

An approach for assessing soil health: a practical guide for optimal ecological management

Yang Gao · Yafeng Wang · Guoming Zhang ·
Jun Xia · Liang Mao · Zhenwu Tang ·
Pei Zhou

Received: 3 February 2010 / Accepted: 21 April 2011 / Published online: 15 May 2011
© Springer-Verlag 2011

Abstract How to integrate environmental geographic information and biodiversity data combined with management measures to effectively assess soil health is still an unresolved problem. This paper suggests an approach for systematically estimating soil quality and guiding ecological management. First, canonical correspondence analysis is used to predict the distributions of plant species or microorganism communities, principle pollutants and environmental variables from which spatial and environmental data are extracted by the geographic information system (GIS). Secondly, geostatistical methodologies are then used to estimate and quantify the spatial distribution characteristic of the species and pollutants and to create

maps of spatial uncertainty and hazard assessment through ArcGis technology. Finally, redundancy analysis provides a suggestion about better management strategy and environmental factor for improving soil health and biodiversity. The combination of these methods with “3S” techniques as an assessment approach effectively meets the challenges for estimation and management in different soil environments.

Keywords Geostatistics · Canonical correspondence analysis · Redundancy analysis · Hazard assessment · Soil health · Ecological management

Y. Wang and Y. Gao contributed equally to this work.

Y. Gao (✉) · J. Xia
Key Laboratory of Water Cycle and Related Land Surface Processes, Institute of Geographic Sciences and Natural Resources Research, CAS, 100101 Beijing, China
e-mail: gaoyang0898@163.com; gaoyang@igsnrr.ac.cn

Y. Wang
State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, 100085 Beijing, China
e-mail: yfwang@rcees.ac.cn

L. Mao · P. Zhou
Key Laboratory of Urban Agriculture (South), Ministry of Agriculture, 200240 Shanghai, China

G. Zhang
State Key Laboratory Earth Surface Process and Resource Ecology, Beijing Normal University, 100078 Beijing, China

Z. Tang
Research Academy of Energy and Environmental Studies, North China Electric Power University, 102206 Beijing, China

Introduction

Many soil properties (biological, chemical and physical) can be seen as spatiotemporal processes and modelling their spatiotemporal distributions is essential in many environmental sciences. Ecologists are often interested in the environmental factors that regulate plant and soil microbe community diversity across temporal and spatial scales, the impact of human activity on this diversity and the consequences of this diversity for ecosystem processes (Cao et al. 2010). Analyses of spatial variation on plant diversity and soil properties responding to soil contamination together with environmental factors may be useful as a tool for developing a sustainable soil management and soil pollution control (Andreasen and Skovgaard 2009; Gao et al. 2010a). Data sets in ecology and evolution often fall outside the scope of the methods taught in introductory statistics classes (Bolker et al. 2009). Where basic statistics rely on normally distributed data, ecology and evolution data are often binary (e.g. presence or absence of a species in a site). Ecologists have more recently begun to

appreciate the importance of random variation in space and time or among individuals.

Geostatistics offer a variety of methods to model such processes as realizations of random functions and there are examples of such geostatistical applications in studies on atmospheric pollution (Rouhani et al. 1992; Vyas and Christakos 1997), earth geophysics (Handcock and Wallis 1994; Bogaert and Christakos 1997), soil moisture content (Goovaerts and Sonnet 1993; Heuvelink et al. 1997; Cao 2008), rainfall or piezometric head fields (Rouhani and Wackernagel 1990; Armstrong et al. 1993), risk for exposure to pollutants (Carroll et al. 1997; Christakos and Hristopoulos 1998), soil impedance (Castrignanò et al. 2002) and ecology (Hohn et al. 1993). An exhaustive review of geostatistical space-time models was given by Kyriakidis and Journel (1999). Moreover, geostatistics offer a lot of methods for modelling spatial uncertainty and risk assessment for both continuous and categorical attributes (Goovaerts 1997). Often geostatistics are combined with principal components analysis (Goovaerts 1997; Wackernagel 2003; Webster and Oliver 2007) and the same could be possible for canonical correspondence analysis (CCA) and redundancy analysis (RDA) (Webster and Oliver 1990).

How to integrate the time and spatial data on soil environment and gain more useful information for effectively assessing soil health effect and realizing optimal management strategy is still an unsolved problem. Gao et al. (2010b) have done the spatial characteristic research on soil enzyme activities and microbial community structure under different land uses in Chongming Island, China. Based on the spatial data from Chongming Island and the multivariate analysis method of ecological data provided by Lepš and Šmilauer (2003), the aim of this paper is to put forward an approach of spatial biostatistics and geostatistics combined with geographic information system (GIS) for better assessing soil health and guidance for land use and ecological management.

Research methods

Data collect and processing

This paper takes the spatial distribution of heavy metals and soil enzyme activities in Chongming Island as an example. The Chongming Island is divided into four major functional districts, including industrial use, agricultural use, commercial use (involving residential area) and wetland (Fig. 1). The detail on the description of each land use and main soil characteristic parameters can be seen in Gao et al. (2010b, c). A total of 294 sample points were collected and located by GPS, and mixed soil samples (about

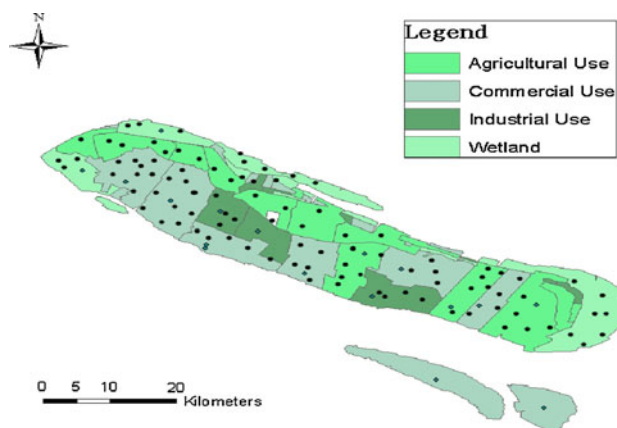


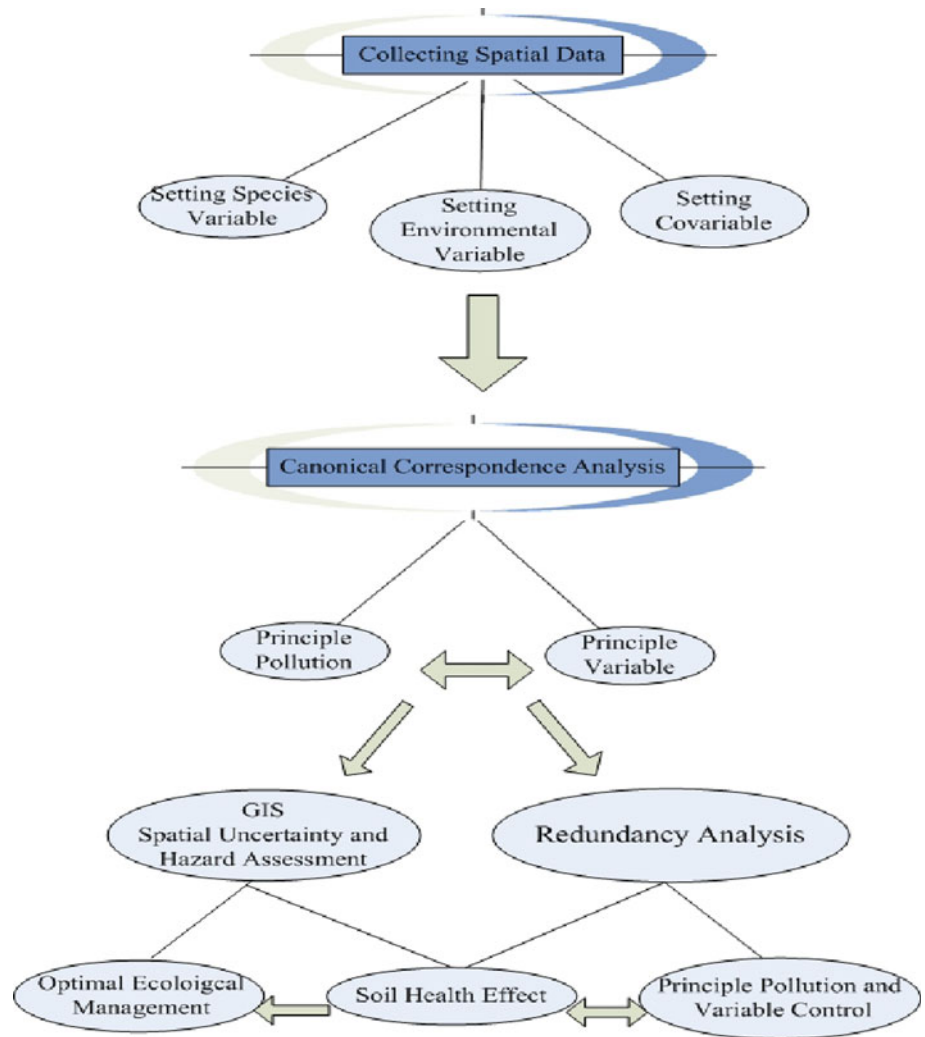
Fig. 1 Land use of Chongming Island, China (121°09'30"–121°54'00"E, 31°27'00"–31°51'15"N)

1 kg soil wet weight for each sampling point) within a radius of 100 m were then collected from each sampling point in sterile plastic bags. After biological, chemical and physical analysis for those samples, the database on the soil properties in Chongming Island was set.

Spatial data on Chongming Island were collected according to a stratification scheme, using elevation (300 m bands), aspect (four classes, centered on N, S, E and W), slope, soil layer (0–20, 20–40, 40–60 cm) and five geologic substrates. Spatial data on Chongming Island was mapped by ArcGis9.0 Software. CCA can be a useful tool to show relationships between microbe communities and soil enzyme activity in response to soil contamination and environmental factors. RDA is used to analyze sampling time as covariables and the interactions between the treatment levels and sampling times represent environmental variables. Moreover, geostatistics are extensively used to assess the level of soil contamination and calculate the risk in contaminated sites, by preserving the spatial distribution and uncertainty of the estimates. It facilitates quantification of the spatial features of soil parameters and enables spatial interpolation. CCA and RDA-based models were calibrated in CANOCO software.

The process of spatially relating data on Chongming Island is concluded in Fig. 2. First, the data on environmental variable (e.g. microorganism community, enzyme activity, soil gene polymorphisms, soil physical and chemical properties, plant diversity, climate, land use and its corresponding geographical coordinate and sample time) are collected. Secondly, predictor and response variables are set, usually as three types of variables including species variable, environmental variable and covariable. Species variable is the evaluated parameter caused by environmental factor, and environmental variable includes the content of heavy metals and organic pollutants in soil. Covariable is binary (i.e. presence/

Fig.2 Flow charts of assessment of soil health



absence), and its related parameters include fertilization, sample time, climate, soil layer depth, different management measure and principle component, and so on.

Canonical correspondence analysis

Canonical correspondence analysis (CCA) is used to predict the distributions of plant species or microorganism communities, principle pollutants, environmental variables and its interaction. In this direct gradient analysis technique, main axes of a correspondence analysis are constrained to be a linear combination of environmental descriptors, which makes it very comparable to linear regression. The ordination diagram of CCA displays species, environmental variable and covariable (Fig. 3). CCA is a multivariate extension of weighted averaging ordination, which is a simple method for arranging species along environmental variables (Andreasen and Skovgaard 2009). CCA constructs those linear combinations of environmental variables, along which the distributions of the

species are maximally separated (Ter-Braak 1986; Guisan et al. 1999). The eigenvalues produced by CCA measure this separation. CCA is a correspondence analysis technique in which the ordination axes are constrained to be linear combinations of environmental variables and the principal axes. In this ordination diagram, the approximated correlation between two variables is equal to the cosine of the angle between the corresponding arrows. Therefore, arrows pointing in the same direction correspond to variables that are predicted to have a large positive correlation, whereas species with a large negative correlation are predicted to have arrows pointing in opposite directions.

The species points can also be projected perpendicularly to the arrows of environmental variables or covariables, which indicates the approximate ordering of the species in order of increasing value of that environmental variable (if proceeding towards the arrow tip and beyond it). The environmental variables (and covariables) are always centered (and standardized) before the ordination model is

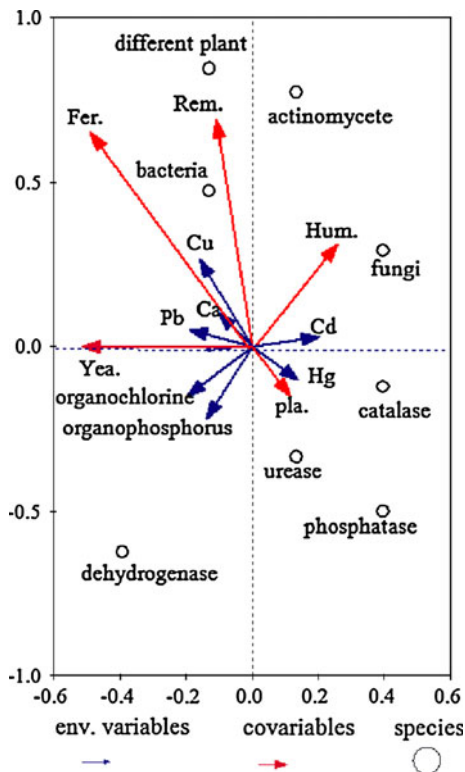


Fig. 3 The relationships of monitored species, environmental variables and covariables

fitted. Thus, similar to projecting the species points on the variables arrows, a projection point near zero (the coordinate system origin) corresponds to the average value of that particular environmental variable in that species.

From Fig. 3, it can be seen that soil health parameter, e.g. microorganism, enzyme activity and plant diversity, significantly correlated with corresponding environmental factor, e.g. Cu, Hg, Pb, Cd and organic pollutants. In addition, it also shows the interaction of inner environmental variable and the relationship between environmental variable and covariable. The species–environment correlation can be measured for each axis as the correlation of the respective multidimensional coordinates of the species occurrences in both the species and the environmental space. The CCA displays a clear pattern for the species and variables. The main source of variation in the occurrence of the different patterns is shown by the clear separation from each other. The CCA indicates a pattern similar to the cluster analyses, but the CCA display is vaguer since it shows all the observations (Andreasen and Skovgaard 2009).

Spatial uncertainty and hazard assessment

After principle specie variable and environmental variable are selected by CCA, the spatial environmental data on

each sample point are extracted by GIS. Geostatistical methodologies are then used to estimate and quantify the species and spatial distribution characteristic for pollutants under different land uses. Principle species and environmental variable and its corresponding geographical coordinate are imported to ArcGis9.0 to create the map of hazard assessment by the geostatistical methods of ordinary kriging and semivariance analysis. Of course, another spatial interpolation method also can be selected according to different spatial distribution characteristics. Phosphatase and Cu are main source of species and environmental variable confirmed by CCA. Figures 4 and 5 show that areas of high activity for phosphatase coincide with areas of high Cu concentrations.

Semivariance is an autocorrelation statistic defined as:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \quad (1)$$

where $\gamma(h)$ semivariance for interval distance class h , x_i measured sample value at point i , $x_i + h$ measured sample value at point $i + h$, and $N(h)$ total number of sample couples for the lag interval h .

Nugget value is generally affected by experimental error and the variance of small sample scale, and sill is generally caused by total variability in system. Figure 4 shows that the sill value for phosphatase distribution is almost three times higher than nugget value for phosphatase distribution, which indicates that the spatial semivariance of enzyme activity is mostly from inner system, not from small sample scale error (Komnitsas and Modis 2006). The ratio of nugget value to sill value, referred to as base effect, can be used to describe the spatial variability characteristic. The higher base effect indicates that spatial variability is more affected by random factors and also can be described as the spatial correlation. The main shortcoming of estimation maps lies in the smoothing effect, which entails

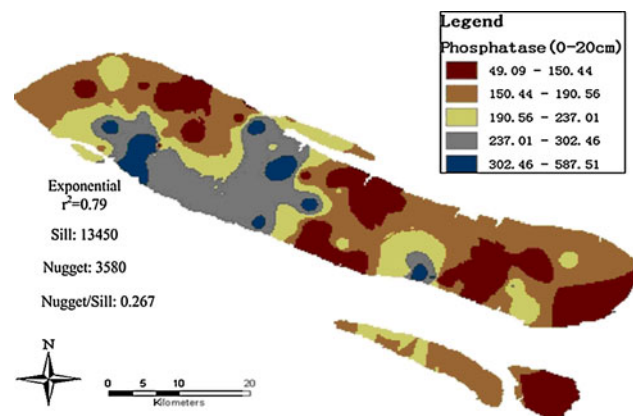


Fig. 4 Spatial distribution characteristic of phosphatase at soil surface layer (0–20 cm) in Chongming Island

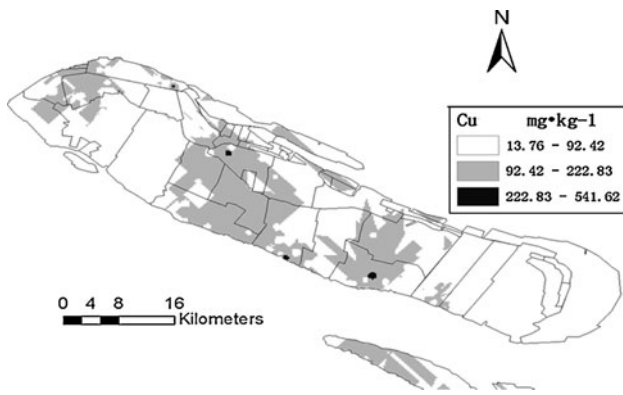


Fig.5 Spatial distribution characteristic of Cu at soil surface layer (0–20 cm) in Chongming Island

overestimation and underestimation of low and high values, respectively. When, in soil quality evaluations, two or more pollutants occur simultaneously, it is necessary to determine the total area affected by any of the pollutants (Franco et al. 2006). This leads to the problem of defining the area contaminated by the different pollutants simultaneously. The geostatistical map for soil quality indicates that there still exists some large areas at low soil quality that need to be mapped and considered in soil management (Rodríguez et al. 2009).

Redundancy analysis

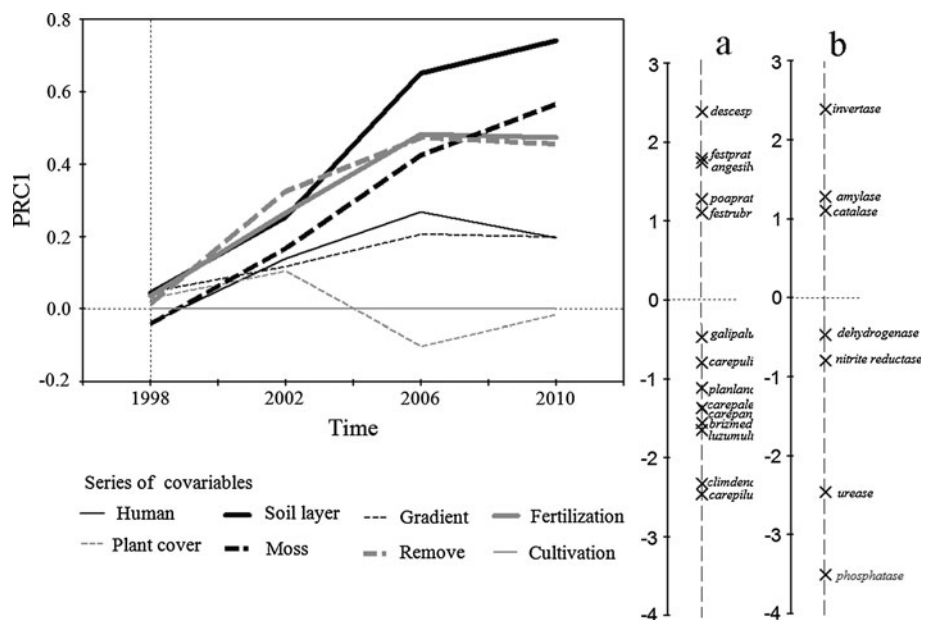
The other substep after CCA is to make redundancy analysis (RDA). Because the data form repeated observations that include the baseline (before treatment) measurements, the interaction of treatment (including land use, fertilization, population and management measure) and time is of

the greatest interest and corresponds to the effect of the experimental manipulation. RDA, a method based on a linear species response, is used because the species composition in the plots is rather homogeneous and the explanatory variables are categorical. By using the various combinations of environmental variables and covariables in RDA with the appropriate permutation scheme, tests analogous to the testing of significance of particular terms in ANOVA models (including repeated measurements) are constructed. From Fig. 6, it can be seen that the covariables change with time. RDA can show the interaction of the covariables or the change of the experimental variables and the interaction of the experimental variables with time.

The diagram of the principal response curves (PRC) shows that there are two directions of departure from covariables on the reference area. Axis *a* and *b* are different plant cover and soil enzyme activity, respectively (Fig. 6). From the PRC diagram and its corresponding axis *a* or *b*, it shows that the effect of different impact factor and its interaction on plant cover and soil enzyme activity with time. For example, the scores of the invertase activity in axis *b* is around +1.5, and if checking the PRC score of the soil layer covariable (*F* in the diagram) in the year 2010, it indicates that the score is approximately +0.5. The estimated change is, therefore, $\exp(1.5 \times 0.5) = 2.12$, so all the invertase activity is predicted to have, on average, more than two times higher activity with soil layer change compared with the effect of cultivation management on invertase activity (Ter-Braak and Šmilauer 2002). Furthermore, the species scores on the first principal response axis can be compared with species traits (Fig. 6).

In this example, the PRC diagram clearly shows that the development of impact factor on species diverges during

Fig.6 Principal response curves created by use RDA, and vertical 1D plot with plant cover (a) and soil enzyme activity (b) scores on the RDA axis



all the years, particularly for the soil layer variable (and in this way confirms that the use of time as a quantitative variable is a good approximation). In a longer run of the experiment, it is easy to find some stabilization and could estimate the time needed to achieve a ‘stable state’ (Vanden-Brink and Ter-Braak 1999; Lepš and Šmilauer 2003). Based on the combined information from all the analyses, many species (particularly the small ones) are suppressed by either environmental factor or by the absence of soil management or by a combination of both factors. Thus, it is easy to diagnose or deduce the degree of soil health and which are major pollutants in this area. Using the hazard assessment maps, e.g. Figs. 4 and 5, decisions can be made on which soil management strategy and environmental control measures to take to improve soil health and alleviate environmental pollution.

Conclusion

Controlling environmental pollution and improving soil health in large scale are still difficult and unsolved environmental problems. Viewing from the results of the map of hazard assessment and RDA, it is easy to diagnose the degree of soil health and predict where the pollution events happen, and search the pollution reason and corresponding source. Based on above diagnosis and assessment, ecologists and environmental managers can realize optimal ecology management through controlling principle pollutants, environmental variables or covariables. Soil management policy can be spatially combined with environmental data reducing the sample work and providing prediction of tendencies for environment factors in this area. Rapid environmental decisions based on visualized management for soil health and contaminations are a long way off, but “3S” technique development shows promise to speed up the process.

Acknowledgments We are grateful to Dr. Jan Lepš and Petr Šmilauer’s work named “Multivariate Analysis of Ecological Data using Canoco” to give us references and enlightenment. Opinions in the paper only reflect the personal views of the authors. This research was supported by the National Natural Science Foundation of China (40901098 and 40871085) and the special program of Water Pollution Control (2009ZX07210-006). Also the authors would like to thank the anonymous reviewers for their remarks that have improved the paper in its present form.

References

- Andreasen C, Skovgaard IM (2009) Crop and soil factors of importance for the distribution of plant species on arable fields in Denmark. *Agric Ecosyst Environ* 133:61–67
- Armstrong M, Chetboun G, Hubert P (1993) Kriging the rainfall in Lesotho. In: Soares A (ed) *Geostatistics Troia '92*, vol 2. Kluwer, Dordrecht, pp 661–672
- Bogaert P, Christakos G (1997) Spatiotemporal analysis and processing of thermometric data over Belgium. *J Geophys Res* 102(D22):25831–25846
- Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, White JSS (2009) Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol Evol* 24(3):127–135
- Cao SX (2008) Why large-scale afforestation efforts in China have failed to solve the desertification problem. *Environ Sci Technol* 42(6):1826–1831
- Cao SX, Chen L, David S, Wang CM, Wang XB, Zhang H (2010) Excessive reliance on afforestation in China’s arid and semi-arid regions: lessons in ecological restoration. *Earth Sci Rev*. doi: 10.1016/j.earscirev.2010.11.002
- Carroll RJ, Chen R, George EI, Li TH, Newton HJ, Schmiediche H, Wang N (1997) Ozone exposure and population density in Harris County, Texas (with discussion). *J Am Stat Assoc* 92(438):392–415
- Castrignanò A, Maiorana M, Fornaro F, Lopez N (2002) 3D spatial variability of soil strength and its change over time in a durum wheat field in Southern-Italy. *Soil Till Res* 65(1):95–108
- Christakos G, Hristopulos DT (1998) Spatiotemporal environmental health modelling: a tractatus stochasticus. Kluwer, Boston
- Franco C, Soares A, Delgado J (2006) Geostatistical modelling of heavy metal contamination in the topsoil of Guadamar river margins (S Spain) using a stochastic simulation technique. *Geoderma* 136:852–864
- Gao Y, Zhou P, Mao L, Zhi YE, Shi WJ (2010a) Assessment of effects of heavy metals combined pollution on soil enzyme activities and microbial community structure: modified ecological dose-response model and PCR-RAPD. *Environ Earth Sci* 60:603–612
- Gao Y, Mao L, Miao CY, Zhou P, Cao JJ, Zhi YE, Shi WJ (2010b) Spatial characteristics of soil enzyme activities and microbial community structure under different land uses in Chongming Island, China: Geostatistical modelling and PCR-RAPD method. *Sci Total Environ* 408:3251–3260
- Gao Y, Miao CY, Mao L, Zhou P, Jin ZG, Shi WJ (2010c) Improvement of phytoextraction and antioxidative defense of *Solanum nigrum* L. under cadmium stress by application of cadmium-resistant strain and citric acid synergy. *J Hazard Mater* 181:771–777
- Goovaerts P, Sonnet P (1993) Study of spatial and temporal variations of hydrogeochemical variables using factorial kriging analysis. In: Soares A (ed) *Geostatistics Troia '92*, vol 2. Kluwer, Dordrecht, pp 745–756
- Goovaerts P (1997) *Geostatistics for natural resources evaluation*. Oxford University Press, Oxford
- Guisan A, Weiss SB, Weiss AD (1999) GLM versus CCA spatial modeling of plant species distribution. *Plant Ecol* 143:107–122
- Handcock MS, Wallis JR (1994) An approach to statistical spatial-temporal modeling of meteorological fields (with discussion). *J Am Stat Assoc* 89(426):368–390
- Heuvelink GB, Musters P, Pebesma EJ (1997) Spatio-temporal kriging of soil water content. In: Baaffi E, Schofield N (eds) *Geostatistics Wollongong '96*, vol 2. Kluwer, Dordrecht, pp 1020–1030
- Hohn ME, Liebhold AM, Gribko LS (1993) Geostatistical model for forecasting spatial dynamics of defoliation caused by the gypsy moth (Lepidoptera: Lymantriidae). *Environ Entomol* 22(5):1066–1075
- Komnitsas K, Modis K (2006) Soil risk assessment of As and Zn contamination in a coal mining region using geostatistics. *Sci Total Environ* 371:190–196
- Kyriakidis C, Journel A (1999) Geostatistical space-time models: a review. *Math Geol* 31(6):651–684

- Lepš J, Šmilauer P (2003) Multivariate analysis of ecological data using Canoco. Cambridge University Press, New York, pp 232–235
- Rodríguez L, Ruiz E, Alonso-Azcárate J, Rincón J (2009) Heavy metal distribution and chemical speciation in tailings and soils around a Pb–Zn mine in Spain. *J Environ Manag* 90:1106–1116
- Rouhani S, Ebrahimpour RM, Yaqub I, Gianella E (1992) Multivariate geostatistical trend detection and network evaluation of space-time acid deposition data-I. Methodology. *Atmos Environ* 26(14):2603–2614
- Rouhani S, Wackernagel H (1990) Multivariate geostatistical approach to space-time data analysis. *Water Resour Res* 26(4):585–591
- Ter-Braak CJF (1986) Canonical correspondence analysis: a new eigenvector technique for multivariate direct gradient analysis. *Ecology* 67:1167–1179
- Ter-Braak CJF, Šmilauer P (2002) CANOCO reference manual and CanoDraw for windows user's guide: software for canonical community ordination (version 4.5). Microcomputer Power, Ithaca
- Van-den-Brink PJ, Ter-Braak CJF (1999) Principal response curves: analysis of time-dependent multivariate responses of a biological community to stress. *Environ Toxicol Chem* 18:138–148
- Vyas VM, Christakos G (1997) Spatiotemporal analysis and mapping of sulfate deposition data over eastern U.S.A. *Atmos Environ* 31(21):3623–3633
- Wackernagel H (2003) Multivariate geostatistics: an introduction with applications. Springer, Berlin
- Webster R, Oliver MA (1990) Statistical methods in soil and land resource survey. Oxford University Press, Oxford
- Webster R, Oliver MA (2007) Geostatistics for environmental scientists, 2nd edn. Wiley, Chichester