ressource en eau
Arrêté n°345/MMEPRH du 13 avril 2004 portant autorisation d'auto-production d'énergie électrique à la SOGARA
Arrêté n°349/MMEPRH du 30 avril 2004 portant agrément technique des organismes ou des cabinets de contrôle et d'expertise électrique
Arrêté n°354/MMEPRH du 30 avril 2004 fixant la périodicité, l'objet et l'étendue des vérifications des installations électriques, industrielles et des réseaux de distribution, de transport et de production d'énergie.
Arrêté 355/MMEPRH du 30 avril 2004 portant contrôle des installations intérieures, des réseaux de distribution et de transport de l'énergie électrique.
Annexe 2. Textes juridiques consultés





Annexe 4. Les trois étapes principales de transformation des points GPS sur ArcGIS

Cours d'eau non navigables par intégration cartographique

Catherine Marlier* — **Vincent Bombaerts*** — **Abdel Mokadem ****
— **Pierre Defourny***

** Earth and Life Institute – Environmental sciences*

UCL-Géomatique Université Catholique de Louvain

Place Croix du Sud, 2 bte L7.05.16 B- 1348 Louvain-la-Neuve (Belgique)

***Service Public de Wallonie - Direction générale opérationnelle de l'Agriculture,
des Ressources Naturelles et de l'Environnement - Département de la Ruralité et
des Cours d'Eau - Direction du Développement Rural*

Avenue Prince de Liège, 15, B-5100 Namur (Belgique)

RÉSUMÉ. Cette recherche a pour objectif le développement d'une approche méthodologique en vue de constituer une base de données géographiques des cours d'eau en Wallonie. La méthode proposée permet, de constituer une base de données unique en intégrant des données provenant de trois sources vectorielles différentes. La méthode proposée comprend deux phases : l'intégration sémantique, qui permet d'unifier les descriptions des objets ; l'intégration des données, qui permet d'unifier des objets identifiés comme sémantiquement équivalents.

ABSTRACT. This research has as objective the elaboration of a methodological approach, in order to build a geographic database of the watercourses in Wallonia. The proposed method enables the creation of a single database integrating data from three different vectorial sources. The proposed method proceeds in two phases: the semantic integration which consists in merging object's descriptions ; the data integration which merge objects identified as semantically identical.

MOTS-CLÉS: Intégration données cartographiques, Cours d'eau, Appariement, Ontologie, SIG, Wallonie

KEYWORDS: Cartographic data integration, Watercourses, Matching, Ontology, Geographic Information System.

1. Introduction

De nombreuses bases de données géographiques sont aujourd'hui disponibles pour représenter les phénomènes du monde réel. Se pose désormais la question de la multiplicité et de la cohérence entre ces différentes sources de données.

Les techniques d'appariement et d'intégration des données géographiques offrent une réponse aux problèmes de cohérence et de redondance engendrés par la multiplicité des sources d'information. Depuis quelques années de nombreuses recherches sont menées sur ce sujet. L'appariement permet d'identifier et relier les différentes représentations des mêmes objets réels contenues dans les diverses sources de données. Il est utilisé dans divers domaines d'application, notamment dans la constitution de bases de données multi échelles [MUS 07], la mise à jour des réseaux routiers [DEV 97, XIO 04, STI 07] ou la détection du changement [BAD 00]. Dans cette recherche, nous utiliserons les techniques d'appariement à des fins d'intégration. L'intégration est un processus qui permet de fusionner l'information provenant de différentes sources [DEV 97].

Ce document présente une méthode semi-automatique d'intégration de trois sources de données vectorielles. L'originalité de cette recherche réside dans le fait que la méthode a été développée dans un contexte opérationnel, impliquant l'application de la méthode et la production de données à l'ensemble d'une région représentant plus de la moitié du territoire de la Belgique.

2. Cadre de l'étude

Le Département de la Ruralité et des Cours d'eau (DRCE) est en charge de la gestion du Réseau Hydrographique Wallon (RHW), document cartographique numérique vectoriel reprenant les caractéristiques attributaires des 15000 km de cours d'eau non navigables classés en Wallonie. Cette donnée cartographique, basée sur les cartes topographiques de l'IGN datant de 1972 et l'Atlas des cours d'eau non navigable, réalisé dans les années 50 et actuellement le document de référence officiel, a besoin d'être actualisée.

Des sources de données géographiques planimétriquement plus précises et/ou plus récentes sont disponibles actuellement en Wallonie ou à l'Institut Géographique National (IGN). Le DRCE souhaite intégrer ces nouvelles données au RHW pour se doter d'une base de données géographiques complète reflétant précisément la situation actuelle du réseau hydrographique wallon.

3. Objectif de l'étude

L'objectif de l'étude est de développer et mettre en œuvre une méthode semi-automatique d'intégration de données permettant de produire une base de données des cours d'eau unique pour la Wallonie.

4. Description des données

La première phase de cette recherche a consisté en un recensement et une caractérisation détaillée des sources de données. La base de données intégrée « cours d'eau » utilise trois sources de données pour sa constitution (Figure 1) : le PICC, le Top10v-GIS et le Réseau Hydrographique Wallon (RHW). Les orthophotographies sont utilisées comme support visuel dans le processus d'intégration.

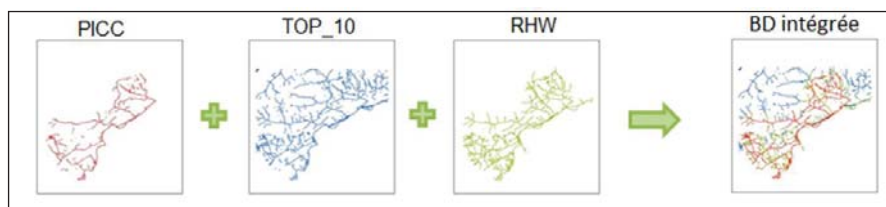


Figure 1. Représentation schématique des données sources et du réseau intégré

4.1. Le Projet Informatique de Cartographie Continue (PICC)

Le Projet Informatique de Cartographie Continue (PICC), initié en 1991 par le Ministère de l'Équipement et des Transports (MET) est une cartographie tridimensionnelle au 1:1000 de la Wallonie réalisée par des techniques de photogrammétrie.

Le PICC est une donnée topographique dont la précision planimétrique est très élevée (entre 15 et 25 cm), mais pour laquelle l'information attributaire est limitée aux routes et aux bâtiments. Le PICC n'est pas structuré en objets pour la thématique des cours d'eau. On remarque une très grande discontinuité des axes dans les zones de couvert dense, un morcellement élevé des axes et l'absence d'attributs autres que le type d'objet restitué.

4.2. Le TOP10v-GIS

Le Top10v-GIS est une base de données géographiques vectorielles au 1:10.000 issue de la carte topographique de base de l'Institut Géographique National belge (IGN). Sa production a débuté en 1989, elle a été élaborée avec des techniques de

photogrammétrie sur base de photographies aériennes au 1:21.000 complétées par des relevés sur le terrain.

Le réseau hydrographique du Top10v-GIS présente une plus grande continuité que celui du PICC, mais dont la précision planimétrique est moins élevée. Le Top10v-GIS couvre l'entière du territoire et possède des attributs intéressants pour la constitution de la couche de référence des cours d'eau. Il sera utilisé pour compléter la composante topographique dans les zones où l'information du PICC n'est pas disponible ou incomplète.

4.3. *Le réseau hydrographique wallon (RHW)*

Le Réseau hydrographique wallon (RHW) est une cartographie numérique qui concerne l'ensemble du réseau hydrographique de la Wallonie. Elle a été établie par numérisation de diverses sources d'informations, dont la principale est le tracé complet des cours d'eau repris sur les anciennes cartes de l'IGN (1972) au 1:10.000.

Le RHW est une donnée dont la précision planimétrique est assez faible comparativement aux autres sources de données (> 1m). Sa qualité est assez variable car il a été constitué à partir de sources différentes. Par contre, il possède une grande richesse attributaire et une grande connectivité car il reprend les axes souterrains. Des informations administratives très complètes relatives à chaque segment de cours d'eau sont associées à cette information graphique. Ces caractéristiques proviennent de l'Atlas des Cours d'Eau Non Navigables (ACENN), document au format papier, réalisé dans les années cinquante, contenant toutes les informations relatives aux cours d'eau, aux ouvrages associés et aux travaux éventuellement réalisés.

4.4. *Les photographies orthorectifiées de la Région Wallonne*

Les orthophotographies couleur et infrarouge sont réalisées sur base de vols aériens réalisés sur deux années et renouvelées tous les trois ans. La résolution des orthophotographies 2006-2007 est de 50 cm pour une précision planimétrique moyenne de 1 m. Pour les vols 2009-2010, ces valeurs sont respectivement de 25 cm et 50 cm. Cette quatrième source de données ne fait pas partie du processus d'intégration. Elle sert de support visuel et permet de vérifier si la situation cartographiée correspond à la situation réellement observée sur le terrain.

5. Méthodologie

Une méthodologie générale d'intégration a été mise au point pour automatiser l'intégration des données. La méthode d'intégration comprend deux phases :

1. L'intégration sémantique, qui permet d'unifier les descriptions des objets ;
2. L'intégration des données, qui permet d'unifier des objets identifiés comme sémantiquement équivalents.

5.1. *L'intégration sémantique*

Un même objet du monde réel, par exemple un « cours d'eau », peut être défini de manière différente dans chaque source de données. Tout schéma conceptuel de base de données s'appuie sur une conceptualisation particulière du monde réel qui induit des choix de modélisation au niveau des objets utilisés pour décrire la réalité de terrain. Ces choix de modélisation étant différents d'une base de données à l'autre, il n'existe généralement pas de correspondance stricte entre les objets des différentes bases de données. L'intégration sémantique a pour objectif d'établir les correspondances entre les descriptions des objets des différentes bases de données.

Le terme « ontologie » est utilisé pour désigner « la spécification explicite d'une conceptualisation » [GRUB 93] ou encore « l'ensemble des choses dont l'existence est admise par une théorie ou un système de pensée donné » [PAR 02]. Les relations existant entre l'ontologie et le schéma conceptuel sont détaillées dans les spécifications de la base de données.

L'intégration sémantique des schémas est une approche tenant compte de la signification des classes (leur sémantique), et de l'information contenue dans les classes (leur population) [SPA 92]. Elle consiste à définir une description unifiée (le schéma intégré), qui regroupe toute la sémantique des schémas initiaux, et des règles de traduction qui vont permettre la transformation des données [DEV 97].

Dans cette méthode, l'intégration sémantique se compose de quatre opérations :

5.1.1. *La Pré-intégration*

La pré-intégration se définit comme l'ensemble des opérations permettant de décrire les schémas conceptuels et les spécifications des bases de données sources.

Les informations sur la signification précise des données sont contenues pour une petite part dans le schéma structurel de la base de données, mais l'interprétation correcte de ce schéma repose largement sur la documentation rédigée par les experts et sur l'expertise des administrateurs et des concepteurs de la base [PAR 00]. Les spécifications rassemblent l'ensemble des règles permettant le passage du terrain conceptualisé à sa représentation dans une base de données.

5.1.2. *La construction de l'ontologie globale*

Pour intégrer de façon efficiente les schémas issus de différentes modélisations du monde réel, il est possible de faire appel à une ontologie de domaine. L'ontologie de domaine peut se définir comme l'expression formalisée d'un ensemble de concepts partagés au sein d'un domaine [GUA 98]. L'intérêt de l'utilisation d'une

ontologie de domaine dans un processus d'appariement réside dans le fait que les concepts qu'elle définit sont partagés par l'ensemble des données sources.

La construction de l'ontologie de domaine a pour objectif de décrire formellement les concepts partagés par les schémas et les ontologies des différentes bases de données à intégrer.

5.1.3. L'appariement des schémas

L'appariement des schémas permet d'établir et de décrire les correspondances interschémas. L'appariement a pour objectif de lier les différentes descriptions des mêmes objets réels contenues dans les diverses sources à intégrer (Figure 2).

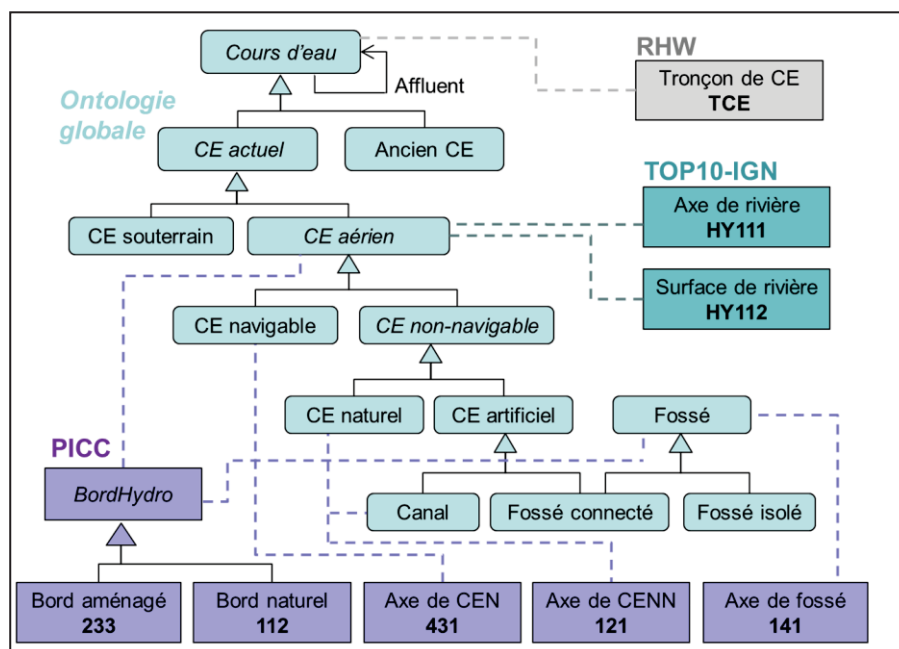


Figure 2. Liens entre les schémas des bases de données sources et l'ontologie globale

5.1.3. L'intégration des schémas

L'intégration des schémas est la description du schéma intégré et des règles de transformation des schémas sources vers le schéma intégré.

En résumé, dans cette recherche, l'intégration sémantique consiste à fusionner les dictionnaires de données des sources dans le dictionnaire intégré. Les

informations intéressantes et intégrables sont sélectionnées et des règles de traduction des modèles sources vers le modèle intégré sont instaurées.

L'intégration de la sémantique permet d'unifier la description, mais ne permet pas de relier les objets représentant les mêmes phénomènes du monde réel (objets homologues). L'intégration sémantique est suivie, dans une deuxième phase, de l'intégration géométrique des objets identifiés comme sémantiquement équivalents. Cette seconde phase est appelée « l'intégration des données ».

5.2. *L'intégration des données*

L'intégration des données, qui consiste à fusionner les données provenant de différentes sources [DEV 97], se déroule ensuite en cinq étapes :

5.2.1. *La Pré-intégration*

La pré-intégration regroupe toutes les étapes de transformation des données, d'ajout des métadonnées et de contrôles de qualité. Les règles de traduction des modèles sources vers le modèle intégré, définies dans la première phase, sont appliquées à ce niveau.

5.2.2. *L'appariement*

Le but de l'appariement est d'identifier et lier les différentes représentations des mêmes objets réels contenues dans les diverses sources à intégrer. Pour appairer une représentation d'un cours d'eau à une autre, nous avons adopté une approche globale qui utilise le PlugIn libre RoadMatcher [RAM 04] (Figure 3). La réussite de l'appariement automatique (85 à 98%) dépend du paramétrage de l'outil et de la préparation des données lors de la pré-intégration. Les appariements restant sont à réaliser manuellement.

5.2.3. *La définition de la précedence*

La définition de la précedence détermine, entre toutes les géométries appariées, la géométrie à conserver en fonction des actualités et des précisions des représentations appariées.

5.2.4. *L'intégration*

L'intégration proprement dite consiste à fusionner et ajuster les géométries précédentes, et à intégrer les attributs des différentes représentations. L'intégration de données pour la création d'un réseau unique implique la fusion des données des différentes représentations (Figure 3).

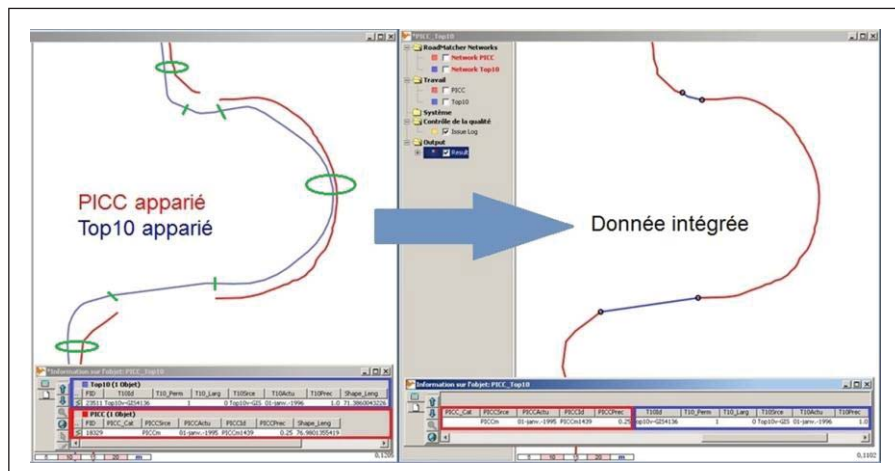


Figure 3. Exemple d'appariement et d'intégration des géométries et des attributs

5.2.5. La post-intégration

La post-intégration comporte toutes les étapes de mise en conformité du résultat de l'intégration au modèle intégré, de recalcul de certains attributs, et de contrôles de qualité.

6. Résultats

Dans un premier temps, la méthode décrite ci-dessus est appliquée pour intégrer les réseaux du PICC et du Top10v-GIS et produire le réseau intégré intermédiaire PICC_Top10v-GIS (Figure 5). Dans un deuxième temps, la mise en œuvre des protocoles développés a pour objectif d'intégrer le RHW au réseau intégré intermédiaire PICC_Top10v-GIS, afin de produire la BD géographique intégrée des cours d'eau. En suivant cette méthode, des protocoles d'intégration ont été développés sur des zones restreintes correspondant à des bassins versants, et ensuite appliqués aux données de l'entière du territoire wallon (Figure 4), afin de constituer la base de données des cours d'eau non navigables de la Wallonie.

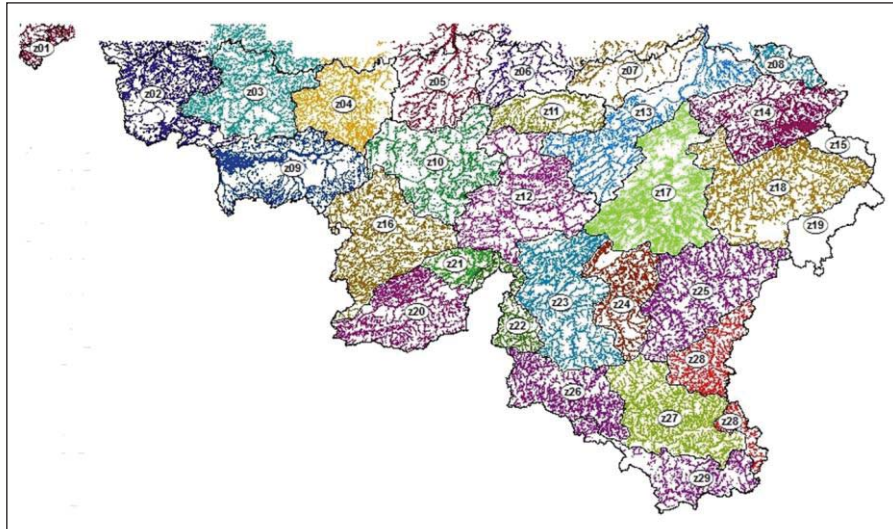


Figure 4. *Le Réseau hydrographique intégré – vue d’ensemble par bassins versants*

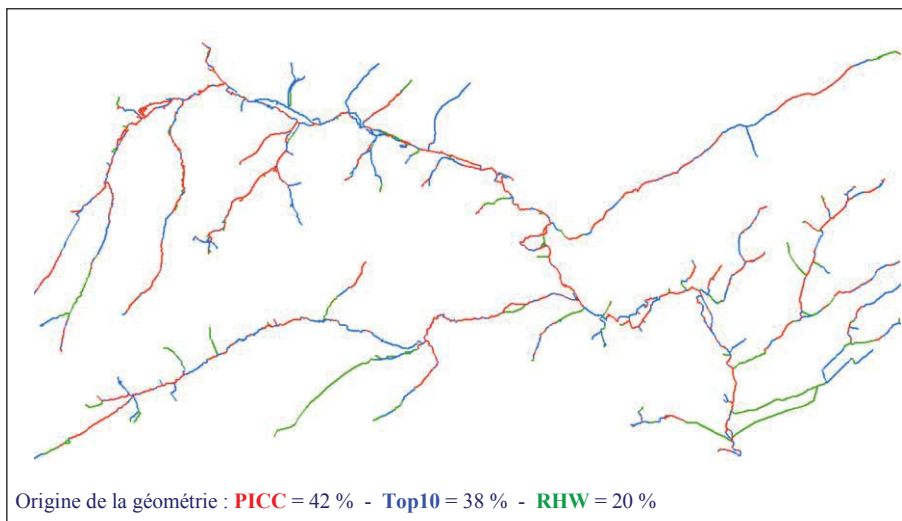


Figure 5. *Extrait du réseau hydrographique intégré – bassin aval de la Meuse*

Le résultat de l’intégration est une donnée plus complète, plus précise planimétriquement, plus riche attributairement et topologiquement et plus actuelle que chacune des données sources prises séparément.

7. Discussion

La mise en œuvre de la méthode et les données produites met en évidence l'intérêt de ce type de méthode pour générer des bases de données. Le travail de production semi-automatisé est en cours de réalisation dans le cadre du projet CENNIC (Cours d'eau non navigables par intégration cartographique). La phase manuelle de production de données reste laborieuse, néanmoins elle apporte l'intérêt de la mise à jour des données. En effet, la confrontation entre les différentes sources de données et des images plus récentes, permet, via la méthodologie appliquée, de créer des produits « dérivés » qui peuvent être utilisés dans d'autres contextes. Les producteurs des données sources sont notamment intéressés par des produits de mise à jour de référentiel ou de détection de zones de changement. La production des données sera suivie d'une phase de validation qui permettra d'intégrer des informations complémentaires provenant du terrain pour résoudre les discordances répertoriées lors de la production.

8. Bibliographie

- [BAD 00] Badard T., Propagation des mises à jour dans les bases de données géographiques multi-représentations par analyse des changements géographiques, Thèse de Doctorat, Université de Marne la Vallée, 2000
- [DEV 97] Devogele T., Processus d'intégration et d'appariement de bases de données Géographiques, Applications à une base de données routières multi-échelles, Thèse de Doctorat, Université de Versailles, 1997
- [GRU 93] Gruber T., Toward principles for the design of ontologies used for know-ledge sharing. In : Guarino, Nicola et Poli, Roberto (dir.) Formal ontology inconceptual analysis and knowledge representation. Dordrecht : Kluwer academic, 1993.
- [GUA 98] Guarino N., Formal ontology and information systems. Proceedings of FOIS'98, Trento, Italy, 6-8 June 1998. Amsterdam : IOS Press, 1998.
- [MUS 07] Mustière S., Devogele T., « Matching networks with different levels of detail », *GeoInformatica*, 2008, p. 765-795.
- [PAR 00] Parent C., Spaccapietra S., Database integration : The key to data interoperability. In : Papazoglou, M. P., Spaccapietra, S. et Tari, Z. (dir.) Advances in object-oriented data modeling. Cambridge (USA) : MIT Press, 2000.
- [PAR 02] Partridge, C., The role of ontology in integrating semantically heterogeneous databases. Rapport technique 05/02, LADSEB-CNR, Padoue, 2002.
- [RAM 04] Ramsey P., The jump project and direction, 2004 [référence du 15 avril 2008]. http://www.jump-project.org/assets/JUMP_Project_and_Direction.pdf
- [SPA 92] Spaccapietra S., Parent C., Dupont Y., Model Independent Assertions for Integration of Heterogeneous Schemas, *Very Large DataBases Journal*, 1992, Vol. 1, Num. 1, pages81-126.

- [STI 07] Stigmar H., Matching Route Data and Topographic Data in a Real-Time Environment, Proceedings, ScanGis2005.

Les paysages numériques en 3D dans les démarches participatives.

Vers une approche par les usages.

Matthieu Noucher*, Sébastien Nageleisen*, Laurent Couderchet*, Julie Pierson*, Maxime Demade*, Irmi Obermaier, Nathalie Pinède***, Véronique Lespinet-Najib******

(*) CNRS / Université de Bordeaux III - UMR ADES
Maison des Suds – 12 esplanade des Antilles – 33607 Pessac – France
m.noucher@ades.cnrs.fr - snageleisen@univ-bordeaux3.fr - lcouderc@u-bordeaux3.fr - j.pierson@ades.cnrs.fr - maxime.demade@etu.u-bordeaux3.fr

(**) Université de Bonn
Department of Geography – Meckenheimer Allee 166 – D-53115 Bonn – Germany
irmi.obermaier@web.de

(***) Université de Bordeaux III – EA MICA – Maison des sciences de l’homme d’Aquitaine – Esplanade des Antilles – 33607 PESSAC cedex – nathalie.pinede@iut.u-bordeaux3.fr

(****) Equipe CIH – Laboratoire IMS (UMR 5218) 146 rue Léo Saignat 33076 Bordeaux Cedex – veronique.lespinet@ensc.fr

RÉSUMÉ. Les paysages numériques en 3D sont aujourd’hui pleinement intégrés dans les démarches participatives des collectivités territoriales. Au delà des discours incantatoires (ils généreraient plus de transparence) ou alarmistes (ils seraient source de confusion), il reste aujourd’hui des enjeux forts pour une recherche sur l’analyse de leurs usages par les citoyens. En mobilisant le cas de la Fabrique Métropolitaine de la Communauté Urbaine de Bordeaux, on se propose ici de présenter les enjeux d’une analyse par les usages, les premiers résultats d’une étude de comparaison (densité physique perçue versus densité virtuelle perçue) et le protocole de recherche inter-disciplinaire (géographie, géomatique, cognitive, sciences de l’information et de la communication) en cours d’expérimentation.

ABSTRACT. Digital 3D landscapes are nowadays completely integrated in participatory approaches of local authorities. Beyond incantatory or alarmist discourses, there are strong issues for research on their uses by citizens. With the case study of Fabrique Metropolitaine, a participatory process in Bordeaux, we propose, here, to present the issues of an uses analysis, the first results of a comparative study (physical density perception versus virtual density perception) and the exploratory and inter-disciplinary protocol being tested (geography, geomatic, cognitive).

MOTS-CLÉS : géovisualisation 3D, usage, démarche participative, densité.

KEY WORDS: 3D geovisualization, use, participatory approach, density.

1. Le contexte : expansion de la 3D, de la communication à la participation

L'expansion de la 3D est perceptible aussi bien au sein de la sphère professionnelle (diversification des usages techniques des structures institutionnelles) que du côté de la sphère citoyenne (émergence de nouveaux usages grand public des internautes). La multiplication des expériences et réalisations dans le monde de la géomatique et des applications 3D du territoire a conduit les collectivités à diversifier leurs usages. Ainsi, au-delà des logiques de communication, les représentations tridimensionnelles du territoire intègrent de plus en plus des dispositifs de concertation et de dialogue avec les citoyens autour des projets d'aménagement. Si ce mouvement est annoncé depuis de nombreuses années, il devient aujourd'hui une réalité sans pour autant susciter une réflexion avancée sur l'appropriation (et la non appropriation) de ces outils par les citoyens et leurs impacts en terme de rapport à l'espace et de participation.

D'une manière générale, on manque encore de cas concrets d'utilisation et de données empiriques correctement collectées pour évaluer l'impact de ces visualisations sur les procédures et les pratiques de ceux qui les produisent comme de ceux qui les utilisent (Joliveau, 2008). En l'occurrence, il semble qu'au delà des discours incantatoires ou alarmistes (Jacquinod, 2011), il reste encore des enjeux forts pour une recherche sur l'usabilité (Nielsen, 1993 ; Tricot, 2003) des paysages numériques en 3D (il s'agit en particulier de comprendre ce que les usagers *font* avec les environnements virtuels tridimensionnels qui leurs sont mis à disposition, quels *usages* ils en construisent et comment ils se les *approprient*) dans le cadre de démarches participatives (il s'agit alors de comprendre dans quelles dynamiques et stratégies organisationnelles s'inscrivent ces dispositifs en amont, au moment de leur conception¹) en adoptant une perspective à la fois cognitive et géographique (c'est-à-dire en s'intéressant à l'évolution des compétences spatiales qu'induisent ces technologies).

2. Les enjeux : évaluer le potentiel de la 3D par une analyse des usages.

Comme pour les études paysagères, les modèles urbains en 3D permettent de naviguer entre l'approche projectionnelle (la « *vue du dessus* ») et l'approche tangentielle (la « *vue du dedans* »). Les paysages numériques en 3D rebattent ainsi les cartes de la participation car la rupture entre le tangentiel et le projectionnel est moins nette, les transitions possibles, les distinctions plus floues avec des vues potentiellement hybrides. Par ailleurs, si le déplacement permet d'enrichir et d'ajouter les expériences spatiales, qu'en est-il de la navigation au sein de paysage numérique en 3D ? Le mode de déplacement par sa nature, impose lui-même au voyageur des sensations et expériences visuelles inédites, chacun « *est*

¹ Voir travaux de recherche réalisés dans le cadre du programme RAUDIN (Recherches Aquitaines sur les Usages pour le Développement des Dispositifs Numériques) – FEDER n°31462, Conseil Régional Aquitaine, Université Bordeaux 3

porteur d'une approche originale de l'espace qui façonne un paysage » (Desportes, 2005). L'analyse des déplacements au sein d'un paysage numérique en 3D peut-elle alors permettre de dégager une rhétorique de la navigation virtuelle, en référence à la « *rhétorique de la marche* » évoquée par Michel de Certeau (1990) ?

Par ailleurs, l'essor des démarches participatives dans le cadre de projets de développement territoriaux conduit progressivement à une revalorisation des représentations spatiales des territoires en jeu. Ces dernières apparaissent comme un support privilégié pour « *faire parler* » l'espace et externaliser les différentes approches du territoire. Si de nombreux auteurs ont montré le « *pouvoir des cartes* » et la difficile appropriation de la vue projectionnelle par certains acteurs (Couderchet et Ormaux, 2010), la question doit être reposée avec les paysages numériques en 3D afin de nuancer les incantations qui en font des outils de transparence démocratique. Ainsi, en complément des questionnements précédents, une analyse des non usages des paysages numériques en 3D pour comprendre les mécanismes de rejet voire d'exclusion des démarches participatives qui reposent sur de tels outils doit permettre d'interroger l'adéquation entre les objectifs qui leurs sont assignés et les pratiques qui en découlent.

In fine, les questions de distanciation, d'hybridation, de navigation, de stratégie organisationnelle et d'appropriation ou de non appropriation individuelle visent essentiellement à comprendre le rapport à l'espace qui peut s'établir entre un usager, un paysage numérique en 3D et un territoire donné. Ces questions devraient permettre d'alimenter des réflexions structurantes en géographie sur l'usage des représentations spatiales dans les démarches participatives, sur les impacts de ces nouveaux environnements virtuels et sur leur pertinence pour « *faire parler* » le territoire actuel ou futur (simulation). L'ensemble des objectifs précités vise ainsi à alimenter une réflexion plus globale sur l'usage des représentations spatiales dans la confrontation des savoirs négociés sur l'espace. Une étude focalisée sur les paysages numériques en 3D devrait ainsi permettre d'analyser l'usage d'un dispositif cartographique spécifique dans le dialogue entre savoirs experts (techniciens) et savoirs ordinaires (citoyens).

3. Le cas d'étude : la Fabrique Métropolitaine de Bordeaux

La Communauté Urbaine de Bordeaux (CUB) a lancé en 2010 une démarche participative sur son projet d'agglomération à l'horizon 2030 : « la Fabrique Métropolitaine ». Cette démarche s'est articulée autour de conférences, d'ateliers-débats, d'enquêtes mais aussi d'une plateforme Web intégrant un paysage numérique en 3D et invitant les Internauts à contribuer à la rédaction du projet (Nageleisen et Noucher, 2012). Lors de l'atelier de restitution de cette dynamique (décembre 2011), plusieurs constats sont apparus dont l'un qui a retenu notre attention : les habitants de la CUB jugent le passage de la métropole millionnaire à l'horizon 2030 (pour 700.000 habitants aujourd'hui) anxiogène (Bosdecher, 2011).

La qualité de vie actuelle est reconnue et les habitants ont peur qu'une densification de la ville la dénature.

La question des densités apparaît au centre des controverses actuelles sur les politiques urbaines. Ressurgie avec une vigueur nouvelle face à la montée du développement durable et à l'importance des besoins en logements, la densité est érigée comme objectif affiché des politiques publiques par la loi SRU (Solidarité Renouvellement Urbain), promulguée en décembre 2000 ou encore plus récemment avec l'annonce de l'ancien Président de la République (janvier 2012) d'augmenter de 30% la constructibilité des terrains pour densifier la ville et éviter de favoriser l'étalement urbain.

Il s'agit alors de questionner le débat sur la densité dans un triple objectif : mettre à mal quelques idées reçues en confrontant l'abstraction des indices de densité aux qualités de formes urbaines dans leur diversité, appréhender la complexité du sujet en croisant les notions de densité bâtie et de densité perçue, confronter l'analyse des densités perçues (sur le terrain) et des densités perçues et simulées (via les paysages numériques en 3D) pour évaluer leur pertinence comme outil de prospective et de participation. C'est sur ce dernier point que se focalisera le poster proposé ici.

4. Les 1ers résultats et le protocole d'observation en cours de définition : traçabilité et analyses des parcours virtuels

À la différence des applications cartographiques « traditionnelles », les paysages numériques en 3D n'offrent pas à l'utilisateur un environnement clos qui imposerait une échelle et une perspective de visualisation du territoire. Au contraire, ils offrent un environnement ouvert, dans la mesure où ils permettent l'usage de multiples points de vue. Ces environnements virtuels agissent ainsi comme une passerelle, entre la « vue du dessus » et la « vue du dedans » et autorisent une multitude de modes de navigations possibles. L'appropriation revient alors pour l'utilisateur à choisir parmi cet ensemble de possibles pour se réinventer son paysage numérique en 3D. En dépassant, le questionnement centré sur l'utilité versus la non utilité des outils, les recherches de Breton et Proulx (2002) ont dévoilé toute la complexité des processus de formation et de stabilisation des pratiques. Ainsi, l'appropriation des dispositifs relèverait davantage de la construction de sens chez les usagers qu'elle ne tiendrait aux caractéristiques techniques propres des technologies. Aussi, nous nous attacherons à l'étude concrète des pratiques déployées par les usagers, c'est-à-dire à ce qu'ils font avec des paysages numériques en 3D, quels usages ils en construisent et comment ils se les approprient.

Une approche quantitative « classique » a permis, dans un premier temps de produire, une série de référentiels cartographiques sur la densité métropolitaine : densité d'habitants par îlot, densité d'habitants par îlot ventilée par bâtiments,

densité de la surface bâtie, etc. Au-delà de cette approche quantitative, l'analyse qualitative en cours (densité perçue) a donné lieu à une étude exploratoire comparative entre l'évaluation de la densité perçue de 8 sites bordelais via l'usage d'un paysage numérique en 3D (Google Earth) et l'évaluation de la densité perçue de ces mêmes 8 sites via une visite sur le terrain par un groupe d'étudiants allemands ne connaissant pas Bordeaux. Un questionnaire (semi directif), des enregistrements vidéos des parcours virtuels réalisés par les cobayes et le tracé des parcours ont permis de disposer d'un corpus de données qui a été analysé. Bien que les résultats ne puissent être extrapolés, quelques éléments intéressants émergent : sur le différentiel entre les critères d'évaluation de la densité dans l'environnement virtuel et dans l'environnement physique, sur la prédominance de l'usage de la « vue du dedans », sur les différents parcours virtuels d'exploration, sur l'impact des pré-visites virtuelles sur les représentations des sites évalués, etc.

Ces premiers résultats seront présentés sur le poster accompagnés également d'un protocole de recherche inter-disciplinaire en cours de définition pour approfondir le sujet et qui regroupe des chercheurs en géographie spécialistes du paysage, en géomatique, en cognitique et en sciences de l'information et de la communication. Il s'agira en particulier d'expérimenter la complémentarité des résultats d'enquêtes qualitatives, d'analyses spatiales, d'analyses d'usabilité à travers notamment l'utilisation de dispositifs d'occulométrie (eye tracking) et de webométrie.

5. Bibliographie

- BRETON P. ET PROULX S., *L'explosion de la communication à l'aube du XXI^e siècle*, Éd. La Découverte, Paris, 2002.
- BOSDECHER L., Il ne faut pas avoir peur du million d'habitants, *Sud-Ouest (Bordeaux)*, édition du 28 octobre 2011.
- COUDERCHET L. ET ORMAUX S., From real-life experience to map, Using landscape as a tool for urban decentralized planning, *ITGO*, Nantes, 2010.
- DE CERTEAU M., *L'invention du quotidien, tome 1 : les arts de faire*, Éd. Gallimard, 1990.
- DESPORTES M., *Les paysages en mouvement*, Éd. Gallimard, 2005.
- JACQUINOD F., The adding value of the notion of intermediate object to describe the of 3D geovisualization. An example to the prevention of flood risk. *25th International Cartographic Conference*, Paris, 2011.
- JOLIVEAU T., Vers une approche géonumérique des paysages In Brossard T. et WIEBER J.-C. *Paysage et information géographique*, Traité IGAT, Hermès, 2008.
- NAGELEISEN S. ET NOUCHER M., Cartographie et participation : au-delà de l'implantation ponctuelle et de la 3D, *Journée de la Cartographie*, Besançon 2012.
- NIELSEN, J. *Usability engineering*. Boston, Academic Press, 1993.

TRICOT A. PLÉGAT-SOUTJIS, F., CAMPS, JF., AMIEL, A., LUTZ, G. & MORCILLO, A. Utilité, usabilité, acceptabilité: interpréter les relations entre les trois dimensions de l'évaluation des EIAH. In *Actes de la conférence EIAH 2003* Strasbourg, 15-17 avril 2003, ATIEF/INRP, Paris, 391-402, 2003

Apports de la technologie LIDAR dans l'objectivation écologique d'un territoire en amont d'un projet d'aménagement : aide à la caractérisation de l'habitat de l'Outarde canepetière dans la ZPS des Costières de Nîmes

Pierre-André Pissard ^{*,**} - Sylvie Vanpeene ^{***} - Sylvie Durrieu ^{**} - Laurent Albrecht ^{**}

** Irstea Centre de Grenoble*

Unité de recherche Ecosystèmes Montagnards, 2, rue de la papèterie - 38402 St Martin d'Hères

*** Irstea Centre de Montpellier*

UMR TETIS - Maison de la Télédétection, 500 rue Jean-François Breton - 34093 Montpellier Cedex 5

{pierre-andre.pissard, sylvie.durrieu, laurent.albrech}@teledetection.fr

**** Irstea Centre d'Aix-en-Provence*

Unité de recherche Ecosystèmes méditerranéens et risques - 3275 route de Cézanne CS 40061 13182 Aix-en-Provence Cedex

sylvie.vanpeene@irstea.fr

RÉSUMÉ. Dans le cadre du programme de recherche INTERMOPES¹, des investigations ont été menées sur les technologies qui permettraient d'améliorer la caractérisation d'habitat d'espèces afin de développer des méthodes et outils d'objectivation écologiques du territoire. Le poster que nous proposons présentera les résultats d'une expérimentation réalisée avec un LIDAR terrestre dans le but d'analyser et de cartographier finement les hauteurs de végétation et de relier les informations acquises avec la présence de mâles ou de femelles d'Outardes canepetière. L'objectif de ces travaux est d'affiner les connaissances sur la répartition de l'oiseau dans le paysage de la Zone de Protection Spéciale des Costières de Nîmes (France) et de développer une méthode généralisable d'analyse instrumentée des potentialités écologiques d'un paysage.

ABSTRACT. In the framework of the INTERMOPES research program, investigations have been lead on technologies allowing to improve characterization of wildlife habitats in order to

¹ INTERMOPES : Infrastructures de transport terrestre Rail et route et modifications induites sur les paysages, les écosystèmes et la société

develop methods and tools of ecological objectivization for territories. The proposed poster will present results from a LIDAR experimentation aiming to map precisely vegetation heights and to link it to the presence of male or female Little Bustard. The goal of the present work is to increase knowledge on bird distribution in Specific Protected Areas (Zone de Protection Spéciale des Costières de Nîmes, France) and develop a global evaluation method of the landscape ecological potential.

MOTS-CLÉS : Territoire, Aménagement, Infrastructure de Transport Terrestre (ITT), Ligne grande Vitesse (LGV), biodiversité, potentialités écologiques, Outarde canepetière, Lidar terrestre, Système d'Information Géographique (SIG), information spatiale, outils d'aide à la décision, diagnostic environnemental.

KEYWORDS: Terrestrial Transport Infrastructure (TTI), high speed railway line, biodiversity, ecological potentialities, Little Bustard (Outarde canepetière), terrestrial Lidar, Geographical Information System (GIS), spatial information, decision support system tools, environmental diagnosis.

1. Introduction

Dans le processus d'aménagement d'une Infrastructure de Transport Terrestre (ITT) sur un territoire, le diagnostic environnemental du site est une étape cruciale. En plus de participer à la définition technique du projet, il permettra d'alimenter l'étude d'impact et, le cas échéant, de définir les mesures compensatoires à mettre en œuvre. Cette phase doit par conséquent tenter de révéler la richesse écologique de l'espace qui sera impacté par l'ITT et les menaces que le projet fait peser sur les milieux qui seront traversés.

Lors de ce diagnostic environnemental, la prise en compte objective de la biodiversité requiert un haut niveau de connaissance des enjeux écologiques du site. Malheureusement, cette connaissance fait parfois défaut et des décisions d'aménagement du territoire sont prises alors que la biologie et l'écologie de certaines espèces présentes sur la zone de projet sont mal connues.

Une telle problématique est rencontrée sur le projet ferroviaire de Ligne Grande Vitesse (LGV) entre Nîmes et Montpellier (France), au niveau de la Zone de Protection Spéciale (ZPS) des Costières de Nîmes, avec un oiseau rare, l'Outarde canepetière (*Tetrax tetrax*).

La hauteur de végétation et plus globalement l'intervisibilité semble être un facteur limitant pour la présence de l'Outarde canepetière dans un paysage. Dans le but d'approfondir cette problématique et de valider des hypothèses de présence d'espèces dans certaines classes de hauteurs de végétation l'expérimentation mise en place a consisté, grâce à un lidar terrestre couplé à un appareil photographique délivrant des images couleurs à très haute résolution, en l'acquisition d'une information fine en 3D de la végétation d'une portion de paysage où la présence de l'oiseau était avérée.

2. Contexte et objectifs

2.1. *L'outarde canepetière*

L'Outarde canepetière est un oiseau steppique à l'écologie complexe menacée en France et qui bénéficie d'une protection internationale². L'effectif national d'Outardes (moins de 2000 mâles chanteurs) est globalement en baisse mais la population sédentaire de la zone méditerranéenne (80% de l'effectif national) est stable voire en augmentation (MEDDTL 2011). Cette région constitue une zone à enjeux particulièrement sensible pour le maintien de l'oiseau en France, avec notamment la ZPS des Costières de Nîmes (site Natura 2000) qui est impactée par un projet de ligne ferroviaire à grande vitesse.

² Convention de Berne annexe 2 ; Convention de Washington annexes 1 et 2 ; espèce quasi-menacée de la Liste Rouge UICN, Directive « Oiseaux » 79/409/CEE annexe 1, protection nationale l'arrêté du 29 octobre 2009

A la fin de l'hivernage, en mai, les Outardes mâles regagnent leurs arènes de parade (ou leks) pour la reproduction. Ces sites sont identiques d'une année sur l'autre sauf si le paysage ou les pratiques agricoles ont changé. Ces espaces bien précis du paysage sont constitués de zones de cantonnement de quelques hectares et de places de chant où les mâles font des parades dans des formations herbacées rases ou peu denses afin d'être vus par les femelles. Celles-ci par contre, très discrètes, préfèrent des milieux à fort recouvrement (MERIDIONALIS, 2004), se cachant dans une végétation plus haute adaptée à la dissimulation des nids et à l'élevage des poussins. Ces milieux peuvent être plus ou moins éloignée des leks qu'elles visitent pour s'accoupler (JIGUET et al., 2000).

Les outardes mâles et femelles opèrent une sélection complexe de leurs habitats durant la période de reproduction. La composition et la structuration du paysage influencent leur distribution dans le paysage (Wolff, 2001). Cette écologie complexe et différenciée entre mâle et femelle complique l'étude de l'espèce et bien souvent seuls les mâles en période de reproduction sont identifiés. Mais comme l'espèce est très territoriale et fidèle à son lieu de reproduction (LETT, 2002), la survie d'une population d'Outarde est intimement liée à l'existence de ces leks donc au maintien de la composition et de la structuration de la mosaïque de milieux qui les constituent.

2.2. Végétation et lidar

La structure de la végétation influence le comportement des oiseaux au point que certains modèles d'habitats d'espèces d'oiseaux reposent sur la caractérisation de cette structure (Bradbury et al. 2005). Cependant réaliser des relevés manuels de cette structure est extrêmement coûteux en temps, surtout si la densité de ces relevés doit permettre une description fine de l'hétérogénéité spatiale (Bradbury et al. 2005).

Le lidar (light detection and ranging) est une technique de télédétection particulièrement prometteuse pour caractériser la structure de la végétation à des échelles et avec des niveaux de précision adaptés aux besoins des études sur les habitats des espèces. Cette technique de télédétection, basée sur l'émission-réception d'un faisceau laser, permet en effet d'acquérir les coordonnées dans l'espace des cibles qui ont intercepté et rétrodiffusé le faisceau laser émis par le système. Les lidars aéroportés ont déjà été utilisés pour caractériser des habitats au travers de l'analyse de la structure de la végétation (Clawges et al. 2008; Goetz et al. 2007). Cependant le coût d'acquisition des données est encore élevé et la souplesse de mobilisation de ces systèmes assez faible. En revanche, les lidar terrestres, développés à l'origine pour le génie civil (Lichti et al. 2002), sont des scanners laser faciles à mobiliser qui permettent d'acquérir des nuages de points très denses. Ils constituent ainsi une alternative intéressante aux mesures manuelles pour mesurer sur des sites de dimension limitée la topographie locale (Resop et al. 2012) ou la structure de la végétation, en particulier forestière (Dassot et al. 2011; Durrieu et al. 2008).

3. Matériel et méthode

Le jeu de données pour cette étude a été acquis par un lidar terrestre VZ 400 de Riegl couplé à un appareil photographique numérique (Nikon D700 + objectif Nikkor très grand angle) afin de coloriser automatiquement le nuage de points. L'expérimentation a été réalisée sur un site avec une présence avérée d'Outardes (localisation visuelle et sonore des individus au moment de l'expérience).

Des cibles ont été réparties sur la zone d'étude et ont été géoréférencées à l'aide d'un GPS différentiel + tachéomètre. Plusieurs scans ont été réalisés pour couvrir la zone d'étude. Les scans ont ensuite été assemblés en s'appuyant sur les cibles homologues présentes dans les différents scans.

A partir du nuage de points ainsi acquis (figures 1 et 2) et d'une grille de taille de cellules de 10 x 10 cm, un modèle numérique de surface (MNS) a été calculé en sélectionnant le point d'élévation maximale dans chaque cellule de la grille.



Figure 1. *Nuage de points acquis par le lidar.*

Un modèle numérique de terrain (MNT) a également été généré à la même résolution en sélectionnant le point le plus bas de chaque cellule de la grille. La différence entre MNS et MNT a permis d'obtenir un modèle de hauteur de la végétation.

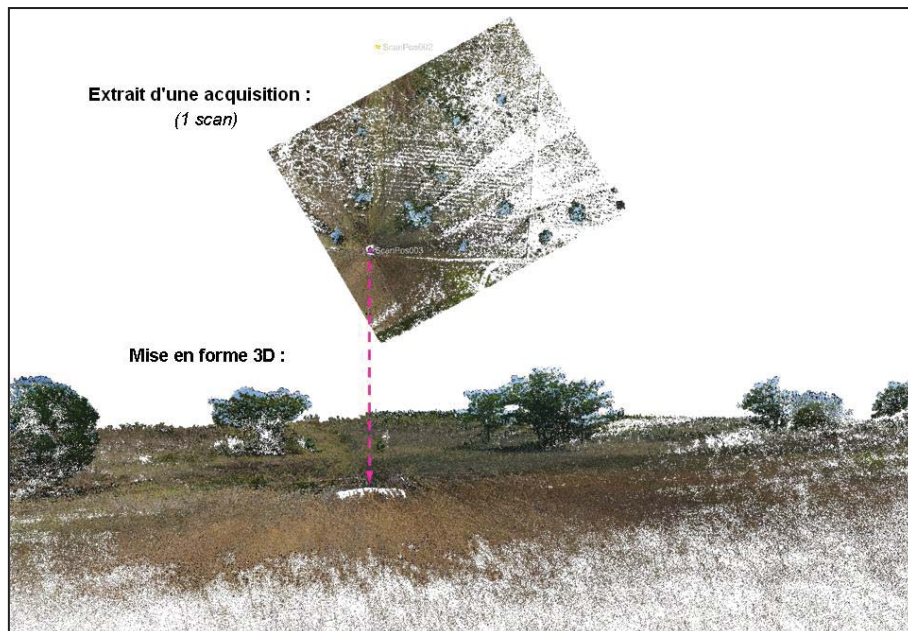


Figure 2. *Restitution en 3D d'un scan acquis par le lidar. L'appareil photographique numérique couplé au lidar permet de coloriser automatiquement le nuage de points.*

4. Premiers résultats

Dans un SIG (Système d'Information Géographique), les travaux ont ensuite consisté à croiser l'information acquise par le lidar avec les connaissances et données locales sur l'Outarde canepetière afin d'affiner les connaissances sur l'écologie de l'espèce en identifiant des corrélations entre densité et hauteur de végétation (figure 3) et présence de l'espèce.

L'objectif opérationnel est de réaliser une hiérarchisation de l'espace et de dresser une cartographie des potentialités écologiques d'un territoire.

La restitution est en 2D sous forme de couches SIG et de cartographies (figure 4) mais des résultats complémentaires en 3D sont envisagés.

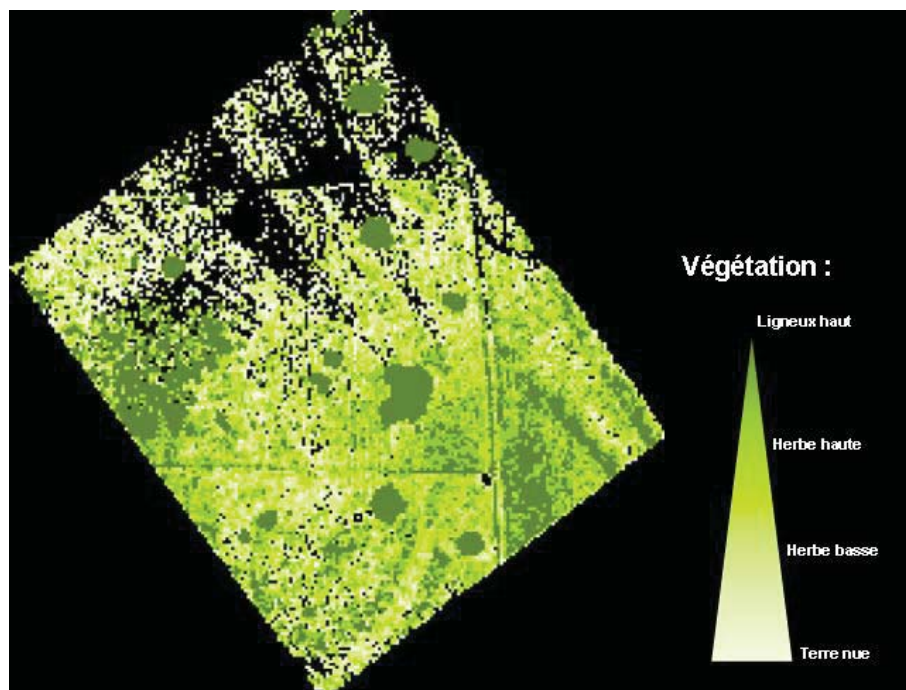


Figure 3. *Cartographie simplifiée des hauteurs de végétation à partir de données lidar (1 scan).*

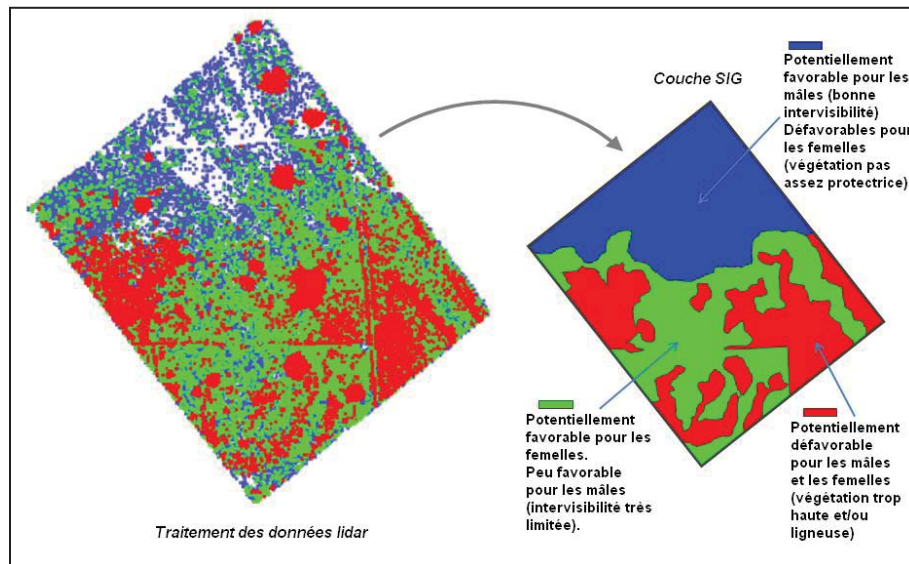


Figure 4. Hiérarchisation du paysage pour l'Outarde canepetière en fonction de la hauteur de végétation.

5. Perspectives

En plus d'apporter de la connaissance et une analyse complémentaire du paysage ciblée sur une espèce, ces travaux sont destinés à développer des méthodes et outils généralisables d'objectivation écologique d'un territoire. Ils doivent également permettre - d'apporter une expertise nouvelle sur l'écopotentialité d'un espace soumis à modifications profondes et - de définir des actions de conservation des patrimoines biologiques d'un territoire.

Une analyse de la mise en œuvre logistique, des résultats obtenus et des perspectives de cette expérience est proposée sur le poster.

6. Bibliographie

- Bradbury, R.B., Hill, R.A., Mason, D.C., Hinsley, S.A., Wilson, J.D., Balzter, H., Anderson, G.Q.A., Whittingham, M.J., Davenport, I.J., & Bellamy, P.E., Modelling relationships between birds and vegetation structure using airborne LiDAR data: A review with case studies from agricultural and woodland environments. *Ibis*, 147, 2005, p.443-452.
- Clawges, R., Vierling, K., Vierling, L., & Rowell, E., The use of airborne lidar to assess avian species diversity, density, and occurrence in a pine/aspen forest. *Remote Sensing of Environment*, 112, 2008, p. 2064-2073.

- Dassot, M., Constant, T., & Fournier, M., The use of terrestrial LiDAR technology in forest science: Application fields, benefits and challenges. *Annals of Forest Science*, 68, 2011, p. 959-974.
- Durrieu, S., Allouis, T., Fournier, R., Véga, C., & Albrech, L., Spatial quantification of vegetation density from terrestrial laser scanner data for characterization of 3D forest structure at plot level. In J.R. Ross Hill, Juan Suárez (Ed.), *Proceedings of SilviLaser 2008* (p. 325-334). Edinburgh, September 17-19, 2008, UK.
- Goetz, S., Steinberg, D., Dubayah, R., & Blair, B., Laser remote sensing of canopy habitat heterogeneity as a predictor of bird species richness in an eastern temperate forest, USA. *Remote Sensing of Environment*, 108, 2007, p. 254-263.
- Jiguet, F. & Wolff, A., Déterminer l'âge et le sexe des Outardes canepetières. *Tetrax tetrax* à l'automne. *Ornithos*, 7 (1), 2000, p. 30-35.
- Lett, J.-M., Fidélité et occupation de l'espace chez l'Outarde canepetière (*Tetrax tetrax*) dans le Boischaut Nord : conséquences sur les futurs aménagements. *Symbioses* 7, 2002, 43-50.
- Lichti, D.D., Gordon, S.J., & Stewart, M.P., Ground-based Laser Scanners: Operation, systems and Applications. *GEOMATICA*, 56, 2002, p.21-33.
- MEDDTL, Deuxième Plan national d'actions en faveur de l'Outarde canepetière 2011-2015, 2011.
- MERIDIONALIS, Programme de conservation de l'Outarde canepetière en Languedoc-Roussillon, dans le cadre du plan national de restauration de l'espèce (2002-2006). Deuxième phase : 2003-2004. Rapport Meridionalis, 2004, Montpellier.
- Resop, J.P., Kozarek, J.L., & Hession, W.C., Terrestrial laser scanning for delineating in-stream boulders and quantifying habitat complexity measures. *Photogrammetric Engineering and Remote Sensing*, 78, 2012, p.363-371.
- Wolff A., Changements agricoles et conservation de la grande avifaune de plaine : étude des relations espèce-habitats à différentes échelles chez l'outarde canepetière. Thèse de Doctorat, Université Montpellier II, 2001, Montpellier.

GéoPeuple: Analyse de l'évolution des territoires et de leur démographie sur 200 ans

Anne Ruas¹; Christine Plumejeaud²; Eric Grosso²; Benoit Costes²; Marie Christine Vouloir³; Claude Motte³; Hervé Le Bras³; Matthieu Cord⁴; Nicolas Thome⁴; Jonathan Guyomard⁴

(1) IFSTTAR Laboratoire MACS 58 Boulevard Lefebvre 75732 Paris Cedex 15

Anne.ruas@ifsttar.fr

(2) IGN Laboratoire COGIT 73 Avenue de Paris 94165 Saint Mandé

(3) EHESS- LaDéHis 190 avenue de France 75013 Paris

(4) LIP6/UPMC Malire 4, place Jussieu 75005 Paris

RÉSUMÉ. Depuis 200 ans, notre territoire s'est considérablement équipé, densifié au fil de l'accroissement démographique et des progrès technologiques. Par ailleurs, l'industrialisation a généré un important exode rural. L'objectif du projet GéoPeuple est d'étudier les relations entre l'évolution du paysage français et l'évolution de la répartition de la population sur le territoire. Ce que nous appelons paysage s'apparente ici à l'espace support ou l'espace topographique. Il s'agit donc de décortiquer puis d'extraire le contenu des cartes de Cassini (échelle 1/86 400) puis d'État-Major (échelle 1/80 000), de constituer des bases de données historiques, et d'étudier ainsi la densification des territoires et de leur population. Nous présentons ici une vue d'ensemble sur le projet GéoPeuple, les différentes tâches et les problématiques associées à la numérisation et à l'analyse de données historiques. Nous présentons des premiers résultats.

ABSTRACT. The landscape changed very significantly for 200 years. Landscape has become denser through demographical growth and technical progress. Moreover industrialization generated strong rural exodus. The aim of the GeoPeuple project is to study the relationships between the evolution of French landscape and its demography. We thus have to extract and analyse the content of Cassini and Etat-Major Maps in order to build topographical data bases necessary to study during to two last centuries the densification of landscape and its population. We present an overview of the project GéoPeuple, the different tasks and issues associated with digitization and analysis of historical data. We present first results.

MOTS-CLÉS : Analyse spatio-temporelle, données historiques, démographie, traitement d'image, ontologie

KEYWORDS: Spatio-temporal analysis, historical data, demography, image analysis, ontology

Le projet GéoPeuple est un projet de recherche français de 30 mois financé en partie par l'ANR. Son objectif est d'analyser conjointement l'évolution de la démographie française (connue depuis 200 ans à l'échelle de la commune) et l'évolution du territoire français. Il s'agit de mieux comprendre la dynamique locale de la démographie en essayant d'identifier les facteurs de différenciation entre communes, qu'ils soient anthropiques (liés par exemple aux réseaux de communication), naturels (liés par exemple au relief) ou éventuellement politiques (liés aux remembrements successifs). Les remembrements sont les redéfinitions des communes au cours du temps. Certaines communes disparaissent et sont absorbées par d'autres. Des communes disparues à une époque peuvent réapparaître. Parfois, plus simplement, des portions de territoires sont échangées entre deux communes. Au niveau démographique, si les grands mouvements de population sont assez bien connus (expansion de la population, exode rural) ainsi que les phénomènes d'extension et de densification de l'habitat et des infrastructures, le projet se propose de concevoir une méthodologie permettant une analyse fine et locale des dynamiques par intégration des données topographiques et démographiques et par comparaison des situations et des évolutions sur 200 ans.

Pour réaliser ces analyses, le projet GéoPeuple est décomposé en plusieurs tâches. Il faut en premier lieu se doter de représentations numériques de l'espace topographique sur des zones tests (Reims, Saint-Malo, Grenoble et Agen) à différentes époques. Puis il s'agit de croiser ces données avec les données démographiques afin d'analyser sur des corrélations par exemple en identifiant des communes de même type (qui se ressemblent sur un plan topographique et/ou démographique) et enfin observer les interactions entre démographie et remembrements territoriaux au cours du temps.

Données administratives et démographiques

Le laboratoire LaDéHis a dépouillé le Bulletin des Lois puis le Journal officiel pour relever les mouvements communaux (fusion, scission de territoires communaux) et les changements de noms officiels des communes. Le LaDéHis a également relevé, à partir des recensements nationaux, les populations communales sur la période 1793 jusqu'au recensement par sondage de l'INSEE. Appuyé sur ces informations complétées par celles provenant d'autres sources (ex: cadastre napoléonien), il a été réalisé une cartographie historique des limites communales au cours des deux derniers siècles.

Saisie des cartes topographiques

Au niveau topographique les sources utilisées sont la carte de Cassini pour la fin XVIII^{ème} siècle, la carte d'État-Major pour le XIX^{ème}, la série bleue de l'IGN pour 1960 et le RGE pour la période actuelle. Au-delà des difficultés liées à la numérisation de cartes anciennes (imprécisions géométriques, temps de saisie,

interprétation visuelle des symboles), une des difficultés vient de la compréhension du contenu des cartes tant au niveau des concepts que de la terminologie qui est peu utilisée aujourd'hui voire plus du tout. Le recours à des sources d'informations spécialisées (exemple : De Dainville 1964) et à l'analyse visuelle des cartes sont indispensables pour écrire un schéma de données permettant la saisie des données (figure 1) et la mise en relation des concepts et des données à différentes époques. Les travaux en cours montrent que si la reconnaissance visuelle de symboles n'est pas toujours aisée sur les cartes de Cassini (figure 2, gauche), les cartes d'État-Major ressemblent bien plus à la cartographie contemporaine. Cependant ces dernières décrivent la sémantique des objets par l'ajout de textes descriptifs et de toponymes qu'il n'est pas toujours facile d'associer aux symboles dessinés. Sur la figure 2, il est impossible d'identifier le(s) bâtiment(s) correspondant au Château de la Molière. L'interprétation du contenu peut être facilitée par l'utilisation d'autres sources telles que les données actuelles ou les minutes de terrain qui sont des cartes à plus grande échelle utilisées par les topographes lors des levés.

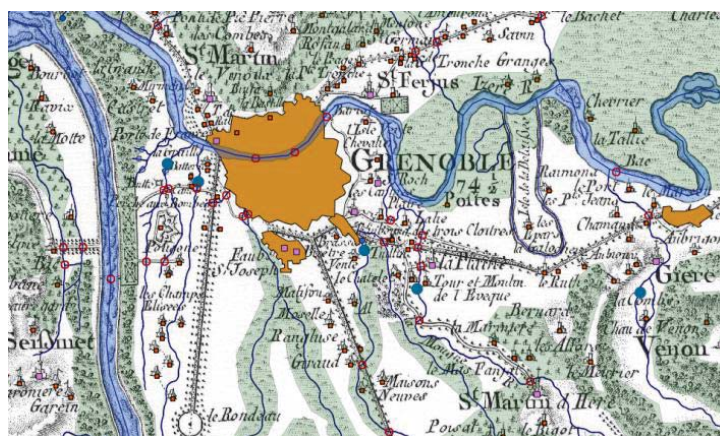


Figure 1. De la Carte de la Cassini à la base de données. Exemple de vectorisation



Figure 2. Difficulté d'identification des symboles sur Cassini (gauche) et des bâtiments nommés sur la carte d'État-Major (droite)

La saisie des données anciennes se fait d'abord par scannage, vectorisation puis mise en projection de la carte dans un système actuel (Lambert93). La mise en projection des cartes de Cassini passe par l'utilisation de points d'appui et d'une fonction de transformation validée par d'autres points d'appui (Costes et al 2012). La vectorisation est soit manuelle soit automatique. L'identification automatique de

symboles est étudiée par l'équipe Malire du LIP6 dans le cas des symboles ponctuels de la carte de Cassini par apprentissage supervisé. Il s'agit de concevoir un algorithme de classification à base d'exemples qui sont des symboles étiquetés manuellement sur les cartes. Ces exemples sont donc des imagettes extraites de cartes et dont la nature (par exemple 'moulin à eau') a été identifiée par un expert du LaDéHis. Le classifieur s'entraîne sur les exemples puis est utilisé sur de nouvelles cartes. Les imagettes sont décrites par des descripteurs qui décrivent les variations de niveaux de gris dans différentes directions à l'aide d'histogrammes de gradients orientés (HoG). Les symboles étant dessinés (et donc différents les uns des autres) l'apprentissage est délicat. Les méthodes en cours de conception font appel à des informations de contexte permettant de désambigüiser la classification en analysant le contexte spatial des symboles (Guyomard et al. 2012). La figure 3 ci-dessous illustre la détection automatique des calvaires avec ou sans information de contexte. En filtrant l'analyse à l'aide du contexte, les églises sont dissociées des calvaires

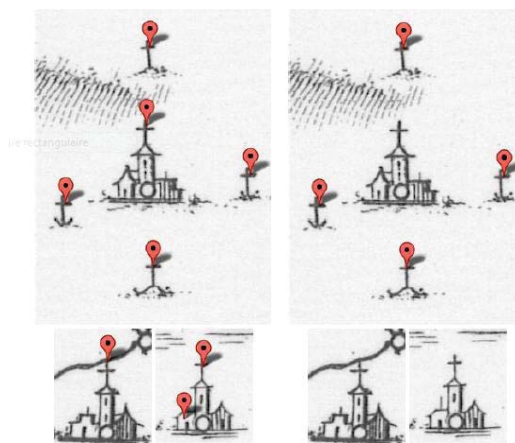


Figure 3. Identification automatique de calvaire sans (gauche) et avec (droite) information contextuelle.

Mise en correspondance des objets pour l'analyse des évolutions

Une fois les données vectorisées de façon manuelle ou automatique il s'agit d'apparier les données soit en interactif soit en automatique (Costes et al 2012). L'appariement des objets repose sur une mise en correspondance des concepts des différentes époques par alignement sémantique et inversement la correspondance entre concepts est confirmée ou infirmée par la correspondance des objets. Le cas des données historiques pose des problèmes liés 1/ à la précision moindre de données plus anciennes, 2/ à la forte différence de niveau de détail dans la description des informations (les objets anciens sont plus agrégés) et 3/ à la généralisation des symboles (la taille des symboles sur les cartes anciennes est très exagérée par rapport à la taille des entités qu'ils représentent). L'échelle de temps étant longue, il

est donc souvent difficile de savoir si des non correspondances proviennent des imprécisions, du niveau de détail ou de la forte évolution du paysage. Malgré ces difficultés, l'analyse des évolutions se faisant au niveau de la commune, certaines imperfections ne sont pas perceptibles. Il s'agit en effet de comparer dans un premier temps les signatures des communes. Pour ce faire la démographie, la nature et la densité des éléments naturels et anthropiques qui la composent, et l'étendue et la forme des tâches urbaines identifiées seront intégrés. La dernière difficulté est de faire émerger des règles en intégrant les effets des redécoupages territoriaux (Plumejeaud et al 2011). Mais puisque l'information démographique est représentée au niveau communal le questionnement double est : 'des territoires de même type (naturel et anthropique) ont-ils le même profil d'évolution (démographique et politique) ; si oui, pourquoi certains de ces territoires ont-ils eu des rythmes différents'.

Les analyses statistiques sont faites à l'échelle de la commune. Les premiers tests intègrent la densité de la population, l'aire du tissu urbain, le nombre d'habitats dispersés, le nombre de kilomètres de routes, le nombre de kilomètres de cours d'eau naturels et artificiels, le nombre de ponts, le nombre de bâtiments industriels, le nombre de constructions militaires, le nombre de postes, le nombre de bâtiments religieux, le statut administratif de la commune, l'altitude et la pente moyenne.

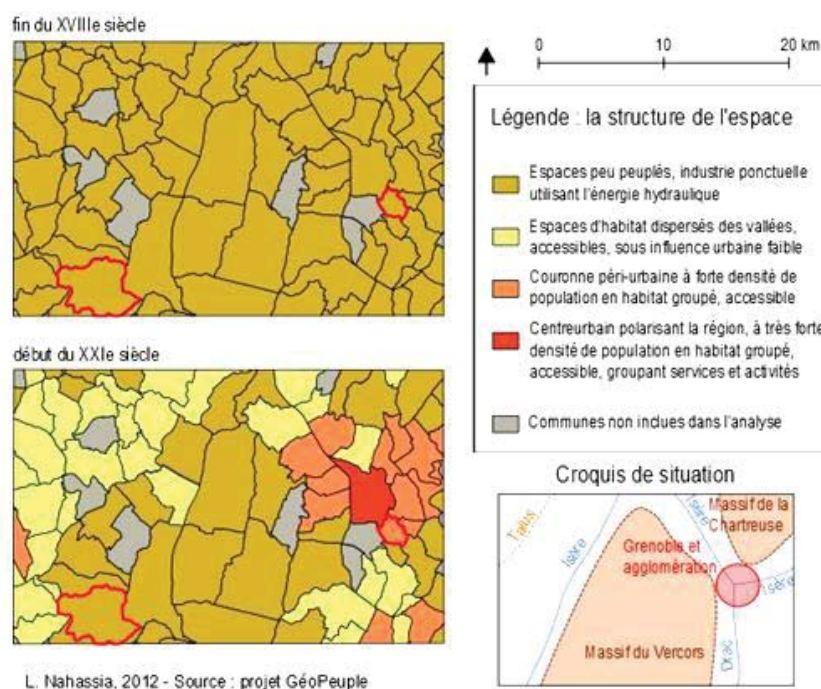


Figure 4. Exemple de classification des communes sous Cassini et aujourd'hui

De nombreuses analyses sont en cours pour faire émerger des connaissances à partir de ces données. Par exemple si l'on fait une classification associant les communes à l'époque de Cassini et à l'époque actuelle, sur la zone de Grenoble par exemple toutes les communes sous Cassini sont dans la même classe alors les communes d'aujourd'hui sont plus différenciées (figure 4, Nahassia 2012). On note cependant que les communes actuelles montagneuses (comme la commune Presles en bas à gauche dans le Vercors) est dans la même classe que les communes de Cassini alors que les communes dans les vallées ou proche d'une grande ville telle que Grenoble (comme la commune Eybens à droite en figure 4) ont changé de classe. On vérifie donc bien ici la spécialisation des communes sur les deux derniers siècles. Des résultats seront disponibles sur le site du projet (<http://geopeuple.ign.fr>) courant 2013. La disponibilité de ces données topographiques et démographiques ouvre un large champ d'investigation.

Bibliographie (extrait)

- Costes B., Grosso E., Plumejeaud C. « Géoréférencement et appariement de données issues des cartes de Cassini » SAGEO 2012.
- De Dainville F. « Le Langage des géographes » Editions A et J Picard 1964.
- Guyomard J., Thome N., Cord M., Artières T.. « Contextual Detection of Drawn Symbols in Old Maps». International Conference on Image Processing (ICIP) 2012
- Motte C., Pélissier J.-P., « Géonomenclature des lieux historiques habités» Archives de France, 2003.
- Nahassia L. «Analyse spatio-temporelle de l'évolution des communes françaises depuis la révolution française à aujourd'hui. Manipulation et exploitation d'une base de données géo-historique» Rapport de Stage Master 2 TGAE 2012
- Plumejeaud C., Mathian H., Gensel J., Grasland C., 2011, "Spatio-temporal analysis of territorial changes from a multi-scale perspective", *International Journal of Geographical Information Systems*, Vol 25, Issue 11, October 2011, p 1597-1612.
- Vouloir M-C., Motte C., « Frontières administratives et identités communales. Le cas de la France, XVIIIe-XXe siècles », *The Historical Review*, Athènes, vol. V, 2008.

Une méthode d'estimation du temps de parcours : application de recherche d'itinéraires dans un contexte d'urgence

Mohamed Ayet Allah Bilel Soussi* — **Jean-Michel Follin*** —
Guillaume Moreau** — **Alain Bouju***** — **Laurent Polidori***

**Ecole Supérieure des Géomètres et Topographes - ESGT - (L2G) - IRSTV
1 boulevard Pythagore - 72000 le Mans {mohamed.soussi, jmfolin,
laurent.polidori}@esgt.cnam.fr*

*** Ecole Centrale de Nantes - CERMA - IRSTV - 1 Rue de la Noë, BP92101, 44321
Nantes Cedex 3 guillaume.moreau@ec-nantes.fr*

****Laboratoire Informatique, Image et Interaction F-17042 La Rochelle cedex 1
alain.bouju@univ-lr.fr*

RÉSUMÉ. De nos jours, une prédiction sur le temps de parcours constitue une information capitale pour les automobilistes. Nous formulons l'hypothèse qu'un calculateur d'itinéraires basés sur l'exploitation de données historiques peut s'avérer plus fiable que les systèmes classiques. Dans cette optique, nous proposons une démarche générale pour estimer le temps de parcours destiné aux services d'interventions. Elle se décline en deux méthodes l'une par mesure directe et l'autre par une extrapolation spatiale en cas d'absence d'observations. Nous présenterons également nos premiers résultats suite à L'application d'un algorithme dynamique de recherche de plus court chemin.

ABSTRACT. Today, predicting travel time is important for motorists. We formulate the assumption that a route search algorithm based on real travel time estimation is expected to be more reliable than their conventional systems. For this purpose, we will show our method to estimate the travel time for French emergency units (SMUR) thanks to direct measurement and spatial extrapolation in the absence of observations. We will also present our first results further to an application of a dynamic algorithm for finding the shortest path.

MOTS-CLÉS: Trajectoire, véhicules d'urgence, estimation du temps de parcours, algorithme de recherche d'itinéraires,

KEYWORDS: Trajectory, emergency vehicles, travel time estimation, route search algorithm.

1. Contexte et objectif

De nos jours, Internet constitue un outil essentiel pour la diffusion d'informations sur les transports, les conditions de circulation et la recherche du meilleur itinéraire. Des services d'informations sur les conditions de circulation et de calcul d'itinéraires sont proposés par des sites web tels que sytadin (www.sytadin.fr) à Paris, coraly (www.coraly.fr) à Lyon ou encore Google Maps (maps.google.fr), Mappy (www.mappy.fr), Viamichelin (www.viamichelin.fr) et V-Trafic (www.V-traffic.com). Sur le portail web geovelo (www.geovelo.fr), adapté aux modes doux, le calcul d'itinéraires propose un meilleur itinéraire en fonction de plusieurs critères tels que la distance totale, la sécurité de l'itinéraire et l'effort fourni par le cycliste. Ces moteurs de recherche d'itinéraires ne sont pas adaptés pour les situations d'urgences. D'une part, ils n'utilisent pas des estimations réelles du temps de parcours destiné aux services d'urgence : ils se basent sur des valeurs standards de limitation de vitesse. D'autre part, les algorithmes de recherche d'itinéraires utilisés ne tiennent pas compte des incertitudes sur l'estimation du temps de parcours.

Dans cette optique, nous proposons un calculateur d'itinéraire pour les services d'urgence (SMUR) du Mans s'appuyant sur l'exploitation des trajectoires de leurs interventions.

Cette exploitation se décline en deux étapes : la première consistant à utiliser ces trajectoires dans le processus d'estimation du temps de parcours, la deuxième pour la validation de notre démarche (comparaison des itinéraires proposés par notre calculateur avec des trajectoires empruntées). Dans (Soussi et al., 2012), nous avons présentés une méthode générale d'enrichissement d'une base de données routière par des mesures issues de trajectoires. La méthode va de la phase de la modélisation des trajectoires jusqu'à l'implémentation dans une base de données spatio-temporelle. Les données sur les vitesses varient en fonction de différentes tranches horaires et des caractéristiques physique du réseau routier.

Nous traitons deux problématiques dans le cadre de nos travaux: la première est liée à la méthode d'estimation du temps de parcours, la deuxième au mode de recherche du plus court chemin le plus adapté pour notre projet. Pour l'implémentation de notre approche, une base de données des trajectoires appariées avec le réseau routier du département de la Sarthe est créée sous le SGBD relationnel libre PostgreSQL/PostGIS. Notre base contient plus de 300 trajets d'interventions.

Dans cet article, nous présentons d'abord quelques techniques pour estimer le temps de parcours et en particulier celle adoptée dans notre projet. La section 3 se focalise sur notre méthode générale d'estimation du temps de parcours. La section 4 concerne notre approche d'estimation du temps de parcours en cas d'absence d'observations. Dans la section 5, nous présentons nos premiers résultats suite à une

application d'un algorithme temporel de recherche du plus court chemin. Enfin la section 6 présente nos perspectives et conclut.

2. Technique de mesure du temps de parcours

Dans le domaine routier, une information fiable sur le temps de parcours s'avère très importante. Pour déterminer ce temps, on distingue deux familles de méthodes : la première est basée sur des mesures issues de détecteurs installés sur le réseau routier tels que les capteurs aux sols, les boucles magnétiques et les caméras. La deuxième se base sur des mesures issues des capteurs mobiles nommés véhicules traceurs (Bréhert et al., 2000). Pour améliorer la fiabilité de la mesure du temps de parcours, les mesures issues des deux techniques peuvent être fusionnées (Lefevre et al., 2006).

Dans la suite, nous présentons une revue de littérature sur l'utilisation des techniques basées sur des capteurs aux sols et sur des véhicules traceurs.

2.1. Capteur au sol

Dans le but d'améliorer la fiabilité de la mesure du temps de parcours, (Danech et al., 2007) proposent un algorithme de calcul: par itinéraire, type de jours et heure de la journée. L'algorithme prend en considération les événements exceptionnels comme les travaux et les accidents. Il exploite des mesures historiques issues de capteurs fixes localisés sur la chaussée (par exemple le débit, la vitesse et le pourcentage de poids lourds).

Pour la recherche d'un meilleur itinéraire dans un contexte multimodal, (Bousquet., 2010) utilise les mesures issues des capteurs aux sols comme technique pour estimer le temps de parcours. Dans le cas d'absence de capteurs sur les routes, l'auteur utilise une approche basée sur une technique par propagation spatiale. La technique consiste à propager les mesures issues de ces capteurs aux routes de même catégorie (autoroute ou route secondaire par exemple) en tenant compte de la proximité spatiale.

En effet, cette méthode est très utile pour mesurer le temps de parcours moyen pour plusieurs véhicules. Cependant, elle nécessite un ensemble d'équipement considérable.

2.2. Véhicules traceurs

Dans la littérature, plusieurs travaux utilisent les véhicules traceurs comme outils pour mesurer le temps de parcours.

(Liu et al., 2009, Schaefer et al., 2002) utilisent les taxis comme véhicules traceurs.

(Liu et al., 2009) montrent ainsi que, pour ces véhicules, il y a une concentration des trajets enregistrés sur certaines routes telles que celles qui relient les gares et les aéroports.

(Uno et al., 2009) utilisent les bus comme véhicules traceurs. Ils montrent que l'estimation des temps de parcours est plus pertinente par rapport à d'autres types de « véhicule traceur » puisque les itinéraires sont bien définis (les trajets habituels des bus).

En effet, ces techniques s'avèrent moins coûteuses et plus adaptées pour réaliser des études sur une catégorie spécifique de véhicule que les techniques basées sur des capteurs fixes.

Pour ces raisons, nous avons opté dans notre projet pour la deuxième technique en intégrant un GPS dans l'unique véhicule du SMUR qui sert à effectuer les interventions primaires¹.

3. Méthode d'estimation du temps de parcours

Notre stratégie d'estimation du temps de parcours peut être réalisée de deux façons : par méthode directe lorsqu'on a un nombre suffisant d'observations et par méthode indirecte dans le cas où nous n'avons pas assez d'observations enregistrées. Dans ce cas, nous procédons par une estimation basée sur une extrapolation spatiale des vitesses.

3.1 Méthode directe

Pour pouvoir procéder à une estimation directe du temps de parcours sur un tronçon routier, une condition doit être vérifiée : le nombre de passage sur celui-ci dans une tranche horaire doit être supérieur ou égal à une valeur seuil.

Dans le cas où cette condition est vérifiée, la vitesse estimée sur ce tronçon routier sera la moyenne des vitesses de tous les tronçons trajectoires qui empruntent ce tronçon routier. La variabilité de la vitesse, que nous nommons incertitude, est mesurée par l'écart type. Dans le cas où notre condition n'est pas vérifiée, on fait appel à la méthode indirecte pour estimer la vitesse sur ce tronçon.

3.2 Méthode indirecte

La méthode d'estimation indirecte est utilisée dans le cas où la condition de la première méthode n'est pas vérifiée (le nombre d'observations est inférieur à un seuil). Pour résoudre ce problème, on procède par une extrapolation spatiale. Notre idée consiste à déterminer une classification particulière en croisant plusieurs facteurs tels que la période d'observation, la largeur des chaussées, l'importance et

¹ Intervention sur les lieux de l'accident (voie publique, travail ou domicile du patient), distincte d'une intervention secondaire (transport d'un patient entre deux centres hospitaliers).

le nombre de voies des tronçons routiers. Pour chaque classification, nous allons créer plusieurs histogrammes qui représentent les effectifs des vitesses observées. La génération de ces histogrammes permet d'évaluer l'impact de plusieurs facteurs sur la variation de la vitesse.

Enfin, après avoir fixé la classification, on pourra déterminer la vitesse estimée sur ce tronçon en fonction de sa classe d'appartenance par l'application de la formule suivante.

$$(1)VitesseT = \frac{((NboT * vitesse_moy_troncon) + (NboC * vitesse_moy_classe))}{(NboT + NboC)}$$

- **NboT** : Nombre de trajectoires qui empruntent ce tronçon routier.
- **Vitesse_moy_troncon** : Vitesse moyenne observée sur ce tronçon routier.
- **NboC** : Nombre de trajectoires observées dans la classe.
- **Vitesse_moy_classe** : Vitesse moyenne observée dans la classe d'appartenance de ce tronçon.

L'équation (1) prend en compte la vitesse observée sur le tronçon routier et celle de sa classe d'appartenance en pondérant davantage la première que la seconde : elle permet donc de tenir compte de la spécificité de chaque tronçon routier.

4. Détermination d'une classification pour la méthode indirecte

Dans cette section, nous allons présenter d'abord les facteurs et les seuils des classes retenus pour notre classification. Ensuite, nous allons étudier l'impact des facteurs retenus sur la variation de la vitesse. Enfin, nous présenterons nos premiers résultats suite à l'adoption d'une classification.

4.1 Analyses des facteurs et les classes retenus pour l'estimation

Les critères adoptés dans notre analyse sont les tranches horaires et les caractéristiques physiques du réseau routier (nombre de voies, largeur des chaussées, etc). Ils sont définis de la façon suivante :

- a) Pour le facteur temporel, on adopte quatre tranches horaires [7h-9h [, [11h30-14h [, [16h30-19h [et une tranche horaire pour le reste de la journée. Le choix de ces tranches horaires est fixé selon le témoignage des ambulanciers (les heures de pointe).
- b) Concernant les classes liées aux largeurs des chaussées, on définit quatre classes (≤ 2 et < 4 mètres), (≤ 4 et < 7 mètres), (≤ 7 et < 10 mètres) et (≥ 10 mètres). Pour déterminer le nombre et les bornes de ces classes, nous avons dressé la courbe de la vitesse observée en fonction des largeurs des chaussées. Nous avons essayé de minimiser la variabilité des vitesses à l'intérieure des classes.
- c) Quant aux classes d'importance des routes, elles sont directement issues de la base de données routière de l'IGN. Les valeurs possibles reflètent le niveau

hiérarchique et vont de 1 (pour les liaisons entre métropoles) à 5 (pour les routes assurant la desserte à l'intérieur d'une commune).

d) Pour le nombre de voies, on définit quatre classes qui sont numérotées de 1 à 4.

4.2 Type de classification adoptée :

Afin de choisir la meilleure classification, nous avons créé plusieurs histogrammes pour chaque classe. Notre idée est de trouver des histogrammes dont la distribution est homogène, relativement concentrée et avec un seul mode si possible. Parmi les différentes classifications pouvant être définies, on recherche celle qui explique le mieux les vitesses observées. Par exemple si on utilise les quatre facteurs, compte tenu du nombre des classes existantes pour chacun d'entre eux, on peut aboutir à 320 classes, dont par exemple la classe **classe1**, telle que **Classe1: (P0, CL1, I1, N1)** (Figure1). L'histogramme issu de **Classe 1** peut être lu de la façon suivante : P0 correspond à la tranche horaire « reste de la journée », CL1 pour les largeurs des chaussées qui sont supérieures ou égales à deux mètres et inférieures à quatre mètres, I1 désigne l'importance de la voie qui est égale à 1 et enfin N1 qui correspond à une route dont le nombre de voies est égal à 1. Actuellement, nous avons retenu un seul critère de description du réseau routier : la largeur des chaussées. Une étude statistique nous a en effet révélé que c'était celui qui était le plus significatif.

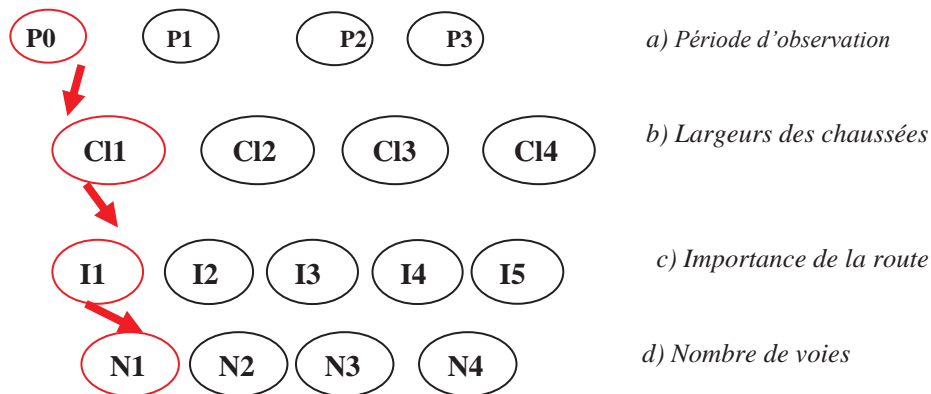


Figure 1. Les facteurs considérés dans la classification multicritères

Pour évaluer notre démarche, nous illustrons les histogrammes des effectifs des vitesses observées selon les classes liées aux largeurs des chaussées (par exemple figure 2, 3).

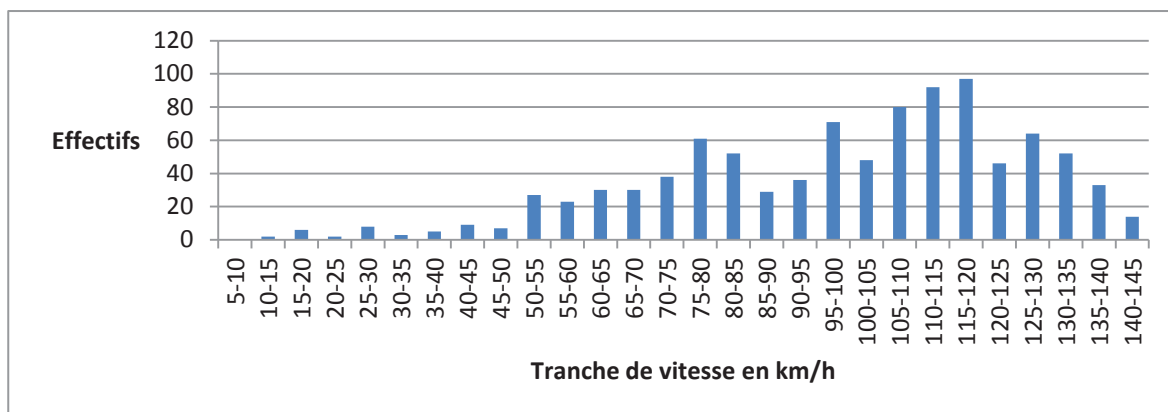


Figure 2. *Histogramme 1 des effectifs des vitesses observées sur des routes avec des largeurs de chaussées ≥ 7 et < 10 mètres*

Dans l'histogramme 1, les observations significatives commencent à partir de la tranche 50-55 km/h jusqu'à la tranche 140-145 km/h. Deux pics sont observés sur les tranches 75-80 et 115-120 km/h. On remarque aussi l'existence de deux modes. Pour avoir un seul mode, il faut intégrer d'autre facteur tel que les tranches horaires dans notre classification.

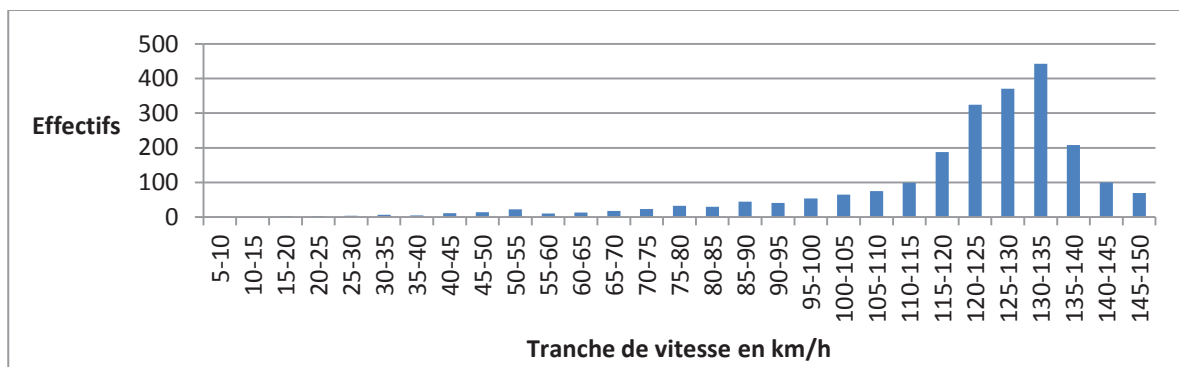


Figure 3. *Histogramme 2 des effectifs des vitesses observées sur des routes avec des largeurs de chaussées ≥ 10 mètres*

En observant l'histogramme 2, on constate une distribution très serrée entre les tranches 115-120 km/h et 145-150 km/h. Un seul mode est observé sur la tranche 130-135 km/h.

Dans les tableaux 1, 2, 3 et 4 nous présentons successivement un récapitulatif des moyennes, écarts types des vitesses ainsi que des effectifs (nombres de tronçons trajectoires) et nombres d'observations (nombres de trajectoires) qui y sont liés en fonction des tranches horaires enregistrées selon les largeurs de chaussées.

Tranche horaire	Vitesse moyenne (km/h)	Ecart type km/h	effectifs	Nombre d'observations
7h-9h	52	20	583	26
11h30-14h	56	27	509	31
16h30-19h	50	22	873	43
Reste de la journée	50	24	2817	137

Tableau 1 *Largeur de chaussées ≥ 2 et < 4 mètres*

Tranche horaire	Vitesse moyenne (km/h)	Ecart type km/h	effectifs	Nombre d'observations
7h-9h	63	29	2665	29
11h30-14h	63	27	3528	40
16h30-19h	67	28	6405	53
Reste de la journée	69	30	20074	171

Tableau 2 *Largeur de chaussées ≥ 4 et < 7 mètres*

Tranche horaire	Vitesse moyenne (km/h)	Ecart type km/h	effectifs	Nombre d'observations
7h-9h	90	26	3880	29
11h30-14h	87	28	4798	30
16h30-19h	93	27	8039	50
Reste de la journée	95	26	26745	163

Tableau 3 *Largeur de chaussées ≥ 7 et < 10 mètres*

Tranche horaire	Vitesse moyenne (km/h)	Ecart type km/h	effectifs	Nombre d'observations
7h-9h	90	26	229	11
11h30-14h	124	26	923	14
16h30-19h	126	27	2050	25
Reste de la journée	122	25	5960	74

Tableau 4 *Largeur de chaussées ≥ 10 mètres*

A partir des tableaux 1, 2, 3, 4 on constate que les valeurs des vitesses observées sont généralement légèrement plus élevées dans les tranches horaires « 16h30-19h » et « reste de la journée » que durant les autres périodes considérées. On remarque aussi que le facteur temporel n'influe pas réellement sur la variation de la vitesse observée sur des routes dont la largeur de chaussées est supérieure ou égale à deux et inférieure à quatre mètres (tableau 1). Concernant les vitesses observées sur des routes dont la largeur de chaussées est supérieure ou égale à dix mètres, seulement deux tranches horaires peuvent être retenues celle de « 7h-9h » et « reste de la journée » (englobant les trois autres classes).

5. Application de recherche d'itinéraires

Dans cette section, nous allons dans un premier temps présenter une revue de littérature sur les différents modes de recherche du plus court chemin afin d'en retenir un pour notre projet. Et dans un deuxième temps, nous allons tester notre méthode d'estimation du temps de parcours.

5.1 Etat de l'art

Dans la littérature, on distingue plusieurs catégories d'algorithme de recherche du plus court chemin : les catégories monocritères et multicritères qui peuvent être déclinées en modes statiques et dynamiques.

Récemment des états de l'art détaillés sur les algorithmes de recherches du plus court chemin en modes monocritère dynamique et multicritère statique ont été présentés respectivement dans (Bousquet, 2010) et (Sauvanet, 2011).

Dans le but de tenir compte de l'incertitude sur les poids des arcs dans un graphe en fonction du temps, (Hizem, 2008) introduit la notion d'un graphe dynamique avec intervalle et celle d'un facteur d'optimisme.

Le but de ce facteur est de rendre le poids d'un arc déterministe. Pour cela, l'auteur définit le poids d'un arc par la fonction suivante :

$p(i, j, t) = a_{ij}^t + (b_{ij}^t - a_{ij}^t) \alpha$, où a_{ij}^t et b_{ij}^t représentent respectivement la borne minimum et maximum d'un intervalle de valeur du poids d'un arc dans une tranche horaire. Concernant le paramètre α il représente le facteur d'optimisme. A titre d'exemple si l'intervalle $[a_{ij}^t, b_{ij}^t]$ représente un intervalle de vitesse observée sur un arc dans une tranche horaire, si on est optimiste alors $\alpha=0$ sinon si on est pessimiste $\alpha=1$.

5.2 Mise en œuvre

Parmi les différents modes présentés dans la section 5.1, on suppose que le mode dynamique avec intervalles (Hizem, 2008) est le plus adapté pour notre projet. En effet, nous avons besoin d'un mode dynamique puisque nos observations sont enregistrées dans quatre tranches horaires [7h-9h [, [11h30-14h [, [16h30-19h [et une tranche horaire pour le reste de la journée. L'utilité d'un graphe avec intervalles se traduit par le besoin de tenir compte de l'incertitude (écart-type) sur les différentes valeurs des vitesses moyennes observées sur les tronçons routiers.

Actuellement, dans notre projet et dans l'application de l'algorithme nous optons pour un degré d'optimisme $\alpha=0,5$. En d'autre terme, le poids d'un arc sera la vitesse moyenne observée dans l'optique où la distribution des vitesses pour chaque classe suit une loi gaussienne.

Pour évaluer notre calculateur d'itinéraire, l'algorithme de recherche d'itinéraire est appliqué sur des points de départ et d'arrivée des trajectoires déjà empruntées.

Avant de commencer notre étude de similarité (distance commune en mètres) entre les itinéraires proposés par notre calculateur et les historiques des trajectoires, nous proposons d'étudier la répartition des effectifs des trajectoires en fonction des tranches horaires étudiées. A partir du tableau 5, nous remarquons que la majorité de nos trajectoires est enregistrée dans la tranche horaire « reste de la journée ». Dans notre test nous n'utiliserons que les trajectoires enregistrées dans la tranche horaire « reste de la journée ».

Tranche horaire	Nombre de trajectoires	Pourcentage %
7h-9h	21	7
11h30-14h	41	13
16h30-19h	55	18
Reste de la journée	183	61

Tableau 5 répartition des effectifs des trajectoires en fonction des tranches horaires

Pour la phase de l'évaluation de notre calculateur, nous avons choisi dix itinéraires qui sont bien appariés avec le réseau routier. Notre objectif consiste à

déterminer le pourcentage de similarité, la différence en termes de temps de parcours entre les itinéraires proposés par notre calculateur et les trajectoires empruntées par le service d'urgence (tableau 6).

Date et heure	Pourcentage de similarité	Différence en termes de temps de parcours
2011-04-10 04 :30	94,3%	-17s
2011-04-30 21 :41	86,05%	-45s
2010-12-26 14 :58	85,9%	+27s
2011-03-08 09 :32	74,5%	-1min, 10s
2011-04-09 20 :05	72,3%	-6min, 3s
2011-04-13 09 :21	69,3%	-14s
2010-12-08 06 :27	67,6%	-31s
2011-02-03 21 :03	59%	+19s
2010-12-01 14 :52	45,5%	+1min46
2010-12-11 15 :07	13%	-2min7s

Tableau 6 Résultats de la comparaison entre les itinéraires proposés par notre calculateur et les trajectoires empruntées par le (SMUR)

A partir du tableau 6, on constate que 70% des itinéraires proposés par notre calculateur ont un pourcentage de similarité supérieur à 60%. Le gain au niveau du temps de parcours est en moyenne 52 secondes. Les 40% de dissimilarité sont dûs à divers sources d'erreurs tels que celles de la précision du GPS et de l'appariement. En effet, une de nos hypothèses de travail est que les ambulanciers connaissent bien le terrain. Autrement dit, l'historique des trajectoires nous sert à valider notre méthode d'estimation du temps de parcours. Pour avoir une meilleure estimation du temps de parcours il faut nécessairement minimiser le pourcentage de dissimilarité entre les itinéraires proposés par notre calculateur et les trajectoires empruntées par le service d'urgence.

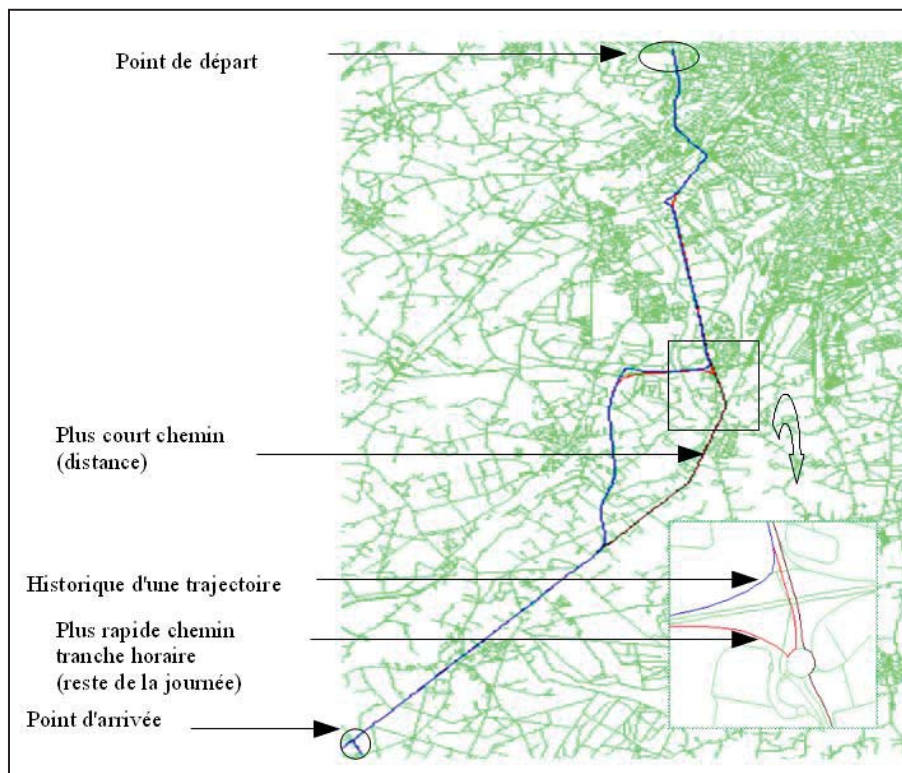


Figure 4. Application d'un algorithme temporel de recherche du plus court chemin

6. Conclusion et perspectives

Dans cet article, nous avons proposé une méthode générale d'estimation du temps de parcours basée sur une technique nommée « véhicules traceur ». En effet, cette technique s'avère la plus adaptée pour estimer un temps de parcours destiné à une catégorie particulière d'utilisateurs de la route : les services d'interventions. Notre méthode se décline en deux types de mesures : une mesure directe de la vitesse dans le cas où nous avons suffisamment d'observations (nombre de passage sur un tronçon routier dans une tranche horaire). Dans le cas contraire, l'estimation du temps de parcours est réalisée à partir d'une étude basée sur une analyse multicritère.

Pour évaluer notre méthode d'estimation du temps de parcours, nous avons appliqué un algorithme temporel de recherche du plus court chemin sur des points de départ et d'arrivée des trajectoires empruntées. Notre idée est de comparer les itinéraires

proposés par notre calculateur et les itinéraires déjà empruntés par le service d'urgence lors de leurs interventions. Nos résultats semblent prometteurs puisqu'on atteint plus de 60% de similarité.

Pour la suite de notre projet, et dans le but d'améliorer la fiabilité des estimations des temps de parcours issues de la méthode indirecte et pour simplifier la méthode de classification, notre idée consiste à dégager les corrélations qui peuvent exister entre les caractéristiques physiques du réseau routier et les vitesses observées. Ainsi, seuls les facteurs ayant un impact sur les vitesses et non corrélés seront retenus. Pour cela, nous pouvons utiliser la méthode statistique ACM (Analyse des correspondances multiples) (Kalbousi, 2010). Et enfin afin d'évaluer notre méthode d'estimation du temps de parcours, nous procédons par une comparaison d'un nombre significatifs des trajectoires empruntées par le service d'urgence et les itinéraires proposés par notre calculateur. Une autre perspective peut être envisagée qui consiste à prévoir un facteur d'optimisme dynamique (Hizem, 2008) en fonction des conditions météorologique. Une autre piste exploitable consiste en l'implémentation d'un système d'aide à la gestion de la perturbation du trafic en temps réel. Deux types d'événements exceptionnels seront pris en considérations : les événements prévus (les manifestations, les travaux publics et les changements des caractéristiques physiques du réseau routiers), et les événements imprévus (les embouteillages et les accidents).

Bibliographie

Bréheret L, Schettini F, Bernauer E, Barbier E « Traitement des données de trafic : besoins, état de l'art, exemple de mise en œuvre ». Etude réalisée dans le cadre du programme national de recherche et d'innovation dans les transports terrestres, 2000.

Bousquet A., Optimisation d'itinéraires multimodaux fondée sur les temps de parcours à l'échelle d'une agglomération urbaine dense. Thèse de doctorat, 2010. Ecole nationale des travaux publics de l'état.

Danech-Pajouh M., Joubert P., Blondeau F., Temps de parcours de préférence, Conventions PREDIM/INRETS 2007.

Hizem M.M. « Recherche de chemins dans un graphe à pondération dynamique : Application à l'optimisation d'itinéraires dans les réseaux routiers ». Thèse de doctorat école centrale de Lille, 2008.

Kalbousi A. « Méthode d'analyse et de classification des segments du réseau routier supérieur de la région de Montréal ». Mémoire de maîtrise en sciences appliquées. Ecole polytechnique de Montréal, 2010.

Lefevre E, El Faouzi N.E, « Estimation et fusion des temps de parcours routiers par la théorie de l'évidence », Workshop Fouilles de données complexes dans un processus d'extraction de connaissances, 6ème journée d'Extraction et Gestion des Connaissances, 2006.

- Liu K., Morikawa T., « Feasibility of using taxi dispatch system as probes for collecting traffic information », *Journal of Intelligent Transportation*. Vol 13, n°1, p16-27, 2009.
- Sauvanet G., « Recherche de chemins multi objectifs pour la conception et la réalisation d'une centrale de mobilité destinée aux cyclistes ». Thèse de doctorat, 2011. Université François Rabelais Tours.
- Schaefer R. P., Thiessenhusen K. U., Wagner P., « A traffic information system by means of real time floating car data ». In *Proceedings of the 9th World Congress on ITS*, Chicago, 2002.
- Soussi M. A. A. B., Follin J. M., Moreau G., Bouju A., Polidori L., « Enrichissement d'une base de données routière à partir de trajectoires GPS de véhicules d'urgence : Application à l'aide au choix d'itinéraires ». *Ingénierie des Systèmes d'information*, vol 17, n°1, p35-54, 2012.
- Uno N., Kurauchi F., Tamura H., Iida Y « Using bus probe data for analysis of travel time variability ». *Journal of Intelligent Transportation Systems*, vol 13 n°1, p2-15, 2009.

EO4REDD

Earth Observation for Reducing Emissions from Deforestation and forest Degradation

Apport de la télédétection dans la Réduction des Emissions provenant de la Déforestation et de la Dégradation des forêts

Lauriane Cayet* — Nathalie Stephenne — Mathieu Rahm*

* Walphot, 97 rue Van Opré, B-5100 Namur (Belgium)
(lauriane.cayet,mathieu.rahm)@walphot.com

ABSTRACT. The International community agreed on the need and emergency of controlling greenhouse gas emissions. Reducing Emissions from Deforestation and forest Degradation (REDD) is one major objective in climate change mitigation. Increasing willingness in setting up REDD policies to compensate for avoiding deforestation need to refer to a robust, reliable, cheap and certified monitoring system of carbon stock in tropical forest. We present the project called of EO4REDD which aims at developing such a system. For best efficiency, cost and large extent applicability, this system relies on Earth Observation (EO). The contribution of geostatistics in the development of this system is evaluated both in the sampling frame and the acquisition, processing and fusion of spatial data.

RÉSUMÉ. La nécessité (et l'urgence) du contrôle des émissions de gaz à effet de serre est désormais admise par la communauté internationale. La lutte contre la déforestation et la dégradation des forêts (REDD / Reduced Emissions from Deforestation and forest Degradation) est apparue comme un élément prioritaire de la bonne gouvernance climatique. La volonté accrue de mettre en place des politiques REDD définissant des incitants économiques à réduire la déforestation présuppose un système robuste et fiable de suivi des stocks de carbone forestiers. Le projet appelé EO4REDD, présenté ici, a pour objectif de développer un service certifié de suivi des stocks de carbone forestiers. Pour des raisons d'efficacité, de coût et d'accessibilité, ce système s'appuie sur le traitement en routine de données d'Observation de la Terre (OT). L'apport de la géostatistique dans le développement de ce système est évalué non seulement dans l'échantillonnage, mais aussi dans les acquisitions, le traitement et la fusion des données spatiales.

KEYWORDS: remote sensing, forest, Central Africa, REDD

MOTS-CLÉS: télédétection, forêt, Afrique centrale, REDD

1. Introduction

Forests play an important role in the global carbon budget, acting either as carbon sinks or sources (Van der Werf et al., 2009). Reducing emissions from deforestation and degradation, as well as conservation, sustainable management of forests and enhancement of forest carbon stocks, known as “REDD+”, is a major objective in climate change mitigation. In 2007, the international community agreed on the need and the emergency to accurately and precisely Measure, Report and Verify (MRV) carbon stocks, sinks and emissions.

Earth Observation (EO) technologies are the most appropriate to provide spatially explicit information and to cover large areas of forest (De Fries *et al.*, 2007). While EO has been extensively used for monitoring deforestation, there is no agreed way for monitoring forest degradation yet. In fact, not all processes associated with forest degradation can be assessed by remote sensing (Herold et al., 2011). Satellite data of multi-spectral and fine spatial resolution satellites (from RapidEye constellation of 6-m to WorldView of 0.5-m resolution) make it possible to monitor and to map not only deforestation but potentially some degradation and forest regrowth (Böttcher et al, 2009). The higher spatial resolution of WorldView-2 could improve the distinction between different forest stages and types.

In this article, we present EO4REDD, a Walloon project, led by WALPHOT to develop a robust, cheap and reliable certified monitoring system of tropical forest carbon using EO inputs. WALPHOT, its partners (Cap conseil, Climact, N.A.D.A.R) and sub-contractor (Université Catholique de Louvain) are almost at the end of the first year of this 3-year project. Tested in the Democratic Republic of Congo (DRC) in the first year of the project, EO4REDD aims to reach the reliability required for an operational solution to implement REDD+. This paper presents here EO4REDD’s framework and methods to conduct a change detection analysis of forest cover and carbon between 2011 and 2012.

2. Study area

The study area is situated in the northeast of the DRC, about 150km northwest of Kinshasa, in the Bandundu province (3°00’S /17°00’E). This area comprises many REDD-related project areas (CBFF agroforestry project area, WWF conservation projects, etc ...). It encompasses about 6,500 km² of dense forests, or 13,000 km² of forests, when counting in forest-savanna mosaics, closed to open deciduous woodland, as well as savanna woodland-tree savanna. The main threat to the dense moist forest is from expanding subsistence agriculture – through slash-and-burn activities, wood extraction (commercial, charcoal production), some through illegal logging, and infrastructure extension.

3. Data

Satellite optical data: For a “wall-to-wall” coverage (i.e. analysis of satellite data that covers the full spatial extent of the forested area), we chose 6.5-m resolution RapidEye data. The “red-edge” spectral region of RapidEye is relevant for vegetation characterization because it is sensitive to changes in chlorophyll content¹. Furthermore, we decided to work with 0.6-m resolution Quickbird-2 and 0.5-m resolution WorldView-2 to be able to detect the more subtle changes. Eckert (2012) found a positive correlation between this band and Above Ground Biomass (AGB) data. All RapidEye images are almost cloud-free, orthorectified and radiometrically calibrated (Level 3A). Quickbird and Worldview images are radiometrically calibrated and orthoready.

Aerial photos: A non-random stratified sampling scheme of aerial photographs supports the acquisition plan of aerial photographs to reduce both costs and the time for analysis. The resolution of the digital photographs is 10-20cm. Derived orthophotos and Canopy Height Models (CHM) are the validation products of the wall-to-wall maps.

Field data: They will be acquired by the *Observatoire Satellital des Forêts d'Afrique Centrale* (OSFAC), during a forest inventory early 2013 to develop stratum-specific allometric equations and apply emission factors to the forest cover change surface area and intensity.

4. Methods

4.1. Methodological framework

EO4REDD's approach is to combine activity data and emission factors to quantify the emissions or removals per forest cover change and land cover change units. We derive carbon net emissions from areas of forest cover changes only (inc. forest cover loss and gain due to deforestation, degradation and forest regrowth in abandoned pastures and agriculture land, as well as plantations). Our methodology is derived from the “Gain & Loss” method for countries with no National Forest Inventories (NFI), such as the DRC (IPCC, 2006) and the methodological framework developed by Maniatis and Mollicone (2010).

4.2. EO4REDD services

¹ <http://www.rapideye.com/solutions/agriculture.htm>

We develop and propose a method based on modular services and products as presented in **Figure 1**, based on the expertise of the EO4REDD partners.

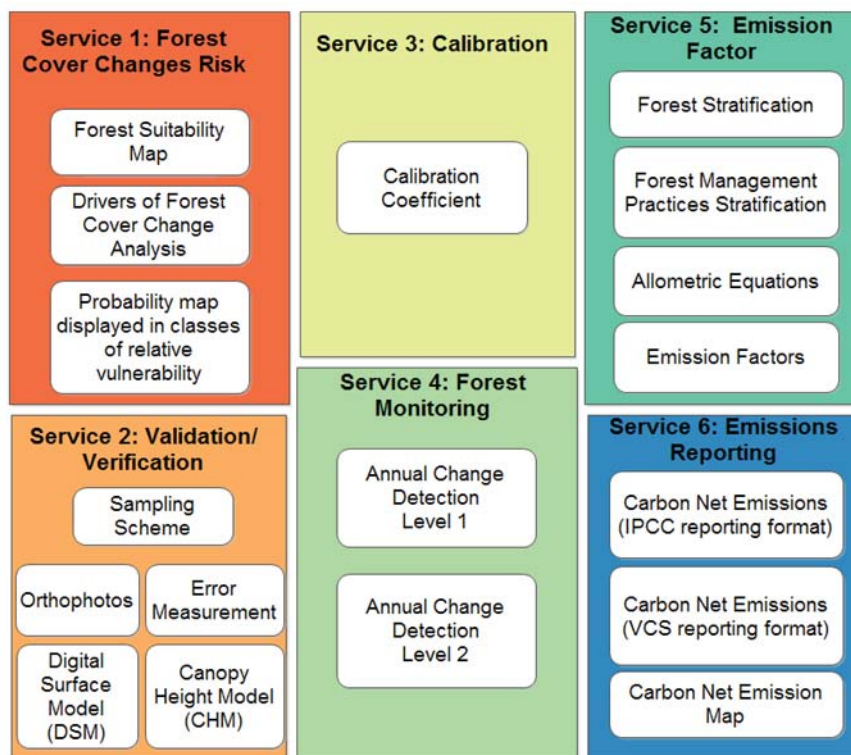


Figure 1. EO4REDD six services

Service 1 consists in producing a map of forest cover changes vulnerability, using forest types, soil and biogeographical regions data, as well as ancillary data (e.g. roads, rivers, populated places) used as proxies for modeling human-induced deforestation/degradation influence. Spatial relationships between forest cover changes and proxy data are explored by geostatistical techniques, such as analysis of covariance to explore the behaviour of the change variable (e.g. canopy gaps) with distance and multivariate interpolation (kriging) using the results of multivariate semi-variogram.

The output of service 1 supports the sampling scheme of the aerial photographs in **Service 2**. The quality of low resolution products is checked by higher resolution data. We intend to use Service 2 outputs (orthophotos, differential DSM) to validate the 2 products of Service 4.

The outputs of **Service 4** (Annual Change Detection maps) are the core products. They are developed using an object-based image analysis approach in the eCognition software, which enables to partition an image into precise object shapes (segments) and assign them a class (Baatz and Schäpe, 2000). The approach is based on multi-date segmentation using two satellite images of the same area at two different dates to create multi-date objects, which group pixels that are spatially, spectrally and temporally similar (Desclée et al., 2006). Firstly, “large objects” (level 1) are delineated with a multi-date segmentation approach, using spectral, scale and compactness parameters. Secondly, for each image, the output of level 1 constrains the detection of small patches of bare soil (canopy gaps) delineated with “smaller objects” (level 2). **Figure 2a and 2b** illustrate the two levels of image segmentation for t_1 and t_2 respectively. We eventually compile statistics of canopy gaps per level 1 object. Five levels of forest degradation are defined for each object according to the percentage difference of canopy gaps between t_1 and t_2 (**Figure 3**). Objects whose percentage of bare soil is higher in t_1 compared to t_2 are classified as regeneration.

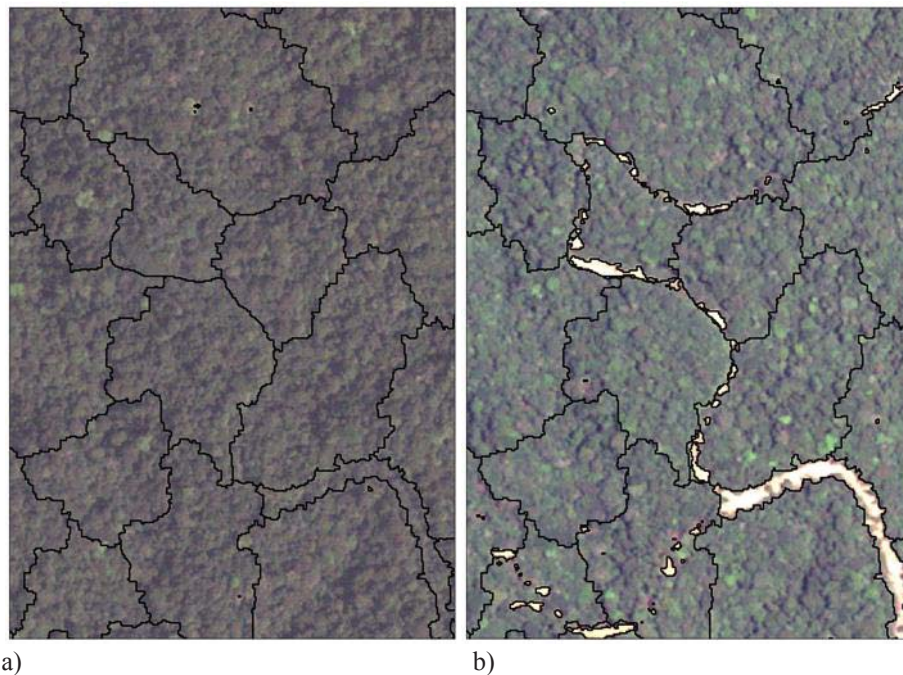


Figure 2. Two levels of image segmentation for t_1 (a) and t_2 (b)

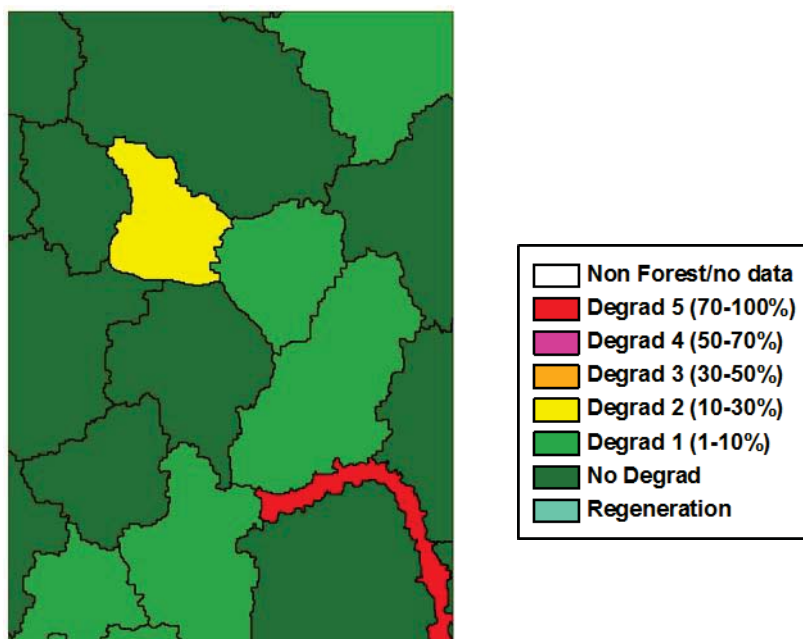


Figure 3. Five levels of degradation based on the percentage difference of bare soil calculated by "large object" between t_1 and t_2

In **Service 3**, Quickbird and WorldView, higher resolution imageries, improve detection made from RapidEye using either an image fusion method to improve the spatial detection of forest cover changes, such as a Bayesian approach (Fasbender, et al., 2009) or improve forest cover change estimations using a linear regression.

Activity data derived from service 4 are multiplied by emission factors derived from **Service 5** (using available regional / country-specific data, completed with local data).

Service 6 presents the results in the specific format of IPCC or Certification Standards such as Verified Carbon Standards.

4.3. Quality Control and Assurance

EO4REDD follows the IPCC recommendations on MRV systems, general quality criteria developed for EO4REDD needs. Furthermore, EO4REDD complies with a rigorous quality check procedure. Additionally, EO4REDD will be using geostatistical techniques to reduce and characterize uncertainties on forest changes and forest emissions estimation.

5. Perspectives

“We need tools, such as EO4REDD, for estimating emissions and removal from deforestation in the DRC” (Christophe Musampa, head of the spatial information unit, Ministry of the Environment, Nature, Conservation and Tourism of the DRC, 2012).

EO4REDD’s potential users include the Ministry of Environment, Nature, Conservation and Tourism of the DRC, as well as service providers of certification, verification, and validation of REDD+ projects in Central Africa

EO4REDD brings EO primarily to the monitoring of global warming, but also addresses global challenges, such as deforestation and forest governance issues (illegal logging or unsustainable practices). Indirectly, EO4REDD provides the civil society with useful information for objective and active engagement in decision-making processes.

A high level of recognition and the certification of this method open opportunities to validate REDD efforts on the carbon markets. The deliverables of this method (maps and reports) will contribute to the current UN-REDD initiatives to establish a reliable MRV system.

6. Bibliographie

- Baatz, M. and Schäpe, A., “Multiresolution Segmentation – an optimization approach for high quality multi-scale image segmentation”, In: *Strbl, J. et al., Angewandte Geographische Informationsverarbeitung XII*, Karlsruhe, 2000, p. 12–23.
- Böttcher, H., Eisbrenner, K., Fritz, S., Kindermann, G., Kraxner, F., McCallum, I., Obersteiner, M., “An assessment of monitoring requirements and costs of Reduced Emissions from Deforestation and Degradation”, *Carbon Balance Management*, 4(7), 2009. Available at: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2741441/pdf/1750-0680-4-7.pdf>.
- De Fries, R., Achard, F., Brown, F., Herold, M., Murdiyarso, D., Schlamadinger, B. and Souza, C. M., “Reducing greenhouse gas emissions from deforestation in developing countries: Considerations for monitoring and measuring”, *Environmental Science and Policy*, 10, 2007, p. 385-394.
- Desclée, B., de Wasseige, C., Bogaert, P., Defourny, P., “Tropical forest monitoring by object-based change detection: Towards an automated method in an operational perspective”, 1st International Conference on Object-based Image Analysis (OBIA). Salzburg University, Austria July 4-5 2006. Available at: http://www.isprs.org/proceedings/XXXVI/4-C42/Papers/13_Automated%20classification%20IC%20I%20%20Forest/OBIA2006_Desclée_et_al.pdf

- Herold, M., Roman-Cuesta, R.M., Mollicone, D., Hirata, Y., van Laake, P., Asner, G.P., Souza, C., Skutsch, M., Avitabile, V. and MacDicken, K., “Options for monitoring and estimating historical carbon emissions from forest degradation in the context of REDD+”, *Carbon balance and management*, 6(1), 2011, p.13.
- Van der Werf, G.R., Morton, D. C., DeFries, R. S., Olivier, J. G. J., Kasibhatla, P. S., Jackson, R. B., Collatz, G. J. and Randerson, J. T., “CO₂ emissions from forest loss”, *Nature*, 2, 2009, p. 737-738.
- Eckert, S., “Improved Forest Biomass and Carbon Estimations Using Texture Measures from WorldView-2 Satellite Data”, *Remote Sensing*, 4(4), 2012, p.810-829. Available at: <http://www.mdpi.com/2072-4292/4/4/810/> [Accessed July 18, 2012].
- Maniatis, D. and Mollicone, D., “Options for sampling and stratification for national forest inventories to implement REDD+ under the UNFCCC”, *Carbon balance and management*, 5(1), 2010, p.9. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3018366&tool=pmcentrez&rendertype=abstract> [Accessed August 29, 2011].
- IPCC, “2006: IPCC Guidelines for National Greenhouse Gas Inventories”, IPCC/IGES, Hayama, Japan, 2006. <http://www.ipcc-nggip.iges.or.jp/public/2006gl/ppd.htm>
- Fasbender, D., Obsomer, V., Bogaert, P. and Defourny, P., “Updating Scarce High Resolution Images with Time Series of Coarser Images: a Bayesian Data Fusion Solution”. In: *Sensor and Data Fusion*, N. Milisavljevic Ed., InTech, Croatia, 2009, p.246–261.

Une nouvelle méthode de calcul d'isovist en 2 et 3 dimensions

SULEIMAN Wassim¹ JOLIVEAU Thierry¹ FAVIER Eric²

¹*EVS-ISTHME CNRS UMR 5600, Université Jean Monnet - Saint-Etienne. was-sim.suleiman,thierry.joliveau@univ-st-etienne.fr.*

²*DIPI EA 3719 ENISE (École Nationale d'Ingénieurs de Saint-Etienne). eric.favier@enise.fr.*

RÉSUMÉ. L'analyse de visibilité est une question importante de la recherche qui peut trouver des applications dans de nombreux domaines: sécurité, conception de réseau sans fil, gestion du paysage, mise en œuvre d'accès piétonniers.... La prise en compte de la troisième dimension dans le calcul de visibilité est un réel défi. Seules quelques solutions peuvent détecter les obstacles 3D qui limitent l'isovist. Dans cette communication, nous présentons un nouvel algorithme qui peut détecter tous les objets qui bloquent la vue dans un environnement 3D reconstitué numériquement intégrant le relief. Une démonstration avec des données SIG est également effectuée.

ABSTRACT. The isovist or vision field is an interesting topic with many applications in different fields: security, wireless network design, landscape management and analysis, pedestrian access.... Taking in account 3D environment is a very challenging task. Only a few solutions can detect 3D obstacles that limit the vision field. We present in this paper a new algorithm that can detect all the objects which block the sight in a 3D environment including the ground surface. A demonstration with GIS data is also given.

MOTS-CLÉS : Isovist, espace visible, SIG, Photométrie.

KEYWORDS: Isovist, GIS, Photometry.

1. Introduction

La question de la délimitation automatique de l'espace visible d'un observateur intéresse plusieurs domaines d'application : la sécurité pour le placement des caméras de surveillance dans une rue, une banque ou un musée ; les télécommunications pour un placement optimisé de bornes wifi afin d'obtenir une bonne couverture avec un nombre minimum de bornes ; la publicité urbaine afin de localiser les endroits les plus favorables à l'installation de panneaux ou d'enseignes lumineuses. Surtout l'analyse de la visibilité est centrale dans le domaine de la gestion des paysages et de l'aménagement urbain. On sait que l'accès à une vue est un facteur important de choix des localisations résidentielles, qu'il s'agisse d'une vue sur l'eau (Luttik 2000), sur la forêt (Gueymard 2006) ou sur le paysage en général (Jim and Chen 2006). Robert (2011) a ainsi bien mis en évidence l'importance de la vue sur mer dans l'urbanisation du littoral de la Côte d'Azur. L'effet sur les prix immobiliers de la nature des vues a aussi été démontré par plusieurs études (Miller 2001), (Cavailhès, Brossard, Hilal, Joly, Tourneux, Tritz, and Wavresky 2007).

Dans la littérature, le premier emploi du terme d'isovist revient à Benedict (Benedict 1979) qui le définit comme l'espace visible à partir du point où se situe l'observateur supposé. Gibson (1983) Conroy Dalton et Bafna (2003) le définissent eux comme l'agrégation des lignes de vue qui partent de l'observateur. On trouve en fait deux termes en anglais pour nommer cet espace visible : « isovist » et « viewshed », dont l'emploi varie en fonction des méthodologies et des applications. « Isovist » est plus utilisé pour les calculs de visibilité dans un environnement urbain ou à l'intérieur d'un bâtiment, avec des données au format vectoriel. « Viewshed » désigne plutôt l'espace visible d'un observateur situé dans un environnement naturel et rural et le calcul est fondé sur des données raster (Floriani and Magillo 2003).

Le calcul automatique de l'isovist pose des problèmes spécifiques dans un environnement en 3 dimensions, en particulier sur un sol qui n'est pas plan.

2. Etat de l'art

2.1. Calcul de l'isovist en 2 dimensions

Dans la littérature, la méthode la plus connue pour calculer l'isovist est le lancement des rayons ou « ray tracing ». Le principe est de lancer M rayons dans toutes les directions autour l'observateur, l'angle entre deux rayons consécutifs étant égal à $360^\circ/M$. Ensuite on cherche le point d'intersection le plus proche avec les objets de l'espace qui font obstacle au regard. L'ensemble de ces points définit le polygone de visibilité. La résolution de ce calcul dépend du nombre des rayons lancés. La Figure 1.a illustre l'algorithme.

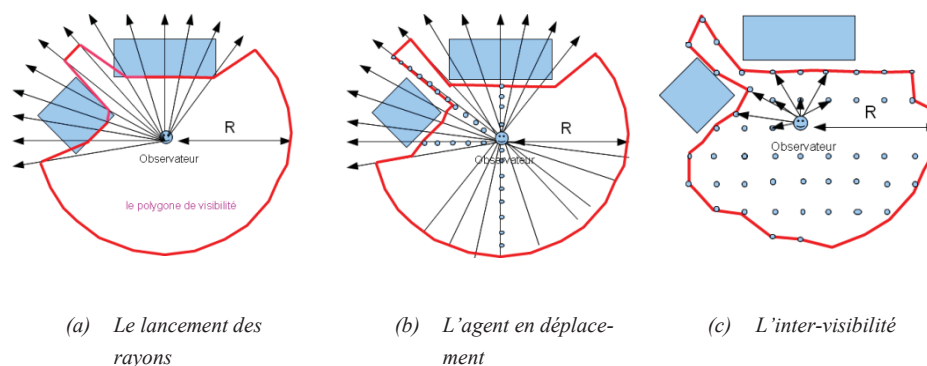


Figure 1. Les méthodes du calcul pour le isovist en dimension 2

Cet algorithme est utilisé par Rana (2006) dans l'application qu'il a développée pour calculer l'isovist en 2 dimensions avec Arcview 3 de la société ESRI. Pour obtenir une bonne précision, il utilise une résolution angulaire de 0.1° et procède donc à $360/0.1 = 3600$ lancements de rayon depuis chaque point. Pour un environnement de N segments, la complexité de cet algorithme est d'ordre $M*N$ intersections mutuelles entre les segments 2D. Le test d'intersection entre ces segments peut être réalisé en utilisant l'algorithme de la ligne de balayage « Sweep line algorithm » (Bentley and Ottmann 1979). La complexité de cet algorithme pour S segments est de l'ordre de $S*\log(S)$.

Une autre méthode a été proposée par Batty and Jiang (1999) Batty (2001), celle de l'agent en marche (walking agent). L'idée n'est pas très différente du lancement des rayons. On imagine un observateur qui marche sur une ligne droite depuis un point donné avec un pas constant jusqu'à ce qu'il arrive à un obstacle. On répète cette opération dans toutes les directions à partir du point de départ. L'angle entre deux directions consécutives est déterminé *a priori* comme dans le cas du lancement des rayons. La méthode est présentée dans la Figure 1.b.

La complexité de cet algorithme dépend du nombre de directions de la marche et de la longueur du pas de l'agent. A chaque pas il faut tester si l'agent est entré en collision avec un objet (polygone) de l'environnement. Cette opération est équivalente à un test d'intersection avec tous les côtés (N segments) des polygones de l'environnement. Dans le cas précis, si l'agent marche sur une distance R dans M directions avec un pas d , on trouve en 2 dimensions $M*R/d*N$ segments dont l'intersection est à tester.

La dernière méthode proposée est le calcul d'inter-visibilité. Cette méthode a été proposée par Turner et al. (2001). L'idée est de générer un maillage régulier de points dans l'espace. Ensuite on teste pour chaque point du maillage l'inter-visibilité avec l'observateur. Cela consiste à vérifier si la ligne entre le point de l'observateur et le point de maillage a une intersection avec les objets de l'espace. La complexité de

cette méthode dépend du nombre des points dans le maillage. Si nous supposons que le nombre des points est k , alors l'ordre de la complexité en dimension 2 pour cette méthode est $k*N$ tests d'intersection mutuelle entre les segments. La Figure 1.c présente la méthode de l'inter-visibilité.

Une liste des programmes pour calculer le champs de vision en dimension 2 est présentée par Do (1997). Ces programmes utilisent des techniques différentes comme le lancement des rayons (Tcl-Light), le « shadow casting » (Do 1994), la suppression des murs occlus dans l'environnement d'AutoCAD par AutoLisp (Do 1994) et la recherche des murs visibles (Do 1995). Christenson (2010) a calculé la perméabilité visuelle dans l'architecture en utilisant l'espace visible. D'autres logiciels sont disponibles en ligne¹.

2.1. Calcul de l'isovist en dimension 3

Quand l'on passe à la troisième dimension, l'idée la plus courante est d'utiliser un MNT (Modèle Numérique de Terrain) en mode raster et de calculer la visibilité par le lancement de rayons. Plusieurs solutions ont été proposées (Franklin and Ray 1994) (Floriani and Magillo 1994) (Van Kreveld 1996) (Lake, Lovett, Bateman, and Day 2000) (Floriani and Magillo 2003) (Brossard, Joly, and Tourneux 2008) (Fishman, Haverkort, and Toma 2009). La Figure 2 explique l'idée du calcul de la visibilité en mode raster.

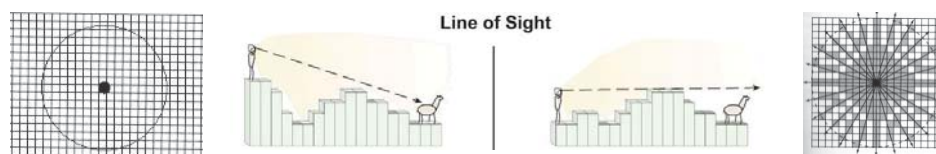


Figure 2. La visibilité en mode raster. Source (Brossard, Joly et Tourneux 2008)

Pour calculer la visibilité, on teste l'inter-visibilité entre le pixel de l'observateur et chaque pixel de l'image raster. Comme nous pouvons voir la précision du calcul dépend de la résolution de l'image raster, et l'augmentation de celle-ci accroît la complexité du calcul. Pour améliorer la qualité du résultat sans augmenter sa complexité dans le même ordre, une approche multi-résolution est possible. On utilise une forte résolution à proximité du point de l'observateur jusqu'à une distance donnée et une résolution plus faible pour le reste de l'image raster. La Figure 3 illustre ce principe.

¹ <http://www.visibility.org/>

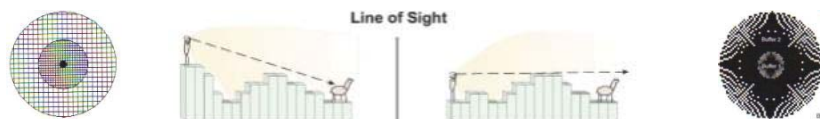


Figure 3. La visibilité en mode raster avec plusieurs résolutions. Source : (Brossard, Joly et Tourneux 2008).

Dans cette approche, l'intégration des éléments du sursol tels que les bâtiments ou la végétation se fait par extrusion en ajoutant à l'altitude des pixels du MNT la hauteur de ces éléments. Cette technique est utilisée par exemple dans les logiciels Arcgis² et SkylineGlobe³.

Le problème de cette méthode est que le modèle de données ne permet pas d'intégrer dans le calcul de visibilité les faces verticales des éléments comme les façades des bâtiments par exemple. Une solution pour le traitement de l'élévation verticale des bâtiments est suggérée par Bartie et al. (2008) (2010). Il s'agit de découper les façades en cellules régulières selon une valeur de hauteur donnée et de procéder au calcul de visibilité pour chaque cellule d'une manière analogue aux pixels du mode raster (Figure 4).

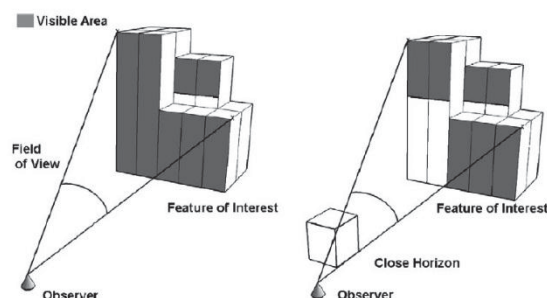


Figure 4. *La visibilité en mode raster avec des façades. Source (Bartie et al. 2010)*

Une autre méthode consiste en l'emploi d'une pixellisation volumique de type « voxel », élément volumique élémentaire d'un maillage tridimensionnel qui permet d'intégrer des éléments comme des bâtiments ou un couvert végétal. On peut aussi associer aux voxels un coefficient de transparence en fonction de la nature des objets

2

<http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/009z000000v8000000>

³ <http://www.skylinesoft.com>

(Pyysalo, Oksanen, and Sarjakoski 2009). Dans la Figure 5, les bâtiments sont intégrés avec des voxels opaques (Morello and Ratti 2009). Ce type de méthode nécessite de disposer de données 3D Lidar, encore assez coûteuses et qui ne sont pas disponibles partout.

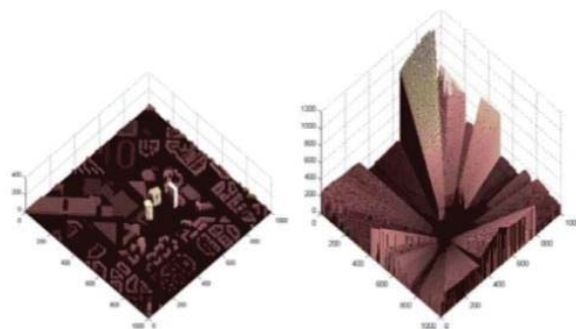


Figure 5. *Modèle voxel avec des bâtiments* (Morello and Ratti 2009)

Une utilisation d'un modèle vectoriel a été proposée par Bilsen (2009). Celui-ci a développé un programme de calcul de visibilité en 3 dimensions qui utilise le lancer de rayons dans un environnement urbain avec un terrain plat (Figure 6).

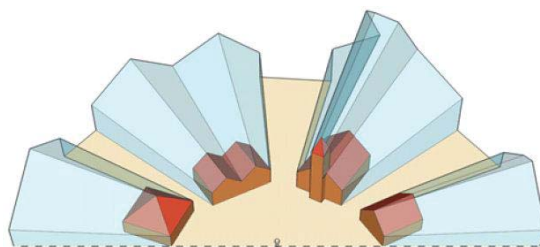


Figure 6. *Le champ de visibilité en 3D avec un terrain plat* (Bilsen 2009)

La précision et la complexité sont les mêmes que dans le cas en 2 dimensions et dépendent du nombre de rayons lancés.

D'autres méthodes de calcul de visibilité ont été proposées plus récemment par Yin and Carswell (2012). Mais il n'existe pas à notre connaissance de solution pour un calcul de visibilité avec des données vectorielles en 3 dimensions sur un terrain avec une topographie quelconque.

3. Un nouvel algorithme de calcul d'isovist en 2 dimensions

Nous proposons un algorithme en dimension 2 fondé sur l'hypothèse que l'espace 2D peut être envisagé comme un ensemble de segments sans intersection, hormis

leurs connexions aux extrémités. Les polygones du modèle de données originel sont alors décomposés en segments 2D. L'objectif est de rechercher parmi ces segments ceux qui bloquent la vue (Figure 7).

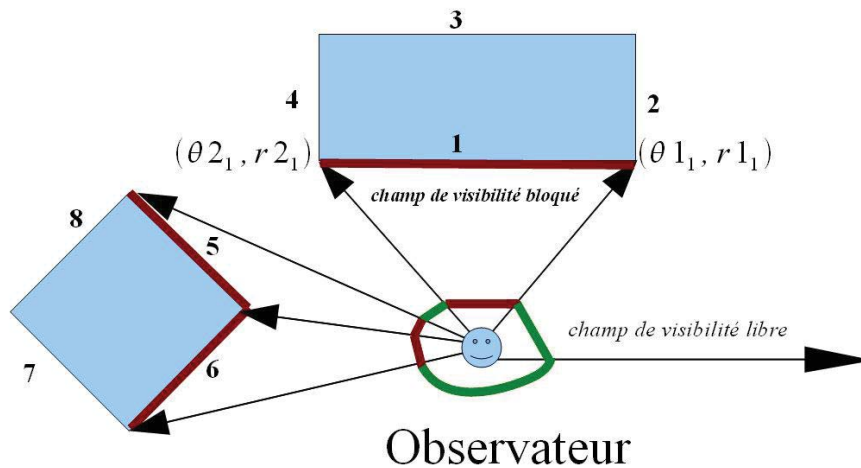


Figure 7. Les angles de vision en rouge sont associés aux segments qui limitent la vision dans ce sens, les angles en vert sont libres.

Si nous considérons cette liste des segments :

$$S = [\text{Segment}_1, \text{Segment}_2, \dots, \text{Segment}_N]$$

Pour chaque segment (Segment_i) nous avons deux extrémités $[(\theta 1_i, r 1_i); (\theta 2_i, r 2_i)]$ en système de coordonnées polaires centré sur le point de l'observateur.

Définition 1 : un angle de vision de Segment_i est l'intervalle entre les angles de ses extrémités, $AVS_i = [\theta 1_i, \theta 2_i]$. Le module de cet angle de vision est : $|AVS_i| = |\theta 1_i - \theta 2_i|$.

Définition 2 : un segment libre est un segment qui a deux extrémités visibles pour l'observateur (segments 1, 5, 6 de la Figure 7).

Définition 3 : un champ de visibilité libre est associé à un angle de vision dans lequel aucun obstacle ne bloque la vue (les angles verts sur la Figure 7).

Définition 4 : un champ de visibilité bloqué est associé à un angle de vision dans lequel il existe un obstacle qui bloque la vue (les angles rouges sur la Figure 7).

Abréviations dans le pseudocode :

FVF : champ de visibilité libre.

BVF : champ de visibilité bloqué.

S : liste des segments.

F : liste libre des segments.

$|AVS_l|$: valeur de l'angle de vision du segment $Segment_l$.

L'algorithme de calcul de visibilité en 2D

Début

Input la liste des segments S , la position de l'observateur.

Initialisation $FVF = [0, 2\pi]$, $BVF = \emptyset$

Tant que ($S \neq \emptyset$)

On cherche les segments libres en S et on les trie par ordre croissant de $|AVS_l|$. cf. Figure 9. $b F = [Seg_c, Seg_v, Seg_t, \dots]$.

Tant que ($F \neq \emptyset$)

Le segment (Seg_c) bloque la visibilité dans l'angle $A_c = \{[\theta_{1c}, \theta_{2c}] \cap FVF\}$

$BVF = BVF \cup A_c$

$S = S - Seg_c$, $F = F - Seg_c$

$FVF = FVF - A_c$.

Fin tant que

On élimine de S tous les segments cachés (c'est-à-dire ceux dont l'intersection entre leur angle de vision et le champ libre de vision est vide)

Fin tant que

Output La résultat = $FVF \cup BVF$.

Fin

Le résultat de cet algorithme est une liste des intervalles angulaires. Chaque intervalle angulaire est associé soit au segment qui bloque la vision dans cette direction soit à l'espace libre. Cela signifie que l'isovist peut être considéré comme une agrégation d'éléments de la forme ci-dessous $[0, 2\pi] * \{\text{Segments}\}$ (Figure 8).

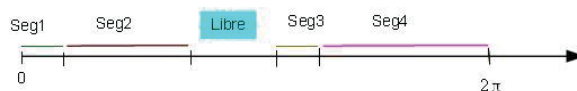


Figure 8. ISOVIST en environnement 2D.

L'algorithme trouve tous les segments qui limitent l'isovist. Le point critique de cet algorithme est de déterminer les segments qui bloquent la vision pour chaque extrémité de la liste. Supposons que nous ayons N segments qui forment les polygones, donc nous avons N extrémités. La complexité de cet algorithme sera

d'ordre N^2 opérations d'intersection entre deux segments 2D (avec N = nombre de segments). Ce calcul peut être amélioré par l'utilisation de l'algorithme « Sweep line algorithm » évoqué plus haut pour trouver l'intersection mutuelle entre les N segments. La complexité de cet algorithme pour N segments est de l'ordre de $N \log(N)$. En reliant les segments (ou les parties de segments) qui bloquent le champ de visibilité, nous obtenons le polygone de visibilité (Rana and Batty 2004). La Figure 9 illustre un exemple de l'algorithme dans un environnement 2D.

Le résultat de la Figure 9 est:

$[0, \theta_{1_1}]$ est un espace libre.

$[\theta_{1_1}, \theta_{1_3}[$ est bloqué par (Segment₁).

$[\theta_{1_3}, \theta_{2_3}]$ est bloqué par (Segment₃).

$] \theta_{1_2}, \theta_{2_2}]$ est bloqué par (Segment₂).

$] \theta_{2_2}, \theta_{2_1}]$ est bloqué par (Segment₁).

$] \theta_{2_1}, \theta_{2_4}]$ est bloqué par (Segment₄).

$] \theta_{2_4}, \theta_{2_5}]$ est bloqué par (Segment₅).

$] \theta_{2_5}, 2\pi]$ est un espace libre.

La Figure 10 présente le résultat de notre méthode dans un environnement virtuel de dimension 2 en utilisant Matlab7. Nous avons choisi cet outil parce qu'il peut traiter les données géographiques (import, export), propose des méthodes de modélisation et des fonctions mathématiques et géomatiques et dispose d'une interface de débogage bien développée. Ces ressources ne sont pas disponibles dans les outils SIG ordinaires.

La méthode employée ici ressemble du point de vue du résultat à celle de Do (1997) mais, à la différence de celle-ci, elle a l'avantage de pouvoir être étendue à un environnement 3D.

La Figure 11 présente les résultats de notre calcul appliqué à une base de données SIG 2D avec extraction du polygone de visibilité et production de quelques mesures morpho-métriques courantes de l'isovist (Rana 2006) que la méthode permet de produire. Dans la Figure 11 nous calculons le secteur visible, la distance, la direction au point le plus proche, la distance et la direction au point le plus lointain, l'index de circularité et l'angle de vue libre ou ouvert.

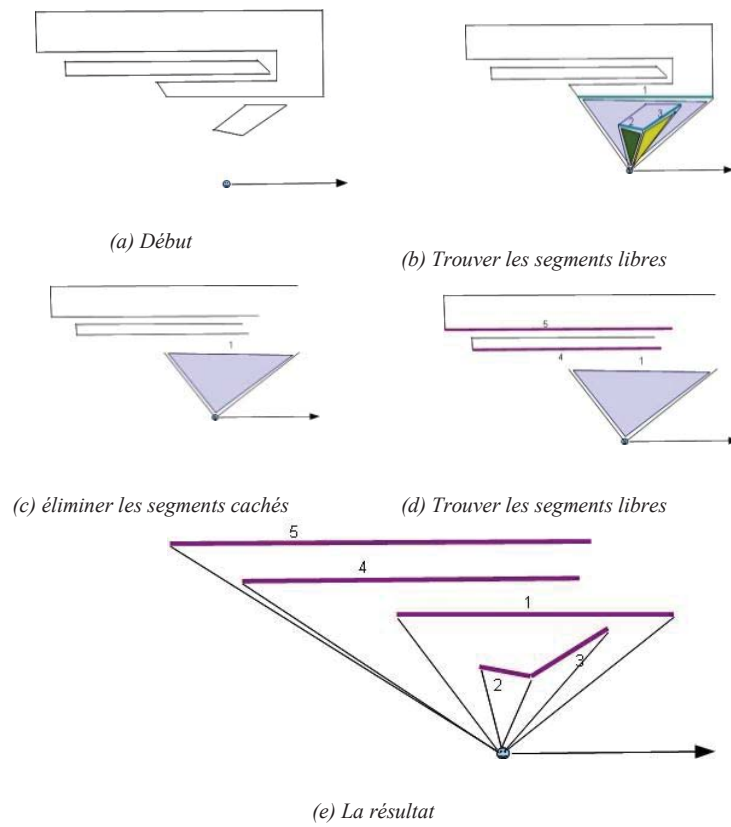


Figure 9. *L'isovist en environnement 2D.*

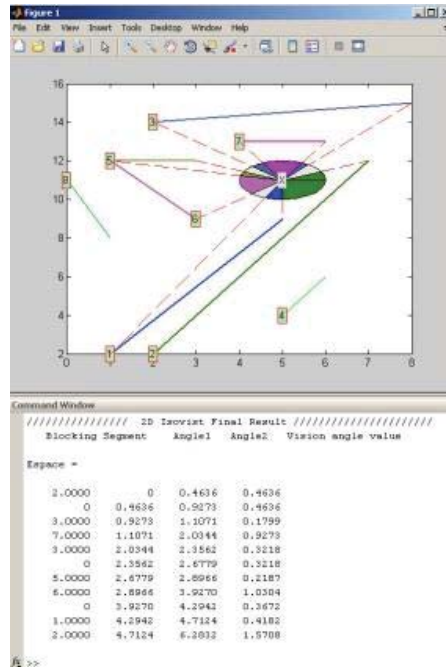


Figure 10. Le résultat de l'isovist en dimension 2: chaque ligne contient l'identifiant de segment, le premier angle et le deuxième angle de l'intervalle de visibilité et la valeur de l'angle de vision.

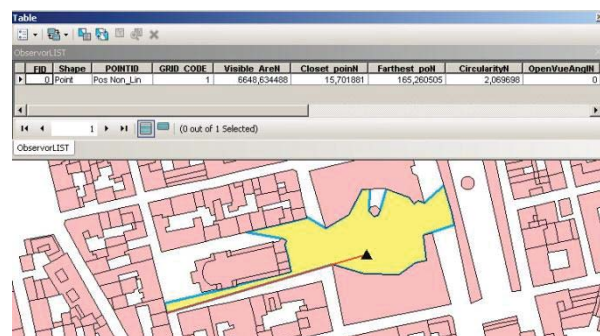


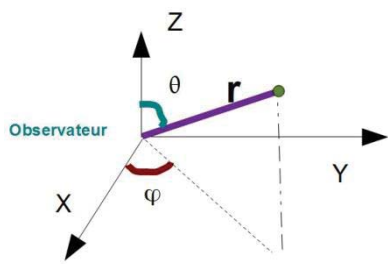
Figure 11. L'isovist en dimension 2 exporté en fichier shape dans ArcGIS.

4. Un nouvel algorithme pour calculer l'isovist en 3 dimensions

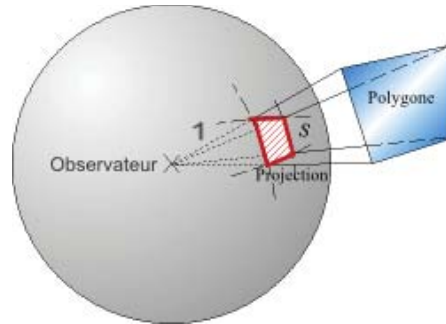
En généralisant notre algorithme précédent en 3 dimensions, nous considérons l'environnement 3D comme une agrégation de polyèdres qui sont décomposés en polygones 3D.

La Figure 12.a présente un système de coordonnées sphériques centré sur le point de l'observateur. Chaque point est défini par (r, φ, θ) avec $r \in \mathbb{R}$, $\varphi \in [0, 2\pi]$, $\theta \in [0, \pi]$. Le polygone Polygone_i est défini par une liste ordonnée des points $\{p_1, p_2, \dots, p_n\}$, qui est définie par des triplets : $\{(r_1, \varphi_1, \theta_1), (r_2, \varphi_2, \theta_2), \dots, (r_n, \varphi_n, \theta_n)\}$. La Figure 12.b présente la projection d'un polygone sur la sphère unitaire (de rayon égale à 1). La projection du polygone Polygone_i est composée des points : $\{(1, \varphi_1, \theta_1), (1, \varphi_2, \theta_2), \dots, (1, \varphi_n, \theta_n)\}$, avec les arcs sphériques qui connectent la liste des points dans l'ordre :

$$\{(1, \varphi_1, \theta_1), (1, \varphi_2, \theta_2)\}; \dots; \{(1, \varphi_{(n-1)}, \theta_{(n-1)}), (1, \varphi_n, \theta_n)\}.$$



(a) Les coordonnées sphériques



(b) Source Wikipédia modifié par Suleiman (les arcs sphériques sont en rouge)

Figure 12. La projection d'un polygone en coordonnées sphériques

Puisque le rayon est égal à 1 pour tous les points, nous pouvons définir la projection comme une série de couples : $\{(\varphi_1, \theta_1), (\varphi_2, \theta_2), \dots, (\varphi_n, \theta_n)\}$. La projection forme l'angle solide de visibilité du Polygone_i et la surface de la projection est la valeur de l'angle solide de visibilité du Polygone_i .

Définition 5: $SAVP_i$ est l'angle solide de visibilité pour le Polygone_i . $|SAVP_i|$ est la valeur de l'angle de visibilité.

Définition 6: un polygone libre est un polygone dont les côtés (les segments de sa frontière fermée) sont visibles par l'observateur.

Définition 7: un isovist libre est associé à un angle solide pour lequel aucun obstacle ne peut gêner la visibilité.

Définition 8: un isovist bloqué est associé à un angle solide pour lequel un obstacle empêche la visibilité.

Abréviations dans le pseudocode :

FS : surface associée au champ de visibilité libre sur la sphère unité.

BS : surface associée au champ de visibilité bloqué sur la sphère unité.

P : liste des polygones.

F : liste des polygones libres.

|SAVP_l| : module de l'angle solide de la visibilité pour la Polygone_l.

L'algorithme de visibilité en 3D

Début

Input le liste des polygones **P**, la position de l'observateur.

Initialisation FS= {Le sphère d'unité entière}, BS = \emptyset .

Tant que (**P** $\neq \emptyset$)

Trouver les polygones libres en **P** et on les trie par ordre croissant de |SAVP_l|, F = [Polygone_c, Polygone_v, Polygone_t,].

Tant que (F $\neq \emptyset$)

Le polygone (Polygone_c) bloque la visibilité dans l'angle solide SA_c = SAVP_c \cap FS.

BS = BS \cup SA_c

P = **P** - Polygone_c , F = F - Polygone_c.

FS=FS - SA_c .

Fin tant que

Élimine tous les polygones cachés de **P**. (c'est-à-dire ceux dont l'intersection entre leur angle de vision et le champ libre de vision est vide)

Fin tant que

Output La résultat = FS \cup BS.

Fin

Cet algorithme produit un ensemble des angles solides associés aux identificateurs des polygones de l'espace.

L'algorithme procède à des tests d'inter-visibilité entre la limite de chaque polyèdre et le point de l'observateur en choisissant un nombre constant *t* de points sur cette limite. La complexité de l'algorithme est donc d'ordre *P* opérations d'intersection entre un polygone et un segment en environnement 3D. La Figure 13.a présente le résultat de notre algorithme de calcul de visibilité sur un cube 3D virtuel en utilisant l'environnement de développement de Matlab.

3.1. L'application dans un environnement 3D

Des tests ont été réalisés avec les bâtiments 3D situés sur une zone de collines à Saint-Etienne (France). Le test utilise les données SIG 3D des bâtiments issues de la base de données de l'IGN et un MNT de résolution d'un mètre pour le terrain issu

d'une campagne de mesure aéroportée. Nous utilisons la triangulation de Delaunay pour modéliser le terrain comme un ensemble de triangles. Les bâtiments sont transformés en un ensemble de rectangles représentant le toit et les façades (Suleiman, Joliveau and Favier 2011). Notre algorithme calcule la visibilité 3D à partir de n'importe quel point de vue situé sur le sol à une hauteur donnée. Le code a été écrit en utilisant les fonctions spatiales et géométriques de Matlab 7. Le rayon de la zone d'étude est de 500 mètres et le modèle contient 17 bâtiments (231 façades). Le temps d'exécution est de 1,5 minutes sur un AMD Athlon 30 Dual Core 4800 +. Les résultats sont illustrés comme un ensemble d'angles solides avec leurs polygones correspondants (façade de bâtiment ou triangle de terrain) qui bloquent la visibilité dans cette direction (Figure 13.b). La Figure 14 présente les résultats de notre programme exportés dans Arcgis et visualisés avec Arcscene comme un fichier des points 3D avec l'attribut de la visibilité.

Pour faciliter les tests et permettre l'usage du programme à des utilisateurs ne disposant pas d'une licence Matlab avec l'extension cartographie, nous avons développé une application autonome en langage C# utilisant la bibliothèque libre shapelib.dll⁴, qui permet d'importer et d'exporter les fichiers shape (Figure 15). Le programme lit en entrée des fichiers type shape, le premier contenant les points MNT avec leur élévation, et le second les bâtiments avec leur identification et un attribut de valeur d'extrusion. Le programme construit le TIN et ajoute par-dessus la couche des bâtiments en fonction de leur valeur d'extrusion. Puis il demande la position et la hauteur de l'observateur et la distance de portée du calcul de visibilité. En option, il est possible de choisir la résolution des points sur les façades des bâtiments et la présence de points virtuels dans le ciel. Le programme dans ce cas calcule la visibilité dans tous les sens autour de l'observateur. On peut aussi calculer la visibilité dans un cône de visibilité déterminé par une direction de vue et un angle de vue vertical et horizontal. Le programme divise l'espace de calcul en quatre secteurs égaux suivant les paramètres de l'utilisateur, puis le calcul s'exécute de manière parallèle pour optimiser le temps de calcul. Le résultat est un fichier de type shape des points 3D portant l'attribut de la visibilité (Figure 14), accompagné d'un fichier texte qui présente les angles solides de la visibilité pour les bâtiments visibles en fonction de leur identification.

Le parallélisme du programme permet de réduire le temps de calcul d'un facteur de 3 à 1 par rapport à notre premier test sur Matlab et ne nécessite aucune ressource logicielle complémentaire pour réaliser le calcul 3D de la visibilité.

La présentation du résultat du calcul de la visibilité en fichier points 3D ne donne pas l'occupation volumique de la visibilité en l'espace. Pour cela on a développé un Arcscene Addin en langage C# en utilisant la bibliothèque ArcObject et le format de données multipatch d'ArcGIS⁵ (Figure 16.a). L'interface demande le fichier des points de la visibilité calculés dans le premier programme avec l'attribut de hauteur et le fichier contenant les paramètres de l'observateur. Le programme forme la triangu-

⁴ <http://shapelib.maptools.org/>

⁵ <http://www.esri.com/library/whitepapers/pdfs/multipatch-geometry-type.pdf>

lation entre les points de la visibilité et les rayons de mire depuis l'observateur. Après le calcul, les résultats peuvent être stockés dans un fichier shape pour être visualisés dans Arcscene. La Figure 16 présente une illustration les formes volumiques de l'isovist 3D à partir d'un point donné avec les points visibles sur les façades.

5. Conclusion

Dans cet article nous avons présenté une nouvelle méthode pour calculer l'isovist avec des données vectorielles dans un environnement 2D et 3D qui prend en compte les formes du relief et les formes géométriques des bâtiments. Nous avons proposé un nouvel algorithme 2D et 3D et l'avons implémenté avec Matlab.

Les premiers tests démontrent la faisabilité et l'intérêt de cette approche en 2D comme en 3D. Une évaluation réelle de cette proposition en termes de temps de calcul nécessiterait un développement spécifique optimisant le matériel (la carte graphique) ou réalisée dans un langage de programmation de bas niveau. Les tests ont été réalisés sur des modèles simples de bâtiment de type LoD1 en City GML, à toit plat, sans décrochement ni chien-assis. La précision des isovists calculés reflète ces limites mais la méthode proposée peut accepter en entrée n'importe quelle forme de bâtiment dès lors qu'il est possible de la décomposer en polygones. La transparence est une limite pour notre algorithme. Tous les objets sont considérés comme solides et opaques. Une amélioration en ce sens est nécessaire pour prendre en compte par exemple les surfaces vitrées non réfléchissantes et, surtout, la végétation.

Le passage à un programme autonome avec la possibilité du parallélisme qui rend le temps de calcul acceptable nous permet d'envisager une application industrielle de la méthode, comme le démontre le premier essai d'intégration dans un environnement logiciel SIG professionnel, en l'occurrence celui d'ESRI.

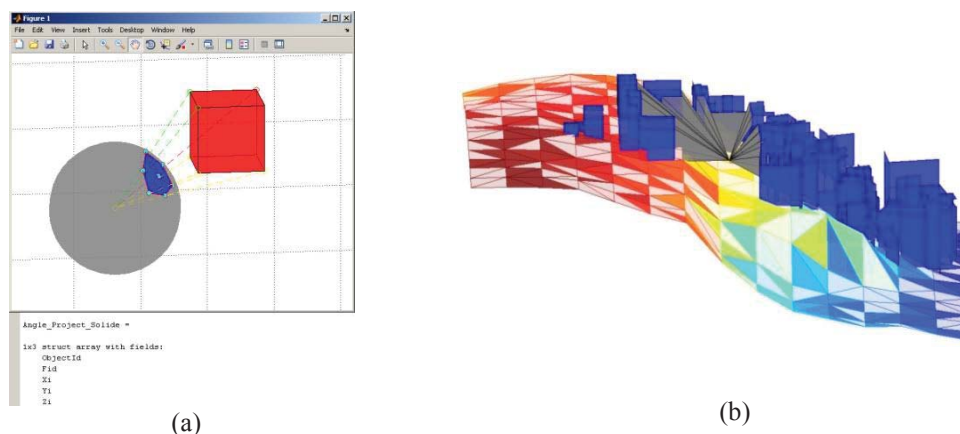


Figure 13. La visibilité 3D avec des données SIG 3D en utilisant Matlab.

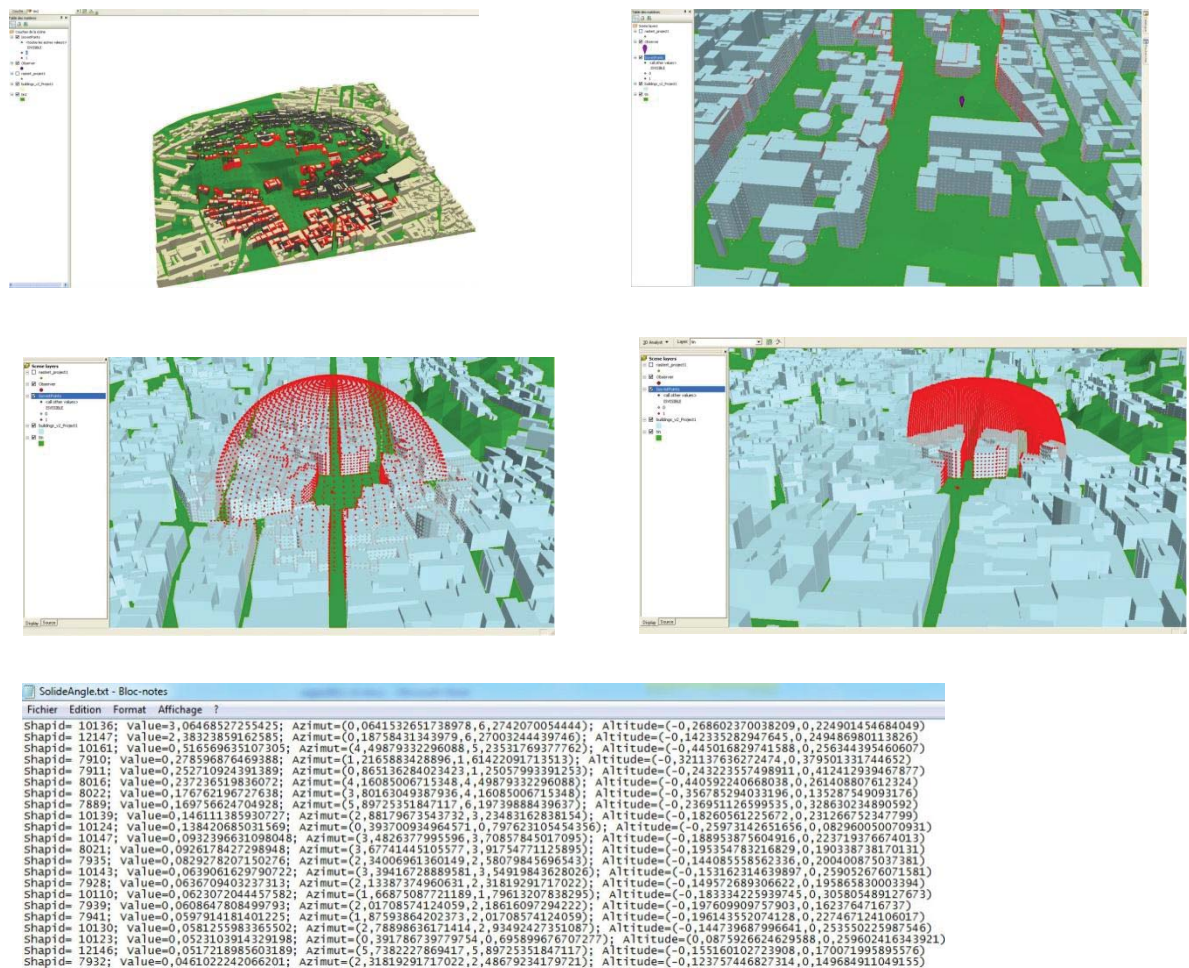


Figure 14. Les résultats du calcul de visibilité en 3D exportés en fichiers shape et visualisés dans Arcgis. Les points correspondant aux angles solides de visibilité sont ajoutés automatiquement selon un treillis régulier sur le sol, dans les airs et sur les façades des bâtiments. Les points rouges sont les points visibles par l'observateur, le fichier texte présente l'angle solide de vue pour les bâtiments visibles.

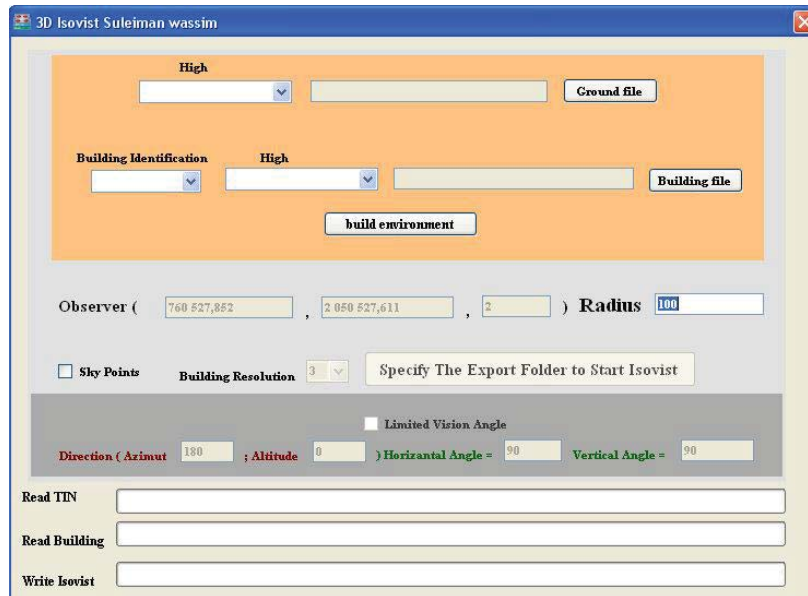
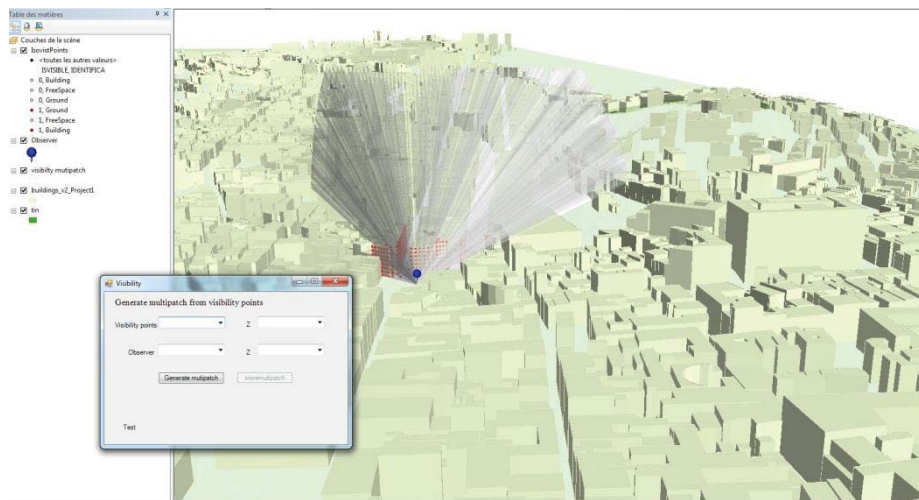
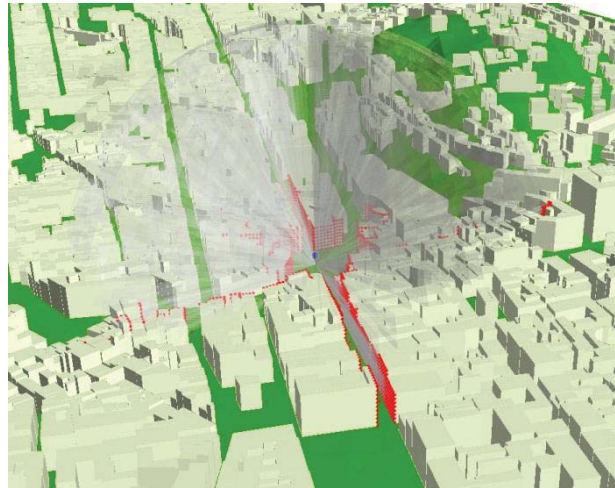


Figure 15. L'interface de la programme C# pour calculer la visibilité en 3D.



(a)



(b)

Figure 16. *L'isovist 3D avec en format mutipatch d'ArcGIS.le fichier mutipatch est visualisé avec une facteur de transparence90%.*

Bibliographie

Bartie P., Mills S., Kingham S., 2008, An Egocentric Urban Viewshed: A Method for Landmark Visibility Mapping for Pedestrian Location Based Services, in *Geospatial Vision*, Springer Berlin Heidelberg, p. 61–85. < <http://www.springerlink.com/content/k421132277816363/abstract/> >

Bartie P., Reitsma F., Kingham S., Mills S., 2010, Advancing visibility modelling algorithms for urban environments, *Computers, Environment and Urban Systems*, November 2010, vol. 34, n°6, p. 518–531. < <http://dx.doi.org/10.1016/j.compenvurbsys.2010.06.002> >

Batty M., Jiang B., 1999, Multi-agent simulation: new approaches to exploring space-time dynamics in GIS, < <http://eprints.ucl.ac.uk/268/> >

Batty M., 2001, Exploring isovist fields: space and shape in architectural and urban morphology, *Environment and Planning B: Planning and Design*, 2001, vol. 28, n°1, p. 123–150.

Benedict M. L., 1979, To Take Hold of Space: Isovists and Isovists Fields. *Environment and Planning*, 1979, p. 6, 47–65.

- Bentley, Ottmann, 1979, Algorithms for Reporting and Counting Geometric Intersections, *IEEE Transactions on Computers*, September 1979, vol. C-28, n°9, p. 643–647. < <http://dx.doi.org/10.1109/TC.1979.1675432> >
- Bilsen V., 2009, How can serious games benefit from 3D visibility analysis?, in *International Simulation and Gaming Association Conference*, Singapore,
- Brossard T., Joly D., Tourneux F., 2008, Modélisation opérationnelle du paysage, in *Paysage et information géographique*, Lavoisier, p. 117–137.
- Cavailhès J., Brossard T., Hilal M., Joly D., Tourneux P.-F., Tritz C., Wavresky P., 2007, Le prix des paysages périurbains, *Économie rurale*, 6 April 2007, n°297-298, p. 71–84.
- Christenson M., 2010, Registering visual permeability in architecture: isovists and occlusion maps in AutoLISP, *Environment and Planning B: Planning and Design*, 2010, vol. 37, n°6, p. 1128 – 1136. < <http://dx.doi.org/10.1068/b36076> >
- Conroy Dalton R., Bafna S., 2003, The syntactical image of the city: a reciprocal definition of spatial elements and spatial syntaxes, < <http://eprints.ucl.ac.uk/1104/> >
- Do E. Y.-L., 1997, Tools for visual and spatial analysis of CAD models, *CAAD Futures 1997 Conference*, 4 August 1997, p. 373–388.
- Do E. Y.-L., 1994, Design and description of form -- using tool command language Tk/Tcl to visualize isovist by lighting and shadow casting analogy., *Computer program ARCH8193A4*, 1994,
- Do E. Y.-L., 1994, Isovist calculation in AutoCAD., *Computer program and independent study*, 1994,
- Do E. Y.-L., 1995, Visual Analysis through Isovist -- building a computation tool, *Working paper and computer program*, 1995,
- Fishman J., Haverkort H., Toma L., 2009, Improved visibility computation on massive grid terrains, ACM Press, p. 121. < <http://dx.doi.org/10.1145/1653771.1653791> >
- Floriani L. D., Magillo P., 2003, Algorithms for visibility computation on terrains: a survey, *Environment and Planning B: Planning and Design*, 2003, vol. 30, n°5, p. 709 – 728. < <http://dx.doi.org/10.1068/b12979> >
- Floriani L. D., Magillo P., 1994, Visibility algorithms on triangulated digital terrain models, *International journal of geographical information systems*, January 1994, vol. 8, n°1, p. 13–41. < <http://dx.doi.org/10.1080/02693799408901985> >
- Franklin W. R., Ray C. K., 1994, Higher isn't Necessarily Better: Visibility Algorithms and Experiments, *ADVANCES IN GIS RESEARCH: SIXTH INTERNATIONAL SYMPOSIUM ON SPATIAL DATA HANDLING*, 1994, vol. 5, p. 751–770.

- Gibson J. J., 1983, *The Senses Considered as Perceptual Systems*, 1st ed. Greenwood Press Reprint,
- Gueymard S., 2006, Facteurs environnementaux de proximité et choix résidentiels, *Développement durable et territoires*, 4 May 2006, < <http://developpementdurable.revues.org/index2716.html> >
- Jim C. Y., Chen W. Y., 2006, Impacts of urban environmental elements on residential housing prices in Guangzhou (China), *Landscape and Urban Planning*, 28 November 2006, vol. 78, n°4, p. 422–434. < <http://dx.doi.org/10.1016/j.landurbplan.2005.12.003> >
- Van Kreveld M., 1996, Variations on Sweep Algorithms: efficient computation of extended viewsheds and class intervals, *IN PROC. 7TH INT. SYMP. ON SPATIAL DATA HANDLING*, 1996, p. 13–15.
- Lake I. R., Lovett A. A., Bateman I. J., Day B., 2000, Using GIS - and large-scale digital data to implement hedonic pricing studies, *International Journal of Geographical Information Science*, 2000, vol. 14, n°6, p. 521. < <http://dx.doi.org/10.1080/136588100415729> >
- Luttik J., 2000, The value of trees, water and open space as reflected by house prices in the Netherlands, *Landscape and Urban Planning*, 1 May 2000, vol. 48, n°3-4, p. 161–167. < [http://dx.doi.org/10.1016/S0169-2046\(00\)00039-6](http://dx.doi.org/10.1016/S0169-2046(00)00039-6) >
- Miller D., 2001, A method for estimating changes in the visibility of land cover, *Landscape and Urban Planning*, 25 May 2001, vol. 54, p. 93–106. < [http://dx.doi.org/10.1016/S0169-2046\(01\)00128-1](http://dx.doi.org/10.1016/S0169-2046(01)00128-1) >
- Morello E., Ratti C., 2009, A digital image of the city: 3D isovists in Lynch's urban analysis, *Environment and Planning B: Planning and Design*, 2009, vol. 36, n°5, p. 837 – 853. < <http://dx.doi.org/10.1068/b34144t> >
- Pyysalo U., Oksanen J., Sarjakoski T., 2009, Viewshed analysis and visualization of landscape voxel models, *24th International Cartographic Conference, Santiago, Chile*, 15 November 2009,
- Rana S., Batty M., 2004, Visualising the structure of architectural open spaces based on shape analysis, < <http://eprints.ucl.ac.uk/187/> >
- Rana S., 2006, Isovist analyst: an arcview extension for planning visual surveillance, < <http://eprints.ucl.ac.uk/2104/> >
- Robert S., 2011, Cartographier la visibilité de la mer pour la gestion d'un littoral: une expérimentation sur la Côte d'Azur, *Espace géographique*, 2011, vol. 40, n°3, p. 215–230.
- Suleiman W., Joliveau T., Favier E., 2011, 3D Urban Visibility Analysis with Vector GIS Data, in *GISRUK 2011*, University of Portsmouth, UK , 27-29 avril 2011, < http://portail.univ-st-etienne.fr/servlet/com.univ.collaboratif.utils.LectureFichier?CODE_FICHIER=1315906066024&ID_FICHE=185671 >

Turner A., Doxa M., O'Sullivan D., Penn A., 2001, From isovists to visibility graphs: a methodology for the analysis of architectural space, *Environ. Plann. B*, 2001, vol. 28, n°1, p. 103–121. < <http://dx.doi.org/10.1068/b2684> >

Do combined atmospheric and topographic correction methods improve land cover classification accuracy in mountain areas?

Steven Vanonckelen^{1*}, Stefaan Lhermitte², Anton Van Rompaey¹

1. Division of Geography, Katholieke Universiteit Leuven
Celestijnenlaan 200E, BE-3001 Heverlee, Belgium

steven.vanonckelen@ees.kuleuven.be
anton.vanrompaey@ees.kuleuven.be

2. Royal Netherlands Meteorological Institute
Wilhelminalaan 10, AE-3730 De Bilt, the Netherlands

lhermitte.stefaan@ceaza.cl

ABSTRACT. Mapping of vegetation in mountain areas based on remote sensing is hampered by atmospheric distortions and differential illumination effects due to topography. A variety of atmospheric and topographic correction techniques has been proposed to minimize atmospheric and topographic effects and should in principle lead to a better classification of land cover. The purpose of this study was to evaluate the effect of different preprocessing methods on land cover classification accuracy. Therefore, three atmospheric and five topographic corrections (including scenarios without atmospheric and/or topographic correction) were applied in all possible combinations in a case study in the Romanian Carpathian mountains. The accuracy of 15 resulting land cover maps was evaluated statistically for a Landsat TM image. First, results showed that all preprocessing methods resulted in a higher overall classification accuracy. The highest accuracy (94.8%) was achieved after combination of an atmospheric correction based on transmittance functions (TF) and a pixel-based Minnaert (PBM) topographic correction. Secondly, class accuracies of coniferous and mixed forest classes were significantly enhanced after correction. The combination of TF with PBM correction produced the best results with an increase in kappa values of about 0.16. Finally, preprocessing methods showed most efficient on weakly illuminated slope segments. After preprocessing, accuracies in the low illumination zone ($\cos \beta \leq 0.65$) were improving more than in the moderate to high illumination zones. After combination of TF correction with a pixel-based C correction or PBM correction, overall classification accuracy were almost doubled (from 48.6% to 79.6% and 80.2%, respectively). Considering all results, we concluded that the topographic component of the corrections contributed more to the improvement of classification accuracies than the atmospheric component.

KEYWORDS: Land cover mapping, Classification accuracy assessment, Carpathians, Atmospheric correction, Topographic correction, Landsat, Mountain area.

Extended abstract

Mapping of vegetation in mountain areas based on remote sensing is hampered by atmospheric distortions and differential illumination effects due to topography. A variety of atmospheric and topographic correction techniques has been proposed to minimize atmospheric and topographic effects and should in principle lead to a better classification of land cover. In literature, though, only a limited number of AC and TC combinations has been tested and described so far. During the past 10 years, several atmospheric and topographic corrections have been evaluated individually. At present, a systematic comparison of the performance of different combined corrections on classification accuracy is lacking. The purpose of this study was to evaluate the effect of different preprocessing methods on land cover classification accuracy based on three statistical analyses and two validation sets (a full set and a subset in the difference zone). Therefore, three atmospheric and five topographic corrections (including scenarios without atmospheric and/or topographic correction) were applied in all possible combinations in a case study in the Romanian Carpathian mountains. The accuracy of 15 resulting land cover maps was evaluated statistically for a Landsat TM image. First, results for the full validation set showed that all preprocessing methods resulted in a higher overall classification accuracy. The highest accuracy (94.8%) was achieved after combination of an atmospheric correction based on transmittance functions (TF) and a pixel-based Minnaert (PBM) topographic correction. The classification accuracy of the un-preprocessed image in the difference area (49.1%) is much lower than the overall classification accuracy (92.2%). All combined corrections result in significantly higher classification accuracies. Secondly, class accuracies of coniferous and mixed forest classes were significantly enhanced after correction. The combination of TF with PBM correction produced the best results with an increase in kappa values of about 0.16. Finally, preprocessing methods showed most efficient on weakly illuminated slope segments. After preprocessing, accuracies in the low illumination zone ($\cos \beta \leq 0.65$) were improving more than in the moderate to high illumination zones. After combination of TF correction with a pixel-based C correction or PBM correction, overall classification accuracy were almost doubled (from 48.6% to 79.6% and 80.2%, respectively). Considering all results, we concluded that the topographic component of the corrections contributed more to the improvement of classification accuracies than the atmospheric component.

1. Introduction

Assessing the rate and spatial pattern of land cover change is challenging given the ruggedness and inaccessibility of mountain areas. Remote sensing techniques are privileged monitoring tools and yet suffer from methodological challenges that have to be resolved by preprocessing techniques. A typical image preprocessing should include sensor calibration, atmospheric and topographic correction (AC and TC), and relative radiometric normalization. Remote sensing-based land cover mapping in mountain areas is especially affected by atmospheric and topographic effects on recorded sensor signals. During the past 10 years, several atmospheric and topographic corrections have been evaluated individually. Some authors (Kobayashi and Sanga-Ngoie, 2008; Richter, 1998) have evaluated the influence of an integrated AC and TC. The integrated methods that have been developed include a specific combination of an atmospheric and a topographic correction. In literature, though, only a limited number of AC and TC combinations has been tested and described so far. At present, a systematic comparison of the performance of different combined corrections on classification accuracy is lacking. The overall research question of this paper is to evaluate the impact of 15 combined AC and TC corrections on the accuracy of land cover classifications in mountain areas. Therefore, a land cover classification analysis is performed on 15 AC and TC combinations. The research is conducted in a study area of a Landsat-5 Thematic Mapper (TM) image in the Romanian Carpathians.

2. Study area and methodology

In order to address the research questions described above, a mountain study area of 915 km² in the Romanian Carpathian mountains was selected. The study area consists of rugged terrain with an elevation varying between 690 and 2,540 m above mean sea level. The Landsat-5 TM image (path 183/row 28) was captured on July 24, 2009 and it was acquired from the archive of the United States Geological Survey (USGS). In this paper, all analyses were performed on the six non-thermal bands. The image was orthorectified with precision terrain correction level L1T by the USGS. Clouds and cloud shadows were ignored since cloud coverage in the study area was below 1%. The solar elevation and solar azimuth angle at image capture were 57.8° and 136.9°, respectively. The co-registered digital elevation model (DEM) used in this study is the space shuttle radar topography mission (SRTM) from CGIARCSI/NASA. For this study, the SRTM version 4.1 was resampled to a

pixel size of 30 x 30 m by means of a bicubic spline interpolation to match the geo-reference of the Landsat dataset. Ground control points (GCP's) for land cover training and validation were gathered through field visits and the analysis of high-resolution satellite imagery. First, eighty-three usable GCP's were recorded through transect walks in the study area during field visits in May 2010 and July 2011. Since the number of field-registered points was insufficient, extra land cover data were derived by a visual interpretation based on high-resolution satellite imagery (WorldView-2, 46 cm resolution, acquisition date October 13, 2010). On the basis of the WorldView-2 image and field expertise, 322 extra reference points were identified. Finally, the reference data gathered during field visits and based on WorldView-2 were randomly subdivided in a training and validation dataset consisting of respectively two-thirds (270 points) and one-third (135 points) of the reference points. This training and validation datasets were used for the classification and validation of all images.

Calibrated radiance values ($L_{s,\lambda}$) were corrected for atmospheric path radiance (L_p). The atmospheric correction was performed by means of 3 different AC methods (including no AC) as described in Table 1. The AC methods were selected with an increasing modeling complexity and necessity of input parameters: i) no AC, ii) DOS and iii) TF correction. Next, the target radiance ($L_{t,\lambda}$) was converted to at-surface reflectance ($\rho_{T,\lambda}$) described by Chander *et al.* (2009). In a final step, the normalized reflectance of a horizontal surface ($\rho_{H,\lambda}$) was calculated using the five topographic corrections described in Table 1. The five selected TC methods were: no TC, band ratioing, cosine correction, pixel-based Minnaert correction (PBM) and pixel-based C-correction (PBC). The three AC and five TC methods – including no AC and no TC – were combined to 15 different combinations of AC and TC methods.

Table 1. Overview of the AC and TC methods with equations and references applied in this study.

Type	Correction	Equation	Reference
AC	No AC	/	/
	DOS	$L_{t,\lambda} = L_{s,\lambda} - L_p$ (1)	Chavez, 1996
	TF	$L_{t,\lambda} = \frac{L_{s,\lambda} - L_p}{0.5(1 + T_{r,\lambda})T_{r,\lambda}T_{w,\lambda}^2}$ (2)	Kobayashi and Sanga-Ngoie, 2008
TC	No TC	/	/
	Band ratio	$\rho_{H,\lambda}^{(i)} = \frac{\rho_{T,\lambda}^{(i)}}{\frac{1}{N} \sum_{j=1}^N \rho_{T,\lambda}^{(j)}}$ (3)	Ono <i>et al.</i> , 2007
	Cosine	$\rho_{H,\lambda} = \rho_{T,\lambda} \frac{\cos \theta_s}{\cos \beta}$ (4)	Teillet <i>et al.</i> , 1982
	PBM	$\rho_{H,\lambda} = \rho_{T,\lambda} \frac{\cos \theta_n}{(\cos \theta_n \cos \beta)^{k_\lambda}}$ (5)	Lu <i>et al.</i> , 2008
	PBC	$\rho_{H,\lambda} = \frac{\cos \theta_s + C_\lambda h_0^{-1}}{\cos \beta + C_\lambda h_0^{-1} h}$ (6)	Kobayashi and Sanga-Ngoie, 2008

$L_{t,\lambda}$ is the corrected radiance value of the image, $L_{s,\lambda}$ is the uncorrected radiance value of the image and L_p represents the minimum radiance of the uncorrected radiance values of the image (calculated as the 1th percentile radiance value over the entire image). $T_{r,\lambda}$ is the Rayleigh scattering transmittance function and $T_{w,\lambda}$ the water-vapor transmittance function. $\rho_{H,\lambda}$ stands for the normalized reflectance of a horizontal surface for a specific spectral band number (N) and $\rho_{T,\lambda}$ for the observed reflectance on an inclined terrain, respectively. β is the incident solar angle, $\cos \beta$ the illumination parameter calculated by $\cos \beta = \cos \theta_s \cos \theta_n + \sin \theta_s \sin \theta_n \cos(\phi_t - \phi_a)$. Where θ_n , ϕ_t and ϕ_a , denote the slope angle of the terrain, the aspect angle of the terrain and the solar azimuth angle, respectively. k_λ is the slope of the regression between $x = \log(\cos \theta_n \cos \beta)$ and $y = \log(\rho_{T,\lambda} \cos \theta_n)$. Parameter C_λ is the quotient of intercept (b_i) and slope (m_λ) of the regression line between x and y , the h -factor represents a topographic parameter deduced from the SRTM ($h = 1 - \theta_n/\pi$) and the h_0 -factor an empirical parameter derived from the regression line between reflectance and $\cos \beta$ ($h_0 = (\pi + 2\theta_s)/2\pi$).

In order to obtain classification accuracies, an appropriate classification algorithm is required. The supervised maximum likelihood classifier based on the Gaussian distribution of the elements in the coherent scattering matrix is used. The six land cover classes are: broadleaved forest, bare soil, coniferous forest, grassland, mixed forest and water surface. The performance of land cover classification maps for each of the preprocessing methods was examined based on three statistical analyses:

(1) Which AC and TC combinations result in the best overall classification accuracy? Overall classification accuracy of land cover maps is examined by assessing the percentage correctly classified pixels from the confusion matrix. Thereby, two validation datasets are used: a set containing all 135 validation pixels and a so called “difference subset”. This subset includes the validation pixels that are classified differently between the classification of one of the combined corrections and the classification of the un-preprocessed image and contains 68 validation pixels.

(2) How big is the impact of AC and TC combinations on land cover class accuracies? Delta-kappa values between the un-preprocessed and preprocessed images are calculated for each land cover type. Positive delta-kappa values indicate a more accurate classification after combined correction.

(3) In which specific illumination conditions are overall classification accuracies after combined corrections increasing most? Therefore, overall classification accuracy is evaluated for three illumination conditions separately. The validation set is divided in three illumination zones based on the illumination parameter $\cos \beta$. The three illumination zones are: low illumination [$\cos \beta \leq 0.65$], moderate illumination [$0.65 < \cos \beta < 0.85$], and high illumination [$\cos \beta \geq 0.85$].

Finally, based on these three accuracy criteria, the individual effect of the two components (AC and TC) within a combined correction is evaluated.

3. Results and discussion

Figure 1 presents the classification accuracy of the un-preprocessed and the different preprocessed classifications using both validation sets. The overall accuracies for the full set are generally high, ranging between 92.2% (no AC and no TC) and 94.8% (TF combined with PBM). The classification accuracy of the un-preprocessed image in the difference area (49.1%) is much lower than the overall classification accuracy (92.2%). All combined corrections result in significantly higher classification accuracies. The highest accuracies in the difference area are achieved after combinations of a TF and PBC correction (83.1%) or a TF and PBM correction (81.4%).

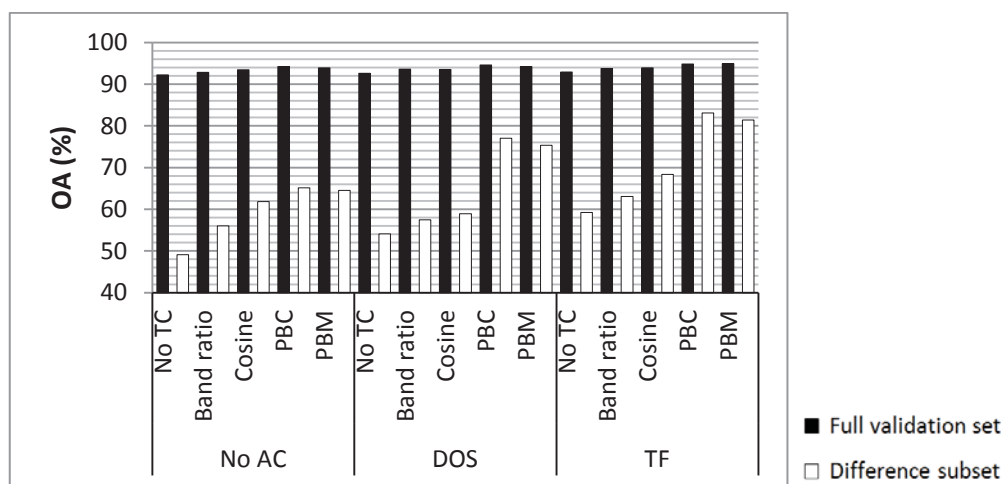


Figure 1. Comparison of the classification results (OA) using the full validation set (black) or subset (white) of the 15 combinations of combined corrections

A visual analysis of land cover maps consists of a qualitative comparison of the different land cover maps. The composites without preprocessing (Figure 2a) show differences in illumination on opposite slopes. Combination of TF and cosine correction (Figure 2b) causes an overcorrection in the visible bands. Combination of TF and PBC correction (Figure 2c) reduces the differential illumination effects. The results of the class accuracies show positive delta-kappa values for coniferous and mixed forest (CF and MX) which implies that for these land cover categories most of the combined corrections improve the classification accuracy. The combination of a TF with a PBM correction produces the best results: an increase in kappa values of about 0.16 for both the CF and MX forest types. The mapping accuracy for the land cover type broadleaved forest is moderately increased while the mapping accuracies for grassland and water surface did not increase significantly after preprocessing. For the land cover type bare soil the results show that preprocessing deteriorates the classification accuracy.

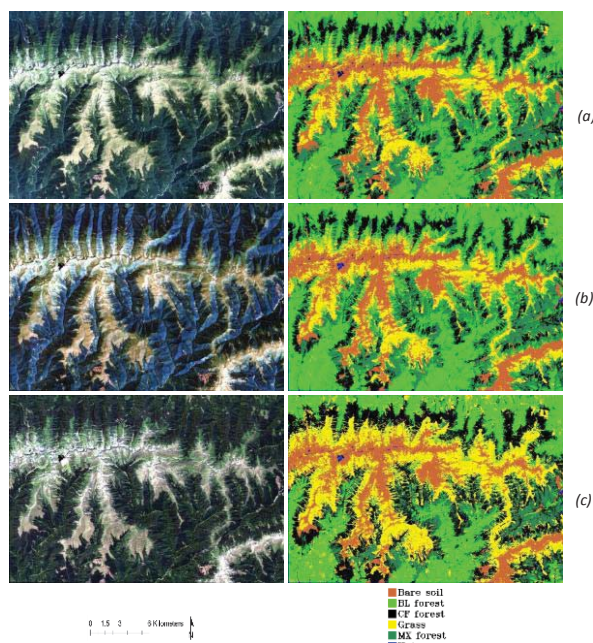


Figure 2. True color composite (RGB: band 3, 2 and 1) and ML classification of the study area with a linear stretching: (a) no AC or TC; (b) TF with cosine correction and (c) TF with PBC correction

The classification accuracy is also evaluated within the difference area for three illumination conditions separately. In the low illumination zone, the accuracy of the un-preprocessed image is low (48.6%). After preprocessing, the accuracy is improving, especially for the combination of TF with PBC or PBM. For those two combinations, accuracy is almost doubling (from 48.6% to 79.6% and 80.2%, respectively). The same trends are visible in the moderate and high illumination zone. Here, accuracies are improving, although the increases are smaller than in the low illumination zone. In the moderate zone, the largest improvement in accuracy is an increase from 56.1% to 84.6% (TF with PBC). Overall, the accuracy is the highest in the high illumination zone (89.0% for the combination of TF with PBM). Finally, the individual effect of the two components (AC and TC) within a combined correction on various accuracy criteria is evaluated. Overall, influence of the TC methods is higher than AC methods. This is concluded from all three statistical analyses. In most cases, implementation of AC methods (DOS and TF) without a TF method has a minor impact on classification accuracy. On the other hand, improvements after implementation of TC correction – especially PBC and PBM – are clear.

References

- Chander G., Markham B.L., Helder D.L. (2009). Summary of current radiometric calibration coefficients for Landsat MSS, TM, ETM and EO-1 ALI sensors. *Remote Sensing of Environment*, vol. 113, n° 5, p. 893-903.
- Chavez P.S. (1996). Image-based atmospheric correction-revisited and improved. *Photogrammetric Engineering and Remote Sensing*, vol. 62, n° 9, p. 1025-1036.
- Kobayashi S., Sanga-Ngoie K. (2008). The integrated radiometric correction of optical remote sensing imageries. *International Journal of Remote Sensing* vol. 29, n° 20, p. 5957-5985.
- Lu D., Ge H. He S. Xu A. Zhou G., Du H. (2008). Pixel-based Minnaert correction method for reducing topographic effects on a Landsat-7 ETM+ image. *Photogrammetric Engineering and Remote Sensing*, vol. 74, n° 11, p. 1343-1350.
- Ono A., Kajiura K., Honda Y. (2007). Development of vegetation index using radiant spectra normalized by their arithmetic mean. *Proceedings of the 42nd Conference of the Remote Sensing Society of Japan*, p. 99-100.
- Richter R. (1998). Correction of satellite imagery over mountainous terrain. *Applied Optics*, vol. 37, n° 18, p. 4004-4015.
- Teillet P.M., Guindon B., Goodenough D.G. (1982). On the slope-aspect correction of multispectral scanner data. *Canadian Journal of Remote Sensing*, vol. 8, n° 2, p. 84-106.

Multifractal portrayal of the distribution of the Swiss population

Vega Orozco C. – Golay J. – Kanevski M.

*Institute of Geomatics and Risk Analysis, Amphipôle building
University of Lausanne
1015 Lausanne
Switzerland*

CarmenDelia.VegaOrozco@unil.ch

Jean.Golay@unil.ch

Mikhail.Kanevski@unil.ch

ABSTRACT: *The present research studies the spatial patterns of the distribution of the Swiss population (DSP). This description is carried out using a wide variety of global spatial structural analysis tools such as topological, statistical and fractal measures, which enable the estimation of the spatial degree of clustering of a point pattern. A particular attention is given to the analysis of the multifractality to characterize the spatial structure of the DSP at different scales. This will be achieved by measuring the generalized q -dimensions and the singularity spectrum. This research is based on high quality data of the Swiss Population Census of the Year 2000 at a hectometric resolution (grid 100 x 100 m) issued by the Swiss Federal Statistical Office (FSO).*

RESUME: *La recherche présentée est consacrée à la description des structures spatiales de la distribution de la population suisse (DPS). Cette description repose sur un ensemble d'indices topologiques, statistiques et fractals qui permettent d'évaluer l'intensité de l'aggrégation spatiale de données ponctuelles. Une attention particulière est accordée à une analyse multifractale permettant de caractériser la structure spatiale de la DSP à différentes échelles. Cette analyse a été réalisée en mesurant les dimensions généralisées de Rényi d'ordre q et le spectre multifractal. D'autre part, cette recherche se base sur les données hectométriques (grille 100 x 100 m) du recensement de la population suisse de l'an 2000 fournies par l'office fédéral suisse de la statistique (OFS).*

KEY WORDS: *multifractal, population, Morisita index, box-counting method, sandbox-counting method, generalized entropy, singularity spectrum.*

MOTS-CLES: *multifractale, population, indice de Morisita, dimensions de Rényi, spectre multifractal.*

1. Introduction

Population in Switzerland has more than doubled its size since the beginning of the 20th century, starting from 3.3 million (1900) to 7.8 million (2009). In the period after the World War II (1950-1970) this country underwent an important population growth with an average annual of about 1.4%. It slowed down (0.6%) during the 1970s to 1990s as a result of immigration restrictions and the economic recession. Since then, population growth has increased again to 0.8% (FSO, 2010). Switzerland can be considered as a densely populated country with an average population density of around 190 people per square kilometre.

Switzerland is geographically divided into three regions: the Alps, the Plateau and the Jura, with a total area of 41,285 km²¹. But not all regions support the same population distribution. For instance, while the Alps occupy the 60% of the total country territory, only 11% of the population live in this mountainous region. The central Plateau, which is the economic epicentre, covers 30% of country's surface area and it holds two-thirds of the total population and most of the main and largest cities. Population growth between 1960 and 1999 was concentrated in smaller centres and agglomeration belts, while, larger urban centres experienced massive population drops in some cases. But since the year 2000, a reversal trend has been observed (FSO, 2010).

The distribution of the Swiss population has been, certainly, shaped and conditioned by the socio-economic history and the complex geomorphology of the country. These factors have rendered a highly clustered, non-homogenous and very variable population distribution at different spatial scales (see Figure 1). Thus, the analysis of the distribution of population is, as well as scientifically, an interesting and challenging task.

The present research studies the spatial patterns of the distribution of the Swiss population (DSP) using a wide variety of global spatial structural analysis tools describing the spatial degree of clustering of a point pattern such as topological, statistical and fractal measures (Kanevski and Maignan, 2004). A particular attention is given to the analysis of multifractality to characterize the spatial structure of the DSP at different scales. This will be achieved by measuring the generalized q -dimensions and the singularity spectrum.

The following measures are applied and the results are compared: Voronoi polygons, Morisita index, and fractal measures (box-counting, sand-box counting). Results of the spatial clustering analysis of the raw data (Swiss population dataset) are compared to the results obtained from the datasets generated under the null hypothesis of spatial randomness. Finally, their multifractal extensions based on either functional measures of clustering or fractal spectra of the corresponding measures are calculated and compared.

¹ <http://www.eda.admin.ch/eda/en/home/repsoceavaus/infoch/chpopu.html> (consulted in 03/2012)

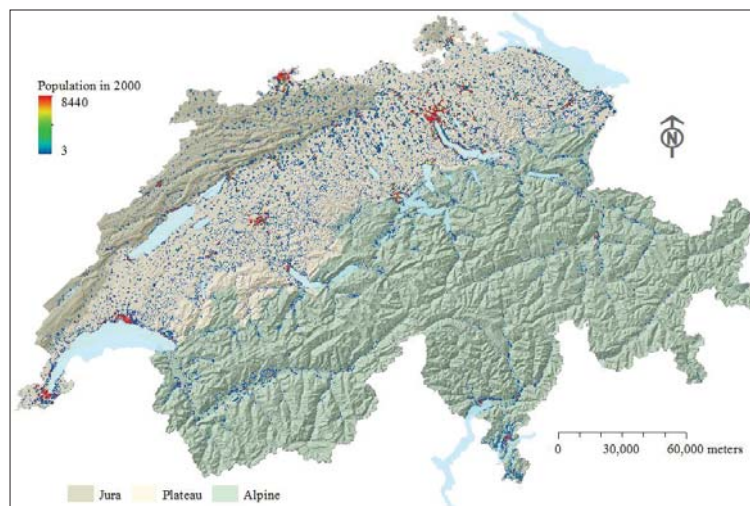


Figure 1. *Swiss population distribution in year 2000 (hectometric resolution)*

Authors intend to demonstrate some preliminary results, rather than to present a comprehensive study of the distribution of the Swiss population. Nevertheless, this study aims at stimulating more elaborated and completed research agenda on Swiss geodemographic analysis.

2. Exploratory analysis and visualisation of data

The population database used in the present study corresponds to the Swiss Census of the Year 2000. This high-resolution data is upheld and managed by the Swiss Federal Statistical Office (FSO, 2010). The census data is presented through points in a hectare-grid resolution (100x100 m) and their XY-coordinates are defined by the lower left angle of the cells; each point portrays the number of people aggregated in a hectare. For privacy protection reasons, the minimum number of inhabitants per hectare is 3 imposed as a restriction for the cartographic representation at a hectare-scale. Figures 1 and 2 display the spatial DSP of the year 2000 and a 3D visualization of the dataset, respectively.

Switzerland is a small country with an extension of 345 km (east-west) and 219 km (north-south). In 2000, Swiss census data counted with 7'351,900 inhabitants distributed in 325,951 points. The point with the maximum population held 8,440 inhabitants and the one with the minimum population carried 3 inhabitants. The global statistics of the Swiss census data are presented in Table 1.

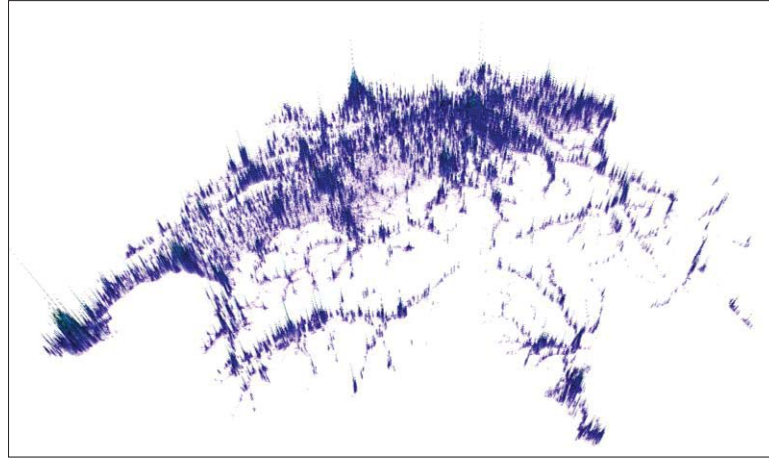


Figure 2. 3D visualization of the *ln*-transformed data.

Total (CH population)	7'351,900	1-st Quartile	4
Mean	22.6	3-rd Quartile	25
Std. deviation	43.5	Skewness	50.90
Median	9	Kurtosis	8615.6

Table 1. Global statistics of the Swiss population data of Year 2000

Moreover, ten artificial random patterns were generated using the same number of points as the raw data (325,951 points) in order to create a confidence interval of spatial randomness. They are used as reference for the comparison with the original pattern (Swiss census dataset).

3. Measures of spatial clustering

To characterize the degree of spatial clustering of the Swiss population pattern different point process measures were implemented. They are divided into topological measures (Voronoi polygons), statistical measures (Morisita Index) and fractal measures (box-counting and sandbox-counting methods) (Tuia and Kanevski, 2008). Each of these measures are described and executed for the raw and reference random pattern datasets.

3.1. Voronoï Polygons

Each object of a point pattern can be inscribed in a polygon in which each polygon's border is always equidistance to the two nearest points. Such regions are called Voronoï Polygons or Dirichlet Cells. The histogram of their area provides a way of evaluating the degree of clustering of a pattern by estimating the homogeneity of its two dimensional covering. The histograms computed from the DSP and from one of the random patterns are displayed in Figure 3. The importance of the highest bin of the raw pattern histogram in comparison with its random pattern counterpart highlights the great rate of small polygons induced by the heavy clustering of Swiss highly urbanised regions.

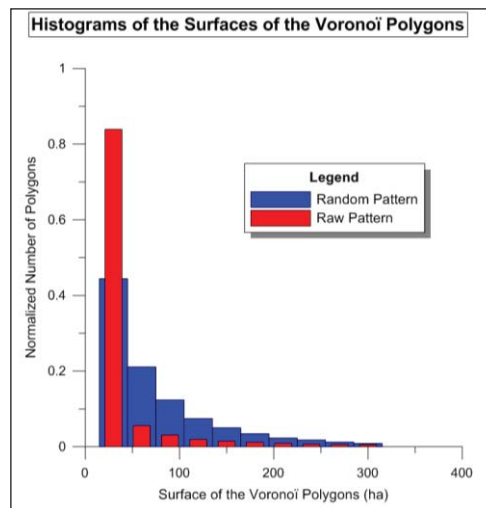


Figure 3. Histogram of the Voronoï polygons for the top decile of a random pattern dataset and the raw one.

3.2. Fractal Measures

Fractals yield a mathematical framework for description of irregular /complex shapes that display similar patterns over a range of scales. This concept, introduced by Mandelbrot (1967), characterises those objects in which properly scaled portions are identical to the original object. Thus, the fractal scaling property refers to the invariance of the probability distributions describing the object's compositions under geometric changes of scale (Rodríguez-Iturbe and Rinaldo, 1967).

3.2.1. Box-Counting Method

In the box-counting method, a regular grid of boxes is superimposed on the region under study and the number N of boxes necessary to cover the whole dataset (e.g. the measurement sites of the DSP where inhabitants have been listed) is counted. Then, the diameter D of the boxes is reduced and the number N is counted again. The algorithm goes on till a minimum diameter D is reached. For a fractal measurement network, the scales and the number of boxes follow a power law:

$$\langle N(D) \rangle \propto D^{-df_{box}} \quad [1]$$

where df_{box} is the fractal dimension measured with the box-counting method (Tuia and Kanevski, 2008). Certainly, in most cases, measurement networks are not fractal and df_{box} must be interpreted as the dimension of the fractal set which is the most similar to the measurement network under study. It is then possible to consider $-df_{box}$ as the slope of the linear regression fitting the data of the plot which relates $\log(N(D))$ to $\log(D)$. If measurement sites are homogeneously/randomly distributed in a studied area the number N of boxes of diameter D necessary to cover the fractal decreases as D^2 . Consequently, the box-counting method enables detecting clustering as a departure from a homogenous situation (Tuia and Kanevski, 2008). Figure 4 shows the implementation of the box-counting method for the DSP. The resulting df_{box} is less than two and indicates a clustering of medium importance.

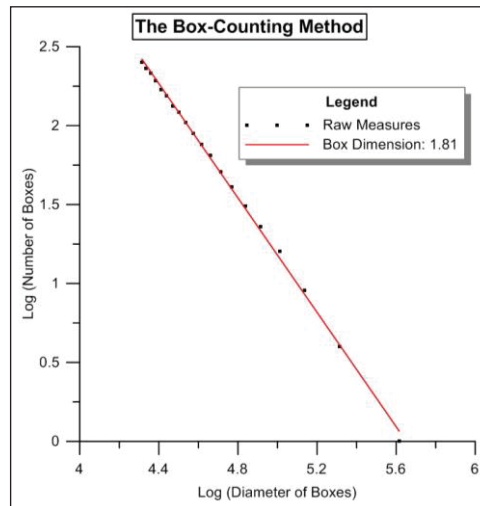


Figure 4. The box-counting method of the Swiss population in the year 2000.

3.2.2. Sandbox-Counting Method

In the sandbox-counting method, neighbours within a circle of radius R centred on a given measurement point are counted. This operation is completed for every point of the measurement network separately and the average number of point per circle is calculated. Then, the radius R is increased and the algorithm goes on till a maximum radius is reached. For a fractal measurement network, the average number of neighbours follows a power law:

$$\langle N(R) \rangle \propto R^{df_{sand}} \quad [2]$$

where df_{sand} is the fractal dimension measured with the sandbox-counting method (Tuia and Kanevski, 2008) and it is sometimes referred to as the correlation dimension (Borgani *et al.*, 1993). Again, fractal measurement networks are hardly encountered and df_{sand} must be interpreted as the dimension of the fractal set which is the most similar to the measurement network under study. Consequently, the value of df_{sand} can be computed as the slope of the linear regression fitting the data of the plot which relates $\log(N(R))$ to $\log(R)$. Again, if points are homogenously/randomly distributed in the studied area, the average number of neighbours N within a circle of radius R increases as R^2 and the intensity of clustering can be assessed through a departure from the homogeneous situation.

Moreover, the method is called functional sandbox-counting if it is applied several times to measurement points above given thresholds (implemented once for each threshold) of a measured function (e.g. number of inhabitants) (Lovejoy *et al.*, 1987). Then, the dimension depends on the threshold, which helps to understand the spatial structure of the phenomenon under study. Functional measures of clustering (FMC) using fractal indices are closely related to the concept of multifractality (Seuront, 2010).

Figure 5 (left) displays the results of the implementation of the sandbox-counting method in the case of the DSP. Again, the resulting dimension indicates a clustering of medium importance. Here, it is interesting to mention that the difference between df_{box} and df_{sand} is also related to multifractality (Borgani *et al.*, 1993) as it will be explained in section 4. Figure 5 (right) shows the results of the functional sandbox-method performed on the raw dataset as well as on the random ones. The chosen thresholds were the deciles and the 95th centile of the DSP. The differences in the evolution of the lines of the raw and random patterns in relation to the thresholds enable to detect the progressive appearance of hotspots (i.e. cities) and allow quantifying their intensity.

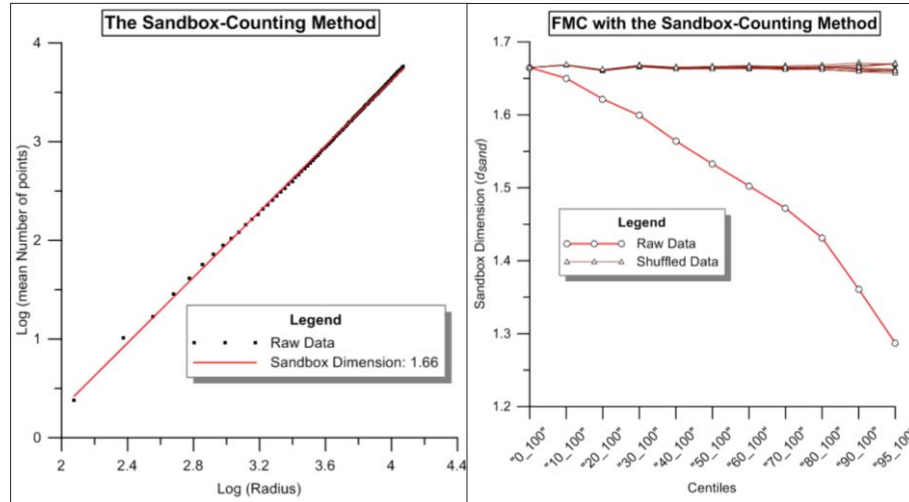


Figure 5. Left: the sandbox-counting method performed on the DSP. Right: the functional sandbox-counting method performed on the DSP

3.3. Morisita Index

Morisita index is computed from the following formula:

$$I_D = Q \frac{\sum_{i=1}^Q n_i(n_i - 1)}{N(N - 1)} \quad [3]$$

where Q is the number of identical quadrats necessary to cover the entire studied area, n_i is the number of points in the quadrat i , N is the total number of points and I_D is Morisita index for a chosen quadrat's dimension D (related to the number of quadrats Q). In other words, this index measures how many times more likely it is to randomly select two points belonging to the same quadrat than it would be if the points were homogeneously distributed (Hurlbert, 1990). I_D is first calculated for a relatively small quadrat's dimension which is then increased till it reaches a chosen value. After that, it is possible to draw a plot which relates every I_D to its matching D . If the points of the pattern are homogeneously distributed on the studied area, every computed I_D varies around the value of 1. If the points are clustered, the empty quadrats at small scales will increase the value of the index (Kanevski and Maignan,

2004). Figure 6 displays Morisita index computed from the DSP data. The results clearly highlight the clustering due to the highly urbanized regions of Switzerland.

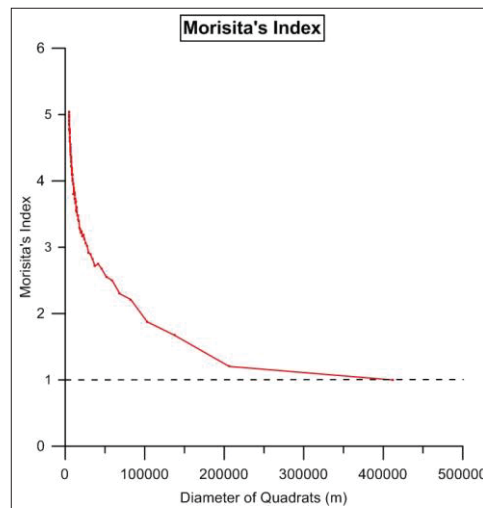


Figure 6. *Morisita's index computed from the DSP data*

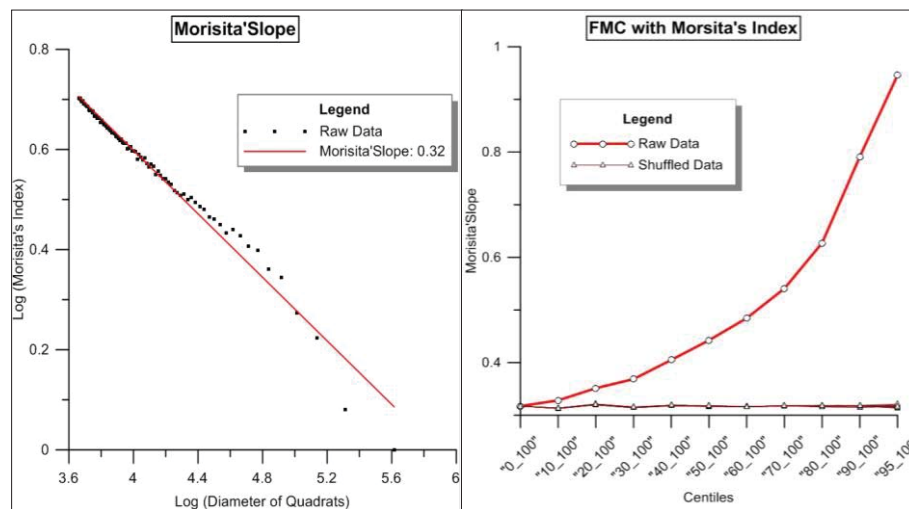


Figure 7. *Left: Morisita slope for the DSP data. Right: Morisita functional index*

Now, it is possible to reduce the information provided by Morisita index to only one number: the slope of the straight line fitting the data of the plot relating $\log(I_D)$ to $\log(D)$ (see Figure 7, left). Finally, Morisita functional index can be coined (see Figure 7, right) and performed on the raw and random patterns using the same thresholds as for functional sandbox-counting. Again, the presence of hotspots of population concentration is demonstrated.

3.4. Multifractal characterization

This study proposes a multifractal formalism to characterize the multiscaling structure of the spatial distribution pattern of the Swiss population in 2000 year. Multifractal measures concern the study of distributions of physical quantities on a geometric support which can be a line segment, a plane, a surface or a volume, and in particular cases the support may be a fractal (Rodríguez-Iturbe and Rinaldo, 1997). While the box-counting measures the fractal dimension of the support because it only considers boxes as empty or not, ignoring the variations in the population inside the boxes; the multifractal analysis allows describing how the pattern distribution fills the geometric support at different moments by presenting interlinked fractal subsets that have different scaling exponents.

To describe the spatial structure a set of generalized q -dimensions, D_q , are measured through a generalization of the box-counting methods. The generalized dimensions are defined in terms of the Rényi information of the q th order moment of the probability distribution. The alpha spectrum was also measured as a complementary analysis. The generalized q -dimensions and the alpha spectrum were calculated for the Swiss Population in 2000 (raw data) and the ten simulated random patterns. Only positive q -dimensions were calculated.

3.4.1. Generalized dimension

The multifractal analysis characterizes the DSP with a set of generalized q -dimensions. To estimate these generalized dimensions, D_q , the fractal object (e.i. DSP) is covered with a regular grid (non-overlapping boxes) of size δ . Let $N(\delta)$ be the number of boxes of the grid needed to cover the fractal object and p_i be the probability of the population in the i th box ($p_i = n_i/N$, where n_i is the population contained in box i and N is the total population in the country). The generalized entropy of the q th order moment is (Appleby, 1996):

$$I_q(\delta) = \frac{1}{(1-q)} \log \sum_{i=1}^{N(\delta)} p_i(\delta)^q \quad [4]$$

The generalized dimensions can be defined with the generalized entropy as (Appleby, 96; Seuront, 2010):

$$D_q = \lim_{\delta \rightarrow 0} \left(\frac{I_q(\delta)}{\log(1/\delta)} \right) \quad [5]$$

The generalized dimension function D_q depends on the order q chosen and it is estimated as the slope of the plot of $\log(I_q(\delta))$ vs $\log(\delta)$. The D_q diagram is obtained by calculating the generalized dimension varying q . When $q=0$, all nonempty boxes are equally weighted ($p_i^0=1$) and D_q is equivalent to the box-counting dimension. For $q>0$, the mass within the boxes (fraction of population) gradually gains more importance in the box contribution to the entropy value; therefore, the larger the population values are in a box, the higher the weight of the box. Thus, D_q indicates the scaling of over-dense regions and strong singularities.

Figure 8 exhibits the manner in which the logarithm of the generalized entropy scales with the logarithm of the boxes' size for the distribution of the Swiss population in 2000. As q increases, the linear behaviour becomes weaker. This is explained by the presence of fewer points with higher density population than there are points with small number of people, and therefore, as q increases, the points with higher number of people become more dominant in the contribution to the entropy even if they may be relatively few in number.

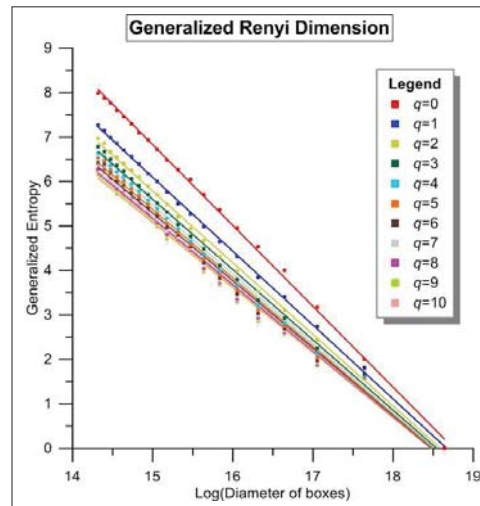


Figure 8. Generalized entropy for each q th of the DSP.

The generalized dimension was estimated for each q moment as the slopes of the linear regressions fitting the data of the plot in Figure 8. Figure 9 shows the variation of the scaling exponents (D_q) at each q th moment.

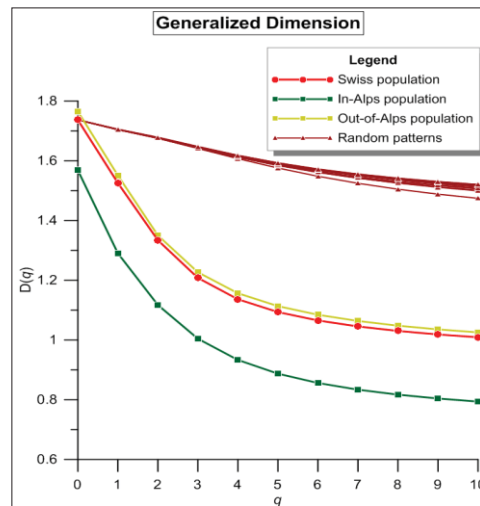


Figure 9. The generalized dimension of the DSP.

The non-linearity of the signature for the DSP (red dot-line) clearly displays the highly multifractal behaviour (scale-invariance) of the Swiss population in 2000. For the DSP, D_q declines faster than for the random patterns (brown triangle-lines). The lower the generalized dimension D_q , the more highly clustered the population. The wide variation of the D_q for the DSP suggests an unevenly and clustered distribution of the population.

The value of D_0 expresses to what extent the distribution fills up the space, that is the dimension of the geometric support of the DSP. Thus, a $D_0 = 2$ means that the population distribution is homogeneous across the whole land and no clusters are found at any scale (Adjali and Appleby, 2001). Looking at Figure 9, we can see that the DSP in the in-Alps region (dark-green dot-line, with $D_0=1.56$) presents a less homogeneously occupation of the territory that the DSP in the out-of-Alps region (yellow dot-line, comprising the Plateau, Jura and Prealps natural regions) which has a $D_0=1.76$. This last distribution has a great influence in the DSP of the entire country ($D_0=1.73$).

The D_q for the DSP in the in-Alps region decreases slightly faster than the out-of-Alps region and the entire country population, meaning that population in the Alpine region is more clustered than in the other country regions. This suggests the

influence of the Alpine rough topography, forcing population in this region to live concentrated around the valleys.

3.4.2. Multifractal spectrum

As a supplementary portrayal to the generalized dimensions, a multifractal spectrum analysis was carried out allowing characterizing the scaling properties of the multifractal ($f(\alpha)$). This is also called the singularity spectrum of the distribution (Seuront, 2010).

The multifractal structure can be considered as the superposition of many homogeneous monofractals. Let $p_i \propto \delta^{\alpha_i}$ be the scaling of the probability around the i th box (Borgani *et al.*, 1993), with α_i being the singularity strength specific to the box size (δ). For every box size δ , the number of boxes with the same α , $N_\alpha(\delta)$, is related to the box size as (Seuront, 2010):

$$N_\alpha(\delta) \propto \delta^{-f(\alpha)} \quad [6]$$

Where $f(\alpha)$ is the singularity spectrum, that is, the scaling exponent of the cells with equals α . It can also be related to the generalized dimension, D_q , as (Adjali and Appleby, 2001):

$$f(\alpha(q)) = q\alpha(q) - (q-1)D(q) \quad [7]$$

The multifractal spectrum is obtained by plotting the singularity spectrum, $f(\alpha(q))$, vs. the singularity exponent, $\alpha(q)$. The multifractal spectrum quantitatively characterizes variability of the probability mass function. In our analysis, we have only used positive moments ($q > 0$), therefore, the multifractal spectrum plot exhibits asymmetry to the left indicating domination of large values.

Figure 10 shows the multifractal spectrum calculated for the DSP in 2000 (the red dot-line), the in-Alps population distribution (dark-green dot-line), the out-of-Alps population distribution (yellow dot-line) and the random patterns (brown triangle-lines). The multifractal spectrum of the DSP is significantly different from the random patterns. The DSP presents an asymmetry more skewed to the left and with lower values than the random patterns; this reflects domination of large population values and a non-homogeneous distribution (highly clustered).

Results from the multifractal spectrum are in agreement with the generalized dimensions presented in section 3.4.1. DSP (red dot-line) presents similar variability patterns than the population in the out-of-Alps region (yellow dot-line). The population in the in-Alps region is the most skewed to the left indicating highly

clustered patterns. This can also be depicted by the fact that the $f(\alpha)$ curve of the in-Alps population is lower and broader than the out-of-Alps region.

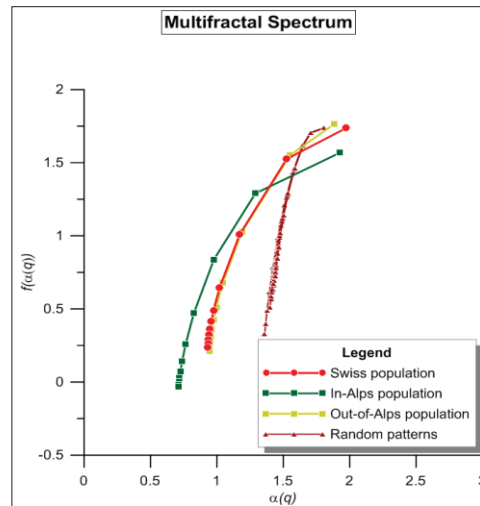


Figure 10. Multifractal spectrum of the DSP.

5. Conclusions

This paper portrays a multifractal analysis of the distribution of the Swiss population using a high-resolution census data from 2000. This analysis allowed the characterization of the spatial Swiss population distribution. The generalized dimensions for the DSP showed a different scaling behaviour of the highly populated areas and less dense regions. The same information was provided by the multifractal spectrum analysis, which turned out to be a useful measure for comparing the degree of clustering of different spatial patterns.

Different researches concerning the scaling structure of the spatial pattern of the population distribution have only considered either a monofractal analysis of the occupied land (a single dimension, i.e. box-counting method) or a multifractal characterization of the population distribution at the city level (i.e. Appleby, 1996). Adjali and Appleby (2001) analysed the multifractal structure of the human population distribution in ten countries (Great Britain, USA, France, Australia, South Africa, Algeria, Colombia, India, Russia and Zaïre), where the demographic information was presented as a grid of 5 x 5 minutes of arc of latitude by longitude cells (~10 x 10 km) with their population counts. Conversely to these papers, our

work goes beyond and deeper analysing the population distribution for Switzerland at the intra-city level using a high-resolution data presented with population counts in a grid of 100 x 100 metres.

Future research will deal with multivariate analyses by taking into account both demographic and socio-economic data.

6. Acknowledgements

This work was partly supported by the SNFS Project No. 200021-140658, “Analysis and Modelling of Space-Time Patterns in Complex Regions”.

7. References

- Adjali I., Appley S., “The multifractal structure of the human population distribution”, in *Modelling scale in geographical information science*, edited by Tate N.J. and Atkinson P.M., West Sussex, John Wiley & Sons Ltd, 2001, p. 69-85.
- Appleby S., “Multifractal characterization of the distribution pattern of the human population”, *Geographical Analysis*, vol. 28 n° 2, 1996, p. 147-160.
- Borgani S., Murante G. Provenzale A. and Valdarnini R., “Multifractal analysis of the Galaxy distribution: reliability of results from finite data sets”, *Physical Review A*, vol. 6 n° 47, 1993, p. 3879-3888.
- Borgani S., Plionis M., Valdarnini R., “Multifractal analysis of cluster distribution in two dimensions”, *The Astrophysical Journal*, vol. 404 n°1, 1993, p. 21-37.
- FSO, “Population”, *Panorama*, Federal Statistical Office, Switzerland, 2010.
- Hurlbert S.H., “Spatial Distribution of the Montane Unicorn”, *Oikos*, vol. 58, 1990, p. 257-271.
- Kanevski M. and Maignan M., *Analysis and Modelling of Spatial Environmental Data*. Lausanne, EPFL Press, 2004.
- Lovejoy S., Schertzer D., Tsonis A., “Functional box-counting and multiple elliptical dimensions in rain”. *Science*, vol. 235, 1987, p. 1036-1038.
- Mandelbrot B., “How long is the coast of Britain? Statistical self-similarity and fractional dimension”, *Science*, vol. 156, 1967, p. 636-638.
- Rodríguez-Iturbe I., Rinaldo A., *Fractal river basins: chance and self-organization*, UK, Cambridge University Press, 1997.
- Seuront L., *Fractals and multifractals in ecology and aquatic science*, Boca Raton, CRC Press, 2010.

Tuia D., Kanevski M., “Environmental monitoring network characterization and clustering”, in *Advance mapping of environmental data. Geostatistics, machine learning and Bayesian maximum entropy*, edited by Kanevski M., London, iSTE and Wiley, 2008, p. 19-46.

8. Bibliography

Carmen Delia Vega Orozco is a PhD student at the IGAR institute, University of Lausanne (Switzerland). She holds a MSc in Environmental Geosciences at the University of Lausanne (Switzerland) in 2010 with a mention in analysis, monitoring and representation of natural hazards; a Master degree in Geography at West Chester University of Pennsylvania (United States) in 2007 with a mention in geographic information systems (GIS) and environmental planning; and a BS degree in Geomatic Engineering from the University of Valle of Cali (Colombia) in 2002. Her research experience focuses in the analysis of environmental phenomena throughout the application of statistical models, GIS and remote sensing.

Jean Golay is a master student in Environmental Geosciences at the University of Lausanne (Switzerland) and has been working as an assistant at the IGAR institute for a couple of years. His research experience involves the exploratory analysis and modelling of geospatial data using point pattern analysis tools, geostatistics and machine learning algorithms.

Prof. Mikhail Kanevski is professor at the Institute of Geomatics and Risk Analysis (IGAR) at the University of Lausanne (Switzerland) within the framework of the Faculty of Geosciences and Environment (FGSE). Professor Kanevski is author and co-author of 3 books and more than 150 publications. He has valuable experience and knowledge in relevant fields including geostatistics, machine learning algorithms (Artificial Neural Networks, Support Vector Machines), geomatics (Geographical Information Systems and remote sensing) and geosimulations providing algorithmic solutions to environmental problems, with application in environmental, socio-economic and financial data mining.

Formalisation des imperfections des données du système d'information agro-environnemental Observox

Gestion des imperfections

Karima Zayrit - Eric Desjardin

*CReSTIC (URCA)
IUT de Reims Châlons Charleville
Rue des Crayères
51687 Reims cedex 2
{karima.zayrit, eric.desjardin}@univ-reims.fr*

RÉSUMÉ. Ce travail s'inscrit dans le projet Observox, un observatoire des pratiques agricoles et leurs pressions sur la ressource en eau au niveau du bassin versant de la Vesle. Il a vu le jour dans le but de comprendre l'impact des pratiques responsables de la contamination de la ressource en eau par les traitements phytosanitaires. La mise en place de l'observatoire, co-construit par les acteurs du territoire, nécessite l'exploitation d'informations hétérogènes, provenant de sources multiples et sujettes à de l'imperfection. Afin de fournir aux utilisateurs des indicateurs plus informatifs, nous intégrons la notion de qualité de l'information via un degré de confiance. Dans cette démarche, deux cadres formels sont respectivement utilisés : la logique floue et les fonctions de croyance pour la propagation de l'imprécision et pour la fusion de différentes sources.

ABSTRACT. This work is a part of OBSERVOX, a community environmental information system, for the monitoring of agricultural practices and their pressure on water resources in the Vesle basin. The purpose of this project is to understand the impact of the practices responsible for the water contamination by pesticides. The setting up of OBSERVOX, co-constructed by the territory's actors, requires the use of heterogeneous informations, which are imperfect and resulting from multiple sources. To provide to the system's users, the most honest indicators, we integrate the notion of information quality by a degree of confidence. Thus two formal frameworks are used respectively: the fuzzy logic and belief functions for propagating imprecision and for merging classifications.

MOTS-CLÉS : Fusion de données, système d'information environnemental, données imparfaites, fonctions de croyances, logique floue, indicateur agroenvironnemental, pratiques agricoles.

KEYWORDS: Data merging, environmental information system, imperfect data, belief functions, fuzzy logic, agro-environmental indicator, agricultural practices.

1. Introduction

La ville de Reims est alimentée par la Vesle, une rivière de faible débit, et présente une activité agro-viticole intense. Cette situation a pour conséquence une pression sur l'eau et a donc un impact sur sa qualité. Afin de gérer au mieux ce risque, il était indispensable de mettre en place un observatoire des pratiques agricoles et leurs impacts sur la ressource en eau. OBSERVOX se base sur une démarche de co-construction de son système d'information (SI) impliquant les différents acteurs du territoire (agriculteurs, viticulteurs, techniciens, gestionnaires de l'eau, citoyens, responsables politiques, chercheurs) en fonction des attentes et des besoins de chacun. Son objectif est d'informer sur l'évolution des pratiques agricoles sur le territoire étudié, en proposant une historisation pérenne de l'information et une reconstruction spatialisée et qualifiée des informations relatives à l'utilisation des pesticides et à leurs effets sur le milieu.

2. Contexte

Dans le cadre de l'observatoire, les sources d'information exploitées sont multiples et variées. Nos données sont principalement issues d'enquêtes terrain, de « dires d'experts » ou fournies par des organismes institutionnels (IGN, Corine Land Cover). Ces informations formelles sont complétées à partir d'analyses par télédétection. Le problème est d'autant plus complexe que ces données peuvent être de natures différentes (quantitatives, qualitative, spatiale, temporelle), collectées à différents niveaux d'échelle (spatiale, temporelle), et présentent des imperfections. Une donnée est dite imparfaite si elle est sujette à de l'imprécision, l'incertitude, l'ambiguïté ou l'absence d'information). De ce fait, et afin de pouvoir offrir aux utilisateurs un système d'information plus informatif, il est nécessaire de prendre en compte la qualité de l'information par la formalisation de ces imperfections.

3. Modélisation des imperfections

La prise en compte de la qualité doit couvrir les différents processus liés à l'acquisition, la gestion, la diffusion et l'utilisation de l'information [DEV05]. Dans cette démarche, nous avons procédé à une première phase de catalogage des sources d'information via l'outil MDweb [DES07]. Ce premier travail avait pour double but d'informer les différents acteurs des données disponibles et leurs caractéristiques, et de travailler sur la qualité externe de la donnée dans le cadre du calcul d'indicateurs et d'une restitution de l'information aux acteurs. La deuxième phase de ce travail consiste à modéliser les informations imparfaites en intégrant la notion de qualité par le biais d'un degré de confiance. A cet effet, deux approches basées sur les théories mathématiques de l'incertain sont proposées.

4. Information qualifiée par la logique floue

La première approche que nous avons proposée vise à modéliser et propager l'imprécision du spatial vers le quantitatif [ZAY11] en nous basant sur la logique floue [DUB08]. Nous avons introduit une représentation des données agronomiques imprécises sous forme d'entité agronomique floue définie par un label LA_i , une forme spatiale FSA_i et une quantité floue de produit FQ_jA_i . Cette représentation combine de l'information sémantique, spatiale et quantitative et permet le chevauchement spatial entre les entités agronomiques floues. En pratique cette méthode permet d'estimer la quantité floue d'un produit chimique diffusé sur des parcelles à frontières mal définies par une agrégation avec l'opérateur T de type T-norme suivant la formule : $\mu_{Q_jA_i}(x, q) = T(\mu_{SA_i}(x), \mu_{Q_jA_i}(q))$. Ainsi pour une position x du territoire, on associe à chaque valeur q de la quantité un degré de confiance $\mu_{Q_jA_i}(x, q)$ qui sera restitué aux utilisateurs.

5. Information qualifiée par les fonctions de croyance

Dans une deuxième démarche, en nous appuyant cette fois sur la théorie des fonctions de croyance [SME00], l'objectif est de fournir aux utilisateurs du SI une information sur l'occupation du sol, qualifiée par un degré de confiance. Concrètement, cette méthode permet de fournir à l'utilisateur une donnée qualifiée par un degré de confiance quant à l'affectation d'un type de culture à un élément de l'espace. Elle s'appuie sur la fusion de résultats de plusieurs classifications crédibilistes, en prenant en compte la fiabilité des sources et le conflit pour la prise de décision finale. Par exemple, comme nous ne disposions pas d'une information exhaustive sur l'assolement sur l'ensemble du territoire, nous utilisons une classification par traitement d'images et une classification par un expert à l'échelle du pixel. Chacune de ces deux sources fournit des masses de croyance quant à l'appartenance d'un type de culture à un pixel donné. Après une étape de combinaison et suivant la règle de prise de décision choisie (maximum de plausibilité, maximum de croyance, maximum de probabilité pignistique), l'utilisateur obtient un résultat qualifié par un degré de confiance.

6. Conclusion

Dans ce travail, nous abordons les problématiques liées à la gestion des imperfections des données agronomiques exploitées dans le cadre de la mise en place d'un observatoire des pratiques agricoles. Afin de modéliser et gérer ces imperfections, nous avons développé deux approches basées respectivement sur la logique floue et les fonctions de croyance en intégrant la notion de qualité de l'information par un degré de confiance. Afin de tester la faisabilité de nos approches, nous avons considéré deux cas d'étude simulés mais bien réalistes ; les premiers résultats obtenus fournissent une information intégrant la notion de qualité

par une mesure de confiance, prête à être stockée dans le système d'information de l'observatoire.

Remerciements

Les auteurs tiennent à remercier le Fond Européen de Développement Régional (FEDER), l'Agence de l'Eau Seine Normandie (AESN) et le Syndicat mixte Intercommunal d'Aménagement du Bassin de la Vesle (SIABAVE) pour le financement de ces travaux.

7. Bibliographie

- [SME00] SMETS P., « Data fusion in the transferable belief model », *Information Fusion*, 2000. *FUSION 2000. Proceedings of the Third International Conference on*, volume 1, pages PS21–PS33. IEEE, 2000.
- [DEV05] DEVILLERS R., JEANSOULIN R., *Qualité de l'information géographique*. Hermès Science Publications, 2005.
- [DES07] DESCONNETS J.C., LIBOUREL T., CLERC S., « Cataloguer pour diffuser les ressources environnementales », *Actes de la conférence INFORSID*, pages 344–361, 2007.
- [DUB08] DUBOIS D., PRADE H., « Gradual elements in a fuzzy set », *Soft Computing-A Fusion of Foundations, Methodologies and Applications*, 12(2):165–175, 2008.
- [ZAY11] ZAYRIT K., DESJARDIN É., DE RUNZ C., AKDAG H., « Propagation of spatial imprecision in imprecise quantitative data in agronomy », *International Symposium on Spatial Data Quality*, pages 145–150, October 2011.