Liet. matem. rink., **44**, *spec. nr.*, 2004, 770–773

Spatial time-series modeling with R system*

Laura ŠALTYTĖ, Kęstutis DUČINSKAS (KU)
e-mail: laura@ik.ku.lt, duce@gmf.ku.lt

Abstract. In this paper we propose modeling technique, which was applied to multivariate time series data that correspond to different spatial locations (spatial time series). ARIMA model class is considered for each location. Forecasting model for new location is built by spatial "connection" of identified models in observed locations. Spatial "connection" is implemented by spatial averaging of the coefficients of models and by ordinary kriging procedure for means. This modeling technique is illustrated by a substantive example using R system.

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

1. Introduction

Research in statistical models that describe the spatio – temporal evolution of a single variable in space and time started in the midseventieth (see Cliff, Ord 1975) and has significantly increased during the last twenty years since it's closely related to the progress in computer technology and the existence of large data bases.

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 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. Modeling procedure

Let $Z_t(s)$ represent an observation of random variable Z at location s and t time.

The whole analyzed data set is represented by the expression $\{Z_{it},\ i=1,\dots,N;\ t=1,\dots,T\}$, where $Z_{it}=Z_t(s_i)$. We assume that mathematical model of Z_{it} is

$$\Phi_{iP}(B^s)\phi_i(B)\nabla_s^D\nabla^d Z_{it}=\alpha_i+\Theta_{iQ}(B^s)\theta_i(B)\varepsilon_{it} \tag{1}$$

and denote it by *ARIMA* $(p,d,q)\times(R,D,Q)_s$, 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

*Darbą finansavo Lietuvos valstybinis mokslo ir studijų fondas.

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 (in R system *arima* package, *stats* procedure) and kriging estimation of mean (*krige* package, *gstat* procedure).

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.

First we should plot the data Z_{it} versus t, and inspect the graph for any anomalies. Next step is to identify initial values of the autoregressive order p, the order of differencing d, and the moving average order q. A time plot of the data will typically suggest whether any differencing is needed. When preliminary values of d have been settled, the next step is to look at the sample autocorrelation function (ACF) and partial autocorrelation function $(PACF)$ of $\nabla^d Z_{it}$ for whatever values of d have been chosen. Also using (ACF) and $(PACF)$ we can choose preliminary values of p and q [10]. At this stage, a several initial values of p, d and q should be at hand, and we can start estimating the parameters.

The next step in model fitting is diagnostics. This investigation includes the analysis of the residuals as well as the model comparisons. Investigation of marginal normality can be accomplished visually by looking at a histogram of the residuals. In addition to this, a $Q - Q$ *plot* can help in identifying departures from normality. There are several test of randomness, for example the runs test, which could be applied to the residuals. Also we can use a general or portmaneu test, i.e., Ljung–Box–Pierce test[10].

The final step of model fitting is model choice. The most popular techniques are AIC, AICc, and SIC also cross validation [10].

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 (*Gstat* package, *variogram, fit.variogram* functions). Spatial connection for new location is proposed to realize by spatial weighted average method with the spatial weights:

$$\delta_{0i} = \frac{\gamma(s_{i0})}{\sum_{l=1}^{N} \gamma_{l0}}. \tag{2}$$

There $\gamma(s_{i0})$ is the semivariogram between i-th and a new location s_0.

Then parameters for new station can be calculated by:

$$\hat{\phi}_{0l} = \sum_{i=1}^{N} \delta_{0i}\hat{\phi}_{il}, \quad l = 1, \ldots, p, \tag{3}$$

(autoregression parameters),

$$\hat{\theta}_{0k} = \sum_{i=1}^{N} \delta_{0i}\hat{\theta}_{ik}, \quad k = 1, \ldots, q, \tag{4}$$

(moving average parameters),

$$\hat{\Phi}_{0L} = \sum_{i=1}^{N} \delta_{0i} \hat{\Phi}_{iL}, \quad L = 1, \ldots, P, \tag{5}$$

$$\hat{\Theta}_{0K} = \sum_{i=1}^{N} \delta_{0i} \hat{\Theta}_{iK}, \quad K = 1, \ldots, Q, \tag{6}$$

(seasonal parameters),

$$\hat{\sigma}_{0\varepsilon}^2 = \sum_{i=1}^{N} \delta_{0i} \hat{\sigma}_{i\varepsilon}^2 \tag{7}$$

(dispersion of residuals).

As we have already fitted semivariogram and nonzero constant for each location, we can find the kriging estimator $\hat{\mu}_k$, and the nonzero constant for model at new location is $\alpha_{0k} = \hat{\mu}_k(1 - \phi_{01} - \ldots - \phi_{0p})$. Then fitted model for new a location will be:

$$\begin{aligned} Z_{0t} &= \alpha_{0k} + \hat{\phi}_{01} Z_{0,t-1} + \ldots + \hat{\phi}_{0p} Z_{0,t-p} + \varepsilon_{0t} + \hat{\theta}_{01} \varepsilon_{0,t-1} + \ldots \\ &\quad + \hat{\theta}_{0q} \varepsilon_{0,t-q} + \hat{\Phi}_{01} Z_{0,t-s} + \ldots + \hat{\Phi}_{0P} Z_{0,t-s-P} + \hat{\Theta}_{01} \varepsilon_{0,t-s} + \ldots \\ &\quad + \hat{\Theta}_{0Q} \varepsilon_{0,t-s-Q}. \end{aligned} \tag{8}$$

If there is not information about observations at new location until time moment T, then time prediction at new location is done by model (8) for simulated data. Simulated data were obtained by simulating several time series (T values for each). Averages of the simulated values for each time moment until T forms a "new" time series which is used for prediction.

As we will have several different models, estimation of prediction at new location can be performed by cross-validation method.

3. Example

Lithuanian Sea research center data was used for illustration of proposed modeling technique. Data set consist of 32 time observation ($t = 1, \ldots, 32$) 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: $ARIMA(2,0,2)$, $ARIMA(2,0,1)$, $ARIMA(2,0,1) \times (1,0,1)_4$, $ARIMA(2,0,0) \times (1,0,1)_4$, $ARIMA(0,0,2) \times (1,0,0)_4$, $ARIMA(1,0,1)_4$.

After diagnostics step the two models $ARIMA(2,0,1) \times (1,0,1)_4$ and $ARIMA(0,0,2) \times (1,0,0)_4$ left.

Thus we can conclude that salinity in Baltic see is seasonal.

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:

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

After spatial connection procedure (3–7) $ARIMA(2,0,1) \times (1,0,1)_4$ model for new station can be written as:

$$Z_0 t = 10.0428 + 0.0546 Z_{0,t-1} - 0.1885 Z_{0,t-2} + \varepsilon_{0,t} - 0.3157 \varepsilon_{0,t-1} + 0.0366 Z_{0,t-4} + 0.1689 \varepsilon_{0,t-4} \quad (9)$$

and $ARIMA(0,0,2) \times (1,0,0)$ model as:

$$Z_0 t = 6.3649 + \varepsilon_{0,t} - 0.2298 \varepsilon_{0,t-1} - 0.2129 \varepsilon_{0,t-2} + 0.0596 Z_{0,t-4}. \quad (10)$$

In case (9) prediction at new station for time moment $T+1$ was $Z_{0,T+1} = 10.0013$, while in case (10) it was $Z_{0,T+1} = 6.0589$.

After cross validation procedure we obtained that in case (9) $MSPE = 4.53$, while in case (10) $MSPE = 3.27$.

Thus we can conclude, that (10) is better model for considered data for prediction of salinity for station s_0.

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, in: *Transactions of the Institute of British Geographers* (1975).
4. N.A.C. Cressie, *Statistics for Spatial Data* (1993).
5. J. Gress, S. Schach, W. Urfer, *Statistical Analysis of Spatial and Temporal Dynamics of a bacterial Layer in an Aquatic Ecosystem*, Department of Statistics, Dortmund (1999).
6. Y. Kamarianakis, *Spatial-Time Series Modeling: a Review of the Proposed Methodologies*, REAL (2003).
7. P.E. Pfeifer, S.J. Deutsch, *A Three-Stage Iterative Procedure for Space-Time Modeling*, Technometrics (1980).
8. P.E. Pfeifer, S.J. Deutsch, *Seasonal space-time ARIMA modeling*, Geofraphical Analysis (1981).
9. B.D. Ripley, *Spatial Statistics*, Geofraphical Analysis (1981).
10. R.H. Shumway, D.S. Stoffer, *Time Series Analysis and Its Applications*, Springer-Verlag, New York (2000).

REZIUMĖ

L. Šaltytė, K. Dučinskas. Erdvės-laiko duomenų modeliavimas sistemos R aplinkoje

Straipsnyje aprašyta nauja erdvinių laiko eilučių modeliavimo technika, kuriuos esminis principas – erdvinis modelių sujungimas. Siūloma technika lengvai realizuojama laisvai platinamos sistemos R pagalba. Modelis realizuotas Naudojant Lietuvos Jūrinių tyrimų centro duomenis.