Geoderma 161 (2011) 107–114

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

Large trends in French topsoil characteristics are revealed by spatially constrained multivariate analysis

Dominique Arrouays [a,*], Nicolas P.A. Saby [a], Jean Thioulouse [b], Claudy Jolivet [a], Line Boulonne [a], Céline Ratié [a]

[a] INRA Orléans, InfoSol Unit, US 1106, CS 40001 Ardon, 45075, Orléans cedex 2, France
[b] Université de Lyon CNRS; Université Lyon 1; CNRS UMR 5558, Biométrie et Biologie Evolutive, F-69622, Villeurbanne cedex, France

ARTICLE INFO

Article history:
Received 12 February 2010
Received in revised form 29 October 2010
Accepted 3 December 2010
Available online 3 February 2011

Keywords:
Topsoil
Mapping
Multivariate analysis
Geostatistics
France

ABSTRACT

Spatially constrained multivariate analysis methods (MULTISPATI-PCA) and classical principal component analysis are applied for the entire country of France to study the main soil characteristics of topsoil and to assess if their multivariate spatial pattern can provide insight on their extent and origin. The results of the MULTSPATI-PCA provided evidence of strong spatial structures attributed to different natural processes. The first axis was interpreted as an axis of global soil richness in clay content. Axis 2 reflected the influence of some parent materials on the geochemical content of K and Al. Axis 3 showed a very large gradient of relative content in coarse silt. Axis 4 was driven by gradients of maritime influence. We show that MULTISPATI-PCA allows better than classical PCA to detect and map large regional trends in the distribution of topsoil characteristics. The two first axes were expected and the maps obtained by both methods were consistent. Interestingly, the other gradients were not expected and were better shown by MULTISPATI-PCA than by classical PCA.

1. Introduction

Topsoil characteristics in temperate Western Europe mainly depend on parent material composition, geomorphology, and past and present climatic conditions. The dominant soil processes are weathering, leaching, illuviation, and, more locally, podzolisation. Among the parent materials, loess deposits are widespread in northern France, where they are locally very thick (Antoine et al., 1998; Jamagne et al., 1981; Lautridou et al., 1986; Lebret and Lautridou, 1991). The map of the quaternary superficial formations of northwest France was updated in 1998 (Antoine et al., 1998), but it still unclear to what extent the silty aeolian sediments may have spatially influenced topsoil characteristics. In a recent study, Saby et al. (2009) showed that the distribution of some trace elements in French topsoil is strongly dependent on the geographical distribution of some specific parent materials. We assume that this dependence is also the case for some major elements. France has about 5500 km of coastline along the Mediterranean Sea, the Atlantic Ocean, and the Channel. Although atmospheric deposition in maritime environments is known to affect terrestrial ecosystems (Farrell, 1995) and soil (Zhang, 2003), to our present knowledge, no national systematic inventory has explored to what extent the coastal sea salt spray may have influenced the topsoil chemistry.

The existence of a systematic sampling grid over the entire French metropolitan territory (Arrouays et al., 2002; Arrouays et al., 2003; Jolivet et al., 2006; Saby et al., 2006) offers an opportunity to tackle these general issues. We study the topsoil characteristics (particle-size distribution, total major elements, organic C and N, Ca, C, N, P, K, cation exchange capacity and exchangeable cations, and pH) to assess if their spatial distribution and correlation can provide any insight into their extent and origin. There are several statistical methods available to study the spatial covariation of variables. One is based on multivariate geostatistics (Webster and Oliver, 2007), where the joint spatial variation and covariation of two variables is expressed as a co-variogram and can be modelled with the linear model of coregionalization (LMCR). This method has been for example successfully used to determine the correlations between yield of wheat and soil nutrients (Bourennane et al., 2003; Webster et al., 1994), in exploratory analysis of trace metal concentrations in soil (Atteia et al., 1994), and to study the distribution of nematodes in soil (Webster and Boag, 1992). Unfortunately, it is strongly dependant on the goodness-of-fit (in least square sense) of the linear model of co-regionalization. Another approach consists in performing in a first step multivariate analysis like Principal Component Analysis (PCA) and to interpret the structures observed on the first few axes. A second step can be achieved by mapping the scores in geographical space or by using geostatistical tools (e.g. (Boruvka et al., 2007; Odlare et al., 2005; Oliver et al., 1997; Satapathy et al., 2009). However, standard PCA does not directly take into account spatial relations and was not specifically designed to identify spatial structures. Therefore, we use here a spatially

* Corresponding author. Tel.: +33 238414802; fax: +33 238417869.
E-mail address: Dominique.Arrouays@orleans.inra.fr (D. Arrouays).

0016-7061/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.geoderma.2010.12.002

constrained multivariate analysis method (MULTISPATI-PCA, (Dray et al., 2008), which is a generalisation of Wartenberg's (1985) Multivariate Spatial Correlation Analysis (MSCA). This technique is a purely descriptive method, based solely on linear algebra and on geometrical properties and implies a compromise between the relationships among many variables (multivariate analysis) and their spatial structure (autocorrelation).

In this paper, we applied MULTISPATI-PCA to a set of topsoil characteristics and we compare the results to classical PCA.

2. Material and methods

2.1. Study area

The French National Soil Quality Monitoring Network, "Réseau de Mesures de la Qualité des Sols" (Jolivet et al., 2006; Saby et al., 2006), consists of soil property observations on a 16-km regular grid covering the French metropolitan territory (550,000 km²). This network was designed to monitor soil evolutions and to identify diffuse contamination due to either atmospheric deposition of trace elements on soils or agricultural practices (e.g., fertilisers, sludge amendments, and inorganic pesticides). The complete inventory consists of around 2200 sites, but in this study, we use measurements from the 2117 sites analysed at present.

The sites were selected at the centre of each 16×16 km cell. In the case of soil being unavailable at the centre of the cell (i.e., urban area, road, river, etc.), an alternative location was selected as close as possible to the centre of the cell, within a 1 km radius, to find a natural (undisturbed or cultivated) soil. However, it was not always possible to find an alternative location. All land cover types were present in the dataset, except industrial sites, which were not sampled. At each site, 25 individual core samples were taken from the topsoil (0–30 cm) layer, using a stratified random sampling design within a 20×20 m area. Core samples were bulked to obtain a composite sample for each site. Soil samples were air-dried and sieved to 2 mm before analysis (AFNOR, 1994). The topsoil was sampled from 0 to 30 cm because it corresponds to the maximal depths affected by ploughing and is a conventional thickness often reported in France (e.g., (Arrouays et al., 2001; Arrouays et al., 2008).

The following soil characteristics were retained: (i) the total organic carbon content (TOC) measured by dry combustion, (ii) the particle-size distribution using five classes [(clay (0–2 μm), fine silt (2–20 μm), coarse silt (20–50 μm), fine sand (50–200 μm) and coarse sand (200–2000 μm)] using wet sieving and the pipette method (NF X 31–107), (iii) cation exchange capacity and exchangeable cations (cobaltihexamin method), (iv) total K, Ca, Mg, Fe and Al determined by ICP-MS after dissolution with hydrofluoric and perchloric acids, (v) pH in water (1 to 5 soil to water ratio), and (vi) extractable P (Olsen method).

Analyses were performed by the Soil Analysis Laboratory of INRA in Arras, France, which is accredited for soil and sludge analysis.

Table 1 shows that the boxcox transformation has removed most of the skewness of the raw data although log transform performs worst. This table shows also that the parameters in the dataset covered a very wide range. The Table 2 shows the linear correlations between the soil parameters. Most of these correlations were expected, such as those between particle-size fractions, or those observed between cation exchange capacity and exchangeable cations. The correlation between clay and C content globally confirms the effect of the fine particle fractions on C stabilisation in soil (Arrouays et al., 2006).

2.2. MULTISPATI-PCA

In this paper we used both classical PCA and MULTISPATI-PCA. The former is a classical analysis. The latter has been fully described by Dray et al. (2008). This paper does not give the full detail of the two procedures that have been already published but we present an outline below.

2.2.1. General analysis of a statistical triplet

Multivariate analysis methods can be described in terms of duality diagram and of their associated statistical triplet (Holmes, 2006; Dray and Dufour, 2007).

Let $\mathbf{X}$ be an $(n \times p)$ matrix (soil parameters). The p variables (in columns) have been centered by substracting the overall mean for the n samples (in rows), and scaled to unit norm by dividing by the standard deviation.

Let $\mathbf{D}_n$ be a scalar product of R^n and let $\mathbf{D}_p$ be a scalar product of R^p. The analysis of the statistical triplet $(\mathbf{X}, \mathbf{D}_p, \mathbf{D}_n)$ is obtained by the eigenanalysis of matrix $\mathbf{S}$:

$$\mathbf{S} = \mathbf{X}^{\mathrm{T}}\mathbf{D}_{\mathrm{n}}\mathbf{X}\mathbf{D}_{\mathrm{p}} \quad (1)$$

Table 1
Descriptive statistics of the soil properties.

Soil properties (unit)	Short name	Mean	Median	10% percentile	90% percentile	Kurtosis	Skewness	Kurtosis log(x)	Skewness log(x)	t (BoxCox parameter)	Kurtosis (boxcox)	Skewness (boxcox)
Clay (g kg^{-1})	Clay	244.11	209.50	106.00	437.90	0.786	0.957	4.406	−1.325	0.441	0.226	0.031
Fine silt (g kg^{-1})	–	227.15	230.00	103.10	352.00	−0252	−0.028	13.979	−3.051	0.929	1.254	0.029
Coarse silt (g kg^{-1})	coarse.silt	176.30	144.00	60.00	358.00	0.292	0.990	5.866	−1.627	0.406	−0.004	0.015
Fine sand (g kg^{-1})	fine.sand	134.15	115.00	40.00	244.00	5.179	1.804	1.472	−0.807	0.296	0.452	0.027
Coarse Sand (g kg^{-1})	coarse.sand	218.29	145.00	16.00	521.90	0.961	1.226	−0.018	−0.744	0.255	−0.833	−0.100
Organic carbon (g kg^{-1})	carbon	25.58	19.50	9.49	48.29	17.777	3.163	0.596	0.214	−0.088	0.919	−0.015
Total nitrogen (g kg^{-1})	tot.N	2.14	1.72	0.85	3.85	11.953	2.644	2.106	−0.318	0.099	1.254	0.029
total calcium carbonate (g kg^{-1})	tot.calc	54.03	0.97	0.60	218.90	10.117	3.121	−0.080	1.234	−0.494	−1.116	0.681
pH water	pH water	6.41	6.20	4.60	8.20	−1.283	0.019	−1.098	−0.230	0.673	−1.240	−0.058
Available phosphorus (g kg^{-1})	available.P	0.05	0.03	0.00	0.12	56.689	4.323	−0.453	−0.610	0.220	−0.601	−0.072
CEC (Cmol+ kg^{-1})	Cec	14.04	10.20	3.81	29.99	1.955	1.422	0.076	−0.283	0.122	−0.345	−0.010
Exchangeable Ca (Cmol+ kg^{-1})	exc.Ca	12.41	8.66	0.71	28.69	1.683	1.350	1.284	−1.175	0.324	−0.565	−0.108
Exchangeable Mg (Cmol+ kg^{-1})	exc.Mg	1.07	0.70	0.23	1.89	32.909	4.907	1.237	0.024	−0.006	1.238	−0.002
Exchangeable K (Cmol+ kg^{-1})	exc.K	0.38	0.32	0.11	0.69	291.788	11.361	1.089	−0.524	0.181	1.132	0.040
Exchangeable Na (Cmol+ kg^{-1})	exc.Na	0.12	0.05	0.02	0.11	1047.680	29.968	10.680	1.792	1.000	9.401	1.551
Exchangeable Al (Cmol+ kg^{-1})	exc.Al	0.61	0.09	0.02	2.10	9.967	2.993	−0.584	0.690	1.000	−0.584	0.690
Total Al (g.100g^{-1})	tot.Al	4.86	4.70	2.25	7.66	−0.463	0.129	4.917	−1.753	0.825	−0.387	−0.080
Total Ca (g.100g^{-1})	tot.Ca	2.56	0.44	0.12	9.02	9.877	3.065	−0.175	0.714	−0.197	−0.185	0.045
Total Fe (g.100g^{-1})	tot.Fe	2.54	2.28	0.94	4.39	5.143	1.548	6.261	−1.839	0.476	0.992	0.083
Total K (g.100g^{-1})	tot.K	1.61	1.45	0.64	2.85	0.574	0.864	3.468	−1.142	0.447	0.053	0.009
Total Mg (g.100g^{-1})	tot.Mg	0.54	0.36	0.12	0.96	104.641	8.227	2.077	−0.480	0.096	2.039	0.059

Table 2
Linear correlation matrix of the boxcox transformed data. Bold print indicates significant linear correlation coefficients at 1% level. Fine silt has been removed due to additive log ratio transformation.

	Clay	coarse.silt*	fine.sand *	coarse.sand *	carbon	tot.N	tot.calc	ph.wate r	available .P	cec	exc.Ca	exc.Mg	exc.K	exc.Na	exc.Al	tot.Al	tot.Ca	tot.Fe	tot.K	tot.Mg
Clay	1.000	**0.085**	**−0.437**	**−0.599**	**0.406**	**0.573**	**0.464**	**0.496**	0.010	**0.835**	**0.757**	**0.555**	**0.508**	**0.306**	**−0.245**	**0.452**	**0.554**	**0.650**	**0.073**	**0.549**
coarse.silt	**0.085**	1.000	−0.018	**−0.662**	**−0.139**	−0.007	−0.048	**0.134**	**0.297**	**0.113**	**0.126**	**0.063**	**0.184**	0.037	**−0.219**	**−0.067**	0.053	**0.100**	**−0.062**	**0.062**
fine.sand	**−0.437**	−0.018	1.000	**0.298**	**−0.254**	**−0.292**	**−0.114**	**−0.099**	−0.033	**−0.321**	**−0.274**	**−0.150**	**−0.170**	**−0.160**	0.022	**−0.250**	**−0.143**	**−0.230**	**−0.064**	**−0.125**
coarse.sand	**−0.599**	**−0.662**	**0.298**	1.000	**−0.095**	**−0.253**	**−0.244**	**−0.387**	**−0.116**	**−0.547**	**−0.532**	**−0.330**	**−0.331**	**−0.233**	**0.312**	**−0.128**	**−0.325**	**−0.379**	**0.102**	**−0.304**
carbon	**0.406**	**−0.139**	**−0.254**	**−0.095**	1.000	**0.909**	**0.073**	**−0.135**	**−0.167**	**0.389**	**0.212**	**0.219**	**0.065**	**0.218**	**0.336**	**0.398**	**0.146**	**0.404**	**0.117**	**0.344**
tot.N	**0.573**	−0.007	**−0.292**	**−0.253**	**0.909**	1.000	**0.183**	**0.086**	**0.057**	**0.562**	**0.430**	**0.381**	**0.267**	**0.275**	**0.094**	**0.511**	**0.331**	**0.565**	**0.191**	**0.503**
tot.calc	**0.464**	−0.048	**−0.114**	**−0.244**	**0.073**	**0.183**	1.000	**0.801**	0.049	**0.619**	**0.656**	**0.190**	**0.379**	0.042	**−0.422**	**−0.067**	**0.801**	**0.134**	**−0.187**	**0.300**
ph.water	**0.496**	**0.134**	**−0.099**	**−0.387**	**−0.135**	**0.086**	**0.801**	1.000	**0.275**	**0.714**	**0.822**	**0.388**	**0.526**	**0.124**	**−0.785**	−0.008	**0.829**	**0.223**	**−0.122**	**0.362**
available.P	0.010	**0.297**	−0.033	**−0.116**	**−0.167**	0.049	0.049	**0.275**	1.000	**0.106**	**0.192**	**0.106**	**0.513**	0.004	**−0.401**	0.025	**0.181**	0.031	**0.131**	0.035
cec	**0.835**	**0.113**	**−0.321**	**−0.547**	**0.389**	**0.562**	**0.619**	**0.714**	**0.106**	1.000	**0.937**	**0.628**	**0.552**	**0.286**	**−0.485**	**0.301**	**0.739**	**0.556**	−0.023	**0.559**
exc.Ca	**0.757**	**0.126**	**−0.274**	**−0.532**	**0.212**	**0.430**	**0.656**	**0.822**	**0.192**	**0.937**	1.000	**0.598**	**0.548**	**0.245**	**−0.650**	**0.205**	**0.779**	**0.463**	**−0.075**	**0.475**
exc.Mg	**0.555**	**0.063**	**−0.150**	**−0.330**	**0.219**	**0.381**	**0.190**	**0.388**	**0.106**	**0.628**	**0.598**	1.000	**0.517**	**0.509**	**−0.372**	**0.426**	**0.451**	**0.549**	**0.187**	**0.671**
exc.K	**0.508**	**0.184**	**−0.170**	**−0.331**	**0.065**	**0.267**	**0.379**	**0.526**	**0.513**	**0.552**	**0.548**	**0.517**	1.000	**0.199**	**−0.413**	**0.250**	**0.513**	**0.365**	**0.175**	**0.413**
exc.Na	**0.306**	0.037	**−0.160**	**−0.233**	**0.218**	**0.275**	0.042	**0.124**	0.004	**0.286**	**0.245**	**0.509**	**0.199**	1.000	**−0.078**	**0.308**	**0.203**	**0.319**	**0.100**	**0.303**
exc.Al	**−0.245**	**−0.219**	0.022	**0.312**	**0.336**	**0.094**	**−0.422**	**−0.785**	**−0.401**	**−0.485**	**−0.650**	**−0.372**	**−0.413**	**−0.078**	1.000	**0.112**	**−0.545**	**−0.112**	**0.144**	**−0.159**
tot.Al	**0.452**	**−0.067**	**−0.250**	**−0.128**	**0.398**	**0.511**	**−0.067**	−0.008	0.025	**0.301**	**0.205**	**0.426**	**0.250**	**0.308**	**0.112**	1.000	**0.138**	**0.737**	**0.712**	**0.672**
tot.Ca	**0.554**	0.053	**−0.143**	**−0.325**	**0.146**	**0.331**	**0.801**	**0.829**	**0.181**	**0.739**	**0.779**	**0.451**	**0.513**	**0.203**	**−0.545**	**0.138**	1.000	**0.348**	**−0.080**	**0.549**
tot.Fe	**0.650**	**0.100**	**−0.230**	**−0.379**	**0.404**	**0.565**	**0.134**	**0.223**	0.031	**0.556**	**0.463**	**0.549**	**0.365**	**0.319**	**−0.112**	**0.737**	**0.348**	1.000	**0.262**	**0.744**
tot.K	**0.073**	**−0.062**	**−0.064**	**0.102**	**0.117**	**0.191**	**−0.187**	**−0.122**	**0.131**	−0.023	**−0.075**	**0.187**	**0.175**	**0.100**	**0.144**	**0.712**	**−0.080**	**0.262**	1.000	**0.382**
tot.Mg	**0.549**	**0.062**	**−0.125**	**−0.304**	**0.344**	**0.503**	**0.300**	**0.362**	0.035	**0.559**	**0.475**	**0.671**	**0.413**	**0.303**	**−0.159**	**0.672**	**0.549**	**0.744**	**0.382**	1.000

where $\mathbf{X}^T$ is the transpose of $\mathbf{X}$. This eigenanalysis leads to the diagonal matrix of eigenvalues Λ and to eigenvectors $\mathbf{U}$. The row scores $\mathbf{R}$ are given by:

$$\mathbf{R} = \mathbf{X}\mathbf{D}_p\mathbf{U} \tag{2}$$

and the variable loadings $\mathbf{L}$, by:

$$\mathbf{L} = \mathbf{U}\Lambda^{1/2} \tag{3}$$

2.2.2. The case of simple PCA

If $\mathbf{D}_n = 1/n\mathbf{I}_n$ and $\mathbf{D}_p = \mathbf{I}_p$, the analysis of the statistical triplet $(\mathbf{X}, \mathbf{D}_p, \mathbf{D}_n)$ is the classical PCA of $\mathbf{X}$. This leads to the eigenanalysis of matrix $\mathbf{X}^T\mathbf{X}$ and to usual row scores and variable loadings. The advantage of the general triplet notation is that it can be used to define many other multivariate analyses, like for example Principal Coordinate Analysis, Correspondence Analysis or Multiple Correspondence Analysis.

The ade4 package for the R statistical software (Chessel et al., 2004) is based on the duality diagram and on the analysis of statistical triplets. For the analysis of one-table data sets alone, it offers 10 different methods.

2.2.3. The MULTISPATI-PCA analysis

The aim of the MULTISPATI analysis is to take into account the spatial location of samples in the analysis of the data table. This information is introduced by the way of a spatial weighting matrix $\mathbf{C} = [c_{ij}]$ that indicates the strength of the relationship between units i and j. This matrix can take many forms, for example a binary connectivity matrix $\mathbf{B}$ ($b_{ij} = 1$ if units i and j are neighbours, else $b_{ij} = 0$). Here, This matrix $\mathbf{B}$ is transformed into a row-standardized spatial weight matrix $\mathbf{W}$ (weight equal to the inverse of the number of neighbours) as:

$$\mathbf{W} = \left[c_{ij} \Big/ \sum_{j=1}^{n} c_{ij}\right] \tag{5}$$

Using $\mathbf{W}$ has several advantages, mainly allowing a decomposition of Moran's I into the product of a spatial smoothing and the Pearson correlation between the variable and its spatial lag (Dray et al., 2008). Another advantage is that, in the case of a correlation matrix PCA of $\mathbf{X}$, the MULTISPATI-PCA analysis is equivalent to Wartenberg's Multivariate Spatial Correlation Analysis (Wartenberg, 1985).

The MULTISPATI-PCA analysis is a fully matched coinertia analysis (Torre and Chessel, 1994) of the data table $\mathbf{X}$ and the « lag vector » $\mathbf{WX}$. It corresponds to the statistical triplet $(\mathbf{X}^T\mathbf{D}_n\mathbf{WX}, \mathbf{D}_p, \mathbf{D}_p)$, and after some transformation, it leads to the eigenanalysis of matrix $\mathbf{H}$ which is symmetric:

$$\mathbf{H} = (1/2)\left(\mathbf{X}^{\mathrm{T}}\left(\mathbf{W}^{\mathrm{T}}\mathbf{D}_{\mathrm{n}} + \mathbf{D}_{\mathrm{n}}\mathbf{W}\right)\mathbf{X}\mathbf{D}_{\mathrm{p}}\right) \tag{6}$$

The row scores of this analysis maximize the scalar product between a linear combination of the original variables and a linear combination of the lagged variables. It is therefore a compromise between the simple PCA of the pedological variables and the maximal spatial autocorrelation obtained by computing the mean of each variable over the neighbours of each point («lagged variables»).

The advantage of MULTISPATI-PCA over PCA is that MULTISPATI-PCA sample scores maximise the spatial autocorrelation between sites, while conventional PCA scores maximise the inertia (i.e., the sum of variances). MULTISPATI-PCA scores are, therefore, "smooth" and show strong spatial structures on the first few axes, while PCA scores can be rough, smooth, or mixed and can show spatial structures on any axis (even distant ones). Moreover, the advantage of MULTISPATI-PCA over Wartenberg's classical MSCA is that MULTISPATI-PCA is not restricted to the case of quantitative normalised variables but

can be applied to any type of variable and any type of analysis (for example, binary variables, counts, or qualitative variables and principal component analysis, correspondence analysis, or multiple correspondence analysis).

Finally, a Monte Carlo test was used to check the statistical significance of the observed structures. This test is a multivariate permutation test against a random distribution of the values of the topsoil characteristics over the sampling sites. The test does not rely on statistical distribution hypotheses.

Computations were conducted with the "ade4" (Chessel et al., 2004) and "spdep" packages (Bivand et al., 2008) for the R statistical software package (R Development Core Team, 20084). MULTISPATI-PCA is a purely descriptive method, based solely on linear algebra and on geometrical properties. The method does not rely on any model fitting.

2.3. Additive log ratio transformation

Component analysis of particle-size distribution involves problem in that the components have a constant sum, 100%, with distributions that are curtailed at the limits of 0 and 100%. Aitchison (1986) drew attention to the matter and described how to deal with the situation using the additive log-ratio (adlr). We therefore transform the particle size data using adlr method before performing the multivariate analysis.

3. Results

3.1. MULTISPATI-PCA Axes

The Monte-Carlo permutation test of the MULTISPATI-PCA was highly significant ($p<0.005$), which showed that the spatial structures exhibited by the soil properties were indeed very strong and could not be attributed to random variations. Table 3 gives a comparison of the results of the 2 PCA in terms of the variance explained. For the "Classical PCA", the variances given in the table are equal to the eigenvalues, and the Moran's index is computed on unit-norm row scores. For the MULTISPATI-PCA analysis, the variances are the spatial variance, which is the variance of the normed MULTISPATI-PCA row scores. The Moran's index is computed on the product of these row scores with the lag vector (mean over neighbouring points). This presentation can be used to compare the loss in inertia (spatial variance vs. total variance) and the gain in spatial information (MULTISPATI-PCA Moran's *I* vs. Classical PCA Moran's *I*). For example, for the first axis, the loss of inertia in the MULTISPATI-PCA is only $(7.597–7.3838)/7.597 = 3\%$, while the gain in spatial information is $(0.42–0.396)/0.42 = 6\%$. Moreover, the gain of spatial autocorrelation (Moran's *I*) is advantageous for axes 2 and 3 (0.644 versus 0.571 and 0.503 versus 0.38 respectively) and negative for axis 4.

Fig. 1 shows the graphical display of the first four axes of the MULTISPATI-PCA and the classical PCA. Axis 1 for both methods was mainly driven by particle-size distribution relative to the clay, fine silt, and coarse sand content. As expected, the scores on the axis 1 were highly and negatively correlated with the clay content and the cation exchange capacity and were positively correlated with coarse sand. We conclude that this axis reflects a global mineral richness (in clay and cations) of the topsoil. Axis 2 appeared to be driven by the relative K and Al contents, no matter the particle-size distribution. Axis 3 was driven by the relative abundance of coarse silt. The Axis 3 scores were positively correlated with coarse silt content and negatively correlated with coarse sand content. This axis indicated samples in which the relative proportion of coarse silt in topsoil was large. We also observed a weak negative correlation between the scores of this axis and phosphorus. Axis 4 appeared driven by the relative abundance of fine sand. This axis was also slightly correlated with total K and most of samples exhibiting the highest exchangeable Na contents were located on the negative part of this axis.

3.2. Maps of the scores on MULTISPATI-PCA and classical PCA axes

Figs. 2–5 show the maps of the scores on the first four axes of the MULTISPATI-PCA and the Classical PCA. As expected, the maps of Axis 1 for the 2 methods are quite similar and nearly match the distribution of topsoil textures in France (Fig. 2). Both axes oppose very coarse sandy materials (e.g. Landes de Gascogne, Sologne, Vosges) to heavy clayey and clayey topsoils (e.g. Lorraine, Jura Mountain, Charentes). The maps of the scores on Axis 2 (Fig. 3) show a spatial distribution linked to specific parent materials, mainly derived from crystalline rocks. The relative abundance of total K and Al could be related to their mineral composition. The map of the scores on the axis 2 of the MULTISPATI-PCA is slightly smoother than the one of the classical PCA as shown by the Moran's indexes in table 3. The maps of the scores on Axis 3 (Fig. 4) exhibit large trends in the proportion of coarse silt relative to particle size fractions. This trend is highly visible in the northwest part of France, and in the extreme Southwest, where soils having very high coarse silt content have been described elsewhere

Table 3
Eigen values and variance of the principal components generated from soil properties analyses for the MULTISPATI-PCA and Classical PCA. Moran's Indexes are also provided.

MULTISPATI-PCA					Classical PCA			
Eigen values	Spatial variance	Percentage	Cumulative percentage	Moran's *I*	Variance	Percentage	Cumulative percentage	Moran's *I*
3.1016	7.3838	36.92%	36.92%	0.420	7.597	37.99%	37.99%	0.396
2.066	3.2047	16.02%	52.94%	0.645	3.362	16.80%	54.79%	0.572
0.8264	1.6511	8.26%	61.20%	0.500	1.792	8.97%	63.76%	0.378
0.6177	1.4364	7.18%	68.38%	0.430	1.549	7.74%	71.50%	0.439
0.2778	1.0267	5.13%	73.51%	0.271	1.093	5.47%	76.97%	0.311
0.2594	1.0414	5.21%	78.72%	0.249	0.919	4.59%	81.56%	0.345
0.2045	1.0383	5.19%	83.91%	0.197	0.902	4.51%	86.07%	0.294
0.1	0.6708	3.36%	87.27%	0.149	0.582	2.91%	88.98%	0.267
0.0827	0.4673	2.33%	89.60%	0.177	0.489	2.44%	91.42%	0.127
0.0554	0.4608	2.31%	91.91%	0.120	0.437	2.19%	93.61%	0.226
0.0409	0.41	2.05%	93.96%	0.100	0.379	1.90%	95.51%	0.176
0.0288	0.2773	1.38%	95.34%	0.104	0.214	1.07%	96.58%	0.143
0.0229	0.1922	0.96%	96.30%	0.119	0.154	0.77%	97.35%	0.152
0.0151	0.1768	0.89%	97.19%	0.085	0.131	0.66%	98.01%	0.126
0.0128	0.155	0.77%	97.96%	0.083	0.103	0.51%	98.52%	0.241
0.0097	0.1196	0.60%	98.56%	0.081	0.097	0.49%	99.01%	0.111
0.0051	0.0812	0.41%	98.97%	0.062	0.078	0.39%	99.40%	0.231
0.0025	0.0589	0.29%	99.26%	0.043	0.051	0.25%	99.65%	0.086
0.0014	0.0394	0.20%	99.46%	0.036	0.042	0.22%	99.87%	0.135
0.0006	0.1084	0.54%	100.00%	0.006	0.027	0.13%	100.00%	0.099

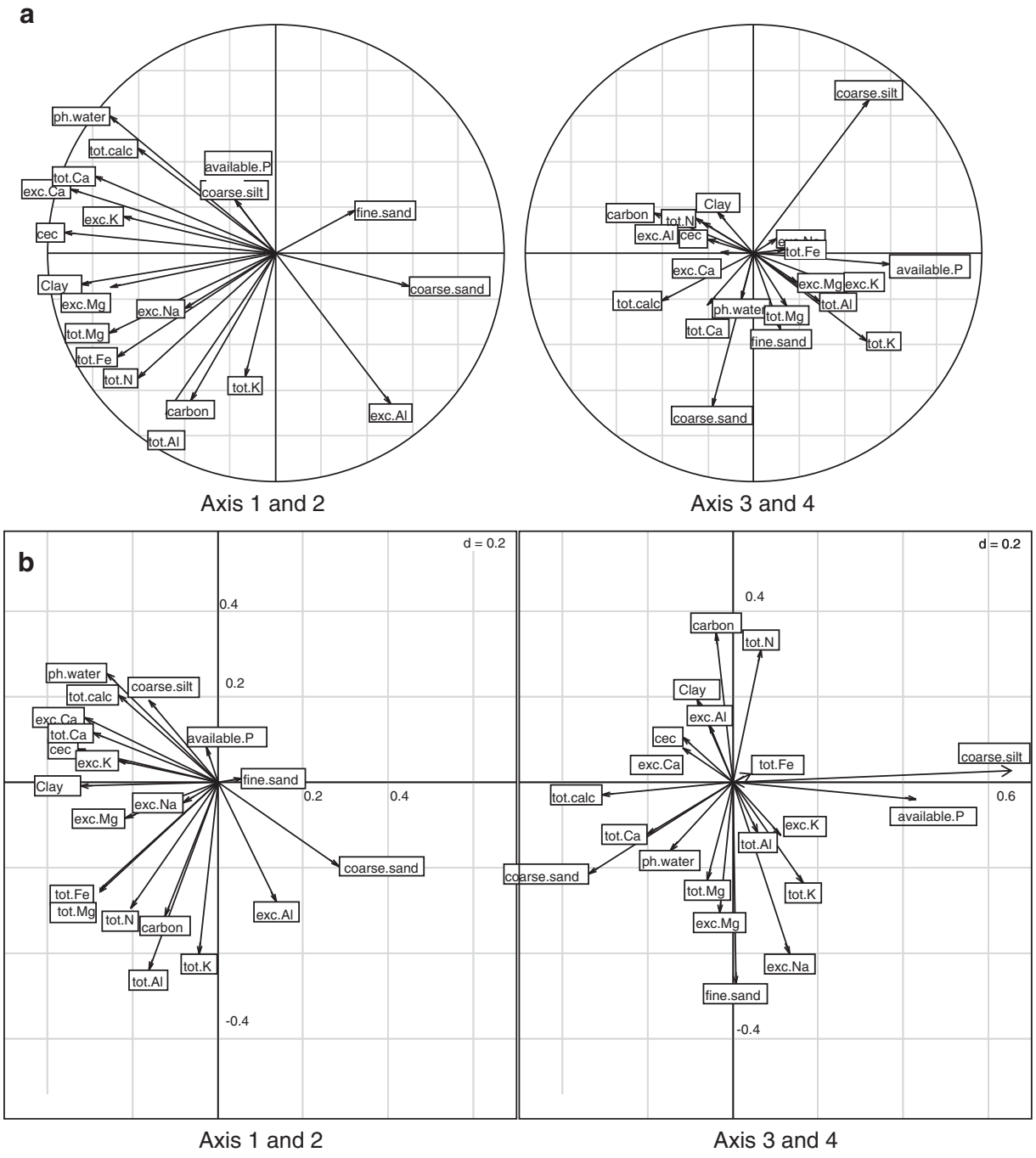

Fig. 1. Graphical display of the first four axes of the classical PCA (a) and MULTISPATI-PCA (b). Correlation between variables and principal components is presented for the classical PCA. Coefficients of variables are presented for the MULTISPATI-PCA.

(Arrouays et al., 1995; Arrouays et al., 1998). Interestingly, this trend in the northwest part of France is more visible on the MULTISPATI-PCA map. The MULTISPATI-PCA map of the scores on Axis 4 (Fig. 5) clearly shows a coastal belt along the coastlines of the Channel, the northern part of the Atlantic Ocean, and the Mediterranean Sea. This belt is quite less visible on the classical PCA map.

4. Discussion

The first two axes were expected and rather easy to interpret. At the national scale, they confirm the major influence of clay content and parent material mineralogy on many topsoil characteristics. The map of the scores on Axis 3 shows that a large part of northwest France exhibits high relative contents of coarse silt. Some of these high coarse silt contents are clearly linked to well known typical Aeolian loess deposits, such as in northern France, Brittany or Normandy (Antoine et al., 1998; Antoine et al., 2003). However, the map suggests that the influence of these Aeolian deposits on topsoil may have influenced a larger territory than what is indicated by the quaternary superficial formations maps (Antoine et al., 1998). Indeed, such a diffuse impact could not have been detected using classical geological and geomorphological observations. We have captured a very large trend that is likely to continue outside of the French borders. Indeed, the loess belt continues in Belgium, and it is well known that typical homogeneous loess occurs along the southeastern coast of England, overlying many different substrata (Antoine et al., 2003; Catt, 1985). The map of the scores on Axis 4 clearly shows a marine influence on soils likely attributable to fine-sized coversands, K and exchangeable

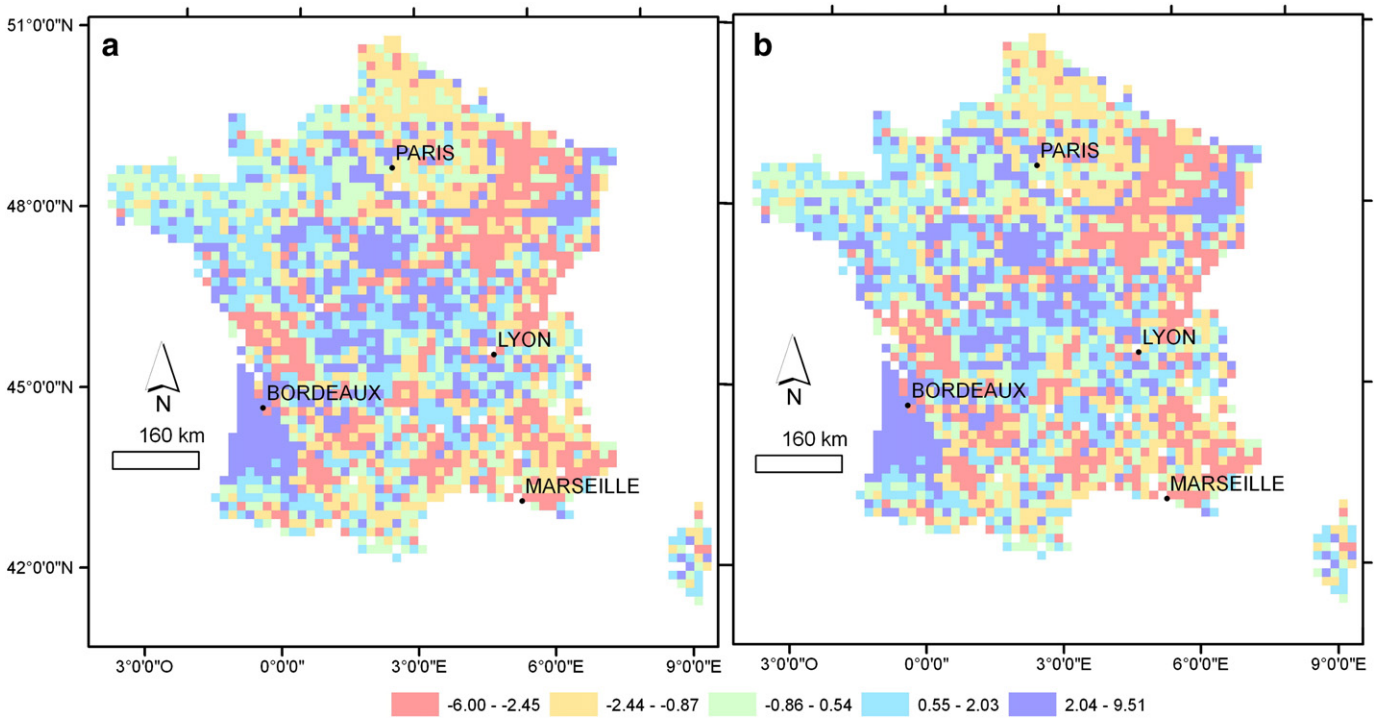

Fig. 2. Maps of the first axis of the MULTISPATI-PCA (a) and classical PCA (b).

Na. On the MULTISPACI-PCA map, a narrow belt is visible close to nearly all of the coastlines and is more pronounced close to the Mediterranean Sea. A noticeable exception is in southwest France, where the soils of the "Landes de Gascogne" are covered by Spodosols almost exclusively composed of coarse quartz sand having a very low cation exchange capacity (Augusto et al., 2006; Jolivet et al., 1997). These results show, that except for this last region, the marine influence can affect soils at a distance that may reach as far as 80 km.

The gain in spatial auto correlation using MULTISPATI-PCA is not as large as the one presented in the results of Dray et al (2008). Indeed, soil properties seem much more spatially correlated than other ecological data. As shown by Dray et al. (2008), MULTISPATI-PCA may

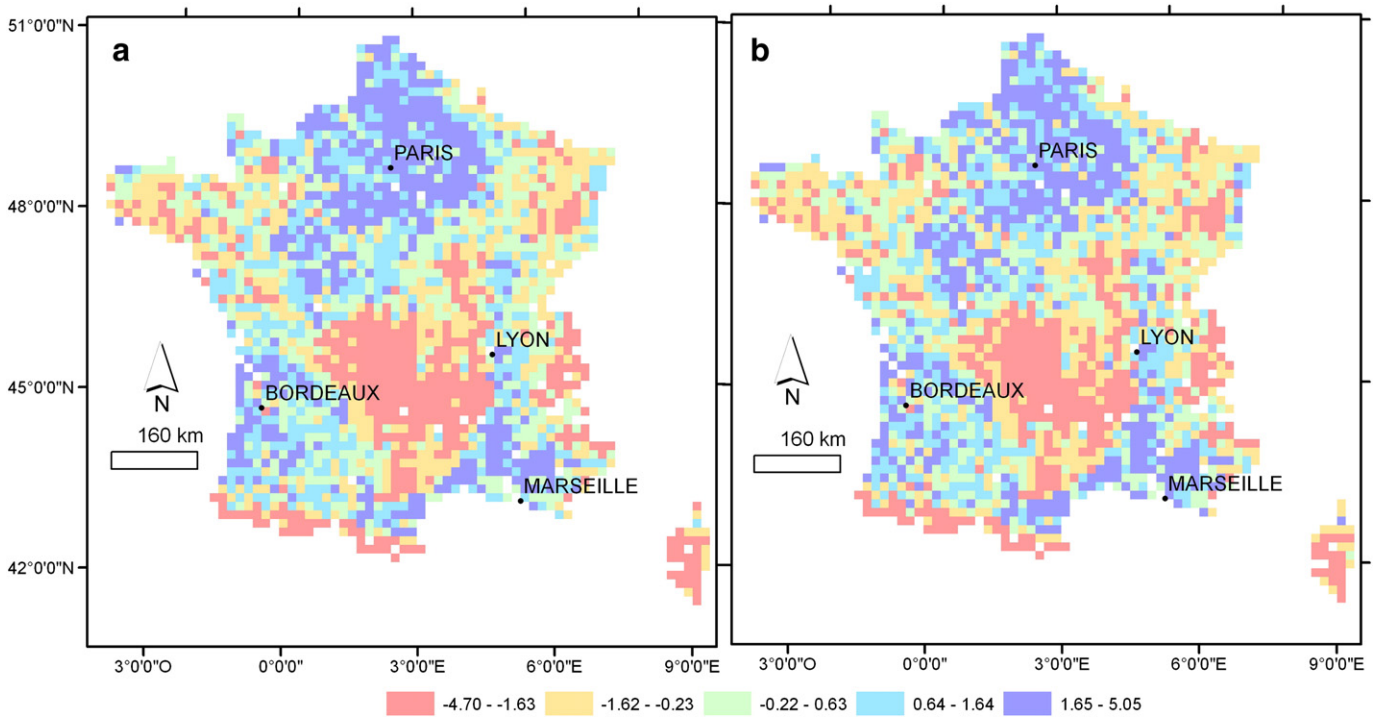

Fig. 3. Maps of the second axis of the MULTISPATI-PCA (a) and classical PCA (b).

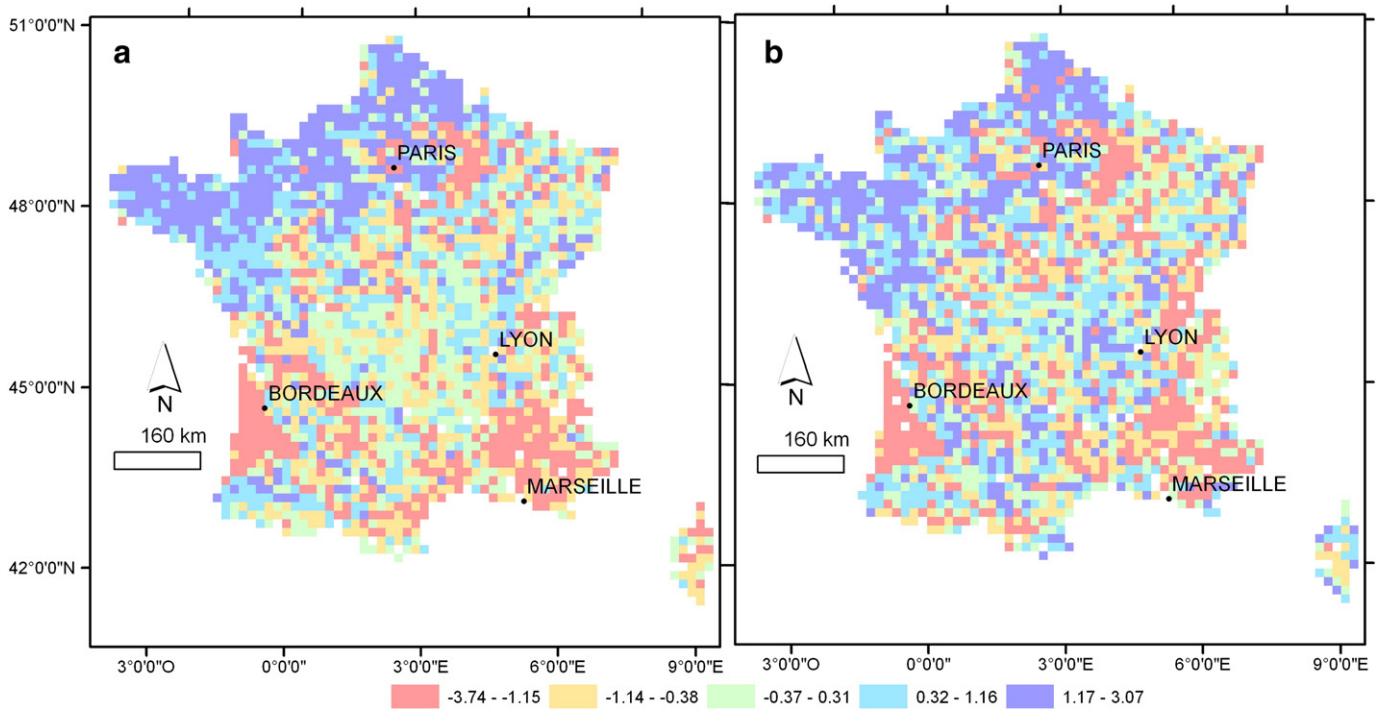

Fig. 4. Maps of the third axis of the MULTISPATI-PCA (a) and classical PCA (b).

produce negative eigen values (due to possible negative autocorrelation in the data) which is not the case for our dataset (Table 3). Yet, MULTISPATI-PCA produces perceptible advantage for pedometricians as shown by the rather easy interpretation of the axis 2 to 4 based on smoother maps. Axis 4 of the MULTISPATI-PCA which represents the marine influence on soils close to the coastlines is interesting because although this axis represents only 8 % of the variance it displays valuable inputs about soil properties distribution. Finally, MULTISPATI-PCA ensures an optimal spatial result for pedometricians who are interested in mapping the multivariate spatial patterns from their datasets.

5. Conclusion

Using the soil samples of a 16 × 16 km grid soil inventory, we performed an estimation of the distribution of the main topsoil characteristics over the entire French territory. This study shows that the spacing of the sampling grid combined with MULTISPATI-PCA

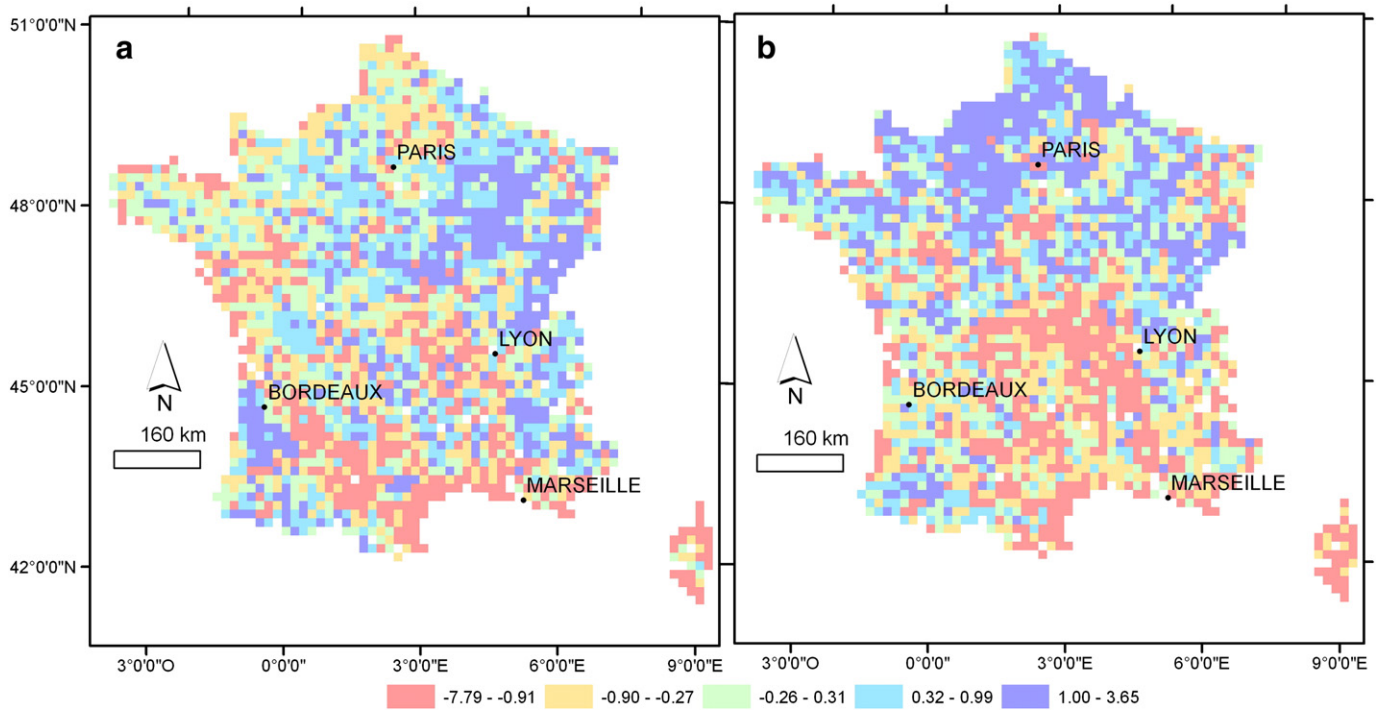

Fig. 5. Maps of the fourth axis of the MULTISPATI-PCA (a) and classical PCA (b).

allows us to detect and map large regional trends in the distribution of topsoil characteristics. Interestingly, the gradients were more visible and pronounced when using MULTISPACI-PCA rather than classical PCA. Some of the gradients that we mapped were expected and could have been drawn without using these techniques. However, the most interesting findings were the large extent of the coarse silt gradient in northwest France and the assessment of the marine influence on soils close to the coastlines.

Acknowledgements

This study was supported by a French Scientific Group of Interest on Soils: the "GIS Sol", involving the French Ministry for Ecology and Sustainable Development, the French Ministry of Agriculture, the French Agency for Energy and Environment (ADEME) , the National Institute for Agronomic Research (INRA), the Institute for Research and Development (IRD) and the National Forest Inventory (IFN). We thank all of the soil surveyors and technical assistants involved in sampling the sites. We are grateful to two anonymous referees and an associate editor of the journal for helpful and stimulating comments on this paper.

References

AFNOR, 1994. NF ISO 11464. Qualité du sol. Prétraitement des échantillons pour analyses physico-chimiques.
Aitchison, J., 1986. The Statistical Analysis of Compositional Data. Chapman & Hall. Chapman & Hall, London.
Antoine, P., Lautridou, J.P., Sommé, J., Auguste, P., Auffret, J.P., Baize, S., Clet-Pellerin, M., Coutard, J.P., Dewolf, Y., Dugué, O., Joly, F., Laignel, B., Laurent, M., Lavollé, M., Levret, P., Lecolle, F., Lefebvre, D., Limondin-Lozouet, N., Locht, J.L., Munaut, A.V., Paris, P., Petit, C., Rousseau, D.D., 1998. Le Quaternaire de la France au Nord-Ouest: limites et corrélations. Quaternaire. Quatern. 9 (3), 227–241.
Antoine, P., Catt, J., Lautridou, J.P., Sommé, J., 2003. The loess and coversands of northern France and southern England. J. Quatern. Sci. 18 (3–4), 309–3118.
Arrouays, D., Vion, I., Kicin, J.L., 1995. Spatial analysis and modeling of topsoil carbon storage in temperate forest humic loamy soil of France. Soil Sci. 159 (3).
Arrouays, D., Daroussin, J., Kicin, J.L., Hassika, P., 1998. Improving topsoil carbon storage prediction using a digital elevation model in temperature forest soils of France. Soil Sci. 163 (2).
Arrouays, D., Deslais, W., Badeau, V., 2001. The carbon content of topsoil and its geographical distribution in France. Soil Use Manag 17 (1), 7–11.
Arrouays, D., Jolivet, C., Boulonne, L., Bodineau, G., Saby, N., Grolleau, E., 2002. A new initiative in France: a multi-institutional soil quality monitoring network. C R Acad Agric Fr 88 (5), 93–105.
Arrouays, D., Jolivet, C., Boulonne, L., Bodineau, G., Saby, N., Grolleau, E., 2003. Le Réseau de Mesures de la Qualité des Sols (RMQS) de France. Etude Gestion Sols 10 (4), 241–250.
Arrouays, D., Saby, N., Walter, C., Lemercier, B., Schvartz, C., 2006. Relationships between particle-size distribution and organic carbon in French arable topsoils. Soil Use Manag 22, 48–51.
Arrouays, D., Morvan, X., Saby, N.P.A., Richer de Forges, A., Le Bas, C., Bellamy, P.H., Berengi Uveges, J., Freudenschuss, A., Jones, R.J.A., Kibblewhite, M.G., Simota, C., Verdoodt, A., Verheijen, F.G.A., 2008. Environmental Assessment of Soil for Monitoring: Volume IIa Inventory & Monitoring. EUR 23490 EN/2A, Office for the Official Publications of the European Communities, Luxembourg.
Atteia, O., Dubois, J., Webster, R., 1994. Geostatistical analysis of soil contamination in the Swiss Jura. Environ. Environ Pollut 86 (3), 315–327.
Augusto, L., Badeau, V., Arrouays, D., Trichet, P., Flot, J.L., Jolivet, C., Merzeau, D., 2006. Caractérisation physico-chimique des sols à l'échelle d'une région naturelle à partir d'une compilation de données, exemple des sols du massif forestier landais. Etude Gestion Sols 13 (1), 7–22.
Bivand, R., Anselin, L., Assunçaõ, R., Berke, O., Bernat, A., Carvalho, M., Chun, Y., Dormann, C., Dray, S., Halbersma, R., Krainski, E., Lewin-Koh, N., Li, H., Ma, J., Millo, G., Mueller, W., Ono, H., Peres-Neto, P., Reder, M., Tiefelsdorf, M., Yu, D., 2008. spdep: Spatial dependence: weighting schemes, statistics and models. R package version 0.4-29. http://cran.at.r-project.org/web/packages/spdep.
Boruvka, L., Mládková, L., Penízek, V., Drábek, O., Vasát, R., 2007. Forest soil acidification assessment using principal component analysis and geostatistics. Geoderma 140 (4), 374.
Bourennane, H., Salvador-Blanes, S., Cornu, S., King, D., 2003. Scale of spatial dependence between chemical properties of topsoil and subsoil over a geologically contrasted area (Massif central, France). Geoderma 112 (3–4), 235–251.
Catt, J.A., 1985. Soil particle size distribution and mineralogy as indicators of pedogenic and geomorphic history: examples from the loessial soils of England and Wales. In: Richard, K., Arnett, R., Ellis, S. (Eds.), Geomorphology and Soils. G. Allen and Unwin, London.
Chessel, D., Dufour, A.B., Thioulouse, J., 2004. The ade4 package — I: one-table methods. R news 4 (1), 5–10.
Dray, S., Dufour, A.B., 2007. The ade4 package: implementing the duality diagram for ecologists. J. Stat. Softw. 22 (4), 1–20.
Dray, S., Said, S., Debias, F., 2008. Spatial ordination of vegetation data using a generalization of Wartenberg's multivariate spatial correlation. J. Veg. Sci. 19, 45–56.
Farrell, E.P., 1995. Atmospheric deposition in maritime environments and its impact on terrestrial ecosystems. Water Air Soil Pollut. 85, 123–130.
Holmes, S., 2006. Multivariate analysis: The French way. In: Nolan, D., Speed, T. (Eds.), Festschrift for David Freedman. IMS, Beachwood, OH, pp. 1–14.
Jamagne, M., Lautridou, J.P., Sommé, J., 1981. Préliminaire à une synthèse sur les variations sédimentologiques des loess de la France du Nord-Ouest dans leur cadre géographique et paléogéographique. Bull. Soc. Geol. Fr. (7), 143–147 XXIII.
Jolivet, C., Arrouays, D., Andreux, F., Leveque, J., 1997. Soil organic carbon dynamics in cleared temperate forest Spodosols converted to maize cropping. Plant Soil 191 (2), 225–231.
Jolivet, C., Arrouays, D., Boulonne, L., Ratie, C., Saby, N., 2006. Le Réseau de Mesures de la Qualité des Sols de France (RMQS), Etat d'avancement et premiers résultats. Etude Gestion Sols 13 (3), 149–164.
Lautridou, J.P., Monnier, J.L., Morzadec-Kerfourn, M.T., Sommé, J., Tuffreau, A., 1986. The Pleistocene of Northern France. Quatern. Sci. Rev. 387–393.
Lebret, P., Lautridou, J.P., 1991. The loess of West Europe. Geo J. 24 (2), 151–186.
Odlare, M., Svensson, K., Pell, M., 2005. Near infrared reflectance spectroscopy for assessment of spatial soil variation in an agricultural field. Geoderma 126 (3–4), 193–202.
Oliver, M.A., Webster, R., Edwards, K.J., Whittington, G., 1997. Multivariate, autocorrelation and spectral analyses of a pollen profile from Scotland and evidence for periodicity. Rev. Of Palaeobot. And Palynol. 96 (1–2), 121–144.
R Development Core Team, 2008. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Saby, N., Arrouays, D., Boulonne, L., Jolivet, C., Pochot, A., 2006. Geostatistical assessment of pb in soil around Paris, France. Sci. Total Environ. 367 (1), 212–221.
Saby, N.P.A., Thioulouse, J., Jolivet, C.C., Ratié, C., Boulonne, L., Bispo, A., Arrouays, D., 2009. Multivariate analysis of the spatial patterns of 8 trace elements using the French soil monitoring network data. Sci. Total Environ. 407 (21), 5644–5652.
Satapathy, D.R., Salve, P.R., Katpatal, Y.B., 2009. Spatial distribution of metals in ground/surface waters in the Chandrapur district (Central India) and their plausible sources. Environ. Geol. 56 (7), 1323–1352.
Torre, F., Chessel, D., 1994. Co-structure de deux tableaux totalement appariés. Rev. Statistique Appl. 43, 109–121.
Wartenberg, D., 1985. Multivariate spatial correlation: a method for exploratory geographical analysis. Geogr. Anal. 17, 263–283.
Webster, R., Boag, B., 1992. Geostatistical analysis of cyst nematodes in the soil. J. Soil Sci. 43, 583–595.
Webster, R., Oliver, M., 2007. Geostatistics for environmental scientists, 2nd Edition. Geostatistics for environmental scientists, 2nd Edition. John Wiley & Sons Ltd, Chichester UK.
Webster, R., Atteia, O., Dubois, J.P., 1994. Coregionalization of trace metals in the soil in the Swiss Jura. Eur. J. Soil Sci. 45 (2), 205–218.
Zhang, Z.H., 2003. Impact of seasalt deposition on acid soils in maritime regions. Pedosphere 13, 375–380.