Journal of Applied Statistics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/cjas20

A new space--time multivariate approach for environmental data analysis

Sandra De Iaco [a]

[a] Dip.to di Scienze Economiche e Matematico-Statistiche, Facoltàà di Economia, Universitàà del Salento

Available online: 12 Apr 2011

Journal of Applied Statistics
2011, iFirst article

A new space–time multivariate approach for environmental data analysis

Sandra De Iaco*

Dip.to di Scienze Economiche e Matematico-Statistiche, Facoltà di Economia, Università del Salento

(*Received 26 February 2010; final version received 7 January 2011*)

Air quality control usually requires a monitoring system of multiple indicators measured at various points in space and time. Hence, the use of space–time multivariate techniques are of fundamental importance in this context, where decisions and actions regarding environmental protection should be supported by studies based on either inter-variables relations and spatial–temporal correlations. This paper describes how canonical correlation analysis can be combined with space–time geostatistical methods for analysing two spatial–temporal correlated aspects, such as air pollution concentrations and meteorological conditions. Hourly averages of three pollutants (nitric oxide, nitrogen dioxide and ozone) and three atmospheric indicators (temperature, humidity and wind speed) taken for two critical months (February and August) at several monitoring stations are considered and space–time variograms for the variables are estimated. Simultaneous relationships between such sample space–time variograms are determined through canonical correlation analysis. The most correlated canonical variates are used for describing synthetically the underlying space–time behaviour of the components of the two sets.

Keywords: space–time random fields; canonical correlation analysis; sample space–time variograms; multivariate environmental data

1. Introduction

Environmental monitoring network usually provides multivariate data which are collected at different survey stations and for a certain period of time (long time series are often available for each monitoring station). Hence, both classical multivariate methods and space–time geostatistical techniques might be applied to analyse, interpret and control the complex evolution of the observed variables. Principal component analysis (PCA) as well as related multivariate techniques, such as canonical correlation analysis (CCA), have been widely applied and a lots of papers and books can be found in literature ([7,10,11] among others). The use of these classical techniques for multivariate spatial and temporal data analysis is mainly due to climatologists. Wackernagel [14] defined a stochastic framework in a multivariate temporal (or spatial) context by using a vector of temporal (or spatial) second-order stationary random functions, whose

*Email: sandra.deiaco@unisalento.it

ISSN 0266-4763 print/ISSN 1360-0532 online

DOI: 10.1080/02664763.2011.559206
http://www.informaworld.com

components are related to different variables at a single station (or time-point). Walden *et al.* [15] used simultaneous R- and Q-mode factor analysis for detecting multivariate patterns in magnetic data sets (multivariate spatial analysis); Hsu [9] studied the interdependence between primary and secondary pollutants by fitting a vector autoregressive model to almost two years of data measured at only two stations, while Statheropoulos *et al.* [13] applied PCA to five years of data for air pollutants and meteorological variables taken at one station (multivariate temporal analysis). De Iaco *et al.* [2,4] considered principal components as measures of total air pollution in lieu of the separate contaminant concentrations. These components were treated as samples from unobserved variates defined over space and time. Space–time variograms were fitted to these new variates using the generalized product-sum model [3]. Besides classical multivariate techniques, multivariate geostatistics provides tools which take into account either inter-variables relations and spatial correlations. Cokriging, for instance, is applied when auxiliary variables may improve the estimation of other variables [12]. Factorial kriging analysis has been developed in a multivariate context to determine how spatial correlation between variables changes from one spatial scale to another [8]. Until very recently, these tools have been widely used primarily in a spatial context rather than in a spatial–temporal one. De Iaco *et al.* [5,6] showed how the generalized product-sum model can be used with a linear model of coregionalization for prediction purposes. The GSLib "COKB3D" program has been recently modified to incorporate the space–time linear coregionalization model, using the generalized product-sum variogram [1].

In this paper, a new space–time multivariate approach is described. A classical multivariate analysis, such as CCA, has been combined with space–time geostatistical tools for detecting possible interactions between two groups of variables, associated with pollutants and atmospheric conditions. Space–time variograms for the variables under study have been estimated and CCA has been used to determine simultaneous relationships between the sample variograms of the pollutants and the sample variograms of the atmospheric variables. In this context, the canonical variates, which are linear combinations of sample space–time variograms, provide a reasonable concise measure of the underlying space–time correlations of the two groups of variables.

The methodology is applied to an environmental data set, composed of hourly averages of three pollutants, such as nitric oxide (NO), nitrogen dioxide (NO_2) and ozone (O_3), and three atmospheric indicators, such as temperature (T), humidity (H) and wind speed (W_s), in February and August 2007, observed at several monitoring stations located in the Milan district and its neighbourhood (Italy).

2. Air quality data in the Milan district and its neighbourhood

Many anthropic factors affect the plume concentrations, such as manufacturing activities, heavy traffic and heating system. Besides anthropic emission sources, other aspects might influence the process of pollution formation in an urban area, such as the geographical position, the anemological field or, more generally, the atmospheric conditions which help some photochemical reactions and might contribute to dispersion or stagnation processes. Hence, the monitoring network in the Milan district and its neighbourhood is composed of several survey stations, classified in high density population stations, industrial stations, high traffic stations and outskirts stations, which provide hourly averages of different types of pollutants and meteorological variables.

These survey stations have been located in the district according to the national law in force and they are sufficiently spread out in the area.

The environmental monitoring network for meteorological and chemical variables in the Milan district and its neighbourhood is shown in Figure 1. As detailed hereafter, there are 44 monitoring stations for pollutants, seven for atmospheric variables and nine for both pollutants and atmospheric variables.

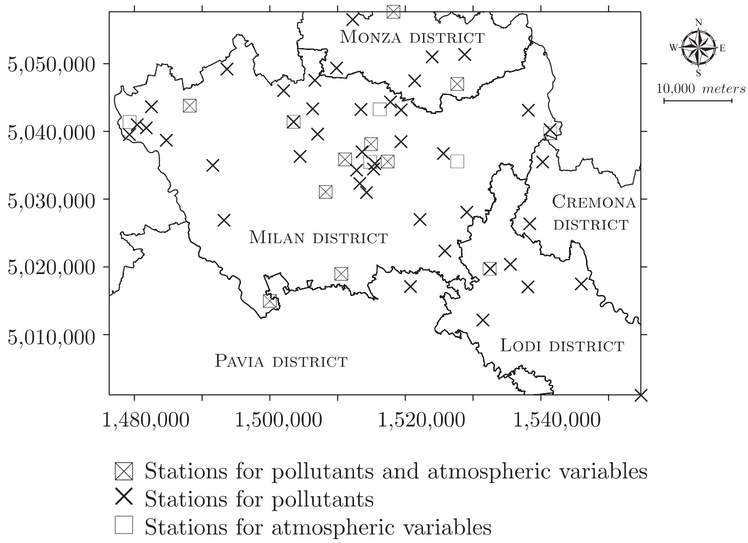

Figure 1. Map of the survey stations in the Milan district and its neighbourhood (Italy).

2.1 *The data set*

Although several pollutants and atmospheric variables are measured at each station of the monitoring network, O_3 and its precursors, such as nitric oxides (NO and NO_2), have been selected among the pollutants, moreover temperature, humidity and wind speed have been chosen among the atmospheric variables. O_3, NO and NO_2 are especially involved in photochemical smog and nowadays they represent a real threat for environmental quality, on the other hand, temperature, humidity and wind speed undoubtedly affect the formation of plumes, their reactions and their stagnation or dispersion.

In particular, the environmental data set consists of hourly averages of the six variables (three contaminants and three meteorological indicators) available at several locations during February and August. These two months are among the most critical periods of the year in terms of plume concentrations: high values of NO and NO_2 are usually recorded during winter time, likewise peaks for O_3 occur during summer time.

As shown in Figure 1, the monitoring network is composed of 60 stations spread out in the area, which are usually set to measure several variables. Regarding the selected variables, there are 53 stations for NO and NO_2 measurements, 18 stations for O_3, 16 for temperature and wind speed, and 13 for humidity.

2.2 *Explorative analysis in space–time*

As a preliminary step, various descriptive statistics were computed. Figure 2 shows the histograms of the variables together with some basic statistics, for February and August. Note that the frequency distributions of the variables are quite different: those related to NO, O_3 and wind speed exhibit a negative skewness and a considerably high variance, especially in February; NO_2 and temperature show a slightly asymmetric distribution during both months; only humidity exhibits a weak positive asymmetry, especially in August when the heat isle phenomenon is more evident. Even if a prior transformation of data might be wise in presence of extreme values, in this case data are preserved for the subsequent multivariate analysis, since CCA is applied to the well-structured sample space–time variograms of the variables.

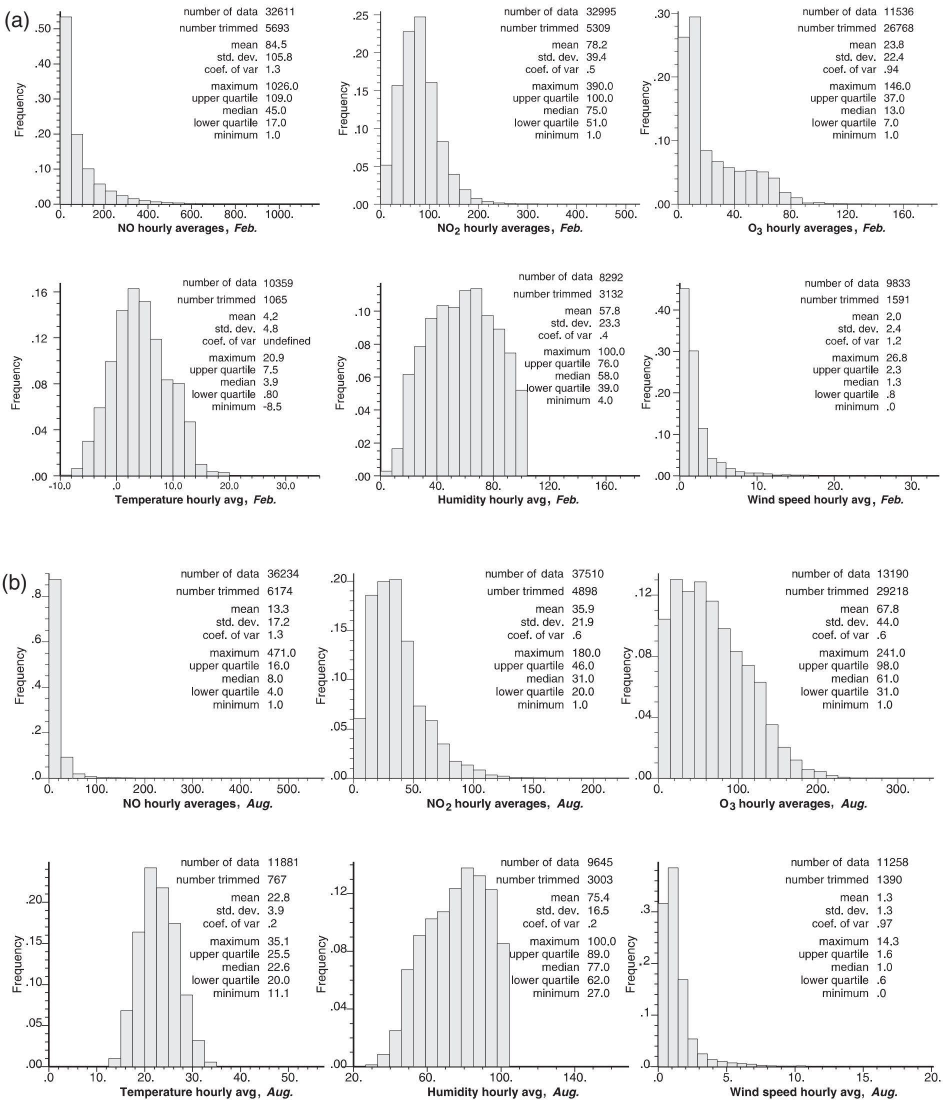

Figure 2. Histograms of the pollutants and atmospheric variables measured in (a) February, (b) August.

Moreover, among various descriptive statistics that were computed, it is worth to discuss the correlation matrices of the two data sets for February and August (Table 1):

- the correlations between NO and NO_2 are positive and relatively high during both months; a weak inverse relationship characterizes NO_2 and O_3, especially in February when the photolysis process does not strongly affect the formation of O_3;
- with respect to the relationships between atmospheric variables, the physical inverse relation between temperature and humidity is strongly confirmed by the correlation coefficients, especially in August;
- correlations between the two sets of variables show how dispersion or stagnation processes and photochemical reactions can aid or prevent the formation of O_3; in February, wind action influences O_3 formation (strong wind throws out NO, then this contributes to increase O_3 level);

Table 1. Correlation matrix of the pollutants and atmospheric variables for (a) February, (b) August.

	NO	NO_2	O_3	T	H	W_s
February						
NO	1.000	0.652	−0.396	−0.197	0.186	−0.409
NO_2	*0.652*	1.000	*−0.532*	0.114	0.035	−0.403
O_3	−0.396	*−0.532*	1.000	*0.502*	*−0.570*	*0.612*
T	−0.197	0.114	*0.502*	1.000	*−0.494*	0.389
H	0.186	0.035	*−0.570*	*−0.494*	1.000	−0.447
W_s	−0.409	−0.403	*0.612*	0.389	−0.447	1.000
August						
NO	1.000	*0.550*	−0.359	−0.124	0.132	−0.008
NO_2	*0.550*	1.000	−0.353	−0.004	0.032	−0.053
O_3	−0.359	−0.353	1.000	*0.717*	*−0.758*	0.339
T	−0.124	−0.004	*−0.717*	1.000	*−0.851*	0.281
H	0.132	0.032	*−0.759*	*−0.851*	1.000	−0.382
W_s	−0.008	−0.053	0.338	0.281	−0.382	1.000

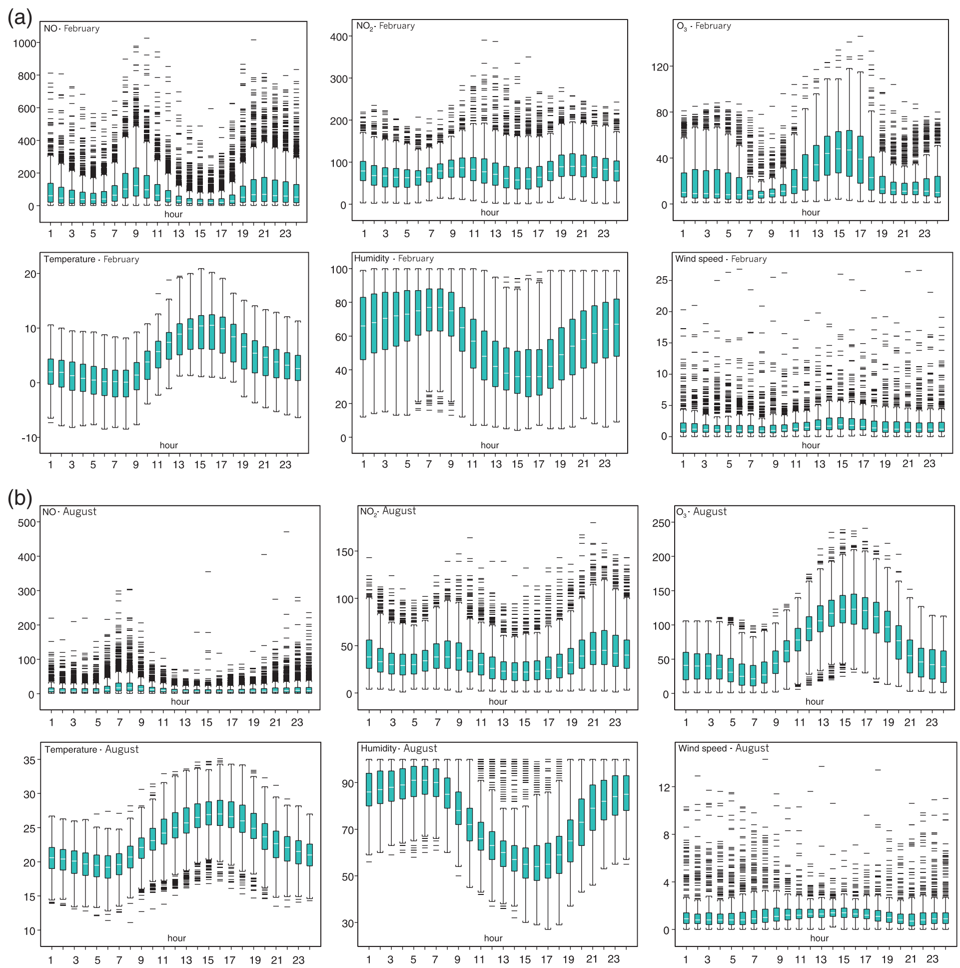

Figure 3. Box plots of the pollutants and atmospheric variables grouped by hour: (a) February, (b) August.

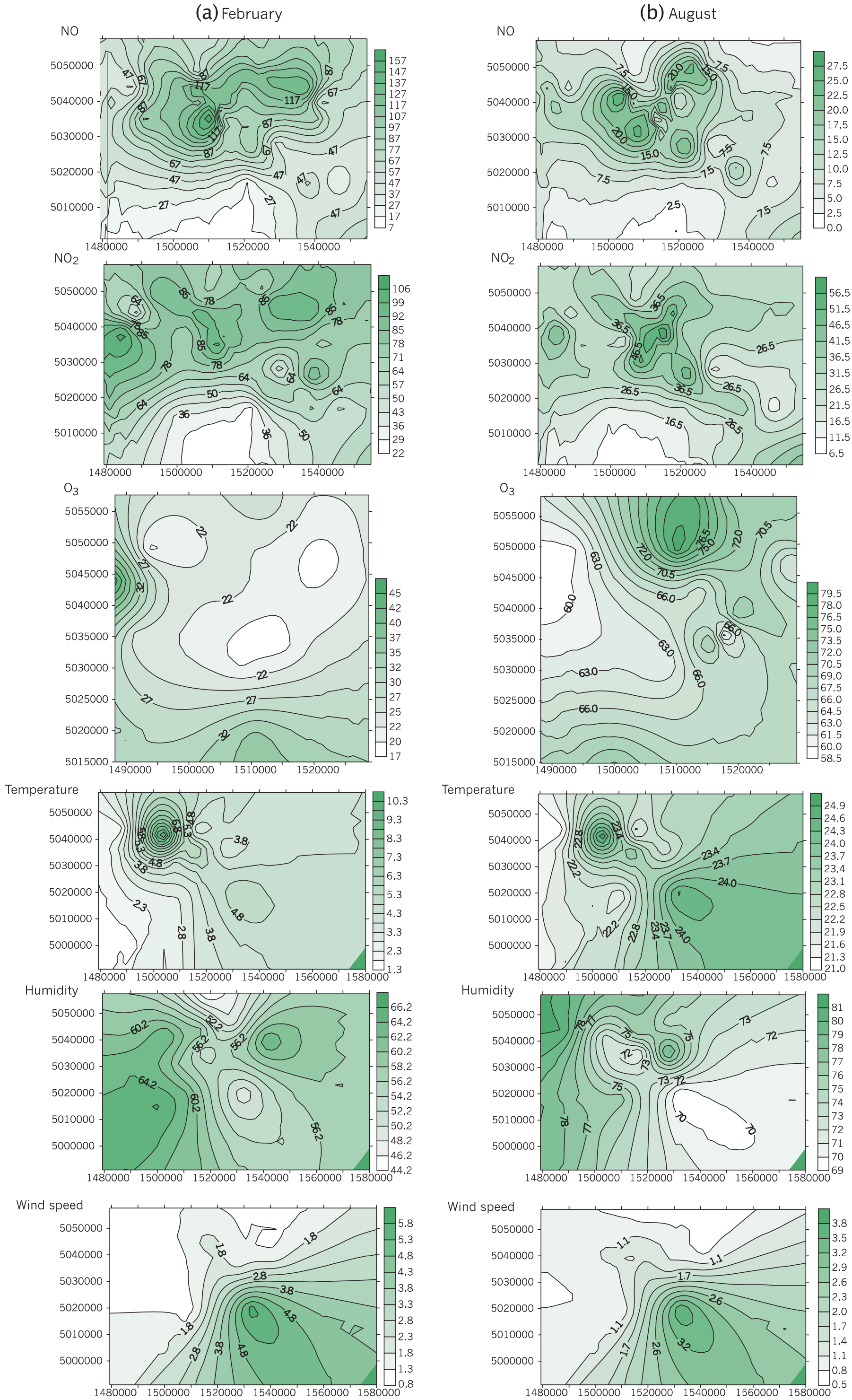

Figure 4. Contour maps of monthly averages of the pollutants and atmospheric variables measured in (a) February, (b) August.

on the other hand, in August, temperature and humidity exert their influence on O_3 formation much more than wind does.

The exploratory analysis was completed by looking at the spatial and temporal profiles of the variables.

Figure 3 shows the box plots of the six variables, grouped by hour, for February and August. They give some insights into both the daily behaviour of the variables and their frequency distributions for each hour of the day. The O_3 daily behaviour contrasts with the daily evolution of NO and NO_2; on the other hand, temperature and humidity have a diametrically opposite cycle. The wind velocity is not characterized by a clear daily periodic component.

Figure 4 illustrates the contour maps for the monthly averages of the variables measured in February and August. These maps were obtained by using ordinary kriging estimator of monthly averages, based on the spatial variogram models of the variables under study. Note that the spatial distribution of nitrogen oxides (NO and NO_2) is more concentrated around urban areas whereas the spatial distribution of O_3 is spread out over peripheral areas. Moreover, the negative correlation between temperature and humidity is confirmed by the spatial profile analyses; especially in August, high temperature values contrast low humidity values in the Southern-East area and in the Northern-West area, and vice versa elsewhere. There is an increasing trend in the spatial distribution of wind speed in the SE-NW direction during both months. This is probably due to the anemological field of the Po Valley which is located in the Southern part of the district.

In the following section, the new space–time multivariate approach, used for detecting possible interactions between environmental variables, has been described.

3. Methodology

Pollutants and atmospheric conditions are characterized by a considerable complexity and their spatial–temporal interactions cannot be clearly ignored. Hence, a new space–time multivariate approach, based on a combined use of multivariate classical techniques and space–time geostatistical techniques, is suggested.

The sample variograms of the variables under study represent the bridge between the two techniques. Possible similarities/dissimilarities of the spatial–temporal evolution of the variables can be:

- interpreted by comparing the sample variograms,
- synthesized by combining the sample variograms.

In fact, one of the multivariate geostatistical method, known as Linear Coregionalization Model, is based on a kind of linear combinations of variograms in order to describe the multivariate structure of the data. Anyway, this has been widely used in a spatial context rather than in the spatial–temporal one [5,6]. In this context, CCA is applied to relate the spatial–temporal behaviour (in terms of space–time correlation structure) of the two groups of selected variables, i.e. pollutants and meteorological variables. Regarding this aspect, the canonical variates, which are linear combinations of sample space–time variograms, reflect the underlying space–time correlations of the two groups of variables.

Note that the collected measurements are not comparable (in terms of scale and measurement units), hence the data for each variable have been standardized by subtracting the monthly mean and dividing by the monthly standard deviation. In all the subsequent analyses, standardized data sets have been used.

3.1 *Space–time structural analysis*

Let $\mathbf{Z}_P(\mathbf{s}, t)$ and $\mathbf{Z}_A(\mathbf{s}, t)$ denote two vector-valued random functions, defined on a space–time domain:

$$\mathbf{Z}_P(\mathbf{s}, t) = \{(Z_{P_1}(\mathbf{s}, t), Z_{P_2}(\mathbf{s}, t), Z_{P_3}(\mathbf{s}, t));\ (\mathbf{s}, t) \in D \times T\},$$

$$\mathbf{Z}_A(\mathbf{s}, t) = \{(Z_{A_1}(\mathbf{s}, t), Z_{A_2}(\mathbf{s}, t), Z_{A_3}(\mathbf{s}, t));\ (\mathbf{s}, t) \in D \times T\},$$

where $D \subseteq \mathbb{R}^2$ and $T \subseteq \mathbb{R}_+$, Z_{P_1}, Z_{P_2} and Z_{P_3} are associated with the standardized chemical variables NO, NO_2 and O_3, respectively, $Z_{A_1}(\mathbf{s}, t)$, $Z_{A_2}(\mathbf{s}, t)$ and $Z_{A_3}(\mathbf{s}, t)$ are associated with the standardized atmospheric variables, temperature, humidity and wind speed, respectively.

Each component is assumed to be an intrinsic random field with variogram

$$\gamma_{Z_{[\cdot]}}(\mathbf{h}_s, h_t) = 0.5\mathrm{Var}(Z_{[\cdot]}(\mathbf{s} + \mathbf{h}_s, t + h_t) - Z_{[\cdot]}(\mathbf{s}, t)), \tag{1}$$

where $(\mathbf{s}, \mathbf{s} + \mathbf{h}_s) \in D^2$ and $(t, t + h_t) \in T^2$. Note that $Z_{[\cdot]}$ is a simplified notation for Z_{P_i} or Z_{A_i}, $i = 1, 2, 3$.

Let the sets of space–time data locations be

$$U_{P_i} = \{(\mathbf{s}_l, t_j),\ l = 1, 2 \ldots, ns_{P_i},\ j = 1, 2, \ldots, nt_{P_i}\},$$

for the pollutants Z_{P_i}, $i = 1, 2, 3$, respectively, and

$$U_{A_i} = \{(\mathbf{s}_l, t_j),\ l = 1, 2 \ldots, ns_{A_i},\ j = 1, 2, \ldots, nt_{A_i}\},$$

for the atmospheric variables Z_{A_i}, $i = 1, 2, 3$, respectively.

The spatial–temporal correlation of each component is estimated by computing the sample space–time variogram, namely:

$$\hat{\gamma}_{Z_{[\cdot]}}(\mathbf{r}_s, r_t) = \frac{1}{2|L_{[\cdot]}(\mathbf{r}_s, r_t)|} \sum_{L_{[\cdot]}(\mathbf{r}_s, r_t)} [Z_{[\cdot]}(\mathbf{s} + \mathbf{h}_s, t + h_t) - Z_{[\cdot]}(\mathbf{s}, t)]^2, \tag{2}$$

where $\mathbf{r}_s$ and r_t are, respectively, the vector lag with spatial tolerance $\boldsymbol{\delta}_s$ and the lag with temporal tolerance δ_t and $|L_{[\cdot]}(\mathbf{r}_s, r_t)|$ is the cardinality of the set $L_{[\cdot]}(\mathbf{r}_s, r_t)$, that is

$$\{(\mathbf{s} + \mathbf{h}_s, t + h_t) \in U_{[\cdot]}, (\mathbf{s}, t) \in U_{[\cdot]} : \|\mathbf{r}_s - \mathbf{h}_s\| < \boldsymbol{\delta}_s \text{ and } \|r_t - h_t\| < \delta_t\}.$$

Note that $U_{[\cdot]}$ and $L_{[\cdot]}(\mathbf{r}_s, r_t)$ are simplified notations for U_{P_i} or U_{A_i} and $L_{P_i}(\mathbf{r}_s, r_t)$ or $L_{A_i}(\mathbf{r}_s, r_t)$, respectively.

At this stage, space–time variograms are estimated for the two groups of variables under study; in the following stage, CCA is proposed in order to determine simultaneous relationships between the sample variograms of the two groups of variables (i.e. pollutants and atmospheric variables).

3.2 *Space–time canonical correlation analysis*

Let X and Y be the data arrays associated, respectively, with the sample space–time variograms for pollutants,

$$\hat{\gamma}_{Z_P}(\mathbf{h}_s, h_t) = (\hat{\gamma}_{Z_{P_1}}(\mathbf{h}_s, h_t), \hat{\gamma}_{Z_{P_2}}(\mathbf{h}_s, h_t), \hat{\gamma}_{Z_{P_3}}(\mathbf{h}_s, h_t)),$$

and the sample space–time variograms for meteorological measurements,

$$\hat{\gamma}_{Z_A}(\mathbf{h}_s, h_t) = (\hat{\gamma}_{Z_{A_1}}(\mathbf{h}_s, h_t), \hat{\gamma}_{Z_{A_2}}(\mathbf{h}_s, h_t), \hat{\gamma}_{Z_{A_3}}(\mathbf{h}_s, h_t)),$$

computed at different spatial and temporal lags $(\mathbf{h}_s, h_t)$. Each data array is a 216×3 matrix, where the number of rows is given by the number of space–time lags (3 spatial lags $\times$ 72 temporal lags) at which the variograms were estimated.

Given the sample covariance matrix:

$$\Sigma = \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}.$$

CCA is used in order to find **a** and **b** such that the correlation between the linear combinations $W = X\mathbf{a}$, and $Q = Y\mathbf{b}$ (with zero mean and unit variance) is maximized [10].

Since the canonical correlation coefficient between W and Q,

$$\rho = \frac{\mathbf{a}^{\mathrm{T}}\Sigma_{xy}\mathbf{b}}{\sqrt{(\mathbf{a}^{\mathrm{T}}\Sigma_{xx}\mathbf{a})(\mathbf{b}^{\mathrm{T}}\Sigma_{yy}\mathbf{b})}},$$

is invariant under scaling of **a** and **b**, W and Q must be constrained to have zero mean and unit variance, hence

$$\mathbf{a}^{\mathrm{T}}\Sigma_{xx}\mathbf{a} = \mathbf{b}^{\mathrm{T}}\Sigma_{yy}\mathbf{b} = 1.$$

Maximizing the correlation between the two linear combinations subject to the above constraints leads to the following eigenvalue problem:

$$\left(\Sigma_{xx}^{-1}\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} - \lambda\mathbf{I}\right)\mathbf{a} = 0, \quad \left(\Sigma_{yy}^{-1}\Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy} - \lambda\mathbf{I}\right)\mathbf{b} = 0,$$

where **I** is the identity matrix and λ the eigenvalues.

Hence, let λ_1, λ_2, and λ_3, be the eigenvalues in decreasing order for the analysed data sets, and $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$, $\mathbf{b}_1, \mathbf{b}_2$, and $\mathbf{b}_3$, the corresponding eigenvectors; the following linear combinations:

$$X\mathbf{a}_i, \quad Y\mathbf{b}_i, \quad i = 1, 2, 3$$

represent the canonical variates and $\sqrt{\lambda_i}$, $i = 1, 2, 3$, the corresponding canonical correlations.

Hence, CCA is applied to the two groups of sample variograms in order to:

- identify simultaneous relationships between them; in the application, sample variograms for the pollutants and sample variograms for the atmospheric conditions have been considered;
- use the most important canonical variates to describe the underlying space–time behaviour of the components of the two groups.

4. Main results and comments

Figures 5 and 6 show the sample space–time variograms (2) for the standardized data of the two groups, in February and August, respectively.

The sample variogram surfaces reflect the spatial–temporal correlation structures of the variables under study. Looking along the temporal lags, the periodic daily components which characterize the corresponding pollutants and atmospheric variables, are clear. In particular:

- a 24 h cycle characterizes all the variables during both months, even if it is quite small for the wind speed, especially in February;
- pollutant variograms reflect a 12 h cycle, especially in February, because of the influence of human activities, such as heavy traffic during opening and closing time of the shops and heating system;

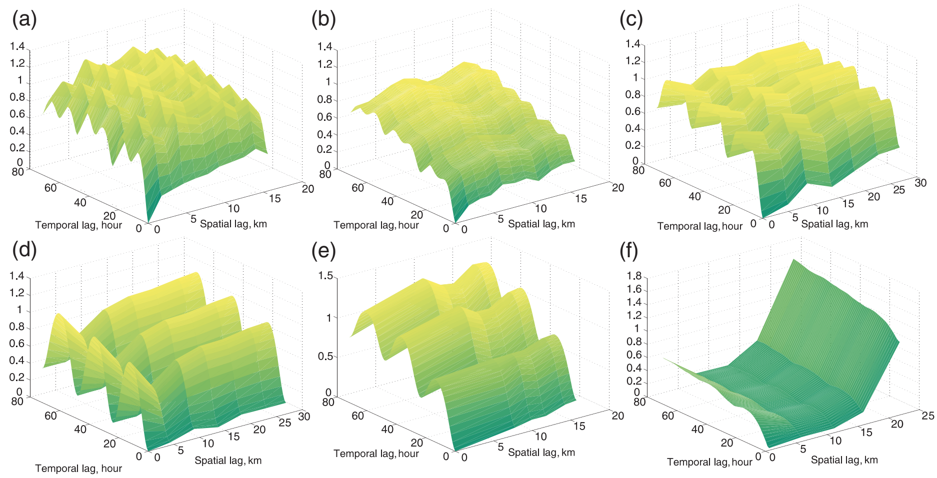

Figure 5. Sample space–time variograms for pollutants and atmospheric variables in February: (a) NO, (b) NO_2, (c) O_3, (d) T, (e) H, (f) W_s.

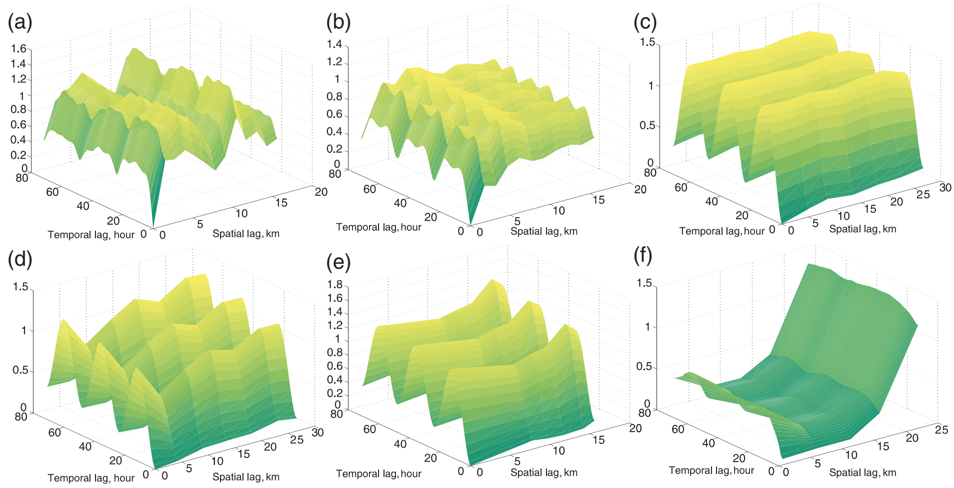

Figure 6. Sample space–time variograms for pollutants and atmospheric variables in August: (a) NO, (b) NO_2, (c) O_3, (d) T, (e) H, (f) W_s.

- O_3 preserves only a periodic component at 24 h in August; it follows the temporal behaviour of the atmospheric variables, such as temperature, since its formation is primarily due to photochemical reactions.

Main results obtained by performing CCA on the two groups of sample space–time variograms, separately for each month, are reported in Table 2.

These results can be interpreted by:

(1) looking at the *magnitudes of the canonical weights*, since the variograms which are most important in maximizing the correlation can be identified;
(2) looking at the *signs of the canonical weights*, since they can give some insights into the contrasts among the variograms;

Table 2. Canonical weights and correlation coefficients.

	February 1st variate	2nd variate	August 1st variate	2nd variate
NO	−0.2185	−0.3832	0.0025	0.2826
NO_2	0.2899	−0.2580	−0.0087	0.0685
O_3	0.2415	0.4564	0.1755	−0.0614
T	−0.0442	0.1460	−0.2931	−0.1134
H	0.2109	−0.1843	0.4283	0.1542
W_s	0.1325	0.2708	0.0593	−0.4128
Corr. coeff.	0.9779	0.8423	0.9845	0.8814

(3) considering the *canonical coefficients*, since they determine how much the linear combinations are correlated.

The weights computed for the first variates in February highlight that:

- the first variate for the group of variograms of pollutants explains the contrast between the secondary pollutants (NO_2, O_3) and the primary pollutant (NO): note that the absolute contribution is almost the same among all pollutants;
- as regards the first variate for the variograms of the atmospheric variables, the greatest contribution is given by humidity and, even if less significant, by wind speed.

On the other hand, from the CCA on the two groups of sample space–time variograms available in August it follows that:

- only the space–time variogram of O_3 characterizes the first linear combination of the group of pollutants and it is highly correlated to the linear combination of the variograms of atmospheric variables ($\rho = 0.98$);
- the weights of humidity and temperature present opposite signs, which is due to the inverse relationship between humidity and temperature.

The second canonical variates reflect:

- the inverse space–time correlation between nitrogen oxides (NO, NO_2) and ozone (O_3). This is more evident in February rather than in August when the weights of NO_2 and O_3 are very close to zero;
- the space–time correlation of wind speed, which has the greatest contribution, with respect to temperature and humidity.

Figures 7 and 8 show the first canonical variates for the two months. Note that the rescaled surfaces provide a synthetic view of the space–time behaviour of the underlying variables.

The first canonical space–time variogram surfaces for February and August are essentially characterized by a common basic structure of two groups of variables which is dominated by a 24-h cycle in time and stationarity in space. Hence, CCA has been very useful to determine the simultaneous relationships between the sample variograms for the pollutants and the sample variograms for the atmospheric variables. Under this viewpoint, the canonical variates, which are linear combinations of sample space–time variograms, provide a concise measure of the underlying space–time correlations of the two groups of variables. Actually, one of the

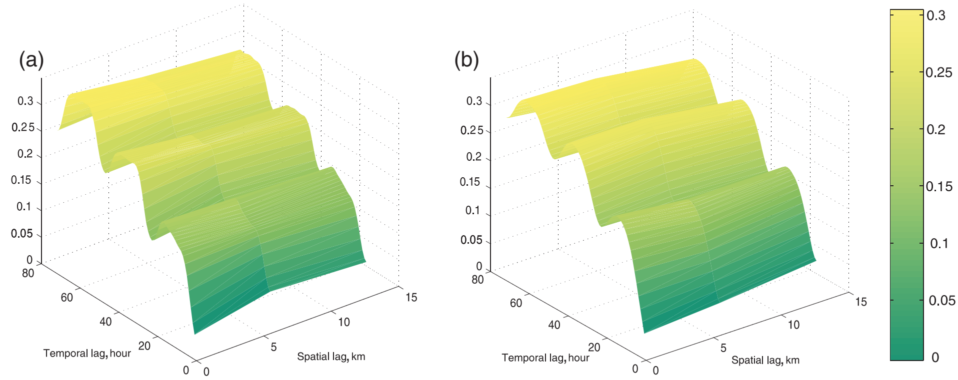

Figure 7. First canonical variates for: (A) pollutants, (B) atmospheric variables (February).

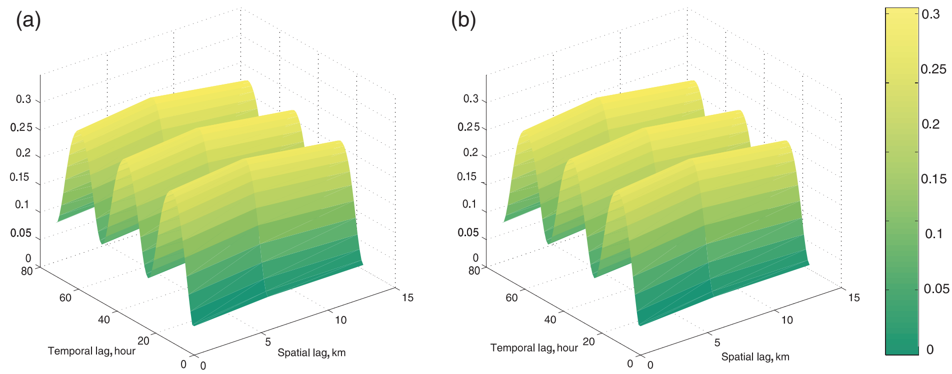

Figure 8. First canonical variates for: (a) pollutants, (b) atmospheric variables (August).

multivariate geostatistical method, known as Linear Coregionalization Model, is based on a kind of linear combinations of variograms in order to describe the multivariate structure of the data.

5. Conclusions

In environmental field and in several other sectors, a detailed analysis often requires the use of multivariate classical techniques as well as the use of Geostatistics: the former for detecting relationships between variables and reducing the number of variables under study, the latter for explaining the spatial–temporal multivariate structure of data and making predictions.

In this paper, the new space–time multivariate approach for an environmental data analysis is based on a combination of classical and geostatistical tools. The proposed methodology enables us to work with a considerably less number of variables as well as to find out the most relevant space–time correlations of the underlying variables. Two sets of correlated variables were analysed: pollutants generated by human activities and photochemical reactions, and meteorological conditions which affect stagnation or dispersion of plume. Sample space–time variograms for the two groups of variables were compared through CCA and the most highly correlated linear combinations of such variograms were considered for characterizing the multivariate spatial–temporal behaviour of the variables. The surfaces, representing the first variates (whose canonical correlation coefficients were significant), were used to synthesize the space–time behaviour of the two groups of variables. These might be clearly used for further investigations about the complex environmental interactions.

The extension of classical techniques to multivariate spatial and temporal phenomena, firstly due to climatologists, deserves attention and a lot can still be done in exploring and understanding spatio-temporal and multivariate patterns. It is clear that further developments should be considered such as modelling the space–time canonical variates by using a product-sum, application of cokriging techniques and comparison with linear coregionalization models in space–time.

Acknowledgements

I am grateful to the reviewers for their helpful suggestions and comments. The author is also indebted to Donato Posa for useful discussions during the preparation of the manuscript. This research has been partially supported by Fondazione Cassa Di Risparmio Di Puglia.

References

[1] S. De Iaco, D.E. Myers, M. Palma, and D. Posa, *FORTRAN programs for space–time multivariate modeling and prediction*, Comput. Geosci. 36(5) (2010), pp. 636–646.

[2] S. De Iaco, D.E. Myers, and D. Posa, *Total Air Pollution and Space-Time Modeling*, in *geoENV III: Geostatistics for Environmental Applications*, P. Monestiez, D. Allard, and R. Froidevaux, eds., Kluwer Academic Publishers, Dordrecht, 2001, pp. 45–52.

[3] S. De Iaco, D.E. Myers, and D. Posa, *Space–time analysis using a general product-sum model*, Statist. Probab. Lett. 52(1) (2001), pp. 21–28.

[4] S. De Iaco, D.E. Myers, and D. Posa, *Space–time variograms and a functional form for total air pollution measurements*, Comput. Statist. Data Anal. 41(2) (2002), pp. 311–328.

[5] S. De Iaco, D.E. Myers, and D. Posa, *The linear coregionalization model and the product-sum space–time variogram*, Math. Geol. 35(1) (2003), pp. 25–38.

[6] S. De Iaco, M. Palma, and D. Posa, *Modeling and prediction of multivariate space–time random fields*, Comput. Statist. Data Anal.48 (2005), pp. 525–547.

[7] W.R. Dillon and M. Goldstein, *Multivariate Analysis-Methods and Applications*, Wiley, New York, 1984.

[8] P. Goovaerts, *Study of spatial relationships between two sets of variables using multivariate geostatistics*, Geoderma 62(1–3) (1994), pp. 93–107.

[9] K. Hsu, *Time series analysis of the interdependence among air pollutions*, Atmos. Environ. 26(4) (1992), pp. 491–503.

[10] J.D. Jobson, *Applied Multivariate Data Analysis, II*, Springer, New York, 1992.

[11] L. Lebart, A. Morineau, and K.M. Warwick, *Multivariate Descriptive Statistical Analysis*, Wiley, New York, 1984.

[12] D.E. Myers, *Estimation of linear combinations and cokriging*, Math. Geol. 15(5) (1983), pp. 633–637.

[13] M. Statheropoulos, N. Vassiliadis, and A. Pappa, *Principal component and canonical correlation analysis for examining air pollution and meteorological data*, Atmos. Environ. 32(6) (1998), pp. 1087–1095.

[14] H. Wackernagel, *Multivariate Geostatistics*, Springer, New York, 1998.

[15] J. Walden, J.P. Smith, and R.V. Dackombe, *The use of simultaneous R- and Q-mode factor analysis as a tool for assisting interpretation of mineral magnetic data*, Math. Geol. 24(3) (1992), pp. 227–247.