SPATIAL-TEMPORAL MODELING OF SO_2 IN MILAN DISTRICT

L. DE CESARE

II Facoltà di Economia, Università di Bari
71100 Foggia (Italy).
I.R.M.A., CNR - 70126 Bari (Italy).

D. E. MYERS

Department of Mathematics, University of Arizona, Tucson, AZ 85721 USA.

AND

D. POSA

Facoltà di Economia, Università di Lecce - 73100 Lecce (Italy).
I.R.M.A., CNR - 70126 Bari (Italy).

1. Introduction

There is an everincreasing number of articles being published about space-time models and analysis: a sampling of such articles in the area of atmospheric pollution includes Bilonick (1985), Eynon and Switzer (1983), Stein (1986) and Le and Petkau (1988).

Most applications of geostatistics have concentrated on spatial dependence and not so many have considered the link with temporal dependence. The above tendency is due to the fact that geostatistical procedures, including kriging, have been almost exclusively designed for the analysis of spatial data; moreover, there are several possible reasons for this neglect (i.e. see Myers, 1992, p. 63).

As noted by Journel (1986), there are major differences between spatial and temporal phenomena. The most important difference is not so much that between spatial and temporal problems, but rather that between spatio-temporal problems and either of the two simpler models.

The extension of methods to combine spatial-temporal dependence require the use of a natural distance function in space-time.

E.Y. Baafi and N.A. Schofield (eds.), Geostatistics Wollongong '96, Volume 2, 1031-1042.

The aim of the present work is to apply kriging in the space-time domain in order to study the phenomena of interest as spatio-temporal variables. An application to SO_2 measurements taken in the Milan district has been described.

2. Modeling space-time data

Space-time data are assumed to be a realization of the stochastic process

$$(Z(s,t); s \in D, t \in T) \tag{1}$$

where the domain $D \subseteq \Re^d, d \leq 3$. Most commonly $T = (1,2,\ldots)$, which allows (1) to be viewed as a time series of spatial processes, each process occurring at equally spaced time points. If the temporal correlation is weak, then sampling across time leads to approximate replication of the spatial error process and more precise inferences. The generic problem is to predict $(Z(s,t); s \in D, t \in \Re_+)$ from data $(Z(s_{1,i},t_i),\ldots,Z(s_{n_i,i},t_i), i = 1,\ldots,m$, and $t_1 < t_2 < \ldots < t_m)$. Assuming that the first and second moments of Z exist, Z can be decomposed as

$$Z(s,t) = m(s,t) + Y(s,t) \tag{2}$$

where Y is the stochastic process with

$$E(Y(s,t)) = 0 \tag{3}$$

and $m(s,t)$ is the mean function of Z. For a second order stationary random field Y, the covariance function

$$C_{s,t}(h) = Cov(Y(s+h_s, t+h_t), Y(s,t)) \qquad h = (h_s, h_t) \tag{4}$$

depends solely on the lag vector h, not on location or time.

The intrinsic hypothesis is a weaker assumption that sets requirements only on differences between variables $Y(s+h_s, t+h_t)$ and $Y(s,t)$, separated by h:

$$E(Y(s+h_s, t+h_t) - Y(s,t)) = 0$$

and the variogram

$$\gamma(h_s, h_t) = \frac{Var(Y(s+h_s, t+h_t) - Y(s,t))}{2}$$

depends solely on h.

In particular $\gamma(h_s, 0)$ is called the *spatial variogram* and $\gamma(0, h_t)$ is called the *temporal variogram*.

In air pollution applications there are some examples of geostatistical methods used to analyze both spatial and temporal variability (Soares et al., 1992).

The simplest way to model a variogram in space-time is to separate the dependence on the two. There are some examples in the literature wherein variograms have been modeled in space-time (Rodriguez-Iturbe et al., 1974; Bilonick, 1987; Rouhani et al., 1989). These include the following:

$$C(u,w) = C_t(w) C_h(u)$$

$$\gamma(u,w) = \delta_0 + \gamma_t(w) + \gamma_h(u) + \gamma_{h,t}(gu^2 + w^2)$$

$$GC(u,w) = GC_t(w) + GC_h(u)$$

where

C = spatiotemporal covariance
C_t = temporal covariance
C_h = spatial covariance
γ = spatiotemporal variogram
γ_t = temporal variogram
γ_h = spatial variogram
$\gamma_{h,t}$ = isotropic spatiotemporal variogram
δ_0 = pure nugget variogram
GC = spatiotemporal generalized (polynomial) covariance
GC_t = temporal generalized covariance
GC_h = spatial generalized covariance
g = geometric coefficient of anisotropy between space and time

Unfortunately, some of these are not completely valid (Myers and Journel, 1990).

2.1. SPACE-TIME KRIGING

If time is simply considered as another dimension, then there is no change in the form of the kriging estimator nor in the kriging equations, that is, if (s,t) is an unsampled location-time and given $(Z(s_{1,i},t_i),\ldots,Z(s_{n_i,i},t_i), i = 1,\ldots,m)$ then $Z(s,t)$ could be estimated by

$$\hat{Z}(s,t) = \sum_{i=1}^{m}\sum_{j=1}^{n_i} \lambda_{i,j} Z(s_{i,j}, t_i)$$

where there are no assumptions about any interrelations between the space and the time coordinates of a point. The estimator will interpolate in either space or time and will extrapolate in either space or time.

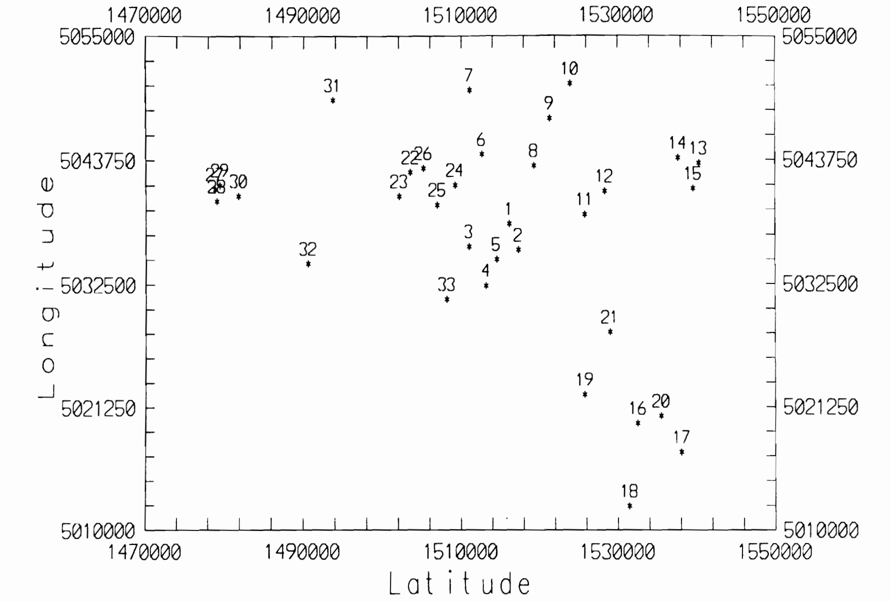

Figure 1. Network of Milan district. The coordinate system of the monitoring stations is referred to the Italian national grid system (Gauss-Boaga).

3. The Data Set

Air pollution in the province of Milan may be attributed to a variety of factors, notably emissions from motor vehicles, manufacturing and heating systems in buildings during winter. Sulphur dioxide (SO_2) is an index of contamination from heating systems not fueled by natural gas and from Diesel-powered motor vehicles.

The data collection network, which has been planned in relation to the standards and information provided by the Lombardy Region's Environmental Town Council, consists of 33 survey stations (Fig. 1). The coordinate system of the monitoring stations is referred to the Italian national grid system (Gauss-Boaga), which is based on the Universal Transverse Mercator (UTM) projection.

The data set described throughout the paper refers to monthly averages of SO_2 from January 1983 to December 1986. The sampling points are not uniformly distributed; the central portion of the study area has a relatively higher concentration of points. Fig. 2 shows the histogram of the data set: there are 8 missing values.

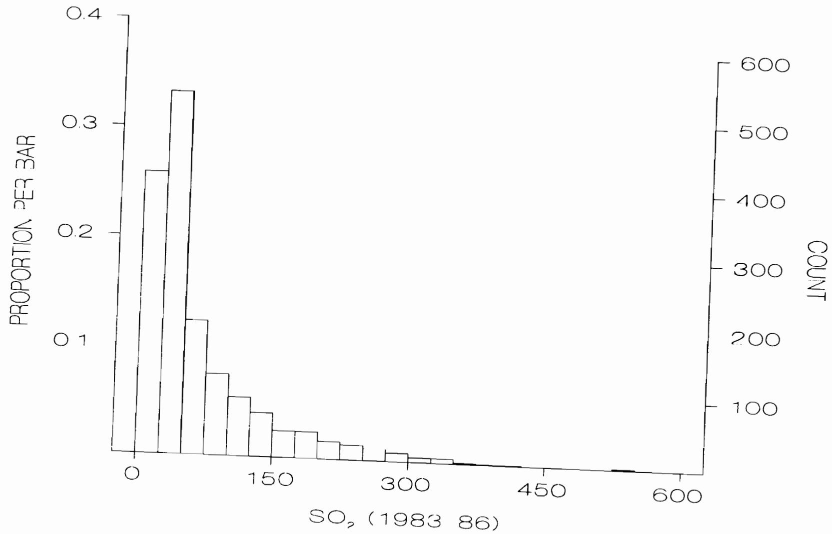

Figure 2. Histogram of the monthly averages of SO_2 from January 1983 to December 1986.

3.1. TREATMENT FOR SEASONAL EFFECT

The seasonal effect has been detected by inspecting the graph of the time series for each monitoring station. Fig. 3 shows such a plot for one of these stations (station 1). Then in model (2) we suppose that the mean function $m(s,t)$ can be decomposed as:

$$m(s,t) = \alpha(s,t) + \mu$$

where

1. $\alpha(s,t) = \alpha(s,t+d) \qquad \forall s \in D \quad \forall t \in T$
2. $\sum_{j=1}^{d} \alpha(s,j) = 0 \qquad \forall s \in D$

$\alpha(s,t)$ is called *seasonal component* with period d and μ is a constant trend. Time series analysis has been carried out for each location. Residuals have been generated for all stations after removal of the seasonal component by the standard technique of moving average estimating (Brockwell and Davis, 1987). Stations 7, 22 and 33 have missing values, then residuals have been generated by removing the first 7 entries and the last 4 values for station 7, by removing the first value for station 22 and by removing the first 5 values for station 33.

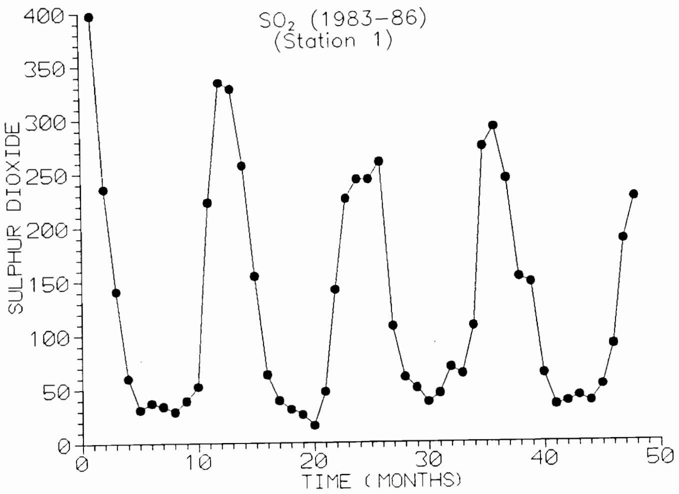

Figure 3. Graph of the time series for station 1.

3.2. STRUCTURAL ANALYSIS

The SO_2 space-time data set has been modelled by the stochastic process described by (2), (3) and (4).

In the present analysis a simple product spatial-temporal covariance model for the residual process Y:

$$C(u,w) = C_t(w)C_h(u) \tag{5}$$

has been used, where spatial dependence has been separated by the temporal one. The previous model could be easily written in terms of the spatial-temporal variogram:

$$\gamma_{st}(u,w) = C_t(0)\gamma_s(u) + C_s(0)\gamma_t(w) - \gamma_s(u)\gamma_t(w)$$

where

γ_{st} = spatiotemporal variogram
γ_t = temporal variogram
γ_s = spatial variogram
C_t = temporal covariance
C_s = spatial covariance

Fig. 4 shows the estimated temporal variogram using the original data: note that the periodic structure of the data set has been reproduced. Fig. 5 and 6 show the estimated spatial and temporal variograms using deseasonalized data respectively, with the fitted models whose analytical expression is

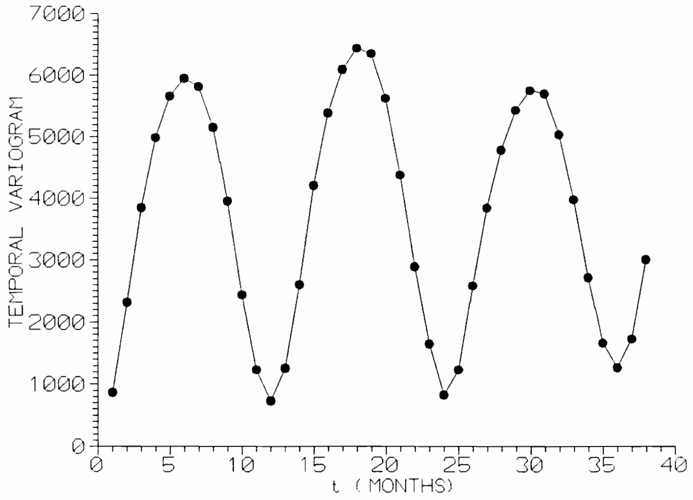

Figure 4. Estimated temporal variogram using the original data set.

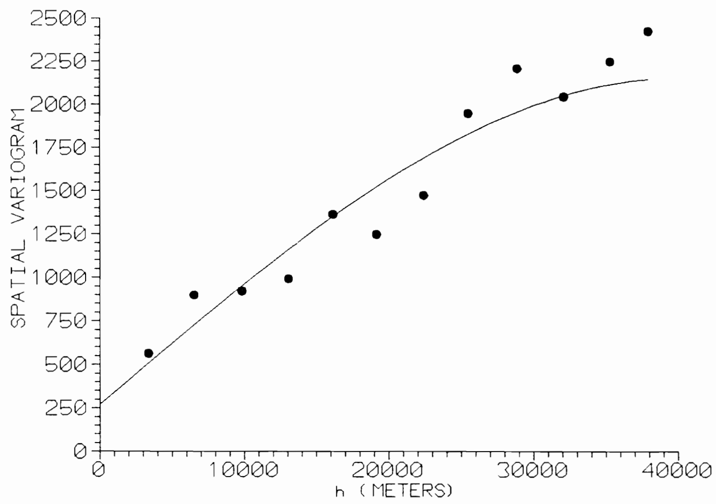

Figure 5. Estimated spatial variogram using the deseasonalized data with the fitted model.

given below:

$$\begin{aligned} \gamma_s(h) &= 270 + 1884 Sph\left(\frac{h}{40000}\right) \\ \gamma_t(h) &= 154 + 462 Sph\left(\frac{h}{4.8}\right) \end{aligned} \tag{6}$$

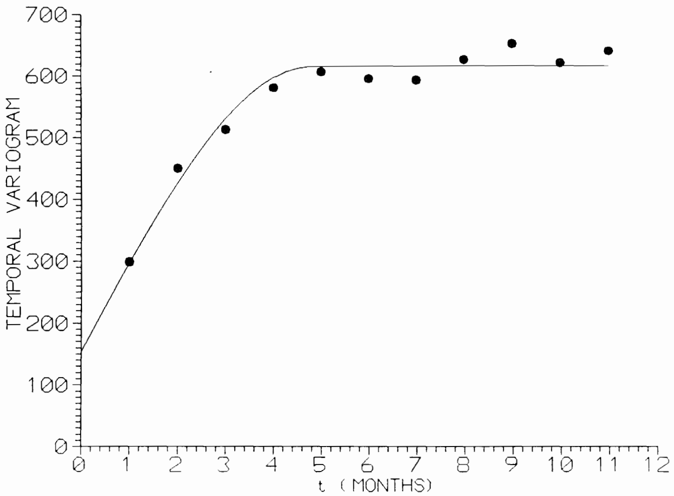

Figure 6. Estimated temporal variogram using the deseasonalized data with the fitted model.

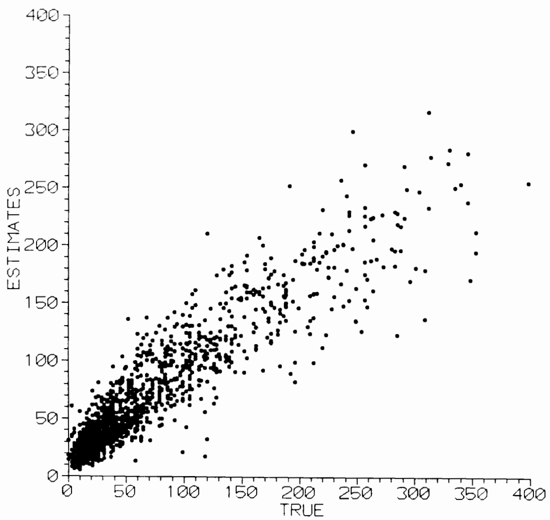

Figure 7. Scatterplot of the true values towards the predicted ones through the cross-validation approach.

In order to obtain a diagnostic check of the fit of the above variogram model towards the raw variogram estimates, the cross-validation approach (Cressie, 1993) has been used. Based on the previous model each datum $(Z(u_j,t_i), j = 1,\ldots,33, t = 1,\ldots 48)$ has been deleted and predicted with $\hat{Z}(u_j,t_i)$ with an associated mean-squared prediction error $\hat{\sigma}^2(u_i,t_j)$. Fig.

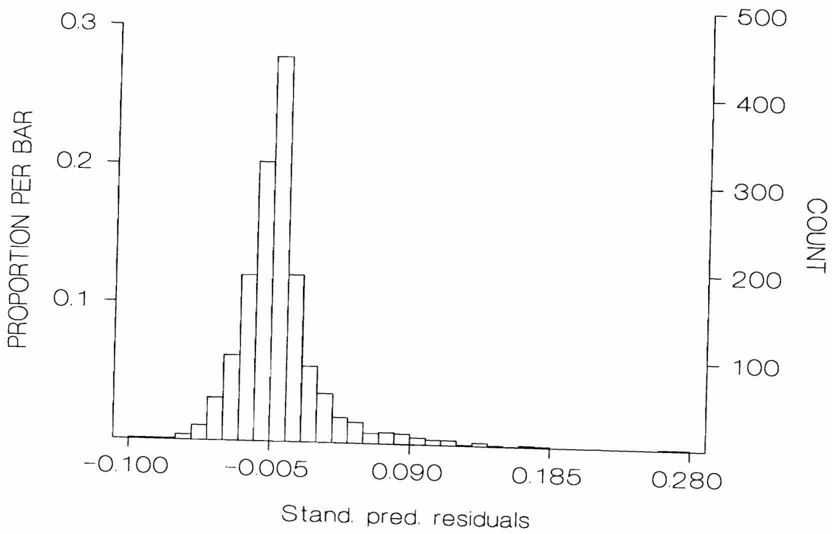

Figure 8. Histogram of the standardized prediction residuals.

7 shows the scatterplot of the true values towards the predicted ones: it exhibits the typical spatial smoothing of kriging, i.e., overestimation of low values and underestimation of high values. The histogram of the standardized prediction residuals (Fig. 8) shows a fairly symmetric distribution that is typically more "peaked" than a normal distribution.

3.3. KRIGING RESULTS

The monthly averages of SO_2 from January 1983 to December 1986 and the previous spatial-temporal variogram model have been used to predict, in January and February 1987, Sulphur-Dioxide by ordinary kriging at the same monitoring stations. The estimated values have been compared to the true ones (Figg. 9-10); of course, these last values have not been used throughout the above analysis. The correlation coefficient between true and predicted values is 0.92 for January 1987 and 0.83 for February of the same year. As expected, the model found gives more plausible estimates in January 1987, because in extrapolation situations, almost by definition, the sample data alone cannot justify the trend model chosen.

Finally, the kriged estimates for January 1987 were obtained using the variogram model given in (6) at the nodes of the kriging grid. Fig. 11 shows the gray-scale map: the highest values are clearly localized in the central part of the graph, corresponding to the city of Milan and some outlying districts, while the peripherical parts of the district are characterized by the lowest values.

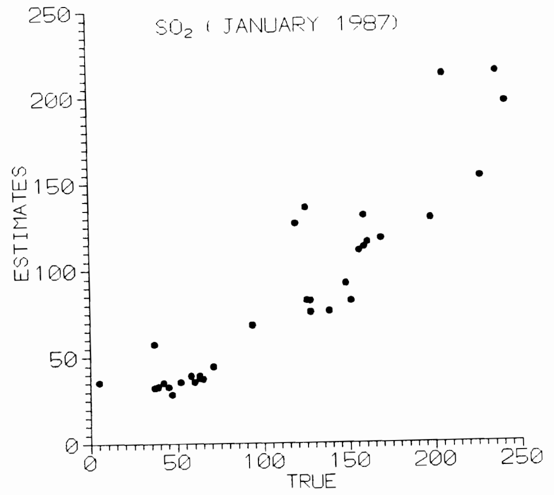

Figure 9. Estimated values compared to the true ones for January 1987.

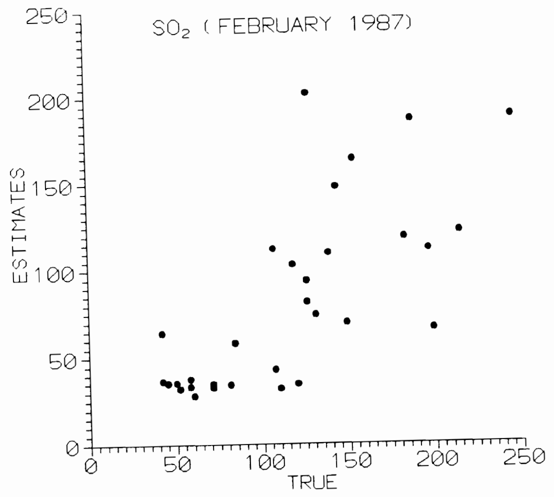

Figure 10. Estimated values compared to the true ones for February 1987.

3.4. SOFTWARE

ITSM software (Brockwell and Davis, 1987) has been used to perform time series analysis at each monitoring station, while GSLIB routines (Deutsch

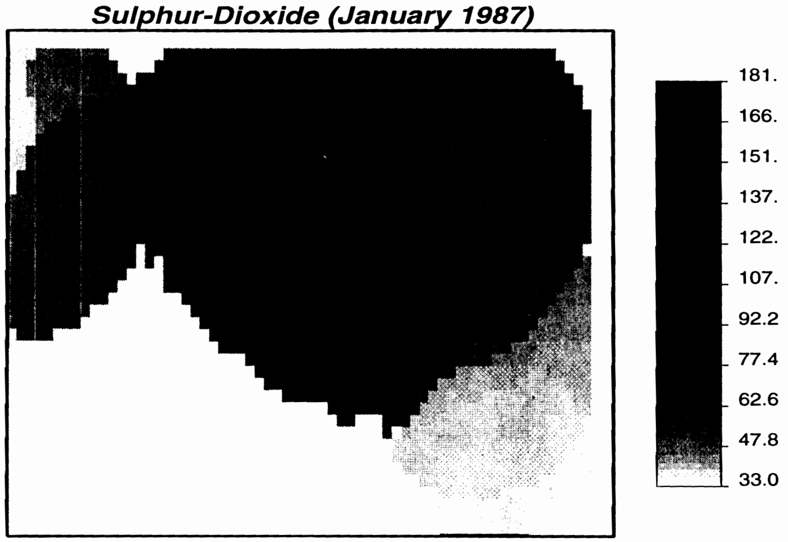

Figure 11. Gray-scale map of the kriged estimates for January 1987.

and Journel, 1992) and VARIOWIN package (Y. Pannatier, 1994) have been used for structural analysis; kriging and cross-validation GSLIB programs have been modified to use the spatial-temporal model given by (6).

4. Conclusions

In this paper ordinary kriging has been applied in the space-time domain in order to study SO_2 measurements in Milan district.

A simple product spatial-temporal covariance model for the residual process has been used to predict, in January and February 1987, Sulphur-Dioxide at the same monitoring stations: the estimated values, compared to the true ones, have shown that the model found, using the data set from January 1983 to December 1986, is quite satisfactory.

References

1. Bilonick, R. A. (1985), The space-time distribution of sulfate deposition in the northeastern United States, *Atmospheric Environment*, 19, 1829-1845.
2. Bilonick, R. A. (1987), Monthly Hydrogen Ion Deposition Maps for the Northeastern U.S. from July 1982 to September 1984, *Consolidated Coal Co.*, Pittsburgh.
3. Brockwell, P. J. and Davis, R. A. (1987) *Time Series: Theory and Methods*, Springer-Verlag, New York.
4. Cressie, N. A. C. (1993) *Statistics for Spatial Data*, J. Wiley and Sons, New York.
5. Deutsch, C. V. and Journel A. G. (1992) *GSLIB: Geostatistical Software Library and User's Guide*, Oxford University Press, New York.
6. Eynon, B. P. and Switzer, P. (1983) The variability of rainfall acidity, *Canadian Journal of Statistics*, 11, 11-24.
7. Journel, A. G. (1986) Geostatistics: Models and Tools for the Earth Sciences, *Mathematical Geology*, 18, 1, 119-140.
8. Le, D. N. and Petkau, A. J. (1988) The variability of rainfall acidity revisited, *Canadian Journal of Statistics*, 16, 15-38.
9. Myers, D. E. and Journel A. G. (1990) Variograms with Zonal Anisotropies and non-invertible Kriging Matrices, *Mathematical Geology*, 22, 779-785.
10. Myers, D. E. (1992) Spatial-Temporal Geostatistical Modeling in Hydrology, *Proceedings of the International Workshop*, Karlsruhe (Germany).
11. Rodriguez-Iturbe, I. and Meija, J. M. (1974) The Design of Rainfall Networks in Time and Space, *Water Resources Research*, 10, 713-728.
12. Rouhani, S. and Hall, T. J. (1989) Space-Time Kriging of Groundwater Data, M. Armstrong (ed.), *Geostatistics*, Kluwer Academic Publishers, 2, 639-651.
13. Soares, A., Tavora, J., Pinheiro, L., Freitas, C. and Almeida J. (1992) Predicting Probability Maps of Air Pollution Concentration: A Case Study on Barreiro/Seixal Industrial Area, A. Soares (ed.), *Geostatistics*, Kluwer Academic Publishers, 2, 625-636.
14. Stein, M. L. (1986) A simple model for spatial-temporal processes, *Water Resources Research*, 22, 2107-2110.
15. *Studi per la valutazione della qualità dell'aria nella provincia di Milano*, 1993, Province of Milan and U.SS.L. 75/III (Local Health Department) of Milan.