Mathematical Modelling and Analysis
Volume 8 Number 4, 2003, pages 283–290
 ISSN 1392-6292

UNIVERSAL KRIGING FOR SPATIO-TEMPORAL DATA

E. LESAUSKIENĖ and K. DUČINSKAS

Klaipėda University faculty of Natural and Mathematical sciences

H.Manto 84, Klaipėda, Lithuania

E-mail: `l.edita@one.lt; duce@gmf.ku.lt`

Received October 10 2003; revised November 18 2003

ABSTRACT

In this article we have used wide applicable classes of spatio-temporal nonseparable and separable covariance models. One of the objectives of this paper is to furnish a possibility how to avoid the usage of complicated covariance functions. Assuming regression model for mean function the analytical expressions for the optimal linear prediction (universal kriging) and mean squared prediction error (MSPE) was obtained. Parameterized spatio-temporal covariance functions were fitted for the real data. Prediction values and MSPE were presented. For visualization of results on graphics are used free available software **Gstat**.

Key words: Spatio-temporal random process, nonseparable and separable covariance functions, universal kriging, temporal independence, seasonal average model

1. INTRODUCTION

A large number of environmental phenomena may be regarded as realizations of space-temporal random process (Eynon and Switezex, 1983: Le and Petkan, 1988). Geostatistics offers a variety of methods to model spatial data: however, applying such space approaches to spatio-temporal random processes, may lead to the loss of valuable information in the time dimension.

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

Let

$$\{Z(s;t) : s \in D \in \Re^d; t \in [0,\infty]\}$$

denote a spatio-temporal random process. Optimal prediction (in space and time) of the unobserved parts of the process, based on the observed part of process is often the ultimate goal, but to achieve this goal, a model is needed to described how various parts of the process co-vary in space and time. In what follows, we assume that the spatio-temporal process $Z(s;t)$ satisfies the regularity condition, $var(Z(s;t)) < \infty$, for all $s \in D, t \geq 0$. Then we can define the mean function as

$$\mu(s;t) \equiv E(Z(s;t))$$

and covariance function as

$$K(s,r;t,q) \equiv cov(Z(s;t), Z(r;q)); \quad s,r \in D,\ t > 0,\ q > 0.$$

Let Z_{it} denote an observation of $Z(s,t)$ at spatial location $\{s_i : i = 1,\ldots,m\}$ and time moments $t = 1,\ldots,T$. Suppose that we have data consisting of $N = \sum_{t=1}^{T} m_t$ observations of $Z(s,t)$. Here m_t denote number of spatial locations observed at time point t and denote it

$$Z = (z_{11},\ldots,z_{m_1 1},\ldots;z_{1T},\ldots,z_{m_T T})'.$$

Furthermore, the optimal (i.e., minimum MSPE, see, e.g., Cressie, 1993) linear predictor of $Z(s_0;t_0)$ is

$$Z^*(s_0;t_0) = \mu(s_0;t_0) + C(s_0;t_0)'\Sigma^{-1}(Z-\mu),$$

where $\Sigma \equiv cov(Z)$,$C(s_0;t_0)' \equiv cov(Z(s_0;t_0), Z(s;t))$, and $\mu \equiv E(Z)$, the MSPE is $C(s_0;t_0)'\Sigma^{-1}C(s_0;t_0)$. In the rest of this article, we assume that the covariance function is stationary in space and time, namely

$$K(s,r;t,q) = C(s-r;t-q),$$

for a certain function C. This assumption is often made so that the covariance function can be estimated from data. For any $(r_1;q_1),\ldots,(r_m;q_m)$, any real $a_1,\ldots,a_m$, and any positive integer m, C must satisfy

$$\sum_{i=1}^{m}\sum_{j=1}^{m} a_i a_j C(r_i - r_j; q_i - q_j) \geq 0. \tag{1.1}$$

To ensure (1.1), one often specifies the covariance function C to belong to a parametric family whose members are known to be positive definite. That is, one assumes that

$$cov(Z(s;t), Z(s+h_s;t+h_t)) = C^0(h_s;h_t|\theta), \tag{1.2}$$

where C^0 satisfies (1.1) for all $\theta \in \Theta \in R^p$.

While there are no difficulties in extending the various kriging estimators and the kriging equations to the spatio-temporal setting, there has been a lack of known valid spatio-temporal covariances and variograms.

In order to estimate the correlation of spatio-temporal process, the main questions are as follows:

1. Is it useful define a spatio-temporal metric, such as

$$d(\mathbf{u}_1, \mathbf{u}_2) = \left(a(x_1 - x_2)^2 + b(y_1 - y_2)^2 + +c(t_1 - t_2)^2 \right)^{1/2},$$

   with $\mathbf{u}_1=(x_1, y_1, t_1)$, $\mathbf{u}_2=(x_2, y_2, t_2)$, where (x_1, y_1), $(x_2, y_2) \in D \subseteq \Re^2$, and $t_1, t_2 \in T \subseteq \Re$, where D and T are the spatial and temporal domains, respectively. In general the units for space and time will be disparate, e.g., meters and hours.

2. How to choose a spatio-temporal covariance or variogram model and how to choose parameters to ensure that the best fit to data is achieved?

In literature most spatio-temporal covariance or variogram models have been derived by utilizing the following theoretical results, since covariances or variograms in $\Re^n$ can be obtained, in general, from other valid functions. We can obtain two parametric families C^0 for (1.2): separable and nonseparable covariance functions.

In the paper [4], we have described the separable covariance functions family, where we have produced some examples for this case. There we have presented the results in case for separate product model. Let C_s be a covariance function on R^m and C_t be a covariance function on T, then the product model is

$$C_{st}(h_s; h_t) = C_s(h_s) C_t(h_t). \tag{1.3}$$

The other family consist of nonseparable functions, when we can't separate the covariance functions for space and for time. In general case, nonseparable stationary covariance functions that model spatio-temporal interactions are in great demand. Using simple stochastic partial differential equations over space and time, Jones and Zhang [3] have developed a four-parameter family of spectral densities that implicitly yield such stationary covariance functions, although not in closed form.

Cressie and Huang [2] are presenting a new methodology for developing whole classes of nonseparable spatio-temporal stationary covariance functions in closed form. But this class of functions also include to our expression autocorrelation function, spectral density and sufficiently complicated form of integral.

One of the objectives of this paper is to furnish a possibility how to avoid the usage of complicated covariance functions. In the case of temporal independent spatio-temporal covariance function is given in the form

$$C_{st}(h_s;h_t) = C^*(h_s). \tag{1.4}$$

Such a case is possible when it is necessary to make prediction for a known season. Then one can use one of four covariance functions $C_i^*(h_s|\theta)$, where $i = 1, \dots, 4$.

After averaging of spatio-temporal covariance model for all seasons we have seasonal average model:

$$C^0(h_s;h_t|\theta) = \sum_{i=1}^{4} \gamma_i C_i^*(h_s|\theta) = C(h_s)\,. \tag{1.5}$$

In this case, when a prediction for concrete season is necessary, it possible to use a spatio-temporal covariance function, that is estimated from concrete season data.

2. MAIN RESULTS

After the covariance function estimation, the interpolation between the measurement points was carried out. For this purpose, different geostatistical methods were used. Kriging is a geostatistical estimation technique, where linear minimum variance unbiased estimation is adopted, this is equivalent to the selection of a functional form, and the estimation of the relevant parameters for the main trend (first order moment) and for the covariance or the semivariogram (second order moments). Kriging is known to be a Best Linear Unbiased Predictor (BLUP), because it minimizes the variance error between the model and the predictor. The kriging equations for spatio-temporal kriging are the same as for purely spatial problems, the difference is in the usage of a spatio-temporal covariance instead of a purely spatial covariance. In the case of regression model of mean function

$$\mu(s;t) := E\big(Z(s;t)\big) = X_{s;t}^T\beta\,,$$

the optimal prediction is called universal kriging (see e.g., N.Cressie,1993).

Let us assume, that $m_t = m$, $t = 1, \dots, T$. Then covariance matrix of Z has the form $C_t \bigotimes C_s$, where C_t is $T \times T$ temporal covariance function, and C_s is $m \times m$ spatial covariance function.

Lemma 2.1. *Optimal linear prediction equation for product covariance function, defined in, is*

$$\widehat{Z}_{UK}(s_0, t_0) = x_{s_0,t_0}^T\widehat{\beta} + \delta^T[Z - X_{s,t}\widehat{\beta}] \tag{2.1}$$

where

$$\widehat{\beta} = (X_{s,t}^T(C_t^{-1} \bigotimes C_s^{-1})X_{s,t})^{-1}X_{s,t}^T(C_t^{-1} \bigotimes C_s^{-1})Z \tag{2.2}$$

$$\delta = (C_t^{-1} \bigotimes C_s^{-1})(C_{t0} \bigotimes C_{s0}) \tag{2.3}$$

where $\bigotimes$ *is the Kronecker product and* C_{s0} *and* C_{t0} *are vectors of spatial and temporal covariances between predicted point with observed points,* x_{s_0,t_0}^T *is a vector of nonrandom regressors.*

Proof. Expressions (2.2) and (2.3) were obtained by using (1.3) in universal kriging equation. Then the mean squared prediction error (MSPE) for the predictor, given in (2.1), is of the form

$$MSPE_{UK} = \delta(0) - 2b^T\big(C_{t0}(h_t) \bigotimes C_{s0}(h_s)\big) + b^T(C_t^{-1} \bigotimes C_s^{-1})b\,, \tag{2.4}$$

where

$$b^T = x_{s0,t0}^T(X_{s,t}^T(C_t^{-1} \otimes C_s^{-1})X_{s,t})^{-1}X_{s,t}^T(C_t^{-1} \otimes C_s^{-1}) +$$
$$+(C_{t0}^T \otimes C_{s0}^T)(C_t^{-1} \otimes C_s^{-1})(I - X_{s,t}(X_{s,t}^T(C_t^{-1} \otimes C_s^{-1})X_{s,t}))^{-1}\cdot$$
$$\cdot X_{s,t}^T(C_t^{-1} \otimes C_s^{-1})$$

$\delta(0) = C_s(0)C_t(0)$ and I – matrix of ones. ■

3. EXAMPLE

In this section we apply the spatio-temporal stationary covariance functions to the problem of prediction at the unobserved locations. The spatio-temporal data, used in this article, was collected in the Baltic sea, where the number of observations are taken regularly every three months during the period of 1994-1998 at six stations in the coastal zone. Salinity is the observed feature.

In order to apply the above statistical methods for data analysis we have used a free available software Gstat. Gstat is a program for the modelling, prediction and simulation of geostatistical data in one, two or three dimensions. All methods mentioned above assume that the residual covariance is known. A common convention is to enter the covariance by ways of the variogram. Gstat calculates direct sample variograms, and can fit models to them. In this approach Gstat has been used for modelling covariance functions.

Spatio-temporal covariance model, described in (1.3), was considered in [4]. In this paper we are presenting expressions of exponential spatio-temporal covariance functions for temporal independence case (1.4):

$$C^*(|h_s|) = \theta_1 \exp(-\theta_2 \cdot |h_s|)\,.$$

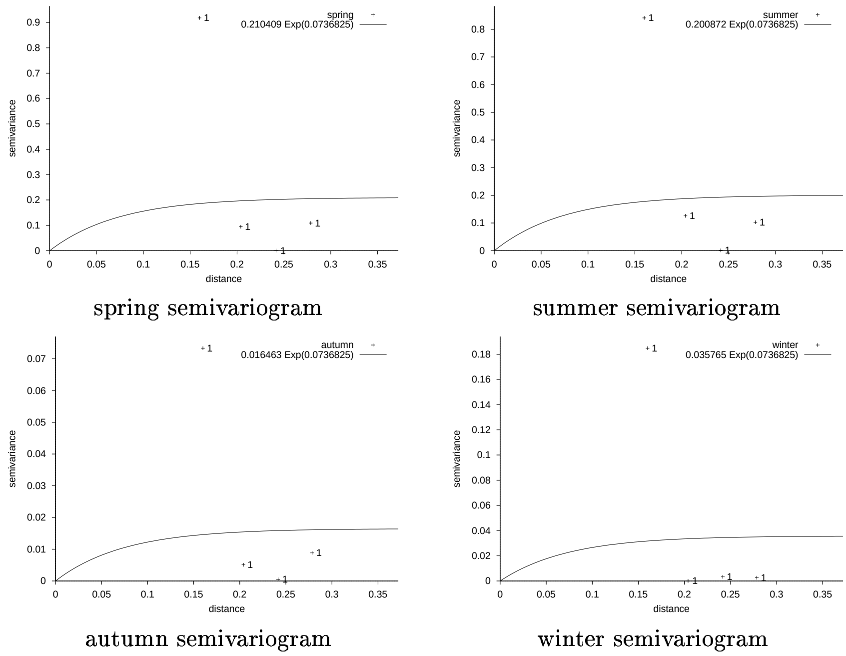


Figure 1. Semivariograms for all seasons

The same exponential model was used for all different seasons. After averaging the spatio-temporal covariance models for all seasons we obtain a seasonal average model (see 1.5).

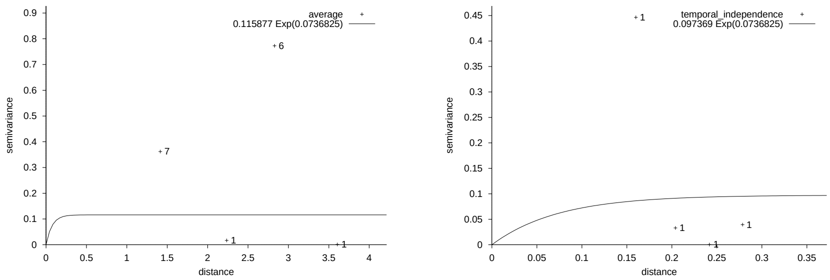
For practical realization we used semivariograms graphics instead of covariances ones. These semivariograms have the following form:

$$\gamma^*(|h_s|) = C^*(0) - C^*(|h_s|) = \theta_1(1 - \exp(-\theta_2 \cdot |h_s|))\,.$$

Figure 1 presents exponential functions of semivariograms for all four seasons. Figure 2 presents exponential semivariogram functions for the seasonal average case and a general semivariogram function in temporal independency case for whole data. As we can see from Fig.1 in winter the spatial dependency is decreasing more rapidly than in other seasons.

Using prediction equation (2.2) and the MSPE equation (2.4) we have calculated our prediction at an unobserved location and presented these results in Table 1. In all cases the exponential covariance model was used in the analysis.

From the presented results it follows that for this real data the temporal independency covariance model is optimal.



seasonal average semivariogram temporal independency semivariogram

Figure 2. Seasonal average and temporal independency semivariograms

Table 1.
Covariance models $C_{st}(h_s; h_t)$ whith parameters θ_1 and θ_2, prediction and MSPE.

Covariance for	θ_1	θ_2	Prediction	MSPE
spring	-13.572	0.210409	6.78512	0.10453
summer	-13.572	0.200872	6.74233	0.10813
autumn	-13.572	0.016463	6.48502	0.10994
winter	-13.572	0.035765	6.15312	0.10322
seasonal average	-13.572	0.115877	6.72793	0.12889
temporal independency	-13.572	0.097369	6.73264	0.10226

REFERENCES

[1] N Cressie. *Statistics for Spatial Data.* John Wiley, New York, 1993.

[2] N. Cressie and H. Huang. Models for continuous stationary space-time processes. *Lecture Notes in Statistics 122*, 569, 1997.

[3] R.H. Jones and Y. Zhang. Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, **94**, 1330, 1999.

[4] Dučinskas K. and Lesauskienė. Separable spatio-temporal covariance functions. *Lithuanian Mathematical Journal*, **42**, 497, 2002.

[5] Dučinskas K. and Šaltytė J. Lesauskienė E. Universal kriging. *Lithuanian Mathematical Journal*, **40**, 277, 2000.

[6] I. Rodriguez-Iturbe and J.M. Mejia. The design of rainfall networks in time and space. *Water Resources Reseach*, 10, 713 – 729, 1974.

Universalaus krigingo taikymas erdvės-laiko duomenims

E. Lesauskienė, K. Dučinskas

Straipsnyje lemos pavidalu pateiktos analitinės išraiškos UK (universalaus krigingo) ir MSPE (vidutinės kvadratinės prognozės klaidos), kai erdvės-laiko kovariacinė funkcija yra atskiriama, naudojant sandaugos modelį. Taip pat gautos kovariacinių modelių išraiškos, eliminavus laiko įtaką stebėjimams bei kovariaciniai modeliai atskiriems sezonams, kurie svorinio vidurkio pagalba gali būti apjungti į sezoninį vidurkio modelį.

Pateiktų formulių pagalba, realiems duomenis (Klaipėdos jūrų tyrimo centro duomenys apie druskingumo kiekį devyniose Baltijos jūros stotyse), įvertinti erdvės ir laiko kovariacinių modelių parametrai ir atlikta optimali prognozė žinomame taške (prieš tai jį eliminavus iš duomenų).

Semivariogramų modelių grafikai gauti programinio paketo **Gstat** pagalba. Lyginant gautus rezultatus, galima teigti, kad šiuos duomenis geriausiai aprašo nepriklausomų laike stebėjimų kovariacinis modelis.