Generalized product-sum variogram models of the data of the Center of Marine Research*

Ingrida KRŪMINIENĖ, Kęstutis DUČINSKAS (KU)
e-mail: duce@gmf.ku.lt, ingrida@ik.ku.lt

Abstract. Environmental data usually depends on both spatial and temporal components. Therefore, it is essential to have statistical models to describe how the data vary across space and time. Structural analysis begins with estimating the space-time covariance or variogram. In this paper a generalised product-sum variogram models has been fitted to the data of the Center of Marine Research by making use of Gstat software.

Keywords: spatial-temporal models, covariance, variogram, product-sum models.

1. Introduction

Random process models for space-time data play increasingly important roles in various scientific disciplines; among them are environmental science, agriculture, climatology, meteorology, and hydrology. In the statistical literature, the recent works of Handcock and Wallis (1994), Kyriakidis and Journel (1999), Christakos (2000), Christakos, Hristopulos, and Bogaert (2000), Brown, Diggle, Lord, and Young (2001), and others, point at the significance of the approach. In order to estimate the correlation of a space-time process, one of the main questions is how to choose a space-time covariance or variogram model and how to choose parameters to ensure that the best fit to data is achieved [7]?

In this paper the generalised product-sum variogram models has been described, in order to estimate and model in a flexible way realizations of spacetime random fields – salinity data of the Center of Marine Research) in each seasons. This generalized model is non-separable, in general is non-integrable and is the simple method for estimating and modeling the covariance or variogram components of the product-sum model using data from realizations of spatial-temporal random fields [6].

More about a various models of variograms, modelling and fiting variogram models, prediction procedures by various kriging, about others methods of geostatistical analysis can be found in the book by J.P. Chiles, P. Delfine "Geostatistics: modelling spatial uncertainly" [1], [2] (the last one in Lithuanian).

Basic concepts of space-time process are introduced in Section 2. Also the generalized product-sum model are presented and comments. The results of the study are presented in Section 3.

*Darbą finansavo Lietuvos valstybinis mokslo ir studijų fondas

2. Basic concepts of space-time process. The generalized product-sum model

Let $Z(\mathbf{s}; t)$ be a random variable at the location $\mathbf{s}$, in space, and t, in time, and let $\{(Z(\mathbf{s}; t)\colon (\mathbf{s}; t)\in\Re^d\times\Re\}$, be a second order stationary spatial-temporal random field, with covariance and variogram, respectively:

$$C_{st}(\mathbf{h}_s, h_t) = Cov\big(Z(s+\mathbf{h}_s, t+h_t), Z(\mathbf{s}, t)\big), \tag{2.1}$$

$$\gamma_{st}(\mathbf{h}_s, h_t) = \frac{Var(Z(\mathbf{s}+\mathbf{h}_s, t+h_t) - Z(\mathbf{s}, t))}{2}, \tag{2.2}$$

where $(\mathbf{s}; \mathbf{s}+\mathbf{h}_s)\in D^2$ and $(t; t+h_t)\in T^2$ [3], [4], [6], [7].

The classes of spatio-temporal covariance models are briefly described in literature [2] (p. 106–109), [3], [4], [7].

The following class of product-sum covariance models was introduced by De Cesare *et al.* (2001):

$$C_{s,t}(\mathbf{h}_s, h_t) = k_1 C_s(\mathbf{h}_s)C_t(h_t) + k_2 C_s(\mathbf{h}_s) + k_3 C_t(h_t), \tag{3.1}$$

in terms of the semivariogram function:

$$\gamma_{st}(\mathbf{h}_s, h_t) = \big[k_2 + k_1 C_t(0)\big]\gamma_s(\mathbf{h}_s) + \big[k_3 + k_1 C_s(\mathbf{0})\big]\gamma_t(h_t) - k_1\gamma_s(\mathbf{h}_s)\gamma_t(h_t), \tag{3.2}$$

where C_s and C_t are covariance functions, γ_s and γ_t are the corresponding semivariogram functions, and $k_1 > 0$, $k_2 \geqslant 0$, $k_3 \geqslant 0$ to ensure admissibility. Note that $C_{st}(\mathbf{0}, 0)$ is the sill of γ_{st} ("global" sill), $C_s(\mathbf{0})$ is the sill of γ_s and $C_t(0)$ is the sill of $\gamma_t(C_s(\mathbf{0})$ and $C_t(0)$ are named "partial" sills) [3], [6].

A generalization of the product-sum covariance model introduced by De Cesare *et al.* (2001) is given by the generalized product-sum model (De Iaco *et al.*, 2001).

$$\gamma_{st}(\mathbf{h}_s, h_t) = \gamma_{st}(\mathbf{h}_s, 0) + \gamma_{st}(\mathbf{0}, h_t) - k\gamma_{st}(\mathbf{h}_s, 0)\gamma_{st}(\mathbf{0}, h_t), \tag{3.3}$$

where $\gamma_{st}(\mathbf{h}_s, 0)$ and $\gamma_{st}(\mathbf{0}, h_t)$ are spatial and temporal bounded variogram functions and

$$k = \frac{(sill\gamma_{st}(\mathbf{h}_s, 0) + sill\gamma_{st}(\mathbf{0}, h_t) - sill\gamma_{st}(\mathbf{h}_s, h_t))}{(sill\gamma_{st}(\mathbf{h}_s, 0)sill\gamma_{st}(\mathbf{0}, h_t))}. \tag{3.4}$$

Since $\gamma(0) = 0$, it follows that

$$\hat{\gamma}_{st}(\mathbf{h}_s, 0) = \big[k_2 + k_1 C_t(0)\big]\gamma_s(\mathbf{h}_s) = k_s\gamma_s(\mathbf{h}_s), \tag{3.5}$$

and

$$\hat{\gamma}_{st}(\mathbf{0}, h_t) = \big[k_3 + k_1 C_s(\mathbf{0})\big]\gamma_t(h_t) = k_t\gamma_t(h_t), \tag{3.6}$$

where k_s and k_t can be viewed as coefficients of proportionality between the space-time variograms $\gamma_{st}(\mathbf{h}_s, 0)$ and $\gamma_{st}(\mathbf{0}, h_t)$ and the spatial and temporal variogram models $\gamma_s(\mathbf{h}_s)$ and $\gamma_t(h_t)$, respectively.

The coefficients k_1, k_2 and k_3 can be solved in terms of the sill values $C_{st}(\mathbf{0},0)$, $C_s(\mathbf{0})$, $C_t(0)$ and the parameters k_s, k_t:

$$k_1 = \frac{k_s C_s(\mathbf{0}) + k_t C_t(0) - C_{st}(\mathbf{0},0)}{C_s(\mathbf{0})C_t(0)}, \tag{3.7}$$

$$k_2 = \frac{C_{st}(\mathbf{0},0) - k_t C_t(0)}{C_s(\mathbf{0})}, \tag{3.8}$$

$$k_3 = \frac{C_{st}(\mathbf{0},0) - k_s C_s(\mathbf{0})}{C_t(0)}. \tag{3.9}$$

In this case $k_1 > 0$, $k_2 \geqslant 0$, $k_3 \geqslant 0$ if and only if k satisfies the following inequality ([6] Theorem 2):

$$0 < k \leqslant \frac{1}{\max\{sill\gamma_{st}(\mathbf{h}_s,0);\, sill\gamma_{st}(\mathbf{0},h_t)\}}. \tag{3.10}$$

Given the set of data locations in space-time

$$A = \big((\mathbf{s}_i, t_j);\, i = 1,2,\ldots,n_s;\, j = 1,2,\ldots,n_t\big);$$

estimating and modeling the spatial and temporal components proceeds as follows [6]:

1. Compute the sample spatial and temporal variograms corresponding to $\gamma_{st}(\mathbf{h}_s,0)$ and $\gamma_{st}(\mathbf{0},h_t)$

$$\hat{\gamma}_{st}(\mathbf{h}_s,0) = \frac{1}{2|N(\mathbf{h}_s,0)|}\sum_{\mathbf{s}\in N(\mathbf{h}_s,0)}\big[Z(\mathbf{s}+\mathbf{h}_s,t) - Z(\mathbf{s},t)\big]^2, \tag{3.11}$$

$$\hat{\gamma}_{st}(\mathbf{0},h_t) = \frac{1}{2|N(\mathbf{0},h_t)|}\sum_{t\in N(\mathbf{0},h_t)}\big[Z(\mathbf{s},t+h_t) - Z(\mathbf{s},t)\big]^2, \tag{3.12}$$

   where $N(\mathbf{h}_s,0)$ and $N(\mathbf{0},h_t)$ are, respectively, the vector lag with spatial tolerance and the lag with temporal tolerance [2] (p. 107).
2. Choose variogram models $\gamma_{st}(\mathbf{h}_s,0)$ and $\gamma_{st}(\mathbf{0},h_t)$ for the above two variogram estimators. Note that at this point, one must use models with sills. Hence estimates for the sill values $k_s C_s(\mathbf{0})$ and $k_t C_t(0)$ should be obtained.
3. Compute the sample variogram corresponding to $\gamma_{st}(\mathbf{h}_s,h_t)$, namely:

$$\hat{\gamma}_{st}(\mathbf{h}_s,h_t) = \frac{1}{2|N(\mathbf{h}_s,h_t)|}\sum_{\mathbf{s},t\in N(\mathbf{h}_s,h_t)}\big[Z(\mathbf{s}+\mathbf{h}_s,t+h_t) - Z(\mathbf{s},t)\big]^2. \tag{3.13}$$

4. Estimate the global sill $C_{s,t}(0,0)$ graphically, by plotting the sample variogram surface $\hat{\gamma}_{st}(\mathbf{h}_s,h_t)$.
5. Once the three sills have been estimated, the value of parameter k is determined, however one must check that (3.10) with γ replaced by $\hat{\gamma}$ is satisfied.

4. Results

The Oceanology Department was established 01 June 1992 and devoted to making observations of the circulation and transformation of water masses in the Baltic Sea and in the Curonian lagoon. The main parameters measuring or estimating are: temperature, salinity, density, water transparency and sea-ice conditions. Some meteorological parameters, such as wind speed and direction, cloud cover, air temperature etc., are also included in the standard set of observations (`http://www1.omnitel.net/juriniai_tyrimai/index.htm`).

The data of depth, temperature and salinity were collected during the period 1994.02.12 to 2001.10.22 in every season. In this research only data of salinity was used. All observed data were collected in 23 stations and their geographic coordinates was transformed to Cartesian coordinates.

Since more than one observation of salinity is collected per day in the each station, we take an average of them. In order to fit semivariograms we have chosen free available software Gstat.

Exploratorary analysis of the time series data of the Center of Marine Research showed that data have seasonal dependence (seasonality). The generalized product-sum variogram models for each season are estimated by Steps 1–5 given.

In order to select the best semivariogram model, we try model linear, spherical and exponential variogram models and calculated (weighted) sum of squared errors of the fitted model. More about parameters of these variogram models and his parameters can be found in the book by Cressie „Statistics for spatial data" [1], K. Dučinskas, J. Šaltytė–Benth "Erdvinė statistika" [2], in research [5].

We list the best chosen spatial (Table 1), temporal (Table 2) and spatial-temporal (Table 3) semivariogram models for all seasons.

The coefficients k_s and k_t was chosen by conditions described in [6] and the coefficients k, k_1, k_2 and k_3 was calculated by formulas (3.4), (3.7), (3.8) and (3.9) (Table 4).

Table 1. Spatial semivariogram models for all seasons

Season	*Variogram model*	C_0	C_1	R
Spring:	Linear	−0.0012	0.0769	1638.56
Summer:	Exponentian	0.0224	0.0475	1353.65
Autumn:	Exponentian	0.0034	0.0581	1601.32
Winter:	Linear	0.0303	0.1044	8242.96

Table 2. Temporal semivariogram models for all seasons

Season	*Variogram model*	C_0	C_1	R
Spring:	Linear	0.0399	0.0176	27470.50
Summer:	Exponentian	0.0233	0.0266	36198.80
Autumn:	Exponentian	0.0287	0.0055	7521.57
Winter:	Linear	0.0877	0.0329	12567.48

Table 3. Spatial-temporal semivariogram models for all seasons

Season	*Variogram model*	C_0	C_1	R
Spring:	Linear	0.0935	0.0549	22986.40
Summer:	Exponentian	0.0879	0.0071	28712.40
Autumn:	Exponentian	0.0851	0.0354	16560.30
Winter:	Linear	0.2077	0.0648	7795.36

Here C_0 – the nugget effect, $C_0 + C_1$ – the sill and R – the range of the variograms [1], [2], [5].

Table 4. The coefficients of generalized product-sum variogram models

	Spring	*Summer*	*Autumn*	*Winter*
k_s coeficient	1.50	1.25	1.50	2.00
k_t coeficient	1.25	1.50	2.00	0.50
k_1 coeficient	0.0008	0.0007	0.0002	0.0058
k_2 coeficient	1.0097	0.2872	0.8470	1.5758
k_3 coeficient	0.6056	0.1528	0.8266	0.0259
k coeficient	0.0005	0.0004	0.0002	0.0001

Season	*Variogram model*	C_0	C_1	R
Spring:	Linear	0.0387	0.0945	8453.08
Summer:	Exponentian	0.0458	0.0741	19811.34
Autumn:	Exponentian	0.0321	0.0636	6521.70
Winter:	Linear	0.1179	0.1373	10882.88

5. Summary

In the research described in this paper the generalized product-sum variogram models for each season are calculated. These type of variogram model are not separable and describes better the spatial-time process. Also the generalised product-sum variogram model not require the use of a complex calculations.

The check of the goodness of the fitted variogram models through cross-validation method planned in future.

References

1. N. Cressie, *Statistics for Spatial Data*, John Wiley, New York (1993).
2. K. Dučinskas, J. Šaltytė–Benth, *Erdvinė statistika*, Klaipėdos universiteto leidykla, Klaipėda (2003).
3. L. De Cesare, D.E. Myers, D. Posa, Estimating and modeling space-time correlation structures, *Statistics and Probability Letters*, **51**(1), 9–14 (2001).
4. N. Cressie, H. Huang, Classes of nonseparable spatio-temporal stationary covariance functions, *Journal of the American Statistical Association*, **94**(448), 1330–1340 (1999).
5. R. Garška, I. Krūminienė, Spatial analysis and prediction of Curonian lagoon data with GSTAT, *Mathematical Modeling and Analysis*, **9**(1), 39–50 (2004).

6. S. De Iaco, D.E. Myers, D. Posa, Space-time analysis using a general product-sum model, *Statistics and Probability Letters*, **52**(1), 21–28 (2001).
7. S. De Iaco, D.E. Myers, D. Posa, Nonseparable space-time covariance models: some parametric families, *Mathematical Geology*, **34**, 23–42 (2002).

REZIUMĖ

I. Krūminienė, K. Dučinskas. Jūrinių tyrimų centro erdvės-laiko duomenų apibendrinti multiplykatyvūs-adityvūs variogramų modeliai

Pagrindinis šio tyrimo tikslas yra Jūrinių tyrimų duomenims parinkti geriausius apibendrintus multiplikatyvius-adityvius erdvės laiko modelius. Esminis apibendrinto ir paprastojo multiplykatyvaus-adityvaus erdvės laiko modelio skirtumas tas, kad apibendrinto modelio koeficientas k priklauso ne tik nuo parametrų k_1, k_2, k_3, bet ir nuo parametrų k_s bei k_t. Siekiant įgyvendinti tikslą k_s ir k_t parametrai buvo parinkti taip, kad būtų įgyvendintos visos būtinos sąlygos, kurios užtikrina apibendrinto multiplykatyvaus-adityvaus erdvės laiko modelio teisingumą. Pagrindinis šio variogramos modelio privalumas – jis aprašo proceso kitimą ir laike, ir erdvėje bei nereikalauja sudėtingų skaičiavimų.