SPATIAL AGGREGATION OF LOW RESOLUTION SATELLITE DATA FOR THE MONITORING OF VEGETATION RESPONSE TO CLIMATIC STRESSES: ANALYSIS OF THE SPATIAL HETEROGENEITY OF AGGREGATED ENTITIES

Horion S.(1), Eerens H.(2), Tychon B. (3) and Cornet Y.(1)

(1) Unit of Geomatics, Department of Geography, University of Liege
17 Allée du 6-Août, 4000 Liège Belgium, shorion@ulg.ac.be and ycornet@ulg.ac.be
(2) Vlaamse Instelling voor Technologisch Onderzoek (VITO),
200 Boeretang, 2400 Mol Belgium, herman.eerens@vito.be
(3) Dept. Environmental Sciences and Management, University of Liege
185 Avenue de Longwy, 6700 Arlon Belgium, bernard.tychon@ulg.ac.be

ABSTRACT

Our PhD research consists in analysing and modelling the vegetation response or sensitivity to climatic stresses with low satellite imagery. In that framework, the selection of optimal calibration sites is very important. These sites should be characterised by a stable and homogenous land cover over large area. Here we analyse the spatial heterogeneity of the aggregation entities (EU-NUTS 2) used by the MARS-FOOD programme for the extraction of regional NDVI-means.

1. INTRODUCTION

The spatial aggregation of low resolution satellite data from the pixel level to coarser units such as geographical or administrative entities is commonly used in real-time monitoring of vegetation at global scale, for different reasons, e.g. work efficiency, availability of agricultural statistics for validation, data management, etc. Several studies have shown that low resolution imagery and aggregated data give precious information on actual and future states of the vegetation at global scale [1, 2]. However, using aggregated data implies several pure spatial problems such as effects of size, scale, surface or shape [3]. Therefore that use must be preceded by a comprehensive analysis of the dataset. For global vegetation monitoring issues and especially for the monitoring of vegetation response to climatic stresses, different questions must be raised at the very beginning of the research:

- Does the aggregation level fit with the objectives of the study?
- What is the intrinsic spatial heterogeneity of the aggregation entities?
- Is this heterogeneity constant in time?

Several spatial analysis techniques can be used to answer these questions. This paper focuses on the characterisation of the spatial heterogeneity of the so-called Regional Unmixed Means, RUM, computed with SPOT-VEGETATION NDVI data [4, 5]. We distinguish the spatial heterogeneity introduced by the presence of different land covers in a region and the spatial heterogeneity of the signal response of a same land cover. Our goal is to identify optimal test sites for the calibration of a forecasting model of vegetation response to climatic stresses at global scale using low resolution satellite data and weather forecasting products.

2. DATA AND METHODOLOGY

2.1. Regional Unmixed Means

Eight years (1998-2006) of SPOT-VEGETATION data were acquired from the Joint Research Centre MARS-FOOD programme, in the form of Regional Unmixed Means (RUM). These RUMs are unmixed values of vegetation indices (Normalized Difference Vegetation Index, NDVI and Dry Matter Production, DMP) averaged per region, i.e. pixels of the region covered totally by a single land cover class are used for the calculation of the regional mean for this land cover [4,5]. This technique allows a better exploitation of the time-series by partly reducing the problem of mixed values.

RUMs are calculated at global scale on 10-day (S10) or 30-day (S30) synthesis of NDVI and DMP for more than 3000 administrative regions (EU-NUTS2) and for 5 different land cover types – 'Cropland', 'Grassland', 'Shrubland', 'Forest' and 'Other land' - corresponding to aggregated classes of the Global Land Cover 2000 global map (Fig.1). The Global Land Cover 2000, GLC2000, was produced by the Global Vegetation Monitoring unit of the European Joint Research Centre in collaboration with 30 research teams [6, 7]. The main data set is the "VEGA 2000" data set, essentially composed of 14 months of daily 1km SPOT4-VEGETATION data acquired over the whole globe. The period covered is 1 Nov.1999-31 Dec.2000. The global land cover product has been created by mosaicing of 21 regional products based on regional expert knowledge.

Proc. 'Envisat Symposium 2007', Montreux, Switzerland
23–27 April 2007 (ESA SP-636, July 2007)

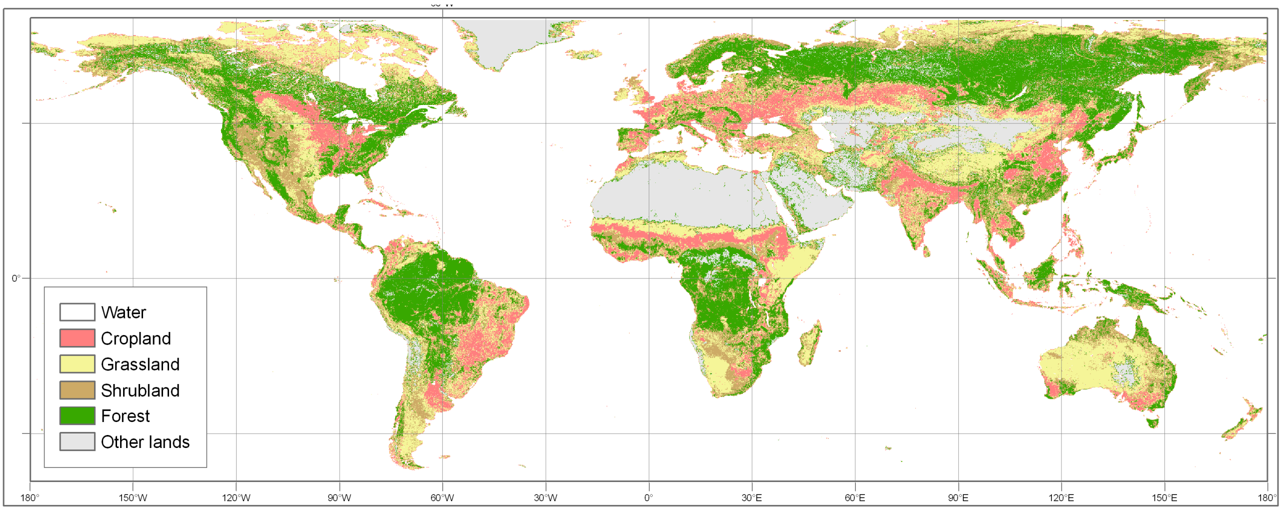

Figure 1. Global Land Cover 2000, generalised version (6 classes).

The GLC2000 legend is composed of 21 classes. To extract RUMs, a simplified version of the GLC2000 is used with the 21 classes generalised to 6 super classes: 'Water', 'Cropland', 'Grassland', 'Shrubland', 'Forest' and 'Other lands' (Tab.1). Further in the text, we refer to that land cover map as GLC2000-6C.

Table 1. Super classes of the generalised version of GLC2000

GLC2000-6C	Original GLC2000
Water	[Water]
Cropland	[Cultivated and managed areas] ; [Mosaic: Cropland/Tree Cover/ Other natural vegetation] ; [Mosaic: Cropland/Shrub and/or grass cover]
Grassland	[Herbaceous Cover, closed-open] ; [Sparse herbaceous or sparse shrub cover]
Shrubland	[Shrub Cover, closed-open, evergreen] ; [Shrub Cover, closed-open, deciduous]
Forest	[Tree Cover, broadleaved, evergreen] ; [Tree Cover, broadleaved, deciduous, closed] ; [Tree Cover, broadleaved, deciduous, open] ; [Tree Cover, needle-leaved, evergreen] ; [Tree Cover, needle-leaved , deciduous] ; [Tree Cover, mixed leaf type]
Other lands	[Tree Cover, regularly flooded, fresh] ; [Tree Cover, regularly flooded, saline water] ; [Mosaic: Tree Cover/Other natural vegetation] ; [Tree Cover, burnt] ; [Regularly flooded shrub and/or herbaceous cover] ; [Bare Areas] ; [Snow and Ice]

2.2. Spatial heterogeneity of land covers and land cover changes

The spatial heterogeneity introduced by the occurrence of different land covers in a region is evaluated with 2 different indices, the Relative Area *RA* and the Fragmentation index *F*, both extracted from the GLC2000-6C.

The Relative Area *RA* (Eq.1) is calculated for each land cover class existing in the administrative regions and gives information on the relative importance of land covers within the region.

$$RA_{r,i} = \frac{Pxl_{r,i}}{Pxl_{r,total}} \qquad (1)$$

Where: $Pxl_{r,i}$ is the number of pixels in the region r completely covered by the class i
$Pxl_{r,\,total}$ is the number of pixels of the region r

The Fragmentation index *F* (Eq.2) refers to the spatial pattern in the surrounding of a pixel [8]. Here we have analysed the fragmentation of land covers (GLC2000-6C) in a neighbourhood 3x3. The overall fragmentation of a region is evaluated by averaging the F_r values of each pixel contained in that region.

$$F_r = \frac{(n_r - 1)}{(c_r - 1)} \qquad (2)$$

Where: n_r is number of different classes observed in a 3x3 kernel,
c_r is number of the cell in the kernel.

Moreover the GLC2000-6C serves as reference for the complete time series (1998-2006) of RUMs although it corresponds to the situation of the year 2000. Land cover changes, LCC, are therefore underestimated. In order to evaluate the error introduced by the use of a static land-cover map, MODIS Land Cover products have been acquired from 2001 up to 2004. Contrary to the GLC2000, the classification scheme does not differ from a region to another [9]. The globe is mapped through a systematic/automatic procedure which allows a yearly update of the product. Amongst the different classification schemes available with the MODIS LC products, we used the primary land cover scheme: 17 classes of land cover defined by the International Geosphere-Biosphere Programme (IGBP) (11 natural vegetation classes, 3 developed land classes, permanent snow or ice, barren or sparsely vegetated, and water). Like the GLC2000, the classification schemes of the MODIS LC are multitemporal classes describing land cover properties as observed during a year (12 months of input data).

Several studies have been realised on the comparison between GLC2000 and MODIS LC products [10, 11]. According to Reference 10, discrepancies between the two land cover products are globally restricted when considering aggregated classes except for savannas/shrublands and for wetlands. Reference 11 shows that the level of agreement between both products at the pixel level vary between 51% and 81% according to the types of operators (Boolean or fuzzy) used for the comparison.

In this paper, the temporal trajectory of each pixel evaluated with MODIS LC for the period 2001-2004 is analysed in order to identify land cover changes. Aberrant behaviours are identified using simple logical rules and such pixels are excluded from the LCC analysis. A temporal trajectory is considered as aberrant or incoherent if one of the following rules is not respected:

- A maximum of two land cover changes is acceptable in the temporal trajectory of the pixel,
- Shrubland is the prior phase to forest,
- Pixels with a temporal trajectory oscillating between water and another class are neglected.

2.3. Spatial heterogeneity of the signal recorded by the sensor

The Coefficient of Variation CV is a measure of the relative dispersion of a variable and is used here to estimate how much is varying the spectral signature of a same land cover inside a same region. It is calculated on the RUM historical dataset (98-05) of monthly SPOT-VGT syntheses (S30) of NDVI.

$$CV_{NDVI,r,i} = \frac{Std_{NDVI,r,i}}{Mean_{NDVI,r,i}} \qquad (3)$$

Where: $Std_{NDVI,r,i}$ is the standard deviation of the S30 NDVI calculated for the region r and the land cover class i,

$Mean_{NDVI,r,i}$ is the averaged value of the S30 NDVI calculated for the region r and the land cover class i

3. Results

The first step of our research is to identify a set of calibration sites which are optimal for the modelling of vegetation-climate interactions, i.e. to select regions where the major part of the NDVI signal variation can be attributed to the climate and not to other phenomena such as land cover changes. We present here the results for the Western Europe.

3.1. Spatial heterogeneity of land covers and land cover changes

As explained above in the section 2.1, a huge improvement was already done with the extraction of the regional statistics by land cover type, i.e. the statistics are calculated per land cover by taking into account only pixels completely covered by the considered land cover. However regional statistics calculated for a dominant land cover are probably more reliable or more representative of the regional vegetation state than statistics extracted for an under-represented land cover because statistics are derived from few pixels inside the region. The under-represented land cover statistics are therefore more sensitive to outliers.

We quantify the spatial heterogeneity of land covers inside the aggregation entities with the Relative Area and the Fragmentation index. Fig. 2 presents the Relative Area calculated for croplands. For the Western Europe, the regions of North France, North Belgium, Denmark and Puglia (Italy) are dominated mostly by croplands, i.e. more than 75% of their area is covered by agricultural pixels (100% of crops).

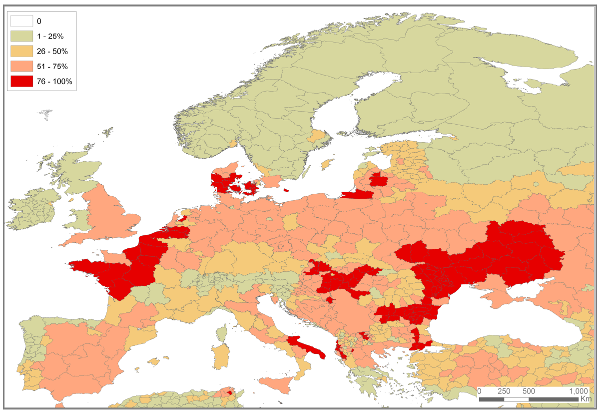

Figure 2. Relative area of croplands, expressed in % of the aggregation entity's area (NUTS-2 regions)

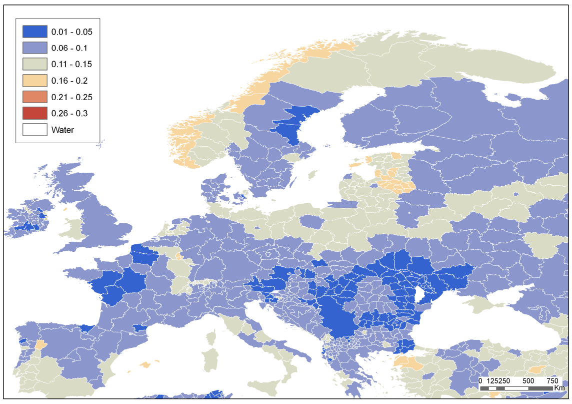

Figure 3. Mean Fragmentation index per NUTS-2 region

The mean fragmentation index per region is also a good indicator of the spatial heterogeneity caused by mixed land covers. Most regions in Western Europe present a very low level of fragmentation, with Fr below 0.125, meaning that in the 3x3 kernel (9km²) only two different land covers are observed in average (Fig.3).

As explained in the section 2.2, LCC are analysed with MODIS LC products. At first we present a quick comparison of the GLC2000-6C and the MODIS LC-6C (generalised version of MODIS LC 17 classes) of 2001. The differences between these two products are quite important (Fig. 4). Main discrepancies are between croplands and grasslands. A simple visual analysis shows clearly that grasslands are underestimated with the MODIS LC, mainly in favour of croplands in Western Europe (Fig.4, zoom) and of shrublands in the Alps. Urban areas are also more extended in the MODIS LC-6C. Therefore it is important to select the calibration sites in regions with low level of discrepancies between these two LC products in order to minimise the risk of misclassification.

The temporal trajectories of the MODIS LC images from 2001 to 2004 have been analysed to identify incoherencies and to locate the pixels where LCC occurred from 2001 to 2004. The most important spots of incoherent changes are situated in the region of the French *Massif Central* (Fig.5), indicating that the spatial heterogeneity of that region is high. Fig. 6 shows pixels where no LCC was observed over the 4 years. Croplands are probably the most stable land cover in Western Europe. The agricultural regions of France (Nord Pas-de-Calais, Picardie, Ile de France, etc.) have recorded few land covers changes as well as the Pô region and the Garonne plain (Fig. 5). In our case, the identification of stable regions is crucial as we

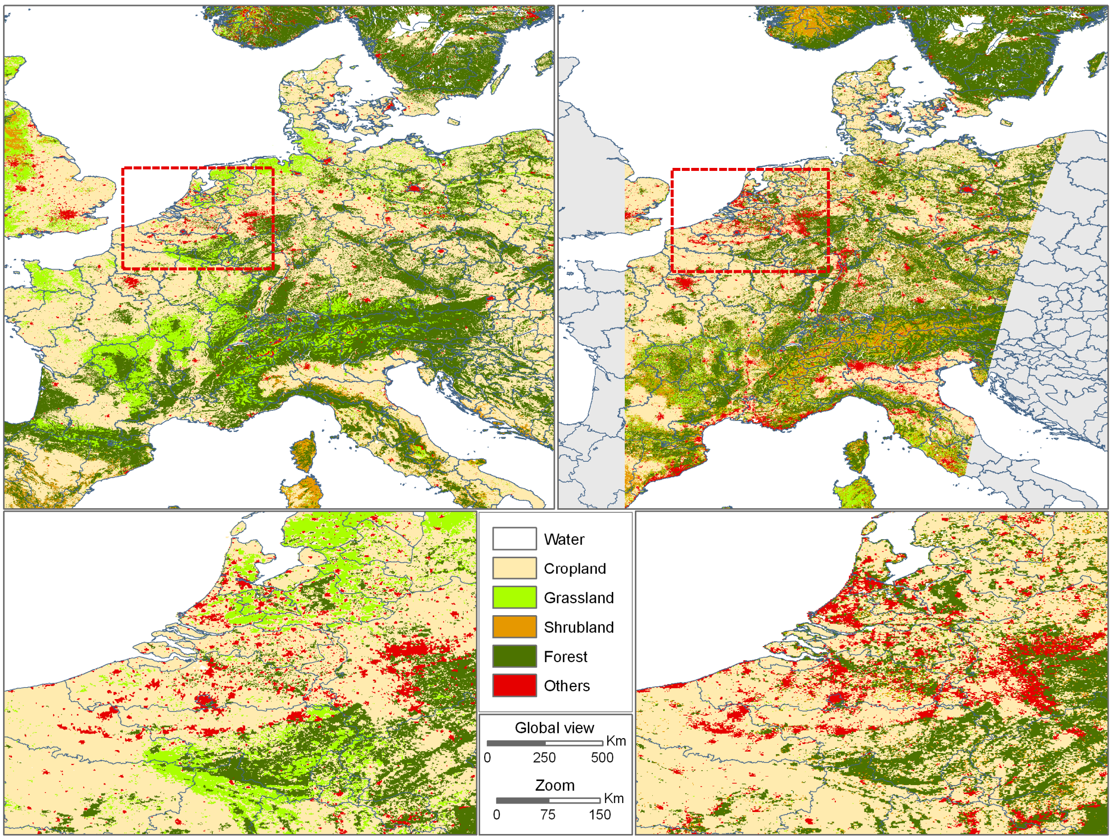

Figure 4. Comparison of GLC2000 (left) and MODIS Land Cover in 2001 (right), generalised versions (6 classes). Zoom over Belgium in red dotted lines.

are working on the response of vegetation to climatic events. Using regions with high LCC dynamic would affect our analysis by introducing extra-variability in the vegetation response retrieved from the sensor.

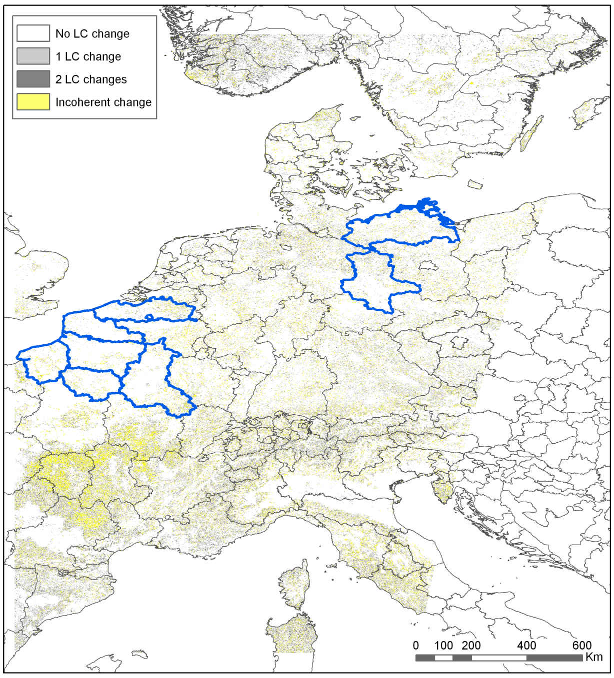

Figure 5. Land cover changes between 2001 and 2004 evaluated with MODIS Land Cover products. In blue, selected regions for croplands monitoring.

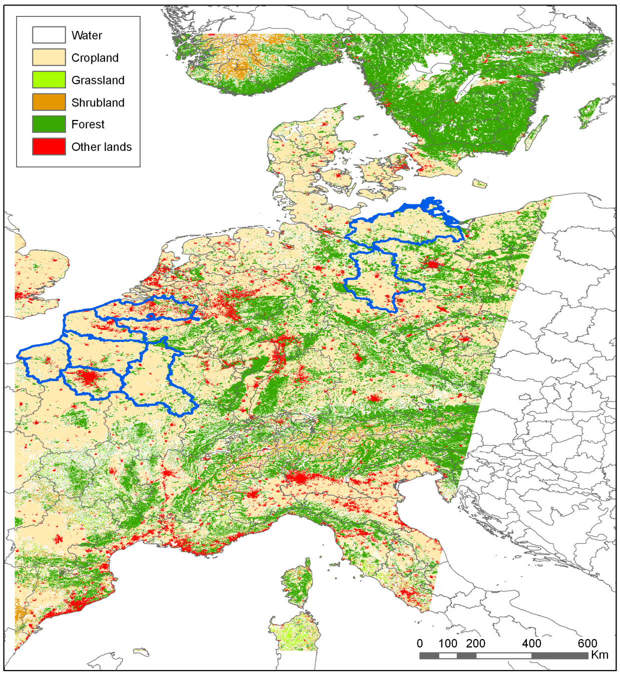

Figure 6. Stable land covers from 2001 to 2004 evaluated with MODLS LC products. In blue, selected regions for croplands monitoring.

3.2. Spatial heterogeneity of the signal recorded by the sensor

The spatial heterogeneity of the vegetation signal recorded by the sensor is evaluated with the coefficient of variation computed respectively for all the land cover classes (mixed), the croplands and the grasslands. In all cases, the coefficient of variation for the Western Europe is globally low. That suggests that the vegetation signal recorded at the sensor and extracted in the form of the RUMs can be considered as homogeneous at the region scale. In other words pixels retained for the calculation of the unmixed NDVI-means of a region have quite similar NDVI values.

Fig 7 shows that the coefficient of variation has also a seasonal variation with a first maximum in February and a second one in August, which correspond quite well for croplands with the harvest.

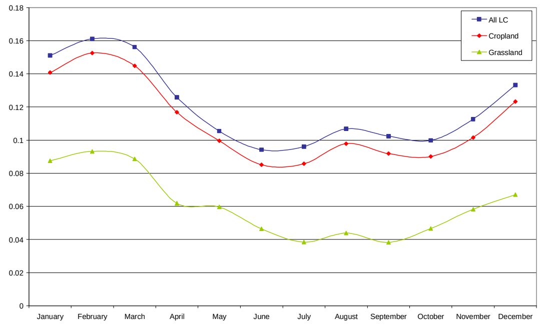

Figure 7. Coefficient of variation for the Flemish region calculated with the RUM of SPOT-VGT S30 NDVI historical data

3.3. Optimal calibration sites in Western Europe

Based on the results presented here above, eight regions have been selected: 5 in France, 1 in Belgium and 2 in Germany (Tab.2). These regions, outlined in blue in Figs. 5 and 6, are characterized by a dominant and very stable agricultural land cover, a low land cover fragmentation and a weak dispersion of the NDVI signal recorded at the pixel level.

4. CONCLUSION AND PERSPECTIVES

The coarse resolution of satellite sensors commonly used for the global vegetation monitoring is sometimes criticised by members of the scientific community: not enough accurate, non pure pixel, etc. However the monitoring of global processes such as vegetation growth or climate change requires an overpass frequency which is unbearable for high resolution satellites and even the spatial scale of such processes is not compatible with the high resolution imagery.

Table 2. Summary table of the scores realised by the selected sites for croplands monitoring

Country	Region	RA (%)	Mean Fr	Incoherent LCC (%)	Stable LC (%)	CVmean per year
Be	Flemish Region	78.1	0.08	3.3	61.0	0.11
Fr	Champagne-Ardenne	68.1	0.07	3.7	75.2	0.14
Fr	Haute-Normandie	85.3	0.06	3.8	83.2	0.11
Fr	Ile-de-france	78.5	0.05	2.5	66.0	0.14
Fr	Nord-Pas-de-Calais	85.6	0.03	1.0	83.4	0.13
Fr	Picardie	87.1	0.04	1.8	87.8	0.12
Ge	Mecklenburg-Vorpommern	68.1	0.12	3.7	63.0	0.09
Ge	Sachsen-Anhalt	72.9	0.08	2.9	63.0	0.12

This study on the Regional Unmixed Means dataset supplied by the MARS-FOOD programme and especially the analysis of the spatial and temporal heterogeneity of the NUTS-2 regions used as aggregation unit have lead to the selection of optimal calibration sites. Our further researches will be dedicated to the analysis of vegetation sensitivity to climatic stresses and to the study of the (dis-) similarities between regions on the basis of seasonal signal response. The results presented here as well as our forthcoming researches are very important for the elaboration of a vegetation monitoring model, as they will provide insight on the geographical applicability and limitations of the model.

5. ACKNOWLEDGEMENTS

This PhD research is financed by a grant from the Belgian « *Fonds pour la formation à la Recherche dans l'Industrie et l'Agriculture* ». We want to thank the MARS-FOOD programme of the EU-JRC and the Vlaamse Instelling voor Technoloogisch Onderzoek, VITO, for the supply of the SPOT-VGT RUM dataset.

6. REFERENCES

1. Nemani, R.R., Keeling, C.D., Hashimoto, H., Jolly, W.M., Piper, S.C., Tucker, C.J., Myneni, R.B. & Running, S.W. (2003). Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *Science*, **300**, 1560-3.
2. Zhou, L., Tucker, C.J., Kaufmann, R.K., Slayback, D., Shabanov, N.V. & Myneni, R.B. (2001). Variations in northern vegetation activity inferred from satellite data of vegetation index during 1981 to 1999, *Journal of Geophysical Research*, vol. **106**(D17), 20 069-20 83.
3. Zaninetti, J.M. (2005). *Statistique spatiale. Méthodes et applications géomatiques*. Hermes Sciences Publishing LTD, Paris, France, pp. 22-29.
4. Piccard, I., Eerens, H., Oger, R., Y. Curnel, Tychon, B., Ozer, P., Van Diepen, K., Boogaard, H., Genovese, G., Nègre, T., Wang, Y. & Li, X. (2005). Use of SPOT/VEGETATION in different versions of the European Crop Growth Monitoring System. In *Proc.2nd International VEGETATION Users Conference*, Antwerp, 24-26 March 2004. Eds.Veroustraete F, Bartholomé E. & Verstraeten W., Luxembourg, pp. 411-420.
5. Eerens, H. Piccard, I. Royer, A. & Orlandi, S. (2004). *Methodology of the MARS Crop Yield Forecasting System (Vol.3). Remote sensing information, data processing and analysis*. Eds. Royer, A. and Genovese, G. European Commission, DG-JRC. EU 21291/EN3. pp 53-56.
6. Bartholomé, E. & Belward, A. (2005). GLC2000: a new approach to global land cover mapping from Earth observation data. *International Journal of Remote Sensing*, **26**(9), 1959-77.
7. Mayaux, P., Strahler, A., Eva, H., Herold, M., Shefali, A., Naumov, S., Dorado, A., Di Bella, C., Johansson, D., Ordoyne, C., Kopin, I., Boschetti, L. & Belward, A. (2006). Validation of the Global Land Cover 2000 Map. *IEEE-Tran. on Geoscience and Remote Sensing*, **44**(7), 1728-1739.
8. Monmonier, M.S. (1974). Measures of Pattern Complexity for Choropleth Maps. The American Cartographer, vol. 1(2), 159-169. Cited in the *User guide of IDRISI Andes version 15*, 2006.
9. Friedl, M.A., McIver, D.K., Hodges, J.C.F., Zhang, X.Y., Muchoney, D., Strahler, A.H., Woodcock, C.E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., Schaaf, C. (2002). Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, **83**(1-2), 287-302.
10. Giri, C., Zhu, Z. & Reed, B. (2005) A comparative analysis of the Global Land Cover 2000 and MODIS land cover data sets, *Remote Sensing of Environment*, **94**, 123-32.
11. See, L.M., Fritz, S. (2006). A method to compare and improve land cover datasets: application to the GLC-2000 and MODIS land cover products. *IEEE-Trans. on Geoscience and Remote Sensing*, **44**(7), 1740-1746.