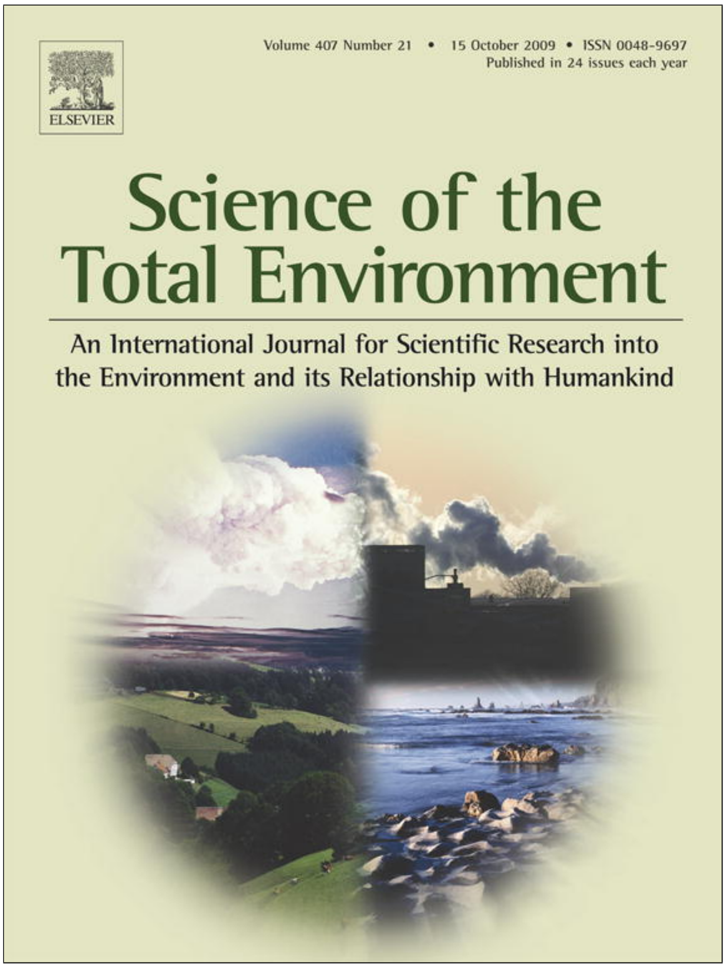
Volume 407 Number 21 • 15 October 2009 • ISSN 0048-9697
Published in 24 issues each year
ELSEVIER
Science of the Total Environment
An International Journal for Scientific Research into the Environment and its Relationship with Humankind

Science of the Total Environment 407 (2009) 5644–5652

Contents lists available at ScienceDirect

Science of the Total Environment

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

Multivariate analysis of the spatial patterns of 8 trace elements using the French soil monitoring network data

N.P.A. Saby [a,*], J. Thioulouse [b], C.C. Jolivet [a], C. Ratié [a], L. Boulonne [a], A. Bispo [c], D. Arrouays [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, Villeurbanne F-69622 cedex, France
[c] ADEME, Département Animation de la Recherche Déchets et Sols, 20, Avenue du Grésillé, BP 90406, 49004, Angers cedex 01, France

ARTICLE INFO

Article history:
Received 8 April 2009
Received in revised form 1 July 2009
Accepted 3 July 2009
Available online 31 July 2009

Keywords:
Soil
Trace elements
Mapping
Multivariate analysis
Geostatistics
France

ABSTRACT

Geostatistical and spatially constrained multivariate analysis methods (MULTISPATI-PCA) have been applied at the scale of France to differentiate the influence of natural background from the pollution due to human activities on the content of 8 trace elements in the topsoil. The results of MULTISPATI-PCA evidence strong spatial structures attributed to different natural and artificial processes. The first axis can be interpreted as an axis of global richness in trace elements. Axis 2 reflects geochemical anomalies in Tl and Pb. Axis 3 exhibits on one hand natural pedogeogenic anomalies and on the other hand, it shows high values attributable to anthropogenic contamination. Finally, axis 4 is driven by anthropogenic copper contamination. At the French territory scale, we show that the main factors controlling trace elements distribution in the topsoil are soil texture, variations in parent material geology and weathering, and various anthropogenic sources.

1. Introduction

Concern about increasing levels of trace elements (TE) in the European soils has led to the development and implementation of numerous national programmes to determine the baseline levels of TE in the Earth's surface (see a review from Morvan et al., 2008). The origin of TE in soils can be studied using various techniques, such as isotope composition (Bacon et al., 1996; Erel et al., 1997), analysis of spatial distribution (Atteia et al., 1994; Facchinelli et al., 2001; McGrath et al., 2004), statistical relationships with other soil characteristics (Basta et al., 1993), and topsoil to subsoil ratios (TSR) of TE concentrations defined as "relative topsoil enhancement" (RTE) (Colbourn and Thornton, 1978; Baize and Sterckeman, 2001; Saby et al., 2006). Multivariate techniques, that can include interpolation, have been used by numerous authors to reduce the dimensionality of datasets and to identify the processes governing the TE spatial distribution (Korre, 1999; Zhang et al., 1999; Rawlins et al., 2003; Lee et al., 2006; Zhao et al., 2007; Imrie et al., 2008; Wannaz et al., 2008).

In this paper, we used data from the French national soil monitoring network (Arrouays et al., 2002) to study the spatial distribution and origin of 8 TE in soil at the national scale (Cd, Co, Cr, Cu, Ni, Pb, Tl, Zn). TE content measurements from total and partial extraction were used because they can bring different insights in TE spatial distribution and origin. We use TE data alone to assess if their spatial distributions and correlations, without confrontation with additional data, can provide an interesting insight on their origin.

Detecting and mapping regional trends in TE distribution over the entire French metropolitan territory could have been done by multivariate geostatistical techniques such as Factorial Kriging Analysis (FKA) (Goovaerts, 1992; Bocchi et al., 2000; Bourennane et al., 2003; Imrie et al., 2008). This technique is suitable to provide quantitative measures of complex interactions between soil properties but it is strongly dependent on the goodness-of-fit (in least square sense) of the linear model of co-regionalization. Alternatively, we use a spatially constrained multivariate analysis method (MULTISPATI-PCA, Dray et al. (2008)), which is a generalization of Wartenberg's (1985) Multivariate Spatial Correlation Analysis (MSCA). This technique implies a compromise between the relations among many variables (multivariate analysis) and their spatial structure (autocorrelation). Moreover, MULTISPATI-PCA was combined with geostatistics to map effective displays of spatial patterns (Thioulouse et al., 1995) of the TE relationships in French soil and to analyse their origin. MULTISPATI-PCA is a purely descriptive method, based solely on linear algebra and on geometrical properties. It does not rely on any model fitting. The geostatistical interpolation technique (kriging) used to draw the geographical maps is used only as a graphical technique here, and this has no consequence on the validity of the MULTISPATI-

* Corresponding author. Tel.: +33 238414822; fax: +33 238417869.
E-mail address: Nicolas.Saby@orleans.inra.fr (N.P.A. Saby).

0048-9697/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.scitotenv.2009.07.002

PCA results. At the French territory scale, we show that the main factors controlling TE distribution in the topsoil are soil texture, variations in parent material geology and weathering, and various anthropogenic sources.

2. Materials and methods

2.1. Study area

The French National Soil Quality Monitoring Network "Réseau de Mesures de la Qualité des Sols" consists of observations of soil properties on a 16 km regular grid across the French metropolitan territory (550,000 km²). This network was designed to monitor soil evolutions and particularly to identify diffuse contamination either due to atmospheric deposition of trace elements on soils or to agricultural practices (e.g. fertilizers, sludge amendments, inorganic pesticides). The complete inventory will consist of 2200 sites but in this study we use measurements of TEs from the 2059 sites analysed at present (Fig. 1).

The sites are 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 is 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 is not always possible to find an alternative location. All land cover types are present in the dataset, except industrial sites which are 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). Topsoil sampled from 0 to 30 cm was chosen because it corresponds in France to the maximal depths affected by ploughing and because it is a quite conventional thickness to report on topsoil properties (e.g. Arrouays et al., 2001; Arrouays et al., 2008).

On one hand, total concentrations of 8 TE (Cd, Co, Cr, Cu, Ni, Pb, Tl, Zn) were determined by inductively coupled plasma mass spectroscopy after dissolution with hydrofluoric and perchloric acids (AFNOR, 2001). On the second hand, 6 TE (Cd, Cr, Cu, Pb, Ni, Zn) were extracted by 0.05 M EDTA solution (ammonium salt) at pH = 7.0. The extraction was performed with a solid/solution ratio of 1/10. 2.5 g of soil, air-dried and sieved at 2 mm, and 25 ml of 0.05 M EDTA solution at pH 7.0 were put in a taped flask for a shake step during 1 h at 20 °C. Then, the mixture was centrifuged at 3000 laps per minute for 10 min. After centrifugation, the extract was filtered through 2 µm filter and transferred in a taped flask before analysis of Cd, Cr, Cu, Ni, Pb and Zn using ICP-AES. Analysis of soil extracts were calibrated using standard solutions made with the extracting agent.

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

2.2. MULTISPATI-PCA

The matrix of the TE contents was analysed by MULTISPATI-PCA (Dray et al., 2008), which is a generalization of Multivariate Spatial Correlation Analysis (MSCA, Wartenberg 1985). This method allowed taking into account the spatial position of sampling sites through a neighbouring relationship between sites (the one-step chess queen's move).

As explained in Dray et al. (2008), the MULTISPATI analysis introduces a spatial weighting matrix **W** in the Principal Component Analysis (PCA) of the data matrix **X**. Here **X** is the matrix of TE contents: it has n rows (soil samples) and p columns (TE). **W** is the row-sum standardized connectivity matrix: if $\mathbf{C} = [c_{ij}]$ is the connectivity matrix (indicating the strength of interactions between units i and j), then $\mathbf{W} = \left[\mathbf{c}_{ij} / \sum_{j=1}^{n} \mathbf{c}_{ij}\right]$. Let **D** be a scalar product of R^n, and let **Q** be a scalar product of R^p. (**X**,**Q**,**D**) is the statistical triplet associated to the PCA of **X**, and the MULTISPATI analysis is the co-inertia analysis (Dray et al., 2003) between **X** and the lag matrix $\tilde{\mathbf{X}} = \mathbf{WX}$. The lag matrix $\tilde{\mathbf{X}}$ is composed of the averages of neighbouring values weighted by the spatial connection

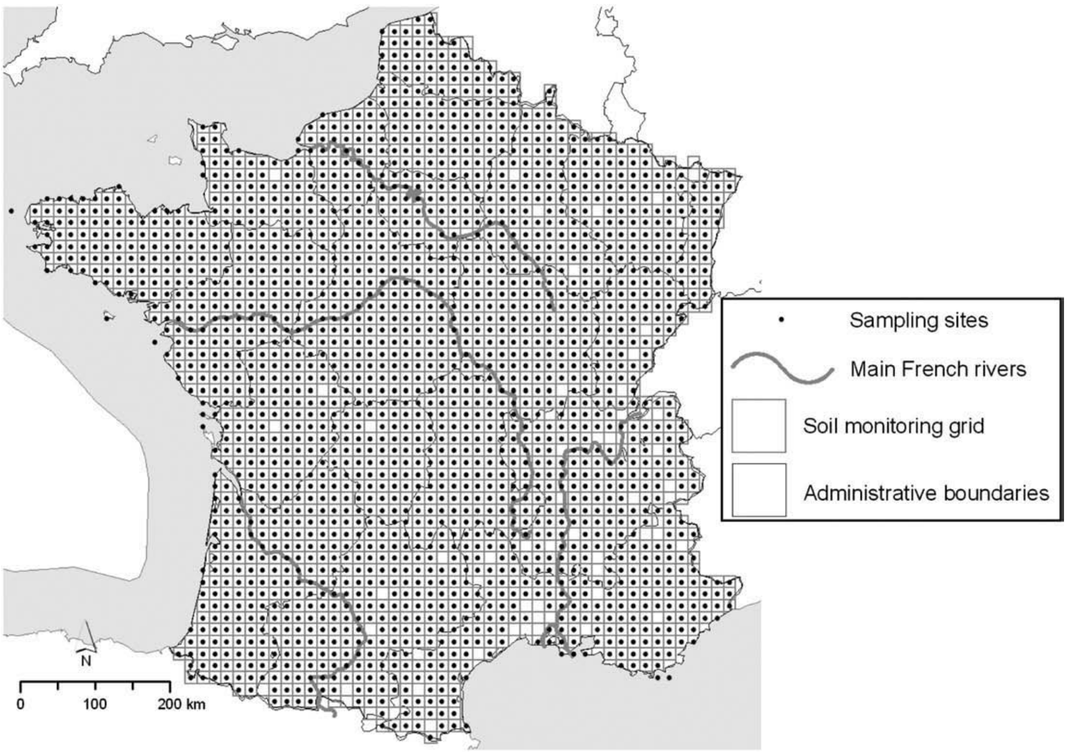

Fig. 1. Location of the study and of the sampling sites.

matrix (this means that only the neighbouring points are taken into account). MULTISPATI maximizes the scalar product between a linear combination of the original variables ($\mathbf{a}_1=\mathbf{XQu}_1$) and a linear combination of the lagged variables ($\tilde{\mathbf{a}}_1=\mathbf{WXQu}_1$). In practice, it is necessary to diagonalize the Q-symmetric matrix $\mathbf{H}=(1/2)(\mathbf{X}^t(\mathbf{W}^t\mathbf{D}+\mathbf{DW})\mathbf{XQ})$ instead of matrix $\mathbf{X}^t\mathbf{DWXQ}$, which is not symmetric.

The advantage of MULTISPATI over PCA is that MULTISPATI sample scores maximize spatial autocorrelation between sites, while conventional PCA scores maximize the inertia (i.e., the sum of variances). MULTISPATI 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 over Wartenberg's classical Multivariate Spatial Correlation Analysis (MSCA, Wartenberg, 1985) is that MULTISPATI is not restricted to the case of quantitative normalized 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 TE over the sampling sites. It does not rely on statistical distribution hypotheses.

Computations were conducted with the "ade4" ("multispati" function, Chessel et al., 2004) and "spdep" packages (Bivand et al., 2008) for the R statistical software (R Development Core Team, 20084).

2.3. Geostatistical interpolation

Kriging or geostatistical interpolation aims at predicting the unknown value of a variable Z at a non-observed location $\mathbf{x}_i$ using the values z_i at surrounding locations. To do this, we use a Stochastic Function (SF) as a model of spatial variation so that the actual but unknown value $z(\mathbf{x}_i)$ and the values at the surrounding location are spatially dependent random variables. We assume that $\mathbf{z}$ is a realisation of a Gaussian random function with a covariance matrix $\mathbf{V}$. If the assumption of Gaussian random realisation is not plausible for a particular data set then that data set should be transformed. We apply the Box–Cox transform (Box and Cox, 1964).

$$z^* = \begin{cases} \log(z) & \text{if } t = 0 \\ \dfrac{z^t-1}{t} & \text{otherwise,} \end{cases} \tag{1}$$

where t is the parameter of the transform.

The elements of $\mathbf{V}$ are expressed as function of the distance separating two observations ($\mathbf{h}$). In the geostatistical literature this function is commonly expressed in terms of the variogram

$$\gamma(\mathbf{h}) = 0.5\mathrm{E}[\{Z(\mathbf{x})-Z(\mathbf{x}+\mathbf{h})\}^2].$$

Full details are given in Webster et al. (2007) for the calculation of $\mathbf{V}$. A number of variogram functions have been suggested to ensure that $\mathbf{V}$ is positive and definite. We retained the Matérn function which has a smoothness parameter ν that gives the flexibility for modelling the spatial covariance, particularly in the decay of the function for small $\mathbf{h}$. When ν is small the spatial process is rough whereas for large ν it is smooth. For particular values of ν the Matérn model is equivalent to other suggested variogram models. Minasny and McBratney (2005) describe different forms the Matérn function can take in greater detail.

The validity of the geostatistical model fitted by ordinary least square may be confirmed by leave-one-out cross validation. For each sampling site location $i=1,\ldots, n$, the value of the property at site $\mathbf{x}_i$ is predicted by ordinary kriging upon $z^*_{(-i)}$, the vector of observations excluding z_i^*. The statistic

$$\theta_i = \frac{\left\{z_i^* - \hat{z}^*_{(-i)}\right\}^2}{\sigma^2_{(-i)}},$$

where $z^*_{(-i)}$ and $\sigma^2_{(-i)}$ denote the kriging prediction and kriging variance at $\mathbf{x}_i$ when z_i^* is omitted from the transformed observation vector, is calculated. If the fitted model is a valid representation of the spatial variation of the soil property then $\boldsymbol{\theta}=(\theta_1\ldots, \theta_n)$ has a χ^2 distribution with mean $\bar{\boldsymbol{\theta}}=1.0$ and median $\tilde{\boldsymbol{\theta}}=0.455$ (Lark, 2002).

At some sites, the interpolated map may not approximate the behaviour of the property due to local contamination by secondary process, also called contaminated process or quasi-point process. Nevertheless, metal concentrations in the soil may also show variation that is a composite of a continuous background and a quasi-point process. The former represents the natural origin of the metal content and diffuse sources of pollution; the latter represents point sources of pollution. To solve this problem, the effects of the background and of the local contamination processes must be separated. Lark (2000) describes three robust variograms. These are designed to estimate the variogram of a background process in the presence of contamination. These include the Dowd variogram (Dowd, 1984). If an appropriate model of the background process is fitted by a robust variogram estimator $\bar{\boldsymbol{\theta}}$ is greater than 1.0 due to large θ_i at the contaminated sites. However $\tilde{\boldsymbol{\theta}}$ will be close to 0.455 because the median is a more robust statistic to outliers that may be present. The mean and the median

Table 1
Descriptive statistics of the trace element metals content in soils.

	Mean	Median	FOREGS median	Standard deviation	10% percentile	90% percentile	FOREGS 90% percentile	Kurtosis	Skewness	Kurtosis (log(x))	Skewness log(x)	t (Box–Cox parameter)	Kurtosis (Box–Cox)	Skewness (Box–Cox)
cd_tot	0.294	0.190	0.284	0.365	0.06	0.61	0.48	45.137	5.165	1.268	−0.395	0.096	1.053	0.032
co_tot	10.328	8.790	7.78	8.515	2.71	18.22	19.7	22.840	3.494	12.063	−3.264	0.440	2.520	0.169
cr_tot	54.530	47.330	60	85.003	18.70	86.52	122.0	821.755	25.416	3.999	−0.911	0.173	7.145	0.216
cu_tot	19.189	13.400	13	26.640	4.76	33.65	34.0	131.490	9.026	17.886	−3.058	0.246	4.541	0.366
ni_tot	24.685	18.600	18	49.309	5.69	44.14	49.8	560.808	21.116	17.706	−3.339	0.277	9.138	0.553
pb_tot	32.478	27.590	22.6	26.724	16.53	48.94	51.1	156.167	9.404	3.441	0.160	−0.044	3.323	−0.026
tl_tot	0.673	0.534	0.66	0.630	0.27	1.19	1.38	235.934	10.863	2.894	−0.349	0.088	2.524	0.048
zn_tot	72.069	61.400	52	58.049	25.64	125.00	111.0	102.337	6.805	36.868	−5.077	0.384	4.936	0.383
cd_ext	0.155	0.099	–	0.201	0.04	0.30	–	47.399	5.492	0.832	0.045	−0.013	0.818	−0.003
cr_ext	0.150	0.110	–	0.183	0.04	0.29	–	153.241	9.205	1.037	−0.394	0.109	1.150	0.033
cu_ext	4.752	2.190	–	12.971	0.53	7.36	–	281.921	13.778	1.223	0.255	−0.052	1.055	−0.021
ni_ext	1.200	0.720	–	2.591	0.25	2.36	–	486.948	19.303	3.606	−0.584	0.093	2.719	0.081
pb_ext	7.423	5.570	–	9.021	2.54	13.04	–	120.678	8.953	4.955	−0.064	0.011	4.601	0.012
zn_ext	3.155	2.080	–	5.803	0.88	5.57	–	290.794	14.260	1.911	0.565	−0.160	1.189	−0.047

FOREGS medians correspond to the published statistics from FOREGS database (Salminen et al., 2005).

Table 2
Linear correlation matrix of the Box–Cox transformed data.

	cd_tot	co_tot	cr_tot	cu_tot	ni_tot	pb_tot	tl_tot	zn_tot	cd_ext	cr_ext	cu_ext	ni_ext	pb_ext	zn_ext
cd_tot	1.000	**0.490**	**0.533**	**0.442**	**0.600**	**0.414**	**0.260**	**0.716**	**0.948**	**−0.057**	**0.465**	**0.458**	**0.339**	**0.548**
co_tot	**0.490**	1.000	**0.830**	**0.578**	**0.867**	**0.391**	**0.317**	**0.700**	**0.421**	**−0.103**	**0.447**	**0.631**	**0.203**	**0.245**
cr_tot	**0.533**	**0.830**	1.000	**0.579**	**0.928**	**0.352**	**0.297**	**0.717**	**0.486**	**0.114**	**0.476**	**0.634**	**0.230**	**0.290**
cu_tot	**0.442**	**0.578**	**0.579**	1.000	**0.613**	**0.381**	**0.285**	**0.606**	**0.394**	**−0.083**	**0.852**	**0.381**	**0.241**	**0.390**
ni_tot	**0.600**	**0.867**	**0.928**	**0.613**	1.000	**0.370**	**0.319**	**0.762**	**0.540**	0.031	**0.501**	**0.678**	**0.242**	**0.324**
pb_tot	**0.414**	**0.391**	**0.352**	**0.381**	**0.370**	1.000	**0.696**	**0.607**	**0.439**	**0.119**	**0.313**	**0.309**	**0.724**	**0.443**
tl_tot	**0.260**	**0.317**	**0.297**	**0.285**	**0.319**	**0.696**	1.000	**0.517**	**0.268**	**0.068**	**0.168**	**0.171**	**0.350**	**0.211**
zn_tot	**0.716**	**0.700**	**0.717**	**0.606**	**0.762**	**0.607**	**0.517**	1.000	**0.668**	0.008	**0.479**	**0.507**	**0.413**	**0.543**
cd_ext	**0.948**	**0.421**	**0.486**	**0.394**	**0.540**	**0.439**	**0.268**	**0.668**	1.000	0.009	**0.473**	**0.519**	**0.433**	**0.615**
cr_ext	**−0.057**	**−0.103**	**0.114**	**−0.083**	0.031	**0.119**	**0.068**	0.008	0.009	1.000	**−0.115**	**0.073**	**0.274**	**0.199**
cu_ext	**0.465**	**0.447**	**0.476**	**0.852**	**0.501**	**0.313**	**0.168**	**0.479**	**0.473**	**−0.115**	1.000	**0.477**	**0.270**	**0.490**
ni_ext	**0.458**	**0.631**	**0.634**	**0.381**	**0.678**	**0.309**	**0.171**	**0.507**	**0.519**	**0.073**	**0.477**	1.000	**0.397**	**0.448**
pb_ext	**0.339**	**0.203**	**0.230**	**0.241**	**0.242**	**0.724**	**0.350**	**0.413**	**0.433**	**0.274**	**0.270**	**0.397**	1.000	**0.499**
zn_ext	**0.548**	**0.245**	**0.290**	**0.390**	**0.324**	**0.443**	**0.211**	**0.543**	**0.615**	**0.199**	**0.490**	**0.448**	**0.499**	1.000

Bold print indicates significant linear correlation coefficients at 1% level.

values of θ were also calculated for 1000 simulated realisations of the fitted model to determine the 90% confidence limits. We used robust variograms to remove the effect of local point-source contamination (i.e. couple of isolated very high values obviously linked to very local anomalies) on variance. From the statistical point of view, a robust estimator is used to identify background variation

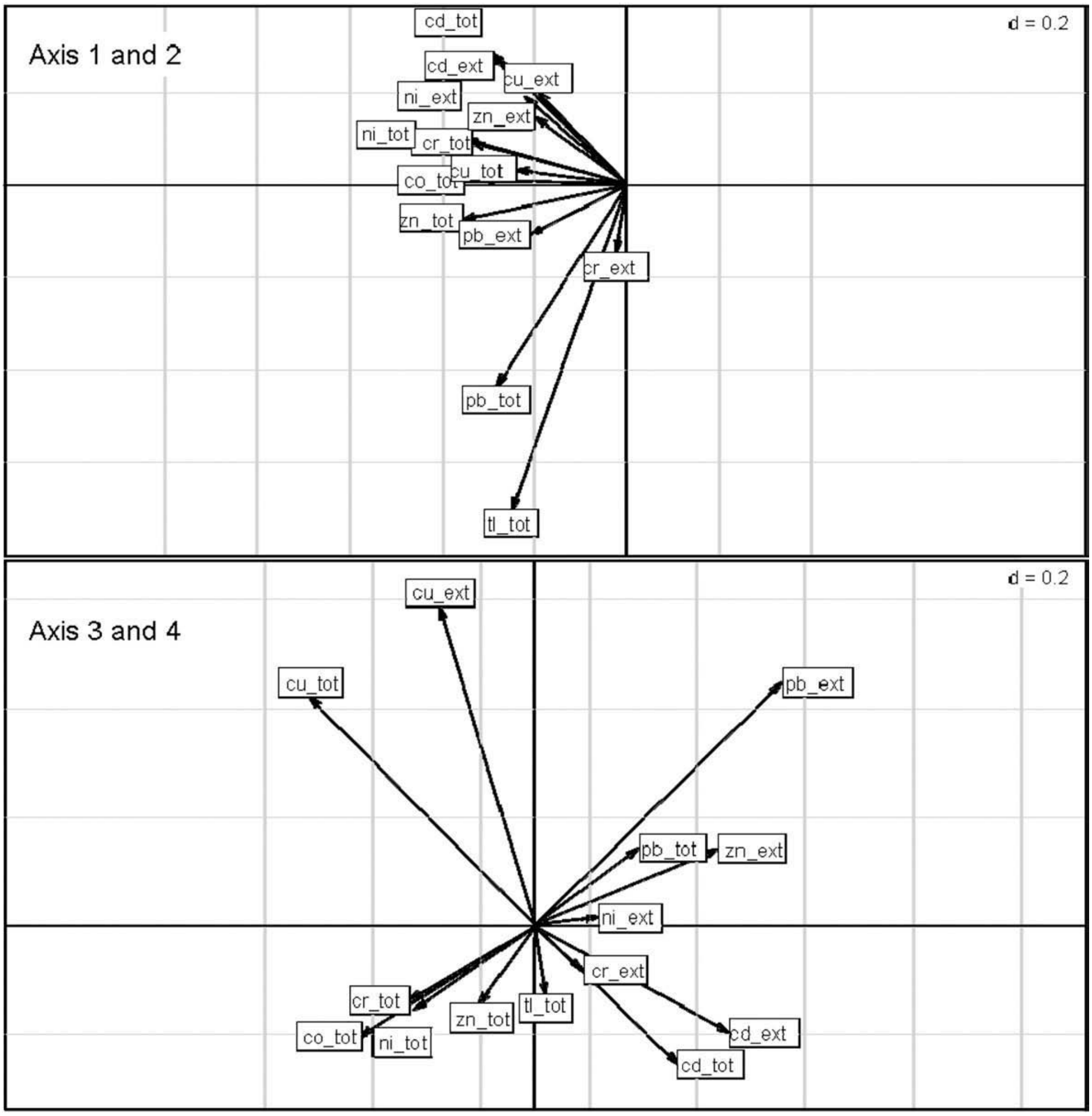

Fig. 2. Graphical display of the first four axes of the MULTISPATI-PCA analysis. In each of the 2 graphs, the labels corresponding to the 14 columns of the data table are placed in the factorial map of axes 1 and 2 and, 3 and 4. The scale (d) is the distance between vertical lines, and the thicker vertical line is the origin ($x=0$ or $y=0$).

of a soil property with the minimum influence of quasi-point process.

The spatial analysis GeoR package (Ribiero and Diggle, 2001) was used for geostatistical interpolation.

2.3.1. Spatial analysis algorithm

A MULTISPATI-PCA was calculated on the Box–Cox transformed data of the TE measurements. The scores of the sites obtained from the first four axes of global MULTISPATI-PCA were then interpolated by robust geostatistic and then geographically mapped to resolve the distribution patterns of TE variability. Box–Cox transformation ensured that both methods performed well.

3. Results

3.1. TE contents

The median observed values for total TE (Table 1) are close to those observed by the FOREGS project (Salminen et al., 2005) in European topsoils. 90% percentile values are also very similar, except for Chromium for which most of the highest values in Europe were observed outside France (http://www.gsf.fi/publ/foregsatlas/maps/Topsoil/t_xrf_cr_edit.pdf). The ranges of observed values (Table 1) are also of the same order of magnitude as those observed by Baize (2007) using a different dataset on the French territory. For the total TE for which national statistics are available in both databases, the median values are very close [i.e. in bold median value for this study, in italic median value from Baize (2007), Cd **0.19**–*0.30*; Cr **47.33**–*3.6*; Cu **13.4**–*13.8*; Ni **18.6**–*24.1*; Pb **27.59**–*25.6*; Zn **61.4**–*59.0*]. However, for some elements, the higher maximal values observed in the database from Baize (2007) indicate the presence of some highly contaminated sites. For Cd, the difference between the two median values is likely attributable to an over-representation of some soils developed on Jurassic calcareous in the database from Baize (2007).

3.2. MULTISPATI-PCA

Table 1 shows that the Box–Cox transformation has removed most of the skewness of the raw data although log transform performs worst. Nearly all TE contents, both in their total and EDTA extractable forms, were positively correlated (Table 2), except for extractable Cr that appeared to have a distinct behaviour. Fig. 2 confirms these positive correlations between all TE, except for extractable Cr which is very fairly correlated to total Cr and shows the best correlations with extractable Zn and Pb. Indeed extractable Cr seems to show a special distribution, in opposition with nearly all the other parameters measured. This is confirmed by the correlation of the variables with the first 4 axes of the MULTISPATI-PCA accounting for over 80% of the total variance.

The Monte-Carlo permutation test of the MULTISPATI-PCA was highly significant ($p<0.005$), which showed that the spatial structures exhibited by TE were indeed very strong and could not be attributed to random variations.

3.3. Cross validation results and variograms

The results of the cross-validations showed that the mean of $\boldsymbol{\theta}$ for every 4 axes is outside the 90% confidence limit although the median is inside. The median and the mean of $\boldsymbol{\theta}$ are respectively (0.460; 1.519) for axis 1, (0.451; 1.343) for axis 2, (0.447; 1.342) for axis 3 and, (0.428; 1.555) for axis 4. Variograms are shown in Fig. 4. The Dowd's variogram provides a lower semi-variance than the one from the classical

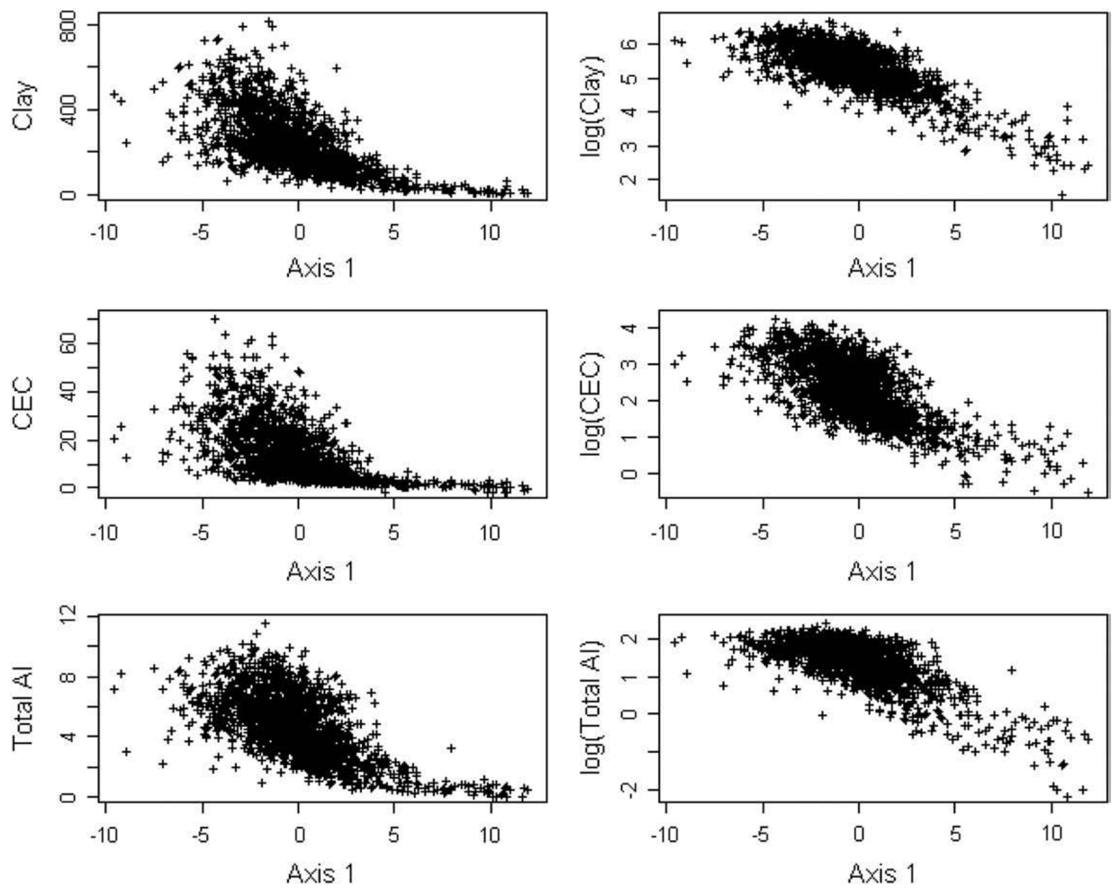

Fig. 3. Relationships between clay content, CEC and total Al and, the scores of axis 1 obtained from the MULTISPATI-PCA of the 8 TE. The second column corresponds to the logarithm of the *y* axis.

Matheron variogram (Matheron, 1971) by removing the effects of local anomalies. This indicates that the outliers have led to large θ_i at some sites and Dowd variogram has then adequately modeled the spatial variation of the background process and authorize us to interpret the factorial map. Although the nugget effects of the fitted Matérn variograms are rather high, they all exhibit a marked increase between 0 and ca 100 km, indicating a strong autocorrelation at rather short distances. Moreover, axes 1, 3 and 4 show a small ν parameter of the Matérn function (0.25, 0.07, 0.15 respectively) indicating that the spatial process is rough at small distance h. Axis 2 shows a ν parameter close to 0.5 (e.g. close to the exponential covariance function) with a continuous and sharp increase till ca 300 km. Finally the largest relative decrease in variance is observed for axis 4 indicating the large effect of quasi-point process.

3.4. Interpretation of the MULTISPATI-PCA axes

The first axis (61.1%) (Fig. 2) can be interpreted as an axis of global richness in TE. The correlations between axis 1 and clay content, cation exchange capacity and Al content (Fig. 3) suggest that trace elements are specifically associated with phyllosilicates of the finest fraction as shown in Northern France by Sterckeman et al. (2006). The map of the scores on axis 1 (Fig. 4) shows a spatial pattern similar to clay content distribution in French topsoil (Fig. 5A). We conclude that this map reveals the spatial trends in the natural abundance of pedogeogenic trace elements in French topsoil. Axis 2 (13.44%) is mainly driven by total Pb and Tl. We investigated if this axis and Pb or Tl concentrations could be linked to organic carbon content in soil and found no significant relationships (Fig. 6). The pattern of the map is clearly linked to the geographical distribution of some specific parent materials (Fig. 5B) and is consistent with the shape of the variogram. The map of the scores of axis 2 shows a spatial distribution linked to old residual soils, developed from rocks highly mineralized by long-term geological processes. Tremel et al. (1997) have shown that very high natural contents of Tl occurred in such soils in France, associated with mineralizations inducing also high natural Pb content. A set of 244 samples, some of which have been collected in the vicinity of potential anthropogenic contamination sources, could not evidence anthropogenic origin. The fact that this axis is mainly driven by total forms of these TE also suggests a natural origin. We conclude that the axis 2 reflects natural geochemical anomalies in Tl and Pb. Axis 3 (8%) is mainly driven by extractable TE, namely extractable Pb, Zn and Cd. The map of axis 3 exhibits two different kinds of processes. It shows on one hand two well known natural pedogeogenic anomalies: the Jura calcareous mountain (Webster et al., 1994; Bourennane et al., 2003; Imrie et al., 2008), and the border of the Cévennes massif in southern France (Tremel et al., 1997). On the other hand, it shows high values around the Paris area and the highly industrialized Northern part of France that are likely to be of anthropogenic origin, as already shown for Pb by Saby et al. (2006) around Paris, and for various TE by

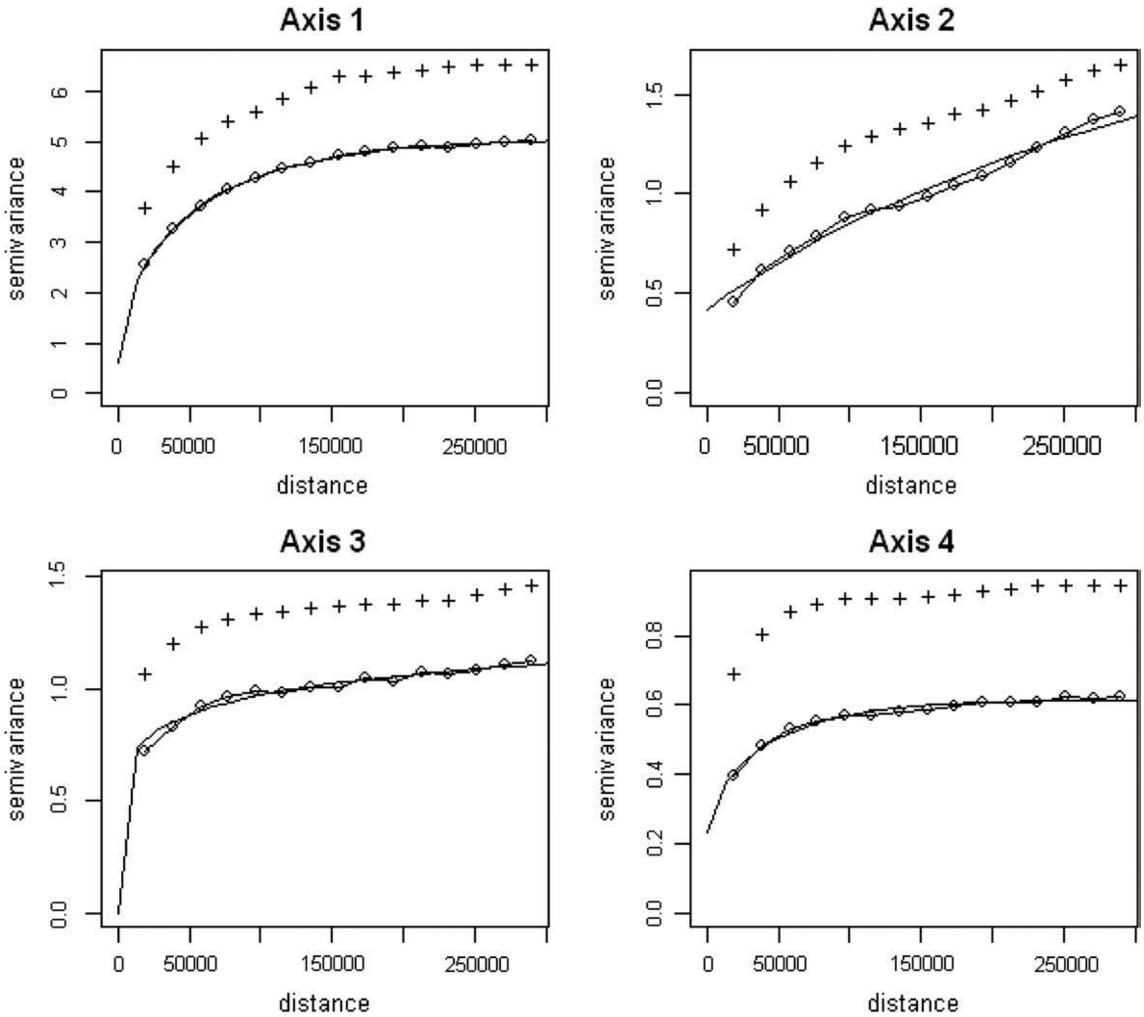

Fig. 4. Variograms of the scores of 4 axis obtained from the MULTISPATI-PCA of the 8 TE. The lines show the fitted models on the Dowd's variograms. The crosses represent the classical Matheron variograms.

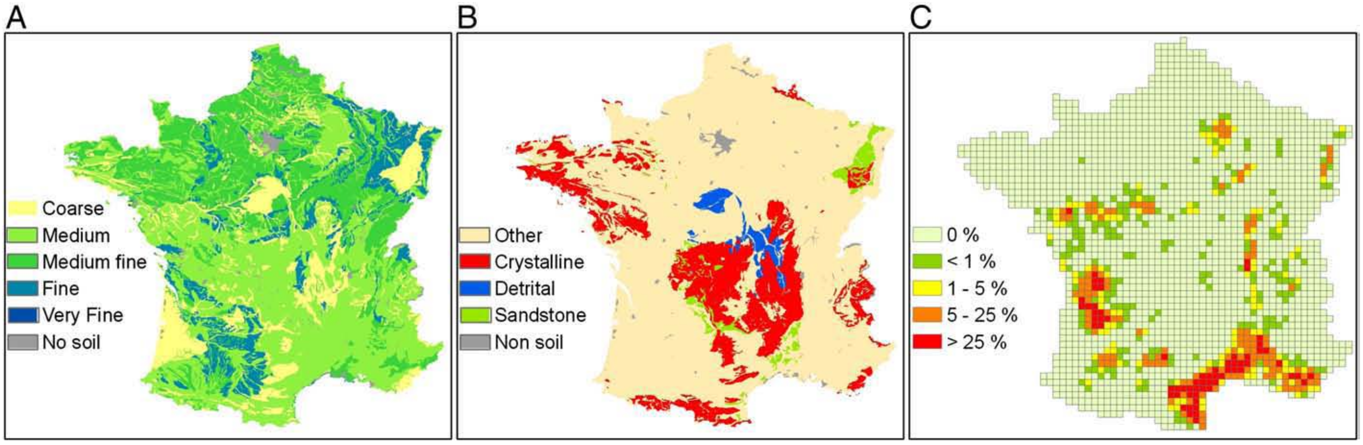

Fig. 5. Maps of (A) dominant topsoil texture extracted for the French 1: 1,000,000 geographical soil database (King et al., 1995), (B) selected dominant parent materials extracted for the French 1: 1,000,000 geographical soil database, (C) percentage of vineyards in land cover derived from agricultural census data and aggregated on the 16 km to 16 km grid.

Sterckeman et al. (2006) in Northern France. This anthropogenic origin is confirmed by high values in extractable forms. Axis 4 (5.6%) is mainly driven by Cu, both in its total and EDTA extractable contents, and its spatial distribution clearly shows two different anthropogenic origins, vine fungicide treatment by $CuSO_4$ (area of Bordeaux, South of France, some sites along the Loire and the Rhône valleys, Languedoc region, see the map of vineyards location in France, Fig. 5C) and industrial origin (Paris region, Northern France). Indeed, in this case, relatively high quantities of the extractable form of Cu can be associated with anthropogenic activity.

4. Discussion

4.1. Limitations due to the sampling scheme

The sampling scheme of the French soil monitoring network is based on a 16×16 km grid, which provides a high nugget/sill ratio, producing smoothed map of PCA axes. Indeed, our sampling scheme and interpolation technique are not suited to detect point-source pollution, therefore local "hot spots" of high TE content could not be detected by our study. The spatial structures we observed reflect only medium range trends in natural or anthropogenic TE distribution. Therefore: (i) the total area of high anthropogenic TE concentration in the topsoil might be underestimated (this is likely to be the case for small vineyard areas and Cu, for instance), (ii) risks due to "hot spots" of TE content cannot be assessed using this method.

4.2. Comparison with other studies

Using factorial kriging on a larger set of elements over Europe, Imrie et al. (2008) found a dominant effect of parent material on topsoil geochemistry. Although soil texture was identified as a possible controlling factor of topsoil geochemistry, its effect was less pronounced than in our study. This difference is likely attributable both to the different set of parameters used and to the different geographical coverage of the studies.

At the French territory scale, we show that the main factors controlling TE distribution in topsoil are soil texture, variations in parent material chemistry and weathering, and various anthropogenic sources. Some of the anthropogenic patterns could not have been detected using a looser sampling design, as suggested by the shape of the variograms. Indeed, some of the patterns that we identified are

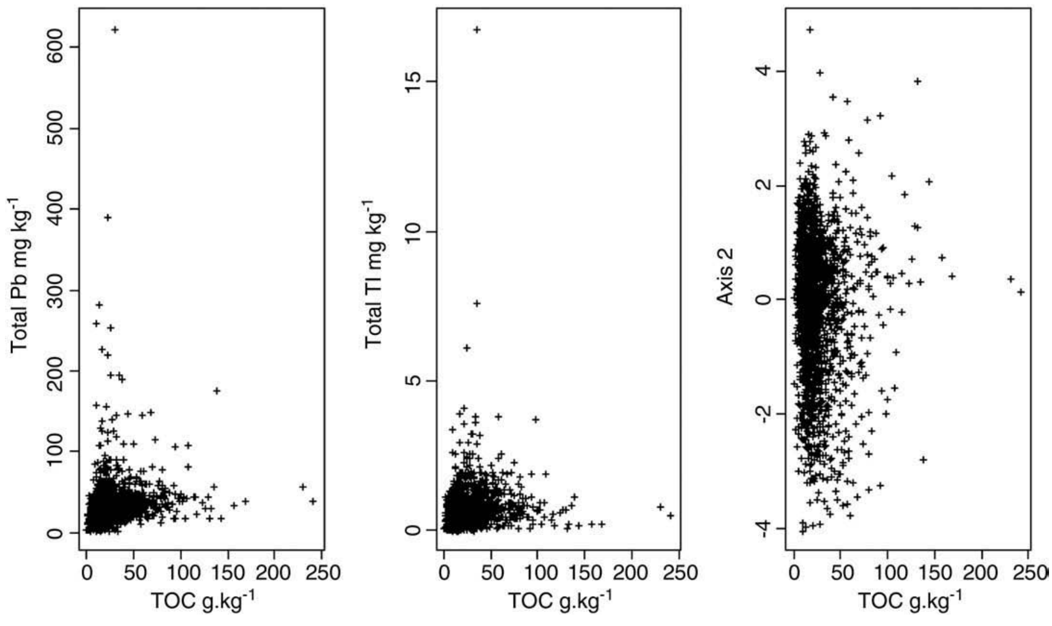

Fig. 6. Plot of total organic carbon (TOC) against total Pb content, total Tl content and the score of axis 2 obtained from the MULTISPATI-PCA of the 8 TE.

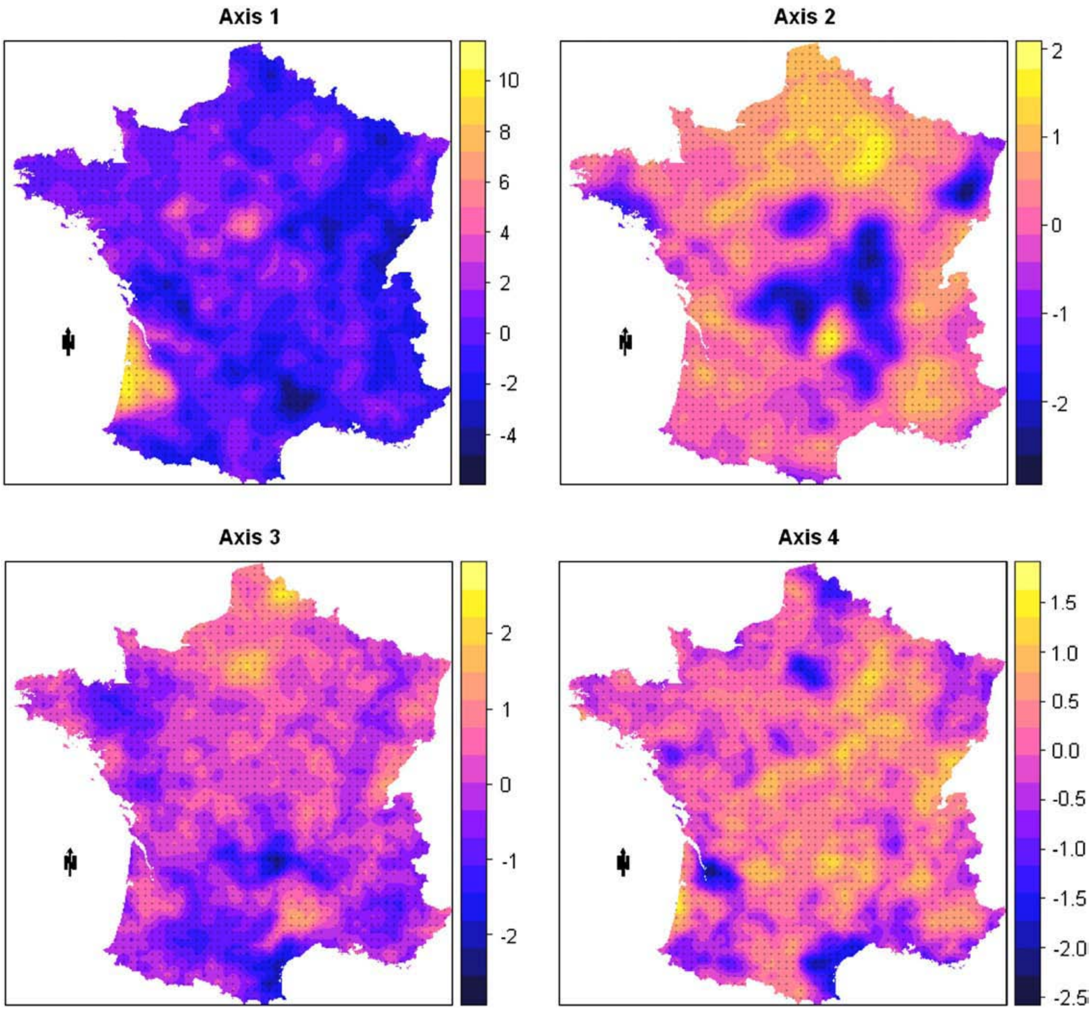

Fig. 7. Interpolated maps of the first four axes of the MULTISPATI-PCA using robust geostatistics.

rather small (see for instance map of axis 4, Fig. 7) and therefore, a looser sampling design may not have captured them. As we stress above, it is even likely that more local anthropogenic or natural structures exist and could not be detected.

5. Conclusion

Using the soil samples of a 16×16 km grid soil monitoring network, we performed an estimation of the distribution of TE contents over the whole French territory. This study shows that the spacing of the sampling grid combined with MULTISPATI-PCA associated with robust geostatistical interpolation techniques allows to detect and to map regional trends in TE distribution. At the French territory scale, we show that the main factors controlling TE distribution in the topsoil are soil texture, variations in parent material geology and weathering, and various anthropogenic sources.

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 the soil surveyors and technical assistants involved in sampling the sites. Thanks are expressed to the anonymous reviewers for their suggestions and to Eva Lacarce for the help with English.

References

AFNOR. NF ISO 11464. Prétraitement des échantillons pour analyses physico-chimiques; 1994. 8 pp.

AFNOR. NF ISO 14869-1. Qualité du sol. Mise en solution pour la détermination des teneurs élémentaires totales. Partie 1: Mise en solution par l'acide fluorhydrique et l'acide perchlorique. 2001, 12 pp.

Arrouays D, Deslais W, Badeau V. The carbon content of topsoil and its geographical distribution in France. Soil Use Manag 2001;17:7-11.

Arrouays D, Jolivet C, Boulonne L, Bodineau G, Saby N, Grolleau E. A new initiative in France: a multi-institutional soil quality monitoring network. C R Acad Agric Fr 2002;88:93-105.

Arrouays D, Morvan X, Saby NPA, Richer De Forges A, Le Bas C, Bellamy PH, et al. Environmental assessment of soil for monitoring: volume IIa inventory & monitoring. Luxembourg: Office for the Official Publications of the European Communities; 2008. p. 188. EUR 23490 EN/2A.

Atteia O, Dubois J, Webster R. Geostatistical analysis of soil contamination in the Swiss Jura. Environ. Environ Pollut 1994;86:315–27.

Bacon JR, Jones KC, McGrath SP, Johnston AE. Isotopic character of lead deposited from the atmosphere at a grassland site in the United Kingdom since 1860. Environ Sci Technol 1996;30:2511–8.

Baize D. Teneurs en huit éléments en traces (Cd, Cr, Cu, Hg, Ni, Pb, Se, Zn) dans les sols agricoles en France. Résultats d'une collecte de données à l'échelon national. Contrats n° 0375C0035 et 0575C0055. ADEME, Angers. 2007, 86.

Baize D, Sterckeman T. Of the necessity of knowledge of the natural pedo-geochemical background content in the evaluation of the contamination of soils by trace elements. Sci Total Environ 2001;264:127–39.

Basta NT, Pantone DJ, Tabatabai MA. Path analysis of heavy metals adsorption by soil. Agron J 1993;85:1054–7.

Bivand R, Anselin L, Assunçaõ R, Berke O, Bernat A, Carvalho M, et al. spdep: Spatial dependence: weighting schemes, statistics and models. R package version 0.4-29. In: http://cran.at.r-project.org/web/packages/spdep., 2008.

Bocchi S, Castrignanò A, Fornaro F, Maggiore T. Application of factorial kriging for mapping soil variation at field scale. Eur J Agron 2000;13:295.

Bourennane H, Salvador-Blanes S, Cornu S, King D. Scale of spatial dependence between chemical properties of topsoil and subsoil over a geologically contrasted area (Massif central, France). Geoderma 2003;112:235–51.

Box GEP, Cox DR. An analysis of transformations. J R Stat Soc 1964;26:211–43.

Chessel D, Dufour AB, Thioulouse J. The ade4 package—I: one-table methods. R news 2004;4:5-10.

Colbourn P, Thornton I. Lead pollution in agricultural soils. J Soil Sci 1978;29:513–26.

Dowd PA. The variogram and kriging: robust and resistant estimators. In: Verly G, David M, Journel AG, Marechal A, editors. Geostatistics for natural resources characterization. Dordercht, The Netherlands: D. Reidel; 1984. p. 91-106.

Dray S, Chessel D, Thioulouse J. Co-inertia analysis and the linking of ecological tables. Ecology 2003;84:3078–89.

Dray S, Said S, Debias F. Spatial ordination of vegetation data using a generalization of Wartenberg's multivariate spatial correlation. J Veg Sci 2008;19:45–56.

Erel Y, Veron A, Halicz L. Tracing the transport of anthropogenic Pb in the atmosphere and in soils using isotopic ratios. Geochim Cosmochim Acta 1997;61:4495–506.

Facchinelli A, Sacchi E, Mallen L. Multivariate statistical and GIS-based approach to identify heavy metal sources in soils. Environ Pollut 2001;114:313–24.

Goovaerts P. Factorial kriging analysis: a useful tool for exploring the structure of multivariate spatial soil information. J Soil Sci 1992;43:597.

Imrie CE, Korre A, Munoz-Melendez G, Thornton I, Durucan S. Application of factorial kriging analysis to the FOREGS European topsoil geochemistry database. Sci Total Environ 2008;393:96-110.

King D, Burrill A, Daroussin J, Le Bas C, Tavernier R, Van Ranst E. The EU soil geographic database. In: King D, Jones RJA, Thomasson AJ, editors. European land information systems for agro-environmental monitoring. Luxembourg: Office for Official Publications of the European Communities; 1995. p. 43–60.

Korre A. Statistical and spatial assessment of soil heavy metal contamination in areas of poorly recorded, complex sources of pollution Part 1: factor analysis for contamination assessment. Stoch. Environ. Res. Risk Assess. 1999;13:260–87.

Lark RM. A comparison of some robust estimators of the variogram for use in soil survey. Eur J Soil Sci 2000;51:137–57.

Lark RM. Modelling complex soil properties as contaminated regionalized variables. Geoderma 2002;106:173–90.

Lee CSL, Li X, Shi W, Cheung SCN, Thornton I. Metal contamination in urban, suburban, and country park soils of Hong Kong: a study based on GIS and multivariate statistics. Sci Total Environ 2006;356:45–61.

Matheron G. La théorie des variables régionalisées et ses applications. Les cahiers du centre de morphologie mathématique, 5. Fontainebleau, 1971, 212 pp.

McGrath D, Zhang C, Cartin OT. Geostatistical analyses and hazard assessment on soil lead in Silvermines area, Ireland. Environ Pollut 2004;127:239–48.

Minasny B, McBratney AB. The Matern function as a general model for soil variograms. Geoderma 2005;128:192–207.

Morvan X, Saby NPA, Arrouays D, Le Bas C, Jones RJA, Verheijen FGA, et al. Soil monitoring in Europe: a review of existing systems and requirements for harmonisation. Sci Total Environ 2008;391:1-12.

R Development Core Team. R: a language and environment for statistical computing http://www.R-project.org, R Foundation for Statistical Computing, 20084.

Rawlins BG, Webster R, Lister TR. The influence of parent material on topsoil geochemistry in eastern England. Earth Surf Processes Landf 2003;28:1389–409.

Ribiero PJ, Diggle PJ. geoR: a package for geostatistical analysis. R-NEWS 2001;1:15–8.

Saby N, Arrouays D, Boulonne L, Jolivet C, Pochot A. Geostatistical assessment of Pb in soil around Paris, France. Sci Total Environ 2006;367:212–21.

Salminen R, Batista MJ, Bidovec M, Demetriasdes A, De Vivo B, De Vos W. Geochemical atlas of Europe. Part 1: background information, methodology and maps. Espoo: Geological survey of Finland; 2005. p. 526.

Sterckeman T, Douay F, Baize D, Fourrier H, Proix N, Schvartz C, et al. Trace element distributions in soils developed in loess deposits from northern France. Eur J Soil Sci 2006;57:392–410.

Thioulouse J, Chessel D, Champely S. Multivariate analysis of spatial patterns: a unified approach to local and global structures. Environ Ecol Stat 1995;2:1.

Tremel A, Masson P, Sterckeman T, Baize D, Mench M. Thallium in French agrosystems—I. Thallium contents in arable soils. Environ Pollut 1997;95:293–302.

Wannaz ED, Harguinteguy CA, Jasan R, Pla RR, Pignata ML. Identification of atmospheric trace-element sources by passive biomonitoring employing PCA and variogram analysis. Int J Environ Anal Chem 2008;88:229–43.

Wartenberg D. Multivariate spatial correlation: a method for exploratory geographical analysis. Geogr Anal 1985;17:263–83.

Webster R, Oliver M. Geostatistics for environmental scientists, 2nd edition. John Wiley & Sons Ltd, Chichester UK, 2007.

Webster R, Atteia O, Dubois JP. Coregionalization of trace metals in the soil in the Swiss Jura. Eur J Soil Sci 1994;45:205–18.

Zhang CS, Selinus O, Kjellstrom G. Discrimination between natural background and anthropogenic pollution in environmental geochemistry — exemplified in an area of south-eastern Sweden. Sci Total Environ 1999;244:129–40.

Zhao FJ, McGrath SP, Merrington G. Estimates of ambient background concentrations of trace metals in soils for risk assessment. Environ Pollut 2007;148:221–9.