

Développement de l'information électronique et Open Access en Géographie

Mise en perspective des performances des outils bibliographiques avec l'usage et la diffusion de la littérature scientifique



Simona STIRBU

Thèse soutenue le 29 mai 2019 en vue de l'obtention du grade de Docteur en Sciences de l'Université de Liège Collège de doctorat en Géographie

Année Académique 2018 - 2019

**Développement de l'information électronique et Open
Access en Géographie**

**Mise en perspective des performances des outils
bibliographiques avec l'usage et la diffusion de la littérature
scientifique**

Simona STIRBU

**Thèse soutenue le 29 mai 2019 en vue de l'obtention du grade de docteur en Sciences de
l'Université de Liège Collège de doctorat en Géographie**

Année Académique 2018 - 2019

Promoteur : Pr. N. GRECO, Université de Liège

Co-promoteur : Pr. S. SCHMITZ, Université de Liège

Membres du Jury

Dr. F. PROSMANS, Université de Liège, *présidente*

Pr. A. FERRARI, Université de Liège, *secrétaire*

Pr. C. MONTES, Université Lumière Lyon 2

Pr. B. DERRUDER, Université de Ghent

Pr. N. GRECO, Université de Liège

Pr. S. SCHMITZ, Université de Liège

Dr. B. POCHET, Université de Liège

RESUME

En recherche scientifique, le développement du support numérique a révolutionné le paysage documentaire. L'objectif de notre travail a été de cerner l'influence du développement de l'information scientifique électronique et de l'Open Access sur l'usage et la diffusion des publications scientifiques en Géographie. Notre dissertation doctorale s'articule autour de deux volets principaux : d'une part, les outils de recherche bibliographique et leur couverture qui influencent l'usage des documents, et d'autre part, la diffusion et la visibilité des résultats de la recherche au travers de ces outils de recherche bibliographique. Différentes approches méthodologiques ont été adoptées et les résultats sont présentés en deux parties et cinq chapitres.

Dans un contexte budgétaire difficile, lié à la hausse permanente des prix des périodiques et à des politiques éditoriales de plus en plus contraignantes, la question de l'efficacité des bases de données bibliographiques commerciales (telles que WoS, Scopus, GeoRef) d'un côté et du moteur de recherche « gratuit » Google Scholar à l'opposé, a fait l'objet de nombreuses publications. Cependant, pour ce qui concerne la Géographie, peu de données sont disponibles quant à l'efficacité de GS en tant qu'outil de recherche bibliographique et moyen de diffusion des publications scientifiques.

Dans notre travail, autant pour les recherches par mots-clés que pour le recouvrement de bibliographies de thèses ou pour le recouvrement d'une sélection de publications des géographes liégeois, les résultats placent numériquement GS largement en tête du classement des outils bibliographiques analysés, le moteur de recherche ayant trouvé ou retrouvé de très nombreuses références, que ce soit en Géographie physique ou en Géographie humaine. Généralement, GS recouvre complètement les contenus des autres bases de données étudiées, avec quelques exceptions pour l'outil spécialisé en Géographie physique qu'est GeoRef.

En ce qui concerne la visibilité de la littérature scientifique produite par les géographes de l'ULiège, nous avons pu constater une excellente performance de GS, suivi par Scopus, WoS et GeoRef. La question du rôle du répertoire institutionnel ORBi, dans la visibilité des publications liégeoises via GS a été posée. C'est grâce au moissonnage d'ORBi que les références sont toutes visibles et accessibles aux chercheurs via GS, même si la performance de GS serait de toute façon supérieure à celle des bases de données sans l'apport d'ORBi.

La représentation cartographique des téléchargements effectués à partir d'ORBi et une approche statistique des données nous ont permis de conclure que la langue des références, la localisation des pays sur les différents continents, ainsi que l'appartenance des pays qui effectuent les téléchargements à la Francophonie influencent positivement le nombre de téléchargements. Les pays francophones développés ainsi que ceux en voie de développement sont les plus grands utilisateurs de ces publications dont une partie importante est de langue française.

En conclusion, l'ensemble de nos résultats nous permet d'avancer que GS peut constituer une alternative solide aux BD classiques pour la Géographie, sans perdre de vue que la « gratuité » de GS n'est qu'apparente et qu'il convient d'être attentif aux enjeux économiques et sociaux liés à son usage. Par ailleurs, le développement du support numérique offre des opportunités au monde académique de se réapproprier, au moins en partie, la diffusion des résultats de la recherche. Ainsi, le dépôt des publications dans les répertoires institutionnels, lesquels font partie intégrante du mouvement Open Access, améliore leur diffusion indépendamment du type de document ou de la langue de publication.

ABSTRACT

In scientific research, the development of digital media has revolutionized the documentary landscape. The objective of our work was to identify the influence of the development of electronic scientific information and Open Access on the use and dissemination of scientific publications in Geography. Our doctoral dissertation is organized around two main matters: on the one hand, bibliographic research tools and their coverage that influence the use of documents, and on the other hand, the dissemination and visibility of the results returned by these bibliographic research tools. Different methodological approaches have been adopted and the results are presented in two parts and five chapters.

In a difficult financial context, linked to the permanent rise in journal prices and to increasingly restrictive editorial policies, the question of the effectiveness of commercial bibliographic databases (such as WoS, Scopus, GeoRef) against the "free" search engine GS, has been the subject of many publications. However, as far as geography is concerned, there is little data available on the effectiveness of GS as a bibliographic research tool and means of dissemination of scientific publications.

In our work, as much for keyword searches, as for its coverage in theses bibliographies or for its coverage in a selection of publications from ULiege geographers, the results numerically place GS largely at the top of the list of the bibliographic tools analyzed. The search engine finds many references, whether in Physical Geography or Human Geography. Generally, GS completely covers the contents of the other studied databases, with some exceptions for GeoRef, a tool that is specialized in Physical Geography.

Regarding the visibility of the scientific literature produced by the geographers of ULiège, we have noted an excellent performance of GS, followed by Scopus, WoS and GeoRef. The role of the ORBi institutional repository in the visibility of the ULiège publications via GS was questioned. The references are all visible and accessible to researchers via GS thanks to ORBi's harvesting, even if the performance of GS would in any case be greater than that of the databases without the contribution of ORBi.

The cartographic representation of downloads made from ORBi and a statistical approach to data allowed us to conclude that the language of references, the location of countries on the different continents, as well as the membership of the countries that make the downloads to the

French-speaking community positively influence the number of downloads. Developed French-speaking countries as well as those in the process of development are the largest users of these publications, of which a large percentage is French-written.

In conclusion, bearing in mind the fact that GS is only apparently "free" and that there are economic and social implications to its use, our results allow us to argue that the search engine can be a solid alternative to classical commercial databases in Geography. Furthermore, the development of digital media offers the academic world the opportunity to regain, at least in part, the dissemination of research results. Thus, making available the publications in institutional repositories, which are part of the Open Access movement, improves their distribution regardless of the type of document or language of publication.

REMERCIEMENTS

Je voudrais dédier cette thèse à mon papa qui aurais été tellement fier de me voir arriver à la fin de ce long et laborieux travail. Merci de m'avoir toujours encouragé de persévéérer dans tout ce que j'ai entrepris.

A la conclusion de ce travail, je pense à toutes les personnes qui m'ont permis d'aboutir cette recherche doctorale.

Je tiens tout d'abord à remercier ma promotrice Madame Ninfa GRECO, pour tous ses conseils, sa patience dans la relecture, ses critiques et ses commentaires qui m'ont permis de progresser, améliorer et finaliser ce travail.

Mes plus vifs remerciements vont à mon co-promoteur Monsieur Serge SCHMITZ, qui a fait preuve d'une grande disponibilité et d'une grande patience et dont les conseils furent précieux.

J'éprouve également une grande reconnaissance envers Monsieur Paul THIRION pour ses avis et conseils tout au long de cette recherche et lors de l'écriture de ce manuscrit. Je le remercie également pour son investissement en tant que membre du comité de la thèse.

Je remercie les membres du jury pour avoir accepté d'en faire partie Mesdames Fabienne PROSMANS et Aurelia FERRARI, et Messieurs Christian MONTES, Ben DERRUDER et Bernard POCHET.

Merci aussi à tous mes collègues Laurent, Thierry, Philippine et Hélène, ainsi qu'à Sandra L. et enfin à l'équipe BST-Sciences pour leur soutien et accompagnement durant le travail.

Je n'oublie pas de remercier ma fille Ilinca d'avoir fait preuve de patience et compréhension pendant mes multiples et innombrables absences. Je te promets de rattraper au moins une partie de ce temps. Je remercie également mon compagnon Diego pour sa patience et son accompagnement durant ces longues années.

Un tout grand merci à ma famille également, qui de loin m'a toujours encouragée.

Je remercie aussi tous mes amis et amies pour le soutien moral, les encouragements et l'aide occasionnelle, avec une mention particulière à Olivier et Silvia, Madame Marechal, Sonia.

Merci à tous les membres du département de Géographie qui ont toujours montré leur intérêt pour le travail effectué.

Pour terminer, je ne peux ni citer ni remercier personnellement l'ensemble des personnes qui ont contribué à la réussite de ce travail mais je ne les oublierai pas.

Merci, Multumesc

SOMMAIRE

INTRODUCTION GENERALE	1
1. Contexte général de la recherche	3
2. Principaux axes de notre recherche	5
3. Les périodiques scientifiques.....	7
4. L'Open Access.....	12
5. La publication des périodiques en Géographie.....	19
6. Présentation des outils bibliographiques	25
7. La structure de la dissertation et le contenu des chapitres.....	35
PREMIERE PARTIE La littérature répertoriée par les bases de données commerciales et par le moteur de recherche Google Scholar en Géographie	39
Chapitre 1 The Utility of Google Scholar When Searching Geographical Literature: Comparison with Three Commercial Bibliographic Databases	41
1. Introduction.....	41
2. Aim and Objectives	42
3. Methods	43
4. Results.....	47
5. Discussion and Conclusion.....	53
Chapitre 2 - Caractéristiques des périodiques associés aux références uniquement fournies soit par GS, WoS ou GeoRef	56
1. Introduction et revue de la littérature.....	57
2. Objectif et hypothèses.....	59
3. Méthodes.....	59
4. Résultats.....	60
5. Discussion et conclusions	73
Chapitre 3 - Usage de la littérature scientifique dans un échantillon de trois dissertations doctorales	76
1. Introduction et revue de la littérature.....	77
2. Objectifs et hypothèses	82
3. Méthodes.....	83
4. Résultats.....	85
5. Discussion et conclusions	97
CONCLUSIONS GENERALES DE LA PREMIERE PARTIE	101
DEUXIEME PARTIE La littérature scientifique produite par les géographes de l'ULiège : diffusion et usage	105

1. Introduction.....	107
Chapitre 1 Diffusion et visibilité des publications en géographie de l'ULiège dans les bases de données commerciales et GS.....	107
1. Introduction et revue de la littérature.....	109
2. Objectifs et hypothèses	109
3. Méthodes.....	113
4. Résultats.....	113
5. Discussion et Conclusion.....	126
Chapitre 2 - Téléchargements des articles en géographie/géographiques déposés dans ORBi.....	131
1. Introduction.....	131
2. Objectifs et hypothèses	134
3. Méthodes.....	134
4. Résultats.....	136
5. Discussion et Conclusion.....	191
CONCLUSIONS GENERALES DE LA DEUXIEME PARTIE ...	198
CONCLUSIONS GENERALES ET PERSPECTIVES	200
PERSPECTIVES	203

TABLE DE CARTES, GRAPHIQUES, FIGURES ET TABLEAUX

Cartes

Carte 1 - Répartition au niveau mondial des pays de publication des périodiques propres correspondant aux articles uniques de GS pour le mot-clé « urbanization »	65
Carte 2 - Répartition au niveau mondial des pays de publication des périodiques propres à WoS pour le mot-clé « urbanization »	66
Carte 3 - Répartition au niveau mondial des pays de publication des périodiques uniquement indexés par GS pour le mot-clé « sedimentation ».....	70
Carte 4 - Répartition au niveau mondial des pays de publication des périodiques uniquement indexés par WoS pour le mot-clé « sedimentation ».....	71
Carte 5 - Répartition au niveau mondial des pays de publication des périodiques uniquement indexés par GeoRef pour le mot-clé « sedimentation ».....	72
Carte 6 - Distribution au niveau mondial des téléchargements ORBi correspondant aux 453 références publiées entre 2000 et 2014	139
Carte 7 - Distribution au niveau européen des téléchargements ORBi correspondant aux 453 références publiées entre 2000 et 2014.	140
Carte 8 - Distribution mondiale des téléchargements ORBi correspondant aux références publiées en anglais	144
Carte 9 - Zoom sur la distribution des téléchargements ORBi correspondant aux références publiées en anglais en Europe.....	145
Carte 10 - Distribution au niveau mondial des téléchargements ORBi correspondant aux références publiées en français.....	147
Carte 11 - Distribution des téléchargements ORBi correspondant aux références publiées en français en Europe.....	148
Carte 12 - Distribution au niveau mondial des téléchargements effectués à partir d'ORBi correspondant aux 115 références « uniques » à GS	150
Carte 13 - Distribution au niveau mondial des téléchargements effectués à partir d'ORBi et correspondant aux références retrouvées par GS uniquement via ORBi	152
Carte 14 - Distribution au niveau mondial des téléchargements réalisé à partir d'ORBi, correspondant aux références retrouvées par tous les outils bibliographiques.....	154
Carte 15 - Répartition des résidus au niveau mondial	174

Carte 16 - Répartition des résidus au niveau mondial pour les téléchargements des références en langue anglaise 181

Carte 17 - Répartition des résidus au niveau mondial pour les téléchargements des références en français 189

Graphiques

Graphique 1 - Evolution des résidus du modèle retenu 170

Graphique 2 - Evolution des résidus du modèle retenu en fonction de variables explicatives 170

Graphique 3 - Résidus du modèle retenu contre la prédiction 173

Graphique 4 - Résidus du modèle retenu contre la prédiction 181

Graphique 5 - Résidus du modèle retenu contre la prédiction 188

Figures

Figure 1 - Plots showing average results over different search times corresponding to different categories of a second variable 47

Figure 2 - Plots showing conditional averages of the results for two explanatory variables 48

Figure 3 - Unique references provided by each bibliographic tool ... 52

Figure 4 - Pourcentage de périodiques avec un contenu académique, *peer-reviewed*, JCR et OA par outils bibliographique pour "urbanization" 63

Figure 5 - Pourcentage de périodiques avec un contenu académique, peer-reviewed, JCR et OA par outils bibliographiques pour "sedimentation" 68

Figure 6 - Types et nombres de documents cités dans les bibliographies de trois thèses 85

Figure 7 - Années de publication des articles cités dans les trois bibliographies 88

Figure 8 - Année de publication des livres et chapitres de livres cités dans les trois bibliographies 90

Figure 9 - Overlap des références dans les quatre outils bibliographiques pour la bibliographie en Climatologie 94

Figure 10 - Overlap des références dans les quatre outils bibliographiques pour la bibliographie en Tourisme 95

Figure 11 - Overlap des références dans les trois outils bibliographiques pour la bibliographie en Analyse spatiale 96

Figure 12 - Nombre de périodiques référencés par les outils bibliographiques et mention *peer-reviewed* dans Ulrichsweb et ORBi et « auteurs » 120

Figure 13 - Pourcentage de périodiques peer-reviewed, JCR et OA par outils bibliographique	121
Figure 14 - Recouvrement entre les différents outils bibliographiques	123
Figure 15 - Graphique des valeurs observées en fonction des valeurs prévues pour le modèle « test »	155
Figure 16 - Distribution du total des téléchargements	159
Figure 17 - Distribution des téléchargements en anglais	159
Figure 18 - Distribution des téléchargements en français	160
Figure 19 - Distribution des variables dépendantes après la transformation logarithmique	161
Figure 20 - Distribution des variables explicatives quantitatives après la transformation	163
Figure 21 - Distribution des trois variables explicatives catégorielles	164
Figure 22 - Matrice de graphiques de nuages de points	165
Figure 23 - Expression de la prévision du modèle retenu	167
Figure 24 - Distribution des résidus et droite de Henry du modèle retenu	169
Figure 25 - Distances de Cook pour le modèle de régression – nombre de téléchargements total.	172
Figure 26 - Matrice de graphiques de nuages de points pour les téléchargements des références en anglais	175
Figure 27 - Expression de la prévision du modèle retenu pour les téléchargements de références en anglais	177
Figure 28 - Distribution des résidus et droite de Henry du modèle retenu	179
Figure 29 - Distances de Cook pour le modèle de régression retenu – nombre de téléchargements des articles en anglais.	180
Figure 30 - Matrice de graphiques de nuages de points pour les téléchargements des références en français	182
Figure 31 - Expression de la prévision du modèle retenu pour les téléchargements des références en langue française.....	184
Figure 32 - Distribution des résidus et droite de Henry du modèle retenu pour les téléchargements des références en français	186
Figure 33 - Distances de Cook pour le modèle de régression retenu pour les téléchargements des références en français.....	187

Tableaux

Tableau 1 - Caractéristiques et spécificités des outils bibliographiques : bases de données multidisciplinaires WoS, Scopus ; moteur de recherche GS ; bases de données en Géosciences GeoRef et FRANCIS	27
Tableau 2 - Overall and geographical literature for the keywords "urbanization" and "sedimentation"	50
Tableau 3 - Results of the bibliographic tool overlap analysis (TO, RO in A, and RO in B— %).....	51
Tableau 4 - Unique references and associated types of literature for each bibliographic tool	52
Tableau 5 - Définition de l'échantillon pour le mot-clé "urbanization", à partir des références « uniques » à chaque outil, tous types de documents confondus, jusqu'au périodiques « propres » trouvés dans <i>Ulrichsweb</i>	60
Tableau 6 - Définition de l'échantillon pour le mot-clé "sedimentation", à partir des références « uniques » à chaque outil, tous types de documents confondus, jusqu'au périodiques « propres » trouvés dans <i>Ulrichsweb</i>	61
Tableau 7- Répartition des périodiques propres aux trois outils bibliographiques pour le mot-clé "urbanization" en fonction des caractéristiques relevées dans <i>Ulrichsweb</i>	61
Tableau 8 - Répartition des périodiques propres aux trois outils bibliographiques pour le mot-clé "sedimentation" en fonction des caractéristiques relevées dans <i>Ulrichsweb</i>	62
Tableau 9 - Classement des pays d'édition des périodiques correspondant aux articles « uniques » de GS, GeoRef et WoS pour les mot-clé « urbanization »	67
Tableau 10 - Classement des pays d'édition des périodiques correspondant aux articles « uniques » de GS, GeoRef et WoS pour le mot-clé « sedimentation »	72
Tableau 11 - Titres de périodiques les plus utilisés, nombre de citations associées et pourcentage cumulatif de citations pour les trois thèses	87
Tableau 12 - Références bibliographiques et type de documents de chaque thèse :, nombre et pourcentage indexés par les outils bibliographiques.	93
Tableau 13 - Disciplines et langues des différents types de documents relevés dans ORBi	114
Tableau 14 - Pourcentage de chapitres ou parties de livres indexés et par les outils bibliographiques	115

Tableau 15 - Nombre et pourcentages des articles et des périodiques retrouvés dans les outils bibliographiques	117
Tableau 16 - Périodiques sans ISSN.....	118
Tableau 17 - Nombre de périodiques référencés par les différents outils bibliographiques et qualifications correspondantes (selon <i>Ulrichsweb</i> pour JCR, OA et disciplinaire)	119
Tableau 18 - Langue de publication des périodiques référencés par les outils bibliographiques.....	122
Tableau 19 - Nombre et langue des références uniques de GS, sans apport d'ORBi ou <i>via</i> ORBi	124
Tableau 20 - Caractéristiques de périodiques propres à GS	124
Tableau 21 - Récapitulatif chiffré des téléchargements effectués par les pays, pour l'ensemble des publications (TOT_TELECH), pour les publications en anglais (TELECH_EN) et pour les publications en français (TELECH_FR).....	141
Tableau 22 - Détails des téléchargements effectués par les pays, pour l'ensemble des publications (TOT_TELECH), pour les publications en anglais (TELECH_EN) et pour les publications en français (TELECH_FR) avec la répartition des téléchargements effectuées par les différentes catégories de pays et les pourcentages associés	142
Tableau 23 - Téléchargements des articles et communications publiés en français ou anglais et retrouvés uniquement par GS (GS_UNIQ), retrouvés uniquement par GS et uniquement <i>via</i> ORBi (GS_UNIQ_ORBi) ou retrouvés par l'ensemble des outils bibliographiques (BD&GS) * % par rapport au nombre total d'articles correspondants	149
Tableau 24 - Variables sélectionnées pour la construction du modèle.....	156
Tableau 25 - Valeurs manquantes dans les observations pour les différentes variables.....	157
Tableau 26 - Statistiques descriptives des trois variables dépendantes.....	158
Tableau 27 - Statistiques descriptives des trois variables dépendantes après la transformation logarithmique	160
Tableau 28 - Statistiques descriptives des variables explicatives quantitatives après la transformation logarithmique.....	161
Tableau 29 - Coefficients de corrélation entre les différentes variables explicatives quantitatives du modèle	165
Tableau 30 - Résumé de l'ajustement : indicateurs globaux de la qualité du modèle retenu.....	166

Tableau 31 - Variables retenues(en gras) lors de la sélection automatique avec les critères BIC, AICc avec procédure ascendante.....	166
Tableau 32 - Variables retenues et estimations des coefficients	167
Tableau 33 - Codage de la variable composée lors de la sélection automatique des variables.....	168
Tableau 34 - Estimations des coefficients et valeurs du test VIF....	171
Tableau 35 - Coefficients de corrélation entre les variables quantitatives du modèle pour les téléchargements des références en anglais	176
Tableau 36 - Résumé de l'ajustement : indicateurs globaux de la qualité du modèle retenu pour les téléchargements de références en anglais	176
Tableau 37 - Variables retenues (en gras) lors de la sélection automatique avec les critères BIC, AICc, avec procédure ascendante.....	177
Tableau 38 - Variables sélectionnées automatiquement avec la méthode pas à pas (et retenues dans le modèle) et estimations des coefficients.....	177
Tableau 39 - Codage de la variable composée lors de la selection automatique des variables.....	178
Tableau 40 - Coefficients de corrélation entre les variables quantitatives du modèle pour les téléchargements des références en français.....	183
Tableau 41 - Résumé de l'ajustement : indicateurs globaux de la qualité du modèle retenu pour les téléchargements des références en langue française.....	183
Tableau 42 - Variables retenues(en gras) lors de la sélection automatique avec les critères BIC, AICc avec procédure ascendante.....	183
Tableau 43 - Variables sélectionnées automatiquement avec la méthode pas à pas (retenues dans le modèle) et estimations des coefficients pour les téléchargements des références en langue française.....	184
Tableau 44 - Codage de la variable composée lors de la sélection automatique des variables.....	185
Tableau 45 - Codage de la deuxième variable composée lors de la sélection automatique des variables.....	185

Tableau 46 - Résultats des modèles de régression retenus prévoyant les téléchargements du nombre total des articles, des articles en anglais et des articles en français	190
--	-----

LISTE DES ABREVIATIONS

AGI : American Geosciences Institute

BT : Bibliographic tools

BD : Bases de données bibliographiques

DOAJ : Directory of Open Access Journals

GS : Google Scholar

GSM : Google Scholar Metrics

GSC : Google Scholar Citations

JCR : Journal Citation Reports

OA : Open Access

OB : Outils bibliographiques

ORBi : Open Repository and Bibliography

SHS : Sciences humaines et sociales

ST : Sciences et techniques

WoS : Web of Science

« La connaissance progresse en intégrant en elle l'incertitude, non en l'exorcisant. »

E. Morin

INTRODUCTION GENERALE

1. Contexte général de la recherche

La recherche d'information et la communication des résultats, lors de conférences ou par la publication d'articles dans des périodiques scientifiques, occupent un temps important dans les activités de la communauté scientifique. La publication est le principal mécanisme reconnu par lequel les chercheurs apportent une contribution à l'ensemble des connaissances scientifiques et s'affirment en tant qu'experts.

Ainsi, les documents constituant la bibliographie d'une discipline peuvent être considérés comme la base des systèmes de production et de diffusion des connaissances. La visibilité des travaux contribue à l'impact des résultats de recherche sur la communauté scientifique.

Durant les trois dernières décennies, la digitalisation de l'information a complètement modifié notre rapport à la lecture, à la communication, à l'apprentissage et au savoir. D'aucuns n'hésitent pas à établir une analogie entre la révolution des nouvelles technologies de l'Internet et la révolution de l'invention de l'imprimerie attribuée à Gutenberg (Vial, 2013). Mais l'impact de la digitalisation de l'information ne se limite pas uniquement à des aspects techniques : elle influence également le développement intellectuel de chaque lecteur depuis son plus jeune âge et on peut facilement envisager qu'il en résulte des effets sur l'organisation sociale, à tous niveaux.

Le paysage documentaire de la communauté scientifique s'est donc profondément transformé avec la généralisation de l'édition en format électronique des périodiques scientifiques, puis d'autres types de documents. Des consortiums se sont mis en place ouvrant l'accès à des collections complètes de différents éditeurs commerciaux, et le mouvement OAI (*Open Access Initiative, 1999*) a contribué à mettre à la portée de tous une masse considérable d'informations. Le contenu de l'information immédiatement disponible sur le web pour le chercheur est parfois devenu très différent de ce qu'il pouvait trouver en format « traditionnel » dans sa bibliothèque de référence.

Parallèlement, de nouveaux outils de recherche documentaire sont apparus sur le web : des bases de données bibliographiques, des moteurs de recherche gratuits, comme Google Scholar (GS) ou plus récemment Microsoft Academic, ou encore des « réseaux sociaux » pour les scientifiques comme ResearchGate.

Ce phénomène n'a pas manqué de susciter la réflexion des professionnels de l'information et des chercheurs et s'est traduit par la production d'une littérature abondante sur le sujet. Néanmoins, les études abordant cette thématique d'un point de vue de l'intérêt du chercheur, dans une discipline scientifique déterminée comme la Géographie, sont relativement plus rares (Schuermans et al., 2010) excepté peut-être dans le domaine des sciences biomédicales.

Contexte général de la recherche

Sur les changements fondamentaux apportés par l'avènement d'Internet et de la digitalisation de la documentation, se greffent des nouveaux outils d'analyse et de mesure. Les nombreuses critiques quant aux limites de l'Impact Factor (le facteur d'impact des revues défini par le *Journal Citation Reports - JCR*) ont conduit à l'apparition de mesures alternatives et ont donné naissance à d'autres indicateurs bibliométriques. Ainsi, après le développement de l'*Open Access* (OA) et l'encodage massif des articles dans les répertoires institutionnels, le développement des réseaux sociaux pour chercheurs et la plus grande importance accordée à la circulation (*downloads*) des documents, des outils comme *Altmetrics* et *PlumX* ont été développés. Les éditeurs commerciaux ont adopté ces indicateurs de nouvelle génération qui s'affichent dorénavant en regard de la plupart des articles.

- Les nouvelles technologies et les moyens modernes de communication ont donc impacté le paysage documentaire, de la recherche bibliographique jusqu'à la diffusion des résultats de recherche. Dans ce contexte, l'apparition de nouveaux et multiples outils de recherche documentaire ont engendré le besoin de quantifier et catégoriser la littérature fournie et son usage pour les différentes disciplines : « *La connaissance n'étant pertinente que dans la mesure où elle est transmise et accessible aux autres* » (Graham et al., 2011).

Notre recherche doctorale s'intéresse à l'impact de cette révolution numérique pour ce qui concerne la Géographie en particulier. La culture documentaire, les spécificités de publication et de communication de la communauté scientifique que constituent les géographes, sont susceptibles d'être bouleversées à des degrés divers par ce qu'on pourrait appeler « *la globalisation* de la communication scientifique ». Dans le cadre de notre thèse de doctorat nous avons souhaité nous pencher sur ces aspects avec pour objectif de donner des réponses ou pistes de réponses aux questionnements que les chercheurs pourraient avoir.

Les résultats de nos recherches, qui se répartissent en cinq chapitres traitant des différentes facettes de la recherche bibliographique et ses spécificités en Géographie, intègrent des aspects bibliométriques et statistiques. Notre ambition est que les géographes puissent exploiter ces résultats directement pour une meilleure maîtrise de la recherche documentaire et de la communication scientifique à l'heure du numérique et de l'OA.

Chaque discipline a ses spécificités de recherche. Pour la Géographie la situation est relativement complexe, avec une division en deux sous-disciplines principales, Géographie physique et Géographie humaine, qui à leur tour sont subdivisées en un certain nombre de champs distincts fonctionnant de manière quasi-indépendante. Nombreux sont les articles et les réflexions qui rendent compte de cette problématique de l'éclatement de la discipline (Castree et al., 2005; Johnston, 2003, 2005; Livingstone, 2010).

2. Principaux axes de notre recherche

Notre recherche doctorale s'inscrit donc dans le cadre des changements induits par la digitalisation de l'information scientifique et les impacts de celle-ci sur la communication scientifique et la diffusion du savoir.

L'idée centrale porteuse de notre recherche appliquée à la Géographie peut se présenter comme suit :

Le processus de digitalisation de la documentation et de « l'infrastructure bibliographique », les moteurs de recherche gratuits et les systèmes alternatifs de publication et de diffusion des documents, ont fondamentalement changé la recherche bibliographique, l'accès à la littérature scientifique ainsi que son usage et sa diffusion.

Cette idée renvoie à deux composantes qui seront traitées dans la première et la deuxième partie de la thèse : d'une part, les outils de recherche bibliographique et leur couverture, d'autre part la diffusion et la visibilité des résultats de la recherche.

Pour ce qui concerne les outils de recherche bibliographique, le moteur de recherche Google Scholar (GS) couvre gratuitement des sources diverses (éditeurs scientifiques, autant gratuits que payants, archives ou répertoires ouverts des universités et d'autres sites internet...), dont la liste n'est pas rendue publique. Cet outil fournit un nombre de références et une variété de types de documents et de sources qui ne seraient pas accessibles par l'intermédiaire des autres bases de données commerciales et que les géographes pourraient exploiter utilement dans le cadre de leurs recherches.

Aussi, le premier objectif de notre recherche doctorale est de mettre en perspective l'efficacité de GS et des outils bibliographiques commerciaux, multidisciplinaires ou spécialisés en géographie, et d'estimer leurs couvertures et leurs apports en termes de résultats de recherche accessibles pour les géographes. Les chapitres 1, 2 et 3 de la première partie constituent trois approches différentes et complémentaires de cette même question.

Pour ce qui concerne la diffusion et la visibilité des résultats, nous postulons que le répertoire institutionnel de l'ULiège, ORBi¹, permet une plus large diffusion des publications des géographes liégeois, entre autres grâce à son moissonnage par GS.

En conséquence, le deuxième objectif de notre recherche doctorale est d'étudier le référencement par les différents outils bibliographiques d'un échantillon de publications des géographes de l'ULiège déposées dans le répertoire institutionnel ORBi ainsi que les téléchargements de ces documents, respectivement dans les chapitres 1 et 2 de la seconde partie. Par rapport à la première partie du travail, nous

¹ <https://orbi.uliege.be/?locale=fr>

Principaux axes de notre recherche

ajoutons donc la composante essentielle de l'*Open Access* et celle des téléchargements de documents qui y sont associés.

La question de l'efficacité de GS en tant qu'outil de recherche bibliographique gratuit et en tant que moyen de diffusion des résultats de la recherche est essentielle dans un contexte budgétaire bouleversé par la hausse permanente des prix des périodiques et des BD bibliographiques. En Géographie, peu d'études ont été entreprises pour répondre à cette question.

Avant d'entrer dans le cœur de notre travail, quelques concepts de base relatifs à la publication scientifique vont être présentés dans la suite de l'introduction générale, y compris dans leurs aspects historiques, comme autant de piliers sur lesquels poser nos résultats.

3. Les périodiques scientifiques

3.1. Introduction et généralités

Les premiers périodiques apparus en 1665, *Le Journal des Savants* et *Philosophical Transactions of the Royal Society of London*, ont vu le jour pour répondre à la nécessité, ressentie par des scientifiques de plus en plus nombreux, de disposer d'un autre moyen de communication de leurs résultats de recherche que les lettres (Meadows, 1998).

« La conception première était celle d'un recueil des interventions publiques dans les réunions de sociétés savantes à destination de ceux qui ne pouvaient pas être physiquement présents. Dans ces conditions les articles reflétaient un travail en cours, alors que l'édition ultérieure d'un ouvrage restait l'objectif scientifique. » (Crosnier, 1997, p. 60)

Comme les deux premiers et ceux qui ont suivi, beaucoup de ces périodiques étaient des publications de sociétés savantes. Les chercheurs se communiquaient les résultats de leurs recherches en échangeant leurs publications, ce qui leur permettait de gagner la reconnaissance de leurs pairs, fondement de la progression de leur carrière (Livingstone, 2010). Le développement de l'enseignement et des différents champs de recherche ainsi que la multiplication des universités corrélée avec la multiplication des sociétés savantes ont conduit implacablement à une multiplication du nombre de périodiques publiés.

Le paysage de la publication scientifique a beaucoup évolué pendant la deuxième partie du XXème siècle. Après les restrictions imposées par la Deuxième Guerre Mondiale, la reprise économique a contribué au développement de la recherche scientifique et parallèlement à l'agrandissement de la communauté mondiale des chercheurs. Le financement de la recherche par les pays était de plus en plus important tandis que les sous-disciplines de disciplines majeures voyaient le jour.

Au gré des variations du climat socio-politique du XXème siècle, sur base d'une analyse du répertoire *Ulrichsweb*, Mabe & Amin (2001) ont noté une augmentation annuelle du nombre de périodiques au niveau mondial de 3,3 % en moyenne entre 1900 et 1945, et de 4,68 % entre 1945 et 1979, puis un retour au taux précédent de 3,3 % jusqu'aux années 2000. Les auteurs ont aussi établi un lien entre la croissance du nombre de périodiques et l'accroissement du nombre de chercheurs.

Si le nombre de périodiques a augmenté, dans de nombreux cas, leur fréquence de publication ainsi que le nombre d'articles et de pages ont également augmenté (Meadows, 2000). Tenopir & King (2000) quant à eux ont montré qu'aux Etats-Unis, le nombre de revues académiques publiées est passé de 4 175 en 1975 à 6 771 en 1995.

Les périodiques scientifiques

La croissance du nombre de résultats de recherche à publier a conduit à une accélération du processus de publication et à un profond changement dans la manière de publier. Ainsi l'article scientifique est devenu « *l'aboutissement d'une recherche, et dans des nombreux secteurs, les monographies tendent à perdre leur aura.* » (Crosnier, 1997, p. 60).

3.2. La commercialisation de la publication des périodiques scientifiques

Jusqu'au milieu du XXème siècle, les périodiques scientifiques ont contribué à la diffusion des connaissances selon un modèle économique qui permettait l'acquisition de ce type de support par la majorité des institutions de recherche et d'enseignement des pays développés et en voie de développement. Les périodiques publiés par des sociétés savantes n'avaient pas pour vocation de faire des bénéfices financiers. Les excédents étaient utilisés pour aider à financer les projets et objectifs de la société savante.

Dans ce modèle de publication qualifié de « *non profit* », les revenus des sociétés savantes provenaient essentiellement des cotisations de leurs membres. En contrepartie les membres recevaient un exemplaire du périodique. Les abonnements des non-membres (institutions et particuliers) complétaient les revenus des éditeurs académiques (Norris, 2008).

Ce modèle n'a cependant pas résisté longtemps. Après la Seconde Guerre Mondiale, le financement de la recherche s'étant accru, la demande pour de nouveaux périodiques spécialisés a suivi. Les sociétés savantes ne pouvant pas répondre à cette demande croissante, des éditeurs commerciaux comme Pergamon, Blackwell et Elsevier, ont vu l'opportunité de développer leurs intérêts sur le marché académique et d'augmenter leur part de marché (Fredriksson, 2001).

Cette pratique était déjà courante en Allemagne, (Sarkowski, 2001) où les éditeurs commerciaux avaient dominé le marché des revues académiques et des livres avant d'être stoppés par la Deuxième Guerre Mondiale, en 1939-1945.

Au fur et à mesure que leur part de marché augmentait, et sans avoir de concurrence de la part des éditeurs allemands, ni des sociétés savantes, les nouveaux éditeurs commerciaux ont pris progressivement le contrôle du marché et ont pu imposer leurs prix et des taux d'inflation excessifs, reposant sur la loi du marché.

3.3. Le développement du support informatique

Après les années '90, le support informatique devient omniprésent dans la recherche. La rapidité des échanges entre chercheurs s'accroît et leurs recherches et expériences progressent également plus rapidement. Le besoin d'accès à la littérature se fait plus pressant et est rencontré par l'édition scientifique électronique. Les périodiques sont

consultables sous format électronique, même si toujours publiés en format papier parallèlement.

Les éditeurs élaborent de nouveaux modèles d'accès et de livraison de leurs « produits » aux institutions et aux lecteurs. Ce changement a eu un effet fondamental sur la façon dont les éditeurs pouvaient commercialiser et vendre leurs périodiques.

Tandis que le nombre de périodiques scientifiques n'a pas cessé d'augmenter, la possibilité d'accéder à leurs contenus à un coût faible a connu un déclin constant (Tenopir & King, 2000), et ce malgré l'apparition d'internet, qui est apparu comme une opportunité pour la diffusion plus aisée du savoir et des résultats de recherche.

L'important nombre d'articles générés par le passage à l'édition électronique et la nécessité de pouvoir y accéder rapidement ont généré le développement de différents modèles de commercialisation, dont :

- a. L'évolution vers des formules d'achat des périodiques par *packages* ;
- b. La constitution des consortiums d'achat.

Les éditeurs ont rapidement adapté leurs services au support électronique afin de répondre à la demande du monde académique. Le besoin de trouver rapidement de l'information dans des sources de plus en plus nombreuses a conduit à la création des différents outils de recherche bibliographique, dont les bases de données bibliographiques, avec des multiples fonctionnalités. Les références ont pu alors être trouvées et localisées rapidement, et les chercheurs ont pu accélérer le processus de recherche bibliographique et d'analyse de citations.

3.4. Les bases de données bibliographiques

Publié dès 1830, « *Chemische Zentralblatt* » a été le pionnier des bulletins bibliographiques (Michel, 1997). Le rôle de ces recueils était de rassembler les références bibliographiques des publications scientifiques spécifiques à un domaine en un seul bulletin. Leur présentation facilitait l'identification des références utiles par l'indication d'une description bibliographique, une indexation et, éventuellement, un résumé.

A la fin du XIX^{ème} siècle, les bulletins bibliographiques constituaient le moyen d'information et de recherche des références bibliographiques dans la plupart des disciplines et domaines. Ce système a commencé à montrer des signes de faiblesse au milieu des années '60, quand le nombre de références contenues dans un bulletin a atteint plusieurs centaines de milliers de références pour certains domaines. La méthode entièrement manuelle ne répondait plus pleinement aux besoins.

Les évolutions en matière d'informatique, logiciels et matériels, ont alors permis la conception des bases de données bibliographiques. D'abord présentées sur disquette,

Les périodiques scientifiques

ensuite sur CD et finalement en ligne, les bases de données ont révolutionné la manière d'accéder aux références bibliographiques : « *des services commerciaux ont été offerts permettant l'accès en ligne, en temps réel et à distance à des millions de références bibliographiques*"(Michel, 1997, p. 102).

En Géographie, la *Bibliographie Géographique Internationale* (BGI) a été créée en 1891 « par les géographes et pour les géographes² », offrant des informations relatives aux principaux périodiques géographiques du monde, mais aussi aux livres, congrès, thèses, cartes et atlas. Les géographes ont utilisé les bulletins en format papier jusqu'à la fin des années '80. Réalisée sous l'autorité d'un comité scientifique international et parrainée par l'Union Géographique Internationale (UGI), la BGI fut intégrée en 1976 à l'ensemble des bases FRANCIS, gérée par l'INIST-CNRS. La mise à jour de BGI fut arrêtée en 2015 suite à un changement de stratégie du CNRS en matière d'information scientifique et technique.

3.5. Le peer-reviewing, les indicateurs d'usage et l'évaluation des chercheurs

Parmi les services fournis aux chercheurs, avancés comme justificatifs des prix pratiqués par les éditeurs commerciaux, se trouve le processus de *peer-reviewing*, traduit en français « évaluation par les pairs ». Le principal but du *peer-reviewing* est de s'assurer que le contenu d'un article est original, que les données sont valides et que les conclusions sont raisonnables et justifiables (de Vries, 2001). Dans la majorité des cas, les *reviewers* sont des scientifiques qualifiés dans un domaine d'expertise et habilités à poser des jugements objectifs. Harnad (1996) décrit le processus comme un contrôle de qualité qui est essentiel à l'intégrité de la communication scientifique, même si le processus est parfois « désordonné et ne fonctionne pas toujours comme il se doit » (Weller, 2001).

Les résultats des recherches relatifs à la nécessité du *peer-reviewing* s'accordent quant à son importance malgré les coûts engendrés, autant en argent qu'en temps (Rowland, 2002; Schroter & Tite, 2006). Les articles publiés dans des périodiques scientifiques évalués par les pairs contiennent, en principe, une information pertinente : l'information présentée a passé un "contrôle de qualité", qui crédibilise l'article et le périodique. Ces « garanties de qualité » se répercuteraient ensuite au niveau de l'impact des résultats de la recherche sur la communauté scientifique. Il s'ensuit que publier dans ce type de périodiques s'est progressivement imposé comme un critère d'évaluation et de reconnaissance du chercheur, dont dépend sa carrière.

² <http://bgi-prodig.inist.fr/>

Les périodiques scientifiques

En fait, un système d'évaluation des chercheurs basé essentiellement sur deux critères s'est mis en place (Castillo, 2010; Kudelka et al., 2016).

Le premier critère, et le plus ancien, est le prestige des périodiques « *peer-reviewed* » dans lequel les articles sont publiés. Le Journal Impact Factor créé par Eugene Garfield dans les années '60 ("The Thomson Reuters Impact Factor," 1994)(Garfield, 2006) a longtemps été le seul indicateur de prestige des périodiques scientifiques.

Le deuxième critère, plus récent, est basé sur le nombre de citations des articles publiés, avec la variante améliorée proposé par John Hirsch en 2005, le *h*-index (Hirsch, 2005). Ce système d'évaluation encourage la publication dans les périodiques *peer-reviewed*, avec un grand impact.

Harnad et al., (2004) définissent l'impact comme une mesure de l'ampleur avec laquelle les résultats des recherches sont lus, utilisés, cités et appliqués dans les recherches futures. L'impact mesure aussi le progrès et la productivité et a des répercussions sur la carrière des chercheurs en question (par exemple, le salaire, le financement des projets, la reconnaissance, les prix, etc.) et des institutions auxquelles il appartiennent (Miguel et al., 2011)

Une analyse du nombre de citations reçues par certains lauréats du prix Nobel pour leurs travaux (Ashton & Oppenheim, 1978; Garfield & Welljams-Dorof, 1992; Ophof, 1997) a permis d'établir un lien certain entre le nombre de citations reçues et la « reconnaissance» de ces chercheurs.

Sous pression, le chercheur est amené à se concentrer sur la publication d'articles plutôt que vers une "véritable démarche de la recherche" (Ducloy, 2004) ce qui conduit à la pratique de la publication de type « *publish or perish* ». Le système d'évaluation, basé sur les indicateurs d'usage, a conforté la mainmise des éditeurs commerciaux sur le point névralgique de la diffusion des résultats de la recherche scientifique.

Prenant conscience de la complexité de la question, des professionnels ont demandé l'utilisation combinée d'indicateurs pour mieux éclairer la productivité scientifique et son impact (Martin, 1996; Moed & Halevi, 2015; Moed, Van Leeuwen, & Reedijk, 1999).

Dans ce contexte, il serait logique de penser que les articles déposés dans des répertoires en OA, donc susceptibles d'être plus lus et plus souvent cités, devraient avoir un impact plus important (Norris, 2008), hypothèse qui a fait l'objet de différentes études (Davis & Fromerth, 2007; Davis & Price, 2006; Davis & Walters, 2011; Davis, 2010; Gorraiz et al., 2014; Guerrero-Bote & Moya-Anegón, 2014; Harnad & Brody, 2004; Moed, 2007).

L'Open Access

4. L'Open Access

Dans les années '90, l'avènement de l'électronique a amené une facilité de consultation que les scientifiques espéraient pouvoir retourner à leur avantage (Benabou, 2010). Malgré les possibilités offertes par les nouvelles technologies, qui permettaient de découpler le contenu du périodique de son support papier et d'y accéder à distance, le modèle traditionnel d'édition des périodiques académiques est longtemps resté intact, les éditeurs commerciaux continuant à contrôler l'accès à la majorité des titres de revues (Cox, 2003).

Mais l'espoir de pouvoir utiliser les nouvelles technologies afin de changer l'organisation de la publication scientifique a grandi parmi les chercheurs, le système commercial en place étant considéré comme sclérosant et inutilement coûteux pour les institutions chargées de diffuser la connaissance (Benabou, 2010). Les chercheurs des pays en voie de développement, ou appartenant à des institutions de recherche et des Universités n'ayant pas le budget nécessaire pour payer l'accès aux périodiques, sont particulièrement pénalisés par le non-accès aux résultats de la recherche.

Début des années 2000, ce besoin des chercheurs de partager librement leurs résultats a ouvert la voie au mouvement OA (Guédon, 2001, (pp. 15-17); Willinsky, 2006 (p.35)). L'idée principale défendue par les promoteurs de l'OA est que la recherche financée par les États doit être accessible à tous ceux qui le souhaitent et sans frais, que ce soit des chercheurs ou simplement des citoyens ayant contribué au financement de cette recherche. L'accès au contenu des revues scientifiques devrait être un service d'usage gratuit, pour ceux qui ont accès à internet.

Il existe de nombreuses définitions de l'OA (les déclarations de Budapest, Bethesda et de Berlin). La plus détaillée et convaincante est la déclaration de Budapest, lancée lors de l'Open Society Institute's Budapest OA Initiative (2002) : « *Par "accès libre" à [la littérature de recherche validée par les pairs], nous entendons sa mise à disposition gratuite sur l'Internet public, permettant à tout un chacun de lire, télécharger, copier, transmettre, imprimer, chercher ou créer un lien vers le texte intégral de ces articles, les analyser automatiquement pour les indexer, s'en servir comme données pour un logiciel, ou s'en servir à toute autre fin légale, sans barrière financière, légale ou technique autre que celles indissociables de l'accès et de l'utilisation d'Internet. La seule contrainte sur la reproduction et la distribution et le seul rôle du droit d'auteur dans ce contexte devrait être de garantir aux auteurs un contrôle sur l'intégrité de leurs travaux et le droit à être correctement reconnus et cités.³*

³ <http://www.budapestopenaccessinitiative.org/>

A ce jour, malgré quinze années d'effort de la part des professionnels des bibliothèques, et des chercheurs intéressés par la problématique, l'accès à la majorité des périodiques académiques est toujours payant et les coûts sont en hausse constante, pour les lecteurs ou leurs institutions. L'accès aux résultats de la recherche se heurte toujours à des obstacles financiers avec une perte du potentiel d'impact de la recherche. Néanmoins, des progrès ont été observés. Afin de pallier aux coûts exorbitants, différents modèles d'accès à la littérature scientifique ont été adoptés, reposant sur des modes de financement différents.

Conscients de l'importance primordiale que constitue l'innovation et la recherche, l'Union Européenne (UE) a mis en place des programmes cadre dont le « FP7 » et plus récemment «*Horizon 2020*» (2014-2020)⁴, qui ont pour objectif de stimuler la croissance économique et de créer des emplois en associant recherche et innovation. Une des priorités du programme, « l'Excellence Scientifique » vise à soutenir et favoriser la recherche et l'innovation avec l'attribution des financements appropriés. Les résultats de la recherche seront ensuite valorisés et transformés en produits et services qui visent à la création de débouchés commerciaux et qui contribueront à l'amélioration du niveau de vie des citoyens⁵.

L'OA permet d'accroître la visibilité des résultats des recherches issues des projets financés par les programmes-cadres de l'UE. Ainsi le programme Horizon 2020 comporte notamment l'obligation d'assurer le libre accès aux publications issues des recherches qu'il aura contribué à financer.

Ainsi la Commission Européenne (CE) a mis en œuvre une plateforme, OpenAIRE⁶, dont le but est d'accompagner l'obligation de dépôt en accès libre (deposit mandates) décidée par la CE et le Conseil Européen de la Recherche (ERC), et de diffuser les publications et les données scientifiques des recherches financées avec des fonds européens et pas seulement. OpenAIRE agrège les données en indexant 1 164 archives⁷ et répertoires institutionnels (data providers).

⁴ <http://www.horizon2020.gouv.fr/cid73300/comprendre-horizon-2020.html>

⁵ Source: https://ec.europa.eu/programmes/horizon2020/sites/horizon2020/files/H2020_FR_KI0213413FRN.pdf

⁶ Open Access Infrastructure for Research in Europe <https://www.openaire.eu/search/find?keyword=Open>

⁷ aout 2018

L'Open Access

4.1. Les différents modèles de l'*Open Access* et les différentes initiatives

Harnad et al., (2004) décrivent deux principaux modèles d'OA. Le premier est connu sous l'appellation « voie verte » ou « *green road* ». Les auteurs, par exemple, qui suivent cette « voie » auto-archivent leurs articles publiés dans des « périodiques traditionnels » sur leurs pages personnelles, dans un répertoire institutionnel ou thématique accessible facilement *via* Internet.

La « voie verte » serait d'après Harnad et al., (2004), la seule option qui conduirait à 100 % d'OA dans un proche avenir car elle n'exige pas une restructuration complète du système de publication scientifique. Pour rencontrer les objectifs de cette option, il a été demandé aux institutions de créer des répertoires conformément au protocole BOAI. Par ailleurs, afin que les chercheurs disposent d'indications claires quant aux politiques des éditeurs en termes de restrictions appliquées pour l'auto-archivage, un site a été développé, *SHERPA-Romeo*, qui représente par un code couleur les différents niveaux d'autorisation de dépôt dans des répertoires institutionnels accordés par les éditeurs. Des 2 559 éditeurs repris dans la base, 81 % autorisaient en août 2018 l'auto-archivage avec ou sans restrictions⁸.

Le second modèle est celui de la « voie dorée » ou « *gold road* » qui implique que les auteurs publient leurs résultats de recherche dans les revues en OA. Pour accroître la visibilité et promouvoir l'utilisation des périodiques de la voie dorée, le « Directory of Open Access Journals » ou DOAJ⁹ a été créé. Ce répertoire rassemble les périodiques internationaux qui contribuent au mouvement OA. Leur nombre est en constante augmentation, progressant de moins de 1 400 titres au début 2005 à 5 138 titres en juin 2010 (Miguel et al., 2011) et à 11 928 titres répartis sur 128 pays en août 2018 (avec un ajout de 2 421 titres sur une année). Cette augmentation n'est pas seulement due à la création de nouveaux titres mais également de la reconnaissance par DOAJ des titres OA existants et pas encore intégrés dans la base.

Il existe des dérives aux modèles OA, des adaptations très connues étant la publication *via* le « *unfair gold* » et le périodique hybride. Dans ce type de périodiques, les auteurs ou leurs institutions doivent payer un montant souvent élevé, nommé « *Article Processing Charges (APC)* », pour publier en OA en plus du maintien de la souscription aux titres de périodiques. Les prix pratiqués ne laissent pas la possibilité à tous les chercheurs de publier de cette manière, notamment à ceux des pays en voie de développement comme Bernard Rentier l'explique sur son blog¹⁰.

⁸ <http://www.sherpa.ac.uk/romeo/statistics.php>

⁹ <http://www.doaj.org/>

¹⁰ <https://bernardrentier.wordpress.com/2018/01/17/science-needs-a-break/>

Une autre dérive à la « voie dorée » sont les « éditeurs prédateurs » qui profitent de la tendance et de la naïveté de nombreux chercheurs pour développer des périodiques en ligne dans des domaines où ils n'ont pas d'expertise. De nouveau, pour publier dans ces périodiques, les auteurs doivent payer des sommes assez élevées. Mais la qualité du *peer-reviewing* chez ces éditeurs est douteuse voire inexistante. Jusqu'en 2017 Jeffrey Beall recensait la plupart des « éditeurs prédateurs » et les périodiques publiés par ces derniers, sur une liste la « *Beall list of predatory journals* ». Cette liste n'est plus actualisée par Jeffrey Beall qui a choisi de ne plus travailler sur le sujet¹¹ en 2017. Actuellement une copie de la liste avec les mises à jour reste consultable sur le Web¹².

Ces dérives de la « voie dorée » renforcent l'idée que la « voie verte » est la plus prometteuse et laisse aux chercheurs la possibilité de publier selon le système traditionnel mais leur permet en plus de diffuser personnellement leurs résultats.

En parallèle, la problématique des pays en voie de développement a conduit à la mise en place d'un accès aux périodiques *via* une version distincte de l'OA, souvent gratuite ou à un coût minimal. Rowland, (2005) décrit deux exemples: HINARI (*Health InterNetwork Access to Research Initiative*) mis en place en 2002 et AGORA (*Access to Global Online Research in Agriculture*) lancé en October 2003. Il existe aussi un troisième projet OARE¹³, mis en place par « *United Nations Environment Programme (UNEP)* » en partenariat avec les grands éditeurs. Comme les deux autres projets, celui-ci lancé en 2006, permet aux pays en développement d'accéder à l'une des plus grandes collections de résultats de recherche en sciences de l'environnement au niveau mondial (“OARE,” 2017).

En Fédération Wallonie-Bruxelles, des avancées importantes concernant l'adoption d'une pratique de publication orientée vers la « voie verte » ont été réalisés le 28 février 2018. Le projet de décret "Open Access" a été adopté par le Gouvernement de la Fédération Wallonie-Bruxelles en dernière lecture. Concrètement, le texte de ce décret prévoit que tous les articles scientifiques subventionnés par des fonds publics doivent être encodées obligatoirement en OA dans une archive ouverte institutionnelle avec un embargo de 6 ou 12 mois maximum. Les listes de publications issues de ces archives ou répertoires institutionnels seront exclusivement utilisées pour l'évaluation des chercheurs¹⁴.

¹¹ <http://retractionwatch.com/2017/01/17/bealls-list-potential-predatory-publishers-go-dark/>

¹² <https://beallslist.weebly.com/contact.html>

¹³ Online Access to Research in the Environment

¹⁴ <https://lib.uliege.be/en/news/l-open-access-adopte-par-le-gouvernement-de-la-federation-wallonie-bruxelles>

L'Open Access

Plus récemment, en Juillet 2018, le Parlement Fédéral de la Belgique a adopté une loi modifiant le droit d'auteur belge et permettant aux auteurs d'articles scientifiques financés par le secteur public de conserver le droit de déposer leurs articles disponibles en libre accès, même s'ils ont cédé leurs droits à un éditeur :

“L'auteur d'un article scientifique issu d'une recherche financée pour au moins la moitié par des fonds publics conserve, même si, conformément à l'article XI.167, il a cédé ses droits à un éditeur d'un périodique ou les a placés sous une licence simple ou exclusive, le droit de mettre le manuscrit gratuitement à la disposition du public en libre accès après un délai de douze mois pour les sciences humaines et sociales et six mois pour les autres sciences, après la première publication, dans un périodique, moyennant mention de la source de la première publication.” (Le Moniteur Belge - Belgisch Staatsblad, 2018, p. 68691, Art. 29)

4.1.1. Systèmes d'archivage de la Voie Verte

Presque cent ans avant le développement de l'Internet, en 1878, Daniel Gilman, premier président de la John Hopkins University, écrivait « *It is one of the noblest duties of a university to advance knowledge, and to diffuse it not merely among those who can attend the daily lectures – but far and wide* ». Gilman se référat certainement à la diffusion du savoir via les périodiques scientifiques et les livres édités par les presses universitaires et les sociétés savantes, mais dans l'ère numérique cette idée est d'application à toute information scientifique sous ses différentes formes (Jones, Andrew, & MacColl, 2006).

Il existe divers types de systèmes d'archivage des documents électroniques dans lesquels différents types de documents peuvent être encodés par les auteurs en OA : les archives disciplinaires, les répertoires institutionnels et les dépôts nationaux généraux. Les limites des systèmes d'archivage ne sont actuellement pas bien définies. Un nombre important de répertoires institutionnels ou d'archives ont été créés à partir des années '90, dans lesquels les chercheurs, dans le respect des conditions posées par les éditeurs, peuvent archiver leur travail.

Une des plus célèbres archives, lancée en 1991 par Los Alamos National Laboratory, ArXiv a accumulé 1 521 586 enregistrements au cours de son existence de vingt-huit ans. Initialement, l'archive comprenait des documents en physique des particules et des hautes énergies ; on pouvait donc la qualifier d'archive disciplinaire. Maintenue et exploitée actuellement par la Cornell University Library, les domaines couverts se sont

élargis à « la Physique, les Mathématiques, l'Informatique, les sciences non linéaires, la Biologie quantitative, les Finances quantitatives et les Statistiques »¹⁵.

Du côté francophone, HAL est un exemple représentatif d'archive nationale. « *L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion d'articles scientifiques, publiés ou non, et de thèses, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés* »¹⁶. Lancée en 2001, le contenu de l'archive s'élevait en avril 2019 à 1 769 901 références dont 47, 4% articles.

Si ces exemples semblent témoigner du succès de la formule, cela est généralement beaucoup plus difficile sur le terrain dans la majorité des cas. Lorsque les gestionnaires des répertoires proposent l'auto-archivage aux auteurs, sans contrainte, peu d'entre eux le font volontairement. Pour les chercheurs, le dépôt dans les répertoires institutionnels apparaît trop souvent comme une procédure laborieuse à suivre, couteuse en temps pour peu de retour. Ils manquent alors de motivation pour déposer.

Cet état de fait a conduit à la multiplication des appels pour que les publications issues des recherches financées par les fonds des États soient régulièrement déposées dans des archives ouvertes (Bosc & Harnad, 2005; Swan & Brown, 2005). Cela a généré une vive controverse entre les éditeurs, les organismes de financement et ceux qui sont à la fois pour et contre l'OA (Norris, 2008).

Avec le lancement du répertoire institutionnel ORBi, l'Université de Liège a contribué au développement de cette solution de la « voie verte ».

4.1.2. ORBi

En Fédération Wallonie Bruxelles, l'émergence de l'OA a conduit à la mise en place d'un projet de la BICfB¹⁷ qui a débuté en 2004 et avait pour but l'implémentation de répertoires institutionnels académiques dans les universités francophones.

En mai 2007, sous la recommandation de Bernard Rentier, recteur de l'ULiège et fervent militant du mouvement OA, le Conseil d'Administration de l'Université a décidé de créer un répertoire institutionnel assorti d'une politique forte de dépôt obligatoire (Liège Mandate) pour les membres du personnel. ORBi a été lancé officiellement le 26 novembre 2008 (Thirion, 2016), et a pour but de collecter, préserver et diffuser la production scientifique des membres de l'Université via l'Internet sans contraintes financières ou techniques.

¹⁵ <https://arxiv.org/>

¹⁶ <https://hal.archives-ouvertes.fr/>

¹⁷ Bibliothèque Interuniversitaire de la Communauté française de Belgique

L'Open Access

Mondialement connu, le mandat de dépôt obligatoire a contribué au succès rencontré par le répertoire. A l'ULiège, l'évaluation du personnel scientifique s'effectue uniquement sur base des rapports de publications générés à partir du répertoire institutionnel. Cette politique motive et responsabilise les scientifiques face au dépôt de leurs travaux.

La politique de dépôt obligatoire, la gestion efficace du système ainsi que l'adoption du dépôt systématique par les chercheurs ont conduit au succès et à la visibilité au niveau mondial du répertoire et de son contenu.

Sur une durée de plus de dix ans, le nombre de références a atteint en avril 2019 une valeur de 172 908 références (“ORBi,” 2019) : 112 302 références sont assorties du texte intégral et pour 64 040 d'entre elles, le texte intégral est directement accessible en OA. Le texte intégral des autres références est accessible sur demande.

L'interface ORBi donne la possibilité aux chercheurs de déposer plusieurs types de publications et matériels de recherche, qu'ils soient publiés ou simplement communiqués, pour autant qu'ils soient correctement décrits par les métadonnées. A chaque référence déposée, une ou plusieurs disciplines thématiques doivent être associées et une licence de diffusion doit être signée. L'exactitude des informations introduites n'est pas systématiquement vérifiée par les gestionnaires d'ORBi. Les auteurs sont entièrement chargés du dépôt et responsables des données correspondant aux références.

ORBi fournit des statistiques spécifiques par article et par auteur, visibles uniquement pour les membres d'ULiège, ainsi que des statistiques générales, publiquement disponibles. Les statistiques montrant le nombre de téléchargements et de visualisations sont basées sur des accès réels, réalisés par l'homme ; les accès par robots, crawler, etc. sont systématiquement filtrés et supprimés.

Dans le « *Ranking Web of World Repositories* », ORBi s'est classé en 19^{ème} position dans la catégorie Classement global des répertoires institutionnels en 2015. Actuellement, le classement des répertoires n'est plus effectué.

ORBi est basé sur le logiciel open source Dspace v1.5, est conforme aux normes W3C et compatible avec XHTML 1.0, CSS 2.1 ainsi qu'avec le protocole OAI-PMH, ce qui permet son indexation par des méta-moteurs de recherche d'OAI (Scientific Commons, OpenAIRE, OAIster, Google Scholar). ORBi est aussi indexé par des moteurs de recherche comme Google, Yahoo, Microsoft Academic, ou des « *discovery tools* » comme PrimoCentral d'ExLibris et EBSCO *Discovery Service*, ou encore d'autres plateformes comme PubMed, Core, BASE et OpenAIRE.

4.1.3. Nouveaux indicateurs

Le passage du support « papier » au support électronique et le développement des réseaux sociaux pour les scientifiques ont permis l'émergence et la diversification d'indicateurs d'usage chiffrés, en la matière.

Différents projets ont été initiés afin d'exploiter le nombre de téléchargements au niveau des périodiques et des articles, ou les données issues des différents réseaux sociaux pour les scientifiques (Bollen & Sompel, 2008; Shepherd, 2007, 2011). Les projets MESUR¹⁸ ou PIRUS¹⁹, par exemple, tendent à quantifier des données d'usage pour compléter les indicateurs de qualité de publications déjà existants.

Dans le cadre du projet MESUR, 39 critères d'impact ou de visibilité ont été analysés afin de vérifier s'ils sont dépendants, complets et avec quel niveau de précision ils représentent l'impact scientifique (Bollen et al., 2009). Les résultats ont permis de constater que le concept de qualité / visibilité est multidimensionnel et ne peut être évalué par un seul indicateur même si certaines mesures se sont montrées plus appropriées que d'autres. Des aspects complémentaires doivent être pris en compte afin de fournir des informations objectives (Gorraiz et al., 2014).

A ce stade, il en ressort que l'OA influence les chances pour qu'un document soit retrouvé, consulté, utilisé et cité. L'OA influence donc l'accessibilité à la science et au savoir et accroît la circulation des connaissances (Davis et al., 2008; Norris et al., 2008; Willinsky, 2006).

5. La publication des périodiques en Géographie

Lors du début de l'enseignement de la Géographie dans les universités, la Géographie a été promue en tant que discipline intégrée. Les chercheurs publiaient dans des périodiques de géographie ne présentant pas d'orientations spécifiques. Ensuite, la Géographie a été divisée en deux sous-disciplines principales qui, à leur tour, ont été divisées en d'autres sous-disciplines : Géographie physique (géomorphologie, hydrologie, climatologie et météorologie) et la Géographie humaine (géographie de la population, la géographie rurale et urbaine, la géographie sociale, la géographie économique et des transports). Les changements intervenus dans les pratiques de la publication scientifique ont influencé la manière de publier des géographes, mais de façon différente en fonction des sous-disciplines et de leurs différentes branches.

Il s'avère que les chercheurs en Géographie physique publient actuellement dans des périodiques autres que ceux à caractère purement géographique (en Sciences et

¹⁸ Metrics from Scholarly Usage of Resources (<http://mesur.informatics.indiana.edu/>)

¹⁹ Publisher and Institutional Repository Usage Statistics (<http://www.cranfieldlibrary.cranfield.ac.uk/pirus2/tiki-index.php?page=pirus2>)

La publication des périodiques en Géographie

Environnement, en général), tandis qu'en Géographie humaine, les chercheurs continuent à publier dans des périodiques du domaine (Castree et al., 2005; Johnston, 2005; Johnston & Sidaway, 2015) ainsi que des livres et chapitres de livres, se rapprochant ainsi de pratiques de publication en SHS .

La séparation des courants de pensée en Géographie entre les pays anglophones (Etats-Unis et Royaume-Uni, Australie, et Nouvelle-Zélande) et les pays francophones d'Europe (France, Belgique, Suisse) ont aussi généré des différences dans la publication. L'hégémonie anglo-américaine, tant pour ce qui est de la langue de publication que du ratio de périodiques publiés par des pays anglophones dans le domaine de la Géographie, et son influence sur les pratiques de publication ont largement été analysées (Aalbers & Rossi, 2006; Bosman, 2009; Foster et al., 2007; Halleux, 2013; Schmitz, 2003; Schuermans et al., 2010).

Ainsi, pour ce qui concerne la Belgique en particulier, Schuermans et al., 2010 ont étudié les bibliographies de chercheurs travaillant en Géographie humaine sur une période de quarante ans. Il en ressort que les géographes belges du nord du pays publient de plus en plus en anglais et dans des revues anglophones éditées par des éditeurs commerciaux, délaissant la publication dans leur langue native et dans les revues éditées par les sociétés savantes belges de la discipline. Dans le sud du pays la tendance est moins accentuée, mais est présente également.

5.1. L'historique de la publication en Géographie

Dans les pays francophones et Outre-Manche, bien connus pour leur tradition de publication, les premiers titres de périodiques de géographie sont apparus au cours du XIX^{ème} siècle.

En **Belgique**, la *Société de Géographie d'Anvers* et la *Société Royale Belge de Géographie* ont été fondées en 1876 à l'initiative du roi Léopold II (Nicolaï, 2004), à la suite du premier Congrès International de Géographie qui eut lieu à Anvers en 1871 et de la Conférence géographique de Bruxelles de 1876 (Vandermotten, 2008). Chacune de ces deux sociétés publiait son propre périodique (Vandermotten & Kesteloot, 2013), respectivement le *Bulletin de la Société de Géographie d'Anvers* et le *Bulletin de la Société Belge de Géographie* (devenu en 1962, la *Revue Belge de Géographie*).

Suite à l'arrêté royal de 1900 qui implémentait le doctorat en Géographie dans les universités d'état (Gand et Liège), le Séminaire de Géographie fut créé en 1903 par Joseph Halkin à Liège (Merenne-Schoumaker, 2004). Puis, le Cercle des Géographes Liégeois a été créé en 1928 et le *Bulletin du Cercle des Géographes Liégeois* a vu le jour. Le Cercle des Géographes Liégeois est devenu, en 1965, la Société Géographique

La publication des périodiques en Géographie

de Liège qui a continué la publication du Bulletin mais sous un autre nom, *Bulletin de la Société Géographique de Liège* (Merenne-Schoumaker, 2004; Nicolaï, 2004).

En parallèle, suite à l'établissement légal de formations universitaires en Géographie en 1929, la *Société Belge d'Etudes Géographiques* a été créée en 1931 en tant qu'association interuniversitaire des géographes académiques. Le *Bulletin de la Société Belge d'Etudes Géographiques*, publié en français et en néerlandais (*Tijdschrift van de Belgische Vereniging voor Aardrijkskundige Studies*) et édité par cette société, se démarquait à l'époque par rapport aux publications des deux sociétés antérieures, qui s'adressaient aux intellectuels, marchands, investisseurs et militaires (Vandermotten & Kesteloot, 2013).

Le *Bulletin de la Société de géographie d'Anvers*, quant à lui, n'est plus publié depuis 1968.

Le paysage de la publication géographique belge a été changé en 2000 par le regroupement du *Bulletin de la Société Belge d'Etudes Géographiques* et de la *Revue belge de Géographie* en une seule publication intitulée *BELGEO, Revue Belge de Géographie*.

Actuellement les deux titres belges subsistant, *BELGEO, Revue belge de Géographie* et le *Bulletin de la Société Géographique de Liège* sont publiés en OA. Le premier sur le portail de ressources électroniques en sciences humaines et sociales *OpenEdition*²⁰, et le deuxième sur le portail PoPuPs²¹ mis en place par l'ULiège Library afin de permettre aux éditeurs de revues scientifiques de l'Université de Liège de diffuser le contenu de ces revues en OA.

En France, le panel de périodiques publiés en Géographie est plus varié. Le nombre de géographes illustres et le plus grand degré d'importance accordé aux voyages et à la découverte d'autres pays ont amené dès 1808 à la création du *Journal des Voyages* (dir. Malte-Brun). Fondée en 1821, la Société de Géographie conduit à l'apparition du *Bulletin de la Société de Géographie* (1822-1899), qui devient *La Géographie : Bulletin de la Société de Géographie* à partir de 1900.

Par la suite, chronologiquement la parution des périodiques en Géographie en France, se présente ainsi (Joseph, 1997):

- 1863-1880 *Année Géographique*
- 1878-1918 *Bulletin de la Société de Géographie Commerciale. Paris* (devient *Revue Economique Française* en 1919)
- 1876 *Revue Géographique Internationale*

²⁰ <http://journals.openedition.org/belgeo/>

²¹ Le Portail de Publication de Périodiques Scientifiques : <http://popups.ulg.ac.be/0770-7576/index.php>

La publication des périodiques en Géographie

- 1877-1924 *Revue de Géographie Annuelle* (dir. Drapeyron)
- 1878 *Bulletin de la Société Languedocienne de Géographie*
- 1891 *Annales de Géographie (Bibliographie Géographique Annuelle* incluse 1891-1915)
- 1891 *Bulletin de Géographie Historique et Descriptive* (CTHS)
- 1910 *Bulletin de la Société des Professeurs d'Histoire et de Géographie* (devient *Historiens et Géographes* en 1966)
- 1913 *Recueil des Travaux de l'Institut de Géographie Alpine* (devient *Revue de Géographie Alpine* en 1920)
- 1922 *Bulletin de l'Association des Géographes Français*

En **Allemagne**, la *Société de Géographie de Berlin (Gesellschaft für Erdkunde zu Berlin)* a été fondée en 1828 à l'initiative de Carl Ritter, Alexander von Humboldt et Heinrich Berghaus. C'est la deuxième plus ancienne société de Géographie au monde après celle de France. Accordant une importance particulière à l'exploration et à la découverte du monde, la société a soutenu différentes expéditions notamment en Afrique ainsi que l'expédition sino-suédoise (1927-1935).

Chronologiquement, la parution des périodiques édités par la Société de Géographie de Berlin se présente ainsi :

- 1834 - 1839 *Jährliche Uebersicht der Thätigkeit der Gesellschaft für Erdkunde in Berlin* → devient en 1839 – 1852 *Monatsberichte über die Verhandlungen der Gesellschaft für Erdkunde zu Berlin* → devient en 1853 - 1865 *Zeitschrift für allgemeine Erdkund* → devient en 1866 - 1944 *Zeitschrift der Gesellschaft für Erdkunde zu Berlin* → devient en 1949 -- (en cours) *Die Erde*.

D'autres périodiques allemands ont été publiés en Géographie :

- 1855 - 1878 *Mittheilungen aus Justus Perthes' geographischer Anstalt über wichtige neue Erforschungen auf dem Gesamtgebiete der Geographie* → devient en 1879 - 1937 *Petermanns Mitteilungen* devient en 1938 - 2004 *Petermanns geographische Mitteilungen*.
- 1866 - 1983 *Geographisches Jahrbuch*
- 1895 - 1917 *Bibliotheca geographica*

Au **Royaume-Uni**, la publication des périodiques en Géographie débute en 1830 avec la fondation de l'*Institute of British Geographers*. Tous les titres étaient publiés par des sociétés savantes - *Geographical Journal* publié à partir de 1830, *Geography and Transactions, Institute of British Geographers*, plus le *Scottish Geographical Magazine* publié à partir de 1884 et le seul périodique « régional » du Royaume-Uni. Plus tard, en 1969, *Area* a d'abord été publié comme une lettre d'information de

La publication des périodiques en Géographie

l'*Institute of British Geographies* et est ensuite devenu le deuxième périodique évalué par les pairs après *Transactions* (Johnston, 2003).

Globalement, comme dans toutes les branches des sciences, le nombre de périodiques en Géographie a augmenté progressivement. Un projet mis en place par l'Union Géographique Internationale (UGI)²² estime à 1 310 le nombre de périodiques publiés en Géographie actuellement, dont la majorité sont en anglais (723). La publication de ces périodiques se répartit de la manière suivante : Afrique, 117 périodiques ; Amérique, 398 périodiques ; Asie et Océanie, 332 périodiques et Europe, 463 périodiques²³. En 2010, approximativement 400 de ces périodiques se retrouvaient dans ISI - Sciences ; et 300 en ISI - Social Sciences.

5.2. La langue de publication

Les pratiques de publication évoluant vers l'anglais sont particulièrement défavorables pour les chercheurs travaillant en Géographie humaine, en particulier pour ceux intéressés par l'articulation de leurs connaissances théoriques avec les situations sociales concrètes et localisées. Dans ce contexte, la question de la langue de publication est cruciale.

En 2003, Schmitz, (2003) affirmait que : « *Le passage vers une écriture du monde dans une langue unique permettrait à priori une communication plus aisée à l'intérieur une grande communauté de géographes. Elle permettrait à notre discipline de progresser plus vite. Néanmoins ce serait oublier la nécessaire pluralité des géographies. Car la langue n'est pas un simple reflet du monde extérieur elle l'habite. Décrire et analyser le monde est une interaction entre une personne, un géographe plongé dans sa propre culture et un milieu. La langue utilisée n'est pas neutre et traduit déjà des ordres de valeurs et des significations. Reconnaissant cette diversité d'approches du monde on ne peut néanmoins négliger l'enrichissement de nos géographies par une comparaison une confrontation avec l'autre et l'ailleurs.* ». ».

Par la suite, dans leur analyse des pratiques de publication des chercheurs en Géographie humaine en Belgique, Schuermans et al., (2010) affirmaient : « *Dans le contexte d'un monde académique en pleine globalisation, les géographes en Europe continentale doivent non seulement produire des formulations théoriques et études empiriques dans leur propre langue et dans leur communauté intellectuelle, avec ses propres traditions, logiques et langages, mais aussi dans un contexte international plus large, avec l'anglais comme lingua franca* ». ²⁴

²² IGU Journal Project

²³ Source : IGU- <https://igu-online.org/wp-content/uploads/2014/08/IGU-JOURNAL-PROJECT.pdf>

²⁴ Citation traduite de l'anglais

La publication des périodiques en Géographie

Selon Vandermotten & Kesteloot (2013), la non-maitrise de l'anglais « *crée un handicap qui est particulièrement sensible et amplifié pour les géographes humains non anglo-saxons* ».

D'avis partagé par différents auteurs, il en ressort que le lieu où est réalisée l'étude et la langue de publication généreront des pratiques d'interprétation et de publication des résultats différentes (Aalbers, 2013; Houssay-Holzschuch & Milhaud, 2013; Rossi, 2008).

Les « contraintes » auxquelles la recherche et la communication des résultats sont soumises actuellement conduisent à négliger des missions et caractéristiques locales inhérentes à certaines disciplines. Dans le numéro inaugural de la revue *Belgeo*, Vandermotten & Kesteloot (2013), ont bien souligné ces aspects critiques de la publication en Géographie :

« *Au-delà de la question de l'hégémonie de certaines revues, la revendication humaniste et citoyenne à la désexcellence scientifique renvoie aussi à celle de la contribution sociale du géographe. Celle-ci est sans doute bien plus grande quand il publie un article critique dans la revue en ligne en libre accès Brussels Studies que lorsqu'il produit un texte bien moins lu, mais qui lui attribue des points pour son curriculum, dans une revue dite internationale, pour lequel il s'est soumis volontairement ou non aux normes auto reproduites et aseptisées.* » (Vandermotten & Kesteloot, 2013, p. 7)

5.3. Le Département de Géographie de l'ULiège

Le choix d'étudier en particulier le département de Géographie de l'Université de Liège, outre l'intérêt institutionnel et un accès plus aisés à l'information, peut être motivé par d'autres raisons scientifiques.

Dans le département de Géographie de l'ULiège, toutes les branches de la Géographie sont enseignées dans le programme Bachelier en Sciences géographiques, orientation générale. Pour se spécialiser par la suite, les étudiants auront le choix entre trois masters à orientation : en **climatologie et géomorphologie**, en **développement territorial ou en didactique**, et en **géomatique et géométrie**. Les deux sous-disciplines majeures de la Géographie sont bien représentées. Le programme des cours est assuré par des géographes faisant partie d'un bon nombre de centres et laboratoires de recherche rassemblés avec les sciences et gestion de l'environnement dans l'Unité de Recherche SPHERES.

En tenant compte de la diversité des thématiques de recherche, représentées par les différents doctorats qui se déroulent au sein de différents centres et laboratoires de recherche, il ressort que les résultats des recherches sont aussi divers, avec des spécificités et orientations différentes. Cette diversité permettra d'explorer des pratiques d'usage de la littérature et de publication relativement variées.

6. Présentation des outils bibliographiques

Les outils bibliographiques multidisciplinaires couramment utilisés par les chercheurs en Géographie pour effectuer les recherches bibliographiques sont Web of Science (WoS), Scopus et Google Scholar (GS). Les bases de données bibliographiques disciplinaires les plus utilisées sont GeoRef et FRANCIS (Mikki, 2010).

WoS a longtemps été considéré comme l'unique outil multidisciplinaire pour l'accès aux références bibliographiques et ce jusqu'en 2004.

Dès 2004, des produits concurrents apparaissent : Scopus, base de données bibliographique commerciale lancée par Elsevier, et GS, version bêta, un moteur de recherche gratuit lancé en novembre par Google.

Ces outils ont tous été utilisés dans notre travail, ainsi que *Ulrichsweb* pour effectuer les relevés des informations concernant les périodiques.

Nous allons procéder à une description de chacun de ces outils, dont les principales caractéristiques sont reprises dans le Tableau 1.

Présentation des outils bibliographiques

Caractéristiques	WoS	Scopus	GS	GeoRef	FRANCIS
Année lancement	2002	2004	2004	1966	1972
Nb. périodiques	33 000 titres de périodiques (240 open access)	22 800 titres en cours (plus de 12 000 arrêtés) + 4 200 périodiques en OA	Aucune donnée fournie (en théorie toutes les ressources électroniques disponibles sur le Web)	3 500 (inclus un nombre inconnu de périodiques en OA)	2 300 périodiques
Langues	Anglais(plus 45 autres langues)	Anglais (textes dans plus de 40 langues)	Anglais (et toutes les autres langues)	Anglais (textes dans plus de 40 langues)	Français (et littérature européenne)
Discipline	Sciences et Technologies, Sciences de la Vie, Sciences sociales, Arts et Sciences Humaines	Sciences et Technologies, Sciences de la Vie, Médecine, Sciences sociales, Arts et Sciences Humaines	Toutes les disciplines	Géosciences	Sciences Humaines et Sociales
Nb. références	100 millions références & actes de conférences (7,4 millions)	60 millions références 7,2 millions actes de conférences	160-165 millions de références**	3,8 millions références (livres, cartes, rapports gouvernementaux, actes de congrès, thèses et mémoires)	2,5 millions de références
Période couverte	1900 – présent	1966 (et jusqu'à 1823)	Inconnu (en principe toutes les références disponibles en format électronique)	1669- présent (géologie pour Amérique Nord) 1933- présent(géologie pour le reste du monde)	1972 - 2015
Possibilité export vers outils de gestion	OUI(EndNote)	OUI(Mendeley)	OUI(gestion et sauvegarde dans MyLibrary)	OUI	--

Présentation des outils bibliographiques

Caractéristiques	WoS	Scopus	GS	GeoRef	FRANCIS
Fonctionnalités de recherche	Interface de recherche avancée Alerte / Enregistrer / Partager Analyse de résultats (type de documents, sujet de recherche, titres, langue) Analyse de citations	Interface de recherche avancée Alerte / Enregistrer / Partager Analyse de résultats (type de documents, sujet de recherche, titres, langue)	Interface de recherche avancée Alerte / Enregistrer / Partager Analyse de résultats (type de documents, sujet de recherche, titres, langue)	Interface de recherche avancée Alerte / Enregistrer / Partager Analyse de résultats (type de documents, sujet de recherche, titres)	Interface de recherche avancée Alerte / Enregistrer / Partager Analyse de résultats (type de documents, sujet de recherche, titres, langue)
Recherche :					
Résumés	OUI	OUI	OUI	OUI	OUI
Auteurs	OUI	OUI	OUI	OUI	OUI
Citations	OUI	OUI	OUI	NON	NON
Brevets	OUI	OUI	OUI	NON	NON
Mise à jour	Hebdomadaire	Hebdomadaire	Pas d'information officielle disponible	Hebdomadaire (100 000 références ajoutées par année)	
Développeur / propriétaire (pays)	Clarivate Analytics, Etats-Unis Précédemment Thomson Reuters	Elsevier, Pays-Bas	Google, Etats-Unis	American Geophysical Union, Etats-Unis	INIST-CNRS, France
Source	https://clarivate.com/products/web-of-science/	https://www.elsevier.com/solutions/scopus/content	https://scholar.google.com/intl/en/scholar/about.html	https://www.americangeosciences.org/georef/about-georef-database	http://www.inist.fr/?FRANCIS&lang=fr

Tableau 1 - Caractéristiques et spécificités des outils bibliographiques : bases de données multidisciplinaires WoS, Scopus ; moteur de recherche GS ; bases de données en Géosciences GeoRef et FRANCIS ** source : Orduna-Malea (2014) et Fagan (2017)

Présentation des outils bibliographiques

6.1. Web of Science

En 2002, *Thomson Reuters* a lancé une plate-forme web appelé **Web of Science (WoS)**. Cette base de données comprenait :

- SCI - Science Citation Index
- A&HCI - Art and Humanities Citation Index
- SSCI - Social Sciences Citation Index
- CPCI-SSH - Conference Proceedings Citation Index- Social Science & Humanities
- CPCI-S - Conference Proceedings Citation Index- Science

Outil pionnier de Institute for Scientific Information (ISI), Science Citation Index (SCI) mis au point par Eugene Garfield en 1955, était auparavant consultable sur papier, puis disquettes et après les années '90 sur CD-ROM.

Sur ces BD bibliographiques se greffent le JCR - *Journal Citation Reports*, qui a été le premier outil d'analyse et de classement des périodiques scientifiques depuis 1975. WoS fournit une couverture rétrospective de 1900 à nos jours et couvre 33 000 périodiques académiques, dont certains en accès libre, et plus de 150 000 actes de conférences.

Depuis octobre 2016, le propriétaire de cette base est *Clarivate Analytics*, et les bases de données spécifiques fournies ont quelque peu changé :

- SCI: 1900 – présent (8 850 périodiques dans 150 disciplines)
- SSCI: 1900 – présent (3 200 périodiques dans 55 disciplines)
- A&HCI: 1975 – présent (1 700 périodiques en arts et humanités)
- CPCI : 1990 – présent
- ESCI (Emerging Sources Citation Index) : 5 000 périodiques
- Book Citation Index : 2005 –présent (80 000 livres)

Les abonnements institutionnels à WoS varient, à la fois en termes de bases de données et de contenu. WoS Platform (une version élargie de WoS Core Collection) indexe les collections spécialisées suivantes : *BIOSIS Citation Index*, *BIOSIS Previews* (thesaurus MeSH), *Biological Abstracts*, *Zoological Record*, *MEDLINE*, *CAB Abstracts*, *CABI Global Health*, *Inspec*, *FSTA*²⁵.

Clarivate Analytics s'est associée avec plusieurs pays pour héberger ou indexer des collections ou bases de données bibliographiques régionales ou nationales : *Chinese Science Citation Database*, *Russian Science Citation Index*, *KCI Korean Journal Database*, *SciELO Citation Index*.

²⁵ <https://clarivate.libguides.com/webofscienceplatform/coverage>

6.2. Scopus

Scopus se veut la plus grande base de données de résumés et de citations de la littérature scientifique évaluée par des pairs. Cette base de données s'est très vite insérée sur le marché et est devenue compétitive. Scopus indexe un total de plus de 34 000 titres de périodiques, dont plus de 21 500 titres en cours et plus de 12 000 arrêtés, ainsi que des livres (116 000) et des actes de conférences dans les domaines des sciences et techniques, médecine, sciences sociales, arts et sciences humaines. Les termes d'indexation de Scopus proviennent de thésaurus et sont ajoutés pour améliorer les résultats des recherches bibliographiques.

Scopus a intégré également les termes d'indexation provenant de :

- *Ei Thesaurus* (Ingénierie, Technologie, Sciences physiques)
- *Emtree medical terms* (Sciences de la vie, Médecine)
- *MeSH* (Sciences de la vie, Médecine)
- *GEOBASE Subject Index* (Géologie, Géographie, Sciences de la Terre et Environnement)
- *FLX terms, WTA terms* (Sciences des fluides, Sciences du textile)
- *Regional Index* (Géologie, Géographie, Sciences de la Terre et Environnement)
- *Species Index* (Biologie, Sciences de la vie)

Comme *Clarivate Analytics* pour WoS, Elsevier a développé plusieurs outils associés à Scopus pour faciliter la recherche bibliographique et pour visualiser, suivre, analyser et évaluer les résultats de la recherche scientifique des institutions ou des auteurs.

Il est possible de faire des recherches en fonction de l'affiliation ou de l'auteur dans Scopus. « *Scopus Author Identifier* » et « *Scopus Affiliation Identifier* » identifient et quantifient automatiquement l'ensemble des résultats de recherche, et les citations, d'un auteur ou d'une organisation ou institution de recherche.

En 2017 *Plum Metrics*, racheté par Elsevier²⁶ et intégré dans Scopus, devient *PlumX Metrics* et fournit des informations sur la façon dont les utilisateurs interagissent avec les résultats de recherche (articles, actes de conférences, chapitres de livres, etc) dans l'environnement web. Concrètement, des statistiques d'usage détaillées sont enregistrées. Ces indicateurs sont répartis en 5 catégories : usage, captures, mentions, médias sociaux et citations.

Dans un effort d'améliorer la couverture de la littérature en Sciences humaines et sociales, en 2009, l'éditeur a identifié les titres relevant de cette discipline que la base Scopus n'indexait pas, à partir de *MUSE* (Etats-Unis) et *ERIH* (*European Reference Index for the Humanities*). Cette démarche a été poursuivie en 2011, avec

²⁶ <https://www.elsevier.com/about/press-releases/corporate/elsevier-acquires-leading-altmetrics-provider-plum-analytics>

Présentation des outils bibliographiques

l'identification de périodiques de « *Social Science Citation Index* », « *Arts & Humanities Citation Index* », AERES (France)²⁷, Cairn et FRANCIS.

Les titres provenant de toutes les régions géographiques sont couverts, y compris des titres en d'autres langues que l'anglais, tant que des résumés en anglais sont fournis. Environ 21% des périodiques indexés par Scopus sont publiés dans d'autres langues que l'anglais.

La couverture de Scopus a atteint près de 8 700 titres de périodiques en Sciences humaines et sociales²⁸ (SHS) (4 200 en incluant les titres connexes aux Sciences humaines). Elle comprend tous les types de publication en série : périodiques, séries de livres et séries de conférences. Plus de 55 % des titres de livres ajoutés (d'un total de 116 000) représentent les Arts et SHS, ce qui élargit considérablement la couverture de ces domaines.

Scopus est devenu très rapidement une alternative bien établie à WoS.

6.3. Google Scholar

Google Scholar (GS) peut être décrit comme un moteur de recherche web, accessible gratuitement, qui indexe la littérature scientifique éditée dans une grande variété de formats et dans la plupart des disciplines. Les informations relatives aux années exactes de couverture et aux éditeurs scientifiques partenaires ne sont pas annoncées par Google²⁹. Les informations fournies par les développeurs de GS et les analyses effectuées par les professionnels de l'information indiquent une importante couverture de la littérature scientifique, en principe illimitée.

Sur base de la reconnaissance automatique des références et des accords avec ses partenaires, GS donne accès à « une grande variété »³⁰ d'articles, thèses, livres et résumés provenant d'associations et d'éditeurs scientifiques (autant gratuits que payants), d'archives ou répertoires ouverts, des universités et d'autres sites Internet.

Les livres et les dissertations indexés dans *Google Book Search* sont automatiquement inclus dans GS, ainsi que des brevets et avis de la cour de justice des Etats-Unis tandis que les articles d'actualité, d'information ou de magazines, les critiques de livres et les éditoriaux sont exclus.

Malgré un nombre certain d'inconvénients relevés dans plusieurs études, il s'avère que l'usage de GS est important chez les chercheurs et étudiants.

²⁷ Agence d'évaluation de la recherche et de l'enseignement supérieur

²⁸ https://www.elsevier.com/_data/assets/pdf_file/0007/69451/0597-Scopus-Content-Coverage-Guide-US-LETTER-v4-HI-singles-no-ticks.pdf

²⁹ <https://scholar.google.com/intl/en/scholar/about.html>

³⁰ <https://www.google.com/intl/en/scholar/help.html#coverage>

Présentation des outils bibliographiques

Parmi les inconvénients mentionnés le plus fréquemment dans la littérature (Bar-Ilan, 2006; Bar-Ilan et al., 2007; Bonato, 2016; Fagan, 2017; Gardner & Eng, 2005; Haddaway et al., 2015; Shultz, 2007) on peut retenir :

- Longueur des équations de recherche limitée à 256 caractères
- Format variable des références et informations incorrectes concernant les auteurs, les revues et les citations (Bar-Ilan, 2006; Gardner & Eng, 2005; Jacsó, 2005)
- Impossibilité d'affichage, visualisation et exportation de plus de 1 000 résultats
- Impossibilité d'affichage de plus de 20 résultats par page
- Citations en double (générées par les multiples « versions » des articles) nécessitant un traitement manuel
- Impossibilité de téléchargement des résultats par lots (pour les charger dans un logiciel de gestion de références bibliographiques (à l'exception de Zotero, par exemple)
- Format des documents indiqué lors de l'export vers un logiciel de gestion de références bibliographiques souvent erroné, et nécessitant une vérification permanente
- Nombres de résultats différents lors des recherches effectuées avec les options Recherche avancée ou Recherche de base
- Interface de la Recherche Avancée trop simple
- Blocage automatique de l'utilisateur lors de l'export de plusieurs dizaines de citations (Haddaway et al., 2015; Ştirbu et al., 2015)³¹.
- Non-divulgation de l'algorithme de tri des résultats de recherche (Bar-Ilan et al., 2007)
- Limitation aux opérateurs booléens basiques lors des recherches
- Temps nécessaire pour le traitement des résultats de GS (exportation vers un outil de gestion de références, identification de doubles) trop important.

Nous-même avons été confrontée à ce dernier problème lors de notre travail (Ştirbu et al., 2015). Levay et al., (2016) ont estimé que le temps nécessaire au traitement des résultats de WoS était de 4 heures contre 75 heures pour GS.

Par ailleurs, une proportion de plus en plus importante, entre 40% (Martín-Martín et al., 2014) et 61 % d'articles de périodiques (Jamali & Nabavi, 2015) et même 74 % (Mikki et al., 2018) est disponible gratuitement en texte intégral. Les trois études désignent nih.gov, ResearchGate, academia.edu et Wiley comme les principaux fournisseurs de texte intégral dans GS.

En 2016, Martín-Martín et al., positionnent GS comme outil de recherche bibliographique le plus utilisé, suivi par ResearchGate.

³¹ Lors de la sauvegarde de références fournies par GS dans les relevés du premier chapitre, nous avons été bloquée également, cette affirmation est vérifique.

Présentation des outils bibliographiques

Malgré les inconvénients mentionnés, la simplicité de son utilisation, sa gratuité et la couverture à la fois des publications les plus anciennes et des plus récentes (Fagan, 2017) expliquent la popularité de GS parmi les étudiants et les scientifiques (Inger & Gardner, 2016; Van Noorden, 2014).

6.3.1. Les fonctionnalités de Google Scholar

Dans une quête d'alignement de ses fonctionnalités d'analyse avec celles des bases de données commerciales, GS a lancé à la fin de 2011 GS *Citations* (GSC)³²(Martín-Martín, Orduña-Malea, et al., 2016). Cette fonctionnalité regroupe les citations des auteurs/chercheurs qui choisissent de créer un profil dans GS. Sur base d'une revue de la littérature ayant utilisé GSC pour l'analyse de citations, Waltman (2016) indiquait une couverture de GS beaucoup plus large que WoS et Scopus dans de nombreux domaines, qui plus est, qui semblait s'améliorer avec le temps. D'autres analyses récentes (De Winter et al., 2014; Harzing, 2014; Harzing & Alakangas, 2016) confirment la compétitivité de GS avec les BD commerciales et signalent que GS fournit des informations plus complètes sur les citations que WoS ou Scopus.

Google Scholar *Metrics* (GSM)³³ (2012³⁴), quant à lui, fournit des données pour des articles et des périodiques qui semblent être fiables (Martín-Martín et al., 2014). La publication de GSM est annuelle (Fagan, 2017), comme *Journal Citation Reports*, *Journal Metrics*, and the *SCImago Journal Rank*.

L'édition 2016 contenait 5 734 publications et 12 classements des publications par langue : anglais, chinois, portugais, espagnol, allemand, français³⁵ et japonais. Cinq langues supplémentaires, le russe, le coréen, le polonais, l'ukrainien et l'indonésien ont été ajoutées en 2016, tandis que l'italien et le néerlandais ont été retirés (Martín-Martín, et al., 2016).

Les publications en anglais sont divisées en domaines de recherche et disciplines.

Parmi beaucoup de recommandations et d'améliorations, les professionnels du domaine demandent à GSM d'afficher le nombre total de documents publiés dans les publications indexées et le nombre total de citations reçues mais aussi l'ajout du pays et de la langue de publication ainsi que les taux d'autocitation (Martín-Martín, et al., 2016).

La fonctionnalité "My Library" de GS, lancée en 2013, est quant à elle en concurrence avec d'autres logiciels de gestion de références bibliographiques. Cette fonctionnalité de GS s'est avérée être la plus utilisée (47 %) par 344 étudiants (majoritairement diplômés) dans un sondage effectué par Conrad et al. en 2015. *EndNote* (37%) et *Zotero* (19%) suivaient en deuxième et troisième position.

³² <https://scholar.google.com/intl/en/scholar/citations.html>

³³ <https://scholar.google.com/intl/en/scholar/metrics.html>

³⁴ Date de lancement incertaine

³⁵ https://scholar.google.com/citations?view_op=top_venues&hl=en&vq=fr

6.4. GeoRef

Base de données bibliographique phare en Géosciences, **GeoRef** a été constituée par l'*American Geosciences Institute* (AGI) en 1966. Cette base contient 4 millions de références provenant de plus de 3500 journaux. La géologie de l'Amérique du Nord est couverte depuis 1699 et la géologie du reste du monde, depuis 1933. GeoRef fournit non seulement l'accès à des références d'articles, mais aussi de livres, de cartes, d'actes de conférence, de rapports et de thèses, y compris dans plus de 40 langues, dont une majorité en anglais. Conformément aux informations fournies par AGI³⁶, un accroissement de 100 000 références est enregistré chaque année. Toutes les publications de US Geological Survey ainsi que les mémoires et les thèses de doctorat des universités américaines et canadiennes sont couvertes.

En outre, la base de données rassemble cinq grandes sources de données pour la géologie (Clement & Ogburn, 1996) : *Bibliography of North American Geology* (1785-1970), « *Bibliography and Index of Geology Exclusive of North America* » (1933-1968), « *Geophysical Abstracts* » (1966-1971) et « *Bibliography and Index of Geology* » (1969- present) » (Clement & Ogburn, 1996).

Le *Thesaurus GeoRef*³⁷, publié par AGI, est un guide précieux des termes utilisés pour l'indexation des documents dans la base de données et peut être consulté avant de commencer une recherche.

En partenariat avec *Geoscience Australia* et *Canadian Federation of Earth Sciences* (CFES), AGI a aussi développé *AusGeoRef* en 2003 et *CanGeoRef* en 2010. Ces bases de données bibliographiques additionnelles couvrent la littérature canadienne et australienne en géosciences, respectivement depuis le début des années 1800 et 1840, et fournissent une alternative moins coûteuse à GeoRef pour les petites entreprises et les particuliers (consultants).

6.5. FRANCIS

Base de données bibliographique multilingue et pluridisciplinaire produite par l'INIST-CNRS³⁸, **FRANCIS** couvre principalement la littérature en français et la littérature européenne en SHS et contient 2,5 millions d'enregistrements. 8 % des références de cette base sont représentées par de la littérature géographique, y compris des livres, des actes de conférences, des dissertations, des articles de revues et des rapports provenant de 2 300 périodiques depuis 1972.

A partir de 2013, à la suite d'un partenariat de valorisation des données du catalogue collectif *Sudoc* conclu entre l'Abes et l'Inist-CNRS, la couverture scientifique de Francis s'est enrichie de plus de 30 000 thèses de doctorat issues du répertoire national et remontant à l'an 2000³⁹. En 2015, FRANCIS ainsi que PASCAL ont été

³⁶ <https://www.americangeosciences.org/georef/about-georef-database>

³⁷ <https://www.americangeosciences.org/georef/georef-thesaurus-lists>

³⁸ Institut de l'information scientifique et technique-Centre national de recherche scientifique

³⁹ <http://www.inist.fr/?FRANCIS&lang=fr>

Présentation des outils bibliographiques

mises en accès libre, un site internet⁴⁰ proposant l'accès aux archives de ces deux bases produites par l'Inist-CNRS entre 1972 et 2015 et jusqu'alors accessibles par abonnement. Trois modes de recherche sont proposés : simple, avancée et experte. Le contenu de FRANCIS et de PASCAL peut également être exploré par mots-clés (vocabulaire) ou par thématiques (classification). Cependant depuis 2015, les mises à jour ne sont plus effectuées.

6.6. Ulrichsweb Global Serials Directory

Autre type de base de données, **Ulrichsweb Global Serials Directory** est considérée comme une des sources principales d'informations bibliographiques relatives à plus de 300 000 périodiques de tous types, publiés dans le monde. L'étendue d'*Ulrichsweb* est très importante, cet outil couvre plus de 900 domaines. Les notices d'*Ulrich's* fournissent des renseignements sur les publications périodiques tels que l'ISSN, l'éditeur, la langue, le sujet, l'indexation et l'accès via les bases de données, la table des matières et des comptes rendus rédigés par les bibliothécaires.

Ulrichsweb attribue également la mention “*peer-reviewed*” à certains périodiques et nous avons retrouvé cette définition sur la page web FAQ d'*Ulrichsweb*: “*As used in the Ulrich's knowledgebase, the term refereed is applied to a journal that has been peer-reviewed. Refereed serials include articles that have been reviewed by experts and respected researchers in specific fields of study including the sciences, technology, the social sciences, and arts and humanities.*” (SerialSolutions., 2016)

Hicks & Wang (2011) avaient trouvé la description suivante quant à l'attribution de la mention peer-reviewed: ”*The Ulrich's editorial team assigns the “refereed” status to a journal that is designated by its publisher as a refereed or peer-reviewed journal. Often, this designation comes to us in electronic data feeds from publishers. In other cases Ulrich's editors phone publishers directly for this information, or research the journal's information posted on the publisher's website. (SerialSolutions, 2010)*”

⁴⁰ <http://pascal-francis.inist.fr/home/>

7. La structure de la dissertation et le contenu des chapitres

Notre dissertation doctorale s'articule donc autour de deux volets principaux de la « communication » scientifique dans le domaine spécifique de la Géographie : d'une part, la recherche d'information, et d'autre part, la diffusion de l'information scientifique. Différentes approches méthodologiques ont été adoptées et les résultats sont présentés en deux parties et cinq chapitres.

Dans la première partie de notre travail nous nous penchons sur le processus qui constitue le point de départ de toute recherche : la recherche de la littérature existante sur un sujet précis, dans les BD traditionnelles multi- et mono- disciplinaires et dans GS. Les spécificités de la littérature fournie et répertoriée par les outils bibliographiques analysés ainsi que leurs apports sont quantifiés et caractérisés dans trois chapitres.

Dans le **premier chapitre** les clés d'accès à l'information, les bases de données bibliographiques traditionnelles (WoS, FRANCIS, GeoRef) et le moteur de recherche GS, ont été comparés quant à leur couverture pour dix mots-clés en Géographie physique et humaine. D'un point de vue méthodologique, nous avons effectué des recherches par mots-clés dans le titre des documents, pour une période de publication limitée (2005 à 2009). Ces recherches ont été répétées à intervalles réguliers afin de s'assurer d'une relative stabilité des résultats, et donc de la couverture des outils bibliographiques, dans le temps.

Le nombre de résultats obtenus a été analysé à l'aide du logiciel R, ce qui constitue en soi une approche originale en la matière, afin de rendre compte des variations liées, entre autres, à l'outil bibliographique interrogé.

Par ailleurs, les résultats obtenus avec deux des dix mots-clés, représentant respectivement la Géographie physique et la Géographie humaine, ont fait l'objet d'une analyse plus détaillée afin de déterminer le recouvrement (*overlap*) entre les outils bibliographiques et leur apport respectif en termes de diversité de types de documents.

Nous escomptons que l'approche adoptée, d'une recherche par mots-clés, favorisera la réappropriation des conclusions par les chercheurs en Géographie, pour une application adéquate à leurs propres recherches bibliographiques.

Ce chapitre a fait l'objet d'une publication sous forme d'article dans *Journal of Academic Librarianship* en mars 2015 (Ştirbu et al., 2015).

Le deuxième chapitre repart du détail des résultats correspondant à deux des dix mots-clés, présentés dans le premier chapitre de la thèse. Ici l'attention se concentre sur les périodiques dans lesquels ont été publiés les articles spécifiquement retrouvés par une des bases de données commerciales ou par GS.

En pratique, les informations relatives à ces périodiques ont été recherchées dans *Ulrichsweb*. Les caractéristiques de ces périodiques : *peer-reviewed, Open Access*,

La structure de la dissertation et le contenu des chapitres

recensement dans JCR, pays de publication, langue du texte ont contribué à l'identification des périodiques régionaux ou nationaux.

L'apport spécifique des différents outils bibliographiques étudiés, en termes d'articles de périodiques, a pu être qualifié pour la première fois en Géographie.

Dans le **troisième chapitre** nous nous sommes penchés sur les bibliographies de trois thèses présentées par des doctorants géographes de l'ULiège durant l'année académique 2014 – 2015. Outre que ces bibliographies sont issues de recherches récentes, elles présentent la garantie d'avoir été établies librement, sans avoir à se plier aux contraintes d'un éditeur commercial.

Pour avoir une vision d'ensemble de la littérature utilisée, compte tenu des orientations sous-disciplinaires des trois thèses, les types de documents utilisés et les années de publication ont été relevés, examinés et comparés.

Par ailleurs, le recensement des trois bibliographies par différents outils bibliographiques a été étudié. L'« overlap » entre les outils bibliographiques ainsi que leurs apports spécifiques ont été représentés à l'aide de diagrammes Venn. L'efficacité avec laquelle des BDs bibliographiques commerciales, d'une part, et GS, d'autre part, répertorient actuellement ces bibliographies a ainsi pu être déterminée.

Il s'agit aussi ici de mettre en perspective un échantillon de la littérature utilisée par les chercheurs avec les conclusions auxquelles nous amènent les deux chapitres précédents, quant à l'intérêt respectif des différents outils bibliographiques.

La deuxième partie de la thèse est axée sur la visibilité et la diffusion des travaux scientifiques publiés par les géographes de l'ULiège, à partir d'un échantillon de 453 articles et communications orales publiés entre 2000 et 2014 et déposés dans ORBi. Pour l'analyse de cet échantillon, nous disposons de données spécifiques supplémentaires, soit diverses données statistiques extraites d'ORBi.

Dans le **premier chapitre**, une méthodologie comparable à celle utilisée dans la première partie de la thèse, nous a permis d'évaluer la contribution du répertoire institutionnel ORBi, dans la visibilité de cette littérature.

Le recensement des 453 articles et communications de l'échantillon issu d'ORBi a été vérifié dans quatre outils bibliographiques : le moteur de recherche GS, qui moissonne ORBi, et les bases de données bibliographiques commerciales Scopus, WoS et GeoRef. Les recouvrements entre les outils bibliographiques et les références « uniques » indexées spécifiquement par chaque outil ont été identifiés. Ainsi nous avons pu estimer l'efficacité de quatre outils quant à l'indexation des articles publiés par les géographes de Liège. Et surtout, en distinguant la littérature provenant d'ORBi dans les références « uniques » de GS, nous avons pu estimer la contribution du répertoire institutionnel dans la visibilité de ces publications.

Les critères de qualité des périodiques dans lesquels les articles ont été publiés ont été relevés dans ORBi et dans *Ulrichsweb* : *peer-reviewing*, nature académique, recensement dans JCR, *Open Access* ou non, langue, pays de publication... Ces

La structure de la dissertation et le contenu des chapitres

caractéristiques ont été examinées en relation avec la couverture des références par les différents outils bibliographiques et l'apport d'ORBi.

Dans le **deuxième chapitre**, sur base de différentes données « statistiques » extraites d'ORBi, la répartition mondiale des téléchargements a été examinée à l'aide d'un support cartographique. Nous en avons tiré des premiers éléments d'analyse « qualitative ».

Puis, par une approche avec les statistiques multivariées, nous avons tenté d'identifier et de quantifier les différents facteurs socio-économiques susceptibles d'influencer les téléchargements des références à partir d'ORBi. Dans ce chapitre, nous avons développé un premier modèle statistique explicatif des téléchargements relatifs aux 453 articles et communications publiés de notre échantillon.

La deuxième partie de notre thèse nous permet de conclure sur l'apport d'ORBi, dont le rôle peut être crucial pour la visibilité et la diffusion de certaines publications auprès d'un certain public.

Dans nos **conclusions générales**, sur base des pratiques de production / consommation ou d'usage / diffusion de l'information scientifique par les géographes, telles que révélées par nos résultats de recherche et confrontées à la littérature existante, nous tentons de dégager les éléments utiles aux chercheurs pour une approche plus consciente de la recherche documentaire et de la communication des résultats scientifiques.

PREMIERE PARTIE

**La littérature répertoriée par les bases de données commerciales et
par le moteur de recherche Google Scholar en Géographie**

Chapitre 1 The Utility of Google Scholar When Searching Geographical Literature: Comparison with Three Commercial Bibliographic Databases

1. Introduction

Analysis of bibliographic tools (BTs) is a growing academic research field nowadays, particularly comparison between traditional commercial databases (DBs) and free web search engines, such as the well-known Google Scholar (GS), launched in 2004.

Studies have employed several search criteria to evaluate BT performance, including citations (Bar-Ilan 2006; Mingers & Lipitakis 2010; Bauer & Bakkalbasi 2005), keywords (Falagas et al., 2008; Gardner & Eng, 2005; Notess, 2005; Robinson & Wusteman, 2007), and author names (Mikki 2010; Bar-Ilan et al., 2007). The elements compared in search results have included citation display, bibliographic tool coverage, result overlap, and h-index values (Bar-Ilan, 2008; Neuhaus et al., 2006; Walters, 2007, 2011).

The fact that GS provides a large number of results has already been discussed in relation to several scientific domains, including medicine, biology, life and environmental sciences, and engineering (Jones, 2005; Löhönen et al., 2010; Meier & Conkling, 2008; Neuhaus et al., 2006). However, at the present time, few existing analyses relate specifically to physical geography, and practically none to human geography. Thus, *GS's coverage of geography is still to be ascertained*.

Across fields, early studies show that despite a high number of results obtained, GS has not performed optimally with regard to update frequency, precision of information, or accuracy of the number of citations (Gardner and Eng 2005; Notess 2005). However, later studies rank GS among the top bibliographic tools in terms of precision, relative recall, and ability to retrieve “Top Ranked Pages” (Robinson and Wusteman 2007; Walters 2007; 2011). Meier and Conkling (2008) and (Cusker, 2013) state that GS is a good option for engineers, and Mikki (2010) argues that it should be included in library programs for information literacy instruction. Further, (Chen, 2010a) shows that GS has become an important tool in bibliographic research and emphasizes the declining value of commercial DBs.

Overall, the evaluation of GS appears to have improved with time. The progressive acceptance of “Google Search Mode” for bibliographic search by the scholarly world could in part explain this progression, as could the gradual improvement of GS's coverage. In fact, in 2010, Chen (2010b) repeated the earlier experimentation of Neuhaus et al., (2006) on seven selected DBs and indicated a “dramatic improvement” of GS's coverage since the earlier study. Since GS collaborates with several scientific editors, and their contract length is not known, GS's coverage

Aim and Objectives

could be much more variable over time than that of traditional DBs. *This condition has to be considered when using GS.*

Going farther into the matter of the coverage of GS as opposed to other bibliographic tools, the previously mentioned Neuhaus et al., (2006) compared a set of 47 bibliographic DBs with GS and discovered that each of them was partially or fully covered by GS (between 6% and 100%). Upload frequency and publication date and publication language bias were checked. “Mean scores of Google Scholar database coverage for all databases assigned to a particular discipline category were calculated” Neuhaus et al., (2006). The mean discipline category scores are high for science and medicine (76%), low for the social sciences and education (39% and 41% respectively), and very low for humanities (10%). The mean score for the multidisciplinary DBs category is 77%, while for discipline-specific DBs, GS’s coverage decreased dramatically to only 26% of GeoRef. Chen (2010b) who repeated this experiment, didn’t consider GeoRef; however, (Musser, 2008) found that the overlap between GS and GeoRef was 55%.

Another study targeting GeoRef was performed by Brown (2009). This study primarily compared GeoRef and GS content with that of randomly selected bibliographies from graduate student theses and dissertations. For searches of the geosciences literature, (Brown, 2009) recommended the use of these tools in the following order: Scopus and/or Web of Science, Google Scholar, GeoRef, WorldCat, and Google Books. Mikki (2010) compared WoS with GS for earth sciences (specifically, she searched the papers of 29 authors working on climate and petroleum geology issues) and found a high similarity of values for the two sources. Recall of WoS documents by GS was 85%, and “the number of search results and their citations was higher in GS.”

In this study, through an investigation of how GS performs in the field of geography, we intended to make some important additions to the existing knowledge on GS. We also wish to determine how the tool’s performance in a given field evolves over time: to our knowledge, this is the first study to tackle bibliographic search tools evolution over time.

2. Aim and Objectives

The aim of this study was to assess the current performance of the GS search engine in the subfields of physical and human geography. To do so, we relied on the search tools currently used in the field as comparison cases: the discipline-specific GeoRef for the scientific discipline of physical geography, and the American and European multidisciplinary databases Web of Science (WoS) and FRANCIS. FRANCIS suits particularly for the social science of human geography although it covers some papers in physical geography too.

Among several established methods, we chose to perform the study by successively searching ten selected keywords (search criteria) using each of the four bibliographic tools. Despite the laborious nature of the processing of results, this approach has the advantage of familiarity with the scholarly researcher's usual search process for bibliographic references. The searches covered several publication years (2005–2009) and were repeated several times to ensure quality of results.

We compared the GS results to those of the commercial DBs and, to allow a more nuanced interpretation, we compared traditional DB results among themselves too.

For each of the ten keywords, we first conducted statistical analysis of the results in order to identify significant variations and to allow the derivation of interpretations. For two of the ten keywords, we then conducted a more detailed comparison of results, eliminating the noise, making allowances for overlap and unique references retrieved by each of the four tools, and determining the different types of unique documents.

Our objective was to answer the following questions:

- 1) In the field of Geography, does GS give many more hits than the commercial bibliographic DBs GeoRef, WoS, and FRANCIS?
- 2) Are GS results very noisy because of duplicates or non-geographical references?
- 3) To what extent do GS results overlap traditional DB results?
- 4) Are the references found uniquely in GS to be considered reliable sources?
- 5) Are results and their implications different for the four bibliographic tools across human and physical geography subfields?
- 6) Are the results reproducible when the search process is repeated over time?

3. Methods

3.1. Bibliographic tools

Google Scholar is a freely accessible web search engine that indexes the full text of scholarly literature across a wide variety of publishing formats and disciplines. The exact years of coverage are not provided by *Google*, since they are in principle unlimited. Based on automatic reference recognition and through agreements with its partners, it provides access to articles, theses, books and abstracts from academic publishers, professional societies, online repositories, universities, and other websites.

Web of Science is a multidisciplinary citation database from *Thomson Reuters* that provides current and retrospective coverage in the sciences from 1900 to the present day. It covers 12,000 academic journals, including open access journals, and over 150,000 conference proceedings.

Methods

GeoRef is a DB specializing in the geosciences. Established by the *American Geosciences Institute* (AGI) in 1966, it contains 3.4 million references from 3500 journals (relating to the geology of North America since 1699 and the rest of the world since 1933). It provides access to journal articles, books, maps, conference papers, reports, and theses, including those in languages other than English, with 80,000 references added annually.

FRANCIS is a multilingual and multidisciplinary DB produced by *INIST–CNRS* (*Institut de l'information scientifique et technique–Centre national de la recherche scientifique*/*Institute of Scientific and Technical Information–National Centre for Scientific Research*), France, which mainly covers French and European literature in the humanities and the social sciences and contains 2.5 million records. A total of 8% of the literature in this DB is geographical literature, including books, conference proceedings, dissertations, journal articles, and reports from 2300 publications dating back to 1972.

The homepages of WoS,⁴¹ FRANCIS,⁴² and GeoRef⁴³ were used to perform the bibliographic searches. For GS, we used the advanced research interface on the *googlescholar.com* webpage.

3.2. Bibliographic search

To decide which keywords to consider, we carried out some tests using simple keywords and also multiple-word (composite) search terms in all the mentioned bibliographic search tools. In all cases, GS provided the highest number of results.

Composite keywords, being too specific, yielded low numbers of results that would not allow a comparative analysis. Therefore, we discarded them as inappropriate.

We selected ten simple keywords corresponding to broad concepts generally studied in geographical scholarship. These keywords are associated with scientific topics in geography, and most of them are currently used in different Departments at the University of Liège (namely, the departments of Geography and ArGENCo—Architecture, Géologie, Environnement & Constructions/Architecture, Geology, Environment & Construction) and around the world. During our tests, we noticed that the keywords we chose are often mentioned in the keyword lists of scientific research papers published in earth science and geography journals (e.g., *Earth Surface Processes and Landforms*, *Catena*, *Geografiska Annaler Series A: Physical Geography*, *Geomorphology*, *Applied Geography*, *Area*, *Urban Studies*, *Geoforum*).

41 http://apps.webofknowledge.com/WOS_GeneralSearch.

42 <http://search.proquest.com/francis/advanced>.

43 [http://web.ebscohost.com/ehost/search/advanced\(georef\)](http://web.ebscohost.com/ehost/search/advanced(georef)).

The keywords finally selected were:

- *urbanization, immigration, transportation, tourism, gentrification* (for human geography)
- *sedimentation, erosion, earthquake, tsunami, flood* (for physical geography)

To obtain comparable results from each of the bibliographic tools, it was essential to execute searches that were strictly the same for each tool. The keywords were placed between quotation marks, so that automatic variations of the searched keyword, imposed by the operating systems of the bibliographic tools, were excluded (e.g., sediment, sedimentary, sedimentous). The keywords were searched exclusively in the “title” field of the search interfaces, so that results would not be distorted by the indexing characteristic of each BT. The bibliographic searches were carried out for the years 2005 to 2009: results are efficiently limited by publication year.

In order to find out more about the number of repeatable results, the identical overall process was carried out eight times, monthly, from November 2010 to July 2011 (one observation per month, with a single observation for May–June).

For the treatment of results, we used the reference manager software *Zotero* and *EndNote* as well as the spreadsheet application *Excel*.

3.3. Analysis of the number of results

The number of hits are represented statistically as the response variable, allowing the observation of any variation according to search characteristics (namely, selected keyword or set of keywords representing either physical geography or human geography, year of publication, database) corresponding statistically to the explanatory variables. In addition, studying the possible time dependence between the repeated counts was a second important objective of this analysis.

In order to gain some insight regarding dependences or interactions, so-called interaction plots have been created using R software. These plots display conditional averages of the number of results for all possible interactions between two explanatory variables (i.e., the variables database and selected keywords yield 4×10 interaction terms). Plotting these conditional averages in an organized way (that is, by using line segments joining them) enabled us to visualize any constant behavior of the response variable across the categories of explanatory variables or, on the contrary, any dependence between them. This is an exploratory tool, often advocated as a preliminary step before the definition of a model.

3.4. Reference analysis—a case study

As several of the bibliographic tools analyzed are multidisciplinary, results related to scientific fields other than geography are likely to be yielded when performing

Methods

searches with the selected keywords. This is a major disadvantage of using keywords as search criteria. It is well known that GS displays a lot of duplicate references; as our analysis required correct lists of unique references exclusively in the field of geography, we had to manually select the results.

Because this selection process took a long time, we worked with two keywords only for this phase: “urbanization” for human geography and “sedimentation” for physical geography. In fact, the other keywords could not be considered anyway, since they provided more than 1000 results per year in GS, which displays only the first one thousand (with the exception of “gentrification,” which yielded too low a number of results for the analysis).

This specific task, requiring both geographical and librarian skills, was completed by the first author, a geographer working as a geosciences librarian. To select for geographical literature across all the bibliographic tools, including WoS, she relied to the “Research Areas” classification of WoS. The reference manager *EndNote* was used for selection of the geographical references and also to remove duplicates for GS.

For the two selected keywords (“urbanization” and “sedimentation”), the percentages of references from each BT that were geographical references were calculated.

3.5. Overlap and specificity

Traditional overlap (TO) and *relative overlap* (RO) were calculated for each pair of BTs, on the basis of the sets of geographical references that were retrieved for the two selected keywords over the five-year target period (2005–2009) and manually selected. *Traditional overlap* (TO) is defined as follows (Bar-Ilan et al., 2007; Gluck, 1990; Hood & Wilson, 2003):

$$\% \text{ TO} = 100 \times \frac{|\text{PUBL A} \cap \text{PUBL B}|}{|\text{PUBL A} \cup \text{PUBL B}|}$$

where PUBL X is the set of publications retrieved from database X (Bar-Ilan et al. 2007)

Relative overlap (RO) is defined as:

$$\% \text{ RO in A} = 100 \times \frac{|\text{PUBL A} \cap \text{PUBL B}|}{\text{PUBL A}} \quad \text{and}$$

$$\% \text{ RO in B} = 100 \times \frac{|\text{PUBL A} \cap \text{PUBL B}|}{\text{PUBL B}}$$

Both TO and RO are 100% if the sets of publications from DBs A and B are equal; if they are completely different, then both TO and RO are equal to 0%. RO is a

more developed measure, expressing the two-way relative overlap of the secondary sources.

By using *EndNote*, we could identify specific references for each of the search tools used. We also determined the corresponding document types (journal articles, theses, reports, conference proceedings, unpublished works, patents, maps, and books or book chapters).

4. Results

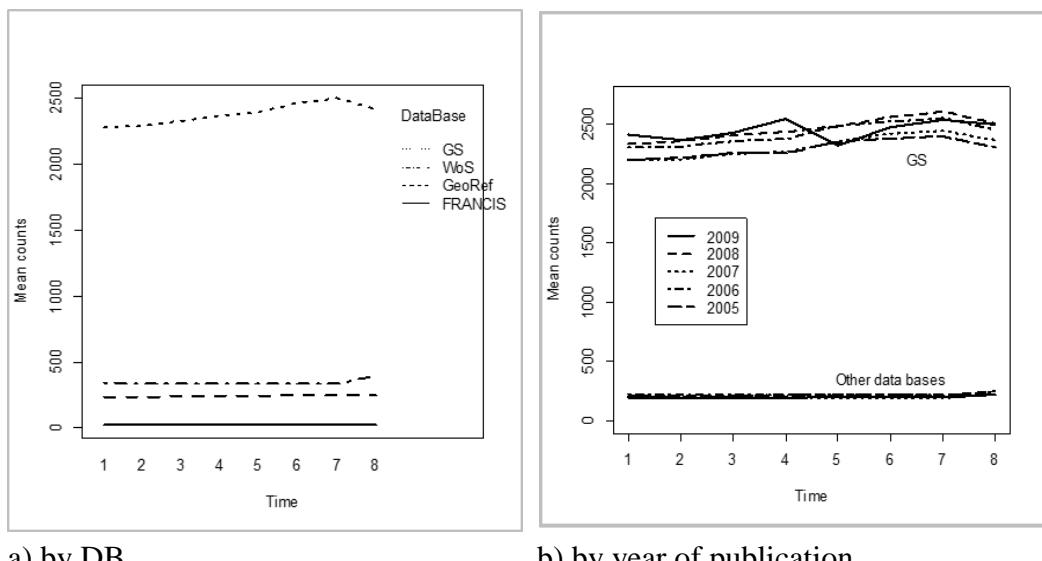
4.1. Analysis of the number of results

4.1.1. Time evolution of search results

In order to determine whether the search results were repeatable or not, interaction plots were drawn up; they are represented in Figure 1. The left panel (Figure 1a) plots the average number of results conditioned by the dates of the search (classified as time categories 1 to 8 on the horizontal axis) and by the different BT.

The right panel (Figure 1b) illustrates the average number of results using the dates and the reference years as conditioning variables, with the additional effect of separating the results obtained from GS on one side from those for the other databases on the other side. One can see that the evolution, for all databases, is relatively constant over time; but one can nevertheless notice that except on the last search, the number of references retrieved by GS increases constantly over time (see Figure 1a).

When decomposing GS results by reference year, the averages computed for 2009 show a more variable pattern than the other years (see Figure 1b), but the number of results remains in the same size order.



a) by DB

b) by year of publication

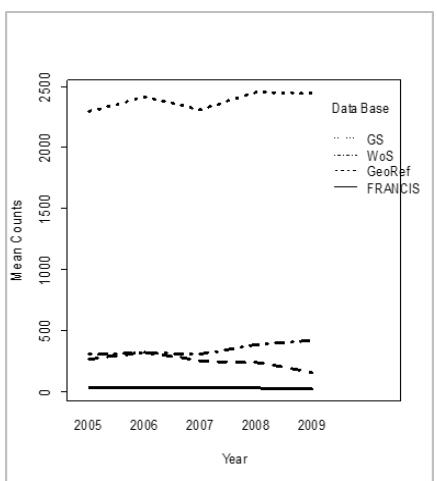
Figure 1 - Plots showing average results over different search times corresponding to different categories of a second variable

Results

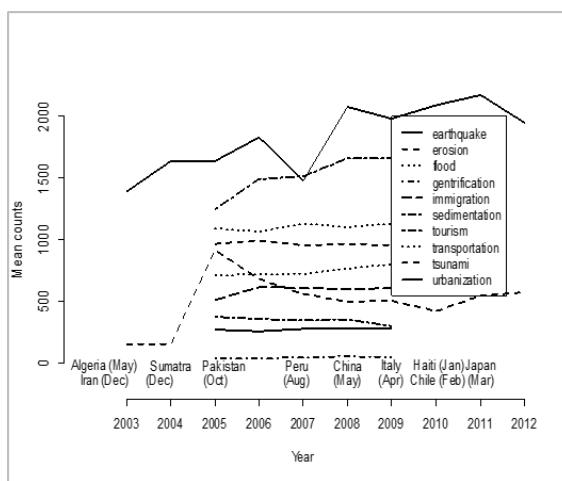
We may therefore conclude that there is little variation in the number of bibliographic references retrieved by the bibliographic tools when the same search is repeated over time. The variability observed in the results obtained by GS illustrates that this tool seems to be more volatile than the others even though the variation is still rather limited when compared to the total number of hits.

4.1.2. Performance of the search tools

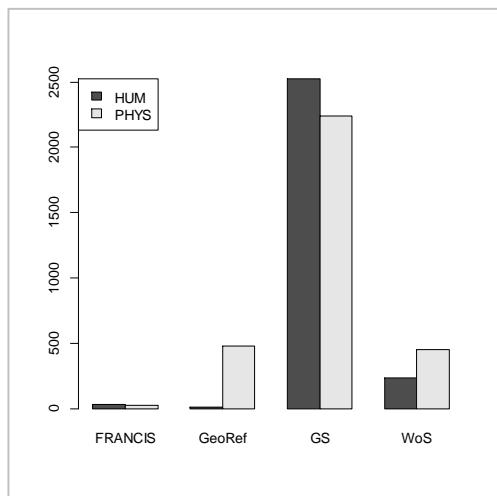
The two figures below (2a and 2b) show the performance of the bibliographic tools over time.



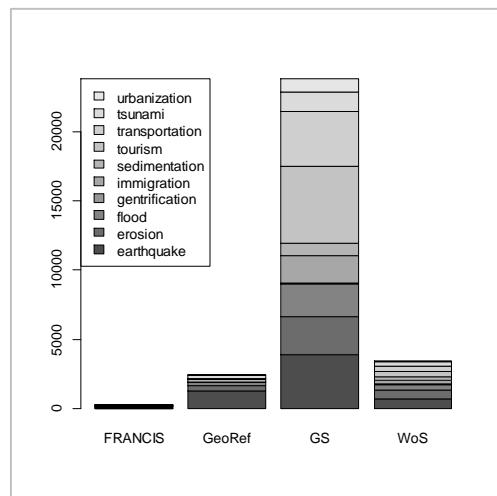
a) Conditioning variables: publication year and database



b) Conditioning variables: publication year and keyword



c) Conditioning variables: database and subfield



d) Conditioning variables: database and keyword

Figure 2 - Plots showing conditional averages of the results for two explanatory variables

Results

Figure 3a plots the average number of hits per year by publication and by bibliographic tool. It clearly shows that GS yields many more results than the other databases. This can also be observed in Figures 3c and 3d, where the averages conditioned not only on the bibliographic tools but also on, respectively, the keywords (3c) and the subfield (physical or human geography, 3d) are represented by stacked or juxtaposed bars.

Compared to GS, the number of references retrieved by the traditional databases is much lower: while GS reached averages between 2000 and 2500 in Figure 1a, these drop to a few hundred for WoS and GeoRef; FRANCIS provides the lowest number of results on average, around a few dozen. This was expected, as it is known that geographical literature only represents 8% of the references included in FRANCIS. However, overall, the huge difference between the numbers of results provided by GS and by the other databases makes us suspect a certain quantity of noise in the former—the nature of which will be explored further over the course of this study.

Figure 3b illustrates changes in the average numbers of hits across keywords over the years of publication. Most keywords remained relatively constant, but three of them, “earthquake,” “tsunami,” and “tourism,” show divergent patterns: erratic spikes and valleys in the cases of the former two, and a continuous rise for the last. To get more insight on these particular keywords, the numbers of results for the keywords “earthquake” and “tsunami” was counted again for each of the four bibliographic tools over an enlarged period of publication (2003–2012). The corresponding averages are added to the interaction plot in Figure 3b. It appears that the variations are directly linked to important events—tsunamis and earthquakes—that occurred at corresponding time. The places where these events occurred are also indicated on the graph. For instance, as can be seen in graph 3b, the December 26, 2004, earthquake that provoked the tsunami in the Indian Ocean had a direct impact on the publications referring to tsunamis. The rise in the keyword “tourism” is probably due to the increasing tourist activity, and the scientific publications studying it, at a global level in the period under examination.

The significant number of results yielded by GS for the most popular keywords, “tourism,” “transportation,” and “earthquake” (Figure 3c), largely explains the overall leading position of GS in comparison to the other databases as well as its especially high performance in the subfield of human geography (Figure 3d). Proportionally to the results, at this stage of the analysis, one might conclude that GS seems to be more efficient in human geography while WoS performs better in physical geography and GeoRef and FRANCIS yield results proportionally in accordance with their main subfield orientations, respectively physical geography and human geography. Nevertheless, to assess the real performance of these search tools in the mentioned subfields, a more thorough analysis is needed; this matter will be further considered in the second part of the article.

Results

In summary, the plots represented in Figure 2 tend to show that the average number of results depends on the chosen bibliographic tool and keyword as well as the subfield.

4.1.3. Case studies: “urbanization” and “sedimentation”

Treatment to compensate for noise

For the total five years results sample (2005–2009), Table 2 shows the initial number of references (No. of Refs.) and the proportions of them accounted for by the selected geographical references (Geogr. Ref.).

The initial number of results returned by GS was the highest, with 4791 references for “urbanization” and 4123 references for “sedimentation.” However, only 57% of the GS results for “urbanization” refer to the geographical literature, and the percentage for “sedimentation” is slightly lower at 45%. The initial results provided by WoS were distinctly lower, with 585 references for “urbanization” and 1445 for “sedimentation,” of which only 47% for “urbanization” and 42% for “sedimentation” were for geographical literature.

“urbanization”		“sedimentation”			
	No. of Refs.		No. of Refs.		
	Geogr. Ref.		Geogr. Ref.		
2005– 2009		2005– 2009			
GS	4791	2732 (57%)	GS	4123	1855 (45%)
WoS	585	275 (47%)	WoS	1445	606 (42%)
GeoRef	99	97 (98%)	GeoRef	1307	1265 (97%)
FRANCIS	64	45 (70%)	FRANCIS	39	33 (85%)

Tableau 2 - Overall and geographical literature for the keywords “urbanization” and “sedimentation”

Although in third position for overall number of results, with 1307 results for “sedimentation” and 99 for “urbanization,” the specialized DB GeoRef quite logically provides the highest percentage of geographical literature: almost 100% (98% for “urbanization” and 97% for “sedimentation”). FRANCIS also provides a good percentage of geographical literature—70% for “urbanization” and 85% for “sedimentation”—but a very low raw number of results.

Thus, with roughly comparable noise rates for the two main multidisciplinary tools, we determine that GS ultimately offers a higher number of potentially relevant results than WoS (and the other databases) for both keywords, especially for “urbanization.” Further, after geographical reference selection, GeoRef outnumbers WoS for “sedimentation” with more than double the number of geographical references, whereas WoS leads GeoRef for “urbanization” results. FRANCIS is ranked far behind for both.

Overlap and specificity

For both selected keywords, “urbanization” and “sedimentation,” the low values for *traditional overlap* (TO) indicate a high difference between the retrieved geographical references over five years (2005–2009) for each pair of bibliographic tools (Table 3). “Sedimentation” percentages show a slightly reduced distance between WoS and GS (25%), GeoRef and GS (33.4%), and WoS and GeoRef (27.5%). However, the distance between FRANCIS and the other BTs is very large, particularly for GS, with very low values for both keywords: 1.1% for “urbanization” and 1.3% for “sedimentation.”

%	WoS - GS	GeoRef - GS	FRANCIS - GS	WoS - GeoRef	FRANCIS - GeoRef	FRANCIS - WoS		
'05-'09	urbaniz.	sedim.	urbaniz.	sedim.	urbaniz.	sedim.	urbaniz.	sedim.
TO	6.5	25	2	33.4	1.1	1.3	7.2	27.5
RO in A	66.5	82	56.7	62	68.8	75.8	9	66.7
RO in B	6.7	26.8	2	42	1.1	1.3	25.8	32
							1.4	2
							2.1	2
							6.6	7.3
							3.7	3.8

Tableau 3 - Results of the bibliographic tool overlap analysis (TO, RO in A, and RO in B— %)

With regard to **GS**, the percentage of ***Relative overlap (RO) in A*** indicates that GS overlaps WoS and FRANCIS very well for the keyword “sedimentation,” at respectively 82% and 75.8% over the study period. For “urbanization,” GS also overlaps FRANCIS and WoS well, at respectively 68.8% and 66.5%. To a lesser extent, GeoRef is relatively well overlapped by GS for both keywords, with 62% for “sedimentation” and 56.7% for “urbanization.”

In contrast, the percentage of ***Relative overlap (RO) in B*** shows that the traditional DBs overlap very few references from GS, except for the keyword “sedimentation,” which is moderately well overlapped by GeoRef (42%) and poorly by WoS (26.8%). The very small values for RO in B are partly due to the very different number of results provided by the bibliographic tools.

Regarding **GeoRef**, RO in A shows that for “sedimentation,” this DB overlaps FRANCIS very well, with 75.7% and WoS well, with 66.7%, but that there is very little overlap for “urbanization,” with 5% for FRANCIS and 9% for WoS. Besides, RO in B values show that GeoRef is only poorly recalled by WoS, with 32% for “sedimentation” and 25.8% for “urbanization.”

FRANCIS is well overlapped by WoS for “sedimentation,” with RO in A of 69.7%, and moderately for “urbanization,” with RO in A of 44.4%. As expected, however, FRANCIS overlaps very few references from GeoRef and WoS, with RO in B ranging from 2% to 7%. It is worth mentioning as well that FRANCIS is 100% overlapped by WoS for 2008 and 2009, GeoRef for 2008, and GS for 2006 (yearly results not shown).

Results

In short, taking into account our findings, we can say that GS is a bibliographic tool that performs well at overlapping retrieved references from traditional DBs, at least for the two selected keywords and perhaps also for the others, while the traditional DBs overlap only a moderate number of documents provided by GS. GeoRef overlaps FRANCIS and WoS very well for “sedimentation,” whereas WoS covers about one-quarter of the documents provided by GeoRef for both keywords and by GS for “sedimentation.”

The specific results by search tool are the corollary of the overlaps (Figure 3 and Table 3).

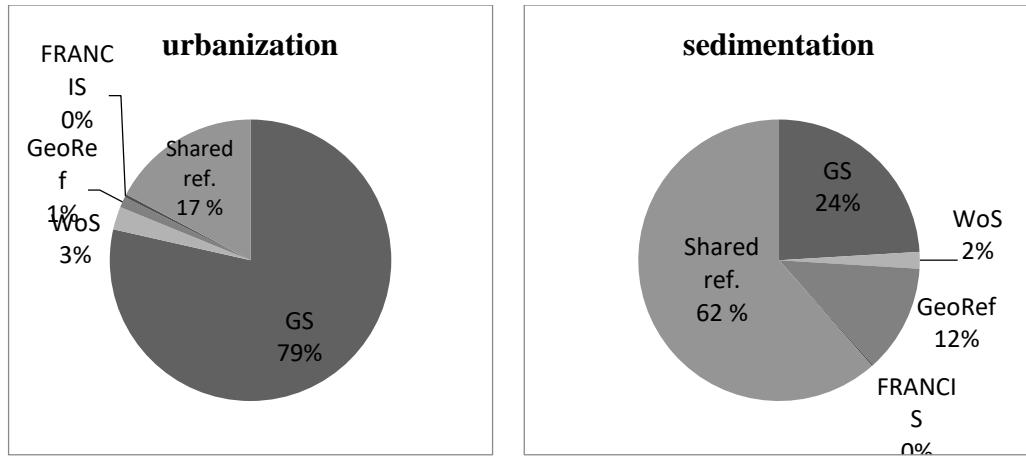


Figure 3 - Unique references provided by each bibliographic tool

In all, 79% (2473) of the references for “urbanization” are unique to GS, and 24% (904) for “sedimentation.” For the latter, a significant number of unique items are also found in GeoRef: 466 references, or 12%.

The percentage of references shared between at least two bibliographic tools is 17% for “urbanization” and 62% for “sedimentation.” This important difference can be interpreted as testimony to GS’s efficiency in retrieving additional unique references in human geography.

2005–2009	Unique Ref/DB	Journal Articles	Conf. Proc.	Theses	Reports	Unpubl. Work	Patents	Maps	Book/Book Chapter
urbanization									
GS	2473	1908	249	99	38	5	--	--	174
WoS	83	83	--	--	--	--	--	--	--
GeoRef	39	19	11	1	6	--	--	--	2
FRANCIS	11	11	--	--	--	--	--	--	--
sedimentation									
GS	904	393	282	107	47	--	5	--	70
WoS	72	71	1	--	--	--	--	--	--
GeoRef	466	243	147	30	26	--	--	2	18
FRANCIS	6	6	--	--	--	--	--	--	--

Tableau 4 - Unique references and associated types of literature for each bibliographic tool

Discussion and Conclusion

Table 4 details the number of unique items provided by each BT and the corresponding types of literature. It appears that grey literature (conference proceedings, theses, reports, and to a lesser extent patents, unpublished works, and maps) and books or book chapters are only returned by GS and GeoRef. Given the significant number of conference proceedings covered by WoS (150,000), we may be surprised at finding a low number of this type of document in WoS.

The percentages of journal articles retrieved by GS are roughly equivalent to those from GeoRef: for “urbanization,” 77% of unique items in GS and 46% in GeoRef are journal articles, while for “sedimentation,” 43% of unique items in GS and 52% in GeoRef are journal articles.

The percentage of references accounted for by grey literature for “urbanization” is low in GS (16%) and equivalent to that for journal articles in GeoRef (46%); and for “sedimentation,” 49% in GS and 44% in GeoRef.

Books or book chapters represent only a few percent of the total, in all cases: for “urbanization,” 7% in GS and 5% in GeoRef, and for “sedimentation,” 8% in GS and 4% in GeoRef.

5. Discussion and Conclusion

With reference to the initial questions asked in this study, our results show that:

- 1) The statistical approach applied to the data consolidates our hypothesis that GS leads the other tools widely on number of results, independently of keyword, subfield, year of publication, or time of search. In contrast, FRANCIS yielded a very low number of results, classifying it as the least useful of the tools in both human and physical geography. The interest of this database is possibly confined to more local and/or non-English literature.
- 2) From the case study (of “urbanization” and “sedimentation”), it appears that GS remains on top even after removing duplicates and selecting geographical references (noise treatment). It is worth noting that GS’s noise rates are similar to or even lower than those of WoS, the other international, multidisciplinary tool assessed. Both tools provided one valid reference out of two. In contrast, the specialized DB GeoRef always retrieves almost 100% geographical literature.
- 3) With regard to overlap, we found that WoS is highly covered by GS, with values (66.5% and 82%) close to those found by Neuhaus and his collaborators (2006) (77%) and by Mikki (2009) (85%). With respect to the specialized DB GeoRef, although GS’s recall performance is reduced, our values (56.7% and 62%) are similar to the most recent ones (55%) found by Musser (2008). That strengthens the argument that GS’s coverage has

Discussion and Conclusion

improved since previous work investigating this point (Neuhaus et al., 2006, with 26% coverage).

- 4) By retrieving a large number of unique references, GS may yield additional information. Like GeoRef but unlike WoS, GS offers a broad range of references (journal articles of course, but also additional types of literature such as reports, theses, conference proceedings and books/book chapters). This additional literature can be very useful to researchers, and can obviously be acceptable as academic literature. Examples of titles present among the unique hits provided by GS are: *IAHS Publications—Proceedings and Reports Series*, *AGU Spring & Fall Meeting Abstracts*, *Geophysical Research Abstracts*, *Geological Society of America—Abstracts*, and *EGU General Assembly Abstracts*. However, some scholars are selectively interested only in peer-reviewed articles and consider that the traditional DBs are the unique way to quickly access to all peer-reviewed articles. But the notable presence of grey literature and books or books chapters is surprisingly not the main explanation for additional hits in GS: the number of unique “journal articles” is impressive. After examination of the lists of unique references, we are currently performing a study to verify the hypothesis that a significant proportion of these articles are published in more international peer-reviewed journals (e.g., the growing amount of literature in English from publishers in Asian countries).
- 5) The performance of the bibliographic tools has to be considered by subfield (human or physical geography). The statistical approach suggests that the number of human-geographical references is higher in GS, whereas GeoRef is clearly geared towards physical geography. However, in absolute terms, GS should be useful to researchers in both fields because of the large number of additional references it provides. However, our results also show that GeoRef provides a significant number of unique items in Physical Geography, putting us in agreement with Mikki (2010), who says that GeoRef, which is highly specialized in geosciences, provides titles and abstracts translated from non-English publications, which could explain the significant number of additional items.
- 6) During the nine-month period of the study, the number of references generated by traditional DBs for the years 2005–2009 proved stable, with an occasional slight increase. In comparison, GS data encountered more variation, and a detailed analysis of the references could possibly reveal even more important turnover within the lists. Nonetheless, there is such a big distance in raw numbers of findings between GS and the traditional DBs that our results and interpretations would not be affected by such an analysis.

On the basis of our findings, and in agreement with Meier and Conkling (2008), we think that GS provides a viable alternative to geographers whose institutions cannot

Discussion and Conclusion

afford or do not wish to pay for access to commercial DBs. When access to commercial DBs is possible, the distance between the results provided by the analyzed tools suggests that geographers should test their specific keywords to determine if the usage of GS remains supplementary or if it has become their best primary bibliographic tool with respect to their individual research areas.

It should be noted that GS does not present only advantages, however. Previous studies have pointed negative aspects such as variable reference format and incorrect information regarding authors, journals, and citations (Bar-Ilan, 2006; Bar-Ilan et al., 2007; Gardner & Eng, 2005; Jacsó, 2005). Criticism of the search interface no longer applies, as Google search mode is increasingly being used even in scholarly libraries; and the simplicity of the search interface (Meier and Conkling 2008) has come to be appreciated. However, simplified queries bring a large number of responses, and the processing or sorting of results remains very time-consuming in GS both due to the overall amount of data and to limited functionality (e.g., there may be limited opportunity to refine results or there may be no mass export function to reference management software).

In conclusion, since the price of commercial DBs is continuously increasing (Chen 2010a; Cusker 2013) and the budgets of academic libraries are often restricted and becoming more so, we think the free access provided by GS makes it a paramount asset for accessing the academic literature.

Chapitre 2 Caractéristiques des périodiques associés aux références uniquement fournies soit par GS, WoS ou GeoRef

1. Introduction et revue de la littérature

Dans le chapitre précédent (Chapitre 1), à partir des résultats des recherches effectuées avec les deux mots-clés, « urbanization » et « sedimentation », nous avons mis en évidence le recouvrement entre les bases de données commerciales WoS et GeoRef ainsi que le moteur de recherche GS, tous types de documents confondus (Figure 3 et tableau 3 du chapitre précédent). Ensuite, nous avons distingué les différents types de documents et répertorié ce qui était propre à chaque outil bibliographique (Tableau 4 du chapitre précédent).

Contrairement à une idée répandue, la littérature grise présente parmi les références « uniques » retrouvées par GS ne constitue pas l'essentiel de celles-ci. Il s'avère que les articles de périodiques représentent de 43 % à 77 % des références « uniques » de GS, respectivement pour les mots-clés « sedimentation » et « urbanization ». Dans les bases de données commerciales, WoS et GeoRef, les références « uniques » sont en majorité des articles de périodiques.

Dans ce chapitre, à partir du même échantillon de résultats, nous allons identifier les périodiques dans lesquels sont publiés les articles « uniques » de GS et des bases de données commerciales. En nous référant à la base de données *Ulrichsweb*, nous allons qualifier ces périodiques ainsi que cela a été fait dans certaines études décrites ci-dessous, en ajoutant aussi la nature de l'accès (OA ou pas). Cette approche permettra d'apporter des compléments à la littérature existante, surtout en ce qui concerne GS pour lequel les informations disponibles sont limitées.

Plusieurs études ont analysé la couverture d'outils bibliographiques traditionnels, dont WoS et Scopus, sur base de la liste des périodiques répertoriés et de leurs caractéristiques fournies par *Ulrichsweb* (Archambault et al., 2006; Braun et al., 2000; Gavel & Iselid, 2008; Hicks & Wang, 2011; Mongeon & Paul-Hus, 2016 ; Moya-Anegón et al., 2007). Ainsi, les caractéristiques des périodiques prises en compte dans ces études sont : la discipline, la nature académique du contenu, l'existence d'une évaluation par des pairs, la langue de publication et la répartition géographique des éditeurs.

Braun et al., (2000) se sont basés sur trois critères (disciplines, pays de publication et éditeurs) pour définir la couverture de WoS en se référant à *Ulrichsweb*. Les auteurs ont montré statistiquement que les SHS étaient sous-représentées dans WoS, tandis que les S&T étaient surreprésentées (Mathématiques, Ingénierie et Sciences de la Terre). Certains pays (Etats-Unis, Royaume-Uni, Allemagne et Pays-Bas) ainsi que certains éditeurs (Elsevier, Kluwer, Springer et Wiley) étaient plus présents que d'autres.

Archambault et al., (2006), ont mis en perspective les pays et langues des périodiques couverts par WoS avec ceux des périodiques *peer-reviewed* repris dans

Introduction et revue de la littérature

Ulrichsweb, dans le but d'évaluer les difficultés liées à l'orientation nationale et régionale de la recherche en SHS. Par rapport à la liste issue de *Ulrichsweb*, leurs résultats indiquent une surreprésentation de 20 à 25 % des revues SHS de langue anglaise dans WoS. En ce qui concerne les pays de publication, le Royaume Uni, la Russie, les Etats-Unis ainsi que la Suisse étaient surreprésentés dans WoS.

Plus récemment, Mongeon & Paul-Hus (2016) ont comparé la couverture des périodiques indexés par WoS (13 605 titres) et Scopus (20 346 titres) avec le répertoire de périodiques *Ulrichsweb* (63 013 titres). Leurs résultats indiquent une surreprésentation des périodiques en langue anglaise et que l'utilisation de WoS ou de Scopus peut créer des biais en faveur des Sciences naturelles et de l'Ingénierie ainsi que des Sciences biomédicales au détriment des Sciences sociales et des Arts et sciences humaines.

Dans le domaine précis des Sciences de la Terre et de l'Atmosphère, toujours en se référant à *Ulrichsweb*, Barnett & Lascar (2012) avaient constaté que Scopus présentait plus de titres propres (340) que WoS (16). Pour les deux bases de données, la majorité des titres propres avaient des « Journal Rank Indicators » (Scopus) et des facteurs d'impact (WoS) assez bas, indiquant, d'après les auteurs, un rôle mineur de ces sources de données. Cependant, la majorité des titres propres à chacune des deux bases étaient évalués par les pairs, publiés par des universités ou des sociétés savantes de pays développés comme les Etats-Unis, les Pays Bas, l'Allemagne, le Japon et le Canada. 30 % de périodiques étaient publiés dans des pays en voie de développement⁴⁴, dont 12 % par la Chine.

Pour ce qui est de GS, ce type d'études est plus rare à cause de l'absence d'une information claire quant aux titres de périodiques couverts (Barnett & Lascar, 2012; Bauer & Bakkalbasi, 2005; Ortega, 2014), incompatible avec les méthodologies appliquées lors des différentes études sur les titres répertoriés par les bases de données.

Néanmoins, en Business et Management (Mingers & Lipitakis, 2010) ont montré que ce sont essentiellement des revues internationales de langue anglaise qui sont couvertes par GS. Cette prédominance de l'anglais a été confirmée par Martín-Martín et al., (2016) sur base de l'étude d'un échantillon de 64 000 documents en provenance de GS, dont 93 % étaient de langue anglaise.

Dans le domaine de la Géographie, qui nous intéresse plus particulièrement, Bosman (2009) avait pointé une faible couverture des périodiques publiés par les sociétés savantes : GS couvrait 9 périodiques d'un échantillon de 26 périodiques tandis que WoS en couvrait 18.

Par ailleurs, si les études sur les titres de périodiques couverts par GS sont rares, d'autres études (De Winter et al., 2014; Harzing, 2013, 2014; Harzing & Alakangas, 2016; Mikki, 2010) se sont basées sur les listes bibliographiques, constituant un

⁴⁴ Terme utilisé par les auteurs

échantillon de documents, pour vérifier la couverture des citations par GS en comparaison à d'autres bases de données commerciales. Les auteurs signalent aussi que GS fournit plus de citations que WoS ou Scopus.

Globalement, il semble que GS ait amélioré sa couverture au fil des ans, en particulier dans le domaine des Sciences sociales (Jamali & Nabavi, 2015), ce qui pourrait distinguer le moteur de recherche des deux bases de données évoquées précédemment.

2. Objectif et hypothèses

Notre objectif est donc de dresser la liste des périodiques propres à GS, WoS et GeoRef, dans lesquels ont été publiés les articles « uniques » fournis par les différents outils bibliographiques, puis de les caractériser afin de disposer d'une base pour vérifier les trois hypothèses ci-dessous.

La première hypothèse précédemment émise (chapitre 1) *est qu'une proportion significative des articles « uniques » de GS sont publiés dans des revues internationales et évaluées par des pairs, en langue anglaise mais par des éditeurs des pays asiatiques. La force de GS serait alors de donner accès à une littérature mondiale de qualité dont une partie échapperait aux bases de données traditionnelles, WoS et GeoRef, centrées sur l'Europe et l'Amérique.*

La deuxième hypothèse *est que les références « uniques » de GS, plus que les références « uniques » de WoS ou GeoRef, comportent des articles de périodiques, traitant de sujets locaux ou régionaux. Il s'agirait d'une conséquence de la grande diversité de sources moissonnées par GS.*

Une troisième hypothèse qui considère le critère OA *est que la supériorité de GS en termes de résultats obtenus serait due aux références retrouvées par le moteur de recherche via des sources en Open Access.*

3. Méthodes

Les références d'articles de périodiques uniquement fournies par un des trois outils bibliographiques, respectivement pour les mots-clés « urbanization » et « sedimentation », ont été sélectionnées « manuellement » (voir chapitre 1) et classées dans six dossiers différents dans l'outil de gestion de références *EndNote*. Les références « uniques » de chaque outil bibliographique ont ensuite été exportées dans un fichier *Excel*. Les erreurs présentes dans les titres des périodiques (différences d'écriture - majuscules ou minuscules, usage de différents caractères et symboles) ont été repérées et corrigées. Ce traitement de données nous a permis de dédoubler les titres de périodiques au sein d'une même liste et de vérifier l'unicité des titres d'une liste à l'autre. Les périodiques dans lesquels avaient été

Résultats

publiés des articles « uniques » mais qui se retrouvaient dans les listes de plus d'un outil bibliographique, ont été écartés.

Une fois les listes de titres de périodiques en ordre, nous avons recherché les informations relatives à ces titres dans l'interface de recherche simple de *Ulrichsweb*⁴⁵ Global Serials Directory :

- le type de contenu (académique ou autre), l'évaluation par les pairs (oui/non), la recension par JCR (oui/non), trois indicateurs de qualité des périodiques ;
- l'Open Access (oui/non), qui pourrait favoriser le moissonnage de certains titres par un moteur de recherche comme GS ;
- la langue du texte, le pays de publication, caractéristiques qui, associés aux indicateurs précédents, peuvent aussi renseigner sur le caractère régional ou global des titres.

Ces informations ont été sauvegardées dans des listes sur la plateforme *Ulrichsweb* et ensuite exportées dans des fichiers *Excel*. Nous avons ainsi pu exploiter les données qui nous permettent de valider ou rejeter les hypothèses.

Les données relatives aux pays de publication des périodiques ont été illustrées à l'aide de cartes mondiales générées avec l'outil ArcGis Desktop 9. 0.

4. Résultats

Pour les deux mots-clés « urbanization » et « sedimentation », les tableaux 5 et 6 font état de la décroissance numérique des résultats associés à chacun des trois outils bibliographiques. Ces tableaux présentent six échantillons de résultats en commençant par les références « uniques » tous types de documents confondus, jusqu'aux périodiques « propres » correspondant aux articles « uniques », qui ont été retrouvés dans *Ulrichsweb*.

Outils	Références "uniques"	Articles de périodiques "uniques"	Périodiques "propres"	Périodiques "propres" trouvés
GS	2473	→ 1908	→ 805	→ 732
WoS	83	→ 83	→ 48	→ 48
GeoRef	39	→ 19	→ 13	→ 12

Tableau 5 - Définition de l'échantillon pour le mot-clé "urbanization", à partir des références « uniques » à chaque outil, tous types de documents confondus, jusqu'au périodiques « propres » trouvés dans *Ulrichsweb*

45 Ulrichsweb Global Serials Directory - <https://ulrichsweb.serialssolutions.com/>

Résultats

Outils	Références "uniques"	Articles de périodiques "uniques"	Périodiques "propres"	Périodiques "propres" trouvés
GS	904	→ 390	→ 260	→ 232
WoS	72	→ 71	→ 30	→ 29
GeoRef	466	→ 243	→ 93	→ 89

Tableau 6 - Définition de l'échantillon pour le mot-clé "sedimentation", à partir des références « uniques » à chaque outil, tous types de documents confondus, jusqu'au périodiques « propres » trouvés dans *Ulrichsweb*

Il apparaît, logiquement, que le nombre de périodiques correspondant au nombre d'articles « uniques » est toujours inférieur à ce dernier, plusieurs articles « uniques » pouvant être issus d'un même périodique.

Sur base de l'examen du nombre de périodiques propres, la dominance de GS par rapport aux deux autres bases de données se confirme avec 805 titres pour « urbanization » et 260 titres pour « sedimentation ». Nous pouvons noter que la dominance de GS s'affaiblit relativement légèrement pour « urbanization » tandis qu'elle se renforce légèrement pour « sedimentation » en comparaison de ce qu'indique le nombre d'articles « uniques ». GeoRef, avec 93 titres, reste en bonne position pour le mot-clé « sedimentation ».

Quasi tous les titres de WoS et de GeoRef ont été retrouvés dans *Ulrichsweb*. Avec des taux d'environ 90 % de recouvrement par *Ulrichsweb*, GS voit 10 % de ses titres de périodiques propres éliminés d'emblée pour la suite de notre étude. GS reste cependant encore dominant avec 732 et 232 titres de périodiques propres correspondant respectivement aux recherches effectuées avec les mots-clés « urbanization » et « sedimentation ».

Les tableaux 7 et 8 reprennent le décompte des périodiques et indiquent la répartition des caractéristiques de qualité et d'accès, telles que relevées dans *Ulrichsweb*, pour les six échantillons de périodiques obtenus.

Outils	Périodiques "propres" trouvés dans <i>Ulrichsweb</i>	Caractéristiques des périodiques trouvés						Texte en autres langues
		Contenu académique	Peer-reviewed	JCR	OA	Texte en anglais	Texte en chinois	
GS	732	587	183	61	22	205	488	39
WoS	48	44	37	36	6	38	--	10
GeoRef	12	9	7	1	--	5	--	7

Tableau 7- Répartition des périodiques propres aux trois outils bibliographiques pour le mot-clé "urbanization" en fonction des caractéristiques relevées dans *Ulrichsweb*

Résultats

Outils	Périodiques "propres" trouvés dans <i>Ulrichsweb</i>	Contenu académique	Peer-reviewed	JCR	OA	Texte en anglais	Texte en chinois	Texte en autres langues
GS	232	209	132	69	17	117	75	40
WoS	29	29	27	20	2	19	1	9
GeoRef	89	75	46	15	5	41	3	45

Tableau 8 - Répartition des périodiques propres aux trois outils bibliographiques pour le mot-clé "sedimentation" en fonction des caractéristiques relevées dans *Ulrichsweb*

4.1. Caractéristiques des périodiques associés aux articles « uniques » pour « urbanization »

4.1.1. Critères de « qualité » et mention OA

Pour le mot-clé « urbanization », les périodiques associés aux articles « uniques » ont souvent un contenu académique, selon *Ulrichsweb* (voir Tableau 7). Avec 80,2 % de périodiques qualifiés d'académiques, GS se situe entre GeoRef (75 %) et WoS (92 %). La Figure 4 illustre ces pourcentages.

Une dispersion plus importante apparaît entre les trois outils en ce qui concerne une qualité essentielle du périodique scientifique, l'évaluation par les pairs (Figure 4). En effet les périodiques de GS se démarquent par un pourcentage relativement bas de 25 %. GeoRef occupe la position intermédiaire avec 58,3 % de *peer-reviewing* tandis que WoS se distingue avec 77 % des périodiques *peer-reviewed*. Néanmoins, GS présente quand-même 183 titres évalués par les pairs (Tableau 7), soit dans l'absolu beaucoup plus que WoS (37 titres) ou GeoRef (7 titres).

Il a été vérifié que tous les périodiques *peer-reviewed* font partie de la catégorie des titres qualifiés d'académiques pour les trois outils. Un seul titre, *Journal of US-China Public Administration*, fait exception pour GS.

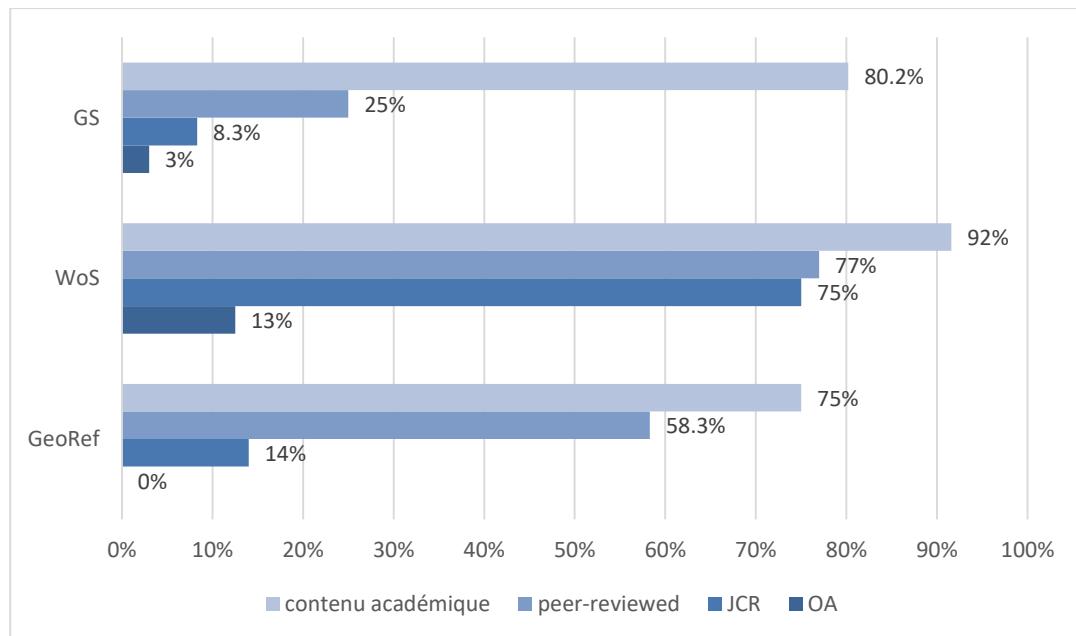


Figure 4 - Pourcentage de périodiques avec un contenu académique, *peer-reviewed*, JCR et OA par outils bibliographique pour "urbanization"

C'est en considérant la mention JCR que l'écart entre les outils est proportionnellement maximum, variant de 8,3 % pour GS à 75 % pour WoS. Toutefois, en examinant ces résultats, il ne faut pas perdre de vue le biais inhérent à ce que WoS et JCR sont produits par le même développeur.

Malgré tout, dans l'absolu, GS reste dominant avec 61 titres repris dans JCR en regard de 36 titres repris dans JCR pour WoS. Un seul titre est repris dans JCR, pour GeoRef, soit *Catena*.

Pour tous les outils, la majorité des titres couverts par JCR sont, en toute logique, issus des listes de titres *peer-reviewed*. Pourtant, certains titres qui ne sont pas qualifiés de *peer-reviewed* par Ulrichsweb se retrouvent dans les listes de titres couverts par JCR, soit 3 titres de GS et 8 titres de WoS. Après vérification, il s'avère que 9 de ces 11 titres sont en réalité *peer-reviewed*.

Peu de titres sont en OA tout outil bibliographique confondu (Tableau 7) et ils ne semblent pas alimenter particulièrement les contenus de GS.

Parmi les 549 titres (75 %) non évalués par les pairs de GS, 6 titres seulement sont en OA et 3 sont finalement quand-même repris dans JCR.

4.1.2. Langue et pays de publication

En ce qui concerne la langue de publication, l'examen du tableau 7 révèle que pour le mot-clé « urbanization » la majorité des titres propres à GS, soit 488 des 732 titres (66,6 %), sont publiés en langue chinoise alors qu'aucun titre publié en langue chinoise n'apparaît dans les titres couverts par WoS ou GeoRef. Cette prédominance se retrouve aussi dans la répartition mondiale de pays d'édition de

Résultats

ces périodiques (carte 1), la Chine se distinguant par le plus grand nombre d'éditions avec 505 titres (69 %) (Tableau 9).

Etant donné le nombre de périodiques publiés en Chine et associés aux articles « uniques » de GS pour « urbanization », il est légitime d'examiner plus attentivement les autres caractéristiques de cette catégorie de périodiques. Il en ressort que 96,6 % des périodiques publiés par la Chine présentent des articles en langue chinoise et que la quasi-totalité sont publiés par des éditeurs commerciaux chinois. 80 % de ces périodiques publiés en Chine auraient un contenu académique mais aucun de ces titres n'est renseigné comme étant évalué par les pairs dans *Ulrichsweb*, ni diffusé en OA, ni repris dans JCR. Ce sont donc ces périodiques qui augmentent l'apport de littérature « unique » de GS pour le mot-clé « urbanization ». Sans ce supplément, GS aurait des performances nettement moins importantes mais qui resteraient supérieures à celles des bases de données commerciales.

L'anglais (Tableau 7) vient en deuxième position pour GS avec 205 des 732 titres propres (28 %) tandis qu'il s'agit de la langue principale pour WoS avec 38 des 48 titres (79,2 %) et GeoRef avec 5 des 12 titres (41,6 %). Nous relevons qu'Archambault *et al.*, (2006) ont calculé des valeurs proches des nôtres pour WoS, avec 89 % des titres en anglais en Sciences et 90 % en SSH.

Les titres publiés en anglais se voient plus fréquemment qualifiés d'académiques que les titres publiés en chinois de GS : 161 des 205 titres en anglais de GS (79 %), 33 des 38 titres en anglais de WoS (87 %) et 3 des 4 titres en anglais de GeoRef (75%).

C'est surtout au niveau des données de *peer-reviewing* et de la mention JCR qu'*Ulrichsweb* marque la différence entre les titres de langue anglaise par rapport à ceux de langue chinoise. La majorité de titres en anglais des trois outils sont *peer-reviewed* : 121 des 205 titres publiés en anglais de GS (59 %), 33 des 38 titres de WoS (87 %) et 4 des 5 titres de GeoRef (80 %).

En ce qui concerne la mention JCR, il s'avère que 57 des 205 titres en anglais de GS (28%) ont la mention JCR et 27 des 38 titres en anglais de WoS (71%). Quant à GeoRef, des 4 titres en anglais, aucun n'a pas la mention JCR.

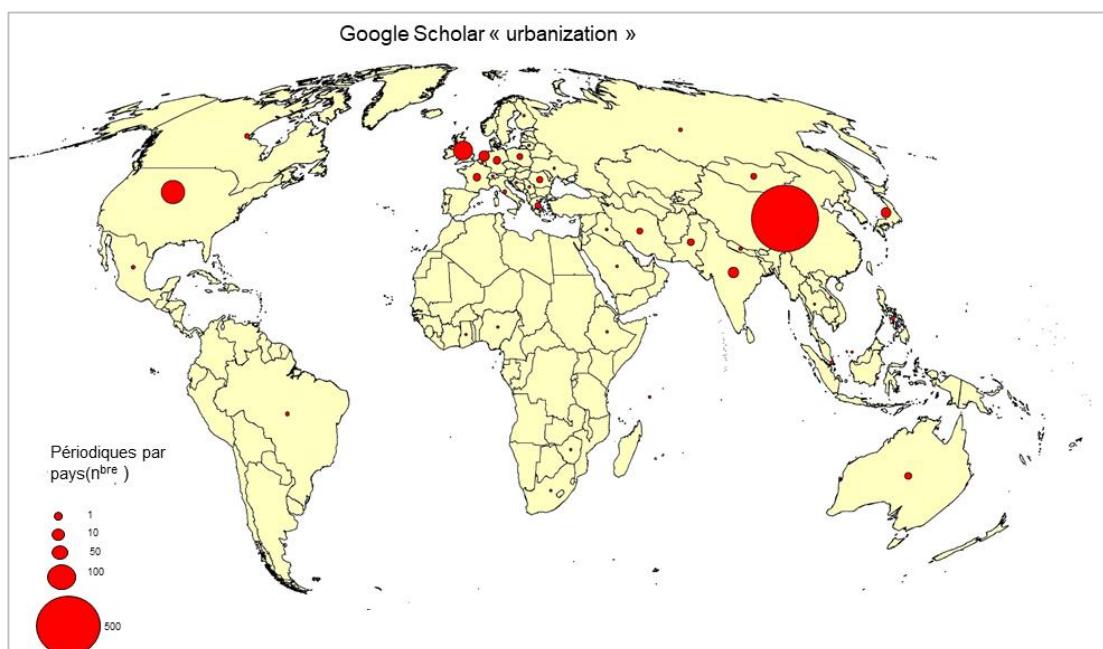
Les autres langues, dont le français, l'italien, le japonais et l'espagnol se dispersent sur les 38 titres restants de GS (5,2 %). Par contre, pour ce qui est de WoS, l'espagnol, l'allemand, le français et le tchèque viennent juste après l'anglais pour les 10 titres (20,8 %) restants. 7 titres propres de GeoRef (58,3 %) sont publiés en allemand, russe, portugais et espagnol.

Le contenu académique des titres publiés dans d'autres langues que le chinois ou l'anglais est régulièrement reconnu par *Ulrichsweb* : 29 des 38 titres de GS (76%), 9 des 10 titres de WoS (90 %) et 4 des 7 titres de GeoRef (57 %). Cependant, ils sont qualifiés de *peer-reviewed* dans une moindre mesure que les titres de langue

Résultats

anglaise : 13 des 38 titres de GS (33,3 %), 2 des 10 titres de WoS (20 %) et 3 des 7 titres de GeoRef (42,8 %).

Dans le classement des pays éditeurs (Tableau 9), pour ce qui est des titres propres à **GS**, les Etats-Unis et le Royaume-Uni suivent la Chine avec 62 (8,5 %) et 41 (5,6 %) titres publiés. A eux deux, ces états anglophones publient 103 des 205 titres de langue anglaise, soit 50 %. Viennent ensuite les Pays-Bas et l'Inde avec 13 titres chacun (1,8 %), le Japon avec 11 titres (1,5 %). 39 pays publient les 87 (12 %) de périodiques restants avec moins de 8 titres par pays. La carte 1 illustre la répartition mondiale de la publication des titres de périodiques propres à GS.

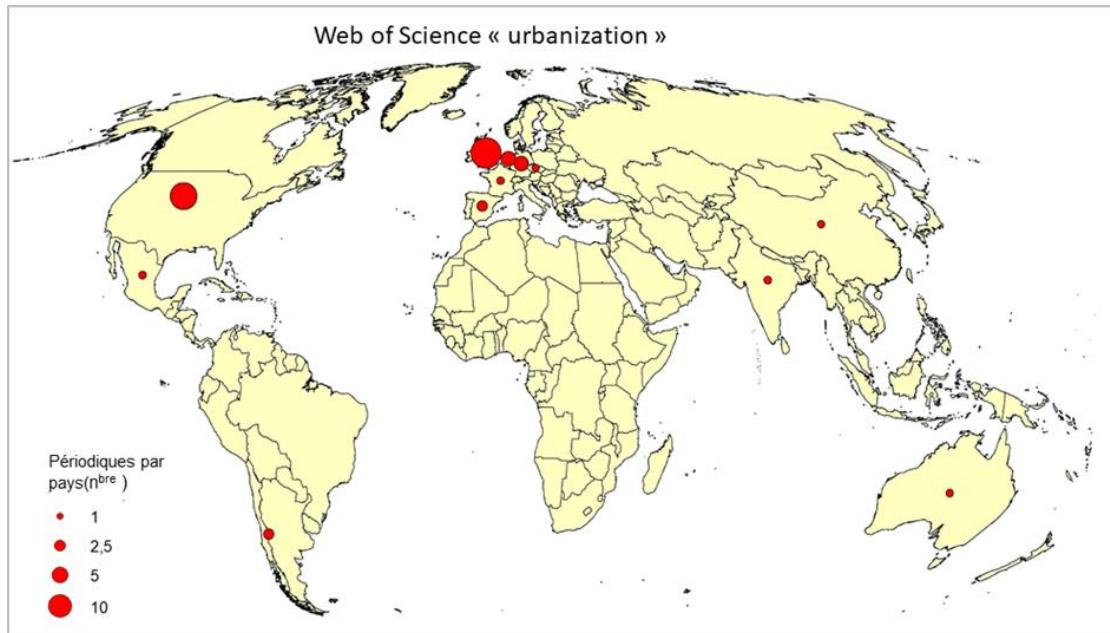


Carte 1 - Répartition au niveau mondial des pays de publication des périodiques propres correspondant aux articles uniques de GS pour le mot-clé « urbanization »

Les périodiques propres à **WoS** pour le mot-clé “urbanization” sont représentés sur la carte 2 en fonction de leur pays de publication, pour illustration des chiffres du tableau 9. La majorité (62,5 %) est publiée par le Royaume-Uni (17 titres) ou les Etats-Unis (13 titres). Archambault, *et al.*, (2006) ont indiqué des résultats similaires pour WoS.

Viennent ensuite l'Allemagne et le Pays-Bas avec respectivement 4 titres chacun. En dehors des pays des deux continents souvent considérés comme leaders de la publication scientifique que sont l'Europe et les Etats-Unis, le Chili publie 2 titres, tandis que la Chine, l'Inde, l'Australie et le Mexique publient chacun un périodique. Au total, 12 pays sont représentés, la plupart d'entre eux publient entre un à deux titres.

Résultats



Carte 2 - Répartition au niveau mondial des pays de publication des périodiques propres à WoS pour le mot-clé « urbanization »

GeoRef n'indexe que très peu de références touchant à d'autres domaines que la Géographie physique et la géologie. Les résultats obtenus pour le mot-clé « urbanization » sont, sans surprise, peu nombreux, soit seulement 12 titres de périodiques propres, que nous avons choisi de ne pas représenter cartographiquement. Les Etats-Unis sont les mieux représentés avec la publication de trois périodiques ; la Russie et le Brésil publient chacun deux titres, tandis que le Royaume-Uni, l'Allemagne, l'Argentine, l'Albanie et le Pays-Bas publient chacun un titre. Les 12 titres sont publiés par des éditeurs commerciaux.

Résultats

"urbanization"									
Pays	GS			GeoRef			WoS		
	N	%	Classement	N	%	Classement	N	%	Classement
Chine	505	69	1				1	2	6
Etats-Unis	62	8,5	2	3	25	1	13	22,4	2
Royaume-Uni	41	5,6	3	1	8,3	4	17	35,4	1
Pays-Bas	13	1,8	4	1	8,3	5	4	8,3	3
Inde	13	1,8	4				1	2,8	5
Japon	11	1,5	5						
Allemagne	7	0,9	6	1	8,3	6	4	8,3	4
France	7	0,9	6				1	2	7
Australie	6	0,8	7				1	2	8
Pakistan	5	0,7	8						
Iran	4	0,5	9						
Pologne	4	0,5	9						
Roumanie	4	0,5	9						
Grèce	4	0,5	9						
Canada	3	0,4	9						
Bresil	2	0,3	10	2	16,6	2			
Russie	2	0,3	10	2	16,6	3			
Italie	2	0,3	10						
Népal	2	0,3	10						

Tableau 9 - Classement des pays d'édition des périodiques correspondant aux articles « uniques » de GS, GeoRef et WoS pour les mot-clé « urbanization »

Résultats

4.2. Caractéristiques de périodiques associées aux articles retrouvés avec « sedimentation »

4.2.1. Critères de « qualité » et mention OA

Pour le mot-clé « sedimentation », plus souvent encore que pour le mot-clé « urbanization », le contenu des périodiques propres à chaque outil est qualifié d'académique, avec des pourcentages allant de 84 % pour GeoRef à 100 % pour WoS. Comme pour le mot-clé « urbanization », GS se situe entre les deux bases commerciales avec un pourcentage de 90 % (Figure 5).

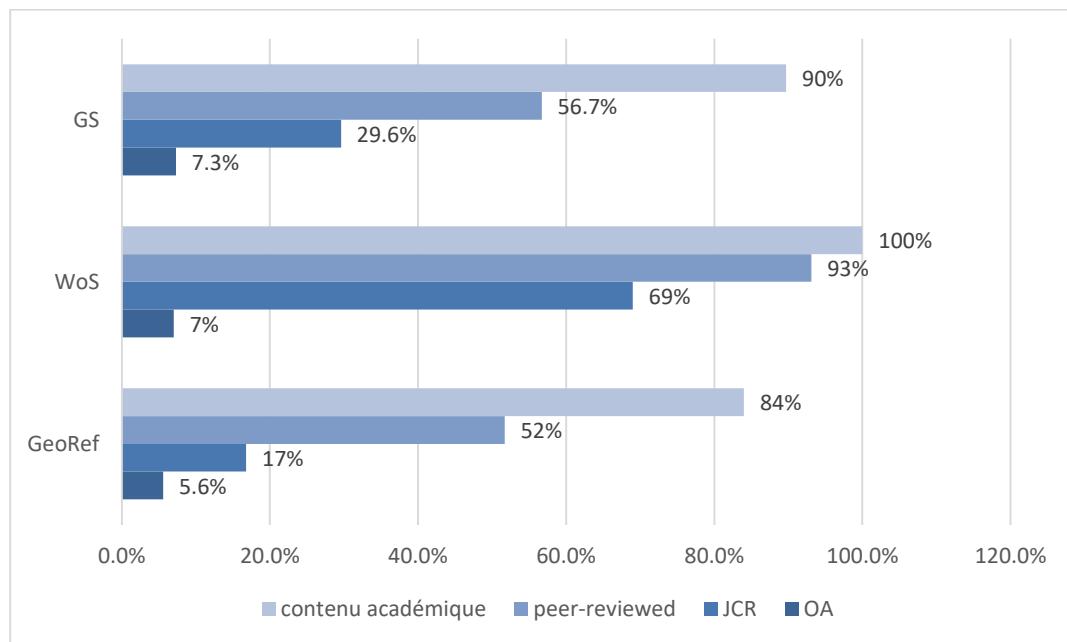


Figure 5 - Pourcentage de périodiques avec un contenu académique, peer-reviewed, JCR et OA par outils bibliographiques pour "sedimentation"

Les pourcentages de périodiques évalués par les pairs pour ce mot-clé sont proches pour GS (56,7%) et GeoRef (52 %). L'écart entre GS et les autres BD pour le mot-clé « sedimentation » est donc moins marqué que pour le mot-clé « urbanization » tandis que GeoRef présente une valeur équivalente à celle d'« urbanization ». Pour WoS les valeurs sont plus élevées qu'auparavant avec 93 % de titres qui ont la mention *peer-reviewed*.

Comme pour le mot-clé « urbanization », dans l'absolu, GS présente plus de titres évalués par des pairs, soit 132 titres (Tableau 8), pour 27 titres dans WoS et 46 dans GeoRef. Il a été vérifié que tous les périodiques qualifiés de *peer-reviewed* avaient aussi un contenu académique, pour les trois outils.

En ce qui concerne la mention JCR, l'écart entre GS et les deux autres BD est également moins marqué avec le mot-clé « sedimentation » qu'avec le mot-clé « urbanization » : 29,6 % pour GS (69 titres propres dont 64 *peer-reviewed*), 17 % pour GeoRef (15 titres propres dont 14 *peer-reviewed*) et 69 % pour WoS (20 titres

propres dont 19 *peer-reviewed*). Ces valeurs sont plus élevées que celles mentionnées pour « urbanization » à la fois pour GS et GeoRef.

La majorité des titres couverts par JCR sont, en toute logique, issus des listes de titres *peer-reviewed*, mais des exceptions persistent pour 5 titres de GS, 1 titre de WoS et 1 titre de GeoRef. Après vérification, il s'avère que 6 de ces 7 titres sont en réalité *peer-reviewed*.

Peu de titres sont en OA pour ce mot-clé ci également, tous outils confondus (Tableau 8). Seulement 6 des 17 titres OA de GS se retrouvent dans la catégorie des 101 titres non *peer-reviewed* de GS.

4.2.2. Langue et pays de publication

Pour « sedimentation », l'anglais est la première langue des périodiques propres à chacun des trois outils : 117 des 232 titres de GS (50 %), 19 des 29 titres de WoS (66 %) et 41 des 89 titres de GeoRef (46%). On peut cependant remarquer que l'anglais n'a pas la majorité absolue face aux autres langues pour GeoRef, ce qui était déjà le cas pour « urbanization » : le nombre de périodiques en autres langues est de 45 titres (50,5 %) (Tableau 10).

Comme pour le mot-clé précédent la majorité des titres en anglais sont qualifiés d'académiques : 108 titres des 117 titres en anglais de GS (92 %), la totalité de 19 titres en anglais de WoS, et 37 de 41 titres en anglais de GeoRef (90 %).

L'analyse de données révèle que la majorité des titres en langue anglaise sont qualifiés de *peer-reviewed* pour GS et WoS avec 94 des 117 titres en anglais de GS (80,3 %) et 17 des 19 titres en anglais de WoS (89,5 %). Pour GS, il est évident, cette fois encore, que la qualité de *peer-reviewed* est plus fréquemment attribuée aux titres de langue anglaise. Pour ce qui est de WoS, la différence n'apparaît pas pour le mot-clé « sedimentation ». Pour GeoRef, 24 titres publiés en anglais soit un peu plus que la moitié (58,5 %) sont *peer-reviewed*. C'est légèrement mieux que la moyenne de l'ensemble des titres.

En ce qui concerne les autres caractéristiques de « qualité » il s'avère que 58 des 117 titres en anglais de GS (49,5 %) ont la mention JCR et 12 des 19 titres en anglais de WoS (63 %). Quant à GeoRef, des 41 titres en anglais, 11(27 %) ont la mention JCR associée. Pour ce critère également, la performance des titres de langue anglaise est plus élevée que celle de l'ensemble des titres de GS.

Les périodiques en langue chinoise pour GS sont moins nombreux pour ce mot-clé mais restent importants avec 75 des 232 titres retrouvés dans *Ulrichsweb* (32,1 %). Parmi les 75 titres de langue chinoise, 67 sont qualifiés d'académiques et 21 sont recensés comme *peer-reviewed* (28 %). GeoRef compte 3 titres (3,4 %) en chinois et WoS un titre (3,5 %).

Les 40 périodiques restants (17,2 %) de GS sont publiés dans les autres langues, dont le français, l'allemand, le japonais, l'espagnol, tandis que 9 titres de WoS (31 %) sont publiés dans d'autres langues, dont le français et l'allemand. Les 45 titres

Résultats

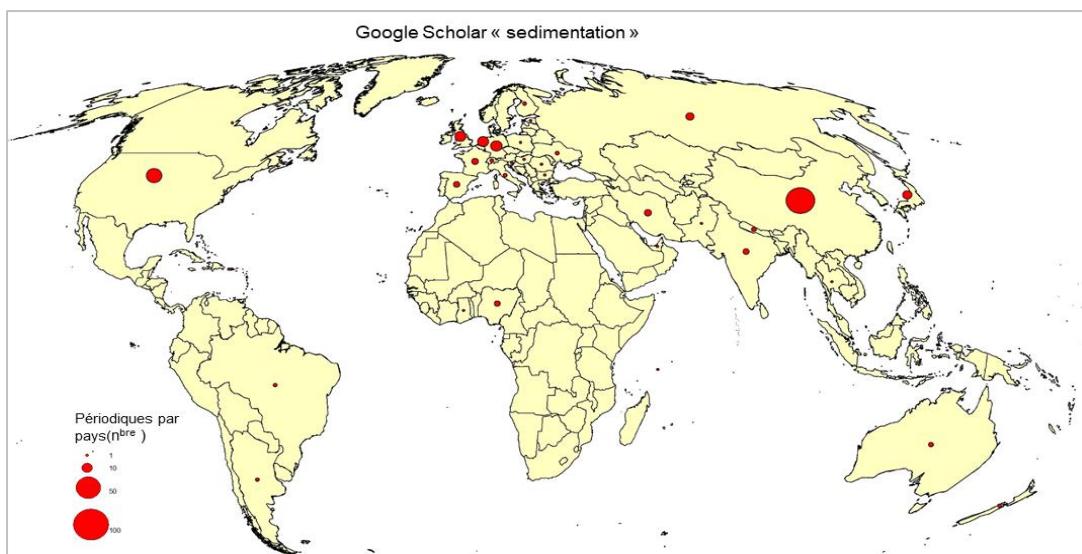
de GeoRef (50,6 %) publiés dans d'autres langues sont répartis entre le russe, l'espagnol, le japonais, l'allemand, l'italien, le polonais et le portugais.

Pour « sedimentation », comme pour « urbanization », les titres publiés dans d'autres langues sont souvent qualifiés d'académiques par *Ulrichsweb* : 36 des 40 titres d'autres langues de GS (90 %), tous les titres de WoS (9 titres) et 37 des 45 titres de GeoRef (82 %).

Par rapport aux périodiques publiés en langue anglaise, les titres publiés dans d'autres langues sont moins souvent qualifiés de *peer-reviewed* quand il s'agit de GS (16 des 40 titres de GS, soit 40 %) ou de GeoRef (22 des 45 titres de GeoRef soit 48,8 %), tandis que tous les titres de WoS sont *peer-reviewed* (9 titres, soit 100 %).

Si l'on considère la répartition mondiale des titres par pays d'édition (carte 3), la Chine se distingue comme le plus gros pays éditeur de périodiques avec 89 des 232 titres propres à GS (38,2 %). Parmi les 89 périodiques publiés par la Chine, 75 titres (84 %) ont le texte des articles en langue chinoise, ce qui n'a pas suffi pour imposer la langue chinoise en première place. La quasi-totalité de ces titres sont publiés par des éditeurs commerciaux.

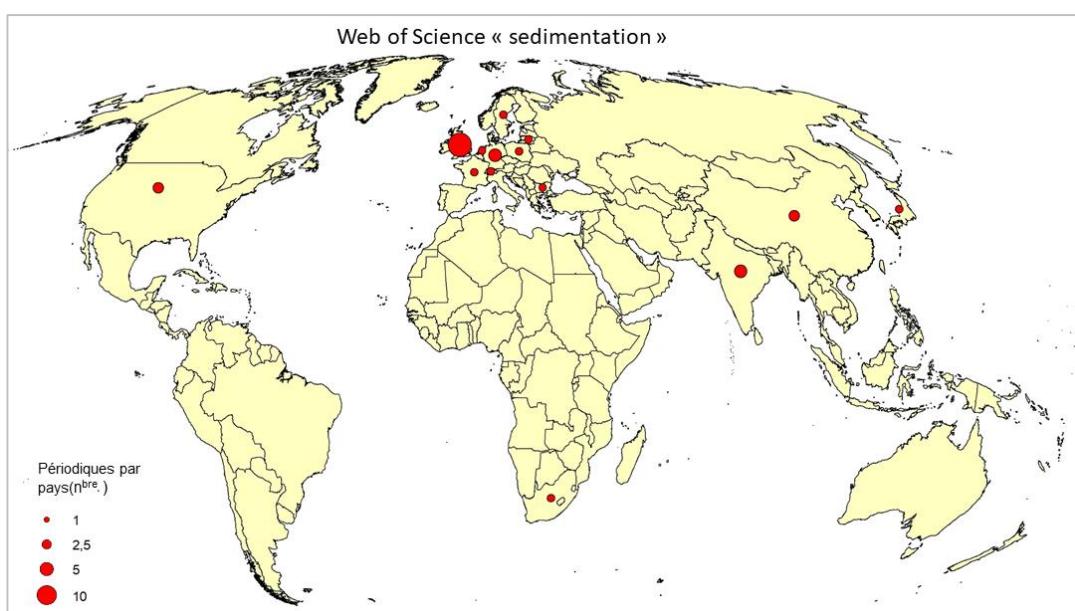
Comparé aux résultats du mot-clé « urbanization », pour lequel aucun des titres n'était évalué par les pairs, 31 titres (34 %) publiés par la Chine ont la mention *peer-reviewed* attribuée dans *Ulrichsweb*, et 6 titres (6,7 %) sont repris dans JCR. Comme pour « urbanization », aucun titre n'est publié en OA. Les Etats-Unis et le Royaume-Uni suivent la Chine avec 27 (11,6 %) et 14 (6 %) titres chacun. A eux deux, ces états anglophones ne peuvent être responsables que de la publication de 41 des 117 titres de langue anglaise, soit 35 % des titres, valeur nettement moins élevée par rapport à « urbanization ». Les Pays-Bas, l'Allemagne et le Japon publient respectivement 15 (6,4 %), 14 (6,3 %) et 10 (4,3 %) titres chacun. Au total 34 pays publient les périodiques propres à GS.



Carte 3 - Répartition au niveau mondial des pays de publication des périodiques uniquement indexés par GS pour le mot-clé « sedimentation »

La liste des pays de publication des périodiques propres à WoS pour le mot-clé « sedimentation », illustrée sur la carte 4, présente un noyau commun à celle d'« urbanization », avec en plus toute une série de pays « faibles producteurs » : Allemagne, Inde, Chine, Pays-Bas, Japon, France, Pologne... Le Royaume-Uni se positionne nettement en première position tandis que les Etats-Unis ne se distinguent guère des autres pays. Le Japon et l'Afrique du Sud ont remplacé le Chili et le Mexique précédemment présents pour le mot-clé « urbanization ». Si on excepte le Royaume Uni, de 14 pays qui publient ces titres, 13 n'en publient qu'entre un et trois.

La dominance de la langue anglaise dans les périodiques propres à WoS s'explique en partie par la majorité de titres publiés par le Royaume-Uni (10 titres) auxquels s'ajoutent les titres publiés par les Etats-Unis (2 titres) soit en tout 41,4 % des titres. Les 7 titres restants, publiés en anglais, le sont par les autres pays (Inde, Japon, Pays-Bas, Pologne, Afrique du Sud).

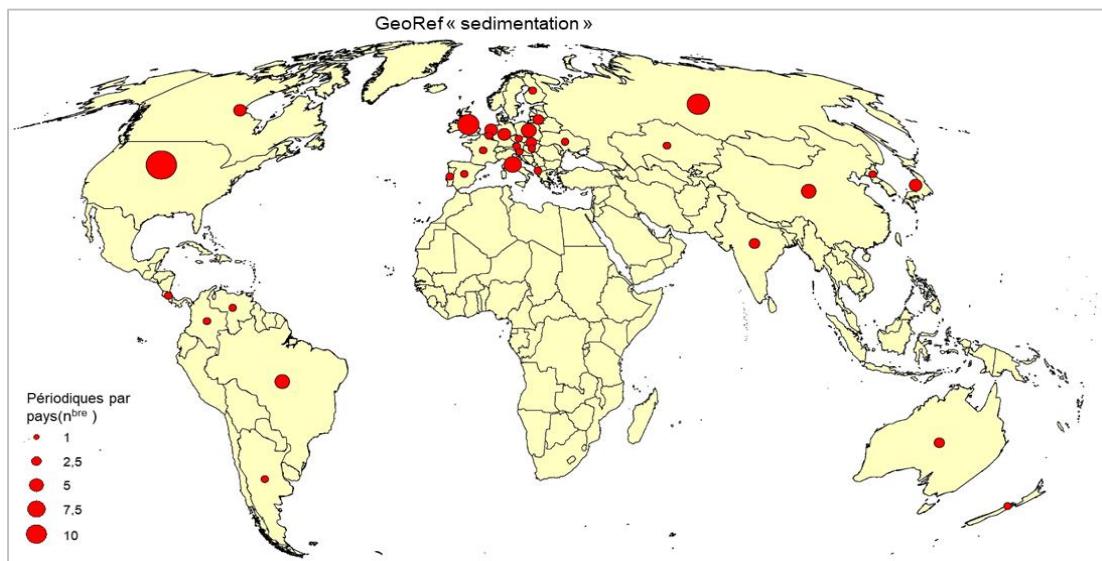


Carte 4 - Répartition au niveau mondial des pays de publication des périodiques uniquement indexés par WoS pour le mot-clé « sedimentation »

Pour GeoRef, avec 89 titres propres correspondant au mot-clé « sedimentation », la liste des pays de publication est plus longue et plus diversifiée que pour le mot-clé « urbanization ». La diversité des pays de publication (33 pays) se marque également au niveau des langues de publication (14 langues). Comme nous pouvons le remarquer sur la carte 5, les titres de périodiques sont publiés par les pays développés et les pays émergents, tels que la Chine, le Brésil et l'Inde. Les Etats Unis sont le mieux représentés avec 16 titres (18 %), la Russie vient ensuite avec 9 titres (10 %) tandis que le Royaume Uni publie 8 titres (9 %). L'Italie publie 5 titres, la Pologne et le Brésil publient 4 titres chacun. A la Chine reviennent 4 titres dont

Résultats

3 sont publiés en langue anglaise et 1 en chinois (tous avec un contenu académique et 2 *peer-reviewed*).



Carte 5 - Répartition au niveau mondial des pays de publication des périodiques uniquement indexés par GeoRef pour le mot-clé « sedimentation »

Les pays africains, à quelques exceptions près pour GS, sont les grands absents du paysage de la publication des périodiques propres à chacun des outils analysés.

"sedimentation"

Pays	GS			GeoRef			WoS		
	N	%	Classement	N	%	Classement	N	%	Classement
Chine	89	38,3	1	4	4,5	5	2	6,9	3
Etats-Unis	27	11,6	2	16	18	1	2	6,9	3
Royaume-U	14	6,3	3	8	9	2	10	34,5	1
Pays-Bas	15	6,4	4	3	3,4	6	1	3,4	4
Japon	10	4,3	5	3	3,4	6	1	3,4	4
Russie	8	3,4	6	9	10,1	3			
France	5	2,1	7	1	1,1	8	1	3,4	4
Iran	5	2,1	7						
Inde	4	1,7	8	2	2,2	7	3	10,3	2
Nigeria	4	1,7	8						
Australie	3	1,3	9	2	2,2	8			
Nepal	3	1,3	9						
Allemagne	14	6,3	10	3	3,4	6	3	10,3	2
Brezil	2	0,5	10	4	4,5	5			
Italie	2	0,5	10	5	5,6	4			
Pakistan	1	0,4	11						
Canada	1	0,4	11	3	3,4	6			
Pologne	1	0,4	11	4	4,5	5	1	3,4	4
Roumanie	1	0,4	11						

Tableau 10 - Classement des pays d'édition des périodiques correspondant aux articles « uniques » de GS, GeoRef et WoS pour le mot-clé « sedimentation »

5. Discussion et conclusions

Les résultats obtenus, que ce soit pour le mot-clé « urbanization » ou pour le mot-clé « sedimentation » rencontrent en partie notre **première hypothèse** de travail.

L'apport de publications chinoises *via* GS s'explique par le moissonnage systématique de plusieurs plateformes et éditeurs chinois, telle par exemple la base de données bibliographique nationale chinoise CNKI⁴⁶, dont les contenus ne se retrouvent que très occasionnellement dans les bases de données commerciales. Ce constat est également valable pour d'autres pays dont les publications sont peu représentées notamment dans WoS ou GeoRef.

Si la littérature chinoise moissonnée par GS est relativement plus importante pour « urbanization », cela peut s'expliquer par la problématique aigue de l'urbanisation dans ce pays, compte tenu de l'industrialisation intensive et de la migration de travailleurs.

Ce nombre impressionnant d'articles pourrait être dû à la croissance et l'investissement de la Chine dans la recherche scientifique. Ce pays a été catégorisé comme « *nouvelle puissance scientifique* », avec un taux de croissance de 298 % de la publication scientifique entre 2000 et 2008, (Baron & Eckert, 2014) et avec et avec près de 401 435 articles par an⁴⁷ d'après les données recueillies sur la plateforme World Bank.

La nature académique des publications chinoises est généralement reconnue par *Ulrichsweb*. Par contre, la mention *peer-reviewed* est rarement attribuée. L'absence de cette mention pose évidemment la question de la qualité de ces publications, que nous ne sommes pas en mesure d'évaluer. Le *peer-reviewing* pratiqué en Chine, se base sur un système à trois niveaux dans lequel les éditeurs des périodiques ont le rôle principal (Qing et al., 2008). Les origines de ce système ont été gardées du système de publication des livres et est différent par rapport à celui appliqué par les éditeurs des périodiques occidentaux.

Nous avons également constaté que la grande majorité des périodiques édités en Chine étaient publiés en langue chinoise, ce qui ne rencontre pas notre hypothèse de départ. Nous n'excluons pas que la langue constitue un obstacle à une analyse complète de ces périodiques par *Ulrichsweb* et que l'absence de la mention *peer-*

⁴⁶ China National Knowledge Infrastructure (<http://oversea.cnki.net/kns55/default.aspx>) est soutenue par l'Université de Tsinghua et par le gouvernement chinois, est un projet à échelle nationale qui a pour but de rendre disponible électroniquement des articles scientifiques chinois, des thèses, des annuaires, des périodiques, et tout autre type de documents qui comprend des résultats de recherches scientifiques. La plateforme de consultation est composée de plusieurs bases de données spécifiques dont une de plus importantes est "China Academic Journals Full-text Database" (CAJ) et "China Doctor/Master Dissertations Full-text Database" (CDMD).

⁴⁷ Données de la World Bank en 2017

Discussion et conclusions

reviewed n'est pas réellement significative d'une non-évaluation des contenus par des experts.

Ce manque de fiabilité présumée de la mention *peer-reviewed* dans *Ulrichsweb* est confortée par certains de nos résultats : les titres qui n'avaient pas reçu de mention *peer-reviewed* dans *Ulrichsweb* alors qu'ils étaient repris dans JCR se sont souvent révélés être *peer-reviewed* après vérification par nos soins. Dans la littérature, Hicks & Wang, 2011 ont pointé des incohérences quant au référencement des titres *peer-reviewed* dans *Ulrichsweb*, plus particulièrement pour les périodiques édités dans d'autres langues que l'anglais. Nous avons remarqué aussi que les titres non chinois publiés dans d'autres langues que l'anglais étaient moins souvent qualifiés de *peer-reviewed*.

Par ailleurs, une proportion importante des périodiques de langue anglaise, propres à GS, sont *peer-reviewed*. Comme nous l'avons mentionné dans notre hypothèse, GS donne accès à une littérature mondiale de qualité qui semble échapper aux bases de données traditionnelles, en particulier à WoS. Une partie de ces périodiques de langue anglaise est produite par des pays qui ne sont pas nécessairement anglophones et qui se répartissent sur divers continents. Pour le mot-clé « sedimentation », GeoRef s'associe d'une certaine façon à GS, avec une diversité de pays d'édition.

Evidemment, les trois outils se différencient par leur contenu, GS ayant pour but de retrouver la totalité des publications scientifiques disponibles sur le web. Le moteur de recherche n'a pas pour but de répertorier un corpus défini de titres de périodiques, comme c'est le cas de WoS et GeoRef.

La première hypothèse selon laquelle *une proportion significative des articles « uniques » de GS sont publiés par des éditeurs asiatiques, chinois en l'occurrence, qui sont très peu couverts par les bases de données que nous utilisons et qui sont centrées sur l'Europe et l'Amérique* ne peut être rejetée.

Aux côtés des périodiques chinois de GS et des périodiques anglophones *peer-reviewed*, il existe des périodiques dont les langues de publication sont diverses, auxquels un contenu plus national ou local pourrait être associé (ex. : *Indonesian Journal of Geography*, *Chinese Journal of Population Science*, *Japanese Progress in Climatology*, *Urban Policy and Research*, *Estudios Demograficos y Urbanos*, *Canadian Journal of Earth Sciences*, *Singapore Journal of Tropical Geography*, *Latin American Journal of Sedimentology and Basin Analysis*). La présence de ceux-ci, en nombre dans GS mais aussi dans GeoRef quand il s'agit de Géographie physique, semble conforter notre **deuxième hypothèse**, complémentaire de notre première hypothèse.

Les périodiques nationaux en Géographie, souvent publiés dans la langue nationale et dont le titre peut faire référence au pays éditeur, se veulent le reflet de la recherche effectuée dans le pays. N'étant pas nécessairement couverts par les BD multidisciplinaires commerciales, ou assortis d'un FI, ces périodiques gardent tout

Discussion et conclusions

leur intérêt pour les géographes qui analysent les interactions entre différentes composantes naturelles et humaines de la Terre, tout en accordant une grande importance aux spécificités locales.

L'intérêt de ces titres pour les géographes ne peut pas être ignoré. Un article traitant un aspect local qui n'est pas publié dans un « grand » périodique en Géographie ne le rend pas moins intéressant pour les utilisateurs potentiels (Schmitz, 2003). Néanmoins, nous ne pouvons pas ignorer le fait que ces articles pourraient généralement être moins utiles pour les géographes qui travaillent sur des sujets de recherche de niveau international.

Nous considérons que la diversité des pays d'édition des périodiques propres à GS et GeoRef (pour « sedimentation ») et des langues associées, constitue un atout pour les géographes. Malgré qu'un nombre important de périodiques propres à GS présentent un texte en d'autres langues que l'anglais et ne sont pas *peer-reviewed*, ceux-ci n'en sont pas moins des fournisseurs de contenu susceptible de répondre aux besoins des géographes en termes de littérature locale ou régionale.

La deuxième hypothèse selon laquelle *les références « uniques » de GS, plus que les références « uniques » de WoS ou GeoRef, comportent des articles de périodiques traitant de sujets locaux ou régionaux* est confirmée.

En référence à la troisième hypothèse qui concerne le caractère OA des périodiques les résultats obtenus indiquent en effet, que tous outils bibliographiques confondus, peu de périodiques « propres » sont en OA. Cela ne signifie pas que d'une façon générale la proportion de périodiques OA est nécessairement faible, mais simplement que la couverture des périodiques OA n'est pas l'apanage d'un ou l'autre des outils.

La troisième hypothèse selon laquelle *selon laquelle la supériorité de GS en termes de résultats obtenus serait due aux références retrouvées par le moteur de recherche via des sources en Open Access* n'est pas confirmée.

En conclusion, la production scientifique et l'impact des résultats de recherche sont devenus des critères très importants dans la hiérarchisation des pays au niveau mondial. Le développement économique d'un pays génère des changements importants et implique, en général, de plus grandes dépenses dans la recherche et le développement, fait qui entraînera une augmentation de la production scientifique (Gonzalez-Brambila *et al.*, 2016). La Chine, et les autres pays émergents comme le Brésil ou encore l'Inde, se retrouvent ainsi parmi les premiers producteurs de littérature scientifique (Baron & Eckert, 2014). En contrepartie, l'Afrique est peu représentée. Cela peut s'expliquer par le fait que la production scientifique d'un pays dépend des plusieurs critères : économiques principalement, mais aussi des traditions de recherche et de publication scientifique, établies historiquement (Barnett & Lascar, 2012; Mongeon & Paul-Hus, 2016; Moya-Anegón & Herrero-Solana, 2013; Salager-Meyer, 2008).

Discussion et conclusions

Dans un contexte de mondialisation, sans négliger l'incidence, au cas par cas, du sujet de recherche lui-même sur le choix des outils de recherche bibliographique et des documents de référence, il nous semble important de pouvoir disposer d'accès à la fois à la littérature internationale et à la littérature locale, produite par tous les continents. GS est certainement un des outils qui peuvent se charger de cette fonction.

-

Chapitre 3 Usage de la littérature scientifique dans un échantillon de trois dissertations doctorales

1. Introduction et revue de la littérature

Dans les deux premiers chapitres de la thèse nous nous sommes questionnée sur l'utilité de GS dans le processus de recherche bibliographique en Géographie par rapport aux bases de données commerciales WoS, GeoRef et FRANCIS. Simulant de façon simplifiée la démarche d'un chercheur, nous avons effectué des recherches bibliographiques par mots-clés, relevant de la Géographie physique et humaine, et nous avons conclu que GS peut fournir une quantité importante de littérature en comparaison des bases de données commerciales. Parmi les références propres à GS, outre des articles de périodiques potentiellement intéressants, nous avons mis en évidence la diversité des types de documents fournis (comptes rendus de congrès, livres et chapitres de livres, thèses, rapports) (Ştirbu *et al.*, 2015). Pour ce qui est des bases de données commerciales, leur apport spécifique est plus réduit.

Dans ce troisième chapitre, nous mettons en perspective les résultats précédents, qui suggèrent que GS pourrait être un outil incontournable d'accès à la littérature, avec l'usage réel que les chercheurs font de la littérature, sur base de l'étude des bibliographies d'un échantillon de trois thèses de doctorat récemment défendues à l'ULiège. Il s'agit d'une étude de cas, basée sur un échantillon limité, dont l'objet est de constater si les chercheurs citent, et dans quelles proportions, des documents spécifiquement couverts par l'un ou l'autre outil bibliographique, en particulier par le moteur de recherche GS.

Ce chapitre s'inscrit donc dans la vaste lignée des études de citations, néanmoins sans ambition de représentativité statistique.

1.1. Etudier les citations des chercheurs pour orienter les choix d'acquisitions en bibliothèque

La première étude des citations a été effectuée par Gross & Gross en (1927), dans le domaine de la Chimie. Ces pionniers de l'analyse des citations dans les mémoires, thèses et publications des membres d'une faculté, avaient pour objectif principal de définir la stratégie d'achat des périodiques indispensables à une discipline. Par la suite, d'autres études du même type ont été reproduites (Edwards, 1999; Enger, 2009; Kaczor, 2014; Kriz, 1978; McCain & Bobick, 1981; McCullough, 2016; M. Sylvia & Lesher, 1995; Waugh & Ruppel, 2004). Ces études se sont avérées efficaces dans la détermination des tendances d'utilisation des documents par les différents groupes d'utilisateurs et dans la détermination de l'utilité d'une collection pour les disciplines afférentes. L'élément de base dans ce type d'analyse est la « citation », définie comme “*a bibliographical entry in a footnote, reference list or bibliography that contains enough information, e.g., author, title, publisher, or journal title to verify the original item*” (Leiding, 2005, p. 418).

Introduction et revue de la littérature

Malgré le fait que beaucoup de ces études aient été réalisées dans le domaine des sciences, peu d'entre elles portaient sur les géosciences. Quelques articles publiés entre 1931 et 1969 ont tenté d'identifier les caractéristiques de la littérature citée en géologie (Craig, 1969; Crissinger, 1981; Gross & Woodford, 1931; Woodford, 1969) en utilisant un petit nombre de périodiques comme sources de données et dans un but de « collection management ». Les résultats de ces études ont montré une constance remarquable dans ce laps de temps, la littérature géo-scientifique citée étant constituée principalement d'articles de périodiques (75 à 86 % des citations) puis de monographies (12 à 22 % des citations). 41 à 49 % de la littérature citée par les géologues provenait d'autres disciplines que les géosciences. La diversité des langues était faible : 79 à 91 % de la littérature citée était dans la langue de l'auteur, soit l'anglais dans la plupart des cas. L'intervalle de temps nécessaire pour couvrir 90 % des citations était de quatre décennies.

Si l'exploitation des citations des travaux scientifiques s'est imposée comme méthode d'analyse objective et indépendante et est devenue très populaire, des limites ont pourtant été pointées par différents auteurs (Sylvia, 1998; Waugh & Ruppel, 2004) :

- les chercheurs sont plus susceptibles de citer des références bibliographiques auxquelles ils ont un accès local et direct ;
- des citations peuvent être ajoutées pour augmenter la taille du manuscrit, surtout lorsque la bibliographie doit inclure un nombre requis de références ;
- les chercheurs peuvent citer des travaux d'importance marginale dans la préparation du manuscrit ;
- les chercheurs pourraient ne pas citer tous les travaux consultés pour la préparation du manuscrit ;
- il existe de multiples publications avec les mêmes données et méthodologies, ce qui peut biaiser les citations ;
- les manuels et les manuels scolaires ne font pas souvent l'objet de citations ;

A ces limites, nous pourrions ajouter des limites liées à une certaine pression éditoriale :

- les chercheurs citent leurs propres publications ou celles des collègues travaillant dans le même laboratoire ;
- certains éditeurs imposent la citation d'articles parus dans leurs périodiques ;
- des reviewers anonymes suggèrent de citer leurs publications ou de citer des chercheurs le leur entourage.

Les thèses constituent une source de données fiable, que ce soit pour évaluer dans quelle mesure les collections répondent aux besoins des doctorants (Edwards & Jones, 2014), ou pour identifier la littérature citée par ceux-ci (Becker & Chiware, 2015; Conkling et al., 2010; Iivonen et al., 2009; Kaczor, 2014; Kayongo & Helm,

2012; Kimball et al., 2013; Kumar & Dora, 2011; Kushkowsky, Parsons, & Wiese, 2003; Nabe & Imre, 2008; Zong et al., 2013).

L'analyse des citations s'est particulièrement intensifiée après les années 1990 lorsque les bibliothèques et la communauté universitaire ont réagi à l'augmentation des prix des périodiques puis à l'introduction et la multiplication des ressources électroniques.

1.2. L'impact de l'émergence de la publication électronique

Dans une enquête sur les pratiques de recherches documentaires, réalisée auprès de 80 chercheurs en Géosciences, Hallmark (1994) constatait « *a relatively conservative picture of the geocientists information seeker* » et « *geoscientists rely heavily on their personal networks, references in the literature, and browsing to discover relevant references, and are very much concerned with their library's budget and journal holdings* » (GSIS Proceedings, page 114).

Néanmoins, avec le développement du support électronique, l'émergence de la publication électronique, la diversification des outils bibliographiques et leur mise à disposition pour une majorité des chercheurs ainsi que la multiplication des ressources en OA, les pratiques ont changé et de nouvelles questions ont été soulevées.

Les thèses de doctorat ont aussi constitué un matériel de choix pour les chercheurs qui ont tenté de faire la part des choses entre les citations de documents provenant des collections « papier » de bibliothèques universitaires et les citations des documents accessibles *via* les moteurs de recherche et les bases de données (Condic, 2015; Conkling et al., 2010; Grigas et al., 2017; Noga et al., 1993; Romic & Mitrović, 2014; Vezzosi, 2009). L'aboutissement d'une thèse représente la fin d'un vaste processus d'éducation et de recherche, qui amène de nouvelles contributions aux connaissances dans une discipline. Ces nouvelles contributions se basent sur les résultats de la recherche et, aussi, sur la littérature déjà publiée sur le sujet. La citation d'une source dans une thèse indique un grand degré d'utilité de la source pour le travail du doctorant, sans que ce dernier ne soit influencé dans ses choix de citations par un éditeur commercial.

Dès 2009, l'étude menée par Vezzosi (2009) a montré que les doctorants s'appuient fortement sur le Web pour leurs recherches, tandis que leur utilisation de la bibliothèque se limite à la fourniture de documents et au prêt entre bibliothèques.

Par la suite, d'autres chercheurs ont examiné l'utilisation de la littérature dans l'environnement pré-web (1990–1993) et post-web, dans les bibliographies des thèses de plusieurs disciplines (Conkling et al., 2010), ou encore dans les publications des académiques et des doctorants du domaine de la « lecture et du leadership éducatif » (Condic, 2015).

Introduction et revue de la littérature

Selon Conkling et al., (2010), les doctorants citent plus des articles de périodiques dans la période « post-web » (2003–2006) tandis que l'usage des monographies, dissertations et rapports diminue.

Condic (2015) relève des pratiques de citation différentes entre les doctorants et les chercheurs, les premiers utilisant un plus large panel de types de documents (actes de conférences, rapports techniques et gouvernementaux) et des périodiques avec des facteurs d'impact plus faibles.

Plus récemment, Grigas et al., (2017) ont analysé les citations de 39 thèses, choisies au hasard, dans différents domaines de recherche soutenues à l'Université de Vilnius à la fin 2014. En moyenne 50 % des citations sont des articles des périodiques évalués par les pairs, les sciences physiques et biomédicales se démarquant respectivement avec des pourcentages de 70 % et 88 %. 80 % des références bibliographiques citées dans les thèses sont couvertes par les collections de bibliothèques et les bases de données payantes tandis que 57 % sont accessibles gratuitement sans l'aide de la bibliothèque.

1.3. La part de Google Scholar

Les catalogues des bibliothèques, ainsi que les BD bibliographiques ne sont plus le premier choix pour les utilisateurs dans la recherche de littérature scientifique (Jamali & Asadi, 2010; Rowlands et al., 2008).

D'après Gullbekk et al., (2013), pour les doctorants, l'aspect le plus important lors du choix des sources d'information est la facilité d'accès aux documents en texte intégral, ceux-ci ayant réduit l'utilisation des ressources imprimées.

Une étude effectuée par Van Noorden (2014) sur plus de 3 500 répondants issus de 93 pays indique que plus de 90 % des chercheurs connaissent GS. Les auteurs précisent aussi que plus de 60 % des chercheurs ST, et plus de 70 % des SHS utilisent régulièrement GS pour effectuer leurs recherches bibliographiques.

Une autre étude réalisée par Bøyum & Aabø (2015) conclut que GS est perçu comme un outil de recherche bibliographique pratique et est largement utilisé par les doctorants en Norvège.

Il semble raisonnable de penser qu'un usage répandu de GS devrait avoir un impact sur les citations apparaissant dans les bibliographies.

Très récemment, Cole et al., (2018) se sont basés sur les bibliographies de dissertations en ingénierie issues de Proquest Dissertations & Theses pour la comparaison des bases de données payantes, Compendex et Scopus vs GS. Il s'avère que les articles de périodiques dominent largement les autres types de documents cités, et que GS surpassé Compendex et Scopus dans la couverture des citations dans les neuf champs de l'ingénierie traités.

Judit Bar-Ilan (2010) a analysé en détail la couverture des citations du livre « *Introduction to informetrics* » de Leo Egghe et Ronald Rousseau, par WoS, GS et Scopus. Pour ce cas, GS fournit bon nombre de citations introuvables dans WoS et Scopus, bien que seulement un quart de celles-ci proviennent de périodiques,

dont seulement 18 titres sur 28 sont *peer-reviewed*. De nombreuses citations (43 sur 108) correspondent à des rapports, manuscrits, bulletins, des entrées encyclopédiques et des brevets. Cette étude constate également que la plupart des citations présentes dans WoS et Scopus sont également couvertes par GS. Elle conclut cependant qu'aucune base de données ne pourrait remplacer les autres.

1.4. Besoins documentaires et citations en géosciences

Deux études ont été effectuées afin de déterminer les périodiques que les principaux groupes d'utilisateurs des bibliothèques de géosciences (étudiants, doctorants et membres du corps professoral) utilisent. Ces deux études aboutissent à des résultats quelque peu contradictoires en ce qui concerne la variation des besoins documentaires en fonction des groupes d'usagers.

Ainsi, Noga et al., (1993) ont comparé les citations des périodiques et des séries de monographies retrouvées dans les bibliographies des publications du corps professoral ainsi que des thèses et des mémoires d'étudiants en géosciences des universités UCLA et Stanford pendant la période 1990-1991. Il s'est avéré que les citations du corps professoral étaient différentes de celles des travaux d'étudiants et que les unes ne pouvaient donc pas être utilisées comme substitut des autres pour la gestion des collections. Les listes de citations obtenues révélaient un besoin différent pour chaque groupe d'utilisateurs.

En se basant sur la méthodologie et les résultats de Noga et al., (1993) et de McCain & Bobick (1981) et en appliquant le coefficient de corrélation de Kendall, Zipp (1995) a également effectué une comparaison détaillée des citations des thèses et des mémoires avec les citations des publications du corps professoral en géosciences et Biologie, cette fois à Iowa State University. Contrairement aux résultats de Noga et al., (1993), elle conclut : “*The most heavily cited journal titles in theses and dissertations can be used as a surrogate for the titles most heavily used by faculty in their publications*” (p. 341).

Outre ces résultats contradictoires, Noga et al., (1993) ont déterminé que les deux périodiques les plus cités par les doctorants des deux universités, *Journal of Geophysical Research* et le *Geological Society of America Bulletin*, sont des titres représentatifs des collections géo-scientifiques. Les autres titres cités sont variables en fonction des centres d'intérêt des recherches de chaque établissement et peuvent changer dans le temps.

Aux Etats-Unis, certains travaux se sont aussi exclusivement consacrés à l'étude des citations de thèses, dans le domaine des géosciences (Walcott (1992)) ou des Sciences atmosphériques (Kaczor (2014)). Walcott (1992) a travaillé sur un échantillon de 500 thèses en provenance de différentes Institutions et défendues entre 1981 et 1985 en géosciences. 59,3 % des citations étaient associées à la géologie, 96,6 % étaient de langue anglaise, 64,3 % ont été publiées aux États-Unis.

Objectifs et hypothèses

Une période de trois décennies couvrait 90 % de l'ensemble des citations et le nombre total de périodiques cités était de 836.

Kaczor (2014) a analysé les citations de 29 thèses défendues entre 2000 et 2010 dans le département de Sciences Atmosphériques et Environnementales à l'Université d'Albany. Les 2 806 articles cités ont été publiés dans 244 périodiques différents tandis que 26 titres suffisaient à couvrir 80,3 % de toutes les citations d'articles. Les articles publiés dans les 9 ans précédant la date de présentation de la thèse représentaient 52,6 % de toutes les citations correspondant aux articles de périodiques.

Malgré le nombre d'années qui séparent les travaux, les résultats s'accordent quant au pourcentage élevé d'articles cités : 79,6 % pour Walcott (1992) et 85,1 % pour Kaczor (2014).

2. Objectifs et hypothèses

Le premier objectif de ce chapitre est d'établir les profils des ressources utilisées par trois doctorants en Géographie sur base des citations relevées dans leurs thèses.

Le second objectif est de vérifier le référencement des échantillons de citations par différents outils bibliographiques : Scopus, GeoRef, WoS et le moteur de recherche GS.

Pour déterminer quelle est la littérature supplémentaire utilisée par ces doctorants et qui n'est pas disponible dans les bases de données commerciales, nous analysons les périodiques propres aux différents outils bibliographiques, notamment ceux de GS.

Sur base de la revue de la littérature et des résultats des chapitres 1 et 2, nous envisageons que GS soit l'outil de recherche bibliographique le plus complet en termes de contenu et nous posons les hypothèses suivantes :

Hypothèse 1 : Les doctorants en Géographie citent plusieurs types de documents dont majoritairement des articles de périodiques en anglais ;

Hypothèse 2 : La loi de Bradford, selon laquelle 80 % des citations d'articles proviennent de 20 % des titres de périodiques, est respectée ;

Hypothèse 3 : GS permet de retrouver une proportion nettement plus élevée des citations issues des trois thèses que les bases de données bibliographiques commerciales ;

Hypothèse 4 : GS retrouve la majorité des citations couvertes par les autres outils lesquels présentent peu de résultats « uniques » ;

Hypothèse 5 : Parmi les citations uniquement retrouvées par GS, outre des articles de périodiques, on retrouve d'autres types de documents cités ;

Hypothèse 6 : Le chercheur en Géographie utilise encore des documents issus de bibliothèques locales ou obtenus *via* son réseau de relations ou repérées au cours de

ses lectures, qui ne sont pas répertoriés par GS ou les principales BDs commerciales ;

3. Méthodes

Les données analysées dans ce chapitre sont constituées par les références bibliographiques issues de trois thèses :

- Belleflamme, A., 2015, *Detection of past and future atmospheric circulation changes over the North Atlantic region with the help of an automatic circulation type classification*
La thèse sera désignée ci-après par la discipline principale dont elle relève : « Climatologie »
- Dubois, C., 2014, *Quels agritourismes pour les campagnes périurbaines ? Les cas de la Wallonie et du Grand-Duché de Luxembourg*
La thèse sera désignée ci-après par la discipline principale dont elle relève : « Tourisme »
- Trotta, M., 2014, *Modelling serial offenders' spatial behaviours: new assumptions for geographic profiling*
La thèse sera désignée ci-après par la discipline principale dont elle relève :
« Analyse spatiale » (pour « Analyse spatiale appliquée à la criminologie »)

Au moment de la mise en place de notre méthodologie notre choix s'est porté sur les thèses les plus récentes présentées dans le département de Géographie de l'ULiège, soit ces trois dissertations présentées durant l'année académique 2014-2015, respectivement en Climatologie, Tourisme et Analyse spatiale.

Les références bibliographiques de chaque thèse nous ont été fournies par les doctorants en format *BibTeX* ou *EndNote* et nous les avons rassemblées dans *EndNote*. Puis, les références ont été exportées dans trois fichiers *Excel*.

Afin de pouvoir repérer qualitativement les variations éventuelles apparaissant entre les résultats liés à chaque thèse, nous avons choisi de ne pas regrouper les données issues des trois thèses mais de les examiner séparément. Le but est essentiellement d'éviter de tirer des conclusions trop hâtives et non-fondées sur base d'un échantillon dont la moyenne n'est pas représentative vu sa petite taille.

Pour chaque référence bibliographique, l'année de publication étant connue, nous avons complété le fichier *Excel* avec l'indication du type de document (article de périodique, livre, chapitre de livre, acte de colloque, thèse ou mémoire, autre type de document). La langue des documents a également été identifiée.

Les périodiques dont sont issus le plus grand nombre d'articles cités ont été identifiés et le suivi de la loi de Bradford a été vérifié.

Pour ce qui concerne le **référencement par les outils bibliographiques**, chaque citation des échantillons a été recherchée directement dans GS, Scopus et GeoRef.

Méthodes

Pour WoS, qui n'était plus accessible en Communauté française de Belgique au moment où l'étude a été réalisée, nous nous sommes référée à un fichier *Excel*, fourni par *Thomson Reuters*⁴⁸ et qui reprenait les périodiques indexés par la base. Enfin, nous avons choisi de ne pas continuer à effectuer les recherches dans Francis qui n'était plus mis à jour. A l'exception de WoS et Francis, la méthodologie de recherche appliquée est identique à celle que nous avons utilisée dans le chapitre 1 de la première partie de la thèse.

Pour chacune des thèses, les nombres et pourcentages de documents (tous types confondus) et d'articles couverts par chaque outil bibliographique ont été calculés. Pour visualiser le recouvrement des résultats des différents outils bibliographiques et leur apport « unique », les « overlap » ont été représentés à l'aide de diagrammes Venn.

Enfin, dans le but d'appuyer les résultats issus de nos recherches sur les trois bibliographies, des entretiens ont été effectués avec les doctorants. Un questionnaire composé de sept questions (en Annexe 1) leur a été soumis oralement. Les réponses alimenteront la discussion à la fin de ce chapitre.

⁴⁸ Au moment où nous avons réalisé nos recherches, WoS était toujours produit par Thomson Reuters

4. Résultats

4.1. Type de documents cités

Le nombre total de références bibliographiques citées dans chaque thèse est de 126 pour la Climatologie, 448 pour le Tourisme et 215 pour l'Analyse spatiale, soit en tout 789 références pour notre échantillon, qui se répartissent entre six catégories de types de documents (Figure 6).

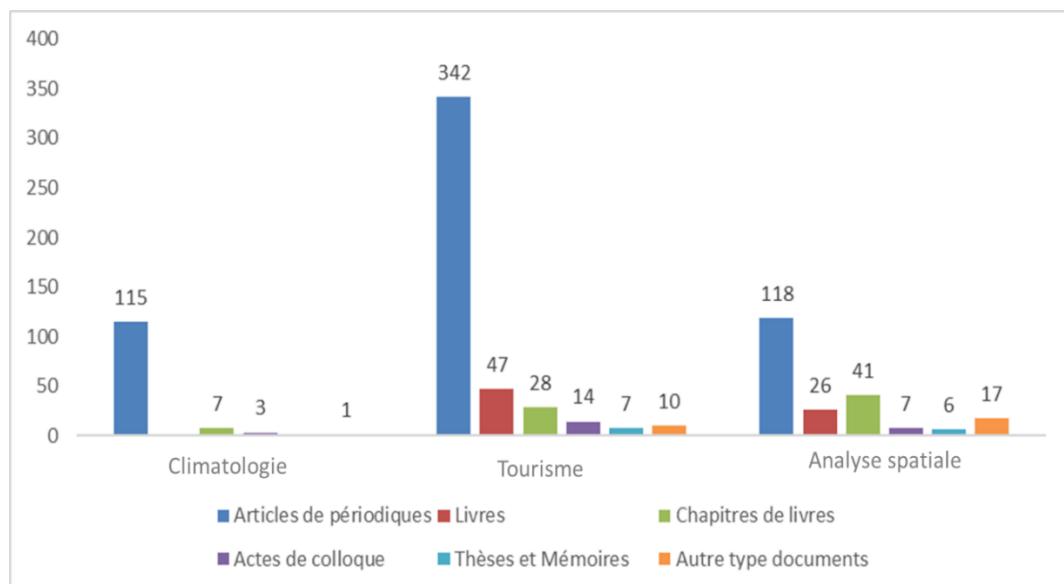


Figure 6 - Types et nombres de documents cités dans les bibliographies de trois thèses

4.1.1. Citation d'articles de périodiques

La majorité des références sont des articles de périodiques (Figure 6), soit 73 % (576) pour l'ensemble de notre échantillon et dans des proportions variables pour les trois thèses : 91% (115) pour la bibliographie en Climatologie, 76,5 % (342) pour le Tourisme et 55 % (118) pour l'Analyse spatiale.

Avec 69,9 % d'articles dans les bibliographies de thèses ayant trait à plusieurs domaines des sciences (Mathématiques, Chimie, Géographie physique, Informatique et Physique), Grigas et al., (2017) ont récemment trouvé une valeur proche de la valeur moyenne de notre échantillon. Globalement, nos résultats peuvent aussi être rapprochés de ceux trouvés par Conkling et al., (2010) avec une moyenne de 79 % d'articles de périodiques pour les sciences, ou de ceux de Walcott (1992) selon lequel 80 % de références citées dans les thèses américaines en géosciences était représentées par des articles de périodiques.

Cependant, si nous considérons individuellement les trois thèses que nous étudions, nous constatons que les pourcentages s'écartent de ces moyennes pour la Climatologie (91%) et pour l'Analyse spatiale (55%). Seuls les chiffres obtenus pour le Tourisme (76,5 %) s'alignent avec ceux de la littérature précitée.

Résultats

Pour les sciences atmosphériques Kaczor, (2014) indiquait que 52,6 % de toutes les citations correspondaient à des articles de périodiques, ce qui s'accorde à nos résultats pour l'Analyse spatiale, mais s'éloigne beaucoup de nos résultats en Climatologie.

Notons aussi que, dans une étude réalisée sur 26 thèses présentées en criminologie entre 1991 et 2004, Sexton (2006) ne recensait que 24,5 % d'articles des périodiques. Même si ce pourcentage est nettement moins élevé que celui que nous avons trouvé pour l'Analyse spatiale, nous relevons quand-même que dans notre échantillon, la thèse en Analyse spatiale appliquée à la criminologie était aussi celle qui présentait la plus faible proportion d'articles. Pour la criminologie, qui fait appel autant aux sciences qu'aux sciences humaines et sociales (psychologie), les pratiques de citation des SHS influencerait la balance entre les articles de périodiques et les livres cités (Grigas et al., 2017; Rosenberg, 2015; Sexton, 2006).

4.1.1.1. Périodiques cités et la loi de Bradford

Les articles cités dans les trois thèses sont issus respectivement de 30 périodiques pour la Climatologie, 111 pour le Tourisme et 62 pour l'Analyse spatiale (un titre est commun aux trois bibliographies, un titre est commun à la Climatologie et au Tourisme, un titre à la Climatologie et à l'Analyse spatiale et deux titres au Tourisme et à l'Analyse spatiale).

Les titres de périodiques les plus utilisés, dont trois articles au moins sont cités dans les thèses étudiées, sont repris par ordre décroissant du nombre de citations dans le tableau 11. Les titres non repris dans les tableaux sont cités moins de trois fois, la grande majorité d'entre eux étant cités une seule fois.

Comme cela peut se remarquer à la lecture des titres du Tableau 8, la majorité des périodiques les plus utilisés dans les trois thèses sont associés aux disciplines de ceux-ci, soit la Climatologie, le Tourisme et l'Analyse spatiale appliquée. Le nombre de citations associées à ces titres suggère qu'ils pourraient faire partie de principaux titres des champs de recherche concernés. Nous avons effectivement pu vérifier que la plupart de ces titres faisaient partie du Q1 du SJR⁴⁹), comme déjà indiqué dans la littérature par (Kaczor, 2014).

Nos résultats ne suivent pas la **loi de Bradford**. En effet, pour couvrir 80 % des articles cités dans chaque thèse, il faut au moins 36,7 % des titres de périodiques pour la Climatologie, 40 % pour le Tourisme, et 66 % pour l'Analyse spatiale, soit beaucoup plus que les 20 % escomptés.

La distribution des citations dans les trois thèses étudiées s'accorde plutôt à celle de Walcott (1992) qui a indiqué qu'un plus grand nombre de périodiques était

⁴⁹ Informations relevées de SJR et non représentées dans les résultats

Résultats

nécessaires pour couvrir la majorité des citations d'articles, 90 % des citations provenant de 53,5 % périodiques en Géosciences.

Titre de périodiques/ Climatologie	Nombre de citations	% cumulatif des citations
<i>Journal of Climate</i>	21	18.2
<i>International Journal of Climatology</i>	16	32.1
<i>Geophysical Research Letters</i>	16	45.2
<i>The Cryosphere</i>	8	52.1
<i>Climate Dynamics</i>	8	59.1
<i>Journal of Geophysical Research</i>	5	63.4
<i>Bulletin of the American Meteorological Society</i>	5	67.7
<i>Theoretical and Applied Climatology</i>	5	72
<i>Quarterly Journal of the Royal Meteorological Society</i>	4	75.5
<i>Climate Research</i>	3	78.1
<i>Nature</i>	3	80.7
Tourisme		
<i>Annals of Tourism Research</i>	48	14
<i>Journal of Travel Research</i>	47	27.7
<i>Tourism Management</i>	41	39.7
<i>Sociologia Ruralis</i>	21	45.8
<i>International Journal of Tourism Research</i>	16	50.5
<i>Journal of Sustainable Tourism</i>	15	54.9
<i>Journal of Rural Studies</i>	14	59
<i>European Countryside</i>	7	61
<i>Progress in Human Geography</i>	4	62.2
<i>Journal of Vacation Marketing</i>	4	63.4
<i>Bulletin de la Société Géographique de Liège</i>	4	64.6
<i>Journal of Agricultural Economics</i>	3	65.5
<i>Economie rurale</i>	3	66.4
Analyse spatiale		
<i>Journal of Investigative Psychology and Offender Profiling</i>	17	14.5
<i>Criminology</i>	9	22.2
<i>Journal of Quantitative Criminology</i>	8	29
<i>Journal of Environmental Psychology</i>	5	33.3
<i>Psychology, Crime & Law</i>	4	36.7
<i>The Professional Geographer</i>	3	39.3

Tableau 11 - Titres de périodiques les plus utilisés, nombre de citations associées et pourcentage cumulatif de citations pour les trois thèses

A côté de la loi de Bradford qui considère uniquement les pourcentages de titres, on peut remarquer que dans l'absolu, il suffit de peu de titres de périodiques pour couvrir une bonne proportion des articles cités : pour la Climatologie 2 périodiques regroupent 32 % des citations et 4 périodiques regroupent 52 % des citations, pour le Tourisme 3 titres regroupent 39,7 % des citations et 5 titres regroupent 50,5 % de citations tandis que pour l'Analyse Spatiale 4 titres regroupent 33,3 % des

Résultats

citations (Tableau 8). Ces chiffres sont relativement proches de ceux trouvés par Kaczor (2014), chez qui 6 titres représentaient 51,4 % de toutes les citations d'articles pour les thèses en Sciences Atmosphériques, ainsi que de ceux indiqués par Crissinger (1981), chez qui 5 titres regroupaient 31 % des citations d'articles des périodiques en Géologie.

4.1.1.2. Années de publication des articles cités

La période de publication des articles cités dans les trois bibliographies s'étend globalement de 1932 à 2015, l'étendue dans le temps différant pour chaque thèse, comme cela peut être remarqué sur la figure 7.

Ainsi pour la Climatologie les années de publication des articles cités vont de 1963 à 2015, avec une moyenne de 4,6 articles par année (médiane : 3 articles), et un âge moyen des articles de 7 ans. Pour le Tourisme, les années de publication des articles s'étendent de 1980 à 2014, avec une moyenne de 12 articles par année (médiane 9) et un âge moyen des articles de 10 ans. Quant à l'Analyse spatiale, les années de publication des articles vont de 1932 à 2014, avec une moyenne de 3,6 articles par année (médiane : 3 articles), et un âge moyen des articles de 12 ans.

L'âge moyen des articles cités dans la thèse en Tourisme et en Analyse spatiale sont de l'ordre de ceux indiqués par Larivière *et al.* (2013), qui étaient de 10 ans pour les S&T et 12 ans pour les SHS. La thèse en Climatologie s'écarte de ces valeurs avec un âge moyen des articles plus court (7 ans).

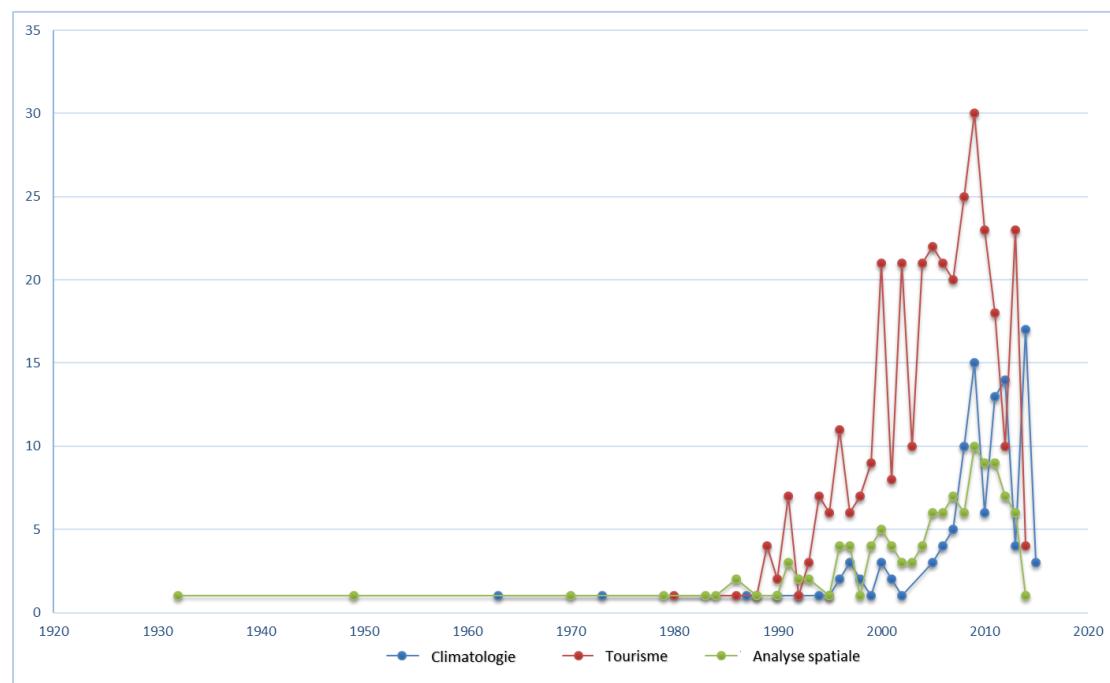


Figure 7 - Années de publication des articles cités dans les trois bibliographies

Une concentration allant de 60 % à 79 % des références peut être remarquée dans les dix ans précédant la défense de la dissertation : 91 articles (79 %) pour la

Climatologie, 217 articles (63,4 %) pour le Tourisme et 71 articles (60 %) pour l'Analyse spatiale. D'autres auteurs (Meier & Conkling, 2008; Nabe & Imre, 2008; Rosenberg, 2015) s'accordent sur le fait que la majorité des articles cités dans les bibliographies des thèses datent des deux dernières décennies, avec des pourcentages de concentration du même ordre de grandeur et une concentration plus importante sur la première décennie avant l'année de défense de la thèse .

Dans son étude sur les thèses en géosciences (Walcott, 1992), Walcott révélait que 23 à 28 ans étaient nécessaires pour couvrir 90 % de toutes les citations, bien que la paléontologie ait besoin de 72 ans. Pour notre échantillon 16 ans sont nécessaires pour couvrir 90 % de citations d'articles en Climatologie, 19 ans en Tourisme et 24 ans en Analyse spatiale.

La période couverte par la bibliographie d'une thèse varie logiquement en fonction du sujet de recherche de chaque dissertation. Ainsi, la plus ancienne citation d'article de la thèse sur le Tourisme, date de 1980, année qui correspond au début des recherches sur l'agritourisme selon l'auteur de la thèse.

4.1.2. Citation des livres et chapitres de livres

Les livres et chapitres de livres sont moins bien représentés que les articles (Figure 6). Leur usage n'est cependant pas insignifiant pour deux de trois thèses et pourrait être dû à l'orientation des thèses vers les SSH : 10 % (47 livres) et 6,2 % (28 chapitres de livres) pour le Tourisme, 12 % (26 livres) et 19 % (41 chapitres de livres) pour l'Analyse spatiale. Ces résultats sont un peu moins élevés que ceux mentionnés dans la littérature : Grigas et al., (2017) ont trouvé des valeurs plus élevées avec 28 % des livres et chapitres de livres en SHS, 17,7 % en Sciences et techniques et 15,4 % en Sciences physiques. Romic & Mitrović (2014), quant à eux, indiquaient 34 % de livres cités dans les thèses à l'Université de Zagreb.

5,5 % seulement (7 références) de la bibliographie en Climatologie sont des chapitres de livres. Pour le domaine des Sciences Atmosphériques, cette valeur correspond à celle trouvée par Kimball et al., (2013) avec 5 % de livres cités dans les bibliographies des chercheurs et Kaczor (2014) qui a trouvé un pourcentage un peu plus élevé de 7,5 % de citations correspondant à des livres.

Pour l'ensemble des trois thèses, les années de publication de livres et chapitres de livres ont un étalement dans le temps semblable à celui des articles et vont de 1927 jusqu'à 2015 (Figure 8). L'examen de la Figure 8 nous permet de constater que les 2 thèses qui relèvent des SHS utilisent ponctuellement une littérature plus datée que la thèse en Climatologie. Mais cette fois, ce sont les citations de la thèse en Tourisme qui remontent le plus loin dans le temps, jusqu'à 1927.

La constante qui peut être relevée est que la thèse de Climatologie fait référence à une bibliographie plus récente, qu'il s'agisse d'articles de périodiques, de livres ou de chapitres de livres.

Résultats

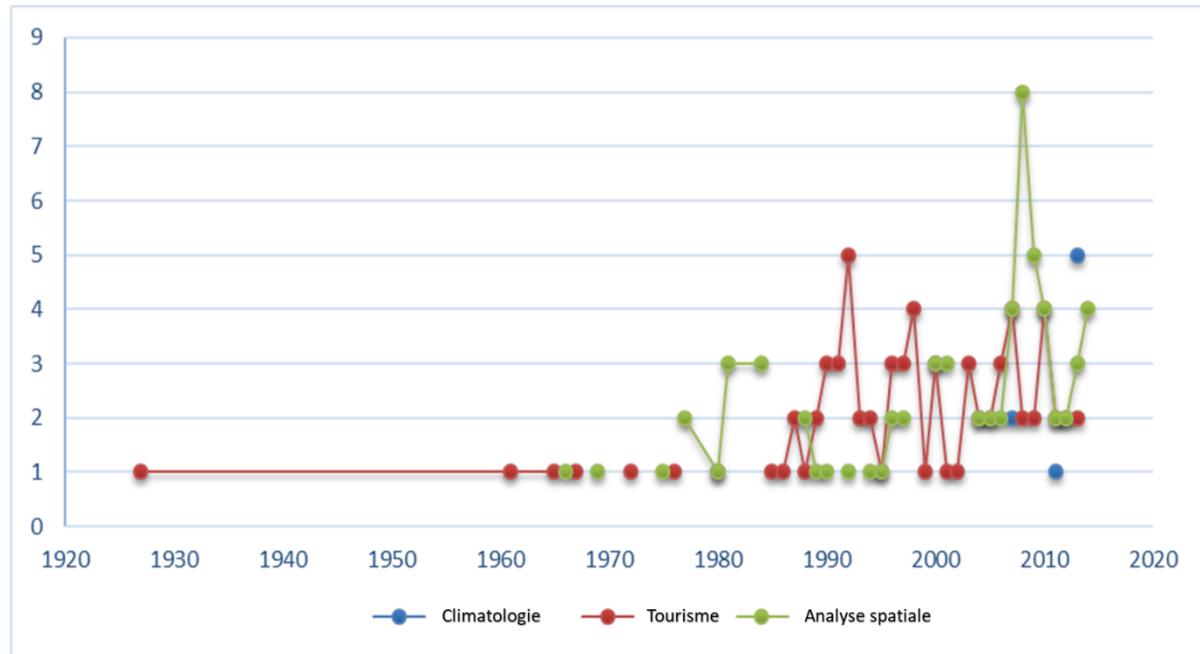


Figure 8 - Année de publication des livres et chapitres de livres cités dans les trois bibliographies

4.1.3. Autres type de documents cités

Les autres types de documents cités, principalement de la littérature grise (actes de conférences, thèses, rapports, travaux non publiés), sont beaucoup moins représentés avec des valeurs < 5 %, comme indiqué dans la Figure 6, et comme rapporté par la majorité des études similaires (Conkling et al., 2010; Feyereisen & Spoiden, 2009; Rosenberg, 2015; Walcott, 1992).

L'étendue des années de publication de la littérature grise est de 1994 à 2014 pour l'ensemble des trois dissertations, soit deux décennies précédant la défense des thèses.

4.2. Langue des citations

Dans notre échantillon, deux thèses, celles en Climatologie et en Analyse spatiale, ont été rédigées en anglais. La thèse en Tourisme a été rédigée en français.

En ce qui concerne les bibliographies des trois thèses considérées dans leur ensemble, la langue des documents cités est majoritairement l'anglais avec une moyenne de plus de **90 %**, soit 98,4 % (124 références) en Climatologie, 87,7 % (393 références) en Tourisme et 96,3 % (207 références) en Analyse spatiale.

L'anglais reste majoritaire quand on prend en considération les articles de périodiques uniquement : 98,2 % des articles (113 références) en anglais pour la Climatologie, 91,5 % des articles (313 références) en anglais pour le Tourisme, et 99 % articles en anglais (117 références) pour l'Analyse spatiale.

La thèse rédigée en français, dont le champ d'étude se situe en Wallonie et au Grand-Duché de Luxembourg, présente le plus de références de langue française soit 11% (49) de la bibliographie en Tourisme. Les références de langue française ne représentent que 3,7 % (8) de la bibliographie en Analyse spatiale, et seulement 1,6 % (2) de la bibliographie en Climatologie.

La thèse en Tourisme est également l'unique thèse dans laquelle on retrouve quelques documents en d'autres langues que l'anglais ou le français : 4 références en allemand et 2 en espagnol, soit 1,3 % de la totalité des documents.

Comme la majorité des recherches publiées se sont effectuées dans des universités des Etats-Unis, la variable langue n'a pas souvent été prise en compte. Néanmoins, les pourcentages relevés dans les trois thèses de notre Institution sont du même ordre de grandeur que ceux indiqués par Walcott (1992), qui avait trouvé 96, 6 % des citations en anglais dans une étude réalisée sur 500 thèses en géosciences aux Etats-Unis. Quant à Rosenberg (2015), il indique des pourcentages de 90 % d'usage de périodiques en anglais pour la Sociologie et l'Anthropologie à l'Université de Jérusalem.

4.3. Référencement des citations dans les outils bibliographiques

Tous types de documents confondus (Tableau 12, 1^{ère} et 2^{ème} colonnes), GS a retrouvé la majorité des références citées dans des proportions comprises entre 100 % pour la bibliographie de la thèse en Climatologie et 84,5 % pour celle en Tourisme. La différence s'explique essentiellement par une couverture partielle des livres et chapitres de livres, relativement nombreux pour la thèse en Tourisme. Il faut cependant noter que les livres et chapitres de livres, tout aussi nombreux dans la thèse en Analyse spatiale, ont bien été retrouvés dans GS.

Considérant les articles de périodiques uniquement (Tableau 12, 3^{ème} et 4^{ème} colonnes), la performance de GS est encore plus élevée avec des pourcentages de 100 % pour les bibliographies des thèses en Climatologie et en Analyse spatiale et de 98 % pour celle en Tourisme.

Scopus se classe en deuxième position, tant pour retrouver les documents de tout type que lorsqu'on considère spécifiquement les articles. Les performances de Scopus sont cependant nettement en-dessous de celles de GS quand il s'agit de tous les types de documents, descendant jusque 44 % pour la thèse en Analyse spatiale, laquelle présente relativement beaucoup de livres et de chapitres de livres que GS trouve mais qui ne sont pas référencés par Scopus. De même pour les articles de la thèse en Tourisme, avec seulement 75 % de couverture par Scopus.

WoS se classe en troisième position, juste après Scopus avec des performances variant de 43 % à 89 % pour ce qui concerne l'ensemble des documents, mais avec une perte de performance encore plus importante par rapport à GS pour les articles

Résultats

cités dans la thèse en Tourisme (59 %). La supériorité de Scopus par rapport à WoS ne semble confirmée ici que par les valeurs trouvées pour la thèse en Tourisme.

GeoRef indexe le moins de références avec des valeurs nettement inférieures aux autres outils bibliographiques : au mieux 13 % pour les articles cités dans la thèse en Climatologie et aucune référence pour le Tourisme et l'Analyse spatiale. Ces valeurs reflètent le peu de couverture de la Climatologie, et la non-couverture de Tourisme et de l'Analyse spatiale par GeoRef.

Malgré le faible nombre de références associé à la littérature grise (actes de conférences, thèses, rapports), il apparaît que ce type de littérature est couvert uniquement par GS, dans de bonnes proportions (Tableau 12, 5^{ème} à 9^{ème} colonnes).

On remarque aussi la supériorité de GS en terme de recensement des livres et chapitres de livres (Tableau 12, 2 dernières colonnes). Toutefois, la couverture n'est pas complète et même très partielle dans le cas des citations de la thèse en Tourisme.

	No. of Bibliogr. Ref.	% indexed in BT	Journal articles	% of JA indexed in BT	Conf. Proc.	Theses	Reports	Unpubl. work	Online course	Computer Program	Books	Book Chapter
Climatologie	126		115		3	--	1	--	--	--	--	7
GS	126	100	115	100	3	--	1	--	--	--	--	7
Scopus	114	90	110	96	--	--	--	--	--	--	--	4
WoS	112	89	110	96	--	--	--	--	--	--	--	2
GeoRef	17	13	15	13	--	--	--	--	--	--	--	2
Tourisme	448		342		14	7	8	2	--	--	47	28
GS	379	84.5	335	98	14	6	3	1	--	--	10	9
Scopus	263	58	260	75	--	--	--	--	--	--	3	--
WoS	208	46	206	59	--	--	--	--	--	--	1	1
GeoRef	0	--	--	--	--	--	--	--	--	--	--	--
Analyse spatiale	215		118		7	6	8	5	1	3	26	41
GS	201	93	118	100	7	3	7	3	--	--	24	39
Scopus	95	44	94	79	--	--	--	--	--	--	--	1
WoS	93	43	85	71	--	--	--	--	--	--	8	--
GeoRef	1	0.4	--	--	--	--	--	--	--	--	--	--

Tableau 12 - Références bibliographiques et type de documents de chaque thèse : nombre et pourcentage indexés par les outils bibliographiques.

Résultats

4.4. Analyse « overlap » entre l'ensemble des outils bibliographiques et apports spécifiques en références « uniques »

Pour la bibliographie de chacune des trois thèses étudiées, les diagrammes de Venn nous permettent d'illustrer les taux de couverture et de recouvrement de tous les outils bibliographiques interrogés. Ils permettent également de visualiser l'éventuel apport spécifique de chaque outil bibliographique par rapport à un ou plusieurs autres outils.

Pour la bibliographie de la thèse en **Climatologie** (Figure 9), 126 références, soit la totalité, sont retrouvées par GS. 10 de ces références sont exclusivement retrouvées par GS (dont 5 articles, 1 chapitre de livre, 3 comptes rendus des conférences et 1 rapport) tandis que 116 références sont retrouvées par au moins un deuxième des outils bibliographiques.

110 références sont communes à WoS, Scopus et GS dont 15 seulement sont aussi répertoriées par GeoRef. Scopus se distingue de WoS par 4 références supplémentaires et, à l'inverse WoS se distingue de Scopus par 2 références supplémentaires. Même GeoRef se distingue de Scopus par 2 références supplémentaires qui sont par contre communes avec WoS. Toutes ces références supplémentaires sont des chapitres de livres et toutes sont retrouvées par GS.

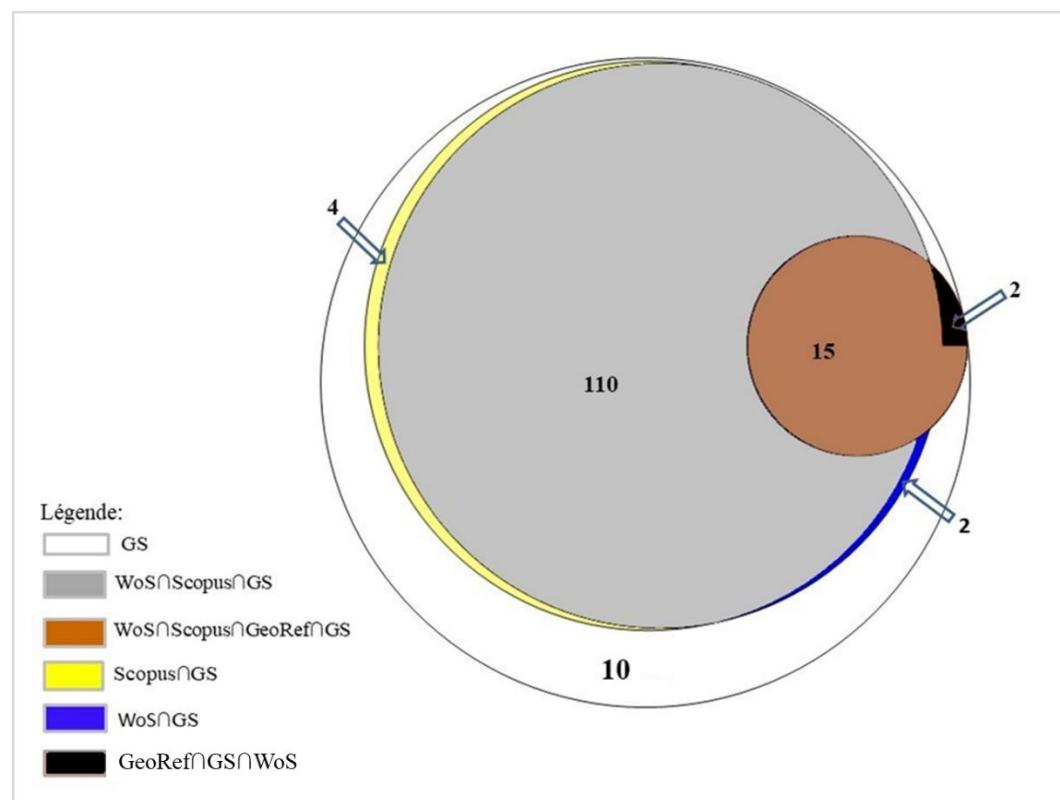


Figure 9 - Overlap des références dans les quatre outils bibliographiques pour la bibliographie en Climatologie

Pour la bibliographie de la thèse en **Tourisme** (Figure 10), qui représente un total de 448 références, GS recense 379 références, soit le nombre le plus important de

références de tous les outils bibliographiques analysés. 111 de ces références sont exclusivement retrouvées par GS, dont 68 articles et 19 livres ou chapitres de livres.

269 références sont retrouvées par au moins deux des outils bibliographiques tandis que 198 références sont communes à tous les outils, abstraction faite de GeoRef qui ne comprend aucune référence de cette bibliographie.

WoS et Scopus ont en commun 200 références. Scopus se distingue de WoS par 63 références supplémentaires dont 62 se retrouvent dans GS. Réciproquement, WoS se distingue de Scopus par seulement 8 références supplémentaires dont 7 se retrouvent dans GS. Pour l'essentiel ces références supplémentaires sont des articles de périodiques.

Sur les trois thèses, celle en Tourisme est le seul cas où la bibliographie comporte quelques références que GS ne recouvre pas mais qui sont répertoriées par au moins une des deux bases de données commerciales, soit 1 référence de livre propre à Scopus, 1 article propre à WoS et 2 références (un chapitre de livre et un livre) communes à Scopus et WoS.

65 références de la bibliographie en Tourisme ne sont retrouvées par aucun des outils bibliographiques, et sont constituées essentiellement de livres (35 références) et chapitres de livres (18 références), 5 rapports, 5 articles de périodiques et 1 travail non publié. Des recherches ont été effectuées dans le catalogue de ULiège Library et il s'avère que 23 des 65 références sont accessibles *via* la bibliothèque institutionnelle. 28 de ces références sont répertoriées par Google Books (dont 9 accessibles également *via* la bibliothèque institutionnelle).

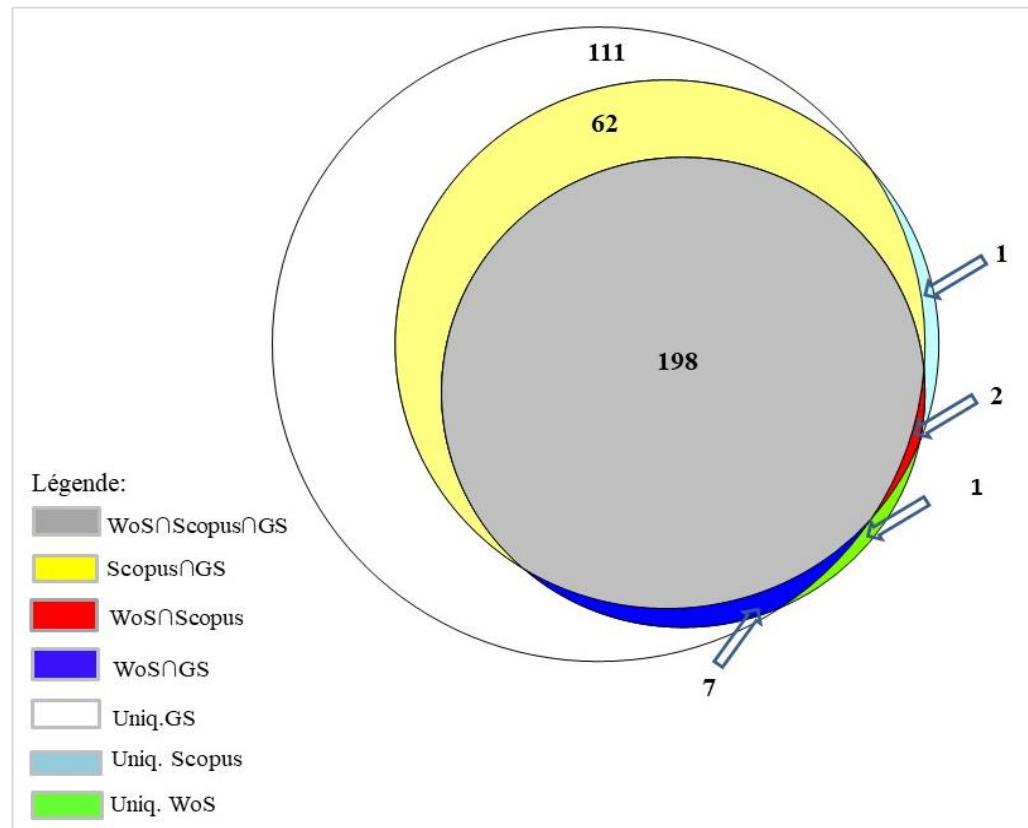


Figure 10 - Overlap des références dans les quatre outils bibliographiques pour la bibliographie en Tourisme

Résultats

D'un total de 215 références que représente la bibliographie de la thèse en **Analyse spatiale** (Figure 11), GS retrouve 201 références. 93 de ces références sont uniquement retrouvées par GS, dont 20 articles et 54 livres ou chapitres de livres.

108 des 201 références de GS sont retrouvées par au moins un deuxième des outils bibliographiques. 80 références sont communes aux trois outils bibliographiques : GS, Scopus et WoS. Il est à noter que pour la thèse en Analyse spatiale, GeoRef (non représenté dans la Figure 11) indexe 1 référence seulement, laquelle est retrouvée par tous les autres outils.

Scopus se distingue des WoS par 15 références qui sont toutes retrouvées par GS (13 articles, 1 proceeding et 1 chapitre de livre). A l'inverse, WoS se distingue de Scopus par 13 références qui sont toutes retrouvées par GS (5 articles et 8 chapitres de livres).

14 références de la bibliographie en Analyse spatiale ne sont retrouvées par aucun des outils bibliographiques. Il s'agit de livres et chapitres de livres, de rapports et de travaux non-publiés.

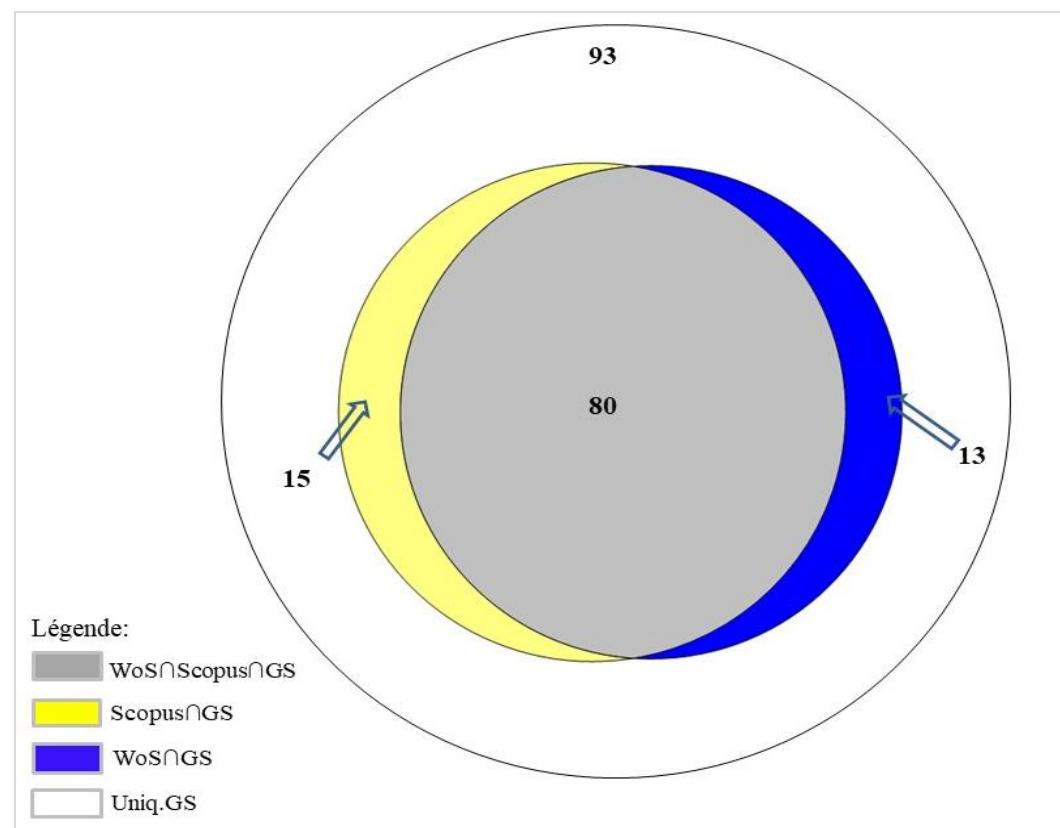


Figure 11 - Overlap des références dans les trois outils bibliographiques pour la bibliographie en Analyse spatiale

5. Discussion et conclusions

Le choix de l'échantillon nous a amené à travailler sur trois thèses aux sujets relativement éloignés et nous avons considéré les résultats obtenus indépendamment les uns des autres. Une certaine variété a été observée dans les résultats, qui peut éventuellement être mise en relation avec les orientations ST ou SHS des thèses et correspond bien à la diversité des pratiques de lecture et de publication des sous-disciplines de la Géographie. Cette variété ne nous met pas en porte à faux quand il s'agit de répondre à nos hypothèses de départ, ce qui conforte nos conclusions faute d'avoir analysé un nombre réellement représentatif de thèses.

En référence à la **première hypothèse** posée dans ce chapitre, nos résultats montrent que plusieurs types de documents sont cités par les doctorants géographes, notamment par deux des doctorants, Dubois (Tourisme) et Trotta (Analyse spatiale), dont les sujets de recherche sont orientés vers les SHS. La distribution des types de documents utilisés dans les trois thèses est différente en fonction de l'orientation du sujet de recherche vers les SHS ou ST. Comme mentionné précédemment les pratiques de citation spécifiques aux SHS (Grigas et al., 2017; Rosenberg, 2015; Sexton, 2006) influencent la balance entre les articles de périodiques et les livres cités. Cependant, les citations correspondant aux articles de périodiques et en langue anglaise sont majoritaires pour les trois thèses.

La première hypothèse selon laquelle *les doctorants en Géographie citent plusieurs types de documents dont majoritairement des articles de périodiques en anglais* ne peut être rejetée.

Les autres types de documents cités, principalement des livres et chapitres de livres et dans une moindre mesure de la littérature grise, sont beaucoup moins nombreux et rejoignent les résultats rapportés par d'autres études similaires (Kaczor, 2014; Romić & Mitrović, 2014; Rosenberg, 2015; Walcott, 1992).

Interrogés sur leur usage préférentiel d'articles de périodiques (voir questionnaires en annexe, question 7), les « doctorants » font la différence entre le contenu généraliste des livres et l'actualité d'un contenu spécialisé pour les articles de périodiques. Selon eux, la recherche est étroitement liée aux informations transmises par les articles de périodiques même si la thématique de recherche peut avoir une influence. Enfin, la facilité d'accès aux articles est également évoquée en comparaison d'un accès moins aisément aux livres et chapitres de livres.

En relation avec la **deuxième hypothèse**, nos résultats indiquent que des pourcentages de périodiques supérieurs à 20 % des périodiques utilisés sont nécessaires pour couvrir 80 % des articles cités dans chaque thèse, en accord avec les résultats obtenus par Walcott (1992) en Géosciences.

La deuxième hypothèse selon laquelle *80 % des citations d'articles proviennent de 20 % des titres de périodiques suivant la loi de Bradford* n'est pas confirmée.

Discussion et conclusions

Même si la petite taille de notre échantillon ne nous permet pas de l'affirmer, il n'est pas exclu qu'un lien existe entre ces pourcentages et le caractère plutôt ST ou SSH (pourcentages de périodiques les plus faibles) de la recherche. En effet, quelques études, spécifiquement en sciences, s'accordent sur le fait que la citation d'articles de périodiques suit la loi de Bradford (Nisonger, 2008; Trueswell, 1969). Par contre, des exceptions à la loi de Bradford sont mentionnées dans des études sur les citations en SHS (Rosenberg, 2015; Sexton, 2006).

Il serait intéressant d'effectuer des comptages sur un nombre plus élevé de thèses en Géographie afin de vérifier que les pourcentages de périodiques concentrant les citations dans les sous-disciplines, Géographie physique ou Géographie humaine, font - ou pas - exception à cette loi.

La Géographie présente aussi la spécificité de souvent étudier une thématique dans un contexte régional, ce qui oblige le chercheur à combiner des lectures sur le thème et sur la région, ce qui accroît nécessairement la liste des périodiques.

Pour ce qui est de la **troisième hypothèse**, en terme de couverture des références, GS est l'outil bibliographique le plus complet, ayant retrouvé la majorité des références recherchées avec des valeurs proches des 100 % pour les bibliographies de trois thèses tous types de documents confondus. Les bases de données WoS et Scopus se comportent relativement bien, mais avec des valeurs moins élevées par rapport à GS, de l'ordre des 90 % pour la thèse à tendance ST et des 60 % à 80 % pour les thèses orientées SHS. GeoRef retrouve beaucoup moins de résultats et ces derniers s'apparentent à zéro pour les thèses à orientation SHS.

Ces résultats confortent la troisième hypothèse selon laquelle *GS permet de retrouver une proportion nettement plus élevée des citations issues des trois thèses que les bases de données bibliographiques commerciales.*

Les taux de couverture des références très élevés de GS en comparaison de ceux des autres bases de données apportent par déduction une réponse quasi automatique à la **quatrième hypothèse** tandis que les diagrammes de Venn réalisés pour les trois bibliographies en sont l'illustration.

La quatrième hypothèse selon laquelle GS retrouve la majorité des citations couvertes par les autres outils lesquels présentent peu de résultats « uniques » est suivie.

Par ailleurs, les taux de recouvrement entre les bases de données commerciales WoS et Scopus sont très élevés. Pour notre échantillon, les apports spécifiques respectifs des deux bases de données commerciales entre elles ou par rapport à GS sont faibles, en particulier pour WoS.

GeoRef se démarque par de très faibles résultats, lesquels sont largement couverts par les autres outils. Dans le cas des trois thèses analysées, assez éloignées des Sciences de la terre *sensus stricto*, cette dernière base de données n'est pas pertinente.

Discussion et conclusions

Sur base de notre échantillon, théoriquement, les auteurs des 3 thèses auraient pu se passer des bases de données commerciales. Ces dernières pourraient être classées dans l'ordre suivant : Scopus, WoS et loin derrière, GeoRef.

Les bonnes performances de GS en ce qui concerne le référencement des citations trouvent écho dans les réponses des « doctorants » au questionnaire qui leur a été soumis. Tous ont privilégié l'usage de GS « *par facilité* » après avoir fait appel aux fonctionnalités des bases de données commerciales (veille, alerte).

Pour ce qui concerne la **cinquième hypothèse**, nos résultats montrent donc un potentiel bien plus élevé de GS à retrouver des références bibliographiques « uniques », non indexées par les bases de données commerciales.

Les références bibliographiques « uniques » de GS sont aussi bien des articles de périodiques que des livres et chapitres de livres ou des actes de conférences, des thèses et des rapports.

Ces observations confirment la cinquième hypothèse selon laquelle *parmi les citations uniquement retrouvées par GS, outre des articles de périodiques, on retrouve d'autres types de documents cités.*

Les livres et les chapitres de livres cités dans les trois thèses sont aussi occasionnellement indexés par WoS et Scopus, dans une moindre mesure que ce que l'on retrouve dans GS.

Toutefois, comme nous l'avons déjà signalé dans les résultats, la couverture de GS n'est pas complète et même relativement partielle pour la thèse en Tourisme, pour laquelle 67 références n'ont été retrouvées par aucun outil bibliographique. Pour la thèse en Analyse spatiale, 14 références de livres et chapitres de livres, de rapports et de travaux non-publiés en parts égales, n'ont pas été trouvées non plus.

Ceci nous amène à la **sixième hypothèse**. Ces références non couvertes par aucun des outils bibliographiques étudiés sont répertoriées en partie sur Google Books et dans le catalogue ULiège Library. Par ailleurs, les réponses des doctorants au questionnaire indiquent qu'ils ont obtenu des documents et informations bibliographiques *via* leurs réseaux de connaissances (promoteurs, collègues de service et collaborateurs étrangers).

La sixième hypothèse selon laquelle *le chercheur en Géographie utilise encore des documents issus de bibliothèques locales ou obtenus via son réseau de relations ou repérées au cours de ses lectures, et qui ne sont pas répertoriés par GS ou les principales BDs commerciales* est rencontrée.

Deux des trois doctorants déclarent qu'ils n'ont pas beaucoup utilisé les ressources imprimées des bibliothèques à part quelques livres de référence et aussi qu'ils ont acheté des livres. Quand les documents n'étaient pas disponibles, ils ont aussi adressé des demandes directement aux auteurs.

Discussion et conclusions

Des analyses effectuées sur la langue des documents cités et les années de publication, il ressort que les trois doctorants ont cité majoritairement des références en langue anglaise (plus de 90 %) et concentrées sur la dernière décennie avant la publication de la thèse. Nos résultats s'accordent à ceux précédemment trouvés dans la littérature (Meier & Conkling, 2008; Nabe & Imre, 2008; Rosenberg, 2015; Walcott, 1992), l'anglais semble tout de même un peu moins dominant en SHS (la bibliographie en Tourisme) qu'en ST. En effet la publication en Sciences, y compris en Géographie, s'effectue majoritairement en anglais actuellement et depuis quelques décennies déjà, comme indiqué dans la littérature (Aalbers & Rossi, 2006; Schuermans et al., 2010).

En conclusion, malgré - ou grâce à - une interface de recherche simplifiée et intuitive, GS constitue un outil de recherche bibliographique apprécié des doctorants. Nos observations et les déclarations des trois doctorants interrogés confirment en cela ce que l'on trouve dans la littérature (Kaczor, 2014; Noga et al., 1993; Rosenberg, 2015; Sexton, 2006; Walcott, 1992). L'abondance des résultats obtenus, la diversité de forme des notices, qui ne répondent pas aux mêmes standards, ne semblent pas décourager les usagers.

Il est vrai que, même s'il ne faut pas ignorer que l'utilisation préférentielle de GS comme outil de recherche bibliographique peut introduire un biais, nous constatons que quels que soient les documents utiles (thématisques, types, âge, langue...), la couverture de GS est excellente et supérieure à celle des bases de données étudiées. Ces résultats et conclusions sont à mettre en perspective avec les résultats et conclusions des chapitres précédents ainsi que dans les conclusions générales de cette recherche.

CONCLUSIONS GENERALES DE LA PREMIERE PARTIE

Les recherches de la première partie de la thèse ont principalement porté sur les outils de recherche bibliographique (GS, WoS et GeoRef, ainsi que FRANCIS et Scopus pour certaines parties du travail) quant à leurs performances en matière de couverture de l'information en Géographie. Après avoir testé les outils à l'aide de mots-clés, les résultats des deux premiers chapitres ont été mis en perspective avec l'usage réel que font les chercheurs de l'information couverte par les différents outils. Nous avons porté notre attention sur les taux de recouvrement et aussi sur les apports spécifiques éventuels de chacun des outils.

Dans le **premier chapitre** les résultats issus des recherches bibliographiques par mots-clés, spécifiques à la Géographie physique ou humaine, sont particulièrement démonstratifs de ce que GS fournit une quantité plus importante de références que les autres bases de données commerciales, même après exclusion des doublons dans les résultats. Il s'agit bien de documents académiques, de type divers y compris des articles de périodiques scientifiques.

Au terme de ce chapitre, nous avions relevé que GS semble encore plus efficace pour retrouver des références supplémentaires en Géographie humaine qu'en Géographie physique, avec un apport de littérature asiatique important.

Nous avons également relevé qu'en Géographie physique, GeoRef fournit aussi un éventail de types de documents différents, cependant en nombre beaucoup plus restreint que GS.

Toutefois, ce chapitre ne démontre pas l'intérêt que l'abondant complément de littérature fourni par GS peut éventuellement représenter pour le chercheur.

Le **deuxième chapitre** repart des résultats correspondant à deux des dix mots-clés utilisés dans le premier chapitre et approfondit les analyses pour les articles et, surtout, les périodiques dans lesquels ces articles ont été publiés.

Les résultats obtenus dans le deuxième chapitre montrent que le nombre de périodiques « propres » à GS est beaucoup plus important que dans les bases de données commerciales étudiées.

Les périodiques édités en Chine constituent une part majoritaire des titres de périodiques spécifiquement couverts par GS. Si ces titres sont reconnus par *Ulrichsweb* comme des titres académiques, la mention *peer-reviewed* leur est rarement attribuée. Il ne nous semble pas opportun d'en tirer des conclusions car la plupart des titres sont publiés en chinois et nous soupçonnons que l'information d'*Ulrichsweb* est incomplète.

Par ailleurs, GS fournit également une proportion importante d'articles issus de périodiques de langue anglaise, qui sont *peer-reviewed*, et qui n'apparaissent pas dans les résultats des bases de données commerciales. Une partie de ces périodiques de langue anglaise est produite par des pays qui ne sont pas nécessairement anglophones et qui se répartissent sur divers continents.

Aux côtés des périodiques chinois de GS et des périodiques anglophones *peer-reviewed*, on retrouve des périodiques dont les langues de publication sont diverses, auxquels nous pouvons associer un contenu plus national ou local.

Il semble donc que GS ouvre l'accès à une littérature plus mondiale, potentiellement de qualité, qui échappe aux bases de données traditionnelles comme WoS. Pour la Géographie physique, GeoRef, base de données traditionnelle et spécialisée, offre également cette diversité mais dans une moindre mesure, grâce au nombre des pays d'édition et des langues couvertes.

Au terme de ce chapitre, nous avons conclu que, selon le sujet étudié, l'apport supplémentaire de GS pourrait présenter un réel intérêt pour les chercheurs. Le chapitre suivant permet d'appuyer concrètement cet intérêt sur base de l'examen d'un petit échantillon de thèses de doctorat.

Dans le **troisième chapitre**, il s'agissait de vérifier si les chercheurs citent, et dans quelle proportion, des documents spécifiquement couverts par l'un ou l'autre outil bibliographique (Scopus y compris), en particulier par le moteur de recherche GS.

Les listes de documents cités dans les bibliographies des trois thèses rendent compte des habitudes de citation différentes en fonction de l'orientation SSH ou ST du sujet de recherche, qui influence les proportions d'articles de périodiques et des livres cités.

Cependant, les citations correspondant aux articles de périodiques en anglais publiés durant la dernière décennie, avant la date de défense des thèses, sont majoritaires pour les trois bibliographies, dépassant largement les autres types de documents cités. De plus, les périodiques qui recensent la plupart des citations sont des titres clés de la discipline pour les trois thèses.

En matière de couverture des références, GS est de nouveau l'outil le plus efficace et retrouve la majorité des citations des trois bibliographies avec des valeurs proches des 100 %, à l'exception des livres et chapitres de livres que GS ne couvre que partiellement en SHS.

Comme dans les deux chapitres précédents, les références bibliographiques « uniques » de GS comportent des articles des périodiques mais aussi de la littérature grise, des livres et des chapitres de livres, non indexés par les bases de données commerciales. L'apport de GS est plus important pour les thèses à orientation SHS, l'écart entre SHS et ST étant moins marqué cependant que ce qu'on pouvait attendre suite aux résultats du premier chapitre.

WoS et Scopus donnent également de bons résultats, mais avec des valeurs moins élevées par rapport à GS tandis que les contenus de GeoRef ne sont pas adaptés pour l'échantillon étudié. Notons que Scopus a une couverture est un peu plus large que WoS pour une des trois bibliographies tandis que WoS n'apporte quasi aucun résultat spécifique. Dès lors, théoriquement, les auteurs des trois thèses auraient pu se passer de ces deux ressources. Les résultats du chapitre 1 ne nous permettaient pas d'arriver à cette conclusion, WoS ayant généré un certain nombre de références « uniques » autant pour les mots-clés SHS que ST.

En conclusion, les résultats des trois chapitres s'accordent quant à la supériorité numérique de la couverture de GS et suggèrent que le moteur de recherche s'imposerait, pour le moins, comme outil complémentaire aux bases de données commerciales pour la recherche bibliographique en Géographie. Le mode de recherche offert par GS correspondrait bien aux habitudes et modes de fonctionnement actuels des chercheurs qui sont plus dans l'immédiateté et l'optimisation du temps dont ils disposent pour la recherche bibliographique.

Les limites de GS sont, selon nous, à chercher plus dans les possibilités de traitement des résultats, lorsque ceux-ci sont très nombreux, que dans l'étendue et la qualité de l'information couverte.

En utilisant ce type d'outil bibliographique on contournerait l'hégémonie⁵⁰ de l'anglais et des courants anglophones et on accède aussi à une littérature mondiale plus diversifiée, autant du point de vue de la langue que du type de documents. GS, à l'instar des outils bibliographiques spécialisés comme GeoRef, donne aussi accès à des titres qui sont susceptibles de répondre aux besoins des géographes en termes de littérature locale ou régionale.

Ces considérations n'empêchent pas qu'à titre individuel, le chercheur soit amené à tester les outils dont il dispose, avec quelques mots-clés bien choisis, pour déterminer si l'utilisation de GS est adaptée à sa recherche à titre principal ou complémentaire des autres bases de données.

⁵⁰ Il faudrait néanmoins rester prudent car derrière cette ouverture et cet outil « miracle » qu'est Google Scholar se cache une société commerciale qui a son propre agenda qui peut conduire à une autre hégémonie, à des arbitrages peu éthiques et à des utilisations très diverses et non transparentes des données générées.

DEUXIEME PARTIE
La littérature scientifique produite par les géographes de
l'ULiège : diffusion et usage

1. Introduction

La première partie de la thèse a été consacrée à l'étude comparative des contenus de bases de données bibliographiques vs le moteur de recherche Google Scholar, avec pour objectif de renseigner le **consommateur d'information**.

L'objectif était plus précisément de comparer ce que les outils offrent en matière de littérature dans le domaine de la Géographie et de mettre les résultats en perspective avec ce que les géographes utilisent concrètement dans leurs publications scientifiques à partir de l'examen des bibliographies de thèses de doctorat.

Dans la seconde partie de notre thèse, nous allons adopter le point de vue du **producteur d'information** et nous intéresser à la visibilité de la littérature scientifique produite par les géographes de l'ULiège, toujours au travers de bases de données bibliographiques traditionnelles et du moteur de recherche Google Scholar mais aussi *via* le répertoire institutionnel ORBi.

Le premier chapitre analyse le référencement d'un échantillon de publications par les bases de données commerciales multidisciplinaires, Scopus et WoS, et spécialisée GeoRef, ainsi que par le moteur de recherche GS.

Le deuxième chapitre traite du nombre et de la répartition mondiale des téléchargements du même échantillon à partir d'ORBi et tente d'identifier les facteurs socio-économiques qui pourraient les influencer.

La méthodologie que nous avons utilisée nous permettra d'estimer la contribution du répertoire institutionnel ORBi dans la visibilité de cette littérature.

Chapitre 1 Diffusion et visibilité des publications en Géographie de l'ULiège dans les bases de données commerciales et GS

1. Introduction et revue de la littérature

Du fait d'aspects similaires entre la première partie du travail et la seconde partie, différentes études déjà citées, comparant GS et les bases de données bibliographiques (Bauer & Bakkalbasi, 2005; Cavacini, 2015; Gavel & Iselid, 2008; Kimball, et al., 2013; Kulkarni, et al., 2009), gardent tout leur intérêt ici.

A celles-ci, s'ajoutent les études qui comparent les différents outils bibliographiques plus dans une optique d'évaluation des chercheurs, des institutions, voire des pays. Ces études ne se focalisent pas uniquement sur le référencement des publications ; elles peuvent aussi prendre en compte le nombre de citations et les divers indicateurs bibliométriques fournis par les outils bibliographiques, pour évaluer la visibilité des travaux au travers d'un ou l'autre outil bibliographique.

Peu après la mise en production de Scopus, une des plus vastes analyses comparatives, dans le sens où elle a prenait en compte plusieurs domaines de recherche, a été effectuée par Bosman et al., (2006) pour Scopus, WoS et GS. Leurs résultats montrent que Scopus a une couverture des disciplines des Sciences de la terre ainsi que de la Géographie sociale et de l'Aménagement du territoire 10 % plus importante que celle de WoS.

Les auteurs précisent que 13 % des publications indexées par GeoRef sont couvertes par Scopus. L'avantage de GeoRef est de renseigner plusieurs autres types de documents « *more non-journal material* » comme des rapports de différents services géologiques, mais aussi des informations comme les coordonnées géographiques. Ils mettent l'accent sur l'absence dans Scopus des termes du thésaurus et des fonctionnalités spécifiques de la base de données disciplinaire.

En revanche, Scopus de par son contenu multidisciplinaire permet la recherche dans plusieurs disciplines, qui sont parfois connexes. Enfin, les auteurs suggèrent que les meilleures performances de Scopus par rapport à WoS pourraient être dues à l'apport de *Geobase*, intégralement incluse dans Scopus, avec les descripteurs et les termes géographiques. Dans les conclusions, les auteurs indiquent que Scopus a une couverture de bonne à excellente pour les Sciences de la Terre, en soi et par rapport à WoS.

Ensuite une série d'études comparatives, sans lien spécifique avec la Géographie, suivent. Meho et Yang (2007) ont comparé WoS, Scopus et GS dans le domaine des Sciences de l'information. Ils ont recherché les citations par d'autres chercheurs des travaux de 25 membres de l'École de Bibliothéconomie et d'Information de l'Université d'Indiana. Les auteurs ont constaté que par rapport à WoS et Scopus, GS se distingue dans la couverture des actes de conférences ainsi que des

Introduction et revue de la littérature

périodiques internationaux publiés dans d'autres langues que l'anglais. Scopus et GS seraient des outils complémentaires à WoS pour la recherche documentaire, tandis que l'usage des trois outils aiderait à avoir une vision plus précise et complète de l'impact scientifique des auteurs. Pour GS le temps consacré au traitement des données est cependant beaucoup plus important.

Vieira et Gomes (2009) ont travaillé sur un échantillon d'articles correspondant à la production scientifique de 16 universités portugaises entre 2000 et 2007. Ils ont compté les documents référencés dans WoS et Scopus. Leur conclusion générale indique qu'environ 2/3 des documents se trouvent dans les deux bases de données alors que 1/3 sont uniquement répertoriés par l'une des deux.

De Sutter et Van Den Oord (2012) ont utilisé WoS, Scopus, ACM⁵¹, GS et CiteSeer⁵² pour répertorier les publications de trois informaticiens belges. Ils ont constaté que les couvertures de WoS, Scopus, et de GS étaient excellentes pour les articles de périodiques. WoS et Scopus indexaient une sélection de conférences et de périodiques à fort impact. Globalement pour les Sciences informatiques la couverture de WoS était comparable à celle de Scopus. GS couvrait plus de « proceedings » et de conférences que les autres bases de données.

Quant aux citations des publications par d'autres chercheurs, les auteurs de l'étude ont indiqué que « *the Scopus and WoS citation counts were only a small fraction of those in GS* » et que GS pourrait être utile pour identifier et remédier aux problèmes de sous-citations dans des bases de données.

Adriaanse et Rensleigh (2013) ont comparé WoS, Scopus et GS afin d'identifier la ressource bibliographique la plus représentative pour les publications en sciences environnementales en Afrique du Sud. Les auteurs se sont intéressés à la couverture des publications par les trois outils et ont aussi analysé le nombre de citations des publications. Leurs résultats indiquent une meilleure performance de WoS en termes de couverture et de références « uniques » et aussi de citations associées aux documents. Viennent ensuite GS, puis Scopus pour ce qui concerne le nombre de citations. Les auteurs ont également relevé que GS présentait plus d'incohérences et d'erreurs quant à l'orthographe des noms d'auteurs et les volumes et numéros de publication.

Bartol et al., (2014) ont analysé le recensement des 50 000 documents de recherche publiés en Slovénie entre 1996 et 2011 dans Scopus et WoS. Ils ont aussi pris en compte les citations reçues par ces documents. Ils ont constaté que Scopus indexe plus de documents et établit des liens avec plus de citations que WoS, dans tous les domaines de recherche. Les différences sont particulièrement évidentes en SHS

⁵¹ Association for Computing Machinery Digital Library: The ACM Digital Library (DL) is the world's most comprehensive database of full-text articles and bibliographic literature covering computing and information technology.

⁵² CiteSeerx is an evolving scientific literature digital library and search engine that has focused primarily on the literature in computer and information science.

Introduction et revue de la littérature

ainsi qu'en Ingénierie et techniques. Par ailleurs, les Sciences humaines ont reçu le moins de citations par document tandis que les Sciences médicales et les Sciences naturelles, qui présentent des résultats similaires, sont beaucoup plus citées. Les documents d'Ingénierie et les techniques sont cités deux fois moins en moyenne que ceux des Sciences médicales et les Sciences naturelles.

Cavacini (2015) a recherché un échantillon d'articles provenant d'un répertoire institutionnel italien dans INSPEC⁵³, Scopus, WoS et DBLP⁵⁴. Leurs résultats montrent une meilleure performance de Scopus pour retrouver les publications en Informatique, par rapport à WoS. Scopus et DBLP indexent le plus grand nombre d'articles « uniques », mais chacune des bases de données indexe sa part de références uniques et elles sont donc complémentaires.

Dans une autre étude récente, réalisée par Gorraiz et al., (2016), on retrouve des données portant sur la Géographie. Cette étude a été réalisée dans le but d'évaluer 40 candidatures à des postes du Département de Géographie de l'Université de Vienne, dans des sous-disciplines variées comme la Géographie de la population et la démographie, la Géographie économique, la Géo-écologie et la Géographie économique et sociale. Il s'agissait de mettre en évidence les différences éventuelles dans les pratiques de publication. Pour la Géographie de la population et la démographie, les publications des candidats ainsi que les citations reçues ont été recherchées en interrogeant en parallèle trois outils bibliographiques : WoS, Scopus et GS. Pour la Géographie économique, WoS et GS ont été utilisés tandis que pour la Géo-écologie et la Géographie économique et sociale, seul WoS a été utilisé.

Notons par ailleurs que les auteurs ont détecté des différences dans les pratiques de publication des géographes relevant de la Géographie physique d'une part et de la Géographie humaine d'autre part. Les chercheurs travaillant dans les sous-disciplines orientées vers les Sciences naturelles ont tendance à publier principalement des articles, dans des périodiques en anglais et à haut facteur d'impact, qui sortent de leur discipline principale. Les chercheurs travaillant dans des sous-disciplines orientées vers les SHS, quant à eux, publient dans des périodiques plutôt orientés vers la Géographie et majoritairement des parties et chapitres de livres. Ils n'ont pas constaté de différence majeure en utilisant WoS ou Scopus, et ils encouragent l'usage de GS en tant qu'outil complémentaire, notamment pour la recherche des livres et de la littérature grise, car ce dernier offre une meilleure couverture.

La revue de la littérature nous a permis de constater que peu d'études portent sur la production scientifique en Géographie et sa visibilité dans les outils bibliographiques. Elles s'accordent cependant sur le fait que les performances des

⁵³ Base de données spécialisé en Ingénierie développé par Elsevier

⁵⁴ Computer science bibliography : ce service fournit des informations bibliographiques des principaux périodiques et comptes rendus en Informatique. DBLP est géré par Schloss Dagstuhl et fut fondé à l'Université de Trèves en 1993.

Objectifs et hypothèses

bases de données commerciales sont similaires et que GS, quand il est pris en compte, fournit des bons résultats et complète les bases commerciales.

Globalement, il semble que GS ait amélioré sa couverture au fil des ans, en particulier dans le domaine des SHS (Jamali & Nabavi, 2015), ce qui pourrait distinguer le moteur de recherche des deux bases de données évoquées précédemment. Un autre atout considérable de GS est le lien direct vers le texte intégral qu'il fournit (lorsqu'il est disponible), que ce soit *via* ResearchGate, les répertoires institutionnels ou directement sur les sites des éditeurs (Mikki et al., 2018). Les références disponibles en OA sont aussi susceptibles de recevoir plus de citations et les dernières études (Archambault et al., 2016; Jamali & Nabavi, 2015; Mikki et al., 2018; Piwowar et al., 2018) font état d'un impact de citation jusqu'à 50 % plus important de ces documents.

Dans ce chapitre, nous voulons vérifier quelle est la visibilité de la littérature scientifique produite par les géographes de l'ULiège dans les bases de données commerciales et dans GS. La méthodologie utilisée nous permettra, d'une part, d'enrichir la littérature existante avec des données relatives à la Géographie et, d'autre part, d'évaluer la contribution du répertoire institutionnel ORBi dans la visibilité de cette littérature.

2. Objectifs et hypothèses

Le premier objectif est d'examiner et de qualifier l'échantillon de publications des géographes de l'ULiège, sélectionné à partir d'ORBi.

Le second objectif est de déterminer dans quelle mesure la littérature scientifique produite par les géographes de l'ULiège est répertoriée par les bases de données commerciales WoS, Scopus, GeoRef et par GS.

En intégrant un nouvel élément, soit le dépôt des références dans ORBi par les membres du Département de Géographie, nous pourrons déterminer la contribution du répertoire institutionnel à la visibilité de cette littérature. A notre connaissance, ceci est la première étude qui intègre cet élément de dépôt dans un répertoire institutionnel.

Sur base de la revue de la littérature et des résultats de la première partie, nous posons les hypothèses suivantes :

Hypothèse 1 : Selon que l'on considère l'orientation vers les SHS ou les ST, la répartition des publications de notre échantillon entre les chapitres ou parties de livres et les articles de périodiques ou communications est nettement différente ;

Hypothèse 2 : Quelle que soit la sous-discipline, l'article de périodique est la principale voie de publication ;

Hypothèse 3 : Quelle que soit la sous-discipline ou le type de document, l'anglais est la principale langue de publication ;

Hypothèse 4 : Grâce au moissonnage d'ORBi, GS retrouve toutes les publications des géographes de ULiège.

Hypothèse 5 : Les références retrouvées uniquement dans GS, exclusivement grâce au moissonnage d'ORBi, sont nombreuses, surtout pour la Géographie humaine.

Hypothèse 6 : Si on fait abstraction des références en provenance d'ORBi, une proportion plus élevée de références est retrouvée par GS en comparaison des bases de données bibliographiques commerciales ;

Hypothèse 7 : Le recouvrement de WoS et Scopus est très important et seul Scopus se distingue par un apport significatif de références supplémentaires par rapport à WoS, relevant souvent des SSH ;

Hypothèse 8 : Dans l'absolu, GeoRef couvre moins de références que WoS et Scopus mais se distingue par un apport de références supplémentaires en Géographie physique par rapport aux deux autres BD commerciales ;

Hypothèse 9 : Les périodiques dans lesquels les géographes de ULiège publient sont majoritairement *peer-reviewed* et de langue anglaise ;

Hypothèse 10 : Du fait de leur présence dans le répertoire institutionnel ORBi, les publications en langue française des géographes liégeois publiées dans les périodiques « locaux » sont plus visibles *via* GS que *via* les BD commerciales ;

3. Méthodes

Un rapport reprenant les références des livres, parties et chapitres de livres ainsi que articles de périodiques et communications, publiés entre 2000 et 2014 par les membres du département de Géographie (ULiège) et déposés dans ORBi, a été généré en septembre 2015.

Les métadonnées (auteurs, titres, années de publication, périodique, ISSN, type de publication) de ces références ont été téléchargées et enregistrées dans des fichiers *Excel*. Les erreurs d'encodage ont été repérées et corrigées.

Une référence pouvant être associée à plusieurs disciplines, pour le descriptif nous avons pris en compte la première discipline indiquée par le déposant, en supposant qu'il s'agit de la discipline principale de la référence.

Afin de statuer sur la couverture des références par les outils bibliographiques, chaque référence de l'échantillon a été recherchée dans GS, GeoRef et Scopus. Par ailleurs, un travail similaire a été effectué pour les titres de périodiques dont relèvent les articles de périodiques et communications.

Notre démarche a été limitée pour les recherches dans WoS, car l'Université de Liège n'a plus accès à cette base depuis 2011. Nous avons adapté notre méthode et nous nous sommes appuyée sur les informations du fichier *Excel* que nous avons pu obtenir par le service « clients » pour la recherche des titres de périodiques. Les

Résultats

champs ISSN ont donc été croisés afin d'obtenir les informations nécessaires. Les recherches pour les parties et chapitres de livres n'ont pas été effectuées dans WoS.

Les listes de références de documents, ou les listes de titres de périodiques, couverts par les différents outils ont été établies et comparées. Les références « uniques » de GS et celles provenant exclusivement d'ORBi ont également fait l'objet d'un relevé.

Les pourcentages des articles et communications couverts par un ou plusieurs outils bibliographiques ont été calculés. Afin d'estimer l'apport d'ORBi, deux types de pourcentages ont été calculés pour GS : ceux qui prenaient en compte les résultats provenant du moissonnage d'ORBi et ceux qui les écartaient. Les recouvrements des outils entre eux ont été représentés avec les diagrammes de Venn.

La base de données *Ulrichsweb* a été consultée afin de relever les caractéristiques (*peer-reviewed*, OA, mention JCR, langue et pays de publication, discipline du périodique) des périodiques retrouvés uniquement par GS en distinguant aussi ceux recensés uniquement par l'intermédiaire d'ORBi.

4. Résultats

4.1. Description de l'échantillon de publications

L'échantillon de publications 2000-2014 issu d'ORBi, *via* un rapport de publication sélectionné sur base de l'affiliation des auteurs, comprend 146 chapitres ou parties de livres ainsi que 405 articles et 48 communications publiés dans des périodiques. La répartition des chapitres et parties de livres par discipline (ORBi), ainsi que des articles et communications, est reprise dans le Tableau 7.

Discipline	Nb.	Ang.	Fr.	Nb.	Ang.	Fr.
	Ch.Livres			Art.		
Arts & sciences humaines	1	1	3	2	1	
Droit, criminologie & sciences politiques	3	1	2	1		1
Ingénierie, informatique & technologie			28	19	9	
Physique, chimie, mathématiques & sciences de la terre	50	39	11	267**	197	68
Sciences sociales & comportementales, psychologie	87*	20	65	82***	11	70
Sciences du vivant (Sciences de l'environnement & écologie)	5		5	36	21	15
Sciences économiques & de gestion				33	3	30
Sciences de la santé humaine				3	3	

*1 référence en espagnol et 1 référence en allemand

**2 références en néerlandais

*** 1 référence en portugais

Tableau 13 - Disciplines et langues des différents types de documents relevés dans ORBi, les intitulés en gras sont celles les disciplines reprises dans la catégorie ST

4.1.1. Les chapitres et parties de livres

Les chapitres et parties de livres représentent 24 % de l'échantillon total. Ce pourcentage est relativement équivalent à celui indiqué par Gorraiz et al., (2016), selon lequel un tiers des publications du département de Géographie et de

Résultats

recherches régionales de l'Université de Vienne étaient des chapitres de livres tandis qu'un dixième étaient des livres.

Les « Sciences sociales & comportementales, psychologie » (voir Tableau 7) comptent le plus grand nombre des publications sous forme de chapitres ou de parties de livres (87), dont la majorité (61) traitent plus précisément de « Géographie humaine & démographie ».

Vient ensuite la « Physique, chimie, mathématiques & sciences de la terre » qui compte 50 références de chapitres et parties de livres, dont 46 en « Sciences de la terre & géographie physique ».

Toutes disciplines confondues, les SHS comptent 91 publications sous forme de chapitres ou parties de livres, soit 43% du total des publications en SHS (210 publications) tandis que les ST comptent 55 publications sous forme de chapitres ou parties de livres, soit 14% du total des publications en ST (389 publications).

Il apparaît donc une nette différence dans la balance entre les chapitres ou parties de livres et les articles de périodiques ou communications, selon que l'on considère les SHS ou les ST.

Pour ce qui concerne la langue de publication, toutes disciplines confondues, une majorité de 57% (83) de chapitres ou parties de livres sont publiés en français tandis que 42% (61) sont publiés en anglais. Pour les SHS, le pourcentage de publications en français augmente à 74% (67) pour 24% (22) en anglais. En ST, le pourcentage de publications en français descend à 29% (16) et le pourcentage de publications en anglais monte à 71% (39). Les autres langues sont utilisées de manière tout-à-fait marginale, en SHS.

La différence entre SHS et ST se confirme également pour ce qui est de la langue de publication.

Seulement 26 des 146 chapitres ou parties de livres se sont vu attribuer la mention *peer-reviewed* par leurs auteurs dans ORBi, dont 18 en SHS et 8 en ST.

Pour ce qui est du **recensement des chapitres ou parties de livres** par les différents outils bibliographiques, GS référence 100% des publications (Tableau 8). Si on faisait abstraction des 96 références retrouvées uniquement grâce à un lien vers ORBi, les performances du moteur de recherche diminueraient considérablement avec seulement 50 chapitres ou parties de livres référencés (33,5 %).

Databases	GS	GS (-) ORBi	Scopus	GeoRef
Nb. ref.	146	50	6	5
(%)	100	33,5	4,1	3,4

Tableau 14 - Pourcentage de chapitres ou parties de livres indexés et par les outils bibliographiques

Des 50 chapitres ou parties de livres retrouvés par GS sans l'apport d'ORBi, 17 sont en « Sciences sociales & comportementales, psychologie », dont 16 traitent

Résultats

plus précisément de « Géographie humaine & démographie ». 31 chapitres et parties de livres sont en « Physique, chimie, mathématiques & sciences de la terre », dont 28 en « Sciences de la terre & géographie physique ».

Des 96 chapitres ou parties de livres retrouvés grâce à ORBi, 74 sont en « Sciences sociales & comportementales, psychologie », dont la majorité (50) traitent plus précisément de « Géographie humaine & démographie ». 19 chapitres et parties de livres sont en « Physique, chimie, mathématiques & sciences de la terre », dont 18 en « Sciences de la terre & géographie physique ».

Scopus ne retrouve que 6 références, toutes en « Physique, chimie, mathématiques & sciences de la terre » et dont 5 plus spécifiquement en « Sciences de la terre & géographie physique ».

Quant à GeoRef, 5 références sont retrouvées par cette BD disciplinaire, toutes en Sciences de la terre & géographie physique.

4.1.2. Les articles de périodiques et les communications

Les articles de périodiques et les communications représentent 76 % (453) de l'échantillon total.

Par rapport aux chapitres et parties de livres, la répartition du nombre de publications entre les deux disciplines principales s'inverse (Tableau 7) avec une majorité de 267 publications en « Physique, chimie, mathématiques & sciences de la terre », dont 247 publications en « Sciences de la terre & géographie physique », et avec 82 publications en « Sciences sociales & comportementales, psychologie » dont 49 publications en « Géographie humaine & démographie ». La sous-discipline « Sciences de la terre & géographie physique » compte le plus grand nombre de références.

Toutes disciplines confondues, les ST comptent 334 publications sous forme d'articles de périodiques ou de communications, soit 86% du total des publications en ST (389 publications) tandis que les SHS comptent 119 publications sous forme d'articles de périodiques ou de communications, soit 57% du total des publications en SHS (210 publications).

Pour ce qui concerne la langue de publication, toutes disciplines confondues, la tendance relevée pour les chapitres et parties de livres s'inverse ici, avec une majorité de 57% (256) de publications rédigées en anglais et avec 43% (194) des publications en français. Les autres langues sont toujours utilisées de manière marginale.

Pour les ST, le pourcentage de publications en anglais monte à 72% (240) pour 28% (92) en français. En SHS, le pourcentage de publications en anglais descend à 13% (16) et le pourcentage de publications en français monte à 86% (102).

Cette fois encore, la différence entre SHS et ST se confirme pour ce qui est de la langue de publication.

360 références, donc près de 80 %, se sont vu attribuer la mention *peer-reviewed* par leurs auteurs dans ORBi, 80 (22%) en SHS, dont 57 en « Sciences sociales & comportementales, psychologie - Géographie humaine & démographie » et 280 (78%) en ST, dont 218 en « Physique, chimie, mathématiques & sciences de la terre ». La mention *peer-reviewed* certifiée par l'équipe ORBi revient à 247 des 360 références, dont 29 en SHS et 181 en ST.

4.2. Recensement des articles et des communications par les outils bibliographiques

Le Tableau 15 reprend les nombres et les pourcentages d'articles et de communications de notre échantillon, répertoriés par les différents outils bibliographiques. En parallèle, sont également repris les nombres et les pourcentages des périodiques correspondants. Pour rappel, afin d'estimer l'apport d'ORBi, deux séries de relevés ont été faites pour GS, qui prenaient en compte ou écartaient les résultats provenant d'ORBi.

GS retrouve 100% des 453 articles et communications de notre échantillon. A l'analyse, il s'avère que 68 de ces références sont retrouvées par GS exclusivement par les liens établis avec ORBi et qu'elles sont toutes « uniques » au moteur de recherche. Ces références seront examinées en détail sous le point 4.2.2.1.

Databases	<i>GS</i>	%	<i>GS (-) ORBi</i>	%	<i>Scopus</i>	%	<i>WoS</i>	%	<i>GeoRef</i>	%
Nb. ref.	453	100	385	84,9	224	49,4	167	36,8	150	32,8
Nb. périodiques	193	100	144	74,6	123	63,7	88	45,6	108	56

Tableau 15 - Nombre et pourcentages des articles et des périodiques retrouvés dans les outils bibliographiques

Même en faisant abstraction des références issues exclusivement d'ORBi, GS présente la meilleure performance avec 385 des références couvertes (84,9 %), dont 252 des 267 (94%) de « Physique, chimie, mathématiques & sciences de la terre » et 55 des 82 (67%) de « Sciences sociales & comportementales, psychologie ». A titre de comparaison, Gorraiz et al., (2016) ont retrouvé dans GS 74,6 % d'articles publiés en Démographie et Géographie de la Population et 91 % en Géographie Economique.

Le moteur de recherche GS est suivi par Scopus avec 49,4 % (224) de références couvertes, dont 162 des 224 (72,3%) en « Physique, chimie, mathématiques & sciences de la terre » et 14 des 82 (17%) en « Sciences sociales & comportementales, psychologie - Géographie humaine & démographie ». Notre valeur en SHS - Géographie humaine & démographie est équivalente à celle trouvée par Gorraiz et al., (2016), qui était de 16,8 % en Démographie et Géographie de la Population.

Résultats

Scopus indexe 12,6 % références de plus que WoS qui couvre 36,8 % (167) des références. Des 167 références retrouvées par WoS, 132 sont issues des 267 (49%) références de « Physique, chimie, mathématiques & sciences de la terre » et 3 des 82 (3,6%) références de « Sciences sociales & comportementales, psychologie ». Globalement, la différence que nous trouvons entre Scopus et WoS est proche de celles trouvées par Bosman et al., (2006) qui indiquaient une supériorité de 10 % pour Scopus pour les Sciences de la terre ainsi que de la Géographie sociale et de l’Aménagement du territoire. Mais si nous entrons dans le détail de nos chiffres par discipline, les proportions de références retrouvées par les outils fluctuent et nous nous écartons de cette valeur de 10%.

Scopus indexe 16,6 % références de plus que GeoRef (150). Des 150 références retrouvées par GeoRef 136 sont en « Physique, chimie, mathématiques & sciences de la terre ». Bosman et al., (2006) avait trouvé 16,3 % en plus pour Scopus par rapport aux titres de périodiques indexés dans GeoRef.

4.2.1. Qualification et recensement des périodiques

Les 453 articles et communications de notre échantillon ont été publiés dans 193 périodiques différents. Dans ORBi, 92 titres de périodiques sont qualifiés *peer-reviewed vérifié par ORBi* et 49 autres titres peuvent être indirectement qualifiés de *peer-reviewed* suite au dépôt d'un article qualifié de *peer-reviewed* par l'auteur.

179 périodiques étaient assortis d'un ISSN dans ORBi. Nos recherches ne nous ont pas permis de retrouver l'ISSN pour les 14 titres restants (Tableau 16) malgré le fait que les articles publiés dans cinq de ces titres était indiqués *peer-reviewed* par les auteurs. Ces 14 titres n'ont pas été retrouvés dans *Ulrichsweb*.

Périodiques sans ISSN

Actes du Geoforum

AGU Abstract

AM/FM-GIS BELUX News

Annales des Sciences et Sciences Appliquées de l'Université Officielle de Bukavu

Bulletin du Cercle historique de Durbuy

Cahiers du Centre Européen de Géodynamique et de Séismologie

La Cemathèque

Landmeter - Géomètre

Les Hors-série de Politique

Papers on Tools and methods of Territorial Intelligence

Proceedings of the AAG Annual Meeting

Puzzle

Revue Hommes Terre et Eaux

TRACeS de ChanGements

Tableau 16 - Périodiques sans ISSN

Résultats

154 des 179 périodiques avec ISSN ont été retrouvés dans *Ulrichsweb*, ce qui nous a permis de compléter les informations du rapport ORBi pour ces titres, concernant la qualité de *peer-reviewing* selon *Ulrichsweb*, la mention JCR, la langue et le pays de publication, ainsi que les disciplines couvertes (Tableau 17).

Selon *Ulrichsweb*, sur l'ensemble de l'échantillon, 129 titres, soit 83,8 % ont un contenu académique et sont associés à la discipline « Géographie ». Le reste des titres relèvent de disciplines connexes, généralement associées à la Géographie, à l'exception de 2 titres qui ont un contenu de vulgarisation et de 2 titres plutôt centrés sur le commerce. 119 (77,3 %) des 154 titres avec un ISSN, tous présentant un contenu académique, sont labellisés *peer-reviewed* par *Ulrichsweb*. 81 de ces titres ont la mention *peer-reviewed vérifié par ORBi* tandis que 18 autres sont indiqués *peer-reviewed* par les auteurs. 9 titres ne sont pas qualifiés de *peer-reviewed* du tout dans ORBi. A contrario, 11 titres sont mentionnés *peer-reviewed vérifié par ORBi* alors qu'ils ne le sont pas par *Ulrichsweb*.

86 des 119 titres labellisés *peer-reviewed* par *Ulrichsweb* ont la mention JCR et 17 des 119 périodiques sont en OA.

Les 25 périodiques non trouvés dans *Ulrichsweb* sont majoritairement francophones et publiés en Europe, à l'exception d'un titre, *Horizons in geography*, publié en Israël. 11 de ces titres sont qualifiés *peer-reviewed* par les auteurs et traitent de sujets de recherche scientifique mais aussi d'aspects locaux, comme par exemple la politique scientifique ou les projets mis en place dans l'enseignement dans une région ou un pays (ex. : *Science Connection*).

Des 154 périodiques, tous référencés pas GS, que nous avons retrouvés dans *Ulrichsweb*, 108 périodiques sont indexés par GeoRef, 123 par Scopus et 88 par WoS (Tableau 17 et Figure 13).

Outils	Périodiques référencés (Nb.)	Peer-reviewed <i>Ulrichsweb</i>	Peer-reviewed "vérifié par ORBi"	JCR	OA	Disciplinaire (%)
Echantil. <i>Ulrichsweb</i>	154	119	85	89	24	129
GS	154	119	85	89	24	129 (83.8)
GeoRef	108	88	66	75	12	93 (86.1)
Scopus	123	105	82	86	19	103 (83.7)
WoS	88	84	70	85	13	74 (84.1)

Tableau 17 - Nombre de périodiques référencés par les différents outils bibliographiques et qualifications correspondantes (selon *Ulrichsweb* pour JCR, OA et disciplinaire)

Dans l'absolu, en ce qui concerne les **critères de qualité**, GS présente 31 titres de périodiques supplémentaires par rapport à Scopus dont 14 sont *peer-reviewed* selon *Ulrichsweb*. De même, GS présente 46 titres supplémentaires par rapport à GeoRef dont 31 sont *peer-reviewed* selon *Ulrichsweb* et GS présente 66 titres

Résultats

supplémentaires par rapport à WoS dont 35 sont *peer-reviewed* selon *Ulrichsweb* (Tableau 11).

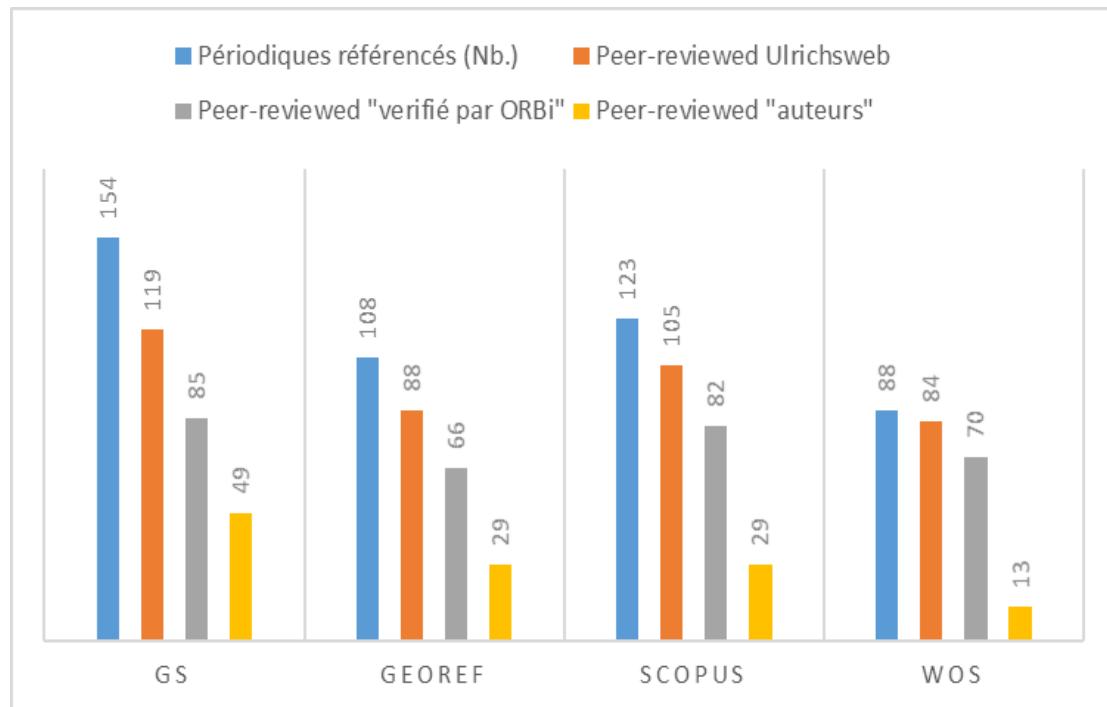


Figure 12 - Nombre de périodiques référencés par les outils bibliographiques et mention peer-reviewed dans Ulrichsweb et ORBi et « auteurs »

Mais si on considère, pour chaque outil bibliographique, les pourcentages de périodiques qui sont évalués par les pairs d'après *Ulrichsweb* (Tableau 11 et Figure 13), le classement des outils bibliographiques est modifié. WoS vient d'abord avec 95,5 % de titres de périodiques *peer-reviewed* (pour 79,5% de *peer-reviewed vérifié par ORBi* et 14,8 % *peer-reviewed « auteurs »*), suivi de Scopus avec 85,4% (pour 66,6% de *peer-reviewed vérifié par ORBi* et 23,5 % *peer-reviewed « auteurs »*), puis GeoRef avec 81,5% (pour 61,1% vérifié de *peer-reviewed vérifié par ORBi* et 26,8 % *peer-reviewed « auteurs »*) et en dernière position GS avec 77% (pour 55,2% de *peer-reviewed vérifié par ORBi* et 31,8 % *peer-reviewed « auteurs »*).

Si la mention de peer-reviewing attribuée par *Ulrichsweb* peut paraître un peu surfaite par rapport à la mention *peer-reviewed vérifié par ORBi*, on constate d'autre part que ce n'est plus le cas si on tient compte de la mention *peer-reviewed* attribuée par les auteurs lors du dépôt de leurs publications dans ORBi (Figure 12).

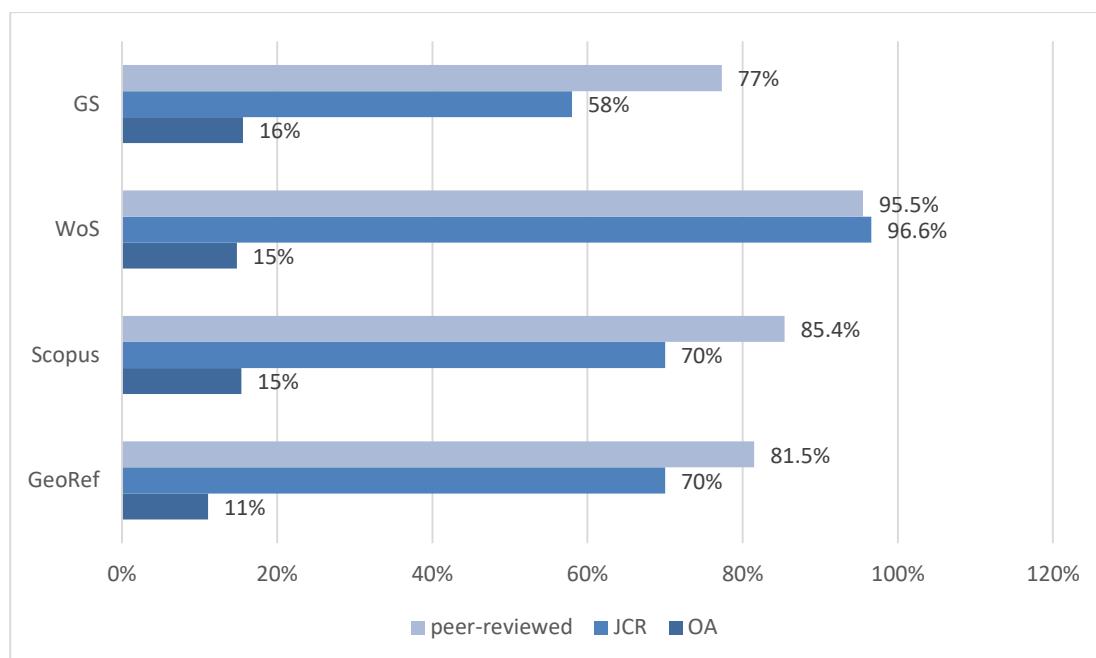


Figure 13 - Pourcentage de périodiques peer-reviewed, JCR et OA par outils bibliographique

En ce qui concerne la mention JCR, dans l'absolu GS se classe en premier avec 89 titres de périodiques, suivi par Scopus avec 86 titres, WoS avec 85 titres et GeoRef, en dernier, avec 75 titres. En données relatives, WoS présente le pourcentage le plus élevé de périodiques repris dans le classement, 96,6 %. Pour les autres outils bibliographiques, la mention JCR est attribuée à 70% des périodiques de Scopus et GeoRef et à 58 % des titres de GS.

Dans l'absolu, on retrouve plus de titres OA (Tableau 11) dans GS (24) que dans Scopus (19), WoS (13) ou GeoRef (12). Pour ce qui est du pourcentage des titres en OA, le classement des outils bibliographiques reste le même et décroît de 16% pour GS à 11% pour GeoRef. Ces valeurs sont proches de celles trouvées par les producteurs de Scopus (19,5 %) et de WoS (10,4 %). Quant à Miguel *et al.*, (2011), ils indiquent 16,2 % des titres OA pour la discipline Sciences de la Terre et Environnement.

Pour ce qui est de la **langue des périodiques**, l'anglais est la première langue pour chacun des quatre outils. Dans l'absolu (Tableau 12), on retrouve plus de titres en anglais dans GS (93) que dans Scopus (82), WoS (74) ou GeoRef (71). Considérant les pourcentages, le classement des outils bibliographiques change : WoS présente le pourcentage le plus élevé de périodiques en anglais (84 %), suivi de Scopus (66,7 %), GeoRef (66 %) et GS (60 %).

Pour tous les outils bibliographiques, l'examen des données révèle que la majorité des titres en langue anglaise sont qualifiés de *peer-reviewed* par Ulrichsweb. WoS vient d'abord avec 72 des 74 titres en anglais (97,2 %), Scopus se classe en deuxième position avec 78 des 82 titres (95 %), GeoRef vient ensuite avec 66 des 71 titres (93 %) et GS en dernier avec 86 des 93 titres (92,4 %).

Résultats

Outils	Périodiques référencés (Nb.)	Texte en anglais	%	Texte en français (+ autre langue)	%	Texte en autres langues	%
Echantil. <i>Ulrichsweb</i>	154	93		52		9	
GS	154	93	60,3	52	33,7	9	5,8
GeoRef	108	71	65,7	32	29,6	4	3,7
Scopus	123	82	66,6	35	28,4	6	4,8
WoS	88	74	84	12	13,6	2	2,2

Tableau 18 - Langue de publication des périodiques référencés par les outils bibliographiques

Le français est la deuxième langue de l'échantillon de périodiques référencés par chacun des trois outils : 52 des 154 titres de GS (33,7 %), 32 des 108 titres de GeoRef (29,6 %), 35 des 123 titres de Scopus (28,4 %), et 12 des 88 titres de WoS (13,6 %).

Les 52 titres de périodiques publiés en français (plus une autre langue dont notamment l'allemand et l'anglais pour 19 des titres), tous référencés pas GS, sont moins souvent qualifiés de *peer-reviewed* que les titres publiés en anglais : *30 des 52 titres* (57,6 %) ont la mention *peer-reviewed* dans *Ulrichsweb*. 15 titres des 30 titres *peer-reviewed* dans *Ulrichsweb* sont qualifiés *peer-reviewed vérifié par ORBi*. Pour les autres outils bibliographiques qui ne couvrent pas tous les titres, le nombre de titres publiés en langue française et *peer-reviewed* est plus faible que dans GS mais les pourcentages de titres *peer-reviewed* augmentent : 19 des 29 (65,5 %) titres en français de GeoRef sont qualifiées *peer-reviewed*, 26 des 35 (74 %) titres en français de Scopus et l'ensemble des titres en français de WoS.

Si l'on considère la répartition mondiale des titres par **pays d'édition**, le Royaume-Uni se distingue comme le plus gros pays éditeur de périodiques avec 33 des 154 titres (21,4 %) de l'échantillon. La France et les Etats-Unis suivent le Royaume-Uni avec 30 (19,4 %) et 24 (15,6 %) titres chacun. L'Allemagne, les Pays-Bas, et la Belgique publient respectivement 18 (11,7 %), 16 (10,4 %) et 15 (9,7 %) titres chacun.

Les autres pays comme le Canada, la Pologne, le Brésil, la Chine, l'Italie... sont beaucoup moins représentés avec un à deux titres publiés, mais il est important de relever leur présence, car cela indique une pratique de publication élargie, malgré la prédominance du Royaume-Uni (Johnston, 2003).

4.2.2. Illustration du recouvrement entre les différents outils bibliographiques

Le diagramme de Venn (figure 14) illustre les taux de recouvrement des outils bibliographiques quant à la recension des 453 articles et communications de l'échantillon.

GS retrouve l'ensemble des 453 références de l'échantillon et le moteur de recherche recouvre à 100 % toutes les autres bases de données. GS recense 183 références « uniques », soit 40 % de l'échantillon. Le corollaire est que 60% des références sont trouvées par au moins deux des outils bibliographiques.

Parmi les 183 références uniques de GS, se trouvent les 68 références qui ont été retrouvées exclusivement par des liens avec ORBi.

97 références (21, 4 %) sont répertoriées par tous les outils bibliographiques. Le pourcentage des références retrouvé par au moins deux des bases de données commerciales est de 38 % (174 références).

Chacune des bases de données commerciales se distingue des autres bases de données commerciales par un apport supplémentaire de références. Scopus et GeoRef se démarquent respectivement par 55 références (12,1%) et 35 références (7,7%) tandis que WoS, par seulement 6 références (1,3%).

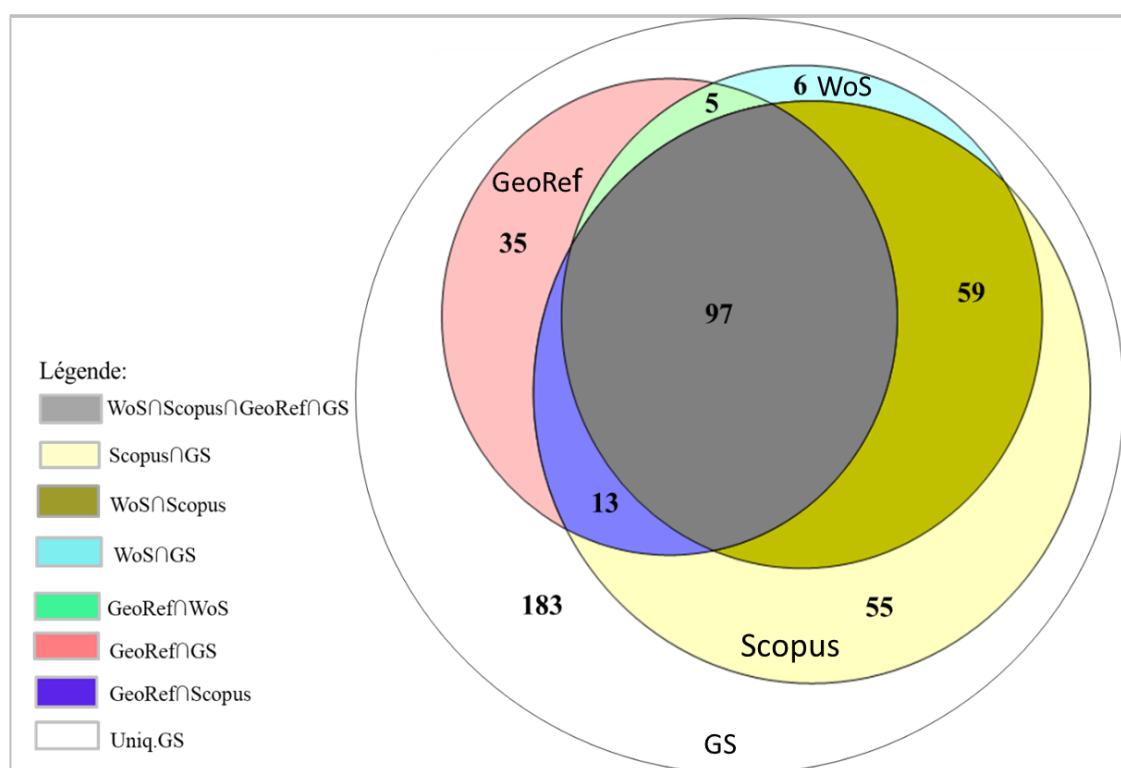


Figure 14 - Recouvrement entre les différents outils bibliographiques

Résultats

4.2.2.1. Caractéristiques des références et périodiques uniquement répertoriés par GS

Les 183 références « uniques » de GS se répartissent en deux lots : 115 références qui seraient de toute façon retrouvées par GS même si elles n'avaient pas été déposées dans ORBi (ci-après : les références de GS sans ORBi) et 68 références qui sont retrouvées par GS uniquement *via* le moissonnage d'ORBi (ci-après : les références de GS *via* ORBi).

	Nb de références uniques	Anglais	%	Français	%
GS sans ORBi	115	20	17,4	94	81,7
GS <i>via</i> ORBi	68	9	13,2	57	83,8

Tableau 19 - Nombre et langue des références uniques de GS, sans apport d'ORBi ou *via* ORBi

A l'examen du Tableau 19, il apparaît que des **115 références uniques à GS sans ORBi**, 94 sont en langue française (81,7 %), 20 en langue anglaise (17,4%) et 1 en portugais.

33 périodiques correspondent à ces références, dont 31 avec un ISSN attribué. Dans le rapport ORBi, 27 titres étaient indiqués comme *peer-reviewed* : 18 indirectement par le déposant et 9 vérifiés par ORBi. Des 23 périodiques retrouvés dans *Ulrichsweb*, 22 titres ont un contenu académique et sont disciplinaires tandis qu'un titre a un contenu de vulgarisation de type magazine. 13 titres sont indiqués *peer-reviewed*, 2 titres ont la mention JCR et 3 périodiques sont en OA.

Pour 10 périodiques, *Ulrichsweb* indique le français comme langue du texte, 7 périodiques ont le texte en anglais, 3 périodiques en français et anglais, 1 périodique en français et allemand, 1 en Slovène et 1 en plusieurs langues.

Quant au pays de publication, à la France reviennent 11 titres ; aux Etats-Unis, 4 titres ; à la Belgique, 3 titres ; au Brésil, 1 titre ; à la Pologne, 1 titre ; à la Slovaquie, 1 titre à l'Allemagne 1 titre et au Pakistan 1 titre.

	Nb de périodiques	Trouvés <i>Ulrichsweb</i>	Peer- Reviewed	JCR	Peer- reviewed vérifiée par ORBi	Open Access	Anglais	Français(+ autre Langue)	Autre langue	Périodiques Disciplinaires
GS uniq.	33	23	13	2	8	3	7	14	2	22
GS uniq. ORBi	49	22	6	0	2	1	2	17	3	18

Tableau 20 - Caractéristiques de périodiques propres à GS

Notons que 62 des 115 références sont publiées dans le seul ***Bulletin de la Société Géographique de Liège***, périodique dont la langue de publication est le français et l'anglais, *peer-reviewed vérifié par ORBi*, en OA et qui est fortement lié au département de Géographie de l'ULiège.

A l'examen du Tableau 19, il apparaît que des **68 références retrouvées par GS via ORBi** se répartissent en 57 références (83,8%) en français, 9 en anglais (13,2%) et 2 en néerlandais.

46 références de ces 68 références sont classées dans les « Sciences sociales & comportementales, psychologie », dont 14 en « Géographie humaine & démographie » et 4 en « Etudes régionales & interrégionales ». Les ST comptent 22 références dont 15 en « Sciences de la terre & géographie physique ».

Ces 68 références ont été publiées dans 49 périodiques dont 37 avec un ISSN attribué (titres de périodiques indiquées dans Annexe 3). Dans le rapport ORBi, 21 périodiques ont reçu la mention *peer-reviewed* : 2 *peer-reviewed vérifié par ORBi* et 19 indiquées indirectement par le déposant. Des 22 titres retrouvés dans *Ulrichsweb*, 18 titres ont un contenu académique et sont disciplinaires, 2 titres ont un contenu de vulgarisation et 2 titres sont plutôt centrés sur le commerce. Toujours d'après *Ulrichsweb*, 6 périodiques ont la mention *peer-reviewed*, 1 périodique est en Open Access et aucun n'est repris dans JCR.

Les relevés obtenus dans *Ulrichsweb* nous indiquent 17 périodiques qui ont le français comme langue de publication, 2 périodiques l'anglais, 2 le portugais et 1 le néerlandais.

Quant au pays de publication, la Belgique se classe en premier avec 10 titres, puis la France avec 7 titres, les Etats-Unis avec 2 titres, le Brésil, le Portugal et le Canada avec 1 titre chacun.

Les 68 références ORBi indexées par GS sont donc, en grande partie, des publications à caractère scientifique publiées dans des périodiques nationaux en langue française (ex. : *Cahiers de l'Urbanisme (Les)*, *Bulletin de la Société Royale Belge d'Etudes Géologiques et Archéologiques*, *G.E.O.*, *Géographes Associés*, *Historiens et Géographes*, *Mosella*, etc.). Le fait que ces titres ne sont pas indexés par les bases de données commerciales constitue un atout pour GS qui donne une meilleure visibilité de cette littérature locale qui ne serait pas visible sur le web autrement.

Discussion et Conclusions

5. Discussion et Conclusions

Les analyses effectuées dans ce chapitre nous ont permis de qualifier les publications des géographes de l'ULiège et de déterminer dans quelle mesure elles sont répertoriées par les bases de données commerciales WoS, Scopus et GeoRef et dans GS. Nous avons pu aussi évaluer la contribution du répertoire institutionnel ORBi dans la meilleure visibilité de cette littérature scientifique.

Le but principal de ce chapitre n'était pas la comparaison de la couverture des bases de données bibliographiques déjà mentionnées et du moteur de recherche GS en Géographie, nos résultats contribuent à renforcer les données existantes dans la littérature.

En référence à la **première hypothèse** posée dans ce chapitre, nos résultats montrent que la distribution des types de documents publiés par les membres du département de Géographie est différente en fonction de l'orientation du sujet des publications vers les SHS ou ST. Comme mentionné précédemment, les pratiques de publication spécifiques à la Géographie (Gorraiz et al., 2016) se retrouvent dans notre échantillon également, une majorité de chapitres ou parties de livres en français étant publiés en SHS. En ST les publications majoritaires correspondent aux articles de périodiques et communications en langue anglaise.

La première hypothèse selon laquelle, pour notre échantillon, *selon que l'on considère l'orientation vers les SHS ou les ST, la répartition des publications de notre échantillon entre les chapitres ou parties de livres et les articles de périodiques ou communications est nettement différente* est confirmée.

Les résultats qui nous ont conduite à la confirmation de notre première hypothèse nous donnent des informations pour tirer les conclusions sur notre **deuxième hypothèse** de ce chapitre : les articles de périodiques sont le type de publication principale quelle que soit la sous-discipline. Sur le total de notre échantillon 75, 6 % sont des articles des périodiques et communications publiées dans un périodique. Si l'on fait la distinction entre les SHS et ST, l'article de périodique reste majoritaire : 56, 5 % des références en SHS et 86 % en ST.

La deuxième hypothèse selon laquelle, pour notre échantillon, *quelle que soit la sous-discipline, l'article de périodique est la principale voie de publication* est confirmée.

En relation avec notre **troisième hypothèse**, qui vient compléter les résultats associés aux hypothèses précédentes en ce qui concerne cette fois la langue de publication, nous avons remarqué une dominance des articles de périodiques en anglais en ST (72%). Moins nombreux dans l'absolu, les articles en SHS sont majoritairement en français (86%). Nos résultats indiquent aussi que 57% de chapitres ou parties de livres sont publiés en français tandis que 42% sont publiés en anglais. Les pratiques de publication du département de Géographie semblent suivre la tendance générale actuelle en termes de publication scientifique, qui

Discussion et Conclusions

consiste principalement dans la publication d'une majorité d'articles de périodiques en anglais en ST tandis que les publications en SHS les publications sont majoritairement en français qu'elles soient des livres et chapitres de livres ou des articles des périodiques (Castree et al., 2005; Johnston & Sidaway, 2015). Cependant, malgré une dominance des articles de périodiques dans les publications des géographes de l'ULiège, quelle que soit l'orientation sous-disciplinaire, les chapitres et parties de livres ainsi que les articles et communications en SHS sont majoritairement en français.

La troisième hypothèse selon laquelle, pour notre échantillon, quelle que soit le type de document ou la sous-discipline, l'anglais est la principale langue de publication n'est confirmée que partiellement ;

En relation avec notre **quatrième hypothèse** nous avons vérifié que GS retrouvait bien 100 % des publications déposées dans ORBi par les géographes de ULiège, ceci grâce au moissonnage du répertoire institutionnel. Sans les références en provenance d'ORBi les résultats de GS serait de 84,8 % de recouvrement pour les articles de périodiques et 33,5 % pour les chapitres ou parties de livres.

Si la production de chapitres ou parties de livres par les géographes de l'ULiège est environ deux fois plus importante que l'usage que l'on semble généralement faire de ce type de document, c'est grâce au moissonnage d'ORBi que ces références sont toutes visibles et accessibles aux chercheurs *via* GS.

De manière générale, le moteur de recherche donne une réelle visibilité à ce type de littérature, sans toutefois la couvrir pleinement (voir chapitre 3 de la première partie), tandis que les outils traditionnels répertorient peu, voire très peu, ce type de documents.

La présence de références du département de Géographie dans ORBi leur donne une plus grande visibilité *via* GS, le répertoire a un rôle d'autant plus important à jouer dans leur diffusion. C'est en effet grâce à ORBi que cette littérature est visible et accessible aux chercheurs *via* GS.

La quatrième hypothèse selon laquelle grâce au moissonnage d'ORBi, GS retrouve toutes les publications des géographes de ULiège est confirmée.

En ce qui concerne les références retrouvées uniquement dans GS, exclusivement grâce au moissonnage d'ORBi, leur nombre est relativement important : pour les chapitres et parties de livres, 96 références (65,7%), dont 67 en français, tandis que les articles ou communications comptent 68 références (15%), dont 57 en français.

Parmi ces références, les « Sciences sociales & comportementales, psychologie » sont bien majoritaires pour les chapitres ou de parties de livres avec 74 références, dont la majorité (50) traitent plus précisément de « Géographie humaine & démographie » ainsi que pour les articles et les communications, 46 références dont 14 en « Géographie humaine & démographie » et 4 références en « Etudes régionales & interrégionales ».

Discussion et Conclusions

La cinquième hypothèse selon laquelle *les références retrouvées uniquement dans GS, exclusivement grâce au moissonnage d'ORBi, sont nombreuses, surtout pour la Géographie humaine* est confirmée.

Si on faisait abstraction des références répertoriées grâce à l'indexation d'ORBi uniquement, les performances de GS diminuent de 15 % pour les articles de périodiques et les communications et de près de 70 % pour les chapitres ou parties de livres. Les performances de GS restent tout de même plus élevées que celles des BD commerciales.

En prenant nos résultats en considération, nous pouvons constater que GS est un outil bibliographique qui recouvre les références indexées par les bases de données commerciales traditionnelles à 100 %. De plus, 183 références, soit 40 % de l'échantillon, sont couvertes par GS uniquement et lui sont « uniques ». GS permet de retrouver des références bibliographiques supplémentaires.

Ces résultats sont comparables aux résultats obtenus dans les chapitres précédents de la thèse, Știrbu *et al.*, (2015) et le chapitre trois de la première partie. Nous avions retrouvé précédemment des pourcentages relativement faibles de références « uniques », distinguant WoS de Scopus entre elles, avec un avantage pour Scopus, tandis que GS englobait les résultats des bases de données commerciales. Pour ce qui est de GeoRef, cet outil se démarquait uniquement pour les résultats « uniques » obtenus avec le mot-clé « sedimentation ».

La sixième hypothèse selon laquelle *si on fait abstraction des références en provenance d'ORBi, une proportion plus élevée de références est retrouvée par GS en comparaison des bases de données bibliographiques commerciales* est confirmée.

Si on regarde le classement des BD commerciales, Scopus et WoS suivent GS pour les articles de périodiques et les communications. D'autres études ayant comparé la couverture des bases de données confirment la supériorité de Scopus par rapport à WoS (Bartol *et al.*, 2014; Bosman *et al.*, 2006; Cavacini, 2015; Gavel & Iselid, 2008; Vieira & Gomes, 2009). Même si GS recouvre Scopus à 100%, 55 références supplémentaires par rapport à WoS et GeoRef sont à signaler pour Scopus. La couverture des BD commerciales dans les deux sous-disciplines principales est plus importante en ST, Scopus se démarque tout de même avec une meilleure couverture en SHS. Nos résultats sont confortés par des études effectuées dans des disciplines différentes (Bartol *et al.*, 2014; Bosman *et al.*, 2006; Cavacini, 2015).

La septième hypothèse selon laquelle, *le recouvrement de WoS et Scopus est très important et seul Scopus se distingue par un apport significatif de références supplémentaires par rapport à WoS, relevant souvent des SSH* est confirmée pour les publications des géographes de l'ULiège.

GeoRef se classe en dernière position pour la couverture des articles de périodiques et des communications. Le classement de GeoRef ne doit cependant pas mener à la conclusion que cet outil bibliographique est à écarter. Même si GS recouvre GeoRef

Discussion et Conclusions

à 100%, 35 références supplémentaires par rapport à WoS et Scopus sont à signaler. 90 % des références retrouvées par GeoRef sont en « Physique, chimie, mathématiques & sciences de la terre », ce qui indique une bonne couverture en ST. Cette performance rejoint les résultats du premier chapitre, où GeoRef occupait la troisième position pour le nombre de références répondant au mot-clé « sedimentation » mais se démarquait par un apport de références uniques et de différents types de documents.

La huitième hypothèse selon laquelle, *dans l'absolu, GeoRef couvre moins de références que WoS et Scopus mais se distingue par un apport de références supplémentaires en Géographie physique par rapport aux deux autres BD commerciales* est confirmée.

Pour ce qui est de la **neuvième hypothèse**, sur base de l'analyse des caractéristiques des périodiques dans lesquels les géographes ont publié, nous avons pu constater que la majorité (77,3 %) sont *peer-reviewed*. Pour ce qui est de la langue des périodiques, l'anglais est la première langue.

Les pratiques de publication du département de Géographie semblent suivre la tendance générale actuelle en termes de publication scientifique dans les périodiques anglophones. Ce type de publications est très bien couvert par l'ensemble des outils tandis que WoS se montre très sélectif pour ce qui concerne les « critères de qualité » traditionnellement reconnus pour les périodiques, qui sont les critères de Thomson et Reuters. Si bien que le nombre de documents couverts par WoS est finalement plus réduit.

La neuvième hypothèse selon laquelle *les périodiques dans lesquels les géographes ULiège publient sont majoritairement peer-reviewed en langue anglaise* est ainsi confirmée.

En ce qui concerne la **dixième hypothèse**, nos résultats indiquent que 43% de l'échantillon d'articles de périodiques et communications sont en français. Même si elle n'est pas majoritaire, par rapport aux publications en langue anglaise, la part des références en français est considérable, soit beaucoup plus que l'usage jamais relevé dans la littérature ou accessoirement dans les bibliographies des trois thèses que nous avons étudiées, pourtant réalisées au sein de l'Institution.

A l'analyse des références et des périodiques « uniques » de GS nous avons constaté aussi que le français était prédominant. Quant aux pays de publication, la France et la Belgique se classent le mieux avec respectivement 11 et 10 titres.

Ce pourcentage important de références en français s'explique par l'affiliation des auteurs à une institution francophone ainsi que par le fait qu'en général, les sujets d'intérêt local et régional sont publiés dans la langue native des auteurs (Gorraiz et al., 2016), dans des périodiques locaux ou nationaux, afin d'atteindre un public local, peut-être autre que les scientifiques et académiques de la discipline. Ainsi, au *Bulletin de la Société Géographique de Liège* revient 62 des références « uniques » à GS.

Discussion et Conclusions

Le bon résultat de GS dans le référencement des périodiques francophones locaux est dû à la présence de cette littérature dans le répertoire institutionnel ainsi qu'au moissonnage d'ORBi.

La dixième hypothèse selon laquelle *du fait de leur présence dans le répertoire institutionnel ORBi, les publications en langue française des géographes liégeois publiées dans les périodiques « locaux » sont plus visibles via GS que via les BD commerciales* est également confirmée.

En conclusion, les performances de GS sont aussi supérieures à celles des BD pour notre échantillon de publications des géographes de l'ULiège. Ceci conforte les résultats obtenus dans les chapitres de la première partie de la thèse. Le répertoire institutionnel ORBi contribue incontestablement à améliorer les performances de GS quant au recouvrement des publications des géographes et à donner une meilleure visibilité de littérature francophone en Géographie.

D'une façon plus générale, le référencement des périodiques dans d'autres langues que l'anglais par GS pourrait constituer un avantage majeur pour la communauté scientifique avec un accès à un panel de publications plus important, avec une plus grande diversité de sujets et lieux traités, et avec une réduction de l'hégémonie anglaise dans la discipline (Meeus et al., 2011 ; Schuermans et al., 2010).

Chapitre 2 - Téléchargements des articles en Géographie déposés dans ORBi

1. Introduction

L’investissement de moyens humain et financiers pour la mise en place d’un répertoire institutionnel tel qu’ORBi fut d’envergure, il est dès lors légitime de s’interroger sur l’impact de cet outil. Les statistiques d’usage du répertoire, soit en substance le nombre de visualisations des références ou de téléchargements des documents et diverses données associées, constituent une mine d’informations.

Le nombre de téléchargements des références d’un répertoire institutionnel, ayant un document associé et téléchargeable, nous paraît particulièrement significatif de l’intérêt que suscite le document auprès de la Communauté scientifique et de l’apport concret d’ORBi dans la diffusion du document. Cependant, tous les documents déposés dans ORBi n’ont pas le même statut quand il s’agit de considérer les téléchargements.

En effet, l’entrave que représente pour l’usager le coût de l’accès aux publications commerciales, peut avoir un impact positif sur le nombre de téléchargements de certains documents. Les publications OA, disponibles par ailleurs, présenteraient alors relativement moins de téléchargements *via* ORBi.

Par ailleurs, du fait de l’indexation partielle ou inexistante des périodiques nationaux ou régionaux par les bases de données traditionnelles, certaines publications sont alors essentiellement accessibles *via* les moteurs de recherche qui moissonnent les répertoires institutionnels, ce qui pourrait également avoir un impact positif sur les téléchargements à partir d’ORBi. Cette dernière considération, compte tenu de ce qui a été développé dans les chapitres précédents, s’applique particulièrement bien à un domaine comme celui de la Géographie.

De plus, l’origine et les catégories d’utilisateurs qui sont intéressés par les publications des géographes liégeois déposées dans le répertoire institutionnel nous paraissent être des informations essentielles. ORBi favorise-t-il une diffusion internationale de publications dont l’impact aurait pu se confiner à une région, d’autant plus si ces publications sont de langue française ?

Considérant la répartition des téléchargements par pays, information disponible dans les statistiques d’ORBi, nous entrevoyons que le nombre de téléchargements par pays pourrait être influencé par différents facteurs socio-économiques.

Dans ce chapitre, à partir d’une partie de l’échantillon de publications du chapitre précédent, soit les articles et les communications publiées dans un périodique, nous allons nous intéresser aux téléchargements effectués à partir d’ORBi et aux facteurs qui pourraient les influencer.

Introduction

Dans un premier temps, le nombre et la répartition mondiale des téléchargements de l'échantillon seront considérés, pour l'ensemble des publications et aussi en établissant une distinction en fonction de la langue de publication ou de leur recouvrement par les différents outils bibliographiques (cf. chapitre précédent).

La seconde partie de ce chapitre est la plus conséquente et consiste dans l'identification des différents facteurs qui pourraient influencer les téléchargements de ces références, par une approche statistique.

1.1. Evaluation de l'impact d'une publication : citations vs téléchargements

Les pratiques de la communauté « publish or perish », bien ancrées depuis quelques décennies dans le milieu académique, ont conduit à une analyse systématique des citations. En effet, celles-ci sont largement prises en compte dans les processus d'évaluation.

En parallèle, avec le développement d'Internet et de la numérisation de l'édition, un nombre croissant de documents universitaires sont accessibles et utilisés sous forme numérique *via* des bases de données bibliographiques ou différents types de répertoires. Les données relatives à l'usage de ces documents sont continuellement enregistrées et stockées : visualisations, téléchargements, pays des usagers, moyen d'accès, etc.

Ces données d'usage, qui permettent l'observation de l'activité scientifique immédiatement après la publication ou le dépôt, ont attiré l'attention des spécialistes de l'information. En effet, elles revêtent un caractère quasi instantané que l'on ne retrouve pas lors de l'examen des citations. Ces dernières apparaissent plus tard après la publication, avec des décalages qui peuvent atteindre plusieurs années. Ainsi, par le biais des citations, les effets d'une publication sont observés et interprétés avec un délai considérable, dans la plupart des cas.

On a donc vu apparaître des études qui tentent de cerner le comportement des usagers (Davis & Price, 2006; Davis & Solla, 2003; Wang et al., 2014; Wang et al., 2013) et les moyens par lesquels les usagers accèdent aux références (Ikeda & Inoue, 2009; Nicholas et al., 2008; Organ, 2006; Robinson, 2009). D'autres chercheurs se sont intéressés aux liens qui pouvaient exister entre le nombre de téléchargements des documents et les citations (Bollen & Sompel, 2008; Davis & Fromerth, 2007; Moed, 2005; Schloegl & Gorraiz, 2010, 2011). Ils ont alors suggéré d'utiliser le nombre de téléchargements en tant que nouveau moyen d'évaluation et proposé un « Download Impact Factor » (Bollen et al., 2009, 2008; Shepherd, 2007).

Les facteurs d'influence des téléchargements sont rarement pris en compte, et les études ont été effectuées pour des articles provenant de périodiques ou de bases de données. Jamali & Nikzad (2011) et Subotic & Mukherjee (2014) ont analysé l'influence que la longueur des titres d'articles de périodiques relevés dans PLoS

(Public Library of Science) et Science Direct pourraient avoir sur les téléchargements. Leurs résultats indiquent un plus grand nombre de téléchargements pour les articles avec des titres courts ou amusants.

Dans une analyse effectuée sur les téléchargements des articles de périodiques chinois publiés entre 2006 et 2008, Duan & Xiong (2017) ont mis en évidence des corrélations faibles entre les téléchargements et les différentes caractéristiques des articles : longueur du titre, nombre d'auteurs, ainsi que le nombre de mots clés. Cependant, des corrélations modérées à élevées existent entre les téléchargements initiaux - définis comme des téléchargements réalisés au cours de la première année après la publication - et des téléchargements « totaux », ce qui suggère qu'il est possible de prévoir des téléchargements « totaux » en fonction des téléchargements initiaux.

Guerrero-Bote & Moya-Anegón (2014) ont pris en compte l'influence sur les téléchargements de la langue de certains périodiques en médecine publiés par Elsevier. Les téléchargements, leur ont été fournis par l'éditeur et les citations ont été relevées dans Scopus et Science Direct. Pour des périodiques en anglais, leurs résultats indiquent des téléchargements importants des Etats-Unis, du Royaume-Uni et de la Chine. Les périodiques en français ont plus été téléchargés par des pays francophones : France, Tunisie, Canada, Algérie, Belgique et la Suisse. Les auteurs ont constaté que les citations revenant aux revues non anglophones étaient moins importantes par rapport aux téléchargements (différentiel supérieur à 50 %). Selon ces auteurs, les revues non anglophones seraient plus citées par des revues non indexées dans Scopus.

1.2. Téléchargements à partir de répertoires institutionnels

Bien qu'une attention croissante soit accordée aux téléchargements, la recherche sur ce type de statistiques issues des répertoires institutionnels est quasi inexistante à l'heure actuelle.

Bonilla-Calero (2008) a analysé les statistiques de téléchargement et les citations des articles de Physique du répertoire institutionnel de l'Université de Strathclyde entre 2006 et 2007. Une corrélation positive entre le nombre de citations et les téléchargements a été trouvée. Il a également été montré qu'un nombre de pays distincts avaient téléchargé et cité les documents, avec les Etats-Unis et le Royaume-Uni en tête du classement. Les résultats indiquaient aussi que les documents les plus fréquemment téléchargés et cités étaient les versions post-print des articles.

La langue des publications n'est pas prise en compte dans l'interprétation de l'étude de Robinson (2009) sur HKIE, répertoire institutionnel de Hong Kong Institute of Education. L'auteur constate que la majorité des accès au répertoire institutionnel proviennent de Hong Kong même, d'autres de la Chine continentale et de Taiwan

Objectifs et hypothèses

ainsi que des Etats-Unis et de l'Australie. La majorité des accès sont effectués à la suite de recherches de texte *via* des moteurs de recherche tels que Google, Yahoo et Baidu, ce dernier étant principalement utilisé pour la recherche de texte en chinois. Mais à la lecture de l'article, nous relevons que le répertoire, qui contient une littérature majoritairement en anglais (70% des articles), fait l'objet de téléchargements majoritairement par des pays où l'anglais est couramment parlé (Chine continentale et Taiwan) ou anglophones (États-Unis et Australie)

La littérature traitant du téléchargement des références présentes dans les répertoires institutionnels est très rare. Cela peut éventuellement s'expliquer par la difficulté de mettre en relation des observations relatives aux téléchargements (nombre, origine géographique) avec d'autres données bibliométriques (comme les citations des références) ou qualitatives (concernant, par exemple, les usagers). En effet, il faut aussi tenir compte de l'existence d'autres facteurs interférant dont, par exemple, la disponibilité - ou pas - d'autres voies d'accès aux documents que le répertoire institutionnel ou des qualités du document lui-même (type, langue...). Nous allons tenter de relever le défi.

2. Objectifs et hypothèses

Les analyses effectuées dans ce chapitre reposent sur les téléchargements d'une partie de l'échantillon utilisé dans le premier chapitre de cette partie, soit 453 articles et communications publiés dans un périodique entre 2000 et 2014 par les membres du département de Géographie, et encodés dans ORBi.

Un **premier objectif** de ce chapitre est de quantifier et représenter la distribution des téléchargements des publications au niveau mondial, dans leur ensemble ou en tenant compte de certaines caractéristiques comme la langue de publication (anglais ou français). En relation avec le chapitre précédent, nous distinguons également les références retrouvées uniquement par GS, celles retrouvées par le moteur de recherche *via* les liens d'ORBi uniquement, et enfin les références communes à tous les outils.

Notre **second objectif**, est de vérifier si des facteurs socio-économiques pourraient influencer les téléchargements de ces articles, par une approche statistique.

Ceci nous amène à poser les hypothèses suivantes :

Hypothèse 1 : Par leur présence dans le répertoire institutionnel ORBi, la diffusion des résultats de recherche des géographes liégeois est assurée au niveau mondial ;

Hypothèse 2 : Les téléchargements des références en anglais sont effectués principalement par les pays anglophones ;

Hypothèse 3 : Les téléchargements des références en français sont effectués principalement par les pays francophones ;

Hypothèse 4 : Du fait que les publications en anglais sont plus souvent référencées par plusieurs plateformes et bases de données, le nombre de téléchargements à partir

d'ORBi est moins important tandis qu'un nombre plus élevé de téléchargements sera enregistré pour les références en français ;

Hypothèse 5 : Les téléchargements d'une part des références « uniques » de GS, d'autre part des références répertoriées par GS uniquement *via* ORBi et enfin des références couvertes par tous les outils bibliographiques, présentent une répartition différente au niveau mondial ;

Hypothèse 6 : Les facteurs socio-économiques des différents pays influencent le nombre total de téléchargements effectués ;

Hypothèse 7 : Les facteurs socio-économiques des différents pays influencent le nombre de téléchargements effectués quelle que soit la langue des références ;

3. Méthodes

3.1. Distribution des téléchargements au niveau mondial

Les nombres des téléchargements correspondant aux *handles*, numéro d'identification unique des références déposées dans le répertoire institutionnel, et les pays associés ont été fournis par l'équipe qui gère la plateforme, dans un fichier *Excel*. Les téléchargements ont été comptabilisés à partir du premier jour de dépôt des articles dans ORBi et jusqu'au 15 décembre 2015. Sur base des données du fichier, nous avons réparti le nombre de téléchargements par pays et distingué ceux correspondant respectivement aux références de langue anglaise et française.

Nous nous sommes basée sur une partie des résultats du précédent chapitre pour distinguer les téléchargements des références retrouvées par GS uniquement et *via* les liens d'ORBi, ainsi que les références communes aux quatre outils bibliographiques (GS, Scopus, WoS, GeoRef). Les nombres de téléchargements de ces différentes catégories de références ont également été répartis par pays.

Les nombres de téléchargements par pays ont été représentés sur des cartes générées avec le programme ArcGis Desktop 9.0.

3.2. Facteurs susceptibles d'influencer les téléchargements

Dans le but de déterminer les facteurs qui influencent les téléchargements des références déposées par les géographes dans ORBi ainsi que les variations relatives aux téléchargements des articles en anglais et des articles en français, des analyses statistiques de régression ont été effectuées.

Les nombres de téléchargements associés à l'ensemble de l'échantillon, aux articles publiés en anglais et à ceux publiés en français constituent les variables dépendantes de notre modèle de régression. Au début de notre analyse statistique, les variables explicatives retenues étaient au nombre de trois : le pays, le PIB et la population.

Méthodes

La liste des variables explicatives retenues a évolué au cours de l’analyse statistique qui est détaillée dans les résultats. L’interface World Bank Data⁵⁵, a été utilisée pour relever les données correspondant à quasi toutes les variables explicatives. Cette interface fournit l’accès gratuit à des statistiques des pays pour différents secteurs tels que l’économie, l’éducation, l’énergie, l’environnement, la santé, etc.

La page web Internet World Stats⁵⁶, qui fournit des statistiques mondiales actualisées sur l’utilisation d’Internet pour plus de 243 pays différents et régions du monde, a aussi été consultée pour une seule des variables explicatives (le nombre d’utilisateurs Internet par pays).

Nos premiers essais pour générer des modèles statistiques, ont été réalisés avec les logiciels XLSTAT et *Statistica*. Ils ne sont pas présentés dans les résultats car ces logiciels ne nous permettaient pas de traiter toutes les variables explicatives. Aussi, nous avons finalement choisi de travailler avec JMP Pro 13, sur le conseil de Madame Timmermans⁵⁷.

3.2.1. Régression multiple

Une fois toutes les variables explicatives potentielles sélectionnées nous avons examiné la distribution des variables et procédé à l’analyse de corrélation entre les variables quantitatives. Une analyse « TreeMap » a été réalisée dans JMP pour détecter les observations manquantes.

Nous avons développé un premier modèle préliminaire au modèle final, dans lequel la variable dépendante « nombre total de téléchargements » est liée à toutes les variables explicatives potentielles. L’algorithme utilisé pour l’ajustement du modèle est la méthode des moindres carrés standard. Les indicateurs globaux de la qualité du modèle et les valeurs p des variables ont été examinés.

Un modèle final a été développé uniquement avec les variables statistiquement significatives. Ensuite, afin de réduire la complexité du modèle final et pour augmenter la précision des estimateurs et produire un modèle plus simple d’utilisation pour l’interprétation et la prédiction, nous avons utilisé un algorithme de sélection automatique de variables ou méthode « pas à pas ». Il existe plusieurs algorithmes de ce type. Celui utilisé dans notre cas est *forward stepwise* avec successivement les critères BIC et AICc avec procédure ascendante (modèle vide qui ajoute progressivement des variables), en commençant pas la variable la plus corrélée avec la variable dépendante (Kutner et al., 2005).

Dans le cadre de ces procédures, des modèles avec une complexité croissante sont considérés et comparés.

⁵⁵ <https://donnees.banquemondiale.org/pays>

⁵⁶ <http://www.internetworkworldstats.com>

⁵⁷ Statisticienne et consultante dans le Service de Statistique Appliquée, Département de Mathématique de l’ULiège

Après la première étape, l'algorithme sélectionne parmi les variables explicatives restantes celle qui fournit la plus grande réduction dans la variance résiduelle (non expliquée) de la variable dépendante. A chaque itération, le programme génère un F-test partiel afin de vérifier si les variables ont une quantité de variation significative par rapport aux variables restantes dans la régression.

Cette méthode a l'avantage de sélectionner automatiquement les variables significatives et de supprimer les variables qui ne le sont pas. De plus, cette méthode nous permet d'éviter le plus possible les problèmes de colinéarité.

Lors de la validation d'un modèle, un certain nombre d'hypothèses doivent être vérifiées :

- existence de données complètes et fiables ;
- linéarité de la relation découlant de la modélisation statistique ;
- caractère additif des termes de l'équation ;
- indépendance des variables explicatives ;
- normalité dans la distribution des résidus ;
- constance dans la variance des résidus ;
- indépendance des termes d'erreurs ;
- représentativité de l'échantillon.

Les hypothèses sous-jacentes ont aussi été vérifiées.

Afin d'éviter des problèmes d'interprétation et de lecture, les résultats du premier modèle ne seront pas présentés dans les résultats. Nous détaillerons le développement du modèle final, que nous avons retenu, en utilisant la méthode de sélection des variables « pas à pas » (*stepwise*).

Les variables dépendantes ainsi que toutes les variables explicatives quantitatives ont subi une transformation logarithmique qui nous permet de mieux percevoir les différences entre les niveaux de téléchargements.

4. Résultats

4.1. Distribution des téléchargements au niveau mondial

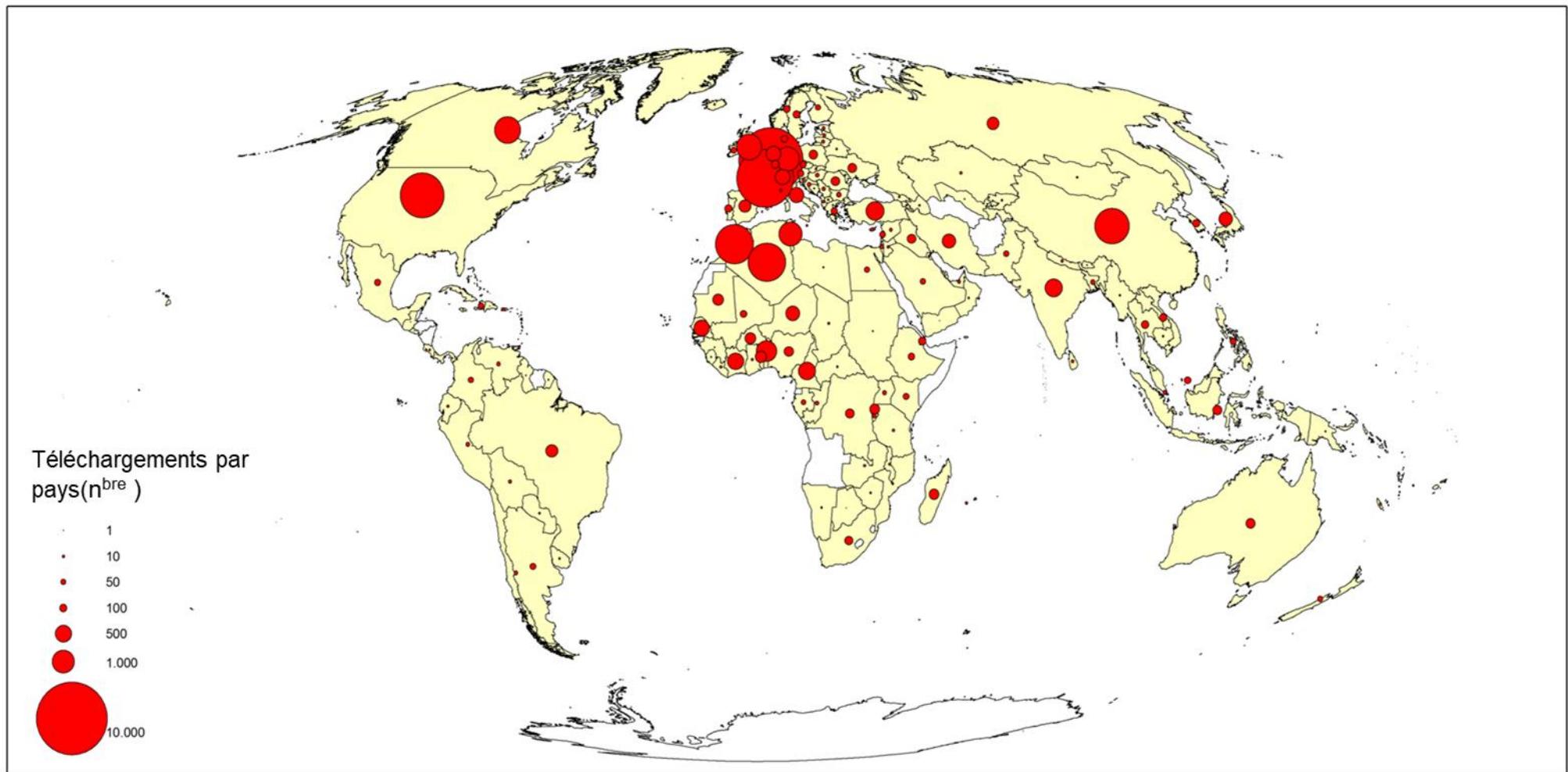
Pour notre échantillon de 453 articles et communications publiées dans un périodique, 44 428 téléchargements ont été enregistrés, dont 43 788 valides, soit en moyenne 97 téléchargements par article.

640 téléchargements ont été exclus de l'analyse car il a été impossible de les situer de manière précise géographiquement, soit 157 téléchargements pour les références en anglais et 483 téléchargements pour les références en français. Au lieu du pays d'origine du téléchargement, nous avons trouvé des indications comme : Yougoslavie, Satellite Provider, Europe, N/A, Anonymous Proxy, Asia/Pacific Région, etc.

Comme on peut visuellement le remarquer sur la Carte 6, l'usage des 453 références d'articles du département de Géographie, déposées dans ORBi, est principalement

Résultats

réparti entre l'Europe, les Etats-Unis, les pays francophones du Maghreb, le Canada et la Chine. Viennent ensuite les pays de l'Afrique Sub-saharienne, l'Inde et la Turquie.



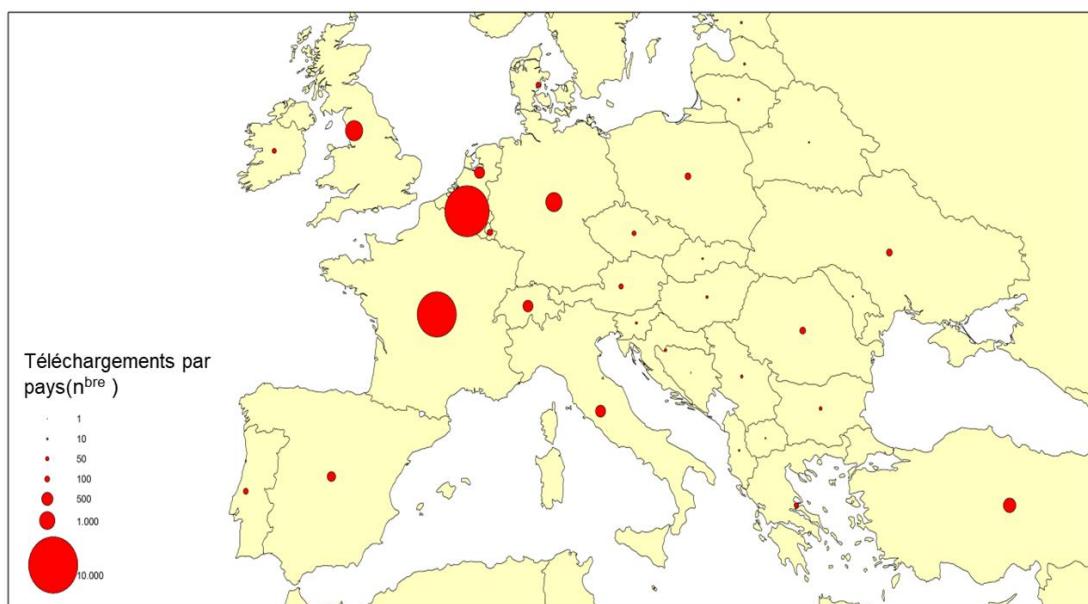
Carte 6 - Distribution au niveau mondial des téléchargements ORBi correspondant aux 453 références publiées entre 2000 et 2014

Résultats

Plus précisément, la Belgique (8 095 téléchargements) et la France (6 348 téléchargements), représentés sur la Carte 7, sont les deux pays ayant téléchargé le plus. En troisième position se trouvent les Etats-Unis avec 3 777 téléchargements. Viennent ensuite les pays du Maghreb, le Maroc (2 833 téléchargements), l'Algérie (2 732), la Tunisie (1 065). Et aussi, la Chine qui enregistre 2 384 téléchargements ainsi que le Canada, le Royaume-Uni et l'Allemagne qui effectuent chacun plus de mille téléchargements.

Il est à noter que, sur un total de 164 pays, les dix pays mentionnés ci-dessus constituent le groupe de pays ayant téléchargé le plus, ils comptabilisent 30 948 téléchargements soit 70,7% du nombre total de téléchargements (Tableau 16). Ce « phénomène » de concentration des téléchargements sur quelques pays se vérifie également si on distingue les publications sur base de la langue : 4 pays sur 146 ont effectué 46,4 % des téléchargements pour les références en anglais et 6 pays sur 127 ont effectué 70,6 % des téléchargements pour les références en français.

Les pays francophones de l'Afrique Sub-saharienne sont aussi à l'origine de nombreux téléchargements, toutes langues de publication confondues : le Benin (806 téléchargements), le Cameroun (575 téléchargements), la Côte d'Ivoire (516 téléchargements) et le Sénégal (466 téléchargements).



Carte 7 - Distribution au niveau européen des téléchargements ORBi correspondant aux 453 références publiées entre 2000 et 2014.

Résultats

	Nb. Téléch./pays	Nb. de Pays	Somme Téléch.	% du total Téléch.
TOT_TELECH	1--10	62	286	0,65
	11--50	35	863	2
	51--100	22	1667	3,8
	101--500	29	6845	15,6
	501--1000	6	3179	7,2
	1001--8095	10	30948	70,7
	164	43 788		100
TELECH_EN	1--10	68	284	2
	11--50	36	1001	6
	51--100	17	1228	7,5
	101--500	17	3386	20,7
	501--1000	4	2835	17,4
	1001--3000	4	7562	46,4
	146	16 296		100
TELECH_FR	1--10	65	273	1
	11--50	27	655	2,4
	51--100	9	704	2,6
	101--500	17	4145	15,2
	501--1000	3	2264	8,2
	1001--3000	6	19337	70,6
	127	27 378		100

Tableau 21 - Récapitulatif chiffré des téléchargements effectués par les pays, pour l'ensemble des publications (TOT_TELECH), pour les publications en anglais (TELECH_EN) et pour les publications en français (TELECH_FR)

Comme nous pouvons remarquer sur la Carte 6 et dans le Tableau 22, les pays francophones dans leur ensemble ont effectué la majorité des téléchargements soit 27 472 téléchargements (62,7% du total des téléchargements). 15 753 téléchargements (57 % des téléchargements effectués par les pays francophones) sont effectués par la Belgique, la France et le Canada. 11 719 téléchargements (43 % des téléchargements effectués par les pays francophones) sont effectués par les autres pays francophones.

Les pays anglophones quant à eux ont effectué 7 794 téléchargements, soit 17,8 % du total des téléchargements.

Les pays qui ne font pas d'affiliation francophone ou anglophone sont en nombre plus important et ont effectué 8 522 téléchargements soit 19,5 % du total des téléchargements.

Résultats

	Nombre d'articles	Moyenne Téléch. par article	Tot. Telech.	% du tot telech	Telech. Pays francophones	%	Telech.Pays anglophones	%	Telech. Autres Pays	%
TOT_TELECH	453	97	43 788	100	27 472	62,7	7 794	17,8	8 522	19,5
TELECH_EN	257	63	16 296	37,2	3 182	19,5	5 439	33,4	7 675	47,1
TELECH_FR	193	142	27 378	62,5	23 868	87,2	1 441	5,3	2 069	7,5

Tableau 22 - Détails des téléchargements effectués par les pays, pour l'ensemble des publications (TOT_TELECH), pour les publications en anglais (TELECH_EN) et pour les publications en français (TELECH_FR) avec la répartition des téléchargements effectuées par les différentes catégories de pays et les pourcentages associés

4.1.1. Distribution des téléchargements des articles en anglais

Pour notre échantillon, 257 articles et communications publiées dans un périodique sont en langue anglaise, soit 56,7% du total de l'échantillon (Tableau 22). Ces références ont fait l'objet de 16 296 téléchargements (37,2 % du total) et se répartissent sur 146 pays, soit en moyenne 63 téléchargements par article.

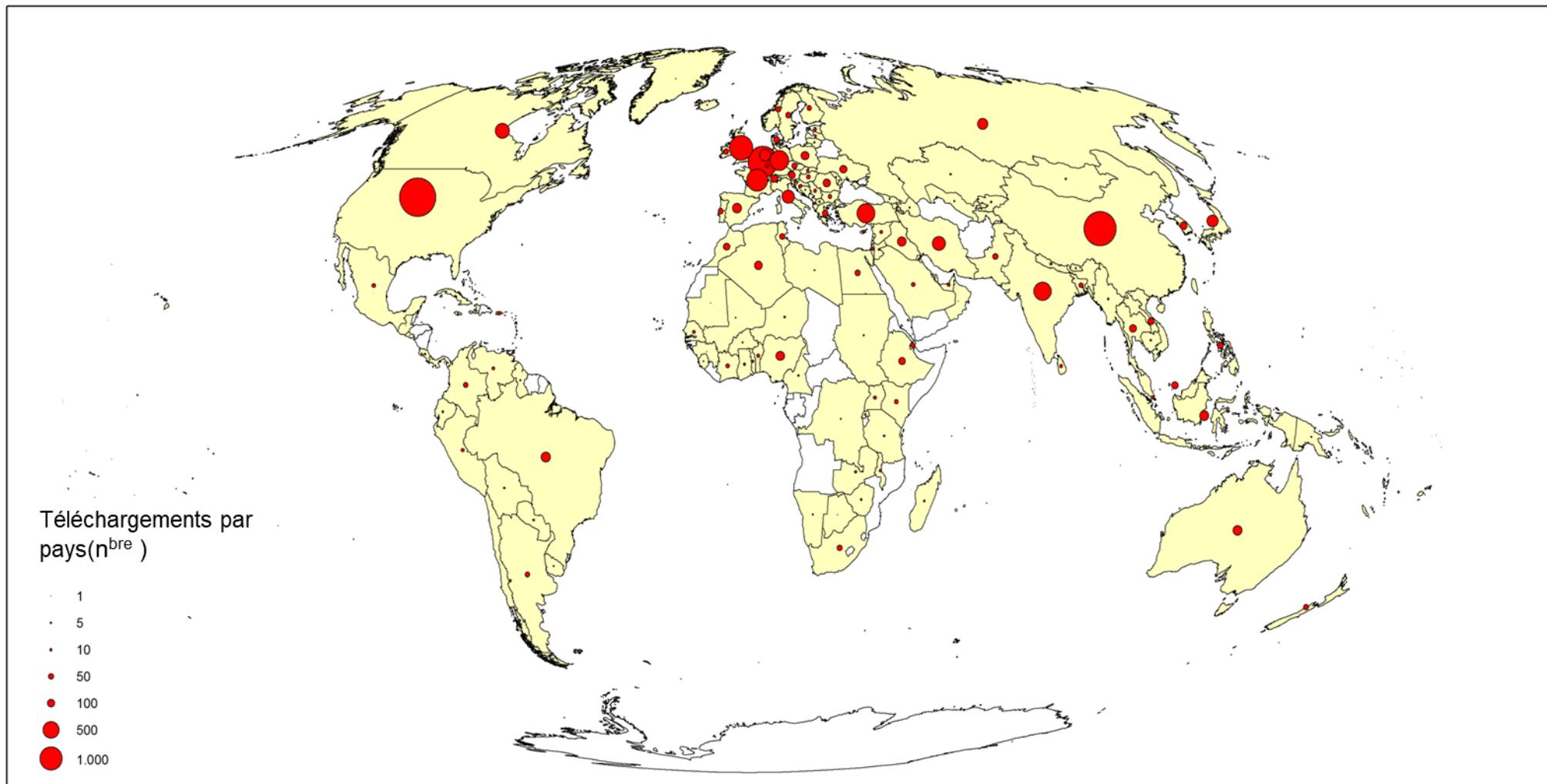
Comme on peut facilement le remarquer sur la Carte 8, l'usage des 257 références est principalement réparti entre les Etats-Unis, l'Asie et l'Europe. Par comparaison avec la Carte 6, qui illustre les téléchargements de l'ensemble de l'échantillon, c'est au niveau des pays d'Afrique que la différence se marque le plus, ces derniers téléchargeant beaucoup moins lorsque l'on fait abstraction des références en français et que seules sont considérées les références en anglais.

Ce n'est pas la seule différence apparaissant dans la distribution des téléchargements au niveau mondial. La Belgique et la France ne sont plus les deux pays qui téléchargent le plus (Carte 9) : 5 439 téléchargements (33,4 %) ont été effectués par des pays anglophones dont les Etats-Unis (2 622 téléchargements), le Royaume Uni (1 118 téléchargements), l'Inde (606 téléchargements), l'Australie (174 téléchargements) et le Nigeria (158 téléchargements).

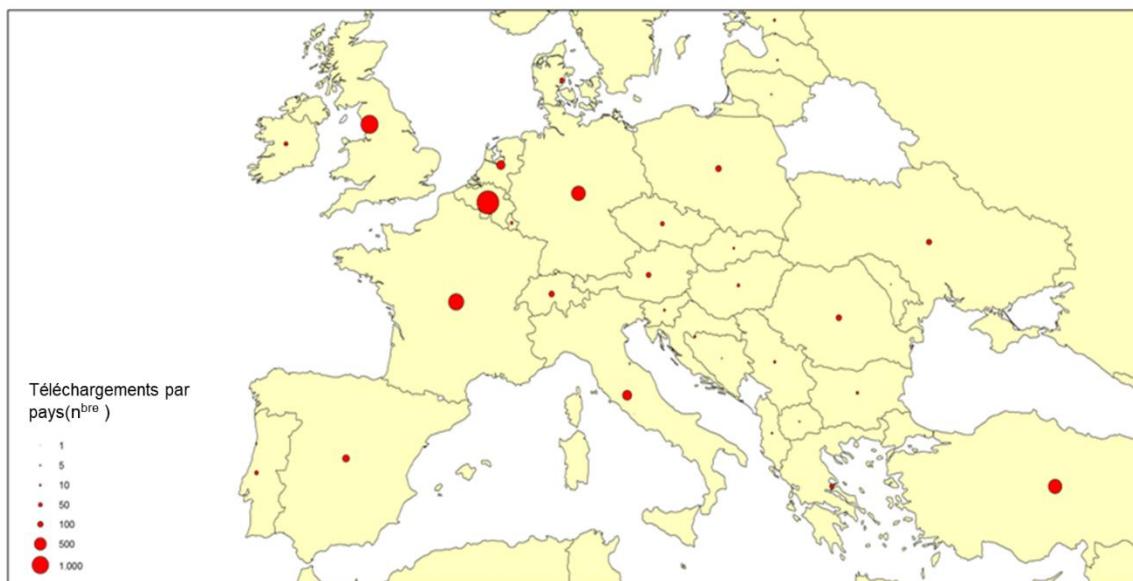
Les pays francophones ont effectué 3 182 téléchargements (19,5 %), dont la Belgique (1 737 téléchargements) et la France (884 téléchargements), l'Algérie (117 téléchargements), la Suisse (108 téléchargements), le Maroc (79 téléchargements), la Tunisie (51 téléchargements).

Les 7 675 téléchargements restants (47 %) ont été effectués par 99 pays, dont la Chine (2 085 téléchargements) qui effectue un nombre de téléchargements très proche du nombre de téléchargements qui lui était attribué pour l'ensemble des références. Viennent ensuite l'Allemagne (701 téléchargements), la Turquie (644 téléchargements), l'Iran (349 téléchargements), l'Italie (325 téléchargements), les Pays-Bas (266 téléchargements), le Japon (266 téléchargements), la Russie (215 téléchargements), etc.

Résultats



Carte 8 - Distribution mondiale des téléchargements ORBi correspondant aux références publiées en anglais



Carte 9 - Zoom sur la distribution des téléchargements ORBi correspondant aux références publiées en anglais en Europe

Résultats

4.1.2. Distribution des téléchargements des articles en français

Les 193 références (42,6%) publiées en français ont enregistré 27 378 téléchargements, soit 62,5 % du total des téléchargements, avec une moyenne de 142 téléchargements par article (Tableau 22). Ces téléchargements sont répartis sur 127 pays (Carte 10).

Ces références sont principalement téléchargées par les pays développés de l'ouest de l'Europe (Carte 11), notamment la Belgique (6 338 téléchargements) et la France (5 462 téléchargements), qui cumulent 11 800 téléchargements, soit 43% des téléchargements des références en français.

Viennent ensuite les pays francophones du Maghreb : le Maroc (2 754 téléchargement), l'Algérie (2 615 téléchargements) et la Tunisie (1 014 téléchargements). Puis, vient le Canada (912 téléchargements) et ensuite, avec des valeurs décroissant mais qui restent relativement importantes, les pays francophones Sub-sahariens : le Benin (786 téléchargements), le Cameroun (566 téléchargements), la Côte d'Ivoire (490 téléchargements), le Sénégal (439 téléchargements), le Niger (402 téléchargements) et le Togo (240 téléchargements). Il faut souligner qu'à peu près la moitié des téléchargements des références en français, 11 488 soit 42 % du total, ont été effectués par les francophones d'Afrique.

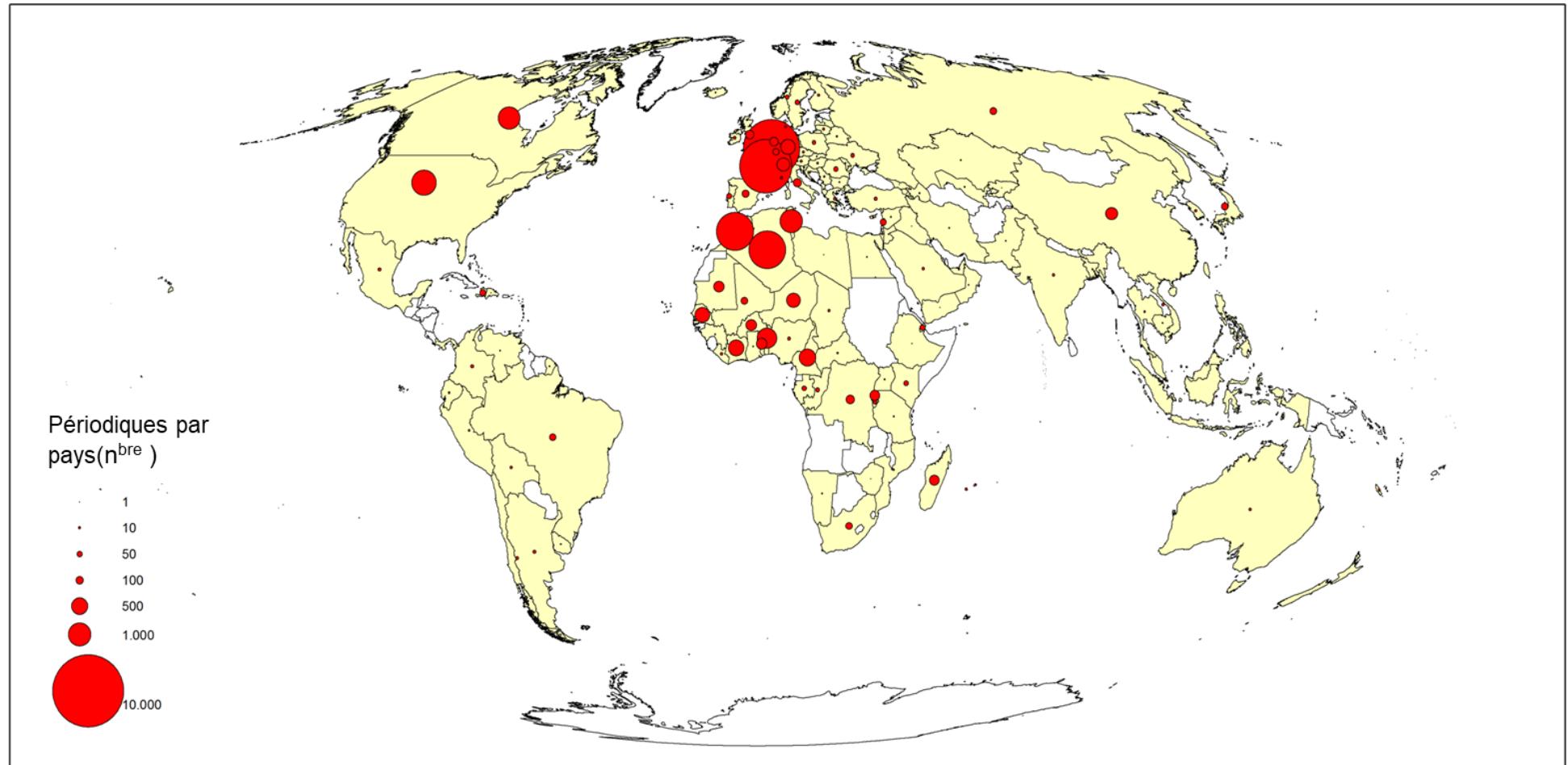
Les pays francophones dans leur ensemble ont effectué 23 858 téléchargements, soit 87,2 % du total des téléchargements des références en français.

Les pays anglophones ont effectué 1 441 de téléchargements pour les références en français, soit 5,5 % du total. Les Etats-Unis enregistrent 1 154 téléchargements, le Royaume-Uni 158 téléchargements, l'Australie 10 téléchargements, etc.

Les pays qui ne rentrent pas dans les catégories des pays « anglophones » ou « francophones », ont effectué 2 069 téléchargements, soit 7,5 % du total des téléchargements des références en français. L'Allemagne comptabilise 425 téléchargements, la Chine seulement 294 téléchargements, les Pays-Bas 140 téléchargements, l'Italie 126 téléchargements, etc.

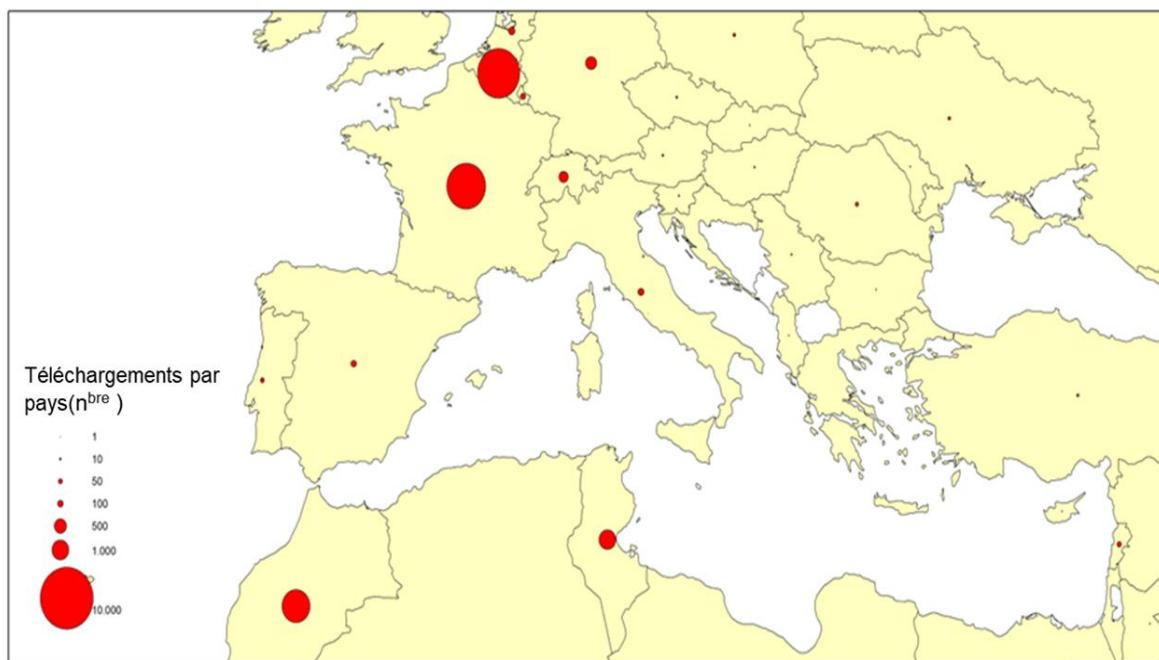
En premières conclusions, nous constatons que notre distribution de téléchargements est semblable à celle de Guerrero-Bote & Moya-Anegón (2014). Dans leur analyse, les auteurs ont trouvé que les pays ayant le plus grand pourcentage de téléchargements dans les revues en français de leur échantillon était la France, la Tunisie, le Canada, l'Algérie et la Belgique, tous des pays francophones.

Nous retenons également que les 193 références publiées en français ont été téléchargées presque deux fois plus que les 257 références publiées en anglais.



Carte 10 - Distribution au niveau mondial des téléchargements ORBi correspondant aux références publiées en français

Résultats



Carte 11 - Distribution des téléchargements ORBi correspondant aux références publiées en français en Europe

4.1.3. Distribution des téléchargements correspondant aux références retrouvées par GS uniquement

Les 115 articles et communications « uniques » à GS, qui représentent 25,4 % de l'échantillon d'articles et communications, ont enregistré 17 572 téléchargements, soit 40 % du nombre total de téléchargements (Tableau 23).

663 téléchargements ont été comptabilisés pour les articles en anglais (20 références) soit en moyenne 33 téléchargements par article, tandis que 11 911 téléchargements ont été comptabilisés pour les articles en français (94 références) soit en moyenne 127 téléchargements par article.

Pour l'article en portugais 55 téléchargements ont été comptabilisés.

367 téléchargements ont été exclus de l'analyse car il a été impossible de les situer géographiquement de manière précise.

La France (3 068 téléchargements) et la Belgique (3 981 téléchargements) sont les plus grands utilisateurs de ces références. Au Canada reviennent 542 téléchargements effectués pour les 115 références (Carte 12).

La Carte 12 indique des nombres de téléchargements importants pour les pays du Maghreb, dont le Maroc (2 131 téléchargements) et l'Algérie (1 598 téléchargements), la Tunisie (544 téléchargements). Les Etats Unis ont effectué 831 téléchargements.

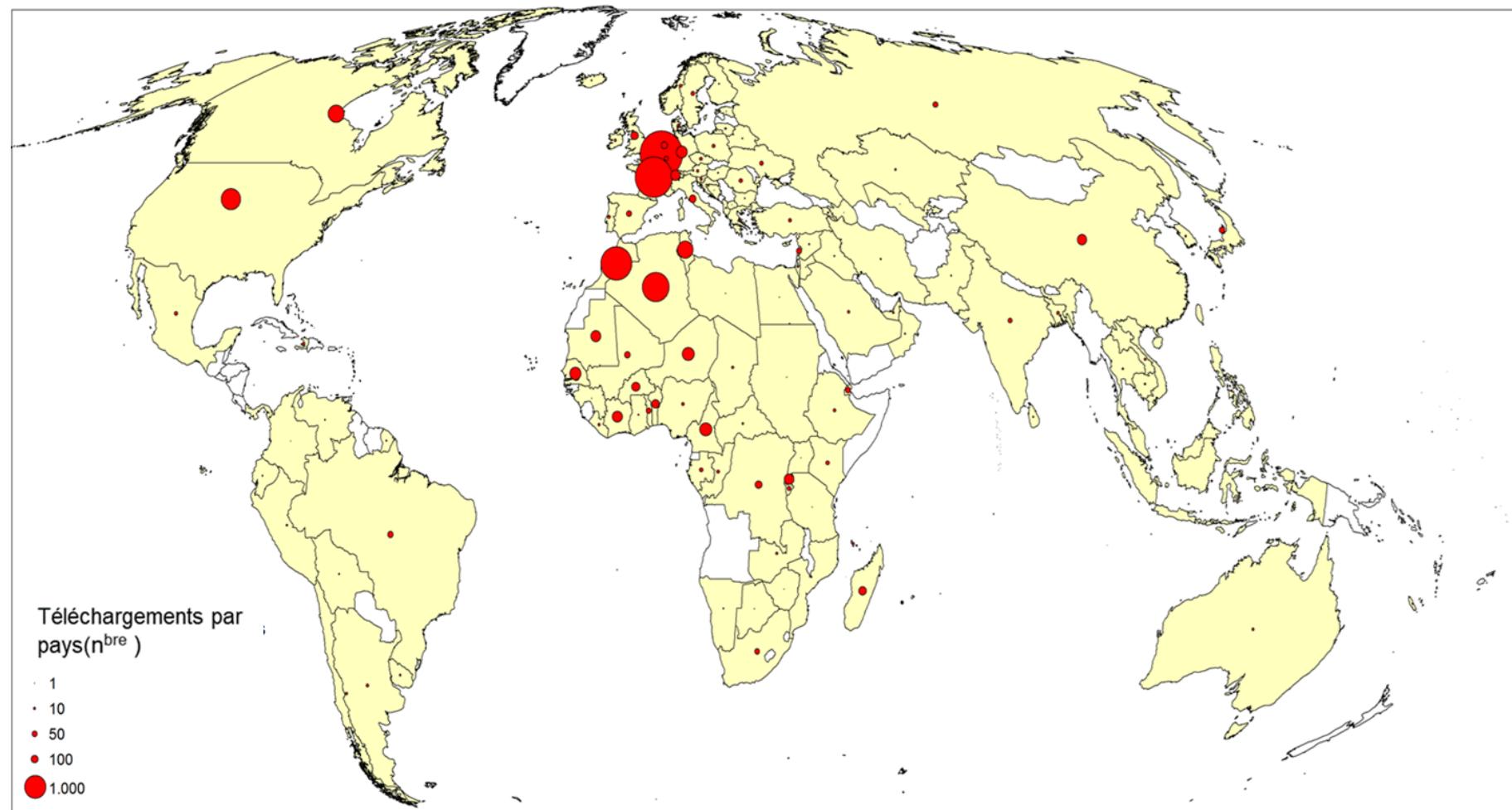
Parmi les pays asiatiques, la Chine se distingue avec 203 téléchargements.

Les pays francophones sont les plus grands utilisateurs de ces références, majoritairement en langue française.

	Nombre d'articles	% du Total	Anglais %*	Français %*	Tot. Telech.	Moyenne par article FR	Moyenne par article EN
GS_UNIQ	115	25,4	20(17,4)	94(81,7)	17 572	127	33
GS_UNIQ_ORBi	68	15	9(13,2)	57(84)	6130	98	61
BD&GS	97	21,4	97	--	6615	--	68

Tableau 23 - Téléchargements des articles et communications publiés en français ou anglais et retrouvés uniquement par GS (GS_UNIQ), retrouvés uniquement par GS et uniquement *via* ORBi (GS_UNIQ_ORBi) ou retrouvés par l'ensemble des outils bibliographiques (BD&GS) * % par rapport au nombre total d'articles correspondants

Résultats



Carte 12 - Distribution au niveau mondial des téléchargements effectués à partir d'ORBi correspondant aux 115 références « uniques » à GS

4.1.4. Distribution des téléchargements correspondant aux références retrouvées par GS via les liens ORBi uniquement

Les 68 articles et communications référencés par GS *via* les liens ORBi uniquement, qui représentent 15 % de l'échantillon d'articles et communications, ont enregistré 6 130 téléchargements, soit 13,9 % du nombre total de téléchargements (Tableau 23).

90 téléchargements ont été exclus de l'analyse car il a été impossible de les situer géographiquement.

Les 9 articles et communications en anglais ont comptabilisé 551 téléchargements, soit en moyenne 61 téléchargements par référence.

Les 57 articles et communications en français ont comptabilisé 5 579 téléchargements, soit en moyenne 98 téléchargements par référence.

La Carte 13 indique un nombre de téléchargements plus important pour les pays européens et plus particulièrement les pays francophones d'Europe, puis du Maghreb et de l'Afrique subsaharienne ainsi que du Canada.

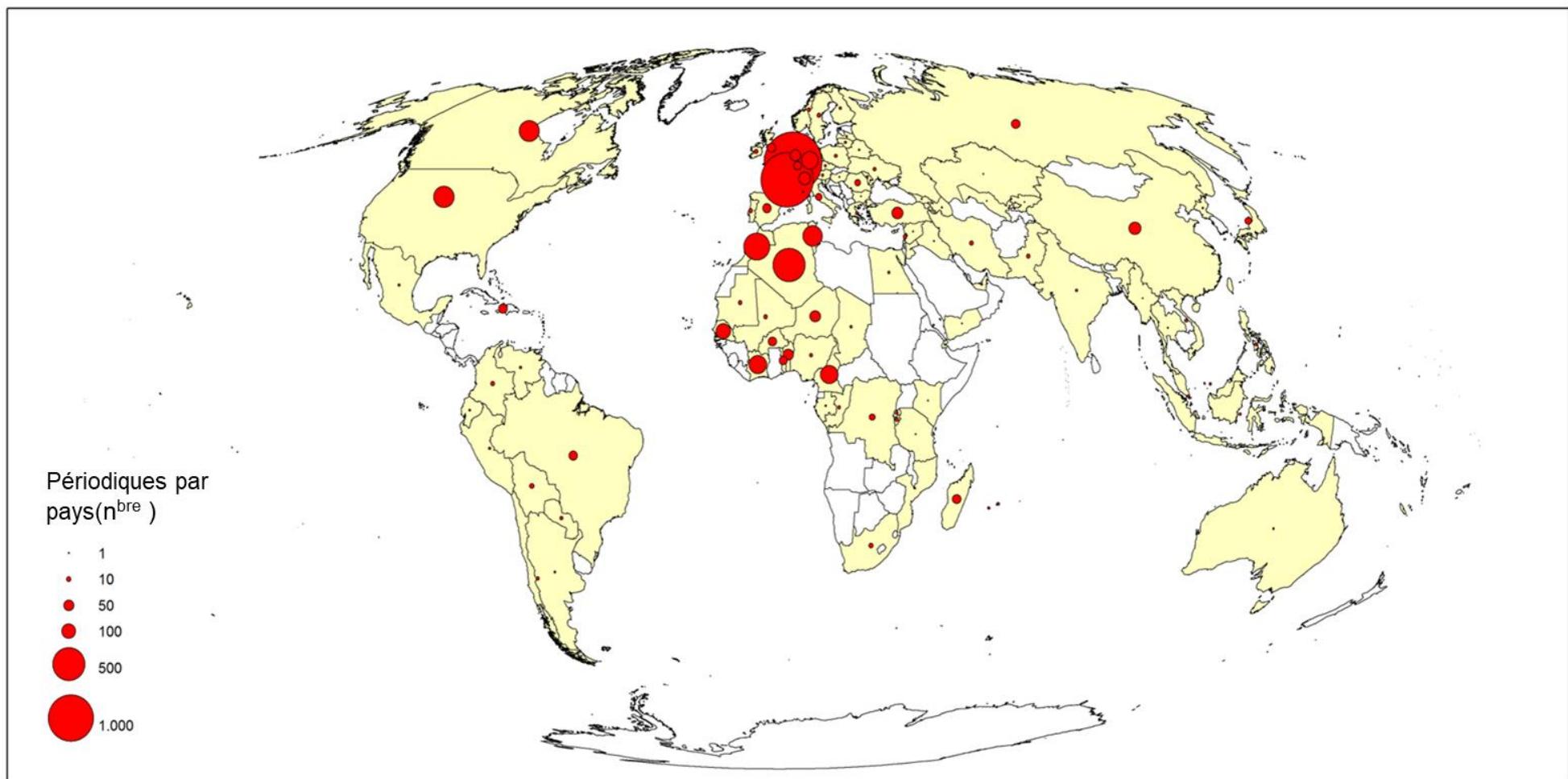
De nouveau, la Belgique (1 646 téléchargements) et la France (1 362 téléchargements) sont les plus grands utilisateurs de ces références parmi les pays francophones. Puis, les pays du Maghreb, l'Algérie (509 téléchargements), le Maroc (332 téléchargements) et la Tunisie (188 téléchargements) ont effectué le plus grand nombre de téléchargements. Viennent ensuite le Cameroun (156 téléchargements) et la Côte d'Ivoire (149 téléchargements). Les Etats Unis, quant à eux, ont effectué 198 téléchargements.

A une échelle plus réduite, la répartition mondiale des téléchargements des 68 articles (Carte 13) est similaire à la représentation des téléchargements de toutes les références (Carte 6).

Parmi les pays non francophones d'Europe, l'Allemagne seule a dépassé les 100 téléchargements. La Chine, la Turquie, le Brésil et la Russie ont aussi téléchargé ces références en moindre nombre, soit moins de 100 téléchargements chacun.

L'usage de ces références se révèle relativement important, le répertoire institutionnel constitue donc un moyen de dissémination efficace de cette littérature qui n'est pas indexée par les bases de données commerciales. Comme précédemment ce sont les pays francophones qui en sont les plus grands utilisateurs, ces références étant majoritairement de langue française.

Résultats



Carte 13 - Distribution au niveau mondial des téléchargements effectués à partir d'ORBi et correspondant aux références retrouvées par GS uniquement via ORBi

4.1.5. Distribution des téléchargements correspondant aux références retrouvées par tous les outils bibliographiques

Les 97 références retrouvées par tous les outils bibliographiques, tous en langue anglaise et qui représentent 21,4 % de l'échantillon d'articles et communications, ont enregistré 6 615 téléchargements, soit 14,9 % du nombre total de téléchargements (Tableau 23).

En moyenne 68 téléchargements ont été effectués par référence.

60 téléchargements ont été exclus de l'analyse car il a été impossible de les situer géographiquement de manière précise.

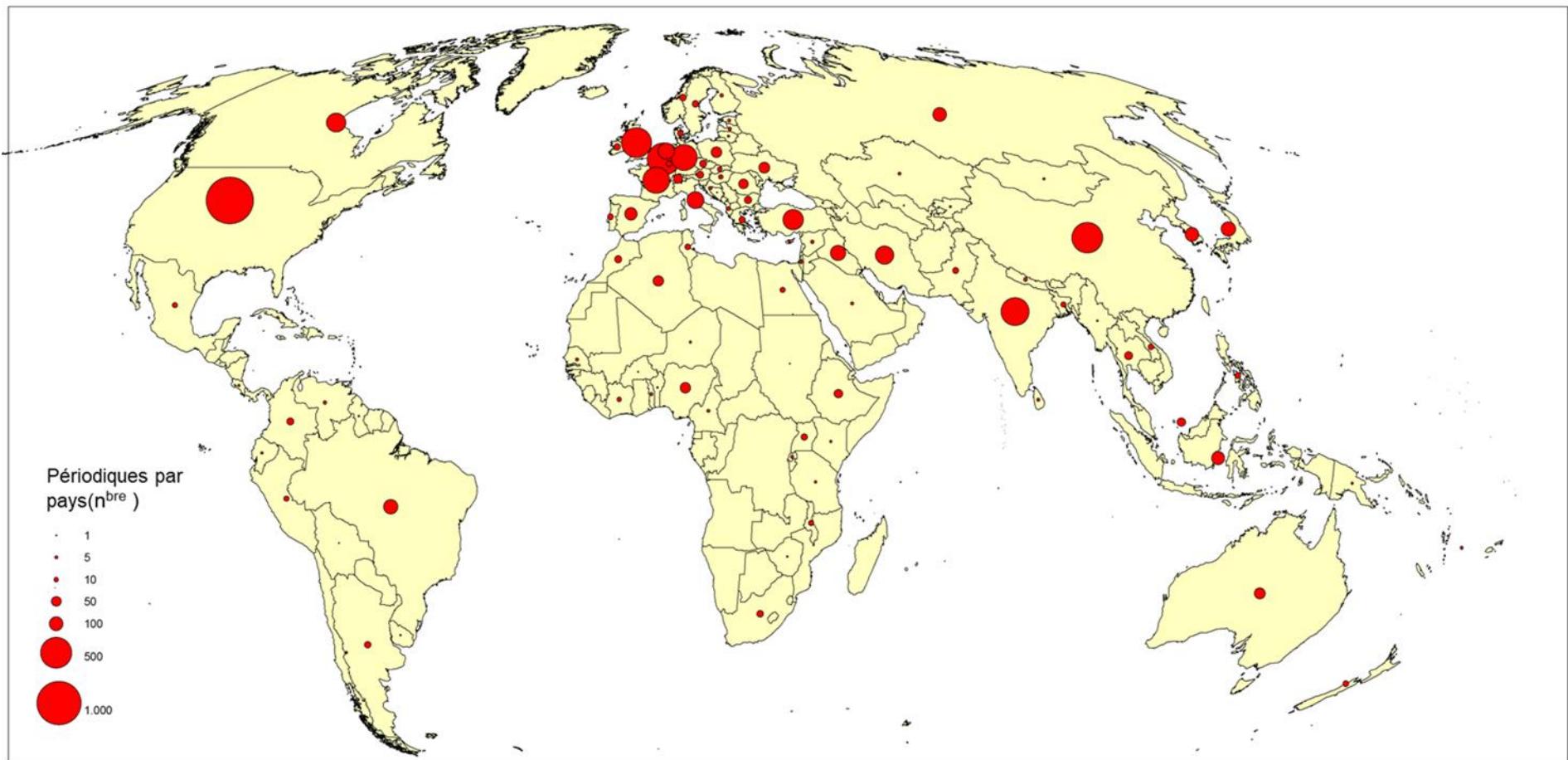
Sur la Carte 14, on observe un nombre plus élevé de téléchargements pour les pays développés de l'Europe de l'Ouest, plus particulièrement les pays francophones comme la France et la Belgique, ainsi que pour le Canada.

Les Etats-Unis se démarquent par un nombre important de téléchargements également.

Les pays asiatiques sont bien représentés par la Chine et l'Inde ainsi que les pays du Moyen-Orient comme la Turquie, l'Irak, l'Iran, etc.

Vu la présence de ces références dans les quatre outils bibliographiques il est intéressant de constater que ces références sont téléchargées *via* ORBi par des pays qui sont susceptibles d'avoir accès à la littérature scientifique « commerciale ».

Résultats



Carte 14 - Distribution au niveau mondial des téléchargements réalisé à partir d'ORBi, correspondant aux références retrouvées par tous les outils bibliographiques

4.2. Facteurs susceptibles d'influencer les téléchargements de la littérature en Géographie

L'inventaire des pays qui téléchargent les articles déposés sur Orbi ne peut négliger la diversité des pays quant à leur volume de population, le nombre de scientifiques, le nombre de connexions internet, les pratiques linguistiques, etc. C'est pourquoi dans cette partie, nous proposons d'explorer le poids de ces facteurs

4.2.1. Description des variables

Initialement les variables pressenties étaient les suivantes :

- Pays (= identifiant)
- Nombre total de téléchargements (T) où $T = F + A$
- Nombre de téléchargements de publications en français (F),
- Nombre de téléchargements de publications en anglais (A),
- PIB 2010 (PIB)
- Population 2010 (POP)

Un premier modèle linéaire fut alors ajusté pour les variables transformées en logarithme :

$$\text{Log}(T) \sim \log(\text{PIB}) + \log(\text{POP})$$

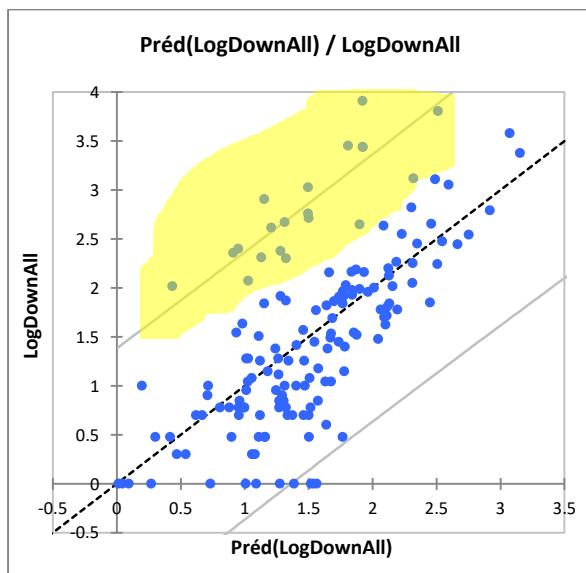


Figure 15 - Graphique des valeurs observées en fonction des valeurs prévues pour le modèle « test »

En analysant les résidus de ce modèle, deux sous-groupes de pays ont été observés (Figure 15). Un premier sous-groupe est identifié (en jaune), l'examen des données nous permet d'identifier directement les pays correspondants dans la base de données : Monaco, Burundi, Gabon, Haïti, Liban, Mali, Djibouti, Luxembourg, Madagascar, Rwanda, Mauritanie, Burkina Faso, Togo, Niger, Suisse, Sénégal, Côte d'Ivoire, Cameroun, Benin, Tunisie, Canada, Algérie, Maroc, France et Belgique. Ce sous-groupe de pays comporte essentiellement des pays francophones et nous avons considéré qu'il serait pertinent d'inclure des variables explicatives supplémentaires dans le modèle, les variables catégorielles qui indiquent l'appartenance à la francophonie ou à l'anglophonie, et qui expliqueraient la subdivision en deux sous-groupes observés dans les résidus du modèle initial.

Résultats

Il nous est aussi apparu intéressant d'inclure des variables supplémentaires tels que le nombre de chercheurs dans les pays, le nombre de techniciens, les dépenses du PIB dans la R&D. Afin d'élargir notre base de données et renforcer le modèle de régression⁵⁸, nous avons ajouté neuf autres variables aux six déjà prises en compte.

Les variables retenues pour l'ajustement du modèle linéaire sont les suivantes :

Nom	Description	Source	Type
PAYS	Nom du Pays	ORBi	Identifiant
TOT_TELECH	Nombre de téléchargements (T) où T = FR+EN	ORBi	Quantitatif
TELECH_FR	Nombre de téléchargements de publications en français	ORBi	Quantitatif
TELECH_EN	Nombre de téléchargements de publications en anglais	ORBi	Quantitatif
PIB	PIB 2010	World Bank	Quantitatif
POP	Population 2010	World Bank	Quantitatif
%PIB->R&D	Pourcentage du PIB consacré à la Recherche et au Développement	World Bank	Quantitatif
%PIB-->ENS	Pourcentage du PIB consacré à l'enseignement	World Bank	Quantitatif
CHERCH	Pourcentage de chercheurs dans la population du pays	World Bank	Quantitatif
TECHN	Pourcentage de techniciens dans la population du pays	World Bank	Quantitatif
INTERNET_U	Nombre d'utilisateurs d'Internet	internet worldstats.com	Quantitatif
ARTICLES	Nombre d'articles scientifiques total par an/pays	World Bank	Quantitatif
CONTINENT	Continent[modalités: Europe / Afrique / Asie / Amérique du Sud / Amérique du Nord / Océanie]	Classement Propre	Catégorielle
FR	Pays Francophones [modalités : Français / Sans affinité]	World Bank	Binaires
ANGL	Pays Anglophones [modalités : Anglais / Sans affinité]	World Bank	Binaires

Tableau 24 - Variables sélectionnées pour la construction du modèle

4.2.1.1. Etude des valeurs manquantes

Dans l'analyse de régression si une valeur associée à l'une des variables présentes dans le modèle est manquante, l'observation correspondante est automatiquement retirée du modèle. La base de données de départ concerne 164 pays pour le téléchargement de tous les articles, 146 pour le téléchargement des articles en anglais et 127 pour le téléchargement des articles en français (Tableau 26). En ce qui concerne les variables explicatives, la situation est totalement différente, beaucoup d'observations étant manquantes (Tableau 25) : PIB->R&D - 46

⁵⁸ Nous avons consulté Mme Timmermans, statisticienne et consultante dans le Service de Statistique Appliquée, Département de Mathématique de l'ULiège, nous a encouragé dans ce sens et nous a recommandé l'usage du logiciel JMP Pro.

Résultats

observations manquantes, PIB-->ENS – 61 observations manquantes, CHERCH - 61 observations manquantes, TECHN – 78 observations manquantes.

Les variables explicatives catégorielles Francophone, Anglophone et Continent ne contiennent pas de valeurs manquantes.

Variables	Nb. valeurs manquantes
LogTOT_TELECH	0
LogTELECH_FR	37
LogTELECH_EN	18
LogPIB	12
LogPOP	7
LogINTERNET_U	0
Log%PIB-->ENSEIGN	61
Log%duPIB-->R&D	46
LogTECHN/Pays	78
LogCHERCH/Pays	61
LogARTICLES/Pays	13

Tableau 25 - Valeurs manquantes dans les observations pour les différentes variables

En prenant en compte les trois variables dépendantes «LogTotalTéléchargements» «LogTéléchargementsFR » et «LogTéléchargementsEN » dans un modèle de régression, il en ressort que seulement 27 pays ne sont associés à aucune valeur manquante pour toutes les variables de la table de données : Argentine, Bolivie, Bulgarie, Chili, Colombie, Chypre, République Tchèque, Danemark, Ecuador, Ethiopie, Ghana, Hongrie, Islande, Irlande, Israël, Kenya, Lettonie, Lituanie, Mali, Mexique, Nouvelle Zélande, Norvège, Pakistan, Singapour, Slovénie, Tanzanie, Uganda.

En utilisant uniquement « LogTotalTéléchargements » comme variable réponse, et non les variables qui nous permettent de faire la distinction selon la langue des articles, le nombre de pays n'ayant aucune valeur manquante augmente à 33 pays, les mêmes que précédemment auxquels s'ajoutent 6 pays : Cap Vert, Estonie, Malawi, Malta, Mauritius, Sri Lanka.

Au vu de ce faible nombre de pays pour lesquels nous disposons de données pour toutes les variables incluses dans le modèle, nous pensons qu'il n'est pas judicieux de créer un modèle liant la variable réponse à toutes les variables explicatives.

Si nous souhaitons pouvoir en déduire des conclusions suffisamment générales quant aux facteurs influençant ou non le nombre de téléchargements, une présélection des variables doit être effectuée sur base de la configuration des valeurs manquantes, afin d'éliminer les variables qui amèneraient à ajuster le modèle sur un nombre de pays trop restreint.

Ainsi sur base du diagramme des cellules des valeurs manquantes et du Tree Map généré, nous avons décidé de ne pas inclure dans notre modèle les quatre variables explicatives avec le plus grand nombre de données manquantes : PIB->R&D, PIB-->ENS, CHERCH et TECHN.

Résultats

Les variables explicatives que nous gardons dans notre modèle sont : PIB, Population, Utilisateurs Internet, Pays Francophones, Pays Anglophones et Continent.

4.2.1.2. Les variables dépendantes

Les variables dépendantes sont représentées par le nombre de téléchargements total (TOT_TELECH), le nombre de téléchargements de publications en français (TOT_TELECH_FR) et le nombre de téléchargements de publications en anglais (TOT_TELECH_EN) par pays. Le Tableau 26 synthétise les statistiques descriptives des trois variables dépendantes, avant la transformation logarithmique.

	Nombre obser.	Donnée(s) manquante(s)	Moyenne	1 ^{er} Quartile	Médiane	3 ^{ème} Quartile	Minimum	Maximum	Écart type
TOT_TELECH	164	0	267	6	24.5	110.5	1	8095	925.29
TELECH_EN	146	18	111.62	3	16	61.5	1	2622	338.92
TELECH_FR	127	37	215.57	3	10	73	1	6338	816.67

Tableau 26 - Statistiques descriptives des trois variables dépendantes

Le nombre de téléchargement total (TOT_TELECH) des références par pays, varient de 1 à 8 095, avec une moyenne de 267 téléchargements et une médiane de seulement 24,5. La valeur la plus élevée, 8 095 téléchargements, revient à la Belgique. L'analyse de distribution des téléchargements confirme que la majorité des pays (120 pays) enregistrent moins de 100 téléchargements par pays.

La moyenne des téléchargements est de 267 avec un écart type de 925. En termes de dispersion, l'écart type est très élevé par rapport à la moyenne en raison de plusieurs valeurs extrêmes. Le graphique nous indique une distribution dissymétrique à droite (positive) avec un coefficient d'asymétrie 6,2 provoquée par les observations extrêmes (Figure 16). Ce sont les pays développés et francophones du Maghreb qui ont téléchargé le plus les références et qui se situent à l'extrême des valeurs positives.

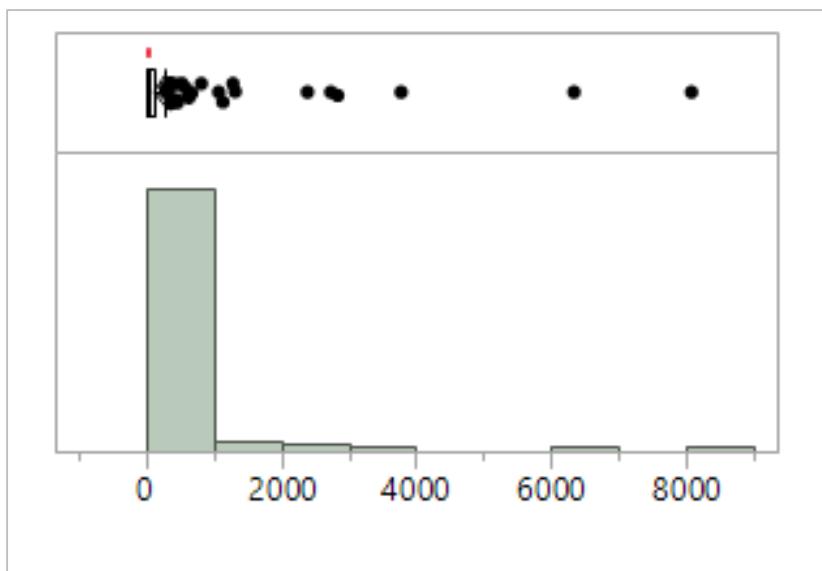


Figure 16 - Distribution du total des téléchargements

Le téléchargement des références en Anglais (TELECH_EN) par pays, varient de 1 à 2 622, avec une moyenne de 112 téléchargements, une médiane de seulement 16 téléchargements et un écart type de 339 (Figure 17). La valeur la plus élevée, 2 622 téléchargements, correspond aux Etats Unis. L'analyse de distribution des téléchargements confirme que la majorité des pays (121 pays) enregistrent moins de 100 téléchargements.

L'écart type très élevé par rapport à la moyenne s'explique par les valeurs extrêmes enregistrées pour les pays développés qui téléchargent plus les références en anglais et qui se situent à l'extrême des valeurs positives.

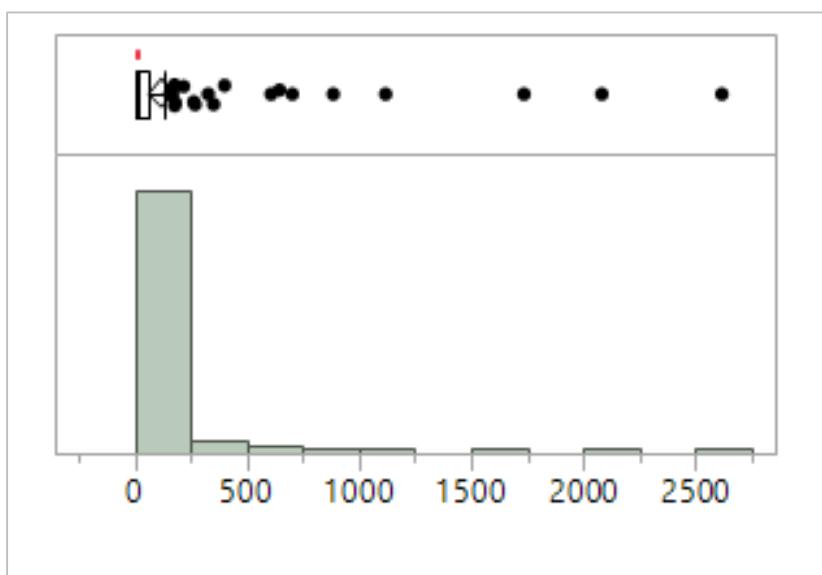


Figure 17 - Distribution des téléchargements en anglais

Le téléchargement des références en Français (TELECH_FR) par pays, varient de 1 à 6 338, avec une moyenne de 216 téléchargements, une médiane de seulement 10 téléchargements et un écart type de 817 (Figure 18). La valeur la plus élevée, 6 338 téléchargements revient de

Résultats

nouveau à la Belgique. Comme pour les deux autres variables, l'écart type est très élevé par rapport à la moyenne en raison de ces valeurs extrêmes.

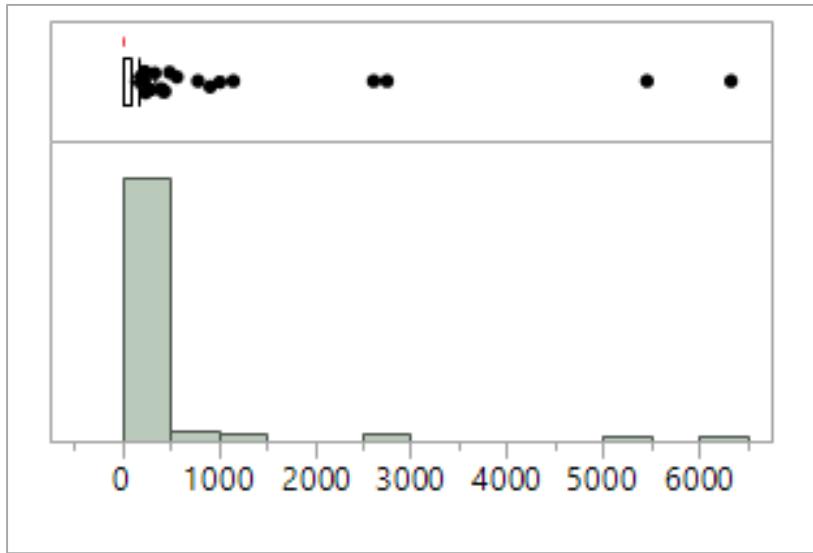
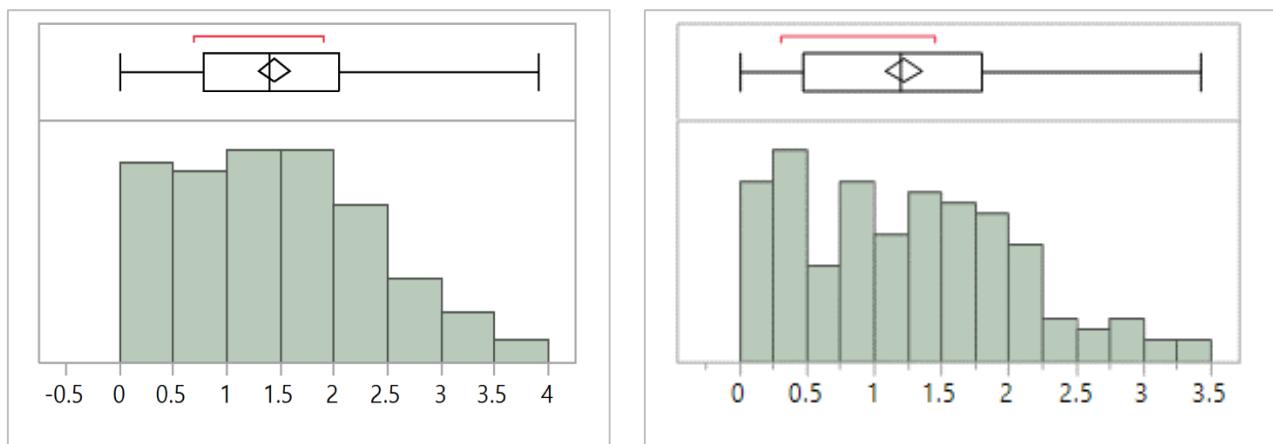


Figure 18 - Distribution des téléchargements en français

Une étape préalable indispensable à la modélisation, est l'analyse statistique descriptive, à l'occasion de laquelle la détection des valeurs aberrantes ou extrêmes est possible. Après la transformation logarithmique, il apparaît (Tableau 27) que les valeurs moyenne et médiane sont assez proches pour toutes les variables dépendantes ce qui est positif pour la suite de notre analyse.

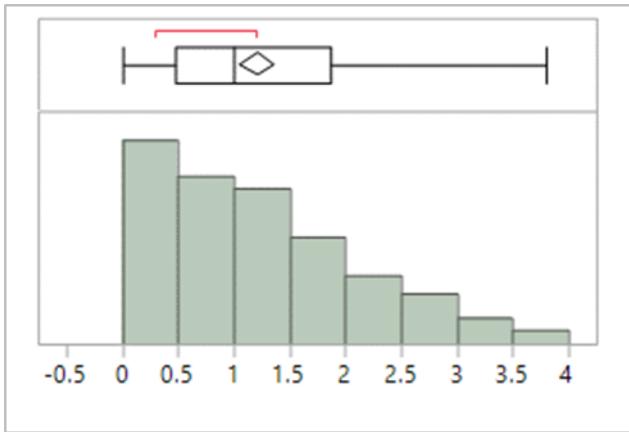
	Nombre d'observations	Donnée(s) manquante(s)	Moyenne	Médiane	Minimum	Maximum	Écart type
TOT_TELECH	164	0	1.45	1.39	0	3.91	0.91
TOT_TELECH_EN	146	18	1.22	1.20	0	3.42	0.83
TOT_TELECH_FR	127	37	1.21	1.00	0	3.80	0.91

Tableau 27 - Statistiques descriptives des trois variables dépendantes après la transformation logarithmique



a) LogTotalTéléchargements

b) LogTéléchargementsENG



c) LogTéléchargementsFR

Figure 19 - Distribution des variables dépendantes après la transformation logarithmique

4.2.1.3. Les variables explicatives

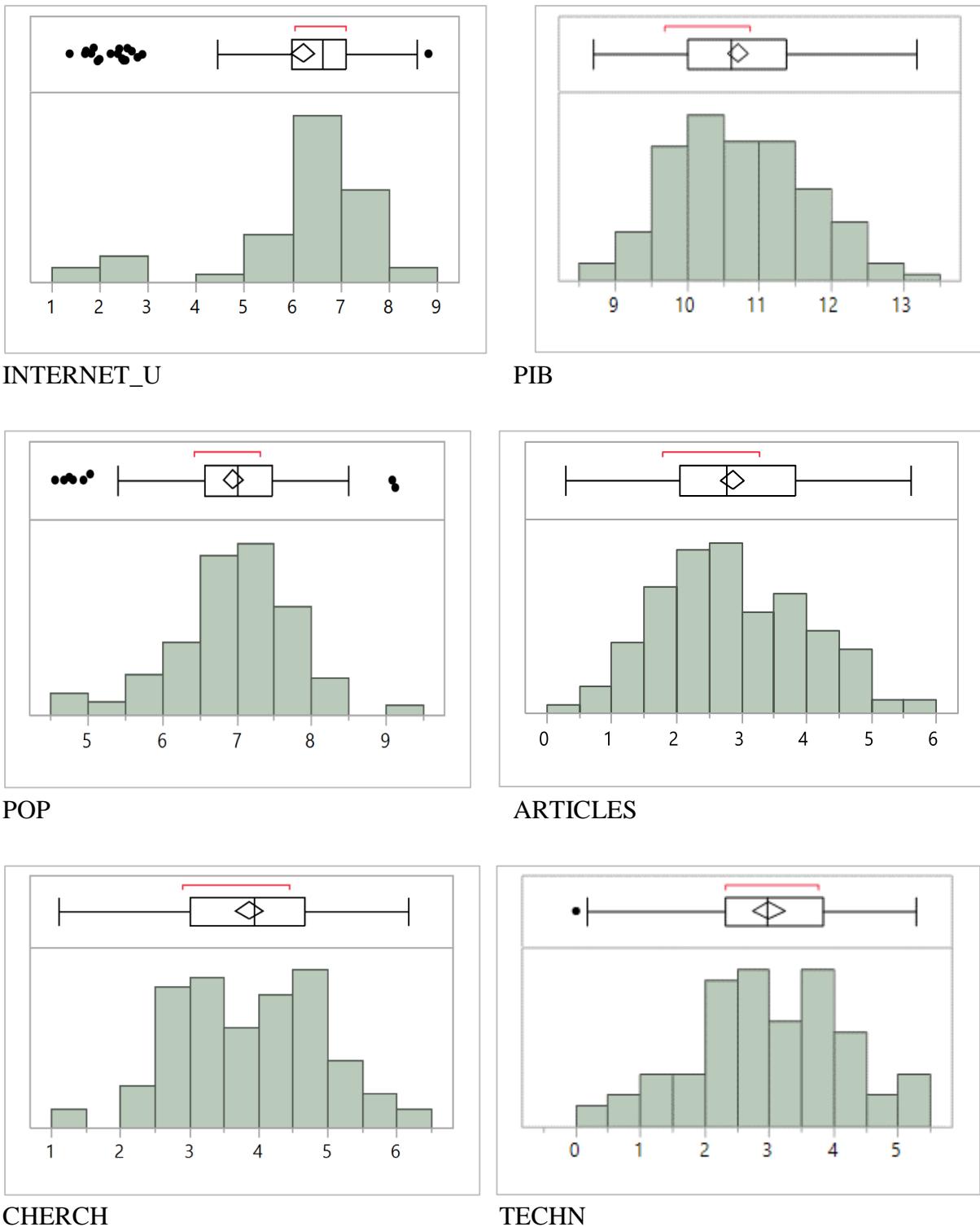
Les variables explicatives prises en compte dans notre modèle étant de plusieurs types nous allons les présenter séparément.

Variables explicatives	Nombre	Donnée(s)	Moyenne	Médiane	Minimum	Maximum	Écart-type
	d'observations	manquante(s)					
INTERNET_U	163	1	6.22	6.62	1.38	8.83	1.56
PIB	152	12	10.71	10.62	8.71	13.18	0.91
POP	157	7	6.96	7	4.57	9.13	0.81
ARTICLES	151	13	2.87	2.77	0.3	5.62	1.13
CHERCH	103	61	3.87	3.93	1.12	6.17	1.04
TECHN	86	78	3	2.98	0.01	5.28	1.18
%PIB-->ENS	103	61	9.37	9.25	7.38	11.91	1.01
%PIB->R&D	118	46	8.61	8.52	6.07	11.61	1.21

Tableau 28 - Statistiques descriptives des variables explicatives quantitatives après la transformation logarithmique

Les variables explicatives ne doivent pas suivre une loi normale, mais cela peut être utile d'avoir un aperçu de la distribution. Les variables explicatives quantitatives ont été soumises à une transformation logarithmique en LOG10 et leurs distributions ont été analysées à l'aide des histogrammes de fréquence. Le Tableau 28 synthétise les statistiques descriptives des variables explicatives prises en compte dans le modèle de régression après la transformation logarithmique. Les valeurs moyenne et médiane de toutes les variables explicatives sont proches.

Résultats



Résultats

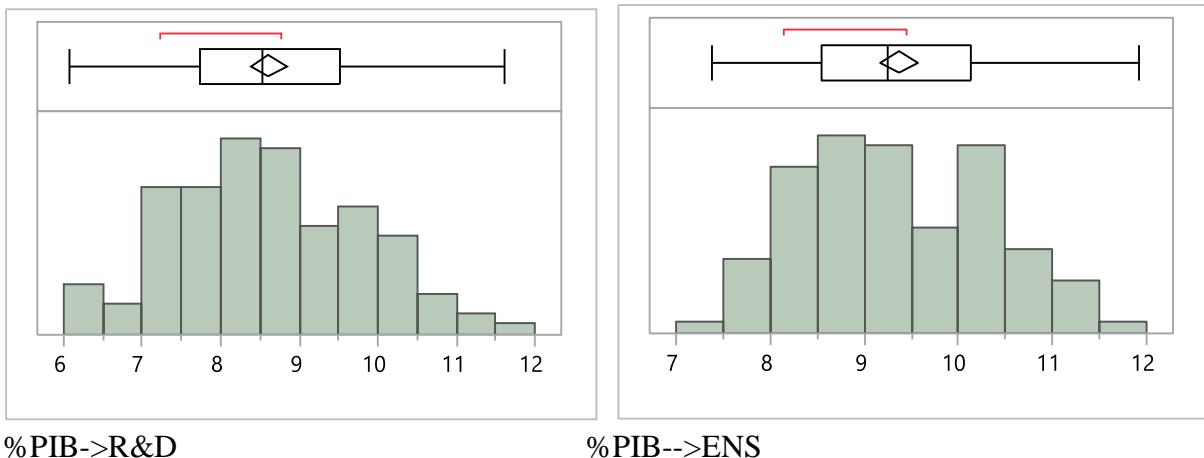
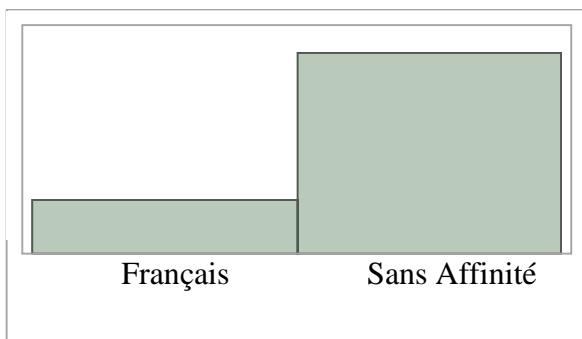


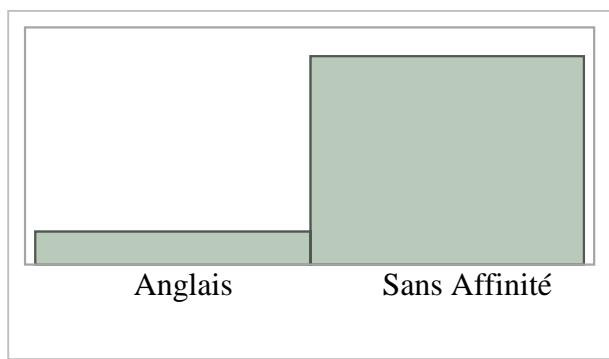
Figure 20 - Distribution des variables explicatives quantitatives après la transformation

Dans la Figure 20, sont représentés les graphiques de distribution des variables explicatives quantitatives après leur transformation logarithmique en base 10. Il apparaît que les distributions de ces variables s'approchent d'une distribution normale dans la plupart des cas. Des valeurs extrêmes peuvent être observées pour les variables Utilisateurs Internet et Population, mais la suppression de ces valeurs extrêmes ne se justifie pas dans notre cas.



a) Francophone

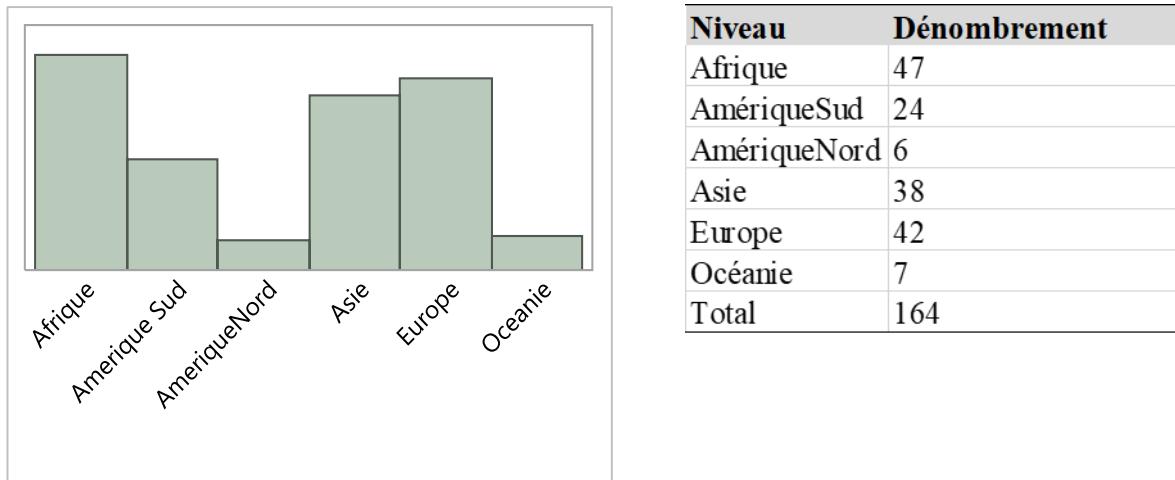
Niveau	Dénombrement
Français	34
Sans Affinité	130
Total	164



b) Anglophone

Niveau	Dénombrement
Anglais	21
Sans Affinité	143
Total	164

Résultats



c) Continent

Figure 21 - Distribution des trois variables explicatives catégorielles

La Figure 21 représente les variables explicatives catégorielles que nous avons prises en compte dans notre modèle de régression et le dénombrement par modalité de ceux-ci.

4.2.1.4. Analyse de corrélation

L'analyse de corrélation est une autre étape recommandée avant la construction d'un modèle de régression. Cette analyse nous permet d'identifier si une relation existe entre les variables dépendantes et les variables explicatives et, par la suite, entre les variables explicatives afin de vérifier leur indépendance les unes par rapport aux autres. De manière générale, ce sont les corrélations excessives en valeurs absolues supérieures à 80 % (Kutner et al., 2005a) qui peuvent impliquer des problèmes de colinéarité, ce qui n'est pas souhaité.

4.3. Les facteurs qui influencent les téléchargements de l'ensemble de l'échantillon de publications déposées dans ORBi par les géographes : développement du modèle retenu et résultats

Notre modèle de régression final retenu a été réalisé au moyen de la sélection automatique des variables explicatives, en utilisant l'ajustement forward stepwise (critères BIC (Bayesian Information Criterion), AICc (Akaike Information Criterion) avec procédure ascendante). Le critère BIC est le plus conservateur, tandis que AICc est contraire au premier.

Résultats

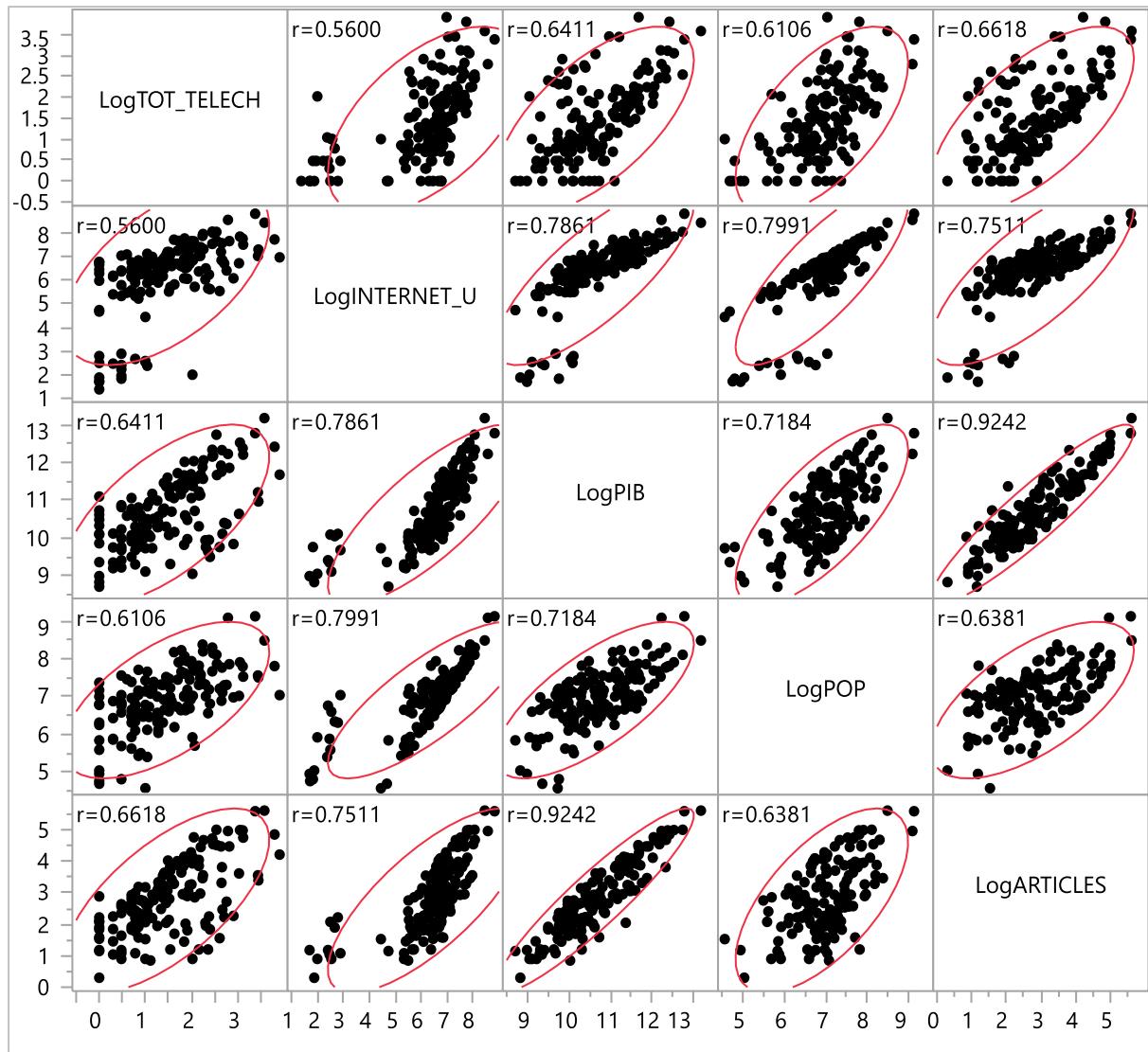


Figure 22 - Matrice de graphiques de nuages de points

Les résultats de l'analyse de corrélation (Figure 22 et Tableau 29) indiquent une relation existante entre la majorité des variables. A l'examen des coefficients résultants, on remarque une intensité de la relation se situant entre 56 et 66% entre la variable dépendante et les variables explicatives. En ce qui concerne la relation entre les variables explicatives elles-mêmes, les coefficients de corrélation obtenus nous indiquent des corrélations plus importantes, se situant entre 64% et 92%, la corrélation la plus importante étant entre les variables LogARTICLES et LogPIB ($r = 0,92$).

	LogTOT_TELECH	LogINTERNET_U	LogPIB	LogPOP	LogARTICLES
LogTOT_TELECH	--	0.56	0.64	0.61	0.66
LogINTERNET_U	0.56	--	0.79	0.80	0.75
LogPIB	0.64	0.79	--	0.72	0.92
LogPOP	0.61	0.80	0.72	--	0.64
LogARTICLES	0.66	0.75	0.92	0.64	--

Tableau 29 - Coefficients de corrélation entre les différentes variables explicatives quantitatives du modèle

Résultats

Cette importante corrélation se traduit par un VIF, Variance Inflation Factor, relativement important de 8,83 mais non strictement problématique pour ces deux variables ainsi que pour la suite du modèle (p. 406 Kutner *et al.*, 2005b). En effet, c'est lorsque ces valeurs dépassent la valeur 10 que cela devient problématique. On peut en conclure que nous ne trouvons pas face à des problèmes de multicolinéarité.

Compte tenu du fait que la valeur du VIF ne dépasse pas le seuil critique pour aucune des corrélations, nous avons décidé de prendre en considération l'ensemble des variables dans un premier temps.

Le Tableau 30 synthétise les indicateurs globaux du modèle réalisé pour comprendre les facteurs qui influencent les téléchargements des références des géographes déposées dans ORBi. Le coefficient de corrélation R^2 est de 80,1%. Cela signifie que la force de la relation entre le nombre de téléchargements totaux et les variables explicatives est élevée : 80,1 % de la variabilité initiale de téléchargements est expliquée par le modèle.

R carré	0.801
R carré ajusté	0.796
Racine de l'erreur quadratique moyenne	0.410
Moyenne de la réponse	1.518
p-value	<0,0001
F	147,34
Observations (ou sommes pondérées)	151

Tableau 30 - Résumé de l'ajustement : indicateurs globaux de la qualité du modèle retenu

L'analyse de la variance et le test de Fisher ($F=147,34$) confirment la pertinence des variables explicatives retenues dans le modèle. Compte tenu de la diversité des pays ayant téléchargé les articles d'ORBi et de leurs caractéristiques fort différentes ces valeurs sont satisfaisantes.

Les coefficients de régression permettent d'interpréter l'importance de l'impact de chaque variable explicative dans la prévision de la variable dépendante. Leurs valeurs représentent l'influence attribuée à chaque facteur.

Le test t de Student est un indicateur de fiabilité statistique des coefficients de régression et la p-value p représente la probabilité que la variable explicative ne soit pas statistiquement significative. En deçà du seuil de 0,05 la variable est considérée comme significative, entre 0,05 et 0,1 la variable n'est pas considérée comme significative. Au-delà de 0,1 les variables sont considérées comme non significatives.

Terme	BIC	AICc	< 0,05
LogINTERNET_U	x	x	
LogPIB			
LogPOP	x	x	x
LogARTICLES	x	x	x
FR[Français]	x	x	x
ANGL[Anglais]		x	
Continent{Asia&Africa-Europe&North America}	x	x	

Tableau 31 - Variables retenues(en gras) lors de la sélection automatique avec les critères BIC, AICc avec procédure ascendante.

Résultats

Le modèle que nous avons retenu comporte quatre variables explicatives significatives (Tableau 31), il s'agit de :

- Pays Francophones [modalités : Français / Sans affinité] - FR (binaire catégorielle)
- Nombre d'articles par pays - LogARTICLES (quantitatif)
- Population par pays (données 2010) - LogPOP (quantitatif)
- Continent {Asia&Africa - Europe&North America} (catégorielle)

A l'examen du Tableau 32, on peut remarquer qu'une variable Continent{Asia&Africa-Europe&North America} a été générée lors du processus de sélection automatique des variables sur base de la variable explicative catégorielle Continent.

Terme	Estimation	Erreur standard	t ratio	Prob. > t
Constante	-2.448996	0.376689	-6.5	<.0001
LogPOP	0.429524	0.065758	6.53	<.0001
LogArticles	0.4737564	0.045494	10.41	<.0001
Francophone[Français]	0.6694495	0.043045	15.55	<.0001
Continent{Asia&Africa-Europe&North America}	-0.114842	0.048448	-2.37	0.0191

Tableau 32 - Variables retenues et estimations des coefficients

Le Tableau 32 détaille les paramètres relatifs à chacune des quatre variables du modèle tandis que la Figure 24 détaille l'expression de la prévision du modèle retenu.

Expression de la prévision							
-	2.4489961900002						
+	0.42952397141123 * LogPOP						
+	0.47375637077357 * LogArticles						
+ Match(Francophone)	<table border="0" style="margin-left: 20px;"> <tr> <td>"Français"</td> <td>⇒ 0.66944947670399</td> </tr> <tr> <td>"Sans Afinité"</td> <td>⇒ -0.669449476704</td> </tr> <tr> <td>autre</td> <td>⇒ .</td> </tr> </table>	"Français"	⇒ 0.66944947670399	"Sans Afinité"	⇒ -0.669449476704	autre	⇒ .
"Français"	⇒ 0.66944947670399						
"Sans Afinité"	⇒ -0.669449476704						
autre	⇒ .						
+	-0.1148417854797 * Continent{Asia&Africa-Europe&North America}						

Figure 23 - Expression de la prévision du modèle retenu

Les variables LogARTICLES et LogPOP influencent le modèle, soit le nombre de téléchargements, et les valeurs des coefficients obtenus sont proches. Si LogARTICLES augmente d'une unité, toutes choses égales par ailleurs, alors le LogTOT_TELECH augmentera de 0,47 unité.

Une augmentation d'une unité de LogPOP, toutes choses égales par ailleurs, augmentera le LogTOT_TELECH de 0,42 unité.

La variable Francophone est définie comme suit : Français - 1 ; Sans Affinité : (-1).

Ainsi si un pays est francophone, le LogTOT_TELECH augmentera de 0,67 unité. Pour les pays ne faisant pas partie de la catégorie francophone le LogTOT_TELECH diminuera de -0,67 unité.

Afin d'interpréter l'impact de la variable composée sur le modèle retenu nous l'avons décomposée selon les modalités et calculée en fonction du coefficient d'estimation associé avec

Résultats

les valeurs du codage de la variable générée lors de la régression et indiqués dans le Tableau 33.

La variable **Continent {Asia&Africa - Europe&North America}** est définie comme suit :

Continent	Codage Variable
Asia	1
Africa	1
Europe	-1
North America	-1
Océanie, South America	0

Tableau 33 - Codage de la variable composée lors de la sélection automatique des variables

Pour les pays se situant en Asie et Afrique, le LogTOT_TELECH est diminué de 0,11 unité (-0,11 x 1). Pour les pays se situant en Europe et Amérique du Nord le LogTOT_TELECH enregistrera une augmentation de 0,11 unité (-0,11 x -1). Il n'y a aucune influence sur le LogTOT_TELECH pour les pays se situant en Océanie et Amérique du Sud.

4.3.1.1. Analyse des résidus du modèle retenu

L'analyse des résidus est une étape de validation et elle permet d'évaluer la qualité de l'ajustement du modèle. Dans le modèle que nous avons développé, un résidu se définit comme la différence entre le nombre de téléchargements enregistré dans ORBi et le nombre de téléchargements calculé par le modèle.

Les hypothèses sous-jacentes à l'ajustement de ce modèle doivent être validées, soit :

- Vérification de la normalité des résidus
- Linéarité de la fonction de régression
- Homoscédasticité
- Multi colinéarité
- Diagnostiques d'influence (Distance de Cook)
- Indépendance des erreurs

Les points suivants ont pour but d'analyser ces hypothèses et ainsi de valider le modèle de régression que nous avons retenu.

4.3.1.2. Vérification de la normalité des résidus

Dans un modèle de régression valide, les résidus doivent suivre une distribution normale (Dagnelie, 2011). Comme cela peut être constaté sur la Figure 24, la distribution des résidus de notre modèle suit bien une loi normale. L'histogramme se superpose bien à la courbe normale théorique. La droite de Henry représentée aussi dans la Figure 24 confirme que les résidus suivent bien la droite à quelques exceptions près. Ces représentations valident ainsi la normalité

de la distribution des résidus, ce qui signifie que la première hypothèse pour la validation du modèle retenu est remplie.

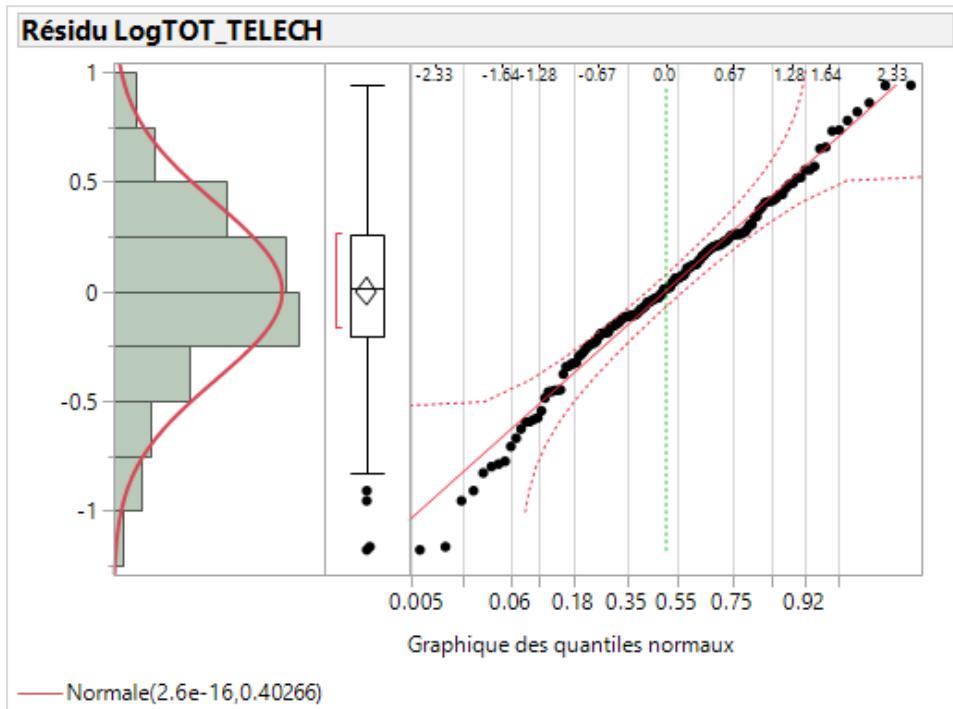
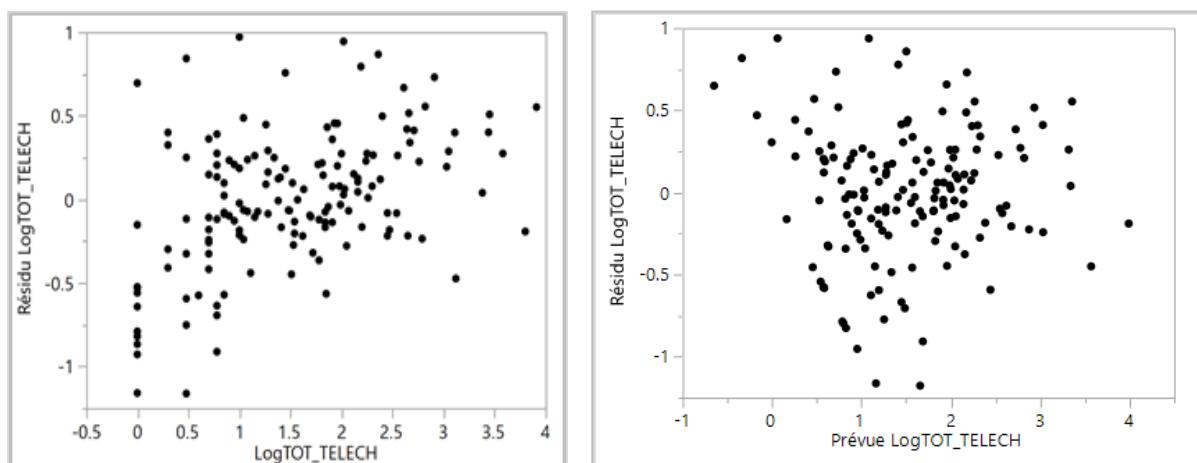


Figure 24 - Distribution des résidus et droite de Henry du modèle retenu

4.3.1.3. Linéarité de la fonction de régression et homoscédasticité

Les résidus doivent être caractérisés par une variance constante, indépendante de la variable dépendante, soit, dans notre cas, du nombre de téléchargements. Autrement, nous sommes en présence d'hétéroscédasticité, ou en présence d'une variance non homogène (Harris & Jarvis, 2014). Si la variance de l'erreur des variables est constante nous sommes en présence d'homoscédasticité, opposé à l'hétéroscédasticité. C'est donc l'hypothèse d'homoscédasticité que nous devons tester afin de vérifier que notre modèle est valide.



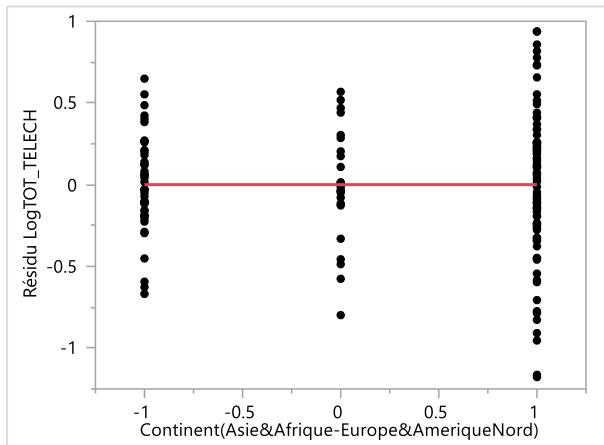
a) Résidus vs variable dépendante

b) Résidus vs prédition

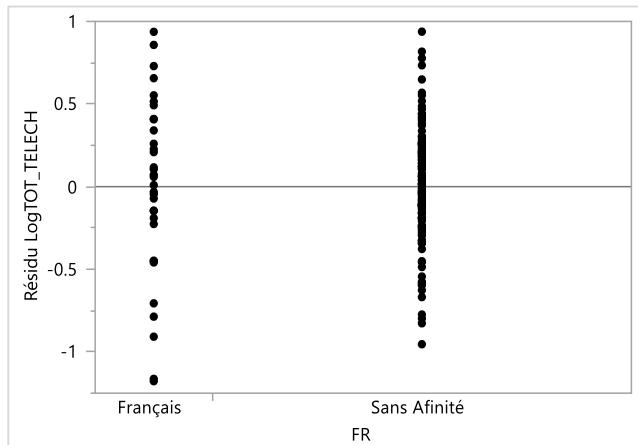
Résultats

Graphique 1 - Evolution des résidus du modèle retenu

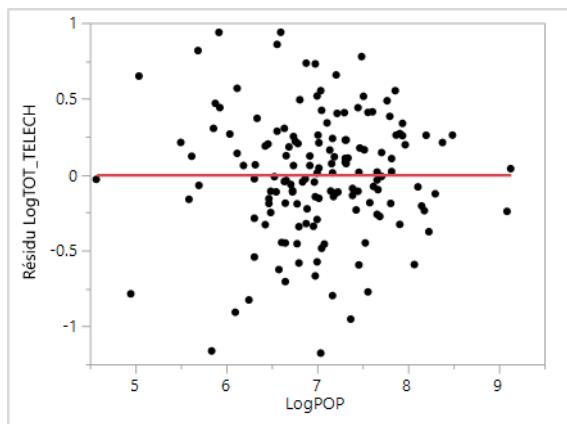
A l'analyse du Graphique 1 a), une très légère tendance dans la distribution de la variable dépendante, LogTOT_TELECH, contre les résidus peut être remarquée. Le nuage de points suit une droite légèrement croissante vers la droite, ce qui signifie que les résidus augmentent légèrement avec le nombre total de téléchargements. Ce fait pourrait nous conduire à conclure à une certaine invalidité du modèle considéré. Il n'y a pas de tendance définie dans le graphique pour la prédiction contre les résidus (Graphique 1 b)) ce qui vérifie notre hypothèse d'homoscédasticité.



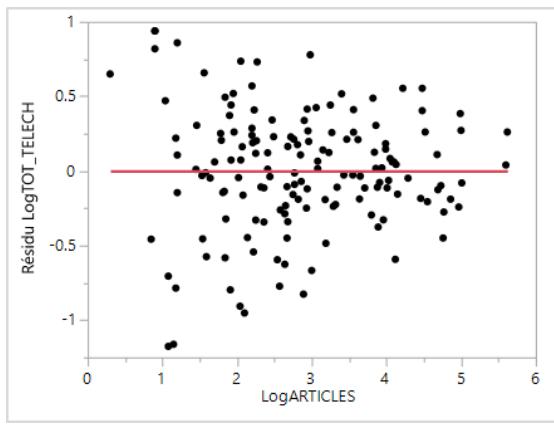
a) Résidus vs variable Continent



b) Résidus vs variable Francophone



c) Résidus vs variable Population



d) Résidus vs variable Articles

Graphique 2 - Evolution des résidus du modèle retenu en fonction de variables explicatives

Afin de vérifier l'hypothèse de linéarité de la fonction de régression, nous avons représenté les résidus du modèle retenu en fonction de chaque variable explicative du modèle. Les Graphiques 2 a), b), c) et d) présentent la distribution des résidus en fonction de la variable Continent(Asie&Afrique-Europe&AmeriqueNord), la variable Francophone, la Population et les Articles. Les nuages de points des résidus sont distribués autour d'une droite de pente nulle pour toutes les variables. Nous sommes donc, dans cette situation-ci, en présence d'homoscédasticité.

4.3.1.4. Test de multicolinéarité

Afin de détecter la multicolinéarité des variables explicatives, nous avons effectué le test « Facteur d'inflation de la variance » ou VIF pour Test Variance Inflation Factor dans JMP.

La valeur du VIF doit être supérieure à 0,1 mais ne doit pas dépasser la valeur 10, afin de pouvoir conclure à l'absence des problèmes de multicolinéarité. Des valeurs élevées de VIF indiquent la présence de multicolinéarité.

Terme	Estimation	Erreur standard	t ratio	Prob. > t	VIF
Constante	-2.449	0.376689	-6.5	<.0001	.
LogPOP	0.429524	0.065758	6.53	<.0001	2.118965
FR[Français]	0.66945	0.043045	15.55	<.0001	1.139518
LogArticles	0.473756	0.045494	10.41	<.0001	2.361103
Continent{Asie&Afrique-Europe&AmeriqueNord}	-0.11484	0.048448	-2.37	0.0191	1.636252

Tableau 34 - Estimations des coefficients et valeurs du test VIF

A la lecture du Tableau 34, on constate que la valeur du VIF de chaque coefficient est supérieure à 0,1 mais largement inférieure à 10. On peut en conclure que nous ne trouvons pas face à des problèmes de multicolinéarité.

4.3.1.5. Diagnostiques d'influence - distance de Cook

Afin de vérifier si certaines observations exercent une influence importante sur l'ensemble du modèle retenu, nous avons procédé à l'analyse de la distance de Cook également. Il existe plusieurs théories pour l'interprétation de ce diagnostic.

Pour juger si la distance C_i est élevée, Cook (1977) propose le seuil $f_{p,n-p}(0,1)$ comme souhaitable et le seuil $f_{p,n-p}(0,5)$ comme préoccupant.

Dagnelie (2011) indique la valeur critique 1 (Distance de Cook >1 alors l'observation i est influente) tandis que la troisième théorie indique le résultat de $D_i > 4/n$ où n i est le nombre d'observations.

L'examen de nos données, Figure 25, n'indique aucune valeur influente au sens de la distance de Cook.

Résultats

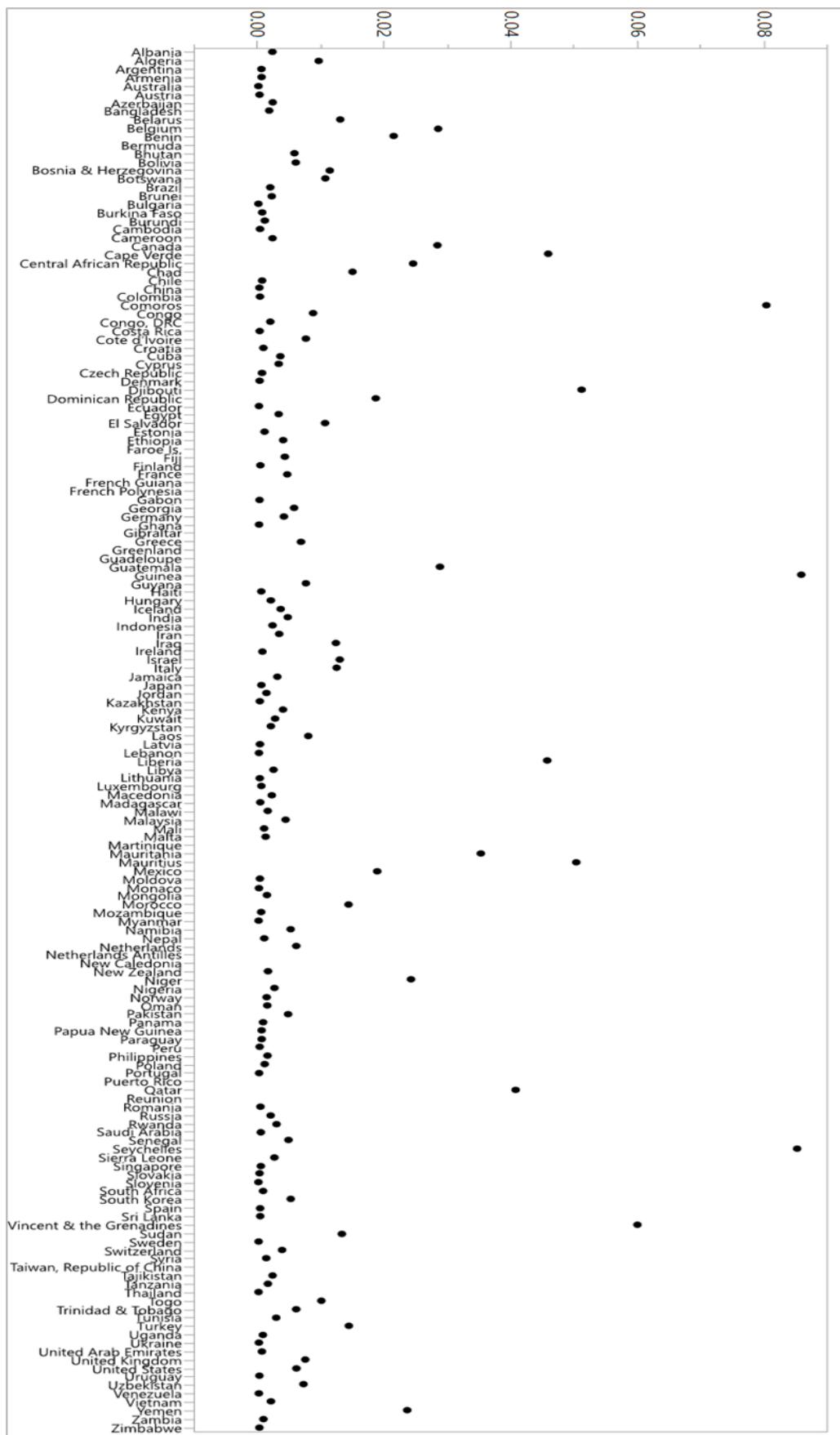
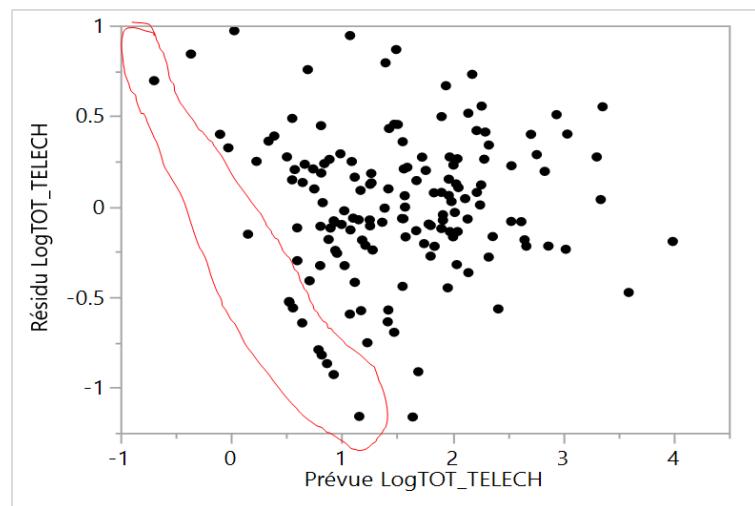


Figure 25 - Distances de Cook pour le modèle de régression – nombre de téléchargements total.

4.3.1.6. Indépendance des erreurs

Les résidus des modèles de régression étant en principe liés au hasard, ils doivent être indépendants les uns des autres, et distribués de manière totalement aléatoire (Harris & Jarvis, 2014). Il ne doit y avoir aucune structure définie dans la distribution des résidus. Les interdépendances pourraient être liées à une autocorrélation spatiale des résidus dans notre cas. Dans le cas d'une autocorrélation spatiale des résidus, le nombre des téléchargements serait réalisé par des pays voisins.

Comme cela peut être remarqué dans les séries de Graphiques 1 et 2 représentées dans les sections précédentes, il n'y a pas de structure marquée qui se distingue dans la distribution des résidus de notre modèle. On peut toutefois relever l'alignement des résidus des pays ayant effectué un seul téléchargement, encerclés dans le Graphique 3, et qui pourrait être considéré comme une faiblesse du modèle. Nous avions envisagé de supprimer ces observations, mais nous ne l'avons finalement pas fait car le faible nombre de téléchargements effectués par un pays est une information et garde donc sa place dans le modèle.

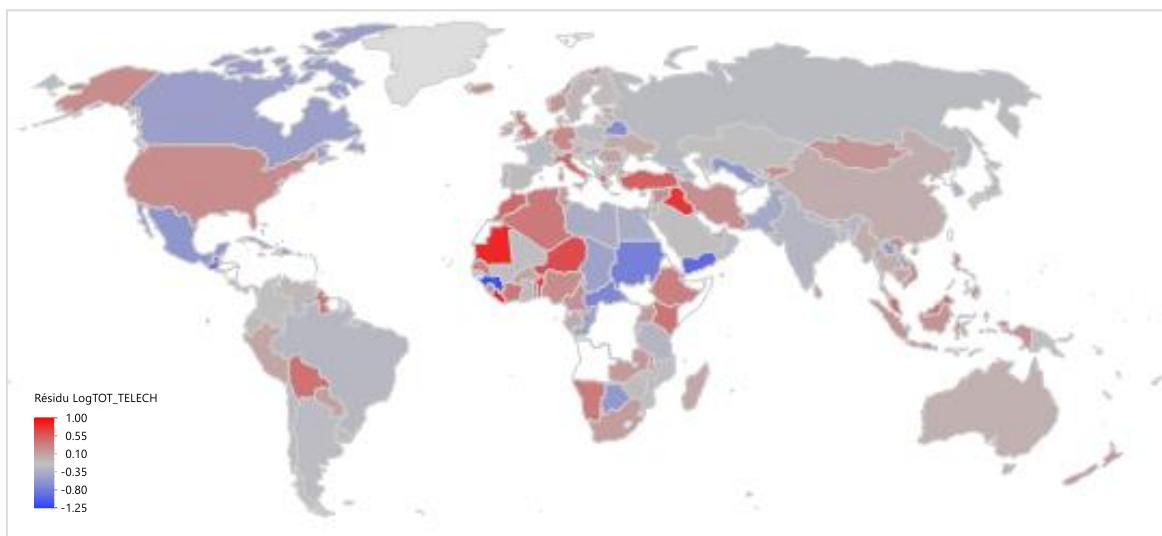


Graphique 3 - Résidus du modèle retenu contre la prédiction

Afin de vérifier l'hypothèse d'indépendance des résidus, nous les avons représentés sur une carte, dans JMP. La Carte 15 représente la valeur des résidus du modèle final associée à chacun des 151 pays ayant téléchargé les références d'ORBi. Malheureusement, les intitulés de quelques pays n'ont pas été reconnus par le logiciel statistique, malgré les nombreux essais de remédiation au problème.

Ajoutons encore qu'un résidu négatif (bleu) signifie que le modèle surestime le nombre de téléchargements tandis qu'un résidu positif (rouge) signifie que le modèle sous-estime le nombre de téléchargements. Les résidus ne semblent pas être dépendants les uns des autres et l'hypothèse d'Indépendance peut être retenue. A l'inspection visuelle de la carte, aucune dépendance spatiale n'est visible (les résidus sont répartis aléatoirement sans concentration particulière), en conséquence l'hypothèse est validée.

Résultats



Carte 15 - Répartition des résidus au niveau mondial

4.3.2. Les facteurs qui influencent les téléchargements des publications en langue anglaise de l'échantillon : développement du modèle retenu et résultats

Afin de déterminer les facteurs qui influencent les téléchargements de la littérature en Géographie en langue anglaise déposée dans ORBi nous avons parcouru exactement les mêmes étapes que pour le développement du modèle avec la variable dépendante LogTOT_TELECH.

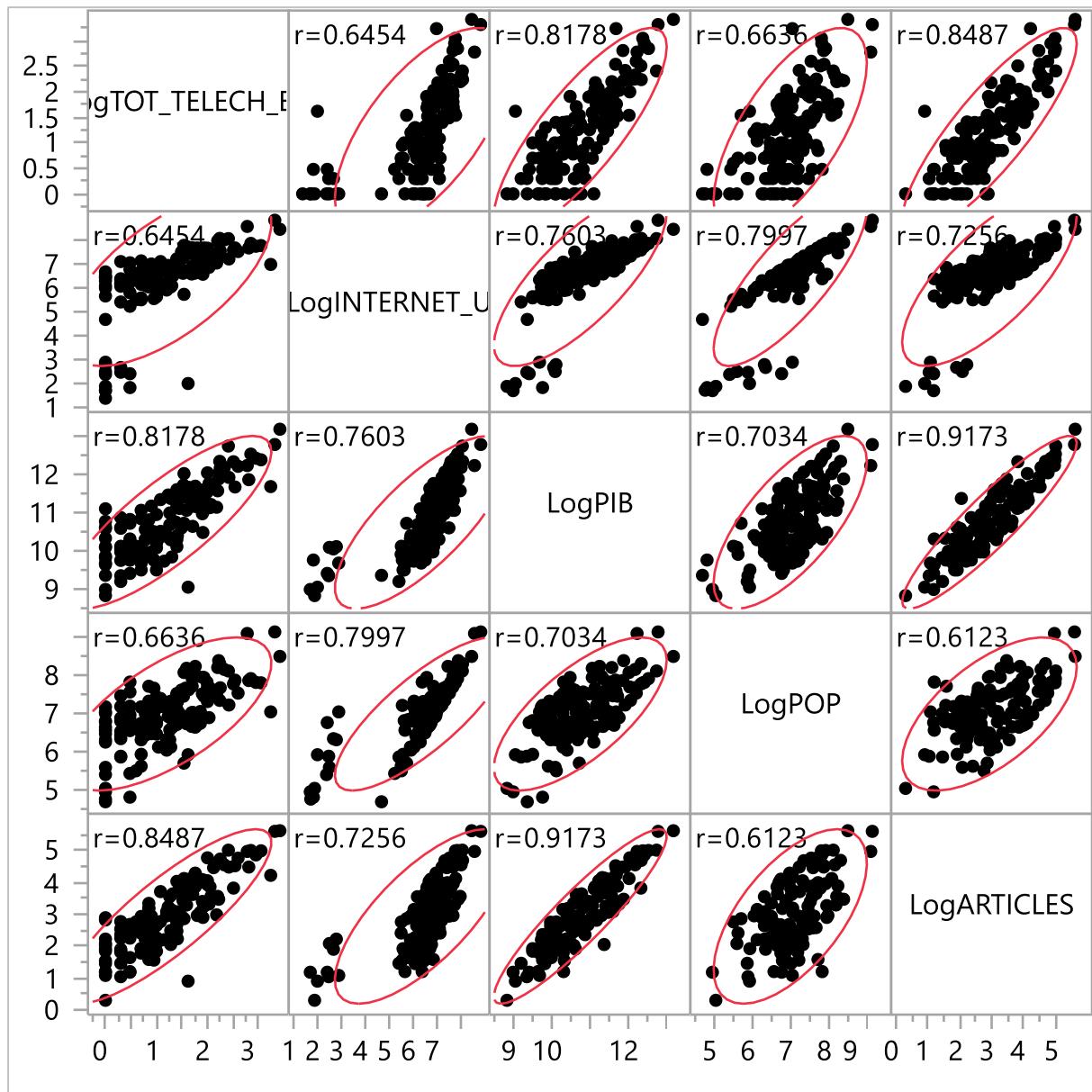


Figure 26 - Matrice de graphiques de nuages de points pour les téléchargements des références en anglais

En ce qui concerne les téléchargements des références en anglais, les coefficients de corrélation sont plus élevés par rapport au modèle précédent indiquant un niveau de corrélation plus fort entre la variable dépendante LogTELECH_EN et les variables explicatives quantitatives. Les résultats de l'analyse de corrélation sont indiqués dans la figure 26 et le Tableau 35. Les coefficients résultants indiquent une intensité de la relation se situant entre 65 et 85 % entre la variable dépendante et les variables explicatives.

Résultats

	LogTELECH_EN	LogINTERNET_U	LogPIB	LogPOP	LogARTICLES
LogTELECH_EN	--	0.65	0.82	0.66	0.85
LogINTERNET_U	0.65	--	0.76	0.80	0.73
LogPIB	0.82	0.76	--	0.70	0.92
LogPOP	0.66	0.80	0.70	--	0.61
LogARTICLES	0.85	0.73	0.92	0.61	--

Tableau 35 - Coefficients de corrélation entre les variables quantitatives du modèle pour les téléchargements des références en anglais

Les coefficients de corrélation obtenus (Tableau 35) nous indiquent des corrélations plus importantes pour les variables explicatives, qui se situent entre 61% et 92% entre les variables LogARTICLES et LogPIB, $r = 0,92$.

Comme pour le premier modèle développé, le VIF est relativement important, 8,20 mais non strictement problématique pour ces deux variables.

Dans le Tableau 36, les indicateurs globaux du modèle retenu sont indiqués. Ainsi, la force de la relation entre les téléchargements des articles en anglais et les quatre variables explicatives reste élevée : R^2 est de 78 %. Le test de Fisher ($F=115,73$) confirme la pertinence des variables explicatives retenues dans le modèle.

R carré	0.777465
R carré ajusté	0.770722
Racine de l'erreur quadratique moyenne	0.392404
Moyenne de la réponse	1.270073
p-value	<0,0001
F	115,29
Observations (ou sommes pondérées)	137

Tableau 36 - Résumé de l'ajustement : indicateurs globaux de la qualité du modèle retenu pour les téléchargements de références en anglais

Les critères utilisés retiennent six variables lors des différentes itérations, mais uniquement quatre s'avèrent statistiquement significatives (Tableau 37).

Les variables explicatives retenues sont :

- Population par pays- LogPOP (quantitatif)
- Nombre d'articles par pays - LogARTICLES (quantitatif)
- Pays Francophones [modalités : Français / Sans affinité] - FR (binaire catégorielle)
- Continent {Amerique Sud&Afrique&Asie&Oceanie-Europe&AmeriqueNord} (catégorielle)

Résultats

Terme	BIC	AICc	< 0,05
LogINTERNET_U	x	x	
LogPIB			
LogPOP	x	x	x
LogARTICLES	x	x	x
FR[Français]	x	x	x
ANGL[Anglais]	x		
Continent{Amérique Sud&Afrique&Asie&Océanie-Europe&AmériqueNord}	x	x	x

Tableau 37 - Variables retenues (en gras) lors de la sélection automatique avec les critères BIC, AICc, avec procédure ascendante.

A l'exception de la variable nouvellement composée, les variables qui influencent le plus ce modèle sont les mêmes que celles retenues dans le modèle précédent dont la variable dépendante était le total de téléchargements. Les poids des coefficients sont différents mais ce sont toujours les variables LogArticles, LogPOP et Francophone, qui influencent le plus le modèle.

La variable explicative **Continent{Amérique Sud&Afrique&Asie&Océanie-Europe&AmériqueNord}** générée lors du processus de sélection automatique des variables sur base de la variable explicative catégorielle **Continent** est différente.

Terme	Estimation	Erreur standard	t ratio	Prob. >	
				t	VIF
Constante	-2.442462	0.38577	-6.33	<.0001	.
LogPOP	0.3345981	0.067551	4.95	<.0001	2.084108
LogARTICLES	0.4990083	0.047996	10.4	<.0001	2.404038
FR[Français]	0.1648502	0.045003	3.66	0.0004	1.075267
Continent{Amérique Sud&Afrique&Asie&Océanie-Europe&AmériqueNord}	-0.118922	0.046013	-2.58	0.0108	1.601794

Tableau 38 - Variables sélectionnées automatiquement avec la méthode pas à pas (et retenues dans le modèle) et estimations des coefficients

Le Tableau 38 détaille les paramètres relatifs à chacune des quatre variables du modèle tandis que la Figure 27 détaille l'expression de la prévision du modèle retenu.

$$\begin{aligned}
 & -2.4424619995532 \\
 & + 0.3345980990735 * \text{LogPOP} \\
 & + 0.4990083119213 * \text{LogARTICLES} \\
 & + \text{Match}\left(\text{FR}\right) \begin{cases} \text{"Français"} \Rightarrow 0.16485021710987 \\ \text{"Sans Afinité"} \Rightarrow -0.1648502171099 \\ \text{autre} \Rightarrow . \end{cases} \\
 & + -0.118922069413 * \text{Continent{Amérique Sud&Afrique&Asie&Océanie-Europe&AmériqueNord}}
 \end{aligned}$$

Figure 27 - Expression de la prévision du modèle retenu pour les téléchargements de références en anglais

Résultats

Compte tenu des valeurs de coefficients de régression (Tableau 38) et de l'expression de la prévision, ce modèle désigne la variable LogARTICLES en tant que variable explicative la plus influente de la régression. Si **LogARTICLES** augmente d'une unité, toutes choses égales par ailleurs, alors le **LogTELECH_EN** augmentera de 0,49 unité.

La variable **LogPOP** vient en deuxième position. Son poids dans le modèle indique que pour une augmentation d'une unité de LogPOP, toutes choses égales par ailleurs, LogTELECH_EN augmentera de 0,33 unités.

La variable Francophone est définie comme suit : Français - 1 ; Sans Affinité : (-)1.

Si un pays est francophone, le **LogTELECH_EN** augmentera de 0,16 unité. Pour les pays ne faisant pas partie de la catégorie francophone, le **LogTELECH_EN** diminuera de 0,16 unité. Ces valeurs sont moins importantes que ceux du modèle précédent. En effet, on peut imaginer que les francophones seront moins tentés de lire la littérature scientifique en anglais. La thématique des articles et les sujets abordés dans la littérature anglophone sont potentiellement moins abordables également.

Le coefficient de la variable composée sur base de la variable **Continent{AmeriqueSud&Afrique&Asie&Oceanie - Europe&AmeriqueNord}** influence moins le modèle. Afin d'interpréter l'impact de cette variable sur le modèle nous l'avons décomposée et calculée en fonction du codage effectuée lors de la régression (Tableau 39).

La variable **Continent{AmeriqueSud&Afrique&Asie&Oceanie-Europe&AmeriqueNord}** est définie comme suit :

Continent	Codage Variable
AmeriqueSud	1
Afrique	1
Asie	1
Oceanie	1
Europe	-1
AmeriqueNord	-1

Tableau 39 - Codage de la variable composée lors de la sélection automatique des variables

Pour les pays se situant en **Amérique du Sud, Afrique, Asie et Océanie**, le **LogTELECH_EN** est diminué de 0,11 unités (-0,11 x 1).

Les pays se situant en **Europe et Amérique du Nord** enregistreront une augmentation des téléchargements de 0,11 unités (-0,11 x -1).

4.3.2.1. Validation du modèle retenu

Les hypothèses sous-jacentes à l'ajustement de ce modèle doivent être validées. Nous allons détailler la distribution des résidus, le diagnostic d'influence (Distance de Cook) ainsi que l'Indépendance des erreurs. Les autres hypothèses (linéarité de la fonction de régression, homoscédasticité, multi colinéarité) ont été vérifiées durant le développement du modèle et sont validées.

4.3.2.2. Vérification de la normalité des résidus

La distribution des résidus de ce modèle suit bien une loi normale (Figure 28). L'histogramme se superpose bien à la courbe normale théorique, à l'exception d'une observation, il s'agit de Djibouti. La droite de Henry, représentée aussi dans la Figure 28, confirme que les résidus suivent bien la droite à quelques exceptions près. Ces représentations valident ainsi la normalité de la distribution des résidus, ce qui signifie que la première hypothèse dans la validation du modèle retenu est remplie.

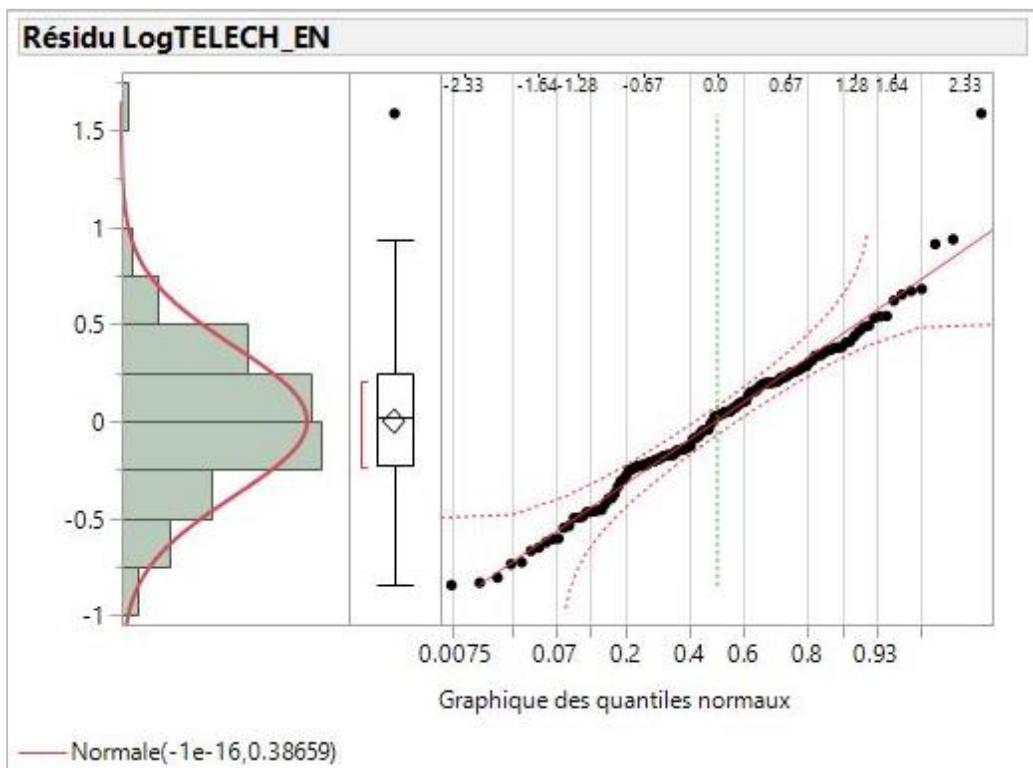


Figure 28 - Distribution des résidus et droite de Henry du modèle retenu

4.3.2.3. Diagnostiques d'influence - distance de Cook

En tenant compte de trois théories qui permettent l'interprétation de la distance de Cook, l'examen de nos données (Figure 29) n'indique aucune valeur influente au sens de la distance de Cook pour ce modèle.

Résultats

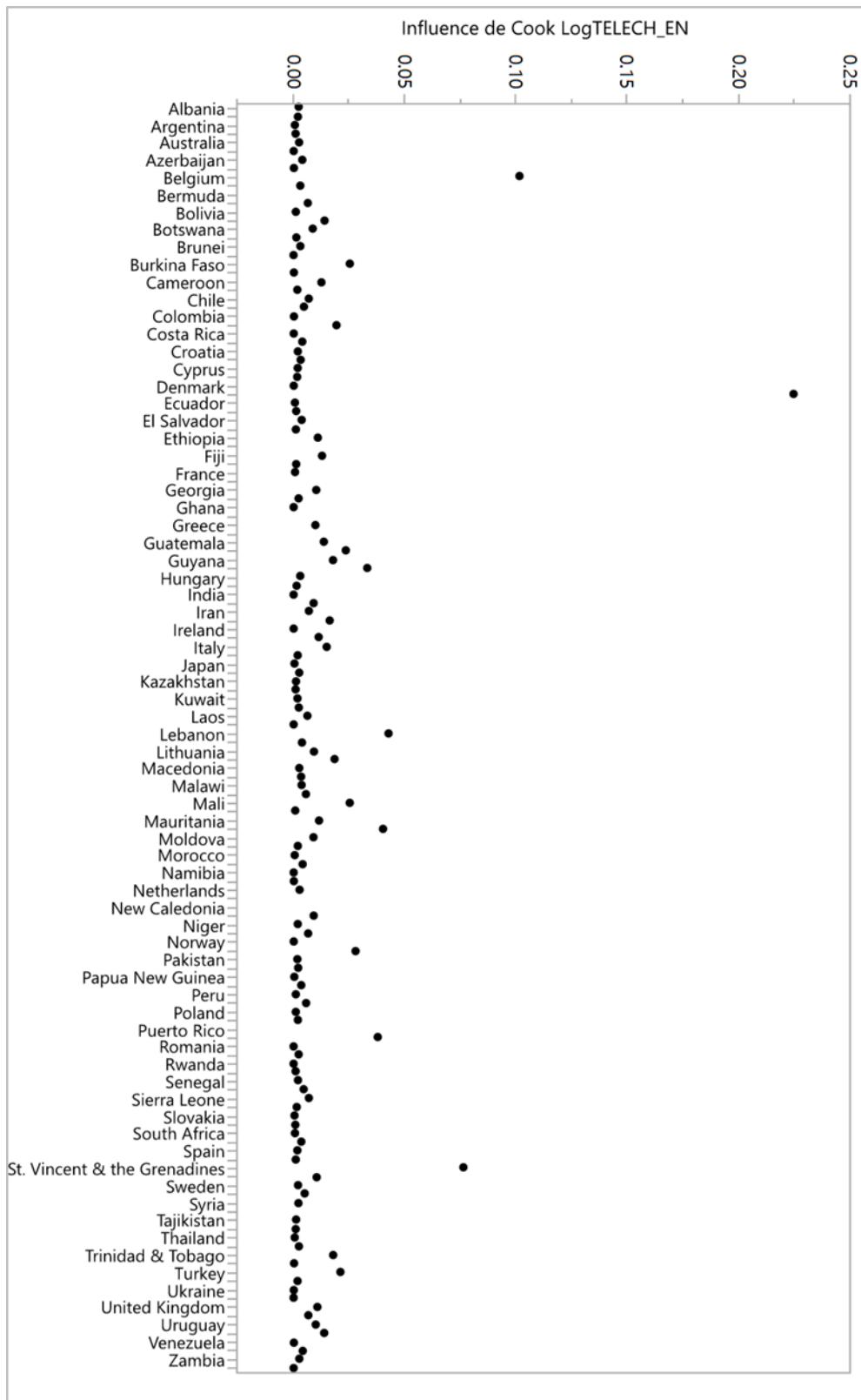
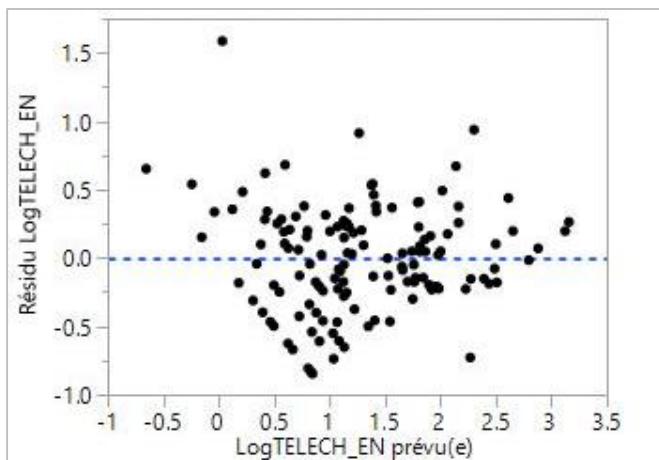


Figure 29 - Distances de Cook pour le modèle de régression retenu – nombre de téléchargements des articles en anglais.

4.3.2.4. Indépendance des erreurs

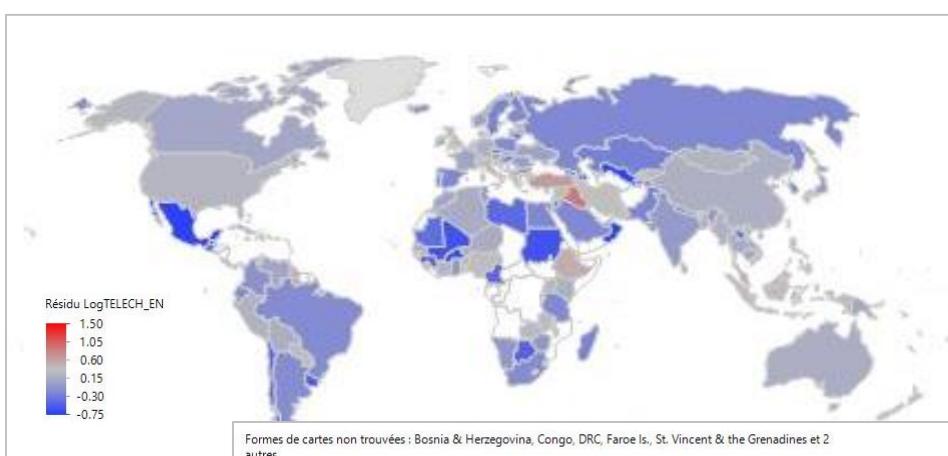
Il n'y a pas de structure marquée qui puisse être remarquée dans la distribution des résidus de notre modèle. Sur le Graphique 4, on peut toutefois relever l'alignement des résidus des pays ayant effectué un seul téléchargement. Cette structure pourrait être considérée comme une faiblesse du modèle.



Graphique 4 - Résidus du modèle retenu contre la prédiction

La Carte 16 représente la valeur des résidus du modèle final associée à chacun des 135 pays ayant téléchargé les références en anglais d'ORBi. Malheureusement les intitulés de quelques pays n'ont pas été reconnus par le logiciel statistique malgré les nombreux essais de remédiation au problème.

Un résidu négatif (bleu) signifie que le modèle surestime le nombre de téléchargements tandis qu'un résidu positif (rouge) signifie que le modèle sous-estime le nombre de téléchargements. Les résidus ne semblent pas être dépendants les uns des autres et l'hypothèse d'Indépendance peut être retenue. A l'analyse de la représentation des résidus au niveau mondial (Carte 16) nous pouvons remarquer que dans ce cas-ci, le modèle surestime légèrement le nombre de téléchargements associés aux articles en Anglais pour la majorité des pays.



Carte 16 - Répartition des résidus au niveau mondial pour les téléchargements des références en langue anglaise

Résultats

4.3.3. Les facteurs qui influencent les téléchargements des publications en langue française de l'échantillon : développement du modèle retenu et résultats

Pour ce modèle-ci, les coefficients de corrélation sont moins élevés par rapport aux deux modèles précédents, indiquant un niveau de corrélation moins fort entre la variable dépendante LogTELECH_FR et les variables explicatives quantitatives. La Figure 30 et le Tableau 35 détaillent les résultats de l'analyse de corrélation.

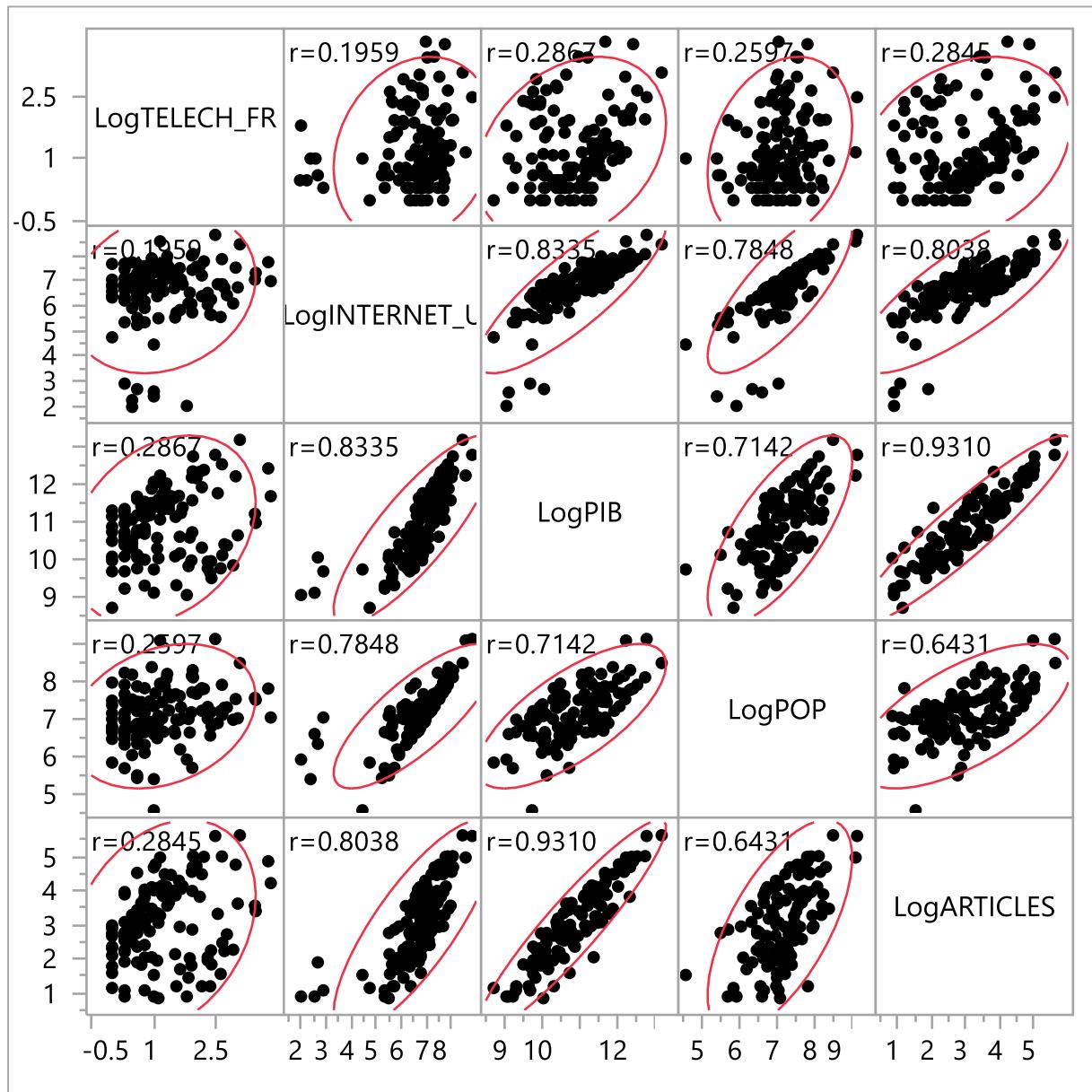


Figure 30 - Matrice de graphiques de nuages de points pour les téléchargements des références en français

Les coefficients résultants indiquent une intensité de la relation se situant entre 20 et 29 % entre la variable dépendante et les variables explicatives. Ces valeurs sont beaucoup moins élevées par rapport aux deux modèles précédents. Etant donné que le VIF ne dépasse pas la valeur de 10 nous poursuivons notre analyse avec toutes les variables à ce stade-ci.

Résultats

Les coefficients de corrélation obtenus (Tableau 40) nous indiquent des corrélations plus importantes pour les variables explicatives, et se situent entre 71 % et 93% entre les variables LogARTICLES et LogPIB, $r = 0,93$.

	LogTELECH_FR	LogINTERNET_U	LogPIB	LogPOP	LogARTICLES
LogTELECH_FR	--	0.19	0.28	0.26	0.28
LogINTERNET_U	0.19	--	0.83	0.78	0.80
LogPIB	0.28	0.83	--	0.71	0.93
LogPOP	0.26	0.78	0.71	--	0.64
LogARTICLES	0.28	0.80	0.93	0.64	--

Tableau 40 - Coefficients de corrélation entre les variables quantitatives du modèle pour les téléchargements des références en français

Comme pour les deux modèles développés précédemment, le VIF est relativement important, 8,70 mais non strictement problématique pour ces deux variables.

Dans le Tableau 41, les indicateurs globaux du modèle retenu sont indiqués. La force de la relation entre les téléchargements des articles en français et les six variables explicatives reste élevée : R^2 est de 75 %. Par ailleurs, 73 % des téléchargements des références en français peuvent être expliqués par les variables qui ont été intégrées dans le modèle avec la sélection automatique des variables.

Le test de Fisher ($F=54,97$) confirme la pertinence des variables explicatives retenues dans le modèle.

R carré	0.746433
R carré ajusté	0.732955
Racine de l'erreur quadratique	
moyenne	0.479688
Moyenne de la réponse	1.247143
p-value	<0,0001
F	54,97
<u>Observations (ou sommes pondérées)</u>	<u>119</u>

Tableau 41 - Résumé de l'ajustement : indicateurs globaux de la qualité du modèle retenu pour les téléchargements des références en langue française.

Dans le Tableau 42, sont reprises les variables qui ont été sélectionnées et retenues lors de la sélection automatique des variables. Les critères utilisés retiennent sept variables lors des différentes itérations, mais six seulement s'avèrent statistiquement significatives (en gras dans le Tableau 42).

Terme	BIC	p-value	AICc	< 0,05
LogINTERNET_U				
LogPIB	x	x	x	x
LogPOP		x	x	x
LogARTICLES	x	x	x	x
FR[Français]	x	x	x	x
ANGL[Anglais]				
Continent{Océanie&Asie&AmériqueSud	-	x	x	x
Europe &Afrique&AmériqueNord}				
Continent{Océanie&Asie - AmériqueSud}	x	x	x	x

Tableau 42 - Variables retenues(en gras) lors de la sélection automatique avec les critères BIC, AICc avec procédure ascendante.

Résultats

Les variables explicatives retenues sont :

- PIB par pays – LogPIB (quantitatif)
- Population par pays - LogPOP (quantitatif)
- Nombre d'articles par pays - LogARTICLES (quantitatif)
- Pays Francophones [modalités : Français / Sans affinité] - FR (binaire catégorielle)
- Continent {Oceanie&Asie&AmeriqueSud - Europe&Afrique&AmeriqueNord}(catégorielle)
- Continent {Oceanie&Asie - AmeriqueSud}(catégorielle)

Terme	Estimation	Erreur		Prob. > t	VIF
		standard	t ratio		
Constante	-3.761	1.132	-3.32	0.0012	.
LogPIB	0.296	0.136	2.18	0.0315	7.866263
LogPOP	0.184	0.084	2.2	0.0298	1.765572
LogARTICLES	0.277	0.102	2.7	0.0079	7.053133
FR[Français]	0.864	0.059	14.68	<.0001	1.408844
Continent{Oceanie&Asie&AmeriqueSud- Europe&Afrique&AmeriqueNord}	-0.135	0.057	-2.39	0.0184	1.475421
Continent{Oceanie&Asie-AmeriqueSud}	-0.209	0.085	-2.47	0.0150	1.181028

Tableau 43 - Variables sélectionnées automatiquement avec la méthode pas à pas (retenues dans le modèle) et estimations des coefficients pour les téléchargements des références en langue française.

Le Tableau 43 détaille les paramètres relatifs à chacune des six variables du modèle tandis que la Figure 31 détaille l'expression de la prévision du modèle retenu.

$$\begin{aligned} \text{LogTELECH_FR} = & -3.761428892681 \\ & + 0.29627257182008 * \text{LogPIB} \\ & + 0.183984787133 * \text{LogPOP} \\ & + 0.27697552728483 * \text{LogARTICLES} \\ & + \text{Match(FR)} \left[\begin{array}{l} \text{"Français"} \Rightarrow 0.86410181460703 \\ \text{"Sans Afinité"} \Rightarrow -0.864101814607 \\ \text{autre} \Rightarrow . \end{array} \right] \\ & + -0.1352227069554 * \text{Continent(Oceanie&Asie&Amerique Sud- Europe&Afrique&AmeriqueNord)} \\ & + -0.2094393011847 * \text{Continent(Oceanie&Asie-Amerique Sud)} \end{aligned}$$

Figure 31 - Expression de la prévision du modèle retenu pour les téléchargements des références en langue française.

Dans ce modèle la variable **LogPIB** a été retenue. Une augmentation d'une unité de cette variable, toutes choses égales par ailleurs, augmentera LogTELECH_FR de 0,29 unité.

Pour une augmentation d'une unité de LogPOP, toutes choses égales par ailleurs, le LogTELECH_FR augmentera de 0,18 unité.

Les coefficients de régression indiquent que pour une augmentation de **LogARTICLES** d'une unité, toutes choses égales par ailleurs, alors le LogTELECH_FR augmentera de 0,27 unité.

Compte tenu des valeurs de coefficients de régression (Tableau 43) et de l'expression de la prévision, ce modèle désigne la variable FR(Francophone), définie comme suit : Français - 1 ; Sans Affinité : (-1), en tant que variable explicative la plus influente de la régression.

Résultats

Un pays francophone aura le LogTELECH_FR augmenté de 0,86 unité. Les pays ne faisant pas partie de la catégorie Francophone auront 0,86 unités en moins pour le LogTELECH_FR.

Les nouvelles variables de ce modèle sont définies comme indiqué dans les Tableaux 44 et 45.

La variable composée **Continent{Oceanie&Asie&AmeriqueSud-Europe&Afrique&AmeriqueNord}** influence moins le modèle. Afin d'interpréter l'impact de cette variable sur le modèle nous l'avons décomposée et calculée en fonction du codage effectué lors de la régression (Tableau 44).

La variable Continent{Oceanie&Asie&AmeriqueSud-Europe&Afrique&AmeriqueNord} est définie comme suit :

Continent	Codage Variable
Oceanie	1
Asie	1
AmeriqueSud	1
Europe	-1
Afrique	-1
AmeriqueNord	-1

Tableau 44 - Codage de la variable composée lors de la sélection automatique des variables

Pour les pays se situant en **Océanie, Asie et Amérique du Sud**, le LogTELECH_FR sera diminué de 0,13 unité (-0,13 x 1).

Les pays se situant en **Europe, Afrique et Amérique du Nord** enregistreront une augmentation du LogTELECH_FR de 0,13 unité (-0,13 x -1). C'est le seul modèle dans lequel les téléchargements des pays d'Afrique sont influencés positivement.

La deuxième variable composée lors de la sélection automatique est **Continent {Oceanie&Asie - AmeriqueSud}**, définie comme suit :

Continent	Codage Variable
Oceanie	1
Asie	1
AmeriqueSud	-1

Tableau 45 - Codage de la deuxième variable composée lors de la sélection automatique des variables

Les téléchargements revenant aux continents **Océanie et Asie** seront diminués de 0,20 unité (-0,20 x 1), tandis que les téléchargements revenant à l'Amérique du Sud seront augmentés de 0,20 unité (-0,20 x -1).

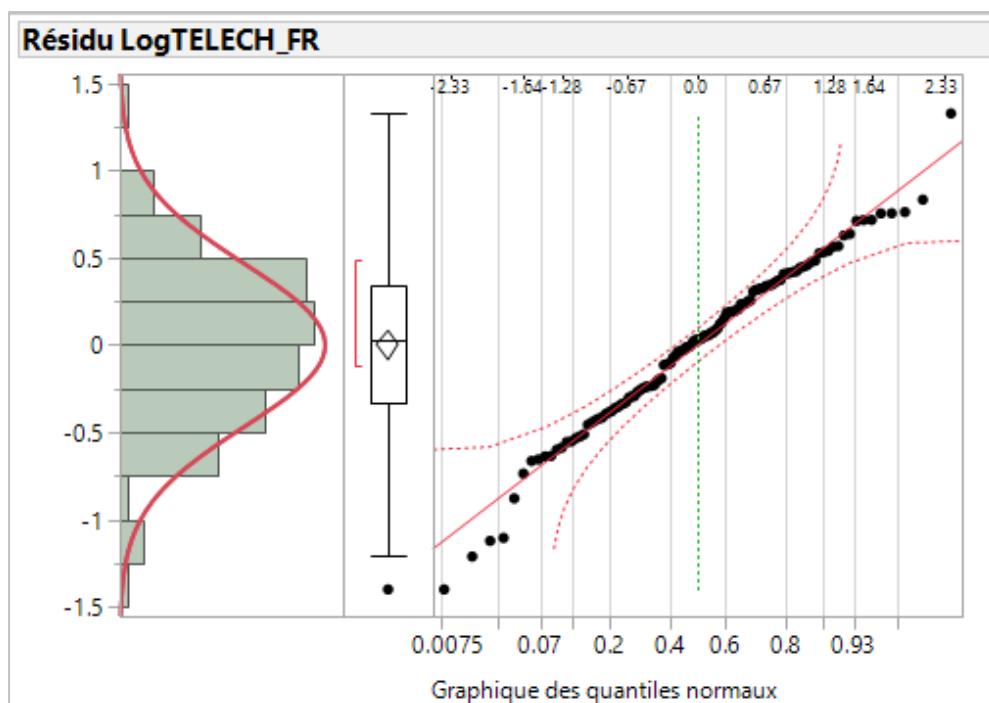
Résultats

4.3.3.1. Validation du modèle retenu

Nous allons détailler la distribution des résidus, le diagnostic d'influence (Distance de Cook) ainsi que l'Indépendance des erreurs. Les autres hypothèses (linéarité de la fonction de régression, homoscédasticité, multi colinéarité) ont été vérifiées durant le développement du modèle et sont validées.

4.3.3.2. Vérification de la normalité des résidus

La distribution des résidus de ce modèle suit bien une loi normale (Figure 32). L'histogramme se superpose bien à la courbe normale théorique. La droite de Henry, représentée aussi dans la Figure 32, confirme que les résidus suivent bien la droite à quelques exceptions près. Ces représentations valident ainsi la normalité de la distribution des résidus, ce qui signifie que la première hypothèse dans la validation du modèle retenu est remplie.



4.3.3.3. Diagnostiques d'influence - distance de Cook

En tenant compte de trois théories qui permettent l'interprétation de la distance de Cook, l'examen de nos données (Figure 33) n'indique aucune valeur influente au sens de la distance de Cook pour ce modèle.

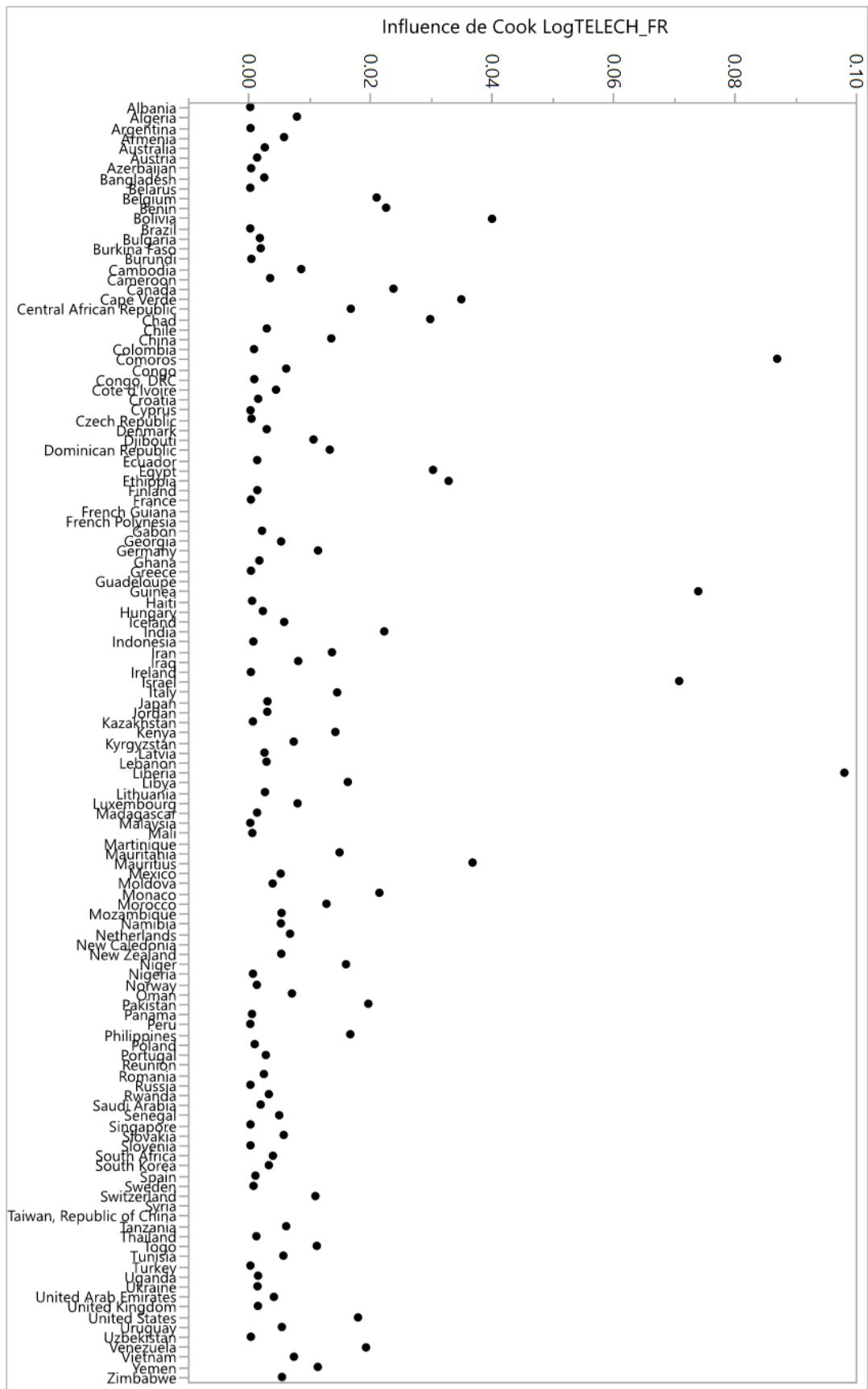
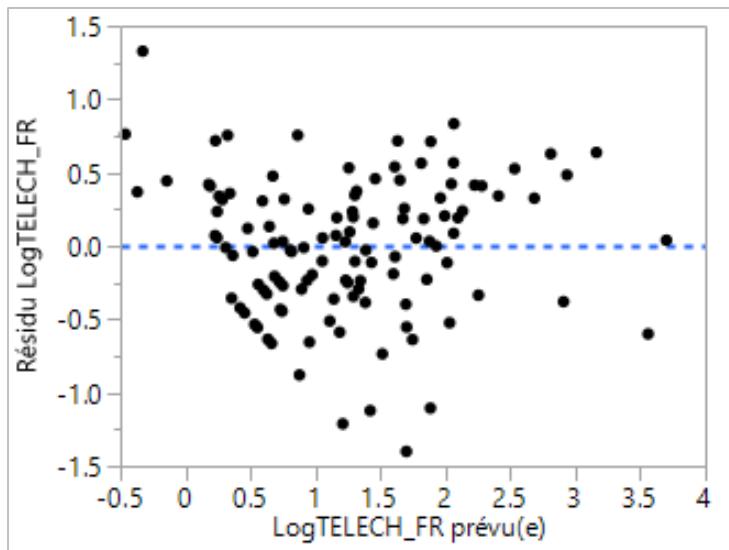


Figure 33 - Distances de Cook pour le modèle de régression retenu pour les téléchargements des références en français

Résultats

4.3.3.4. Indépendance des erreurs

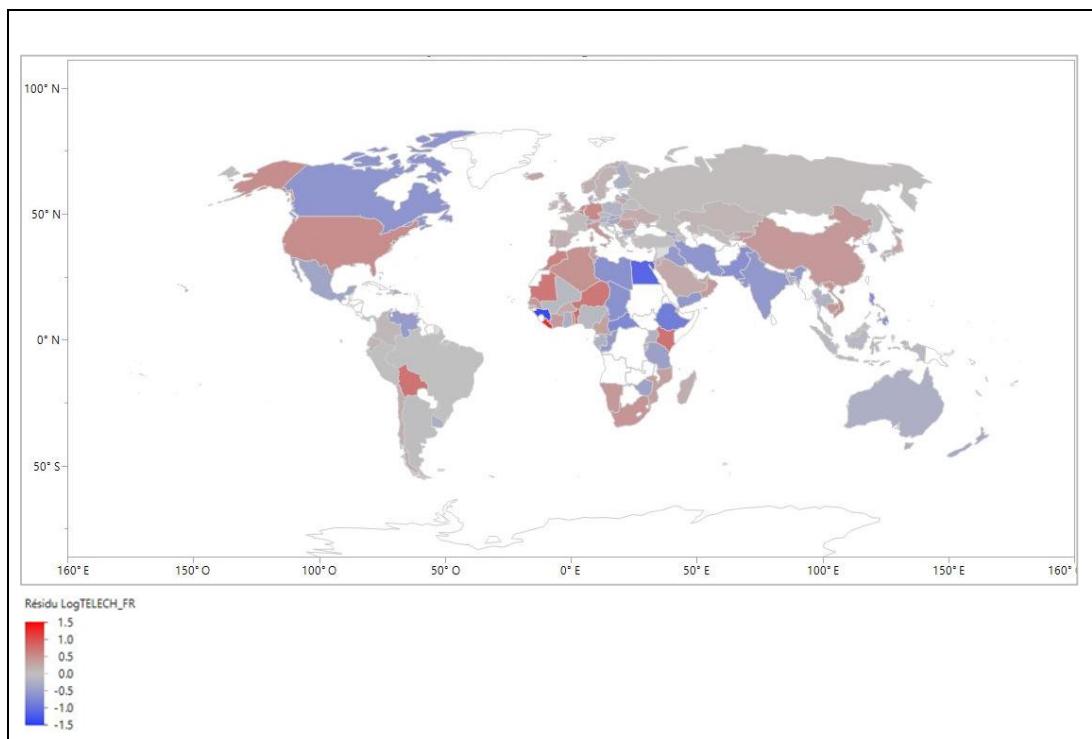
Il n'y a pas de structure marquée qui puisse être repérée dans la distribution des résidus de notre modèle. Sur le Graphique 5, on peut toutefois remarquer l'alignement des résidus des pays ayant effectué un seul téléchargement. Comme pour le modèle précédent (téléchargement d'articles en anglais), cette structure pourrait être considérée comme une faiblesse du modèle.



Graphique 5 - Résidus du modèle retenu contre la prédiction

La Carte 17 représente les résidus du modèle retenu, associés à chacun des 119 pays ayant téléchargé les références en Français d'ORBi. Un résidu négatif (bleu) signifie que le modèle surestime le nombre de téléchargements tandis qu'un résidu positif (rouge) signifie que le modèle sous-estime le nombre de téléchargements. Les résidus ne semblent pas être dépendants les uns des autres et l'hypothèse d'Indépendance peut être retenue. A l'analyse de la représentation des résidus au niveau mondial, Carte 17, nous pouvons remarquer que dans ce cas-ci, le modèle sous-estime le nombre de téléchargements associés aux références en français pour les pays francophones du Maghreb, les Etats Unis et la Chine.

A l'exception des Etats-Unis, il est évident que la langue, française dans ce cas, constitue le facteur majeur influençant le téléchargement des références en français.



Carte 17 - Répartition des résidus au niveau mondial pour les téléchargements des références en français

Résultats

	Tot. Télécharg.		Télécharg. Art. Anglais		Télécharg. Art. Français	
	Coefficient	valeur t	Coefficient	valeur t	Coefficient	valeur t
Constante	-2,44	-6,5	-2,442	-6,33	-3,761	-3,32
LogARTICLES	0,473	10,41	0,499	10,4	0,277	2,7
LogPOP	0,429	6,53	0,334	4,95	0,184	2,2
Francophone(FR)	0,669	15,55	0,164	3,66	0,864	14,68
LogPIB	--	--	--	--	0,296	2,18
Continent{Asia&Africa-Europe&North America}	-0,114	-2,37				
Continent{Amerique Sud&Afrique&Asie&Oceanie-Europe&AmeriqueNord}			-0,118	-2,58		
Continent{Oceanie&Asie&AmeriqueSud-Europe&Afrique&AmeriqueNord}					-0,135	-2,39
Continent{Oceanie&Asie-AmeriqueSud}					-0,209	-2,47
F	147,34		115,29		54,97	
R ² ajusté	0,79		0,77		0,73	
Nombre observations	151		137		119	

Tableau 46 - Résultats des modèles de régression retenus prévoyant les téléchargements du nombre total des articles, des articles en anglais et des articles en français

5. Discussion et Conclusions

Les analyses effectuées dans ce chapitre nous ont permis de visualiser la répartition mondiale des téléchargements effectués à partir d'ORBi d'un échantillon de publications des géographes de l'ULiège ainsi que de déterminer dans quelle mesure ces téléchargements sont influencés par différentes caractéristiques des pays, soit des facteurs socio-économiques.

Pour à la **première hypothèse** posée dans ce chapitre, nos résultats montrent que les références de notre échantillon ont été téléchargées par 164 pays sur les 243 pays pour lesquels *Internet Word Stats* fournit des statistiques. Ces pays appartiennent aux 5 continents tandis que les 10 pays qui téléchargent le plus et effectuent à eux seuls 71% des téléchargements appartiennent à 4 continents, par ordre décroissant des téléchargements continentaux : l'Europe (4 pays), l'Afrique (3 pays), l'Amérique (2 pays) et l'Asie (1 pays).

Si l'intérêt des pays d'Afrique est plus marqué pour la littérature de langue française qu'anglaise, l'intérêt des pays d'Asie se renforce proportionnellement pour la littérature de langue anglaise. L'Australie, qui ne se classe pas dans le peloton de tête, est proportionnellement plus intéressée également par la littérature anglaise, comme d'autres pays qui ne sont ni anglophones ni francophones.

L'approche statistique réalisée dans la deuxième partie de ce chapitre rend compte des variations d'intérêt des pays des différents continents lorsqu'une l'appartenance à la Francophonie est prise en compte et lorsqu'une variable Continent est composée lors du processus de sélection automatique des variables. Les observations de la partie statistique confirment les observations de la première partie. Qu'il s'agisse du nombre total de téléchargements, du nombre de téléchargements des documents en anglais ou du nombre de téléchargements des documents en français, l'appartenance à la Francophonie influence positivement le nombre des téléchargements. Aussi, les téléchargements des pays d'Europe ou d'Amérique du Nord, sont toujours influencés positivement. Pour le nombre de téléchargements en français, les pays d'Afrique et d'Amérique du Sud rejoignent ceux d'Europe et d'Amérique du Nord.

La première hypothèse selon laquelle *par leur présence dans le répertoire institutionnel ORBi, la diffusion des résultats de recherche des géographes liégeois est assurée au niveau mondial*, est confirmée.

En relation avec la **deuxième hypothèse** du chapitre, si nous faisons une distinction sur base de la langue de publication des références, nous constatons que le nombre de téléchargements des références en anglais se répartit sur 146 pays. Les pays ayant effectué les téléchargements sont diversifiés avec une majorité des téléchargements effectués par les Etats-Unis et différents pays d'Europe et d'Asie. Les téléchargements effectués par les pays d'Afrique sont en nombre beaucoup moins importants par rapport aux téléchargements de l'ensemble des références.

Un tiers des téléchargements ont été effectués par des pays anglophones et à peu près un cinquième par les pays francophones. A peu près la moitié des téléchargements sont effectués par de pays pratiquant d'autres langues ce qui n'est pas surprenant compte tenu que l'anglais est la « langue universelle » dans la publication la communication scientifique.

A titre de comparaison, dans son analyse d'un échantillon d'articles en Physique en anglais provenant d'un répertoire institutionnel américain, Bonilla-Calero (2008) a trouvé que les Etats -Unis et le Royaume-Uni étaient les pays qui téléchargeaient le plus les références, suivis par le Canada, la Chine, l'Inde, le Japon et la France.

D'après le modèle statistique, l'appartenance d'un pays à la **Francophonie** augmente le nombre de téléchargements de 0,16 unité pour le LogTELECH_EN, tandis que pour les pays non francophones, le LogTELECH_EN diminue d'autant.

La variable explicative **Anglophone** n'a pas été retenue dans le modèle statistique. Les pays anglophones ne sont pas majoritaires dans le téléchargement des références en anglais. Le fait que les Etats Unis et le Royaume-Uni effectuent des téléchargements des références en anglais n'a pas suffi pour que la variable Anglophone soit retenue dans le modèle.

La deuxième hypothèse selon laquelle les téléchargements des références en anglais sont effectués principalement par les pays anglophones n'est pas confirmée.

En relation avec la **troisième hypothèse** du chapitre, les téléchargements des références en français se répartissent sur 127 pays et ce sont principalement les pays francophones qui téléchargent les articles en français (87% des téléchargements), tandis que les pays anglophones, les Etats Unis principalement dont le nombre de téléchargements n'est pas négligeable, effectuent 5,3% des téléchargements.

La coïncidence de la langue des pays et des références joue donc un rôle important dans le téléchargement des références en français.

De fait, dans la partie statistique, le modèle désigne la variable FR(Francophone) en tant que variable explicative la plus influente de la régression. D'après le modèle, l'appartenance d'un pays à la **Francophonie** augmente le nombre de téléchargements des références en français de 0,86 unité pour le LogTELECH_FR. Les pays ne faisant pas partie de la catégorie Francophone ont 0,86 unité en moins pour le LogTELECH_FR.

L'appartenance à la francophonie influence le LogTELECH_FR (+0,86), de façon plus marquée que les LogTOT_TELECH (+0,67) et les LogTELECH_EN (+0,16). Les francophones seront plus facilement tentés de télécharger les articles de Géographie en français présents dans ORBi. Il est important de mentionner que les articles qui abordent des sujets locaux et généraux, d'intérêt commun pour les scientifiques francophones sont publiés en français majoritairement, ce qui constitue une explication du poids important de la variable Francophone dans ce modèle.

La troisième hypothèse selon laquelle les téléchargements des références en français sont effectués principalement par les pays francophones est confirmée.

Pour ce qui est de la **quatrième hypothèse**, nos observations indiquent 97 références communes à l'ensemble des outils bibliographiques qui sont toutes de langue anglaise tandis que parmi les références uniquement couvertes par GS, on retrouve plus de documents en français (94) qu'en anglais (20).

Cela peut-il favoriser les téléchargements des documents en français à partir d'ORBi, aux dépends du téléchargement des documents de langue anglaise ?

Si on considère les téléchargements des références en anglais, qui représentent 56,7 % de l'échantillon, nous constatons que ces références ont enregistré 37,2 % du total des téléchargements, avec une moyenne de 63 téléchargements par référence.

Malgré un nombre moins important de références en français, soit 42,6% de l'échantillon, celles-ci enregistrent des téléchargements en nombre plus important, soit 62,5% du total des téléchargements. La moyenne des téléchargements par article est aussi beaucoup plus élevée pour les articles en français : 142 téléchargements par référence.

Ces résultats pourraient être liés à une disponibilité plus réduite des références en français *via* les bases de données ou les éditeurs commerciaux. Cependant, du fait que l'analyse statistique a montré que le nombre de téléchargements étaient positivement influencé par la variable FR (Francophone), il serait hasardeux de tirer des conclusions.

Cela semble d'autant plus hasardeux que le nombre moyen de téléchargements des articles en français uniquement disponibles *via* GS (127) ou *via* GS par le seul lien ORBi (98) est inférieur au nombre moyen de téléchargements du total des articles en français (142).

*La quatrième hypothèse selon laquelle *du fait que les publications en anglais sont souvent référencées par plusieurs plateformes et bases de données, le nombre de téléchargements à partir d'ORBi est moins important tandis qu'un nombre plus élevé de téléchargements est enregistré pour les références en français* n'est ni confirmée, ni infirmée.*

En relation avec notre **cinquième hypothèse**, dans la première partie du chapitre, nous avons pu constater que les références « uniques » à GS et celles retrouvées par GS *via* les liens ORBi uniquement, majoritairement de langue française, sont principalement téléchargées par les pays européens et plus particulièrement les pays francophones d'Europe, du Maghreb et de l'Afrique sub-saharienne, ainsi que du Canada.

Les références couvertes par l'ensemble des outils bibliographiques, toutes de langue anglaise, sont aussi téléchargées par les pays développés de l'Europe de l'Ouest, particulièrement la France et la Belgique, ainsi que par le Canada. Cette fois, les Etats-Unis se démarquent par un nombre important de téléchargements ainsi que les pays asiatiques comme la Chine et l'Inde et les pays du Moyen-Orient comme la Turquie, l'Irak, l'Iran, etc.

Encore une fois nous pouvons constater que la langue des références impacte la répartition des téléchargements au niveau mondial. Les répertoires institutionnels permettent la diffusion au niveau mondial d'une information écrite dans une langue nationale autre que l'anglais, qui autrement resterait beaucoup moins accessible.

*La cinquième hypothèse selon laquelle *les téléchargements d'une part des références « uniques » de GS, d'autre part des références répertoriées par GS uniquement via ORBi et enfin des références couvertes par tous les outils bibliographiques, présentent une répartition différente au niveau mondial* est confirmée.*

Pour ce qui est de la **sixième hypothèse**, en ce qui concerne les facteurs socio-économiques, compte tenu des valeurs de coefficients de régression pour LogTOT_TELECH, nous avons vu que la variable FR(**Francophone**) est la variable explicative la plus influente (+0,67).

Le nombre total de téléchargements effectués par un pays est également lié statistiquement au **nombre d'articles publiés** par ce pays. Le LogTOT_TELECH des pays augmente de 0,47 unité pour une unité d'augmentation de LogARTICLES. Il semble logique qu'une nouvelle recherche, qui aboutira avec la publication d'un article, booste la recherche bibliographique, autant *via* GS que *via* les outils bibliographiques traditionnels, et aboutisse globalement à une augmentation des téléchargements à partir des répertoires institutionnels et des plateformes d'éditeurs. Ce processus peut être générateur d'un grand nombre de téléchargements pour des articles en OA présents dans des répertoires institutionnels. Le poids relativement élevé de la variable LogARTICLES sur notre modèle est donc explicable.

La variable **Population** influence aussi le nombre total de téléchargements effectués par un pays : pour une unité d'augmentation de LogPOP, le LogTOT_TELECH augmente de 0,43 unité. En effet, plus un pays a une population importante, plus il y aura de chances que parmi les habitants il y ait des utilisateurs potentiels de ces références, travaillant dans l'enseignement, la recherche ou dans des domaines connexes, ou encore des citoyens intéressés par les sujets traités. Sur base des recherches déjà effectuées, qui montrent que les utilisateurs finaux potentiels de références présentes dans des répertoires institutionnels sont des citoyens (Zuccala, 2010), la contribution de la variable Population au modèle s'explique facilement.

Les variables composées sur base de la variable **Continent**, indiquent que la localisation des pays sur un continent aurait un impact sur le nombre des téléchargements effectués par les pays qui y sont localisés, impact modéré qu'il soit positif ou négatif.

Pour les pays d'Europe et d'Amérique du Nord, la valeur du coefficient de régression sera positive, avec 0,11 unité en plus pour LogTOT_TELECH. Ceci peut être mis en relation avec la représentation cartographique de la distribution des téléchargements. En effet, comme nous l'avons constaté précédemment ce sont les pays développés d'Europe et des Etats-Unis qui sont les plus grands utilisateurs de références de notre échantillon. Les explications tiennent au développement économique des pays se situant sur les deux continents et au nombre de pays ayant effectué des téléchargements. La thématique des recherches ainsi que la proximité linguistique et scientifique des pays favorisent des téléchargements plus importants.

Pour les pays localisés en Asie et Afrique, l'influence est négative avec -0,11 unités pour LogTOT_TELECH. Ce résultat n'est pas immédiatement conforté par l'examen de la distribution des téléchargements sur carte. Comme nous l'avons vu dans l'illustration cartographique, les pays du Maghreb sont de grands utilisateurs des références du département de Géographie présents dans ORBi. Cependant, pour le restant des pays africains les valeurs enregistrées sont moindres, ce qui, compte tenu du nombre relativement élevé de pays en Afrique, expliquerait une valeur négative du coefficient de régression.

La sixième hypothèse selon laquelle *les facteurs socio-économiques des différents pays influencent le nombre total de téléchargements effectués* est vérifiée sur notre échantillon pour un nombre déterminé de facteurs d'influence.

Pour ce qui est de la **septième hypothèse** posée dans ce chapitre, nous avons établi une distinction entre la langue des références (anglais ou français) pour vérifier l'influence des facteurs socio-économiques sur les téléchargements.

D'après les modèles statistiques, le **nombre d'articles publiés par un pays** a une influence positive, à la fois sur les téléchargements des publications de langue française (+ 0,27 unité pour LogTELECH_FR) et de langue anglaise (+0,49 unité pour LogTELECH_EN) pour une unité d'augmentation de LogARTICLES.

La variable **Population** intervient également positivement. Pour une unité d'augmentation de la variable LogPOP, le nombre de téléchargements augmente de 0,18 unité pour LogTELECH_FR et de 0,33 unité pour LogTELECH_EN.

Nous relevons que l'influence du nombre d'articles publiés par le pays et de la population du pays serait moins élevée pour les téléchargements des publications de langue française que pour les téléchargements des publications de langue anglaise.

L'influence positive d'une variable supplémentaire, celle du **PIB**, est mise en évidence uniquement pour le téléchargement des articles en français. Selon le modèle, le LogTELECH_FR augmente de 0,29 unité pour une unité d'augmentation du LogPIB des pays. L'explication pourrait tenir à ce que plus un pays est économiquement développé, plus son PIB est élevé et plus grande est la somme susceptible d'être allouée à l'enseignement et la recherche (équipement des écoles avec du matériel informatique et accès à Internet, développement de la recherche, etc). Il en découlerait une plus grande probabilité que ce pays se classe parmi ceux qui téléchargent le plus de références scientifiques, comme c'est le cas pour la Belgique et la France. Cependant, plus un pays est riche et son enseignement bien doté, plus il peut avoir recours à des bases de données et de la littérature payantes, donc moins de téléchargements de références des répertoires institutionnels.

Comme ce fut le cas pour le nombre de téléchargements total, la localisation des pays sur les différents continents a une influence sur les téléchargements de références en langue anglaise et en française. Les pays se situant en **Europe et Amérique du Nord** enregistreront une augmentation des téléchargements de 0,11 unités pour LogTELECH_EN et 0,13 unités pour le LogTELECH_FR.

Les pays se situant en **Afrique** enregistreront une augmentation des téléchargements de 0,13 unités pour le LogTELECH_FR. Le modèle des téléchargements de publications en français est le seul modèle dans lequel le critère « Afrique » influence positivement les téléchargements.

Comme décrit pour le premier modèle qui se rapportait au nombre total de téléchargements, pour les téléchargements des publications de langue française, avec les pays francophones du Maghreb et l'Afrique Sub-Saharienne, ce sont les pays développés d'Europe et des Etats-Unis qui téléchargent le plus.

Pour l'Europe et les Etats Unis, les explications tiennent au développement économique et de la recherche scientifique, l'intérêt partagé pour des sujets de recherche communs à des pays leaders quant au nombre de publications scientifiques. De fait, les Etats Unis restent les leaders en ce qui concerne la production scientifique au niveau mondial (Baron *et al.*, 2014). Les données recueillies sur la plateforme de la World Bank utilisées dans l'analyse de régression l'indiquent également. Malgré une très belle avancée de la Chine qui est désormais en seconde place après les Etats Unis, les pays d'Europe avec de la tradition dans la publication scientifique se maintiennent (Royaume Uni, Allemagne, France) restent dans le top 10 des pays producteurs

de publications scientifiques comme indiqué dans le rapport « La position scientifique de la France dans le monde 2000-2015 » édité par l’Observatoire des Sciences et Techniques, Paris, en 2018).

En ce qui concerne l’Afrique, c’est principalement l’appartenance à la francophonie de beaucoup de pays ainsi que les collaborations institutionnelles et économiques avec l’Europe qui expliquent les téléchargements importants.

Une influence positive sur le nombre de téléchargements de publications de langue française (+0,20 unité pour le LogTELECH_FR) est également mise en évidence pour les pays **d’Amérique du Sud**, par la deuxième variable composée du modèle. En ce qui concerne l’Amérique du Sud, une explication serait donnée par la montée du Brésil en tant que puissance économique ainsi qu’en tant que productrice de publications scientifiques, 13^{ème} place sur le top 20 de pays producteurs⁵⁹. Sinon, les pays en voie de développement d’Afrique et d’Amérique du sud, ayant moins facilement accès à de bases de données bibliographiques payantes, sont aussi des utilisateurs potentiels des références en français présentes dans de répertoires institutionnelles.

Le téléchargement important effectué par les utilisateurs américains, et par ceux des pays en voie de développement, confirme l’intérêt que cette littérature en français suscite.

Pour les pays se situant en **Amérique du Sud, Afrique, Asie et Océanie** le LogTELECH_EN est diminué de 0,11 unité. Ce résultat est de nouveau conforté par l’analyse de la distribution de téléchargements. A l’exception de la Chine et de l’Inde, peu de téléchargements sont effectués depuis les pays situés sur ces continents pour les articles en anglais.

Les téléchargements des continents **Océanie, Asie et Amérique du Sud** sont diminués de 0,13 unités pour le LogTELECH_FR, résultat conforté par l’analyse de la distribution de téléchargements. Comme nous l’avons constaté avec l’illustration cartographique des données, les pays d’Océanie, Asie et Amérique du Sud, sont les plus faibles utilisateurs des références en Géographie présentes dans ORBi.

Les téléchargements revenant au continents **Océanie et Asie** seront, d’après le coefficient de la deuxième variable composée du modèle, diminués de 0,20 unités pour LogTELECH_FR.

Les valeurs obtenues pour ces variables sont le reflet d’un intérêt moins marqué des pays se trouvant sur ces continents pour les références en Géographie présentes dans ORBi, ce qui pourrait être dû à une affiliation à la francophonie beaucoup moins marquée par rapport à l’Afrique mais aussi à des différenciations plus importantes dans les sujets de recherche. L’impact négatif sur le nombre de téléchargements pour les pays se situant sur ces continents pourrait être traduit aussi par une production scientifique moins importante comme le montrent Baron & Eckert, (2014) mais aussi à des collaborations moins fréquentes.

⁵⁹ Source : Rapport :« La position scientifique de la France dans le monde 2000-2015 » édité par l’Observatoire des Sciences et Techniques, Paris, en 2018.

La septième hypothèse selon laquelle *les facteurs socio-économiques des différents pays influencent le nombre de téléchargements effectués quelle que soit la langue des références* est confirmée.

En **conclusion**, les résultats de ce chapitre nous permettent de conclure qu'une diffusion mondiale des résultats de recherche des géographes liégeois est assurée par leur présence dans ORBi.

La langue des documents influence le nombre de téléchargements effectués et la répartition au niveau mondial de ceux-ci.

En moyenne 76 % des téléchargements des références en Géographie présentes dans ORBi sont expliqués par les caractéristiques des pays ayant effectué ces téléchargements dont principalement la langue, la production scientifique et la taille des pays.

Il convient de relever toutefois que le modèle développé ne tient pas compte du fait que les publications examinées sont belges et qu'il ne faut pas perdre de vue le biais quant au poids attribué aux téléchargements des pays francophones d'Europe.

Malheureusement, il existe peu d'études disponibles basées sur des données de téléchargement globales ou spécifiques au niveau des répertoires institutionnels (Bonilla-Calero, 2008) pour des buts comparatifs.

La difficulté d'accès à la littérature scientifique pourrait être comblée par la disponibilité de ces références dans des répertoires institutionnels. L'important rôle des répertoires institutionnels dans la diffusion de la littérature scientifique est d'autant plus important pour les pays en voie de développement qui n'ont en général pas accès aux bases de données payantes. La problématique de l'écart Nord - Sud pour l'accès à la littérature scientifique est ainsi en partie résolue par la mise à disposition en libre accès de la littérature dans des répertoires institutionnels.

CONCLUSIONS GENERALES DE LA DEUXIEME PARTIE

Les recherches effectuées dans la deuxième partie de la thèse ont porté sur la visibilité de la littérature scientifique produite par les géographes de l'ULiège, toujours au travers des outils bibliographiques mais aussi *via* le répertoire institutionnel ORBi.

Dans le premier chapitre, nous avons vérifié le référencement d'un échantillon de publications par les bases de données commerciales multidisciplinaires, Scopus et WoS, et spécialisée, GeoRef, ainsi que par le moteur de recherche GS, et relevé les particularités de publication. Dans le second chapitre nous nous sommes intéressée à la répartition mondiale des téléchargements effectués à partir d'ORBi, en établissant une distinction en fonction de la langue des publications ou du recouvrement par les différents outils bibliographiques, et en nous penchant sur les facteurs qui pourraient les influencer.

Dans le **premier chapitre**, les résultats renforcent l'idée selon laquelle GS offre une visibilité supérieure aux publications par rapport aux BD commerciales. ORBi contribue incontestablement au renforcement de la performance de GS puisque le moissonnage du répertoire par GS lui permet de retrouver la totalité des références de notre échantillon, quel que soit le type de document.

Par ailleurs, nos résultats montrent que la distribution des publications des membres du département de Géographie entre les types de documents est différente en fonction de l'orientation du sujet vers les SHS ou les ST, même si les articles de périodiques sont le principal type de publications quelle que soit la sous-discipline. Notons qu'en SHS, les chapitres et parties de livres ainsi que les articles et communications sont majoritairement en français.

L'examen des titres de périodiques de notre échantillon nous a également permis de constater qu'au travers du moissonnage d'ORBi, GS étend l'accès à une littérature « locale », qui échappe en grande partie aux bases de données traditionnelles.

Au terme de ce chapitre, nous pourrions affirmer que GS, par l'intermédiaire d'ORBi, fournit une visibilité totale des publications des géographes liégeois et permet un meilleur référencement des publications locales en français.

Dans le **second chapitre**, il s'agissait d'abord de vérifier la répartition mondiale des téléchargements effectués à partir d'ORBi d'une partie de l'échantillon de publications du chapitre précédent, soit les articles et les communications publiés dans un périodique.

La répartition mondiale des téléchargements montre que du fait de leur présence dans le répertoire institutionnel ORBi, la diffusion des résultats de recherche des géographes liégeois est assurée au niveau mondial et que les téléchargements sont principalement répartis entre l'Europe, les Etats-Unis, les pays francophones du Maghreb, la Chine et le Canada. Les pays de l'Afrique Sub-saharienne, l'Inde et la Turquie ont également effectué des téléchargements dans une moindre mesure.

Dans le cas où une distinction de la langue de publication des références est prise en compte, il ressort que les références en français sont téléchargées en plus grand nombre que les références en anglais. Les téléchargements des références en français sont majoritairement effectués par

les pays francophones. *A contrario*, le constat, qui pourrait sembler évident, n'est pas le même pour les références en anglais : seulement un tiers des téléchargements reviennent aux pays anglophones.

En établissant une distinction en fonction du recouvrement des références par les différents outils bibliographiques, nous avons pu constater que les références « uniques » à GS et celles retrouvées par GS *via* les liens ORBi uniquement, sont majoritairement de langue française et principalement téléchargées par les pays européens et plus particulièrement les pays francophones d'Europe, puis par le Maghreb, l'Afrique sub-saharienne et le Canada. Les références couvertes par tous les outils bibliographiques, sont toutes de langue anglaise et sont aussi téléchargées par les pays développés de l'Europe de l'Ouest, particulièrement la France et la Belgique, ainsi que par le Canada.

La localisation des pays sur certains continents ainsi que l'appartenance à la Francophonie influencent positivement le nombre de téléchargements. Les pays francophones tant développés ainsi que ceux en voie de développement sont les plus grands utilisateurs.

En ce qui concerne les résultats fournis par les modèles de régression, nous retenons que le nombre de téléchargements des articles des géographes déposés dans ORBi dépend principalement de la langue du document, la variable Francophone, ayant du poids dans les trois modèles. La localisation des pays sur les continents influence le nombre des téléchargements : les pays se situant en Europe et Amérique du Nord vont enregistrer plus de téléchargements par rapport aux autres continents. Il convient de mentionner que l'Afrique et l'Amérique du Sud se démarquent aussi par rapport aux autres continents pour les téléchargements des références en français. Le nombre d'articles publiés ainsi que la population des différents pays sont également des variables qui influencent positivement le nombre de téléchargements des articles d'ORBi, mais dans une moindre mesure par rapport aux premiers facteurs mentionnés. L'augmentation du PIB augmentera le nombre de téléchargements des articles en français uniquement.

En conclusion, les résultats de l'ensemble des analyses effectuées dans la deuxième partie de la thèse s'accordent sur le fait que la présence des références en Géographie dans le répertoire institutionnel ORBi augmente leur visibilité au niveau mondial *via* GS, quel que soit le type des documents et la langue de publication. Cela est particulièrement important pour les références, principalement francophones, qui ne sont pas répertoriés dans les bases de données commerciales qui gagnent ainsi en visibilité. Les pays ayant effectué le plus de téléchargements sont les pays francophones d'Europe et du Maghreb, les Etats Unis, et dans une moindre mesure les pays asiatiques. La localisation des pays en Europe, Amérique du Nord, Afrique et en Amérique du Sud et l'appartenance à la Francophonie, constituent des facteurs qui contribuent au nombre de téléchargements des références en Géographie d'ORBi.

CONCLUSIONS GENERALES

L'objectif de notre recherche doctorale était de cerner l'influence du développement de l'information scientifique électronique et de l'Open Access sur l'usage et la diffusion des publications scientifiques en Géographie.

Tout au long de notre travail, nous avons comparé différents outils bibliographiques quant à la couverture de l'information scientifique en Géographie. Qu'il s'agisse de retrouver de l'information utilisée par les géographes (première partie de la thèse) ou de l'information produite par les géographes (deuxième partie de la thèse), nous avions toujours d'un côté Google Scholar, outil de l'ère du numérique dont l'objectif est de référencer « une grande variété » de littérature scientifique, et de l'autre les bases de données traditionnelles, qui s'appuient sur une expertise du référencement dès la publications héritée du passé qu'elles adaptent aux nouvelles technologies.

Autant pour les recherches par mots-clés (Partie 1, Chapitre 1), que pour le recouvrement de bibliographies de thèses (Partie 1, Chapitre 3) ou le recouvrement d'une sélection de publications de géographes (Partie 2, Chapitre 1), les résultats placent **GS** en tête du classement des outils bibliographiques analysés d'un point de vue quantitatif, le moteur de recherche ayant trouvé ou retrouvé de très nombreuses références, que ce soit en Géographie physique ou en Géographie humaine. GS se démarque non seulement par un apport important de références « uniques », mais également par les pourcentages élevés de recouvrement des bases de données bibliographiques étudiées. Les références « uniques » sont en majorité des articles de périodiques mais aussi des livres et chapitres de livres, des actes de conférences, des rapports et autres types de documents que les bases de données commerciales n'indexent généralement pas, mais qui peuvent être utiles pour les géographes.

En tant que base de données disciplinaire, **GeoRef** couvre également un plus large éventail de périodiques (dont des périodiques à portée nationale), mais uniquement en Géographie physique. Plusieurs types de documents (« more non-journal material ») sont recensés dans GeoRef ainsi qu'une diversité importante des langues de publication. Aussi, GeoRef peut, dans certains cas, fournir un nombre relativement important de références « uniques » et constitue une ressource qui n'est pas à négliger par les chercheurs travaillant en Géographie physique.

Les performances des deux bases de données commerciales multidisciplinaires, **WoS** et **Scopus**, sont très similaires. L'apport en références « uniques » des deux bases de données commerciales multidisciplinaires n'est pas significatif par rapport à GS. Scopus se démarque cependant de WoS par un plus grand nombre de périodiques indexés avec une plus grande variété de pays d'édition et de langues des textes. Comme nous l'avons déjà signalé dans l'introduction générale, Scopus, qui est apparu sur le marché beaucoup plus récemment que WoS, étend progressivement son corpus de références et inclut de plus en plus des périodiques « nationaux » et de la littérature en SHS. Les périodiques indexés par les deux bases de données sont majoritairement de langue anglaise et *peer-reviewed* (Partie 1, Chapitres 2 et 3).

Les chercheurs acquis au système de publication de type « publish or perish », pour lesquels le facteur d'impact des périodiques constitue un critère important et qui se soumettent à l'hégémonie anglo-américaine, autant pour la langue que pour les sujets de recherche,

trouveront probablement les références dont ils ont besoin dans WoS, plateforme la plus traditionnelle pour laquelle les origines remontent à 1955, année de création du Science Citation Index qui était consultable sur papier.

Les chercheurs intéressés par des sujets plus divers, au niveau régional et dans des langues autres que l'anglais, publiés sous d'autres formes que des articles dans des périodiques à facteur d'impact, sont susceptibles de trouver plus de références dans GS ainsi que, dans une moindre mesure, dans Scopus ou même GeoRef pour la Géographie physique.

Il n'est pas dans nos propos de remettre en question l'intérêt et la qualité de la littérature scientifique internationale, telle celle répertoriée par WoS. Il semble plus légitime de se poser des questions sur l'intérêt et la qualité de la littérature scientifique qui n'est pas répertoriée par les bases de données traditionnelles, mais à laquelle un outil tel que GS donne accès.

Dans notre travail, les listes de documents cités par les trois doctorants indiquent, même s'il faut être prudent compte tenu du petit échantillon, des habitudes de citation différentes en fonction de l'orientation SHS ou ST du sujet de recherche, qui influence la balance entre les articles de périodiques et les livres cités. Les citations correspondant aux articles de périodiques publiés en anglais durant la dernière décennie avant la date de défense des thèses sont majoritaires pour les trois bibliographies, dépassant largement les autres types de documents cités. L'apport en références « uniques » de GS est plus important pour les thèses à orientation SSH, avec le bémol que les livres et chapitres de livres ne sont couverts que partiellement par GS.

Pour ce qui est des publications des chercheurs travaillant dans le département de Géographie, il s'agit principalement d'articles autant en SHS qu'en ST, parus dans des périodiques majoritairement *peer-reviewed*. Cependant, malgré une dominance des articles de périodiques dans la publication scientifique des géographes quelle que soit la discipline, les chapitres et parties de livres ainsi que les articles et communications en SHS sont majoritairement en français.

En ce qui concerne la **visibilité** de la littérature scientifique produite par les géographes de l'ULiège, nous avons pu constater une excellente performance de GS, suivi par Scopus, WoS et GeoRef. C'est grâce au moissonnage d'ORBi que les références sont toutes visibles et accessibles aux chercheurs *via* GS, même si la performance de GS est de toute façon supérieure à celle des bases de données. Du fait de leur présence dans le répertoire institutionnel ORBi, les publications en langue française des géographes liégeois, publiées dans les périodiques « locaux », sont particulièrement plus visibles notamment *via* GS qui leur assure une diffusion internationale.

La représentation cartographique des téléchargements nous a permis de constater que la langue des références impacte leurs téléchargements. Les résultats de la partie statistique renforcent ces constatations et nous permettent de conclure que la langue des références, la localisation des pays sur les différents continents, ainsi que l'appartenance des pays qui effectuent les téléchargements à la Francophonie influencent positivement le nombre de téléchargements. Les pays francophones développés ainsi que ceux en voie de développement sont les plus grands utilisateurs.

L'émergence et la multiplication de répertoires institutionnels qui facilitent l'accès en Open Access aux publications et l'existence de moteurs de recherche académiques modifient le paysage documentaire. Même si les échantillons considérés dans notre étude sont de petite taille, l'ensemble des résultats obtenus et leur confrontation à la littérature préexistante confortent l'idée que l'information électronique et de l'Open Access répond à un besoin supplémentaire en Géographie, que ne remplissent pas les bases de données bibliographiques traditionnelles. Ce besoin a peut-être été masqué pendant des années par des éditeurs de bases de données, les mêmes qui ont convaincu une partie de la communauté scientifique que seuls les périodiques qu'ils éditaient étaient porteur d'information scientifique de qualité. En ce qui concerne la langue de publication, par exemple, il est interpellant de constater que dans une discipline internationale comme la médecine, la « *Revue médicale de Liège* » dont les articles sont en français occupe la première place pour ce qui est du nombre de publications des chercheurs liégeois toutes disciplines confondues⁶⁰ et une place de choix pour ce qui est des citations⁶¹.

Face à l'augmentation constante des coûts de l'accès à la littérature scientifique, les institutions autant que les chercheurs se voient obligés de chercher des alternatives. En prenant en considération l'ensemble des résultats de ce travail, nous pouvons avancer que GS peut constituer une alternative aux BD classiques pour la Géographie. Le mode de recherche offert par GS correspondrait bien aux habitudes et modes de fonctionnement actuels des chercheurs qui sont plus dans l'immédiateté et l'optimisation du temps dont ils disposent pour la recherche bibliographique.

Pourtant, comme nous l'avons signalé à plusieurs reprises au long des différentes parties de notre thèse, il ne faut pas oublier les inconvénients de GS, que nous avons détaillés dans la description de l'outil. Les limites de GS sont, selon nous, à chercher plus dans les possibilités de traitement des résultats, lorsque ceux-ci sont très nombreux, que dans l'étendue et la qualité de l'information couverte.

Par ailleurs, comme déjà mentionné, il ne faudrait pas remplacer le monopole détenu par les éditeurs commerciaux par le monopole des GAFAM⁶², entreprises de l'économie Internet les plus puissantes auquel appartient Google. La gratuité de Google ou de Google Scholar n'est qu'apparente. Les GAFAM ont la main mise sur les données des usagers d'Internet, lesquelles constituent un enjeu majeur de l'économie d'aujourd'hui et de demain. L'Union Européenne s'attaque régulièrement à Google à qui elle reproche son hégémonie et un abus de position dominante. Le monde académique doit être conscient de sa responsabilité dans la mise en place de monopoles et des dangers qu'ils représentent.

En **conclusion**, durant les trois dernières décennies, nous avons été confrontés à une commercialisation excessive et à une marchandisation mondiale de la production scientifique qui nuit à la diffusion de la science. Cette marchandisation doit être évaluée et repensée afin de

⁶⁰ 92 publications en 2017 d'après Scopus (donnée fournie par Dr Fabienne Prosmans, ULiège)

⁶¹ 97ème position sur plus de 8 444 titres de périodiques, avec 99 citations d'articles en 2017 par les chercheurs de l'ULiège, d'après Scopus (donnée fournie par Dr Fabienne Prosmans, ULiège)

⁶² Acronyme pour Google Apple Facebook Amazon Microsoft

ramener la littérature scientifique à la portée des scientifiques sans coûts exorbitants. Etant donné que le prix des bases de données bibliographiques commerciales ne cesse d'augmenter et que les budgets des bibliothèques académiques n'arrivent plus à suivre ces augmentations, nous pensons que des solutions doivent être trouvées dans la diversité. L'accès gratuit, fourni autant par GS que par les répertoires institutionnels tels qu'ORBi, sont actuellement des atouts primordiaux pour accéder à la littérature scientifique et diffuser ses publications. Par ailleurs, le développement du support numérique offre des opportunités au monde académique de se réapproprier, au moins en partie, la diffusion des résultats de la recherche. Mais il faut continuer à positionner les universités, leurs bibliothèques et les chercheurs qui ont un rôle important à jouer dans le futur de la production et de la publication scientifique. Ce sont les chercheurs qui sont à la base de la production scientifique, qui génèrent la "matière première" : ceux-ci doivent agir et retourner à leur avantage les changements apportés par la digitalisation et les différents mouvements qui tentent de faciliter l'accès à la littérature scientifique.

PERSPECTIVES

Pour en revenir à des considérations pragmatiques, ce travail a apporté quelques éclairages sur l'efficacité des outils bibliographiques à retrouver des résultats lors de recherches bibliographiques en Géographie ainsi que sur la diffusion et la répartition mondiale des téléchargements d'un échantillon de références déposés dans ORBi par les géographes de l'ULiège. Compte tenu de la complexité du sujet, nous pouvons affirmer que beaucoup reste à faire. Ainsi, nous souhaitons évoquer dans cette dernière section les perspectives de recherche qui sont apparues au fil de nos analyses.

Premièrement, nous avons souligné, dans le premier chapitre, le manque de transparence qui persiste quant au contenu de GS. Notre recherche confirme sans nul doute l'absence des informations et la difficulté du traitement des données. Nous avons en effet été plusieurs fois contraints de revoir nos ambitions en raison du nombre trop important de données et de la durée du temps que leur traitement aurait engendré. Par exemple, dans le premier chapitre nous avons dû nous contenter de faire une étude de cas avec les résultats correspondant à deux mots-clés. Initialement les résultats des dix mots-clés, en Géographie physique et humaine, devaient être analysés, ce qui nous aurait permis de mieux percevoir le contenu de GS dans les deux sous-disciplines. Toutefois, en raison du nombre trop important des résultats obtenus et du fait qu'uniquement les premiers mille résultats s'affichent et peuvent être exportés vers un outil de gestion de références bibliographiques (avec les contraintes détaillées dans le Chapitre 1 de la Partie 1), nous n'avons pas été en mesure de contourner cette difficulté. Ainsi, persévérer dans la recherche de données et leur analyse pour évaluer l'apport de GS en termes de résultats de recherches bibliographiques en Géographie apparaît comme une piste à envisager pour de futures recherches.

Deuxièmement, notre travail s'est focalisé sur les résultats obtenus avec de mots-clés en Géographie physique et humaine (dans le premier chapitre de la première partie) et sur la littérature utilisée par les doctorants géographes (dans le troisième chapitre de la première partie

de la thèse). Malgré le fait que la revue de la littérature a mis en lumière des recherches effectuées sur les outils bibliographiques dans d'autres disciplines, des analyses supplémentaires devraient être réalisées afin de découvrir le potentiel de GS en comparaison d'autres bases de données, pour d'autres champs de recherche que la Géographie.

Troisièmement, nous souhaitons revenir sur les modèles de régression créés afin d'identifier les facteurs qui influencent les téléchargements des publications de géographes déposées dans ORBi. Compte tenu que les références examinées sont celles des chercheurs belges d'une université francophone, nous avons constaté que la langue de publication représente un facteur important dans le nombre de téléchargements. La langue des pays effectuant les téléchargements à un poids important dans les modèles de régression. Aussi, nous pensons qu'il serait intéressant de mener plus d'investigations pour identifier les facteurs qui influencent les téléchargements des articles des géographes issus d'une université anglophone, par exemple, ou d'une discipline moins liée à la diversité des contextes régionaux, comme la Chimie ou la Physique, ou encore la Médecine. Nous serions ainsi en mesure de vérifier si ce sont les mêmes facteurs qui influencent le nombre de téléchargements. Enfin, n'oublions pas un aspect déjà été maintes fois démontré dans la littérature, l'augmentation du nombre de citations pour les références qui se trouvent dans les répertoires institutionnels en libre accès. Il serait dès lors très intéressant d'examiner les citations des références que nous avons analysées dans la deuxième partie de notre travail.

Quatrièmement, les analyses développées dans ce travail sont principalement de type quantitatif. Nous pensons qu'une manière de donner du poids à nos conclusions serait de poursuivre les recherches par des analyses qualitatives plus étendues. Concrètement, nous pensons à la réalisation d'enquêtes auprès de chercheurs, de doctorants et d'étudiants dans le cadre de cours en recherche bibliographique. Au final, ce sont eux les premiers acteurs touchés par l'accessibilité ou l'inaccessibilité de la littérature scientifique. Ainsi, les doctorants et les chercheurs semblent être les plus habilités à identifier les manquements concernant l'accès au savoir. Car si la révolution numérique dans la recherche bibliographique change la recherche, ce sont aussi les chercheurs, les institutions et les grands groupes commerciaux qui orienteront la façon de faire de la recherche et de la communiquer.

BIBLIOGRAPHIE

- Aalbers, M. B. (2013). What if we all work in between? Notes on the geography of geographical knowledge production and consumption. *Dialogues in Human Geography*, 3(2), 209–212. <https://doi.org/10.1177/2043820613492640>
- Aalbers, M. B., & Rossi, U. (2006). Beyond the Anglo-American hegemony in human geography: a European perspective. *GeoJournal*, 67(2), 137–147.
- Archambault, É., Côté, G., Struck, B., & Voorons, M. (2016). Research impact of paywalled versus open access papers. *Science-Metrix and Iscience*.
- Archambault, É., Vignola-Gagné, É., Côté, G., Larivi? re, V., & Gingrasb, Y. (2006). Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics*, 68(3), 329–342.
- Ashton, S. V., & Oppenheim, C. (1978). A method of predicting Nobel prizewinners in chemistry. *Social Studies of Science*, 8(3), 341–348.
- Bar-Ilan, J. (2006). An ego-centric citation analysis of the works of Michael O. Rabin based on multiple citation indexes. *Information Processing & Management*, 42(6), 1553–1566.
- Bar-Ilan, J. (2008). Which h-index?—A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, 74(2), 257–271.
- Bar-Ilan, J., Levene, M., & Lin, A. (2007). Some measures for comparing citation databases. *Journal of Informetrics*, 1(1), 26–34.
- Barnett, P., & Lascar, C. (2012). Comparing unique title coverage of Web of Science and Scopus in Earth and atmospheric sciences. *Issues in Science and Technology Librarianship*.
- Bartol, T., Budimir, G., Dekleva-Smrekar, D., Pusnik, M., & Juznic, P. (2014). Assessment of research fields in Scopus and Web of Science in the view of national research evaluation in Slovenia. *Scientometrics*, 98(2), 1491–1504.
- Bauer, K., & Bakkalbasi, N. (2005). An examination of citation counts in a new scholarly communication environment. *D-Lib Magazine*.
- Becker, D. A., & Chiware, E. R. T. (2015). Citation Analysis of Masters' Theses and Doctoral Dissertations: Balancing Library Collections With Students' Research Information Needs. *The Journal of Academic Librarianship*, 41(5), 613–620.
- Benabou, V.-L. (2010). Les publications scientifiques: faut-il choisir entre libre accès et libre recherche? *Hermès, La Revue*, (2), 95–106.
- Bernadette Joseph. (1997). Liste chronologique des principaux périodiques français de géographie. *Cybergeo : European Journal of Geography*. <https://doi.org/10.4000/cybergeo.5399>
- Bollen, J., & Sompel, H. Van de. (2008). Usage impact factor: the effects of sample characteristics on usage-based impact metrics. *Journal of the Association for Information Science and Technology*, 59(1), 136–149.

- Bollen, J., Van de Sompel, H., Hagberg, A., & Chute, R. (2009). A principal component analysis of 39 scientific impact measures. *PloS One*, 4(6), e6022.
- Bollen, J., Van de Sompel, H., & Rodriguez, M. A. (2008). Towards usage-based impact metrics: first results from the MESUR project. In *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital libraries* (pp. 231–240). ACM.
- Bonato, S. (2016). Google Scholar and Scopus for Finding Gray Literature Publications. *Journal of the Medical Library Association*, 104(3), 252–255.
- Bonilla-Calero, A. I. (2008). Scientometric analysis of a sample of physics-related research output held in the institutional repository Strathprints (2000-2005). *Library Review*, 57(9), 700–721.
- Bosc, H., & Harnad, S. (2005). In a paperless world a new role for academic libraries: providing open access. *Learned Publishing*, 18(2), 95–100.
- Bosman, J. (2009). The changing position of society journals in geography. *Tijdschrift Voor Economische En Sociale Geografie*, 100(1), 20–32.
- Bosman, J., Mourik, I. van, Rasch, M., Sieverts, E., & Verhoeff, H. (2006). *Scopus reviewed and compared: The coverage and functionality of the citation database Scopus, including comparisons with Web of Science and Google Scholar*. Utrecht University Library.
- Bøyum, I., & Aabø, S. (2015). The information practices of business PhD students. *New Library World*, 116(3/4), 187–200.
- Braun, T., Glänzel, W., & Schubert, A. (2000). How balanced is the Science Citation Index's journal coverage? - A preliminary overview of macrolevel statistical data. In *The Web of Knowledge: A Festschrift in Honor of Eugene Garfield* (pp. 251–278).
- Brown, B. N. (2009). A comparative analysis of geoscience literature databases. In *Geological Society of America, 2009 annual meeting*.
- Castillo, M. (2010). Measuring academic output: the H-index. Am Soc Neuroradiology.
- Castree, N., Rogers, A., & Sherman, D. (2005). *Questioning geography: fundamental debates*. Wiley-Blackwell.
- Cavacini, A. (2015). What is the best database for computer science journal articles? *Scientometrics*, 102(3), 2059–2071. <https://doi.org/10.1007/s11192-014-1506-1>
- Chen, X. (2010a). Google Scholar's dramatic coverage improvement five years after debut. *Serials Review*, 36(4), 221–226.
- Chen, X. (2010b). The declining value of subscription-based abstracting and indexing services in the new knowledge dissemination era. *Serials Review*, 36(2), 79–85.
- Clement, E., & Ogburn, J. L. (1996). Searching GeoRef for archaeology. *Behavioral & Social Sciences Librarian*, 14(1), 1–10.
- Cole, C., Davis, A. R., Eyer, V., & Meier, J. J. (2018). Google Scholar's Coverage of the Engineering Literature 10 years Later. *The Journal of Academic Librarianship*. [https://doi.org/https://doi.org/10.1016/j.acalib.2018.02.013](https://doi.org/10.1016/j.acalib.2018.02.013)

- Codic, K. S. (2015). Citation analysis of student dissertations and faculty publications in reading and educational leadership at Oakland University. *The Journal of Academic Librarianship*, 41(5), 548–557.
- Conkling, T. W., Harwell, K. R., McCallips, C., Nyana, S. A., & Osif, B. A. (2010). Research material selection in the pre-web and post-web environments: An interdisciplinary study of bibliographic citations in doctoral dissertations. *The Journal of Academic Librarianship*, 36(1), 20–31.
- Conrad, L. Y., Leonard, E., & Somerville, M. M. (2015). New Pathways in Scholarly Discovery: Understanding the Next Generation of Researcher Tools. In *Association of College and Research Libraries Conference* (pp. 25–28).
- Cox, J. (2003). *Scholarly publishing practice: the ALPSP report on academic journal publishers policies and practices in online publishing*. Association of Learned and Professional Society Publishers London (England).
- Craig, J. E. G. (1969). Characteristics of use of geology literature. *College & Research Libraries*, 30(3), 230–236.
- Crissinger, J. D. (1981). The use of journal citations in theses as a collection development methodology. In *Proceedings of the Fifteenth Meeting of the Geoscience Information Society* (Vol. 11, pp. 113–124).
- Crosnier, H. Le. (1997). Le périodique scientifique: son rôle dans la communication de la science. In *L'information scientifique et technique : Nouveaux enjeux documentaires et éditoriaux : Compte-rendu d'un colloque national sur l'information scientifique et technique, Tours (France), 21-23 octobre 1996*. (p. 59). Paris: INRA.
- Cusker, J. (2013). Elsevier Compendex and Google Scholar: a quantitative comparison of two resources for engineering research and an update to prior comparisons. *The Journal of Academic Librarianship*, 39(3), 241–243.
- Dagnelie, P. (2011). *Statistique théorique et appliquée: Inférence statistique à une et à deux dimensions* (Vol. 2). De Boeck Université.
- Davis, P., & Fromerth, M. (2007). Does the arXiv lead to higher citations and reduced publisher downloads for mathematics articles? *Scientometrics*, 71(2), 203–215.
- Davis, P. M. (2010). *Access, readership, citations: A randomized controlled trial of scientific journal publishing*. Cornell University. Retrieved from <https://ecommons.cornell.edu/bitstream/handle/1813/17788/Davis%2CPhilip.pdf?sequence=1>
- Davis, P. M., Lewenstein, B. V., Simon, D. H., Booth, J. G., & Connolly, M. J. L. (2008). Open access publishing, article downloads, and citations: randomised controlled trial. *BMJ*, 337, a568.
- Davis, P. M., & Price, J. S. (2006). eJournal interface can influence usage statistics: implications for libraries, publishers, and Project COUNTER. *Journal of the Association for Information Science and Technology*, 57(9), 1243–1248.
- Davis, P. M., & Solla, L. R. (2003). An IP-level analysis of usage statistics for electronic journals in chemistry: Making inferences about user behavior. *Journal of the Association*

- for Information Science and Technology*, 54(11), 1062–1068.
- Davis, P. M., & Walters, W. H. (2011). The impact of free access to the scientific literature: a review of recent research. *Journal of the Medical Library Association: JMLA*, 99(3), 208.
- de Moya-Anegón, F., Chinchilla-Rodríguez, Z., Vargas-Quesada, B., Corera-Álvarez, E., Muñoz-Fernández, F. J., González-Molina, A., & Herrero-Solana, V. (2007). Coverage analysis of Scopus: A journal metric approach. *Scientometrics*, 73(1), 53–78. <https://doi.org/10.1007/s11192-007-1681-4>
- De Sutter, B., & Van Den Oord, A. (2012). To be or not to be cited in computer science. *Communications of the ACM*, 55(8), 69–75.
- de Vries, J. (2001). Peer review: the holy cow of science. *A Century of Science Publishing*, 231–244.
- De Winter, J. C. F., Zadpoor, A. A., & Dodou, D. (2014). The expansion of Google Scholar versus Web of Science: a longitudinal study. *Scientometrics*, 98(2), 1547–1565.
- Duan, Y., & Xiong, Z. (2017). Download patterns of journal papers and their influencing factors. *Scientometrics*. <https://doi.org/10.1007/s11192-017-2456-1>
- Edwards, S. (1999). Citation analysis as a collection development tool: A bibliometric study of polymer science theses and dissertations. *Serials Review*, 25(1), 11–20.
- Enger, K. B. (2009). Using citation analysis to develop core book collections in academic libraries. *Library & Information Science Research*, 31(2), 107–112.
- Fagan, J. C. (2017). An evidence-based review of academic web search engines, 2014–2016: Implications for librarians' practice and research agenda.
- Falagas, M. E., Pitsouni, E. I., Malietzis, G. A., & Pappas, G. (2008). Comparison of PubMed, Scopus, web of science, and Google scholar: strengths and weaknesses. *The FASEB Journal*, 22(2), 338–342.
- Feyereisen, P., & Spoiden, A. (2009). Can local citation analysis of master's and doctoral theses help decision-making about the management of the collection of periodicals? A case study in psychology and education sciences. *The Journal of Academic Librarianship*, 35(6), 514–522.
- Foster, J., Muellerleile, C., Olds, K., & Peck, J. (2007). Circulating economic geographies: citation patterns and citation behaviour in economic geography, 1982–2006. *Transactions of the Institute of British Geographers*, 32(3), 295–312.
- Fredriksson, E. H. (2001). The Dutch Publishing Scene: Elsevier and North-Holland. In *Idem (red.), A century of science publishing. A collection of essays* (pp. 61–76).
- Gardner, S., & Eng, S. (2005). Gaga over Google? Scholar in the social sciences. *Library Hi Tech News*, 22(8), 42–45.
- Garfield, E. (2006). The history and meaning of the journal impact factor. *Jama*, 295(1), 90–93.
- Garfield, E., & Welljams-Dorof, A. (1992). Of Nobel class: A citation perspective on high impact research authors. *Theoretical Medicine and Bioethics*, 13(2), 117–135.

- Gavel, Y., & Iselid, L. (2008). Web of Science and Scopus: a journal title overlap study. *Online Information Review*, 32(1), 8–21.
- Gluck, M. (1990). A review of journal coverage overlap with an extension to the definition of overlap. *Journal of the Association for Information Science and Technology*, 41(1), 43–60.
- Gonzalez-Brambila, C. N., Reyes-Gonzalez, L., Veloso, F., & Perez-Angón, M. A. (2016). The scientific impact of developing nations. *PloS One*, 11(3), e0151328.
- Gorraiz, J., Gumpenberger, C., & Glade, T. (2016). On the bibliometric coordinates of four different research fields in Geography. *Scientometrics*, 107(2), 873–897. <https://doi.org/10.1007/s11192-016-1864-y>
- Gorraiz, J., Gumpenberger, C., & Schlögl, C. (2014). Usage versus citation behaviours in four subject areas. *Scientometrics*, 101(2), 1077–1095. <https://doi.org/10.1007/s11192-014-1271-1>
- Grigas, V., Juzeniene, S., & Velickaite, J. (2017). “Just Google It”—The Scope of Freely Available Information Sources for Doctoral Thesis Writing. *Information Research: An International Electronic Journal*, 22(1), n1.
- Gross, P. L. K., & Gross, E. M. (1927). College libraries and chemical education. *Science*, 66(1713), 385–389.
- Gross, P. L. K., & Woodford, A. O. (1931). Serial literature used by American geologists. *Science*, 73(1903), 660–664.
- Guédon, J.-C. (2001). *In Oldenburg's long shadow: Librarians, research scientists, publishers, and the control of scientific publishing*. Association of Research Libr.
- Guerrero-Bote, V. P., & Moya-Anegón, F. (2014). Relationship between downloads and citations at journal and paper levels, and the influence of language. *Scientometrics*, 101(2), 1043–1065. <https://doi.org/10.1007/s11192-014-1243-5>
- Gullbekk, E., Rullestad, T., & Calvo, M. C. T. (2013). PhD candidates and the research process. *The Library's Contribution. University of Oslo, Oslo*.
- Haddaway, N. R., Collins, A. M., Coughlin, D., & Kirk, S. (2015). The role of Google Scholar in evidence reviews and its applicability to grey literature searching. *PloS One*, 10(9), e0138237.
- Halleux, J.-M. (2013). La structuration du domaine scientifique du développement territorial et l'importance à accorder aux revues internationales. *Géo-Regards: Revue Neuchâteloise de Géographie*, (6), 151–159.
- Hallmark, J. (1994). Scientists' access and retrieval of references cited in their recent journal articles. *College & Research Libraries*, 55(3), 199–209.
- Harnad, S. (1996). Implementing peer review on the net: scientific quality control in scholarly electronic journals. In *Scholarly publishing: the electronics frontier*. MIT Press, Cambridge, MA.
- Harnad, S., & Brody, T. (2004). Comparing the impact of open access (OA) vs. non-OA articles

in the same journals. *D-Lib Magazine*, 10(6).

- Harnad, S., Brody, T., ValliÃ“res, F. § ois, Carr, L., Hitchcock, S., Gingras, Y., ... Hilf, E. R. (2004). The access/impact problem and the green and gold roads to open access. *Serials Review*, 30(4), 310–314.
- Harris, R., & Jarvis, C. (2014). *Statistics for geography and environmental science*. Routledge.
- Harzing, A.-W. (2014). A longitudinal study of Google Scholar coverage between 2012 and 2013. *Scientometrics*, 98(1), 565–575.
- Harzing, A.-W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics*, 106(2), 787–804.
- Hicks, D., & Wang, J. (2011). Coverage and overlap of the new social sciences and humanities journal lists. *Journal of the Association for Information Science and Technology*, 62(2), 284–294.
- Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569.
- Hood, W. W., & Wilson, C. S. (2003). Overlap in bibliographic databases. *Journal of the Association for Information Science and Technology*, 54(12), 1091–1103.
- Houssay-Holzschuch, M., & Milhaud, O. (2013). Geography after Babel-a view from the French province. *Geographica Helvetica*, 68, 51–55.
- Iivonen, M., Nygrén, U., Valtari, A., & Heikkilä, T. (2009). Library collections contribute to doctoral studies: Citation analysis of dissertations in the field of economics and administration. *Library Management*, 30(3), 185–203.
- Ikeda, D., & Inoue, S. (2009). Access flows to a repository from other services. In *4th International Conference on Open Repositories*. Georgia Institute of Technology. Retrieved from <http://hdl.handle.net/1853/28422>
- Inger, S., & Gardner, T. (2016). How readers discover content in scholarly publications. *Information Services & Use*, 36(1–2), 81–97.
- Jacsó, P. (2005). Google Scholar: the pros and the cons. *Online Information Review*, 29(2), 208–214.
- Jamali, H. R., & Asadi, S. (2010). Google and the scholar: the role of Google in scientists' information-seeking behaviour. *Online Information Review*, 34(2), 282–294.
- Jamali, H. R., & Nabavi, M. (2015). Open access and sources of full-text articles in Google Scholar in different subject fields. *Scientometrics*, 105(3), 1635–1651.
- Jamali, H. R., & Nikzad, M. (2011). Article title type and its relation with the number of downloads and citations. *Scientometrics*, 88(2), 653–661.
- Johnston, R. (2003). *Geography: a different sort of discipline? Transactions of the Institute of British Geographers* (Vol. 28). Wiley Online Library.
- Johnston, R. (2005). Geography–coming apart at the seams? In *Questioning Geography*:

- Fundamental Debates* (pp. 9–25). Wiley-Blackwell.
- Johnston, R., & Sidaway, J. D. (2015). *Geography and geographers: Anglo-American human geography since 1945*. Routledge.
- Jones, D. Y. (2005). Biology article retrieval from various databases: making good choices with limited resources. *Issues in Science and Technology Librarianship*, 44.
- Jones, R. E., Andrew, T., & MacColl, J. (2006). *The institutional repository*. Elsevier.
- Kaczor, S. (2014). A citation analysis of doctoral dissertations in atmospheric science at the University at Albany. *Science & Technology Libraries*, 33(1), 89–98.
- Kayongo, J., & Helm, C. (2012). Relevance of library collections for graduate student research: A citation analysis study of doctoral dissertations at Notre Dame. *College & Research Libraries*, 73(1), 47–67.
- Kimball, R., Stephens, J., Hubbard, D., & Pickett, C. (2013). A citation analysis of atmospheric science publications by faculty at Texas A&M University. *College & Research Libraries*, 74(4), 356–367.
- Kriebel, L., & Lapham, L. (2008). Transition to electronic resources in undergraduate social science research: A study of honors theses bibliographies, 1999–2005. *College & Research Libraries*, 69(3), 268–284.
- Kriz, H. M. (1978). Subscriptions vs. Books in a Constant Dollar Budget. *College and Research Libraries*, 39(2), 105–109.
- Kudelka, M., Plato, J., & Krömer, P. (2016). Author Evaluation Based on H-index and Citation Response. In *Intelligent Networking and Collaborative Systems (INCoS), 2016 International Conference on* (pp. 375–379). IEEE.
- Kulkarni, A. V., Aziz, B., Shams, I., & Busse, J. W. (2009). Comparisons of citations in web of science, Scopus, and Google Scholar for articles published in general medical journals. *JAMA - Journal of the American Medical Association*, 302(10), 1092–1096. <https://doi.org/10.1001/jama.2009.1307>
- Kumar, H. A., & Dora, M. (2011). Citation analysis of doctoral dissertations at IIMA: A review of the local use of journals. *Library Collections, Acquisitions, and Technical Services*, 35(1), 32–39.
- Kushkowski, J. D., Parsons, K. A., & Wiese, W. H. (2003). Master's and doctoral thesis citations: Analysis and trends of a longitudinal study. *Portal: Libraries and the Academy*, 3(3), 459–479.
- Kutner, M. H., Nachtsheim, C., Neter, J., & Li, W. (2005a). *Applied linear statistical models* (Fifth Edit). McGraw-Hill Irwin.
- Kutner, M. H., Nachtsheim, C., Neter, J., & Li, W. (2005b). Building the regression model I: Model selection and validation. In *Applied linear statistical models* (Fifth Edit). New-York: McGraw-Hill Irwin.
- Kutner, M. H., Nachtsheim, C., Neter, J., & Li, W. (2005c). Building the Regression Model II: Diagnostics. In *Applied linear statistical models* (Fifth Edit). New-York: McGraw-Hill

Irwin.

Le Moniteur Belge - Belgisch Staatsblad. (n.d.).

Levay, P., Ainsworth, N., Kettle, R., & Morgan, A. (2016). Identifying evidence for public health guidance: a comparison of citation searching with Web of Science and Google Scholar. *Research Synthesis Methods*, 7(1), 34–45.

Livingstone, D. N. (2010). *Putting Science in Its Place: Geographies of Scientific Knowledge*. Chicago: University of Chicago Press.

Löhönen, J., Isohanni, M., Nieminen, P., & Miettunen, J. (2010). Coverage of the bibliographic databases in mental health research. *Nordic Journal of Psychiatry*, 64(3), 181–188.

Mabe, M., & Amin, M. (2001). Growth dynamics of scholarly and scientific journals. *Scientometrics*, 51(1), 147–162.

Martín-Martín, A., Ayllón, J. M., Orduña-Malea, E., & López-Cózar, E. D. (2016). 2016 Google Scholar Metrics released: a matter of languages... and something else. *ArXiv Preprint ArXiv:1607.06260*.

Martín-Martín, A., Orduna-Malea, E., Ayllón, J. M., & López-Cózar, E. D. (2016). A two-sided academic landscape: portrait of highly-cited documents in Google Scholar (1950-2013). *ArXiv Preprint ArXiv:1607.02861*.

Martín-Martín, A., Orduña-Malea, E., Ayllón, J. M., & López-Cózar, E. D. (2014). Does Google Scholar contain all highly cited documents (1950-2013)? *ArXiv Preprint ArXiv:1410.8464*.

Martín-Martín, A., Orduña-Malea, E., Ayllón, J. M., & López-Cózar, E. D. (2016). The counting house: measuring those who count. Presence of Bibliometrics, Scientometrics, Informetrics, Webometrics and Altmetrics in the Google Scholar Citations, ResearcherID, ResearchGate, Mendeley & Twitter. *ArXiv Preprint ArXiv:1602.02412*.

Martin, B. (1996). The use of multiple indicators in the assessment of basic research. *Scientometrics*, 36(3), 343–362.

McCain, K. W., & Bobick, J. E. (1981). Patterns of journal use in a departmental library: A citation analysis. *Journal of the Association for Information Science and Technology*, 32(4), 257–267.

McCullough, I. (2016). Journal usage at department and research group level. *Journal of Electronic Resources Librarianship*, 28(2), 73–83.

Meadows, A. J. (1998). *Communicating research*. San Diego: Academic Press.

Meadows, J. (2000). The growth of journal literature: A historical perspective. In H. Barsky Atkins & B. Cronin (Eds.), *The web of knowledge: A festschrift in honor of Eugene Garfield* (pp. 87–109). Information Today Medford, NJ.

Meho, L. I., & Yang, K. (2007). Impact of data sources on citation counts and rankings of LIS faculty: Web of Science versus Scopus and Google Scholar. *Journal of the Association for Information Science and Technology*, 58(13), 2105–2125.

Meier, J. J., & Conkling, T. W. (2008). Google Scholar's coverage of the engineering literature: 212

- an empirical study. *The Journal of Academic Librarianship*, 34(3), 196–201.
- Merenne-Schoumaker, B. (2004). 1903-2003. Cent ans de géographie à l’Université de Liège: changements et permanences. *Bulletin de La Société Géographique de Liège*, (44), 125–127.
- Michel, J. (1997). Des bases de données bibliographiques aux bases multimédias. In *L’information scientifique et technique: Nouveaux enjeux documentaires et éditoriaux. Tours (France), 21-23 octobre 1996* (p. 101). Editions Quae.
- Miguel, S., Chinchilla-Rodriguez, Z., & de Moya-Anegón, F. (2011). Open access and Scopus: A new approach to scientific visibility from the standpoint of access. *Journal of the Association for Information Science and Technology*, 62(6), 1130–1145.
- Mikki, S. (2010). Comparing Google Scholar and ISI Web of Science for earth sciences. *Scientometrics*, 82(2), 321–331.
- Mikki, S., Ruwehy, H. A. Al, Gjesdal, Ø. L., & Zygmuntowska, M. (2018). Filter bubbles in interdisciplinary research: a case study on climate and society. *Library Hi Tech*, 36(2), 225–236.
- Mingers, J., & Lipitakis, E. A. (2010). Counting the citations: a comparison of Web of Science and Google Scholar in the field of business and management. *Scientometrics*, 85(2), 613–625.
- Moed, H. F. (2005). Statistical relationships between downloads and citations at the level of individual documents within a single journal. *Journal of the Association for Information Science and Technology*, 56(10), 1088–1097.
- Moed, H. F. (2007). The effect of “open access” on citation impact: An analysis of ArXiv’s condensed matter section. *Journal of the Association for Information Science and Technology*, 58(13), 2047–2054.
- Moed, H. F., & Halevi, G. (2015). Multidimensional assessment of scholarly research impact. *Journal of the Association for Information Science and Technology*, 66(10), 1988–2002.
- Moed, H., Van Leeuwen, T. H., & Reedijk, J. (1999). Towards appropriate indicators of journal impact. *Scientometrics*, 46(3), 575–589.
- Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, 106(1), 213–228. <https://doi.org/10.1007/s11192-015-1765-5>
- Moya-Anegón, F., & Herrero-Solana, V. (2013). Worldwide topology of the scientific subject profile: A macro approach in the country level. *PloS One*, 8(12), e83222.
- Musser, L. R. (2008). Comparison of GeoRef and Google Scholar. In *43rd meeting of the Geoscience Information Society, symposium on Libraries in transformation; exploring topics of changing practices and new technologies* (p. 77). Houston, TX, United States.
- Myriam Baron, Denis Eckert, L. J. (2014). Un « tout petit » monde scientifique. In F. G. Grataloup Christian (Ed.), *L’atlas global, 60 cartes inédites : un autre monde émerge sous nos yeux* (p. 71).

- Nabe, J., & Imre, A. (2008). Dissertation citations in Organismal Biology at Southern Illinois University at Carbondale: implications for collection development. *Issues in Science & Technology Librarianship*, 55.
- Neuhaus, C., Neuhaus, E., Asher, A., & Wrede, C. (2006). The depth and breadth of Google Scholar: An empirical study. *Portal: Libraries and the Academy*, 6(2), 127–141.
- Nicholas, D., Huntington, P., & Jamali, H. R. (2008). User diversity: as demonstrated by deep log analysis. *The Electronic Library*, 26(1), 21–38.
- Nicolaï, H. (2004). Geography in Belgium. *Belgeo. Revue Belge de Géographie*, (1), 33–44.
- Nisonger, T. E. (2008). The “80/20 rule” and core journals. *The Serials Librarian*, 55(1–2), 62–84.
- Noga, M. M., Derksen, C. R. M., & Haner, B. (1993). Characteristics of geoscience serial use by faculty and students. In *Finding and Communicating Geoscience Information: Proceedings of the 28th Meeting* (Vol. 24, pp. 61–97).
- Norris, M. (2008). *The citation advantage of open-access articles*. Loughborough University. Retrieved from <https://dspace.lboro.ac.uk/2134/4089>
- Norris, M., Oppenheim, C., & Rowland, F. (2008). The citation advantage of open-access articles. *Journal of the American Society for Information Science and Technology*, 59, 1963–1972.
- Notess, G. R. (2005). Scholarly web searching: Google Scholar and Scirus. *Online Information Review*, 29, 39–41.
- OARE. (n.d.). Retrieved July 26, 2017, from <https://www.unep.org/oare/>
- Ophof, T. (1997). Sense and nonsense about the impact factor. *Cardiovascular Research*, 33(1), 1–7.
- Organ, M. K. (2006). Download statistics—What do they tell us? The example of Research Online, the open access institutional repository at the University of Wollongong, Australia. *D-Lib Magazine*, 12(11). <https://doi.org/http://dx.doi.org/10.1045/november2006-organ>
- Ortega, J. L. (2014). *Academic search engines: a quantitative outlook*. Elsevier.
- Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., ... Haustein, S. (2018). The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ*, 6, e4375.
- Qing, F., Lifang, X., & Xiaochuan, L. (2008). Peer-review practice and research for academic journals in China. *Journal of Scholarly Publishing*, 39(4), 417–427.
- Robinson, M. (2009). Promoting the visibility of educational research through an institutional repository. *Serials Review*, 35(3), 133–137.
- Robinson, M. L., & Wusteman, J. (2007). Putting Google Scholar to the test: A preliminary study. *Program: Electronic Library and Information Systems*, 41(1), 71–80.
- Romić, K., & Mitrović, G. (2014). Using citation checking of Ph. D. dissertation references as a tool for evaluating library collections of the National and University Library in Zagreb.

Libraries in the Digital Age (LIDA) Proceedings, 13.

- Rosenberg, Z. (2015). Citation analysis of MA theses and Ph. D. dissertations in sociology and anthropology: An assessment of library resource usage. *The Journal of Academic Librarianship*, 41(5), 680–688. <https://doi.org/http://dx.doi.org/10.1016/j.acalib.2015.05.010>
- Rossi, U. (2008). Being here and there: in-betweenness, double absence, and the making of a multi-layered academic citizenship. *Area*, 40(3), 401–406.
- Rowland, F. (2002). The peer-review process. *Learned Publishing*, 15(4), 247–258.
- Rowland, F. (2005). Journal access programmes for developing countries. *Serials*, 18(2), 104–106.
- Rowlands, I., Nicholas, D., Williams, P., Huntington, P., Fieldhouse, M., Gunter, B., ... Tenopir, C. (2008). The Google generation: the information behaviour of the researcher of the future. In *Aslib proceedings* (Vol. 60, pp. 290–310). Emerald Group Publishing Limited.
- S. Adriaanse, L., & Rensleigh, C. (2013). Web of Science, Scopus and Google Scholar: A content comprehensiveness comparison. *The Electronic Library*, 31(6), 727–744.
- Salager-Meyer, F. (2008). Scientific publishing in developing countries: Challenges for the future. *Journal of English for Academic Purposes*, 7(2), 121–132.
- Sarkowski, H. (2001). The growth and decline of German scientific publishing 1850–1945. In *A Century of Science Publishing: a collection of essays* (pp. 25–34). IOS Press.
- Schloegl, C., & Gorraiz, J. (2010). Comparison of citation and usage indicators: the case of oncology journals. *Scientometrics*, 82(3), 567–580.
- Schloegl, C., & Gorraiz, J. (2011). Global usage versus global citation metrics: The case of pharmacology journals. *Journal of the Association for Information Science and Technology*, 62(1), 161–170.
- Schmitz, S. (2003). La géographie humaine et ses revues «internationales»: globalisation ou fragmentation?/ Human geography and its “international” journals: globalisation or fragmentation? *Annales de Géographie*, 112(632), 402–411.
- Schroter, S., & Tite, L. (2006). Open access publishing and author-pays business models: a survey of authors' knowledge and perceptions. *Journal of the Royal Society of Medicine*, 99(3), 141–148.
- Schuermans, N., Meeus, B., & De Maesschalck, F. (2010). Is there a world beyond the Web of Science? Publication practices outside the heartland of academic geography. *Area*, 42(4), 417–424.
- Sexton, E. (2006). Journal Use by Graduate Students as Indicated by Master's Theses Bibliographies at an Urban Commuter College, 1991–2004. *Behavioral & Social Sciences Librarian*, 24(2), 93–111.
- Shepherd, P. T. (2007). The feasibility of developing and implementing journal usage factors: a research project sponsored by UKSG. *Serials*, 20(2).

- Shepherd, P. T. (2011). PIRUS2: individual article usage statistics—developing a practical, global standard. *Learned Publishing*, 24(4), 279–286.
- Shultz, M. (2007). Comparing test searches in PubMed and Google Scholar. *Journal of the Medical Library Association: JMLA*, 95(4), 442.
- Sous la direction de : Frédérique Sachwald. (2018). *La position scientifique de la France dans le monde 2000-2015*. Paris.
- Ştirbu, S., Thirion, P., Schmitz, S., Haesbroeck, G., & Greco, N. (2015). The utility of Google Scholar when searching geographical literature: comparison with three commercial bibliographic databases. *The Journal of Academic Librarianship*, 41(3), 322–329.
- Subotic, S., & Mukherjee, B. (2014). Short and amusing: The relationship between title characteristics, downloads, and citations in psychology articles. *Journal of Information Science*, 40(1), 115–124.
- Swan, A., & Brown, S. (2005). *Open access self-archiving: An author study*. UK FE and HE funding councils.
- Sylvia, M. J. (1998). Citation analysis as an unobtrusive method for journal collection evaluation using psychology student research bibliographies. *Collection Building*, 17(1), 20–28.
- Sylvia, M., & Lesher, M. (1995). What journals do psychology graduate students need? A citation analysis of thesis references. *College & Research Libraries*, 56(4), 313–318.
- Tenopir, C., & King, D. W. (2000). *Towards electronic journals: realities for scientists, librarians, and publishers*. Special Libraries Association Washington, DC.
- Thirion, P. (2016). L'open access, au service de la visibilité de la recherche et des chercheurs? L'expérience d'ORBi. In *Open-access à l'université Paris Nanterre : Inauguration du portail HAL*.
- Trueswell, R. L. (1969). Some behavioral patterns of library users: The 80/20 rule. *Wilson Libr Bull*.
- Van Noorden, R. (2014). Scientists and the social network. *Nature*, 512(7513), 126.
- Vandermotten, C. (2008). L'histoire de la géographie belge à travers les biographies nationales. *Belgeo*, 1. Retrieved from <http://journals.openedition.org/belgeo/10209>
- Vandermotten, C., & Kesteloot, C. (2013). Belgeo et les quatre crises de la géographie. *Belgeo*, 2012(2), 2–9.
- Vezzosi, M. (2009). Doctoral students' information behaviour: an exploratory study at the University of Parma (Italy). *New Library World*, 110(1/2), 65–80.
- Vial, S. (2013). *L'être et l'écran: comment le numérique change la perception*. Presses Universitaires de France.
- Vieira, E. S., & Gomes, J. A. N. F. (2009). A comparison of Scopus and Web of Science for a typical university. *Scientometrics*, 81(2), 587.
- Walcott, R. (1992). Characteristics of citations in geoscience doctoral dissertations accepted at

- United States academic institutions 1981-1985. *Science & Technology Libraries*, 12(2), 5–16.
- Walters, W. H. (2007). Google Scholar coverage of a multidisciplinary field. *Information Processing & Management*, 43(4), 1121–1132.
- Walters, W. H. (2011). Comparative recall and precision of simple and expert searches in Google Scholar and eight other databases. *Portal: Libraries and the Academy*, 11(4), 971–1006.
- Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, 10(2), 365–391.
- Wang, X., Wang, Z., Mao, W., & Liu, C. (2014). How far does scientific community look back? *Journal of Informetrics*, 8(3), 562–568.
- Wang, X., Wang, Z., & Xu, S. (2013). Tracing scientist's research trends realtimely. *Scientometrics*, 95(2), 717–729.
- Waugh, C. K., & Ruppel, M. (2004). Citation analysis of dissertation, thesis, and research paper references in workforce education and development. *The Journal of Academic Librarianship*, 30(4), 276–284.
<https://doi.org/http://dx.doi.org/10.1016/j.acalib.2004.04.003>
- Weller, A. C. (2001). *Editorial peer review: Its strengths and weaknesses*. Information Today, Inc.
- Willinsky, J. (2006). *The access principle: The case for open access to research and scholarship*. Cambridge, Mass.: MIT Press.
- Woodford, A. O. (1969). Serial literature used by American geologists, 1967. *Journal of Geological Education*, 17(3), 87–90.
- Zipp, L. S. (1995). Identifying core geologic research journals: A model for interlibrary cooperative collection development. In *Proceedings of the Twenty-Ninth Meeting of the Geoscience Information Society* (Vol. 25, pp. 59–65).
- Zong, Q.-J., Shen, H.-Z., Yuan, Q.-J., Hu, X.-W., Hou, Z.-P., & Deng, S.-G. (2013). Doctoral dissertations of Library and Information Science in China: A co-word analysis. *Scientometrics*, 94(2), 781–799.
- Zuccala, A. (2010). Open access and civic scientific information literacy. *Information Research: An International Electronic Journal*, 15(1).

Table des matières

RESUME.....	i
ABSTRACT	iii
REMERCIEMENTS	v
SOMMAIRE	vii
TABLE DE FIGURES, GRAPHIQUES ET TABLEAUX	ix
LISTE DES ABREVIATIONS	xvi
INTRODUCTION GENERALE.....	1
1. Contexte général de la recherche	3
2. Principaux axes de notre recherche	5
3. Les périodiques scientifiques	7
3.1. Introduction et généralités	7
3.2. La commercialisation de la publication des périodiques scientifiques	8
3.3. Le développement du support informatique.....	8
3.4. Les bases de données bibliographiques.....	9
3.5. Le <i>peer-reviewing</i> , les indicateurs d'usage et l'évaluation des chercheurs	10
4. L'Open Access	12
4.1. Les différents modèles de l' <i>Open Access</i> et les différentes initiatives	14
4.1.1. Systèmes d'archivage de la Voie Verte.....	16
4.1.2. ORBi.....	17
4.1.3. Nouveaux indicateurs	19
5. La publication des périodiques en Géographie	19
5.1. L'historique de la publication en Géographie	20
5.2. La langue de publication	23
5.3. Le Département de Géographie de l'ULiège	24
6. Présentation des outils bibliographiques	25
6.1. Web of Science.....	28
6.2. Scopus	29
6.3. Google Scholar	30
6.3.1. Les fonctionnalités de Google Scholar.....	32
6.4. GeoRef	33
6.5. FRANCIS	33
6.6. Ulrichsweb Global Serials Directory	34
7. La structure de la dissertation et le contenu des chapitres	35
PREMIERE PARTIE La littérature répertoriée par les bases de données commerciales et par le moteur de recherche Google Scholar en Géographie	39
Chapitre 1 The Utility of Google Scholar When Searching Geographical Literature: Comparison with Three Commercial Bibliographic Databases	41
1. Introduction	41

2. Aim and Objectives	42
3. Methods	43
3.1. Bibliographic tools	43
3.2. Bibliographic search.....	44
3.3. Analysis of the number of results.....	45
3.4. Reference analysis—a case study.....	45
3.5. Overlap and specificity.....	46
4. Results	47
4.1. Analysis of the number of results.....	47
4.1.1. Time evolution of search results.....	47
4.1.2. Performance of the search tools	48
4.1.3. Case studies: “urbanization” and “sedimentation”	50
5. Discussion and Conclusion	53
Chapitre 2 - Caractéristiques des périodiques associés aux références uniquement fournies soit par GS, WoS ou GeoRef	56
1. Introduction et revue de la littérature	57
2. Objectif et hypotheses	59
3. Méthodes	59
4. Résultats	60
4.1. Caractéristiques des périodiques associés aux articles « uniques » pour « urbanization »	62
4.1.1. Critères de « qualité » et mention OA	62
4.1.2. Langue et pays de publication	63
4.2. Caractéristiques de périodiques associées aux articles retrouvés avec « sedimentation »	68
4.2.1. Critères de « qualité » et mention OA	68
4.2.2. Langue et pays de publication	69
5. Discussion et conclusions.....	73
Chapitre 3 - Usage de la littérature scientifique dans un échantillon de trois dissertations doctorales	76
1. Introduction et revue de la littérature	77
1.1. Etudier les citations des chercheurs pour orienter les choix d'acquisitions en bibliothèque.....	77
1.2. L'impact de l'émergence de la publication électronique	79
1.3. La part de Google Scholar.....	80
1.4. Besoins documentaires et citations en géosciences.....	81
2. Objectifs et hypothèses.....	82
3. Méthodes	83

4. Résultats	85
4.1. Type de documents cités	85
4.1.1. Citation d'articles de périodiques.....	85
4.1.1.1. Périodiques cités et la loi de Bradford.....	86
4.1.1.2. Années de publication des articles cités	88
4.1.2. Citation des livres et chapitres de livres	89
4.1.3. Autres type de documents cités	90
4.2. Langue des citations	90
4.3. Référencement des citations dans les outils bibliographiques	91
4.4. Analyse « overlap » entre l'ensemble des outils bibliographiques et apports spécifiques en références « uniques »	94
5. Discussion et conclusions.....	97
CONCLUSIONS GENERALES DE LA PREMIERE PARTIE	101
DEUXIEME PARTIE La littérature scientifique produite par les géographes de l'ULiège : diffusion et usage	105
1. Introduction	107
Chapitre 1 - Diffusion et visibilité des publications en géographie de l'ULiège dans les bases de données commerciales et GS	107
1. Introduction et revue de la littérature	109
2. Objectifs et hypothèses.....	112
3. Méthodes	113
4. Résultats	114
4.1. Description de l'échantillon de publications	114
4.1.1. Les chapitres et parties de livres.....	114
4.1.2. Les articles de périodiques et les communications.....	116
4.2. Recensement des articles et des communications par les outils bibliographiques	117
4.2.1. Qualification et recensement des périodiques	118
4.2.2. Illustration du recouvrement entre les différents outils bibliographiques	123
4.2.2.1. Caractéristiques des références et périodiques uniquement répertoriés par GS .	124
5. Discussion et Conclusion	126
Chapitre 2 - Téléchargements des articles en géographie/géographiques déposés dans ORBi	131
1. Introduction	131
1.1. Evaluation de l'impact d'une publication : citations vs téléchargements	132
1.2. Téléchargements à partir de répertoires institutionnels.....	133
2. Objectifs et hypothèses.....	134
3. Méthodes	134

3.1. Distribution des téléchargements au niveau mondial.....	135
3.2. Facteurs susceptibles d'influencer les téléchargements	135
3.2.1. Régression multiple.....	136
4. Résultats	136
4.1. Distribution des téléchargements au niveau mondial.....	137
4.1.1. Distribution des téléchargements des articles en anglais.....	143
4.1.2. Distribution des téléchargements des articles en français	146
4.1.3. Distribution des téléchargements correspondant aux références retrouvés par GS uniquement	149
4.1.4. Distribution des téléchargements correspondant aux références retrouvés par GS via les liens ORBi uniquement.....	151
4.1.5. Distribution des téléchargements correspondant aux références retrouvés par tous les outils bibliographiques	153
4.2. Facteurs susceptibles d'influencer les téléchargements de la littérature en géographie	155
4.2.1. Description des variables.....	155
4.2.1.1. Etude des valeurs manquantes.....	156
4.2.1.2. Les variables dépendantes	158
4.2.1.3. Les variables explicatives	161
4.2.1.4. Analyse de corrélation.....	164
4.3. Les facteurs qui influencent les téléchargements de l'ensemble de l'échantillon de publications déposées dans ORBi par les géographes: développement du modèle retenu et résultats.....	164
4.3.1.1. Analyse des résidus du modèle retenu.....	168
4.3.1.2. Vérification de la normalité des résidus	168
4.3.1.3. Linéarité de la fonction de régression et homoscédasticité	169
4.3.1.4. Test de multicolinéarité	170
4.3.1.5. Diagnostiques d'influence - distance de Cook :	171
4.3.1.6. Indépendance des erreurs	173
4.3.2. Les facteurs qui influencent les téléchargements des publications en langue anglaise de l'échantillon : développement du modèle retenu et résultats.....	175
4.3.2.1. Validation du modèle retenu.....	178
4.3.2.2. Vérification de la normalité des résidus	179
4.3.2.3. Diagnostiques d'influence - distance de Cook :	179
4.3.2.4. Indépendance des erreurs	181
4.3.3. Les facteurs qui influencent les téléchargements des publications en langue française de l'échantillon : développement du modèle retenu et résultats.....	182
4.3.3.1. Validation du modèle retenu.....	186
4.3.3.2. Vérification de la normalité des résidus	186

4.3.3.3. Diagnostiques d'influence - distance de Cook :	186
4.3.3.4. Indépendance des erreurs	188
5. Discussion et Conclusion	191
CONCLUSIONS GENERALES DE LA DEUXIEME PARTIE	198
CONCLUSIONS GENERALES.....	200
PERSPECTIVES.....	203
BIBLIOGRAPHIE	205
ANNEXES	223

ANNEXES

ANNEXE 1 - Chapitre 3 Partie 1 : Questionnaire soumis aux doctorants

- 1) Quelle stratégie de recherche bibliographique avez-vous appliquée dans le cadre de votre recherche doctorale ?
- 2) Quelles sont les sources bibliographiques que vous avez utilisées (bases de données bibliographiques, moteurs de recherche, catalogue de la bibliothèque) ?
- 3) Est-ce que vous privilégiez l'un ou l'autre de ces outils de recherche bibliographique ? Si oui pour quelle raison ?
- 4) Dans le cadre de vos recherches documentaires êtes-vous un utilisateur de la bibliothèque et de ses ressources imprimées, livres, cartes, autres documents ?
- 5) Avez-vous obtenu des documents ou informations bibliographiques via votre réseau de connaissances (promoteurs, collègues de service et collaborateurs étrangers) ???
- 6) Comment avez-vous procédé pour consulter les documents auxquels vous n'avez pas pu accéder en électronique ou papier (p.ex livres) ?
- 7) Nous avons remarqué une majorité d'articles scientifiques cités dans vos bibliographies, y a-t-il une raison particulière ou c'est dû aux pratiques de recherche et citation actuelles ?

ANNEXE 2 - Chapitre 1 Partie 2

Titres des périodiques avec ISSN non trouvé dans <i>Ulrichsweb</i>	ISSN
<i>Annales de l'Université Abdou Moumouni de Niamey, Série B Lettres et Sciences Humaines</i>	1859-5014
<i>Bulletin de la Société Royale Belge d'Etudes Géologiques et Archéologiques</i>	0776-0086
<i>Bulletin de liaison des membres de la Société de Géographie de Paris</i>	1964-8995
<i>Bulletin des Séances de l'Académie Royale des Sciences d'Outre-Mer</i> = <i>Mededelingen der Zittingen van de Koninklijke Academie voor Overzeese Wetenschappen</i>	0001-4176
<i>Cahiers de l'Urbanisme (Les)</i>	0775-5317
<i>Cahiers Nouveaux (Les)</i>	0775-5317
<i>De la Meuse à l'Ardenne</i>	0773-8781
<i>G.E.O</i>	1780-5945
<i>Géographes Associés, AFDG</i>	1266-4618
<i>GeoPlatform</i>	1872-5953
<i>GEORAMA Revue trimestrielle pour le géomètre-expert</i>	1372-6285
<i>Géo-Regards : Revue Neuchâteloise de Géographie</i>	1662-8527
<i>Hautes Fagnes</i>	0770-2507
<i>Horizons in geography</i>	0034-3774
<i>La Lettre « Changement Global » PIGB - PMRC France</i>	1261- 4246
<i>Les Cahiers du CEPRESS</i>	0007-9642
<i>Mutations. Mémoires et perspectives du Bassin minier</i>	2078-7634
<i>Point sur la Recherche en Education (Le)</i>	1376-8425
<i>Politique : Revue de Débats</i>	1372-908X
<i>Publications de l'Association Internationale de Climatologie</i>	1140-0307
<i>Revue d'écologie - environnement</i>	1112-5888
<i>Sciences Connection</i>	1780-8456
<i>Territoire(s) Wallon(s)</i>	1784-4991
<i>Territoires en mouvement</i>	1954-4867
<i>Urbia</i>	1661-3708

ANNEXE 3 - Chapitre 1 Partie 2 : Titres non retrouvés dans Ulrichsweb

A+

Cahiers Pédagogiques

Mosella

XYZ : la Revue de l'Association Française de Topographie

De la Meuse à l'Ardenne

Cahiers de l'Urbanisme (Les)

Bulletin de la Société Royale Belge d'Etudes Géologiques et Archéologiques "Les Chercheurs de Wallonie"

La tribune de l'eau

La Lettre « Changement Global » PIGB - PMRC France

Géographes associés, AFDG , Université Lyon II

Education relative à l'environnement - Regards, recherches, réflexions

Ruimte & Planning

Research in Geographic Education

Urbia

Géo-Regards : Revue Neuchâteloise de Géographie

Territoire(s) Wallon(s)

Annales de l'Université Abdou Moumouni de Niamey, Série B Lettres et Sciences Humaines

Bulletin du Cercle historique de Durbuy

Cahiers du Centre Européen de Géodynamique et de Séismologie

TRACeS de ChanGements

Historiens et Geographes

Revue Internationale de Géomatique

Les Cahiers du CEPESS

Distribution d'Aujourd'hui

Abstracts with Programs - Geological Society of America

Journal des Ingénieurs (Le)

Travaux de l'Institut de Géographie de Reims

Acta Geographica Lovaniensia

Revista de Geografia

Revue du CEBEDEAU

Education-Formation

GeolNova

Revue d'écologie - environnement

Revue trimestrielle pour le géomètre-expert

Politique : Revue de Débats

Point sur la Recherche en Education (Le)

GEO

GeoPlatform

Bulletin de liaison des membres de la Société de Géographie de Paris

Mutations. Mémoires et perspectives du Bassin minier

Actes du Geoforum

AGU Abstract

AM/FM-GIS BELUX News

La Cemathèque

Landmeter - Géomètre

Les Hors-série de Politique

Papers on Tools and methods of Territorial Intelligence

Proceedings of the AAG Annual Meeting

Puzzle
