ELSEVIER

Statistics & Probability Letters 51 (2001) 9–14

STATISTICS & PROBABILITY LETTERS

www.elsevier.nl/locate/stapro

Estimating and modeling space–time correlation structures

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

[a] *II Facoltà di Economia, Università di Bari, Italy*
[b] *Istituto per Ricerche di Matematica Applicata (CNR), Bari, Italy*
[c] *Department of Mathematics, University of Arizona, 617 N. Santa Rita, Tucson, AZ 85721, USA*
[d] *Facoltà di Economia, Università di Lecce, Italy*

Received November 1999; received in revised form March 2000

Abstract

In this paper, a class of product–sum covariance models has been introduced for estimating and modeling space–time correlation structures. It is shown how the coefficients of this class of models are related to the global sill and "partial" spatial and temporal sills; moreover, some constraints on these sills have been given in order to assure positive definiteness of the product–sum covariance model. A brief comparative study with some other classes of spatial–temporal covariance models has been pointed out.

Keywords: Space–time correlation; Geostatistics

1. Introduction

A large number of environmental phenomena may be regarded as realizations of space–time random fields (Eynon and Switzer, 1983; Le and Petkau, 1988). Geostatistics offers a variety of methods to model spatial data; however, applying such space-oriented approaches to spatiotemporal processes, may lead to the loss of valuable information in the time dimension.

One obvious solution to this problem is to consider the spatiotemporal phenomenon as a realization of a random field defined in $\Re^{d+1}$ (i.e. d is the physical space dimension plus one time dimension). This approach demands the extension of the existing spatial techniques into the space–time domain. Despite the straightforward appearance of this extension, there are a number of theoretical and practical problems that should be addressed prior to any successful application of geostatistical methods to space–time data.

In order to point out one of such difficulties let $Z = \{Z(s),\ s \in D \subseteq \Re^d\}$ be a real-valued spatial random field defined on the domain D of the d-dimensional space $\Re^d$, where $d \leqslant 3$. If Z is second-order stationary,

* Corresponding author. Tel.: +1-520-621-6859; fax: +1-520-621-8322.
E-mail address: myers@math.arizona.edu (D.E. Myers).

0167-7152/01/$ - see front matter
PII: S0167-7152(00)00131-0

its covariance function

$$C(h_s) = \mathrm{Cov}[Z(s+h_s), Z(s)], \quad \forall s,\ s+h_s \in D \tag{1}$$

depends only on the separation vector between a pair of points $s+h_s$, s. In geostatistics the second-order stationarity assumption is ordinarily weakened slightly to require only that first-order increments be second-order stationary, then the spatial correlation is quantified by the variogram of Z

$$\gamma(h_s) = \tfrac{1}{2}\,\mathrm{Var}[Z(s+h_s) - Z(s)], \quad \forall s,\ s+h_s \in D. \tag{2}$$

When Z is second-order stationary, $\gamma(h_s) = C(0) - C(h_s)$; moreover, if the covariance or variogram is only a function of the length of h_s, then the covariance or variogram is called isotropic. In practice, anisotropic models are constructed as positive linear combinations of isotropic models where the "lengths" are obtained by affine transformations (Journel and Huijbregts, 1981): this is called a geometric anisotropy because the range of each component changes with direction. Since the variogram or covariance must satisfy a certain positive definiteness condition (variograms must be conditionally negative definite), one ordinarily fits a sample function to standard isotropic models (Cressie, 1991; Myers, 1991).

Unfortunately, the technique of reducing the modeling process down to a one-dimensional problem does not work so well for spatial–temporal models. If time is simply considered as another "dimension" then it would be necessary to have an appropriate metric in space–time (Bilonick, 1985). The technique of separability is used instead. Note that, in general, the sum of a spatial covariance (or variogram) and a temporal covariance (respectively, a temporal variogram) will not be strictly positive definite and in that case the coefficient matrix in the kriging equations may be non-invertible for some sample location patterns (as shown in Myers and Journel, 1990; Rouhani and Myers, 1990; Myers, 1992). While the product of two covariances will be a covariance, in general, the product of two variograms is not a variogram.

In this paper, a class of product–sum covariance models has been introduced, in order to estimate and model in a flexible way realizations of space–time random fields, which are very common in environmental applications (De Cesare et al., 1999). Some constraints on the coefficients of this class of models have been given in order to guarantee positive definiteness of the covariance model.

An overview and a short comparative study with some classes of space–time covariance models used in some applications has been given.

2. Some classes of spatiotemporal covariance models

In this section some classes of spatiotemporal covariance models, which have been used in literature, are discussed, in order to compare these last with the product–sum model proposed by the authors. In order to distinguish between space and time, let $Z = \{Z(s,t),\ (s,t) \in D \times T\}$ be a second-order stationary spatial–temporal random field, where $D \subset \Re^d$ and $T \subset \Re_+$, with expected value

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

In this case, the covariance

$$C_{st}(h) = \mathrm{Cov}(Z(s+h_s, t+h_t), Z(s,t)) \tag{4}$$

where $h = (h_s, h_t)$, $(s, s+h_s) \in D^2$ and $(t, t+h_t) \in T^2$ and the variogram

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

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

The function C_{st} in (4) must satisfy a positive definiteness condition in order to be a valid covariance function. That is, for any $(s_i, t_i) \in D \times T$, any $a_i \in \Re$, $i = 1, \ldots, n$, and any positive integer n, C_{st} must satisfy:

$$\sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j C_{st}(s_i - s_j, t_i - t_j) \geqslant 0.$$

For continuous functions, positive definiteness is equivalent to the process having a spectral distribution function (Matern, 1960).

1. *The metric model.* It is assumed (Dimitrakopoulos and Luo, 1994) that

$$C_{st}(h_s, h_t) = C(a^2|h_s|^2 + b^2 h_t^2) \tag{5}$$

where the coefficients $a, b \in \Re$. Note that in (5) the same type of model is assumed for the spatial and temporal covariances, with possible changes in the range.

2. *The product model.* One of the simplest ways to model a covariance in space–time is to separate the dependence on the two (Rodriguez-Iturbe and Meija, 1974; De Cesare et al., 1996). The product spatial–temporal covariance model is

$$C_{st}(h_s, h_t) = C_s(h_s) C_t(h_t), \tag{6}$$

where spatial dependence is separated by the temporal one. Posa (1993) used the above model which assumes a range of influence independent of the time.

The previous model could be easily written in terms of the spatial–temporal variogram

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

where γ_{st} is the spatiotemporal variogram, γ_t the temporal variogram, γ_s the spatial variogram, C_t the temporal covariance and C_s the spatial covariance.

In (6) C_s is a positive-definite function in $\Re^d$ and C_t is a positive-definite function in $\Re$; admissible spatial covariance models and admissible temporal covariance models are readily available (Cressie, 1991) and hence they can be combined in product form to give spatiotemporal covariance models.

However, class (6) 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.

3. *The linear model.* Another type of separability involves adding spatial and temporal covariances (Rouhani and Hall, 1989), that is

$$C_{st}(h_s, h_t) = C_s(h_s) + C_t(h_t). \tag{7}$$

For this model, covariance matrices of certain configurations of spatiotemporal data are singular (Myers and Journel, 1990; Rouhani and Myers, 1990): in this case, the covariance function is only positive semidefinite and it is unsatisfactory for optimal prediction.

4. *The nonseparable model.* A new approach that allows to obtain classes of nonseparable, spatiotemporal stationary covariance functions has been derived by Cressie and Huang (1999). The authors assume that

$$H(\omega, h_t) = \rho(\omega, h_t) K(\omega), \tag{8}$$

where

$$H(\omega, h_t) = (2\pi)^{-d} \int \mathrm{e}^{-\mathrm{i} h_s^T \omega} C_{st}(h_s, h_t)\, \mathrm{d}h_s$$

and the following two conditions are satisfied:

- for each $\omega \in \Re^d$, $\rho(\omega, .)$ is a continuous autocorrelation function and $K(\omega) > 0$;
- the positive function $K(\omega)$ satisfies:

$$\int K(\omega)\,\mathrm{d}\omega < \infty.$$

3. The product–sum model

A very general model combining products and sums can be obtained in the following way:

$$C_{st}(h_s, h_t) = k_1 C_s(h_s) C_t(h_t) + k_2 C_s(h_s) + k_3 C_t(h_t), \tag{9}$$

or equivalently

$$\gamma_{st}(h_s, h_t) = [k_2 + k_1 C_t(0)]\gamma_s(h_s) + [k_3 + k_1 C_s(0)]\gamma_t(h_t) - k_1 \gamma_s(h_s)\gamma_t(h_t), \tag{10}$$

where C_s and C_t are covariance functions and γ_s and γ_t are the corresponding variogram functions. Note that $C_{st}(0)$ is the "sill" of γ_{st}, $C_s(0)$ is the sill of γ_s and $C_t(0)$ is the sill of γ_t. Note also that by definition $\gamma_{st}(0,0) = \gamma_s(0) = \gamma_t(0) = 0$.

3.1. Estimation and modeling

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

$$k_1 C_s(0) C_t(0) + k_2 C_s(0) + k_3 C_t(0) = C_{st}(0,0). \tag{11}$$

Note also from Eq. (10) that

$$\gamma_{st}(h_s, 0) = [k_2 + k_1 C_t(0)]\gamma_s(h_s), \tag{12}$$

$$\gamma_{st}(0, h_t) = [k_3 + k_1 C_s(0)]\gamma_t(h_t). \tag{13}$$

It is assumed that

$$k_2 + k_1 C_t(0) = 1, \qquad k_3 + k_1 C_s(0) = 1 \tag{14}$$

in order to estimate and model $\gamma_s(h_s)$ and $\gamma_t(h_t)$ by $\gamma_{st}(h_s, 0)$ and $\gamma_{st}(0, h_t)$, respectively.

Let H be the set of data locations, then the standard moment estimator for the spatial variogram at the vector lag r_s, with spatial tolerance δ_s, is

$$\hat{\gamma}_s(r_s) = \hat{\gamma}_{st}(r_s, 0) = \frac{1}{2|N(r_s)|} \sum [Z(s + h_s, t) - Z(s,t)]^2, \tag{15}$$

where the summation is over the set

$$N(r_s) = \{(s + h_s, t) \in H \text{ and } (s,t) \in H \text{ such that } \|r_s - h_s\| < \delta_s\},$$

and $|N(r_s)|$ is the cardinality of this set. Similarly

$$\hat{\gamma}_t(r_t) = \hat{\gamma}_{st}(0, r_t) = \frac{1}{2|M(r_t)|} \sum [Z(s, t + h_t) - Z(s,t)]^2, \tag{16}$$

where

$$M(r_t) = \{(s, t + h_t) \in H \text{ and } (s,t) \in H \text{ such that } \|r_t - h_t\| < \delta_t\}$$

and $|M(r_t)|$ is the cardinality of this set. Usually, the spatial locations need not be on a regular grid, while the temporal points are regularly spaced and hence it is not necessary to use temporal distance classes.

3.2. Results on positive definiteness

From Eq. (9) it is clear that $k_1 > 0$, $k_2 \geqslant 0$ and $k_3 \geqslant 0$ is a sufficient condition for positive definiteness. From Eqs. (11) and (14) we obtain

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

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

$$k_3 = [C_{st}(0,0) - C_s(0)]/C_t(0). \tag{17}$$

In modeling the separate spatial and temporal variograms it is necessary to ensure that the sills are chosen so that the numerators in Eqs. (17) remain positive.

4. Some general comments

A brief comparison between the classes of models described in Section 2 and the product–sum covariance model proposed by the authors is made in this section.

- note that if the autocorrelation function ρ in (8) is purely a function of h_t, then the product covariance model is obtained;
- the product model and the linear model are easily obtained by the product–sum covariance model setting, respectively, $k_2 = k_3 = 0$ and $k_1 = 0$;
- the product–sum covariance model is more flexible than the nonseparable covariance model for estimating and modeling spatial–temporal correlation structures, as described in Section 3.1.

References

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

Cressie, N., 1991. Statistics for Spatial Data. Wiley, New York, reprinted 1993.

Cressie, N., Huang, H., 1999. Classes of Nonseparable, Spatiotemporal Stationary Covariance Functions. J. Amer. Statist. Assoc. 94, 1330–1340.

De Cesare, L., Myers, D., Posa, D., 1996. Spatial–Temporal Modeling of SO_2 in Milan District. In: Baafi, E.Y., Schofield, N.A. (Eds.), Geostatistics Wollongong'96. Kluwer Academic Publishers, Dordrecht, Vol. 2, pp. 1031–1042.

De Cesare, L., Myers, D., Posa, D., 1999. Product–sum covariance for space–time modeling: an environmental application, submitted for publication.

Dimitrakopoulos, R., Luo, X., 1994. Spatiotemporal Modeling: Covariances and Ordinary Kriging Systems. Geostatistics for the Next Century. Kluwer Academic Publishers, Dordrecht, pp. 88–93.

Eynon, B.P., Switzer, P., 1983. The variability of rainfall acidity. Canad. J. Statist. 11, 11–24.

Journel, A.G., Huijbregts, C.J., 1981. Mining Geostatistics. Academic Press, London.

Le, D.N., Petkau, A.J., 1988. The variability of rainfall acidity revisited. Canad. J. Statist. 16, 15–38.

Matern, B., 1960. Spatial Variation-stochastic Models and their application to some problems in forest surveys and other sampling investigations. Springer, Berlin–Heidelberg, 1986, 2nd edition.

Myers, D.E., Journel, A.G., 1990. Variograms with Zonal Anisotropies and Non-Invertible Kriging Systems. Math. Geology 22, 779–785.

Myers, D.E., 1991. On Variogram Estimation. In: Öztürk, A., van der Meulen, E.C. (Chief Eds.), The Frontiers of Statistical Scientific Theory and Industrial Applications, Dudewicz, E., Nelson, P.R. (Eds.), Proceedings of the First International Conference on Statistics Comp., Cesme, Turkey, 30 March–2 April 1987, Vol II, American Sciences Press, New York, pp. 261–281.

Myers, D.E., 1992. Spatial–temporal geostatistical modeling in hydrology. Proceedings of the International Workshop. Karlsruhe, Germany.

Posa, D., 1993. A simple description of spatial–temporal processes. Comput. Statist. Data Anal. 15, 425–437.

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

Rouhani, S., Hall, T.J., 1989. Space–Time Kriging of Groundwater Data. In: Armstrong, M. (Ed.), Geostatistics. Kluwer Academic Publishers, Dordrecht, Vol. 2, pp. 639–651.

Rouhani, S., Myers, D.E., 1990. Problems in space–time kriging of hydrogeological data. Math. Geology 22, 611–623.