

Modified *STARIMA* model for space-time data *

Laura ŠALTYTĖ (KU)

e-mail: laura@ik.ku.lt

Abstract. In this paper we propose spatial time series model. *ARIMA* model class is considered for each location. Model for each location is built by spatial “connection” of identified *ARIMA* models in observed locations. Spatial “connection” is implemented by spatial averaging of the coefficients of *ARIMA* models and by ordinary kriging procedure for means. Comparison between proposed model which can be considered as modified *STARIMA* model and general *STARIMA* model developed by Pfeifer and Deutch is presented. Mean square prediction error (MSPE) for proposed procedure of prediction is presented. Values of MSPE for real data are calculated.

Keywords: spatial connection, spatial time series modeling, *ARIMA*, kriging, semivariogram.

1. Introduction

Spatial-time series called *STARIMA* model class developed at early eighties by Pfeifer and Deutch (1980). But these are still not implemented in the widely applicable computer program systems such as SPSS, STATISTICA, S-PLUS and R. We have developed spatial-time series modeling technique (see L. Šaltytė, K. Dučinskas [9]) which could be easily implemented by software with *ARIMA*, ordinary kriging and semivariogram fitting procedures (i.e., GEOSTAT, R, S-PLUS). The proposed technique based on spatial “connection” of *ARIMA* fitted to observed data.

2. *STARIMA* model

STARIMA model class, developed by Pfeifer and Deutch, is characterized by linear dependence lagged in both space and time [6]. Weight matrix $W^{(l)}$ is a square $N \times N$ l th order weight matrix with elements $w_{ij}^{(l)}$ that are nonzero only if the measurement locations i and j are l th order neighbors. Autoregressive and moving average parameters are estimated by space-time autocovariance function.

Seasonal multiplicative *STARIMA* $(p_{\lambda}, d, q_m) \times (P_{\Lambda}, D, Q_M)_S$ model:

$$\Phi_{P,\Lambda}(B^S)\phi_{p,\lambda}(B)\nabla_S^D\nabla^dZ_t=\Theta_{Q,M}(B^S)\theta_{q,m}(B)\varepsilon_t, \quad (1)$$

where $\phi_{p,\lambda}=1-\sum_{k=1}^p\sum_{l=0}^{\lambda_k}\phi_{kl}W^{(l)}B^k$ and $\theta_{q,m}(B)=1-\sum_{k=1}^q\sum_{l=0}^{m_k}\theta_{kl}W^{(l)}B^k$ – autoregressive and moving average parameters. ϕ_{kl} and θ_{kl} – autoregressive and

* Darbą finansavo Lietuvos valstybinis mokslo ir studijų fondas (programos registr. nr. T-05258).

moving average parameters at temporal lag k and space lag l ; p is the autoregressive order; q is the moving average order; λ_k – spatial order of the k th autoregressive term; m_k – spatial order of the k th moving average term; d – number of nonseasonal differences required, such that $\nabla^d = (1 - B)^d$; $B\varepsilon_{it} = \varepsilon_{i,t-1}$; ε_t – “white noise”; $\Phi_{P,\Lambda} = 1 - \sum_{k=1}^P \sum_{l=0}^{\Lambda_k} \Phi_{kl} W^{(l)} B^{kS}$ and $\Theta_{Q,M}(B^S) = 1 - \sum_{k=1}^Q \sum_{l=0}^{M_k} \Theta_{kl} W^{(l)} B^{kS}$ are seasonal autoregressive and moving average parameters $\nabla_S^D = (1 - B^S)^D$ – required seasonal difference; S – seasonal lag. $W^{(l)} = (w_{ij}^l)$ – spatial weights matrix with $\sum_{j=1}^N w_{ij}^l = 1$.

When $d = 0$ and $D = 0$ the model collapses to the easier to interpret *STARIMA* model:

$$\begin{aligned} Z_t = & \sum_{k=1}^p \sum_{l=0}^{\lambda_k} \phi_{kl} W^{(l)} Z_{t-k} - \sum_{k=1}^q \sum_{l=0}^{m_k} \theta_{kl} W^{(l)} \varepsilon_{t-k} \\ & + \sum_{k=1}^P \sum_{l=0}^{\Lambda_k} \Phi_{kl} W^{(l)} Z_{t-S-k} - \sum_{k=1}^Q \sum_{l=0}^{M_k} \Theta_{kl} W^{(l)} \varepsilon_{t-S-k} + \varepsilon_t. \end{aligned} \quad (2)$$

3. Spatial connection method

Let $Z_t(s)$ represent an observation of random variable Z at location s and time t , and $Z_{it} = Z_t(s_i)$, $i = 1, \dots, N$; $t = 1, \dots, T$ describe whole analyzed data set.

We assume that mathematical model of Z_{it} is

$$\Phi_{iP}(B^S)\phi_i(B)\nabla_S^D\nabla^d Z_{it} = \alpha_i + \Theta_{iQ}(B^S)\phi_i(B)\varepsilon_{it} \quad (3)$$

and denote it by *ARIMA* $(p, d, q) \times (P, D, Q)_s$, i.e., multiplicative seasonal autoregressive moving average model with nonzero mean. In the above equations and notational expression, the ordinary autoregressive and moving average components are represented by polynomials $\phi(B)$ and $\theta(B)$ of orders p and q respectively the seasonal autoregressive and moving average components by $\Phi_P(B^S)$ and $\theta_Q(B^S)$ of orders P and Q and ordinary and seasonal difference components by $\nabla^d = (1 - B)^d$ and $\nabla_S^D = (1 - B^S)^D$, respectively.

Spatial-time model fitting consists of two parts: *ARIMA* model fitting in each location and kriging estimation of mean.

There are several basic steps of fitting *ARIMA* models to time series data. These steps involve plotting the data, possibly transforming the data, identifying the dependence orders of the model, parameter estimation, and diagnostics [8].

The final step of model fitting is model choice. The most popular techniques are AIC, AICc, and SIC also cross validation [8].

Thus in each location we should fit *ARIMA* model with the same number of parameters and nonzero constant.

For kriging estimator and for spatial “connection” we need to fit semivariogram.

Spatial connection between i th and j th locations is realized by spatial weighted average method with the spatial weights:

$$\delta_{ij} = \frac{1/\gamma(s_{ij})}{\sum_{i=1}^N 1/\gamma(s_{ij})}, \quad (4)$$

there $\gamma(s_{ij})$ is the semivariogram between i th and j th locations.

Inverse distance is used, while semivariogram is measure of dissimilarity and inverse dimension let us consider that stronger dependence is between nearest locations. Also $\gamma(h) = 0$, when $h = 0$ it means that main diagonal of weight matrix is zero.

Then parameters for new station can be calculated by

$$\hat{\phi}_{il} = \sum_j \delta_{ij} \phi_{jl}, \quad l = 1, \dots, p, \quad i, j = 1, \dots, N \quad (\text{autoregression parameters}), \quad (5)$$

$$\hat{\theta}_{ik} = \sum_j \delta_{ij} \theta_{jk}, \quad k = 1, \dots, q, \quad i, j = 1, \dots, N \quad (\text{moving average parameters}), \quad (6)$$

$$\hat{\Phi}_{iL} = \sum_j \delta_{ij} \hat{\Phi}_{jL}, \quad L = 1, \dots, P, \quad i, j = 1, \dots, N, \quad (7)$$

$$\hat{\Theta}_{iK} = \sum_j \delta_{ij} \hat{\Theta}_{jK}, \quad K = 1, \dots, Q, \quad i, j = 1, \dots, N, \quad (\text{seasonal parameters}). \quad (8)$$

As we have already fitted semivariogram and nonzero constant for each location, we can find the kriging estimator μ_k , and the nonzero constant for model at new location is $\alpha_{0k} = \mu_k(1 - \phi_{01} - \dots - \phi_{0p})$. Then fitted model for i th location:

$$\begin{aligned} Z_{it} = & \alpha_{ik} + \sum_{l=1}^p \phi_{il} Z_{i,t-l} + \varepsilon_{it} + \sum_{k=1}^q \theta_{0k} \varepsilon_{i,t-k} \\ & + \sum_{l=1}^p \Phi_{il} Z_{i,t-S-l} + \varepsilon_{it} + \sum_{k=1}^q \Theta_{0k} \varepsilon_{i,t-S-k}. \end{aligned} \quad (9)$$

In vector form:

$$Z_t = \alpha_k + \sum_{l=1}^p A_l Z_{t-l} + \varepsilon_t + \sum_{k=1}^q B_k \varepsilon_{t-k} + \sum_{L=1}^P C_L Z_{t-S-L} + \sum_{K=1}^Q D_K \varepsilon_{t-S-K}, \quad (10)$$

where Z_t – vector of observations; ε_t – vector of residuals; ϕ_l , Φ_l and θ_k , Θ_k respectively non seasonal and seasonal parameter matrixes of autoregression and moving average estimated using spatial weights;

$$A_l = \text{diag}(\phi_{1l}, \phi_{2l}, \dots, \phi_{Nl}); \quad B_l = \text{diag}(\theta_{1k}, \theta_{2k}, \dots, \theta_{Nk});$$

$$C_l = \text{diag}(\Phi_{1L}, \Phi_{2L}, \dots, \Phi_{NL}); \quad D_l = \text{diag}(\Theta_{1K}, \Theta_{2K}, \dots, \Theta_{NK}).$$

Proposed model for each location is built by spatial connection of identified *ARIMA* models in observed locations. Spatial “connection” is implemented by spatial averaging of the coefficients of *ARIMA* models and by ordinary kriging procedure for means. Spatial weights are estimated using semivariogram. In opposite to *STARIMA* model, in model for i th location we are using only lagged in time observations from the same location, whereas *STARIMA* model takes observations lagged both in space and time. Using *STARIMA* model we should fit $2N$ cross covariances. Using spatial connection model there was $2N$ autocorrelation and partial autocorrelation functions. So spatial connection model can be considered as modified *STARIMA* model.

4. Example

Lithuanian Sea Research Center data was used for illustration of proposed modeling technique. Data set consist of 32 time observation ($t = 1, \dots, 32$) of the salinity in Baltic coastal zone in 9 ($N = 9$) station.

Using plotted data, ACF and PACF we selected several most appropriate models for all stations.

After diagnostics step model *ARIMA* $(1, 0, 1) \times (1, 0, 0)_4$ left. I. e. In each location we have model:

$$Z_{it} = \alpha_{ik} + \phi_{i1}Z_{it-1} + \varepsilon_{it} + \theta_{0k}\varepsilon_{i,t-k} + \Phi_{i1}Z_{it-4}. \quad (11)$$

Predictor for this model, taking conditional expectation is:

$$\begin{aligned} E_t[Z_{it+j}] &= E_t[Z_{it+j}/Z_{it}, Z_{it-1}, \dots, \varepsilon_{it}, \varepsilon_{it-1} \dots] \\ &= \alpha_{ik} + \phi_{i1}E_t[Z_{it+j-1}] + \Phi E_t[Z_{it+j-4}], \quad j > 4 \end{aligned} \quad (12)$$

Then MSPE can be written:

$$\begin{aligned} MSPE &= E(Z_{it+j} - E_t[Z_{it+j}])^2 \\ &= E(Z_{it+j} - \alpha_{ik} - \phi_{i1}[Z_{it+j-1}] - \Phi E_t[Z_{it+j-4}])^2, \end{aligned} \quad (13)$$

where parameters are fitted using (5)–(8).

To obtain spatial weight, we need semivariogram. Semivariogram is fitted using the all data. Minimum of Sum square error (SSE) is used for selection criterion. The Spherical semivariogram is optimal:

$$\gamma(|h|) = \begin{cases} 0, & \text{when } |h| = 0, \\ 0.208 + 0.06 \left(\frac{3}{2} \frac{|h|}{16456.81} - \frac{1}{2} \left(\frac{|h|}{16456.81} \right)^3 \right), & \text{when } 0 < |h| < 16456.81, \\ 0.208 + 0.06, & \text{when } |h| \geq 16456.81. \end{cases}$$

Using fitted semivariogram we can estimate spatial parameters for each location and fit spatial time series model.

For fitted spatial time series models in each location MSPE was calculated. For calculating cross validation method was used. Results of MSPE in each location are: 0.064312, 0.225515, 0.423063, 0.912638, 0.434277, 0.677374, 0.076821, 0.386017, 0.494622.

As we can see MSPE are quite small. So we can conclude that proposed modified model is applicable in some practical situations.

References

1. T.W. Anderson, *An Introduction to Multivariate Statistical Analysis*, Wiley & Sons (2003).
2. J.P. Chiles, P. Delfiner, *Geostatistics: Modeling Spatial Uncertainty* (1999).
3. A.D. Cliff, J.K. Ord, Space-time modeling with an application to regional forecasting, *Transactions of the Institute of British Geographers*, **64**, 119–128 (1975).
4. N.A.C. Cressie, *Statistics for Spatial Data* (1993).
5. Y. Kamarianakis, *Spatial-Time Series Modeling: a Review of the Proposed Methodologies* (2003).
6. P.E. Pfeifer, S.J. Deutsch, A three-stage iterative procedure for space-time modeling, *Technometrics*, **22**(1), 35–47 (1980).
7. P.E. Pfeifer, S.J. Deutsch, Seasonal space-time ARIMA modeling, *Geographical Analysis*, **13**(2), 35–47 (1981).
8. R.H. Shumway, D.S. Stoffer, *Time Series Analysis and Its Applications*, Springer-Verlag, New York. Inc. (2000).
9. L. Šaltytė, K. Dučinskas, Spatial-time series modeling with *R* system, *Liet. matem. rink.*, **44** (spec. nr.), 770–774 (2004).

REZIUOMĖ

L. Šaltytė. Modifikuotas STARIMA modelis erdvės-laiko duomenims

Straipsnyje aprašyta nauja erdvinių laiko eilučių modeliavimo technika, kuriuos esminis principas – *ARIMA* modelių erdvinis sujungimas. Siūlomas modelis gali būti nagrinėjamas kaip modifikuotas *STARIMA* modelis. Darbe pateiktos modifikuoto *STARIMA* modelio vectorinė išraiška ir vidutinė kvadratinė prognozės paklaida (MSPE). Siūlomas modelis sudarytas Jūrinių tyrimų centro duomenims apie druskingumą Baltijos jūroje ir jo efektyvumas įvertintas, apskaičiuojant MSPE kryžminio patikrinimo metodu.