FORTRAN programs for space-time modeling [☆]

L. De Cesare[a,b], D.E. Myers[c], D. Posa[a,d,*]

[a] *Facoltà di Economia, Dipartimento di Scienze Economiche, e Matematico-Statistiche, Via per Monteroni, ECOTEKNE, 73100 Lecce, Italy*
[b] *I.R.M.A.-CNR, Bari, Italy*
[c] *Dept of Mathematics, University of Arizona, Tucson AZ, USA*
[d] *I.R.M.A.-CNR, Bari, Italy*

Abstract

Modified GSLIB FORTRAN 77 routines are given in this paper for estimating and modeling space-time variograms. Two general families of models are incorporated in the programs: these are the product model and the product-sum model, both based on the decomposition of the space-time covariance in terms of a space covariance and a time covariance. The GSLIB kriging program has also been modified to incorporate these space-time models. One of the programs detects and removes temporal periodicities in the data. The program removes them and generates residuals for all monitoring stations, in order to estimate and model the spatial-temporal variogram using residuals. The modified kriging program also allows the use of cross-validation in conjunction with fitting of space-time variogram models. The trend component and the residual variogram model can be used for prediction. To illustrate the use of the programs, hourly averages of NO_2 for the first ten months of 1998 in Lombardy were used. © 2002 Elsevier Science Ltd. All rights reserved.

Keywords: Geostatistics; Space-time models; Time series; Product models; Product-sum models

1. Introduction

Recently considerable attention has been given to methods for analyzing spatial-temporal phenomena (Host et al., 1995; Meiring et al., 1998; Rouhani and Wackernagel, 1990; Solna and Switzer, 1996). For example, a large number of environmental data sets (Bilonick, 1985; Christakos and Vyas, 1998; Eynon and Switzer, 1983; Le and Petkau, 1988; Soares et al., 1993) have been treated as realizations of space-time random

functions. A recent review of geostatistical space-time models was given by Kyriakidis and Journel (1999). De Cesare et al. (2001) include a brief comparison of the product-sum covariance model and other classes of spatial-temporal covariance models that have appeared in the literature. These include the metric model (Dimitrakopoulos and Luo, 1994), the product model (Rodriguez-Iturbe and Meija, 1974), the linear model (Rouhani and Hall, 1989), the product model (De Cesare et al., 1997) and the Fourier product model (Cressie and Huang, 1999).

FORTRAN 77 programs are described for estimating and modeling the product and the product-sum covariance models (in variogram form). To simplify determining the three coefficients in the product-sum model, two constraints are imposed on these coefficients. These constraints ensure that the model is valid and simplify estimating the coefficients, i.e., fitting the model to the data.

Although geostatistics offers a variety of methods to extend existing spatial techniques to the space-time domain, there are a number of theoretical and practical issues that must be addressed prior to successful application of geostatistical methods to space-time data. These include

- identification and removal of daily cycles and seasonal trends, in order to estimate and model the space-time variogram of the residuals,
- choosing valid space-time covariance and variogram models,
- applying some form of validation, to check the fit of the variogram model versus the estimated variogram values (Cressie, 1993; Myers, 1991).

The data is treated as a non-random sample from a realization of a spatial-temporal random field

$$(Z(s,t); s \in D, t \in T) \qquad (1)$$

with $D \subseteq \mathbb{R}^2$ and $T \subseteq \mathbb{N}$, which allows (1) to be viewed as a time series of spatial processes, each process occurring at equally spaced time points. Assuming that the first and second moments of Z exist, Z can be decomposed as

$$Z(s,t) = m(s,t) + Y(s,t), \qquad (2)$$

where Y is a second order stationary stochastic process with

$$E(Y(s,t)) = 0$$

and $m(s,t)$ is the mean function of Z; the spatial-temporal covariance function

$$C_{st}(h) = \mathrm{Cov}(Y(s+h_s, t+h_t), Y(s,t)), \qquad (3)$$

where $h = (h_s, h_t) \in \mathbb{R}^2 \times \mathbb{R}$, $(s, s+h_s) \in D^2$ and $(t, t+h_t) \in T^2$ and the spatial-temporal variogram

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

depend solely on the lag vector h, and not on the location in space or time.

2. Missing values and removal of seasonal components

A space-time data set can be viewed as a collection of long time series, one for each spatial location. For example, hourly data available for one or more years. For this kind of data set there are likely two deterministic cycles, namely the annual cycle (due to the alternation of seasons) and the diurnal one with period d. Hence, to take into account the above features, it is possible to split the time series for each location in p pieces (i.e. months). That is, the time interval T is partitioned such that $T = [\tau_0, \tau_1[\cup[\tau_1, \tau_2[\cup \cdots \cup [\tau_{p-1}, \tau_p]$ where $\tau_0 < \tau_1 < \cdots < \tau_p$, then the trend $m(s,t)$ can be modeled in the following way:

1. $\alpha_i(s,t) = \alpha_i(s, t+d) \ \forall s \in D \ \forall t, t+d \in [\tau_{i-1}, \tau_i[,$ $i = 1, \ldots, p$.
2. $\sum_{j=1}^{d} \alpha_i(s,j) = 0 \ \forall s \in D, \ i = 1, \ldots, p$.
3. $m(s,t) = \alpha_i(s,t) + \mu_i \ \forall s \in D \ \forall t, t+d \in [\tau_{i-1}, \tau_i[,$ $i = 1, \ldots, p$.

$\alpha_i(s,t)$ is called *seasonal component* for the ith "month" and μ_i is a constant trend for the same "month". In order to estimate the seasonal component, time series analysis can be applied to each monitoring station using the standard technique of moving average estimation (MAE) (Brockwell and Davis, 1987). To implement such a technique, values are needed at equally spaced time points.

In many data sets there are missing values, hence interpolation in time is required in order to apply the time series analysis. The FORTRAN routine REMO-VE.FOR performs time interpolation and removes the trend component.

Input parameters for the program REMOVE

- the number p of "months",
- the values $\tau_0, \tau_1, \ldots, \tau_p$,
- the period d of the seasonal component,
- the minimum number n of consecutive values to which the MAE method is applied,
- the maximum number k of consecutive missing values to which the linear interpolation is applied.

Procedure description

(a) Sequences with fewer than k consecutive missing values are linearly interpolated (there is no interpolation if there are more than k consecutive missing values).
(b) The MAE method is applied to sequences with more than n consecutive values: for each of these sequences the seasonal components are separately computed. Sequences with fewer than n consecutive values are not considered: for such sequences there are not enough values to estimate the seasonal component.

3. Some space-time covariance models

For the process Y denote by:
γ_{st}, the spatiotemporal variogram, γ_t, the temporal variogram, γ_s, the spatial variogram, C_t, the temporal covariance, C_s, the spatial covariance, $C_{st}(0,0)$, the "sill" of γ_{st}, $C_s(0)$, the sill of γ_s, $C_t(0)$, the sill of γ_t.

The FORTRAN 77 programs described in this paper incorporate the following spatial-temporal covariance/variogram models for Y.

- *The Product model*. One of the simplest ways to model a covariance in space-time is to separate the

dependence in space and in time (Rodriguez-Iturbe and Meija, 1974; De Cesare et al., 1997). The product spatial-temporal covariance model is:

$$C_{st}(h_s, h_t) = kC_s(h_s)C_t(h_t). \tag{4}$$

The product covariance model could be re-written in terms of the spatial-temporal variogram

$$\gamma_{st}(h_s, h_t) = k(C_t(0)\gamma_s(h_s) + C_s(0)\gamma_t(h_t) - \gamma_s(h_s)\gamma_t(h_t)).$$

The parameter k is determined by Eq. (4):

$$k = \frac{C_{st}(0,0)}{C_s(0)C_t(0)}.$$

Hence, it can be estimated by estimating $C_{st}(0,0)$ (called the "global" sill) and the spatial and temporal sills $C_s(0)$ and $C_t(0)$, respectively.

If in Eq. (4) C_s is a positive-definite function in $\mathbb{R}^d$ and C_t is a positive-definite function in $\mathbb{R}$, then the product model is positive definite in space-time (Cressie, 1993). However, the class (4) is severely limited, since for any pair of spatial locations the cross-covariance function of the two time series always has the same shape. In fact, for any two fixed spatial lags h_1 and h_2

$$C_{st}(h_1, h_t) \propto C_{st}(h_2, h_t).$$

A similar result holds for any pair of time points and the cross-covariance function of the two spatial processes.

- *The Product-sum model.* The product model can easily be extended as follows:

$$C_{st}(h_s, h_t) = k_1 C_s(h_s)C_t(h_t) + k_2 C_s(h_s) + k_3 C_t(h_t). \tag{5}$$

This is a valid model in space time if the separate space and time models are valid and if $k_1 > 0$. This model can be re-written in terms of the spatial-temporal variogram

$$\gamma_{st}(h_s, h_t) = (k_1 C_s(0) + k_3)\gamma_t(h_t) + (k_1 C_t(0) + k_2)\gamma_s(h_s) - k_1\gamma_s(h_s)\gamma_t(h_t). \tag{6}$$

In order to assure the admissibility of model Eq. (4), C_s and C_t must be positive-definite functions (Cressie, 1993). In addition, the coefficients k_2 and k_3 must be non negative while k_1 must be strictly positive.

The following condition is implicit in the transformation from covariance form to variogram form:

$$k_1 C_s(0)C_t(0) + k_2 C_s(0) + k_3 C_t(0) = C_{st}(0,0). \tag{7}$$

Note also from Eq. (6) that

$$\gamma_{st}(h_s, 0) = [k_2 + k_1 C_t(0)]\gamma_s(h_s). \tag{8}$$

and

$$\gamma_{st}(0, h_t) = [k_3 + k_1 C_s(0)]\gamma_t(h_t). \tag{9}$$

To determine the coefficients, three equations are needed. Two of these are generated by assuming that

$$k_2 + k_1 C_t(0) = 1$$
$$k_3 + k_1 C_s(0) = 1. \tag{10}$$

Thus to estimate and model $\gamma_s(h_s)$ and $\gamma_t(h_t)$ it is sufficient to estimate and model $\gamma_{st}(h_s, 0)$ and $\gamma_{st}(0, h_t)$, respectively.

From Eq. (7), and Eq. (10) we obtain

$$k_1 = [C_s(0) + C_t(0) - C_{st}(0,0)]/C_s(0)C_t(0),$$

$$k_2 = [C_{st}(0,0) - C_t(0)]/C_s(0),$$

$$k_3 = [C_{st}(0,0) - C_s(0)]/C_t(0). \tag{11}$$

In modeling the separate spatial and temporal variograms it is necessary to ensure that the sills are chosen so that the numerators in Eq. (11) remain positive.

The main advantage of using models Eqs. (4) and (5) is that these are completely determined by the temporal variogram γ_t and the spatial variogram γ_s. The severe shortcomings of model Eq. (4) were mentioned above; on the other hand, restrictions Eq. (10) on model Eq. (5) impose a form of symmetry on the same model, i.e., symmetry between the impact of the spatial component and the temporal correlation component.

The spatial-temporal sill $C_{st}(0,0)$ can be estimated from a plot of the spatial-temporal variogram surface γ_{st}: this variogram surface can be obtained by using a modified version of the GAMV routine of the GSLIB package (Deutsch and Journel, 1997). This modification computes the values of the sample space-time variogram, γ_{st}, at the spatial lag h_s and temporal lag h_t. The empirical spatial and temporal variograms can be then be easily extracted (put $h_t = 0, h_s = 0$, respectively).

The COVA3 routine of the GSLIB package, which computes the covariance function, has been modified. The new routine allows the use of a product variogram model (4) or a product-sum variogram model Eq. (5). The input parameters for this routine include the "global" sill $C_{st}(0,0)$, the spatial and the temporal nugget effect and the spatial and the temporal variograms. The corresponding parameter file and the input routine READPARM have been modified in order to take into account the above corrections.

4. Cross-validation and prediction

To implement cross-validation and spatial-temporal prediction using a product or a product-sum variogram

model, the KT3D GSLIB program is modified by substituting the new COVA3 subroutine. The only other modification to KT3D concerns the substitution at the beginning of the line code *covmax = c0(1)* with *covmax = cc(nst(1) + 1)* because the maximum covariance actually is the "global" sill. As noted above this parameter can be estimated by plotting the spatial-temporal variogram surface.

Because space and time are treated quite differently, it is necessary to choose the search radii and the anisotropy ratios separately. The variogram models available for use either as space variograms or as time variograms are the same ones as in the GSLIB software library. In the program code the spatial variogram models are identified by the GSLIB code number, while the temporal ones are specified by adding 5 to the previous code number.

5. A numerical example

The data collection network, planned to the standards and information provided by the Lombardy Region's Environmental Town Council, consists of 52 survey stations for hourly averages of NO_2 measured from the 1st to the 31st of October 1998.

Residuals were generated for all stations after removal of the seasonal component by the standard technique of moving averages described in Section 2. The parameter file used by the FORTRAN routine REMOVE.FOR is shown in the Appendix.

Fig. 1(a) shows the sample temporal variograms of the original data and the residuals with the fitted model: Note the periodic structure at 24 h exhibited by the variogram of the original data. Fig. 1(b) shows the sample spatial variogram of the residuals with the fitted model.

The "global" sill value $C_{st}(0,0)$ of γ_{st} has been estimated through the spatial-temporal variogram surface of the residuals (see Fig. 2).

Cross-validation, using the modified KT3D routine, was used to check the fit of the above variogram model versus the sample variogram(s). The parameter file used for cross-validation is shown in the Appendix: The correlation coefficient between the residual values and the predicted ones is $\rho = 0.9$.

5.1. Spatial-temporal prediction

The hourly average data for NO_2 measured during October 1998 together with the product-sum variogram model given in (6) were used to predict the hourly averages of NO_2 on the 1st of November 1998, using ordinary kriging at the same monitoring stations available on the first of November. The daily averages of the predicted hourly values were then compared to the daily averages of the true hourly values (Fig. 3); this latter data was not used in making the predictions. The correlation coefficient between true and predicted values is 0.9.

In order to evaluate the evolution of the hourly averages of NO_2 for the 1st of November 1998, the predicted values, at the same locations where the true values are available, were computed in the following way:

1. the residual product-sum variogram model was used to predict by ordinary kriging the residuals for the 1st
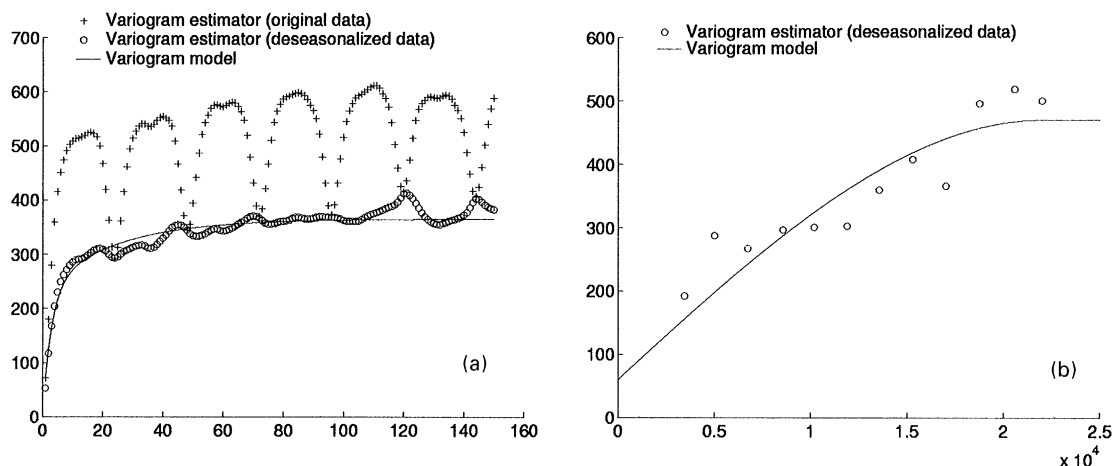

Fig. 1. Sample temporal variograms of original data and residuals with fitted model (a); sample spatial variogram of residuals with fitted model (b).
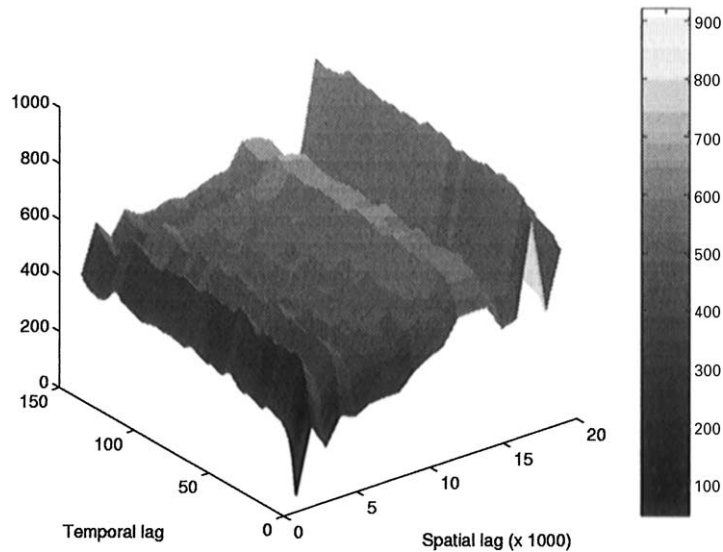
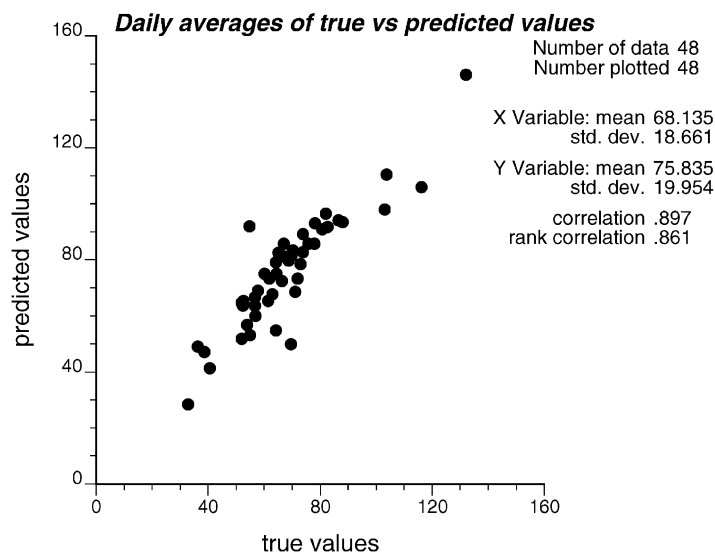Fig. 2. Spatial-temporal variogram surface of residuals.

Fig. 3. Daily averages of true hourly values versus daily averages of predicted hourly values on 1st November 1998.

of November 1998 at the same locations where the true values are available,

2. the seasonal component corresponding to October 1998, which was previously estimated, has been added to the predicted residuals in order to obtain the predicted values.

In the above, it is assumed that the trend component does not change from one month to the next.

Fig. 4 shows the spatial plots of the daily averages of the true hourly values and the daily averages of the predicted hourly values on the 1st of November 1998. The highest values are clearly localized in the central part of the graph, corresponding to the city of Milan and outlying districts, while the peripheral parts of the district are characterized by the lowest values.
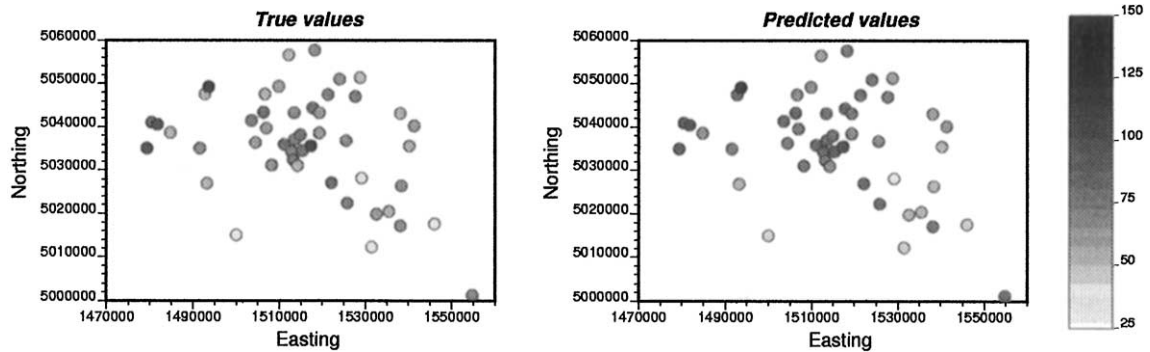
Fig. 4. Postplots of daily averages of true hourly values and daily averages of predicted hourly values on 1st November 1998.

Appendix

6.1. GAMV.PAR

```
                         Parameters for GAMV
                         ********************

START OF PARAMETERS:
ott-98.dst                    \datafile
2  3  4                       \column for x, y, t coordinates
1  4                          \nvar; column numbers ...
−9000.   1.0e21               \tmin, tmax (trimming limits)
vmapdest.var                  \output file for variograms
18                            \nlag – the number of spatial lags
1700                          \xlag – unit separation distance
850                           \xltol – lag tolerance
144                           \ntlag – the number of temporal lags
1.                            \tlag – unit separation distance
1                             \ndir – number of directions
0.0  90.0  15000.  0. 0. 0.   \azm, atol, bandh, dip, dtol, bandv
1                             \standardize sills? (0=no, 1=yes)
1                             \number of variograms 1 1
1                             \tail, head, variogram type
```

6.2. KT3D.PAR

```
                         Parameters for KT3D
                         ********************

START OF PARAMETERS:
ott-98.dst                    \file with data
1  2  3  4  0                 \columns for X, Y, Z, var, sec var
−2000 1.0e21                  \trimming limits
1                             \option: 0=grid, 1=cross, 2=jackknife
xvk.dat                       \file with jackknife data
1  2  0  3  0                 \columns for X, Y, Z, vr and sec var
2                             \debugging level: 0, 1, 2, 3
ckt3d.dbg                     \file for debugging output
```

```
ckt3d.dat                       \file for kriged output 1
1460000   750000               \nx,xmn,xsiz
1  5000000  700000             \ny,ymn,ysiz
1  15313   400                 \nz,zmn,zsiz
1  1  1                        \x,y and z block discretization
2  10                          \min, max data for kriging
0                              \maxper octant (0-> not used)
15000   15000   50             \maximum search radii
0.0   0.0   0.0                \angles for search ellipsoid
1  6.                          \0=SK,1=OK,2=non-st SK,3=exdrift
0 0 0 0 0 0 0 0 0              \drift: x, y, z, xx, yy, zz, xy, xz, zy
0                              \0, variable; 1, estimate trend
extdrift.dat                   \gridded file with drift/mean
4                              \column number in gridded file
3  60.                         \nst, spatial nugget e⌷ect
1  410. 0.0 0.0 0.0            \it, cc, ang1, ang2, ang3
   22000   22000   22000       \a_hmax, a_hmin, a_vert
7  235.   0.0   0.0   0.0      \it, cc, ang1, ang2, ang3
    9.0  9.0  9.0              \a_hmax, a_hmin, a_vert
7  130.   0.0   0.0   0.0      \it, cc, ang1, ang2, ang3
    72.0  72.0  72.0           \a_hmax, a_hmin, a_vert
0  500                         \temporal nugget, global sill
1                              \0=product, 1=sum-product
```

6.3. REMOVE.PAR

```
                    Parameters for REMOVE
                    *********************


START OF PARAMETERS:
ott-98.txt                     \data file
2   3   4                      \column for x, y, t coordinates
5                              \var column number
−9000.    1.0e21               \tmin, tmax (trimming limits)
ott-98.dst                     \output file
5                              \Max number of missing consecutive values
24                             \period
60                             \Min number of data for deseasonalization
1                              \time unit
1                              \number of "months"
15313                          \months+1 values of months denition
16056
```

References

Bilonick, R.A., 1985. The space-time distribution of sulfate deposition in the northeastern United States. Atmospheric Environment 19, 1829–1845.

Brockwell, P.J., Davis, R.A., 1987. Time Series: Theory and Methods, Springer, New York, 577pp.

Christakos, G., Vyas, V.M., 1998. A composite space-time approach to studying zone distribution over eastern United States. Atmospheric Environment 32 (16), 2845–2857.

Cressie, N.A.C., 1993. Statistics for Spatial Data, Wiley, New York, 900pp.

Cressie, N.A.C., Huang, H., 1999. Classes of nonseparable, spatio-temporal stationary covariance functions. Journal of the American Statistical Association 94, 1330–1340.

De Cesare, L., Myers, D., Posa, D., 1997. Spatial-temporal modeling of SO₂ in Milan district. In: Baafi, E.Y., Schofield, N.A. (Eds.), Geostatistics Wollongong '96. Kluwer Academic Publishers, Dordrecht, Vol. 2, 1031–1042.

De Cesare, L., Myers, D., Posa, D., 2001. Estimating and modeling space-time correlation structures. Statistics and Probability Letters 51 (1), 9–14.

Deutsch, C.V., Journel, A.G., 1997. GSLIB: Geostatistical Software Library and User's Guide, 2nd edn. Oxford University Press, New York, 369pp.

Dimitrakopoulos, R., Luo, X., 1994. Spatiotemporal modeling: covariances and ordinary kriging systems In: Dimitrakopoulos, R. (Ed.), Geostatistics for the Next Century, Kluwer Academic Publishers, Dordrecht, 88–93.

Eynon, B.P., Switzer, P., 1983. The variability of rainfall acidity. Canadian Journal of Statistics 11, 11–24.

Host, G., Omre, H., Switzer, P., 1995. Spatial interpolation errors for monitoring data. Journal of the American Statistical Association 90 (431), 853–861.

Kyriakidis, P.C., Journel, A.G., 1999. Geostatistical space-time models: a review. Mathematical Geology 31 (6), 651–684.

Le, D.N., Petkau, A.J., 1988. The variability of rainfall acidity revisited. Canadian Journal of Statistics 16, 15–38.

Meiring, W., Guttorp, P., Sampson, P.D., 1998. Space-time estimation of gridcell hourly ozone levels for assessment of a deterministic model. Environmental and Ecological Statistics 5, 197–222.

Myers, D.E., 1991. On variogram estimation. Proceedings of the First International Conference Statistical Computing, Cesme, Turkey, 30 March–2 April 1987, Vol. II. American Sciences Press, pp. 261–281.

Rodriguez-Iturbe, I., Meija, J.M., 1974. The design of rainfall networks in time and space. Water Resources Research 10, 713–728.

Rouhani, S., Hall, T.J., 1989. Space-time kriging of groundwater data. Proceedings 3rd International Geostatistics Congress. Avignon, France, Kluwer Academic Publishers, Dordrecht, Vol. 2. pp. 639–651.

Rouhani, S., Wackernagel, H., 1990. Multivariate geostatistical approach to space-time data analysis. Water Resources Research 26 (4), 585–591.

Soares, A., Tavora, J., Pinheiro, L., Freitas, C., Almeida J., 1993. Predicting probability maps of air pollution concentration: A case study on Barreiro/Seixal industrial area. In: Soares, A. (Ed.), Geostatistics Troia '92. Kluwer Academic Publishers, Dordrecht, Vol. 2. pp. 625–636.

Solna, K., Switzer, P., 1996. Time trend estimation for a geographic region. Journal of the American Statistical Association 91, 577–589.