Extracting Dynamics of Multiple Indicators for Spatial recognition of Ecoclimatic zones in Circum-Saharan Africa

Didier Leibovici[1], Gilbert Quilevere[2]
[1]Centre for Geospatial Sciences, University of Nottingham
Nottingham NG7 2RD, U.K.
Telephone: +44 - (0)115 84 66058
Fax: +44 (0)115 95 15249
didier.leibovici@nottingham.ac.uk
[2]IRD-US166 Désertification, Maison de la Télédétection
500 av JF Breton, 34093 Montpellier cedex 05, France
gil.quillevere@mfr.asso.fr

1 Introduction

Focusing on ecoclimatic variations defined by the global physical and climatic conditions characterising arid and semi-arid zones the aim of this paper is to identify spatially the different patterns of ecoclimatic variations. For most of the indicators of arid or semi-arid zones, spatio-temporal variations through a typical year are observed . In order to take into account these dynamics in the clustering approach one must use a methodology that captures interactions between spatial-location, time of measurement and indicator measured. For this purpose a multiway analysis(Leibovici (2004)), generalising PCA, has been used on internationally recognised ecoclimatic indicators Lehouerou (2004, 1989) that characterise arid and semi-arid zones.

2 Clustering the dynamics of mutiple indicators

WORLDCLIM database Hijmans *et al.* (2005) (1950-2000, see www.worldclim.org/current.htm for the most recent one) at resolution 5 minutes of arc ($0.08dd$), was used to derive the necessary climatic parameters except for potential evapotranspiration of Penman-Monteih data provided from FAO. All the parameters were averaged for each month over 50 years in order to ensure large stability of the results in first approximation ignoring inter-annual variations. From these parameters Table 1 lists the monthly versions of the classical indicators used in the analysis.

2.1 Capturing the dynamics features

To take into account the dynamics of indicators we need a methodology that allows analysis, synthesis and extraction of interactions between $spatial\text{-}location$, $measurement\text{-}times$ and $indicators\text{-}measured$. The method used is taking advantage of the tensorial

Table 1: 10 indicator variables used in the analysis

Indicators	*Description*
P_m	monthly rainfall (mm)
T_{max}	monthly maximum of temperature($°C$)
T_{min}	monthly minimum of temperature($°C$)
T_{ave}	monthly average of temperature ($°C$)
ETo_m	monthly potential evapotranspiration
	of Penman-Monteih (mm)
P_m/ETo_m	monthly aridity index
Alt_{ave}	average altitude for the pixel grid considered(m)
$dM2T_{nb}$	number of of dry months according to the
	criterion $P_m < 2T_{ave}$
$Q3_m$	monthly simplified Emberger's pluviothermal index $Q3$
	$Q3_m = 3.43P_m/(T_{max} - T_{min})$ $(mm.°C^{-1})$
$dMETo_{nb}$	number of of dry months according to
	the criterion $P_m/ETo_m < 0.35$

structure of the data and can be considered as one generalisation of PCA for multi-array data: the method PTAk Leibovici and Sabatier (1998). It has been programmed as an R add-on package and is available online (Leibovici (2004), Leibovici (2007)). PTAk offers a decomposition similar to what is obtained from a Principal Component Analysis, but working on multiple-entries table (seen as tensors), instead of matrices. In our current case there are three entries: *spatial-location*, *month*, *indicator*, and each cell of the table contains the value of one indicator for a given month at a specific location. In order to describe the generalisation proposed with PTAk model let us first rewrite the PCA method within a tensorial framework.

For a given matrix X of dimension $n \times p$, the first principal component is a linear combination (given by a p-dimensional vector φ_1) of the p columns ensuring maximum sum of squares of the coordinates of the n-dimensional vector obtained. The square root of this sum of square is called the first singular value σ_1. One has: ${}^t(X\varphi_1)X\varphi_1 = \sigma_1^2$ and $X\varphi_1/\sigma_1$ is the principal component normed to 1. This maximisation problem can be written either in matrix form or tensor form:

$$\begin{aligned}\sigma_1 &= \max_{\substack{\|\psi\|_n=1\\ \|\varphi\|_p=1}} ({}^t\psi X \varphi) = \max_{\substack{\|\psi\|_n=1\\ \|\varphi\|_p=1}} X..(\psi \otimes \varphi) \\ &= {}^t\psi_1 X \varphi_1 = X..(\psi_1 \otimes \varphi_1)\end{aligned} \tag{1}$$

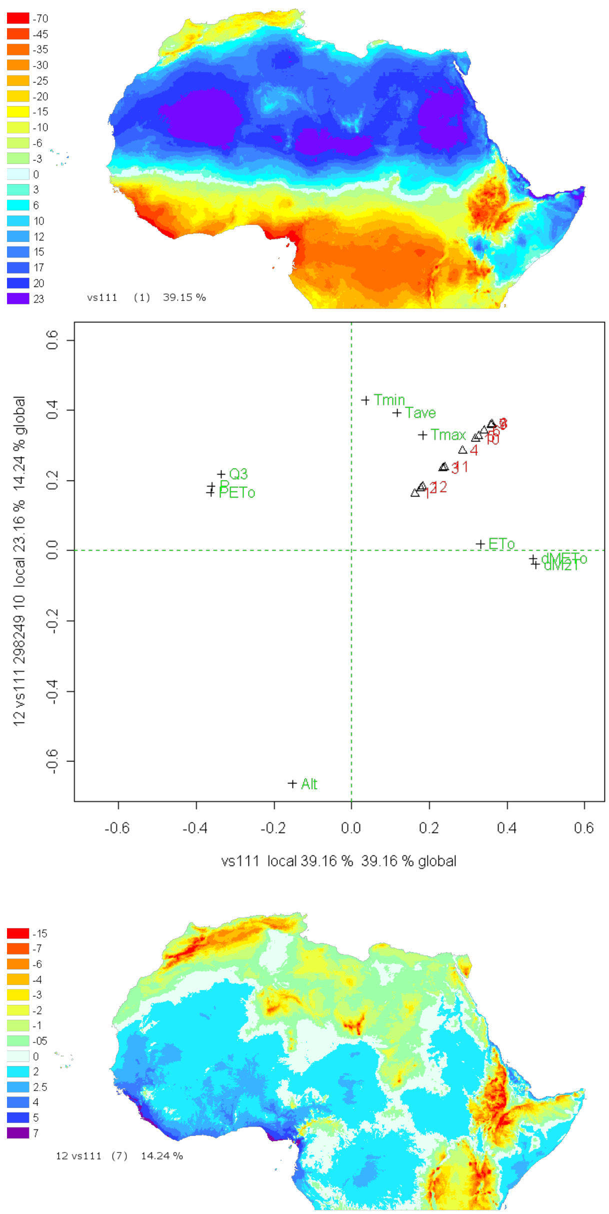
In equation 1 X is used either for the matrix or the tensor. An easy way of understanding computationally the operators ".." and "$\otimes$" is to see them as the following operations: $\psi_1 \otimes \varphi_1$ is a np vector of the n blocks of the p vectors $\psi_{1i}\varphi_1$, $i = 1, ...n$; ".." called a contraction generalises the multiplication of a matrix by a vector, and in the case like here of equal dimensions of the two tensors (np), corresponds to the natural scalar product (X is then also seen an np vector). ψ_1 is termed first principal component, φ_1 first principal axis, $(\psi_1 \otimes \varphi_1)$ is called first principal tensor. Now if X is a tensor of higher order, say 3 here with the modes: time or *month* ($t = 12$),variable or *indicator* ($v = 10$) and space or *spatial-location* ($s = 298249$), we can look for the first principal tensor associated with the singular value with the optimisation form:

$$\begin{aligned}\sigma_1 &= \max_{\substack{\|\psi\|_s=1\\ \|\varphi\|_v=1\\ \|\phi\|_t=1}} X..(\psi \otimes \varphi \otimes \phi) \\ &= X..(\psi_1 \otimes \varphi_1 \otimes \phi_1)\end{aligned} \tag{2}$$

Adding an orthogonality constraint allows to carry on the algorithm. Following a recursive algorithm scheme Leibovici (2007) the decomposition obtained offers a way of synthesising the data according to uncorrelated sets of components ordered by the percent of total sum of squares.

On figure 1 we have the plots of components ("loadings") of two different tensors. Temporal variations and ecolimatic variables associations, *i.e.* the $month$ and $indicator$ modes are plotted on the same scatter plot and their $spatial$-$location$ mode component can be read simultaneously to explain the variability captured. For the tensor n°1 ($vs111$) one can see a spatial separation between the Saharan zone positively weighted, with North Maghreb, Sahelian zone and central Africa negatively weighted. This appears mainly like a latitude gradient North and South from the Sahara. This is associated with the opposition on one side of drought and extreme dry condition indicators (ETo_m, $dM2T_{nb}$, $dMETo_{nb}$, $Tmax$) and on the other side rain related indicators ($Q3m$, Pm, $PETom$) ; and this occurs all year around, and especially during rain seasons (May(5) to October(10)). The vertical axis on figure 1 shows an opposition between Altitude and temperature ($Tmin$ more strongly) also persistent all year and more likely during rainy seasons (May(5) to October(10)). This vertical axis is read with the bottom spatial picture showing high relief associated with it.

Figure 1: Spatio-temporal association of ecoclimatic indicators captured in the first principal tensor representing 39.16%(vs111) and on the first $month$ mode associated principal tensor representing 14.25% of variability.(*Labels of indicators are the first letters of the names given in Table 1; scatter plot and spatial values are the "loadings" or components values of the tensors*)

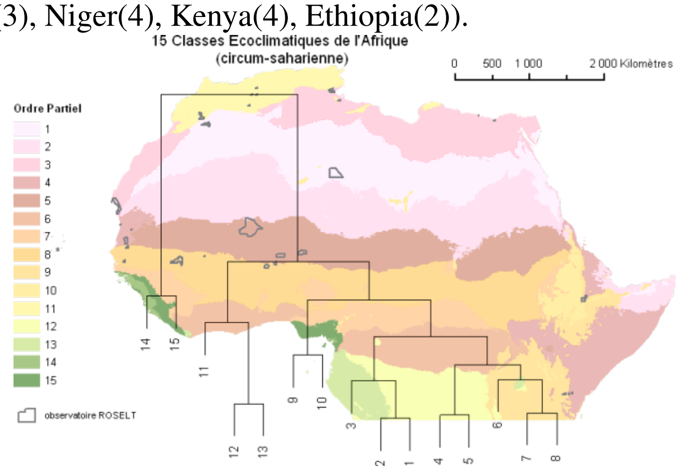


Other tensors will be shown at the conference expressing different spatio-temporal patterns altogether capturing various ecoclimatic aspects.

2.2 Ecoclimatic zones and their proximities

Once meaningful Principal Tensors are selected, it is possible to perform a multivariate clustering on the corresponding spatial components to obtain spatial classes of zones with similar ecoclimatic dynamics.Figure 2 shows the 15 classes we obtained with a $k$means procedure. In order to reinforce the ecoclimatic proximity of the classes obtained we performed a hierarchical clustering on the centroids of the classes. The dendogram obtained is also used to calibrate the colour range by matching it with a "pseudo" ordering of the classes read or computed from the dendogram history.

Figure 2: Ecoclimatic Classes with their aggregation tree illustrating hierarchical climatic proximities (based on WorldClim 2004 parameters) and 34 ROSELT pilots observatories polygons (Egypt(2), Tunisia(3), Algeria(5), Morocco(3), Mauritania(3), Cap-Verde(2), Senegal(3), Mali(3), Niger(4), Kenya(4), Ethiopia(2)).



3 Perspectives and Conclusion

The results are very encouraging but some issues may be relevant depending on the use of this classification. The ecoclimatic characteritics captured with the PTAk method would need physical process assessment for validation of the classification obtained at a biometeorological level . So far some experts including HN Le Houérou found coherence in the results but full validation in comparison with other known classifications has to be addressed. It is very interesting that the spatial coherence and homogeneity is well achieved with the herein method without any spatial constraint other than actual indicators measurements natural spatial multiple autocorrelation. Fuzziness of the borders can be addressed when dealing with ecoregion borders and some methods could be applied *a posteriori* Hargrove and Hoffman (1999), Hargrove and Hoffman (2002). Other fuzzy algorithms are available but the intensive computing needed for this massive dataset may preclude their use. Averaging initial parameters over 50 years for stability of results could be compared with an approach considering a more realistic range of different stable periods, say before 1970 and after at least. This modified approach would now consider adding a *period* mode, then a tensor of order 4 to analyse. PTAk can in fact decompose a tensor of any order (k>2). Some scale issues are relevant to this methodology either when looking at the available resolutions of the WorldClim data but also on the extent analysed.

4 Acknowledgements

The authors whishes to thank Henry Noël Le Houérou who showed a stimulating interest in this work and for helpful comments about the report Quillevere (2004). Warm thanks go also to Hélène Fonta who finalised an atlas Fonta (2005) available online within http://mdweb.roselt-oss.org and containing maps issued from this methodology also at National levels.

References

Fonta, H. (2005). Création d'une base de données géographique régionale dans le cadre du programme ROSELT. mémoire de D.U Cartographie des territoires et Systèmes d'Information Géographique Université Montpellier III, Maison de la Télédétection.

Hargrove, W. W. and Hoffman, F. M. (1999). Using multivariate clustering to characterize ecoregion borders. *Computers in Science and Engineering Special Issue on Scientific Visualization of Massive Data Sets*, **1**(4), 18–25.

Hargrove, W. W. and Hoffman, F. M. (2002). Representativeness and network site analysis based on quantitative ecoregions. web, Oak Ridge National Laboratory Environmental Sciences Division, http://research.esd.ornl.gov/ hnw/network/.

Hijmans, R., Cameron, S., Parra, J., Jones, P., and Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land area. International Journal of Climatology, **25**, 1965–1978.

Lehouerou, H. N. (1989). Classification écoclimatique des zones arides (s.l.) de l'Afrique du nord. *Ecologia Mediterranea*, **XV (3/4)**, 95–143.

Lehouerou, H. N. (2004). An Agro-bioclimatic Classification of Arid and Semi-arid lands in the Isoclimatic Mediterranean Zones. *Arid Land Research and Management*, **18**, 301–346.

Leibovici, D. (2004). PTA-*k* add on R-package version 1.1-12. `http://cran.r-project.org/src/contrib/PACKAGES.html`, **version 1.1-12**.

Leibovici, D. (2007). An R-package for a generalisation of PCA to multiway data: PTA-*k*modes. *Journal of Statistical Software*, **in preparation**.

Leibovici, D. and Sabatier, R. (1998). A Singular Value Decomposition of *k*-Way Array for a Principal Component Analysis of Multiway Data, PTA-*k*. *Linear Algebra and Its Applications*, **269**, 307–329.

Quillevere, G. (2004). Representativité circum-saharienne du réseau d'observatoires ROSELT. Mastère SILAT ENSAM/INRA, Montpellier, Maison de la Télédétection.

ROSELT/OSS (1995). Conception, organisation et mise en oeuvre de ROSELT. *Collection ROSELT/OSS, Document Scientifique*, **1**, 130.

Biography

Dr Leibovici has a PhD in Applied Mathematics from the University of Montpellier and worked for some years as a Statistician Researcher in epidemiological and medical imaging contexts in France and in England. More recently after completing a Masters degree in Information Technology he worked in geomatic modelling for landscape changes at the IRD (Institute of Research for Development) in France.