VILNIUS UNIVERSITY

LAURA PAULIONIENĖ

STATISTICAL MODELLING OF SPATIO-TEMPORAL DATA BASED ON SPATIAL INTERPOLATION OF TIME SERIES PARAMETERS

Summary of doctoral dissertation

Physical sciences, Mathematics (01 P)

Vilnius, 2013

Doctoral dissertation was prepared at Faculty of Mathematics and Informatics of Vilnius University in 2012 - 2013.
Dissertation is defended by an external student.

Academic Consultant
Prof. Dr. Kęstutis Dučinskas (Klaipėda University, Physical Sciences, Mathematics – 01 P).

The dissertation will be defended at the Council of Scientific Field of Mathematics at the Vilnius University:
Chairman
Prof. Dr. Habil. Remigijus Leipus (Vilnius University, Physical Sciences, Mathematics – 01 P).
Members:
Prof. Dr. Habil. Algimantas Bikelis (Vytautas Magnum University, Physical Sciences, Mathematics – 01P),
Prof. Dr. Habil. Vydas Čekanavičius (Vilnius University, Physical Sciences, Mathematics – 01P),
Prof. Dr. Habil. Alfredas Račkauskas (Vilnius University, Physical Sciences, Mathematics – 01P),
Prof. Dr. Tonu Kollo (Tartu University, Physical Sciences, Mathematics – 01P).
The dissertation will be defended at the public meeting of the Council of Mathematics Science in the auditorium number 102 at the Faculty of Mathematics and Informatics of Vilnius University, at 2 p.m. on 8 November 2013.
Address: Naugarduko st. 24, LT-03225 Vilnius, Lithuania.
The summary of the doctoral dissertation was distributed on 8th of October 2013. A copy of the doctoral dissertation is available for review at the Library of the Vilnius University.

VILNIAUS UNIVERSITETAS

LAURA PAULIONIENĖ

ERDVĖS - LAIKO DUOMENŲ STATISTINIS MODELIAVIMAS, PAGRĮSTAS LAIKO EILUČIŲ PARAMETRŲ ERDVINIU INTERPOLIAVIMU

Daktaro disertacijos santrauka

Fiziniai mokslai, Matematika (01 P)

Vilnius, 2013

Disertacija rengta 2012–2013 metais Vilniaus universiteto Matematikos ir informatikos fakultete.
Disertacija ginama eksternu.

Mokslinis konsultantas
prof. dr. Kęstutis Dučinskas (Klaipėdos universitetas, fiziniai mokslai, matematika – 01P).

Disertacija ginama Vilniaus universiteto Matematikos mokslo krypties taryboje:

Pirmininkas:
prof. Dr. Habil. Remigijus Leipus (Vilniaus Universitetas, fiziniai mokslai, matematika – 01P).
Nariai:
prof. habil. dr. Algimantas Bikelis (Vytauto didžiojo universitetas, fiziniai mokslai, matematika – 01P),
prof. habil. dr. Vydas Čekanavičius (Vilniaus universitetas, fiziniai mokslai, matematika – 01P),
prof. habil. dr. Alfredas Račkauskas (Vilniaus universitetas, fiziniai mokslai, matematika – 01P)
prof. dr. Tonu Kollo (Tartu universitetas, fiziniai mokslai, matematika – 01P),

Disertacija bus ginama viešame Vilniaus universiteto Matematikos mokslo krypties tarybos posėdyje 2013 m. lapkričio 8 d. 14 val. Vilniaus universiteto Matematikos ir informatikos fakulteto 102 auditorijoje.
Adresas: Naugarduko g. 24, LT-03225, Vilnius, Lietuva. Disertacijos santrauka išsiuntinėta 2013 m. spalio 8 d. Disertaciją galima peržiūrėti Vilniaus universiteto bibliotekoje.

General Characteristic of the Dissertation

Topicality of the problem – Space-time data modelling problem is not entirely new, however, in light of evolving technologies, there are new opportunities to model, so this topic is still quite relevant. Often spatial data sets are relatively small, and the points, where observations are taken, are located irregularly. When collecting data for a period of time (usually equal intervals of time) we can have many observations at each point. In such case usually we have not more than hundred spatial locations (points) but in each of them quite long time series are obtained. When solving spatial task, usually we are interpolating or estimating the spatial average. Time series data usually are used to predict future values. Meanwhile, the space - time tasks combines both types of tasks. One of the obvious solutions is to analyse spatial data at each time moment separately, also we can analyse time series at each point, i.e. multiple time series. However in this case it‘s impossible to predict and estimate values at the points outside the sample. In general, it is necessary to take into account the correlations in space and time, and to identify relations between them.

Aim of the work. To create time series model at new unobserved location by incorporating spatial interaction thru spatial interpolation of estimated time series parameters.

Tasks of the work:

To create algorithms of space-time modelling,

To verify the proposed models for space-time data realization,

To perform a detailed comparison of the proposed models with real data by using R program.

Scientific novelty

1. Several original space-time data modelling techniques based on spatial interpolation of time series parameters were proposed and their properties were investigated.

2. Forecasting properties of proposed space-time models were investigated.

3. Diagnostic checking and comparison of Space-time models was performed by cross-validation method by modelling wind speed data in Republic of Lithuania.

Methodology of the research. The scientific work uses the following research methods: analysis of scientific literature on the topic of the dissertation, classical statistical methods, and spatial statistics. The proposed methods are realized by using open-source system R.

Practical value

1. Quite simple modelling technique for space-time data is proposed, which can be easily implemented by software.
2. Model for spatial time series is designed which allows calculating prediction at new spatial location s_0 and for new time moment $T + k$.

Defended propositions

1. Often space-time random field $Z(\boldsymbol{s}; t)$ can be modelled by separating spatial and temporal components.
2. Proposed spatial weights and ordinary kriging methods based on semivariances quite well describe analysed space-time data. These methods are more accurate then trend surface and weighted inverse distance methods.
3. Proposed space-time modelling technique can be widely applied in different research fields.

The scope of the scientific work. The scientific work consists of the general characteristic of the dissertation, 3 chapters, conclusions, list of literature and list of publications. The total scope of the dissertation – 84 pages, 66 equations, 20 pictures and 13 tables.

1. Space-time data

Spatio – temporal stochastic process can be described:

$\{Z(\boldsymbol{s}; t): \boldsymbol{s} \in D_s, t \in D_t\}, (\boldsymbol{s}; t) \in D_S \times D_t$.

Set $D_S \times D_t$ is called space-time set.

Space and time can be continuous and discreet, depending on the analysed data.

Spatio-temporal process $Z(\boldsymbol{s};t)$ can be written:

$$Z(\boldsymbol{s};t) = \mu(\boldsymbol{s};t) + \delta(\boldsymbol{s};t), \qquad (1)$$

here $\boldsymbol{s} \in D_s, t \in D_t$; $\mu(\boldsymbol{s};t)$ is deterministic mean function; $\delta(\boldsymbol{s};t)$ – zero-mean random effect describing the variation of the trend in time and space, which does not describe a deterministic function of the mean $\mu(\boldsymbol{s};t)$. It is usually considered that space and time in (1) model are continuous. Second-order moment of space - time process (1) can be written:

$$cov\big(Z(\boldsymbol{s};t), Z(\boldsymbol{x};r)\big) \equiv C_Z(\boldsymbol{s},\boldsymbol{x};t,r). \qquad (2)$$

There is another way related to the covariance function and helping to define dependence. *Semivariogram* of stocastic spatio-temporal process is described:

$$var\big(Z(\boldsymbol{s};t) - Z(\boldsymbol{x};r)\big) \equiv \gamma(\boldsymbol{s},\boldsymbol{x};t,r), \qquad (3)$$

here γ is called semivariogram.

Space - time covariance function can be used not only for statistical dependencies of spatial and temporal characterization. In spatial statistics there is well-known optimal linear spatial prediction method – *kriging*.

There is a number of space - time data models, but the most popular are the vector autoregressive moving average model (VARMA) and the space - time autoregressive moving average model (STARMA), and their modifications.

The n – dimensional process $\boldsymbol{Z}_t$ is called vector autoregressive moving average process if $\{\boldsymbol{Z}_t\}$ is stationary and for any t:

$$\boldsymbol{M}_0\boldsymbol{Z}_t + \boldsymbol{M}_1\boldsymbol{Z}_{t-1} + \cdots + \boldsymbol{M}_p\boldsymbol{Z}_{t-p} = \boldsymbol{A}_0\boldsymbol{\epsilon}_t + \boldsymbol{A}_1\boldsymbol{\epsilon}_{t-1} + \cdots + \boldsymbol{A}_q\boldsymbol{\epsilon}_{t-q}, \qquad (4)$$

here $\boldsymbol{Z}_t \equiv \big(Z_t(\boldsymbol{s}_1), \dots, Z_t(\boldsymbol{s}_n)\big)'$, $t = 1,2,\dots,T$; $\boldsymbol{A}_0, \boldsymbol{A}_1, \dots, \boldsymbol{A}_p, \boldsymbol{M}_0, \boldsymbol{M}_1, \dots, \boldsymbol{M}_q$ are $n \times n$ order weight matrixes, and $\{\boldsymbol{\epsilon}_t\}$ n – dimensional zero mean withe noise sequence with covariance matrix $\boldsymbol{\Sigma}$, i.e. $\boldsymbol{\epsilon}_t \sim WN(\boldsymbol{0},\boldsymbol{\Sigma})$. Such process usually is denoted as $VARMA(p,q)$.

The n – dimensional process $\boldsymbol{Z}_t$ is called pace - time autoregressive moving average process ($STARMA$) if $\{\boldsymbol{Z}_t\}$ is stationary and for any t:

$$\boldsymbol{Z}_t = \sum_{k=1}^{p}\left(\sum_{l=0}^{\lambda_k}\phi_{kl}\boldsymbol{W}_l\right)\boldsymbol{Z}_{t-k} + \sum_{k=1}^{q}\left(\sum_{l=0}^{\mu_k}\theta_{kl}\boldsymbol{W}_l\right)\boldsymbol{\epsilon}_{t-k} + \boldsymbol{\epsilon}_t,\ t = 0,1,\ldots, \quad (5)$$

here $\{\boldsymbol{W}_l\}$ is known weights matrix; p and q autoregressive and moving average order respectively; $\{\phi_{kl}\}$ and $\{\theta_{kl}\}$ – model parameters; λ_k – spatial order for $k-th$ autoregressive element; μ_k –spatial order for $k-th$ moving average element; $\{\boldsymbol{\epsilon}_t\}$ – zero mean white noise process with covariance matrix $\boldsymbol{\Sigma}_\varepsilon = \sigma_\varepsilon^2 I$.

The essential weakness in $VARMA$ and $STARMA$ models is a large number of estimated parameters. Methods described in 2 section of this dissertation work do not take into account space-time cross-correlations what significantly reduces number of estimated parameters.

2. Modelling of spatial time series

First time series model at each spatial location independently of spatial information is considered.

Denote by $Z_i(t)$ a time series in the spatial location $s_i \in D, i = 1,\ldots,n$, $t = 1,2,\ldots,T$. In this dissertation work two time series models are described: first model decomposes time series into trend component, seasonal component and $ARMA$ process, second model is classical $ARIMA$ process.

Model for $Z_i(t)$ can be written:

$$Z_i(t) = L_i(t) + S_i(t) + \sum_{k=0}^{p}\phi_k^i\big(Y_i(t-k)\big) + \sum_{l=0}^{q}\theta_l^i\big(\epsilon_i(t-l)\big) + \epsilon_i(t), \quad (6)$$

here $L_i(t) = \sum_m a_m^i t^m, m = 0,1,\ldots$ is $m-th$ order polynomial, which describes trend of time series; $X_i(t) = Z_i(t) - L_i(t);\ Y_i(t) = X_i(t) - S_i(t)$; ϕ_k^i - parameters of autoregressive $AR(p_k)$ process, and θ_l^i – parameters of moving average $MA(q_l)$ process. $S_i(t)$ – seasonal component:

$$S_i(t) = b_0^i + \sum_{r=0}^{R}\left[b_{2r+1}^i cos\left(\frac{(2r+2)\pi t}{365}\right) + b_{2r+2}^i sin\left(\frac{(2r+2)\pi t}{365}\right)\right]. \quad (7)$$

It‘s assumed, that process $\epsilon_i(t)$ is of the following form

$$\epsilon_i(t) = \sigma_i(t)\varepsilon_i(t), \quad (8)$$

where $\sigma_i(t)$ is (possibly) time-dependent standard deviation function, and $\varepsilon_i(t)$ – is as zero-mean temporally independent Gaussian random process with standard deviation equal to one.

One way to model the variance is to fit a truncated Fourier series

$$\sigma_i^2(t) = c_0^i + \sum_{j=1}^{J} \left[c_{2j}^i cos\left(\frac{2j\pi t}{365}\right) + c_{2j+1}^i sin\left(\frac{(2j+1)\pi t}{365}\right)\right], \quad (9)$$

to the empirical daily variance. The alternative is to use the empirical daily variance. Errors are normalized by dividing them by one of the variance functions.

As alternative to (6) model, integrated autoregressive moving average $ARIMA$ model to each time series $Z_i(t)$ can be fitted

$$\phi^{*i}(B)\nabla^d Z_i(t) = \theta^{*i}(B)\epsilon_i(t), \quad (10)$$

here $\phi^i(B) = \left(1 - \phi_1^{*i}B - \cdots - \phi_p^{*i}B^p\right)$ and $\theta^i(B) = \left(1 - \theta_1^{*i}B - \cdots - \theta_q^{*i}B^q\right)$ autoregressive and moving average operators respectively; $\nabla^d = (1-B)^d$ - $d - th$ order difference operator.

As time series behaviour in the different spatial locations is quite similar at all locations s_i the same order $ARIMA$ model is fitted. However coefficients ϕ_k^i and θ_l^i depend on spatial location. Thus in each location model is individual but with the same number of parameters.

The estimated models are validated for out-of-sample time series data.

After fitting time series model (6) and (10) at each spatial location we have sets of parameters $\hat{\eta}^i = \{\hat{a}_m^i, \hat{b}_r^i, \hat{\phi}_k^i, \hat{\theta}_l^i : m = 0,1,2 \ldots; r = 0,1,2, \ldots; k = 0,1,2, \ldots; l = 0,1,2, \ldots\}$ or $\hat{\eta}^{*i} = \{\hat{\phi}_k^{*i}, \hat{\theta}_l^{*i} : k = 0,1,2, \ldots; l = 0,1,2, \ldots\}$ respectively. For spatial interpolation of fitted time series parameters four methods are proposed.

In order to incorporate spatial dependence, empirical semivariogram to normalized errors of time series is fitted. It is shown, that semivariogram estimator based on normalised errors is *unbiased, consistent* and *asymptotically normal*, i.e. $\sqrt{T}\left(\hat{\gamma}_{ij} - \gamma_{ij}\right) \sim N\left(0, 2\gamma_{ij}^2\right)$.

Matheron estimator for normalized residuals at locations i ir j can be written:

$$\hat{\gamma}_{ij} = \sum_{t=1}^{T} (\varepsilon_i(t) - \varepsilon_j(t))^2 / (2T).$$

It is shown, that $E(\hat{\gamma}_{ij}) = \gamma_{ij}$; $D(\hat{\gamma}_{ij}) = \frac{2\gamma_{ij}^2}{T}$.

In this scientific work parameters of empirical semivariogram are fitted by using Matheron estimator for normalized residuals, i.e. $\hat{\varepsilon}_i(t)$:

$$\hat{\hat{\gamma}}_{ij} = \sum_{t=1}^{T} (\hat{\varepsilon}_i(t) - \hat{\varepsilon}_j(t))^2 / (2T)$$

As distributions of normalized time series residuals $\hat{\varepsilon}_i(t)$ in case of finite sample are complicated, in this scientific work their properties are not analysed.

First proposed method for spatial interpolation, based on semivariogram, is *spatial weights* method.

Spatial weights between fixed spatial location s_0 and any other spatial location s_j are described in such way:

$$\hat{\delta}_{0j} = \frac{1/\gamma(h_{0j}; \hat{\boldsymbol{\vartheta}})}{\sum_{i=1}^{n} 1/\gamma(h_{0i}; \hat{\boldsymbol{\vartheta}})}, \tag{11}$$

here $\gamma(h_{ij}; \hat{\boldsymbol{\vartheta}})$ semivariogram between $i-th$ and $j-th$ locations; $h_{ij} = s_i - s_j$.

Set of parameters for new location s_0 is then calculated:

$$\hat{\eta}_{0(ES)} = \hat{\boldsymbol{\eta}} \cdot \hat{\delta}_0{}', \tag{12}$$

here $\hat{\boldsymbol{\eta}} = (\hat{\eta}^1, \ldots, \hat{\eta}^n)$, were $\hat{\eta}^i$ vector of fitted parameters in $i-th$ location and $\hat{\delta}_0 = (\hat{\delta}_{10}, \ldots, \hat{\delta}_{n0})$ vector of spatial weights.

Kriging is well known spatial prediction (interpolation) method which provides the best linear unbiased prediction (BLUP). In this scientific work *formal ordinary kriging* is used.

Estimators of *formal ordinary kriging* can be calculated:

$$\hat{\eta}_{0(OK)} = \hat{\boldsymbol{\eta}} \cdot \hat{\lambda}_0', \tag{13}$$

here $\hat{\lambda}_0$ are also weights: $\hat{\lambda}_0' = \frac{1_n'\mathbf{\Gamma}^{-1}}{\Gamma_{\cdot\cdot}} + \gamma_0'\mathbf{\Gamma}^{-1}\left(I - \frac{\boldsymbol{J}\cdot 1_n'\mathbf{\Gamma}^{-1}}{\Gamma_{\cdot\cdot}}\right)$, $\Gamma_{\cdot\cdot} = 1_n'\mathbf{\Gamma}^{-1}1_n$, $\boldsymbol{J} = 1_n 1_n'$; $\gamma_0 = \gamma\left(h_{0j}; \hat{\boldsymbol{\vartheta}}\right)$, and $\mathbf{\Gamma}$ – matrix of semivariograms.

Alternatively to spatial interpolation methods described above simpler methods can be used. These are methods which interpolates just by geometrical layout of locations. One of them is *trend surface* model fitted by least squares method. This model depends on coordinates of point (location). For location $s_i = (x_i, y_i) \in D, i = 1,2, \dots, n$, model is described:

$$P^{\hat{\beta}_i} = \sum_{\substack{f+u \le d \\ 0 \le f,u \le d}} \lambda_{fu}^{\hat{\beta}_i} x_i^f y_i^u, \quad (14)$$

here $\hat{\beta}_i$ – component of parameters vector $\hat{\eta}^i$ or $\hat{\eta}^{*i}$ at location s_i; $\lambda_{fu}^{\hat{\beta}_i}$ - parameter of suggested model, d – order of trend surface model.

Modelling of spatial weights is analogue of *inverse distance weighting.* Inverse distance weighted methods are based on the assumption that the interpolating surface should be influenced most by the nearby points and less by the more distant points. The simplest form of inverse distance weighted interpolation is sometimes called "Shepard's method". The equation used is as follows

$$\widehat{w}_{0j} = \frac{h_{0j}^{-p}}{\sum_{i=1}^{n} h_{0i}^{-p}}, \quad (15)$$

here $p > 0, p \in \mathbb{R}$; h_{0i} distance between each and interpolated point, i.e. $h_{0i} = \sqrt{(x_0 - x_i)^2 + (y_0 - y_i)^2}$.

In order to select the best space-time model cros-validation procedure is used.

3. Spatial time series model for wind speed data

Proposed space-time modelling techniques were realized to wind speed (WS) data. The records of (WS) data provided by the Lithuanian Hydrometeorological Service in Vilnius, Lithuania, were used in this dissertation work. Data were collected in 20 meteorological stations (see Fig. 1). Each measurement represents the average WS measured in meters per second. Data were collected since 1977 01 01 until 2007 12 31.

Figure 1. Map of Lithuania with measurement stations.

Modelling univariate time series. The model was estimated on in-sample data from the period 1977 01 01 to 2006 12 31, consisting of 10,950 observations recorded in 18 stations. As there was no trend, first the seasonal function (7) for WS data at each station separately was estimated. The function with only three parameters was sufficient to remove seasonal fluctuations in data. As ACF (PACF) of residuals showed seasonal variations were removed in the data, however, the residuals still exhibited strong autocorrelations. As indicated by the PACF plot, we needed a higher-order AR process to explain the autocorrelations. The AIC (the lower the better) was used together with ACF (Box-Ljung test) and histograms of the residuals when deciding on the order of the AR process. We started with a low order AR process and proceeded with estimating the processes with an increasing order. The order $p_k$was bounded to 10. For the most of the stations, $AR(3)$ or $AR(4)$ was sufficient to capture the variations in the residuals. However, in four stations AR order higher than four was needed. An $ARMA$ process with the combination of AR and MA components did not provide better fit. The residuals after fitting AR became not autocorrelated (Fig. 2) according to the Box-Ljung statistic. Even though a negative skewness and positive kurtosis in residuals in all stations was observed, the distribution seemed to be rather symmetric (Fig. 3). The mean of the residuals was essentially zero for all stations with standard deviations varying from 0.36 to 0.48. Normal distributions (and Gaussian random fields) are easier to analyse and estimate.

In this case, the deviations from normality are not big and we believe that the normality assumption is a reasonable approximation.

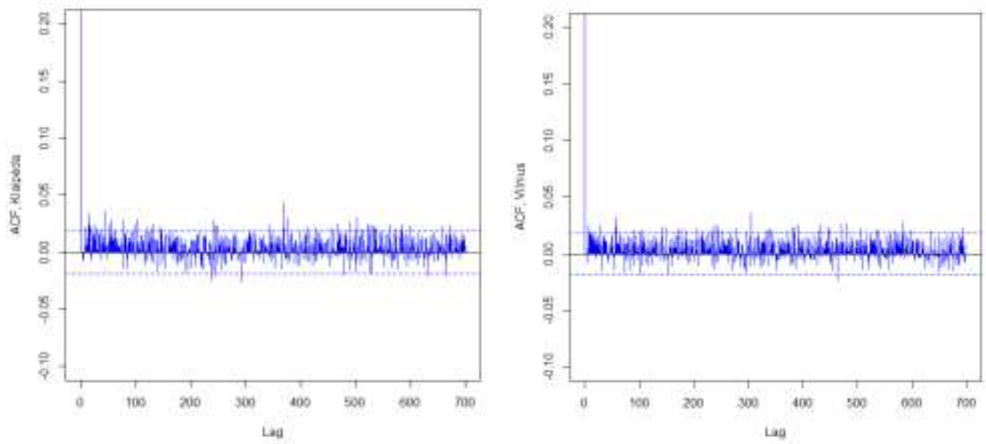

Figure2. ACF with 95% confidence bounds for residuals after seasonal and AR components were eliminated from WS in Klaipeda (left) and Vilnius (right).

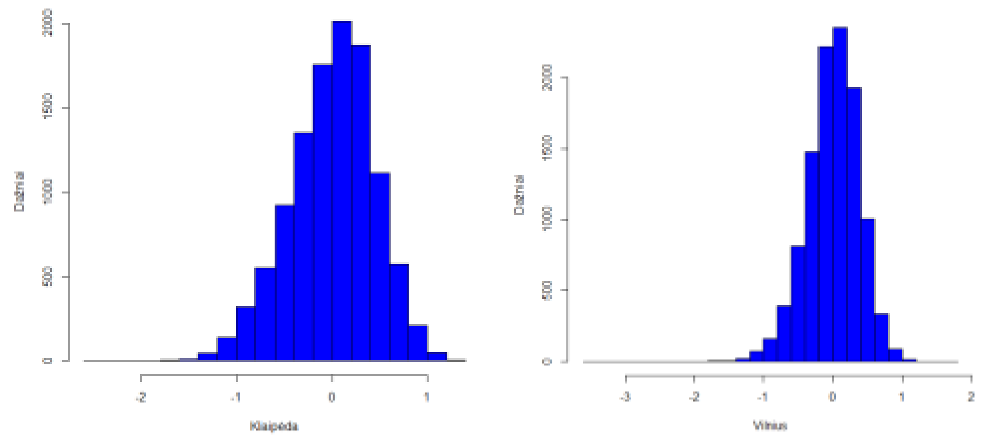

Figure 3: Histogram of residuals after seasonal and AR components were eliminated from WS in Klaipeda (left) and Vilnius (right).

The autocorrelations seemed to vary randomly around zero, however the ACF of the squared residuals revealed time dependency in the variance (see Fig. 4). Therefore next the empirical daily variance of the residuals obtained after the seasonal effects and AR process were eliminated from WS data was examined. The empirical daily variance was calculated by averaging the squared residuals for each day, resulting in 365 observations. For wind the ACF function of the squared residuals exhibits a clear seasonal pattern, although the residuals themselves are uncorrelated. Following the analysis of the empirical variance, we started with modelling it as a function defined in (9) with $J = 1$. Also the daily empirical variance for modelling the seasonality in the residuals was used. The residuals were first normalized by dividing them by the square root of one of the two variance functions. For model validation 365 out-of-sample WS observations from the year 2007 available in all 18 stations were used. To validate the temporal model, one-step-ahead predictions for out-of-sample observations were generated and the prediction errors (PE) defined as the differences between the observations and predictions were calculated. Simple statistical tests were performed on PEs. The PEs were normally distributed for all but two. Even though the normality assumption was not verified at the 5% significance level in Siauliai and Utena, the histograms demonstrated a reasonable symmetry in these stations.

Following the same steps of time series modelling in each station $ARIMA$ models were fitted. We started with $ARIMA(1,1,0)$ process and continued until $ARIMA(3,1,3)$ at most of the stations $ARIMA(3,1,1)$ was sufficient. The residuals also were normalised by dividing them by square root of empirical daily variance function and variance estimated via (9) formula.

Spatial interpolation

For estimated time series residuals exponential semivariogram was fitted:

$$\gamma(h;\hat{\boldsymbol{\vartheta}}) = \begin{cases} 0, if\ h = 0 \\ \hat{\vartheta}_0 + \hat{\vartheta}_1\left(1 - exp\left(-\frac{1}{\hat{\vartheta}_2}\sqrt{h_1^2 + h_2^2}\right)\right), if\ h > 0 \end{cases}.$$

In case of time series model (6) estimated semivariogram parameters are: $\hat{\vartheta}_0 = 0{,}031$; $\hat{\vartheta}_1 = 0{,}005$; $\hat{\vartheta}_2 = 0{,}192$ su $MSE = 0{,}013$.

In case of model (10) estimated semivariogram parameters are: $\hat{\vartheta}_0 = 0{,}032$; $\hat{\vartheta}_1 = 0{,}004$; $\hat{\vartheta}_2 = 0{,}214$ su $MSE = 0{,}015$.

After semivariogram was fitted spatial interpolation for Trakų Vokė and Palanga stations was performed. Results of spatial interpolation are shown in Table 1 and Table 2 for time series models (6) and (10) respectively. In first column estimated parameters with standard errors(SE) of univariate time series are shown. Most of the estimated values are within two SEs from the fitted values, demonstrating a rather good fit of the model. As expected closest to the actual values are values estimated by using spatial weiths and ordinary kriging, i.e. estimates based on spatial correlations.

	Palanga					Trakų Vokė				
	Estimates (SE)	ES	OK	TPM	IDW	Estimates (SE)	ES	OK	TPM	IDW
b_0	1,414 (0,006)	1,266	1,316	1,257	1,136	1,014 (0,005)	1,118	0,997	0,958	1,112
b_1	0,141 (0,009)	0,165	0,151	0,148	0,161	0,175 (0,007)	0,168	0,183	0,167	0,167
b_2	0,008 (0,009)	0,005	0,011	−0,029	0,024	0,310 (0,007)	0,360	0,395	0,483	0,224
ϕ_1	0,478 (0,014)	0,498	0,464	0,494	0,475	0,433 (0,010)	0,471	0,459	0,466	0,468
ϕ_2	-0,046 (0,015)	-0,033	-0,045	−0,031	-0,029	-0,020 (0,010)	-0,036	-0,015	−0,028	-0,029
ϕ_3	0,067 (0,015)	0,062	0,073	0,057	0,059	0,037 (0,010)	0,058	0,061	0,057	0,055
ϕ_4	0,055 (0,014)	0,026	0,035	0,023	0,023					
σ^2	0,173 (0,005)	0,198	0,193	0,186	0,181	0,219 (0,005)	0,232	0,301	0,198	0,193

Table 1. Estimates of time series parameters with standard errors (SE) for model (6) also estimates calulated by using spatial weights (ES), ordinary kriging (OK), trend surface model (TPM) and inverse distance weights (IDW)

	Palanga					Trakų Vokė				
	Estimates (SE)	ES	OK	TPM	IDW	Estimates (SE)	ES	OK	TPM	IDW
ϕ_1^*	0,429 (0,014)	0,450	0,437	0,418	0,461	0,415 (0,010)	0,446	0,438	0,419	0.487
ϕ_2^*	-0,062 (0,015)	-0,047	-0,051	-0,059	-0,048	-0,027 (0,011)	-0,049	-0,041	-0,022	-0,052
ϕ_3^*	0,065 (0,014)	0,042	0,071	0,019	0,034	0,018 (0,010)	0,035	0,027	0,034	0,055
θ_1^*	-0,977 (0,004)	-0,972	-0,981	-0,965	-0,971	-0,975 (0,003)	-0,970	-0,961	-0,976	-0,974
σ^{*2}	0,187 (0,013)	0,191	0,201	0,151	0,154	0,244 (0,009)	0,198	0,203	0,215	0,190

Table 2. Estimates with standard errors (SE) of model (10) also estimates calulated by using spatial weights (ES), ordinary kriging (OK), trend surface model (TPM) and inverse distance weights (IDW)

Cross-validation results in Table 3 shows that the model $\hat{Z}_{ES}(s_0;t)$ fits best to the given data. It is a model with seasonal component and $AR(p)$ process and for spatial interpolation spatial weigths method was performed.

Model	RMSE
$\hat{Z}_{TPM}(s_0;t)$	0,57 m/s
$\hat{Z}_{ES}(s_0;t)$	0,14 m/s
$\hat{Z}_{IDW}(s_0;t)$	0,21 m/s
$\hat{Z}_{OK}(s_0;t)$	0,16 m/s
$\hat{Z}^*_{TPM}(s_0;t)$	0,48 m/s
$\hat{Z}^*_{ES}(s_0;t)$	0,17 m/s
$\hat{Z}^*_{IDW}(s_0;t)$	0,28 m/s
$\hat{Z}^*_{OK}(s_0;t)$	0,23 m/s

Table 3. Results of cross-validation. RMSE – root mean square error.

General conclusions

1. Quite simple methods for spatial interpolation of time series parameters where proposed. Given models describe WS dynamics in space and time quite accurately.

2. There are two methods for univariate time series proposed. First method decomposes time series is into seasonal component and $AR(p)$ process. Application of time series decomposition makes it quite easy to model wind speed over time. After removing all deterministic components, we were left with the residuals which essentially were white noise.

3. Second proposed model – classical integrated autoregressive moving average model $ARIMA(p,d,q)$. Residuals of this model after estimating seasonality of variance and removing it were white noise.

4. It is then simple to estimate spatial structure based on time uncorrelated residuals. Empirical semivariogram to time series residuals was fitted.

5. Four methods for spatial interpolation were proposed, two of them take into account spatial correlations: spatial weights method and ordinary kriging method. The other two methods only take into account the layout between points.

6. The WS is a very noisy meteorological parameter, which makes it difficult to model. The provided models thus contains quite many parameters, however they are clear and easily applicable for space-time prediction, which is an important input in many applications. After establishing the form of different model components, a maximum likelihood estimation approach could be implemented in a straightforward manner by appealing to the normality assumption. However, it can be a priori difficult to parameterize the model of such a noisy phenomena as WS in an optimal way.

7. In proposed models we are able to separate time and space components by applying well-established $ARIMA$ process in time and Gaussian random field in space. The $ARIMA$ parameters are quite stable over the space, all noise is left to the residuals. This can be attributed to homogeneous weather patterns in Lithuania. Models over the bigger and/or more complicated geographical area would most likely contain more components.

8. After cross-validation procedure it was found that wind speed in Republic of Lithuania is most accurately described by seasonal $AR(p)$ model were spatial interpolation was performed by using spatial weitghs method.

List of Published Works on the Topic of the Dissertation
In the reviewed scientific periodical publications

J. Šaltytė – Benth, L. Šaltytė. (2011). Spatial-temporal model for wind speed in Lithuania. *Journal of Applied Statistics*. Vol 38. No 6, p. 1151-1168, ISSN 0266-4763 (print), ISSN 1360-0532 (online)

Šaltytė L., Dučinskas K. (2004). Spatial time series modeling with system R. *Lietuvos Matematikos Rinkinys. LMD darbai*, Vol. 44, 770-773, ISSN 0132-2818

Šaltytė L. (2005). Modified STARIMA model for space – time data. *Lietuvos Matematikos Rinkinys. LMD darbai*, Vol. 45, 375-379, ISSN 0132-2818

Paulionienė L. (2012). Spatial ARMA model for wind speed data. *Lietuvos Matematikos Rinkinys. LMD darbai*, Vol. 53, 102–107, ISSN 0132-2818

In the reviewed scientific publications of conference proceedings

Šaltytė L**.** (2009). Adequacy Testing of Two Methods for Modeling Spatial Time Series**.** *In Proceedings of the 13th International Conference „Applied Stochastic Models and Data Analysis" (ASMDA-2009): selected papers,* 33-36, Vilnius: Technika. ISBN 978-9955-28-463-5.

About the author

Laura Paulionienė was born in Kretingale on 10 of September 1980.

Graduated from Klaipėda University Faculty of Nature and Mathematics in 2002 acquiring Bachelor's Degree in Mathematics. Gained Master's Degree in Mathematics at Klaipėda University Faculty of Nature and Mathematics in 2004. Since 2004 works as assistant in Department of Statistics at Klaipėda University. Since 2007 works as lecturer in Department of Statistics at Klaipėda University. Since 2012 to 2013 was an external student at Vilnius University Faculty of Mathematics and Informatics.

Erdvės - laiko duomenų statistinis modeliavimas, pagrįstas laiko eilučių parametrų erdviniu interpoliavimu

Nagrinėjama problema ir darbo aktualumas. Darbe nagrinėjama erdvės – laiko duomenų modeliavimo problema nėra visiškai nauja, tačiau dėl besivystančių technologijų, atsiranda vis naujų modeliavimo galimybių, todėl ši tema vis dar yra gana aktuali. Dažnai erdvinių duomenų rinkiniai yra gana nedideli, o taškai, kuriuose pasklidę stebėjimai, išsidėstę netaisyklingai. Renkant duomenis tam tikrą laiko periodą (paprastai vienodais laiko intervalais), jų gali būti net ir labai daug. Monitoringo sistemą tokiuose stebėjimuose paprastai sudaro ne daugiau kaip šimtas taškų, tačiau kiekviename taške gaunamos gana ilgos laiko eilutės. Sprendžiant „erdvinį" uždavinį, paprastai siekiama inerpoliuoti arba įvertinti erdvinį vidurkį. Laiko eilučių duomenys dažniausiai naudojami ateities reikšmėms prognozuoti. Tuo tarpu erdvės – laiko uždaviniai jungia abu uždavinių tipus. Vienas iš akivaizdžių sprendimo būdų yra analizuoti erdvėje rinktus duomenis kiekvienu atskiru laiko momentu, t.y. ignoruoti reiškinio kitimą laike. Taip pat galima analizuoti laiko eilutes skirtinguose taškuose, t.y. daugiamates laiko eilutes. Tačiau tokiu atveju neįmanoma modeliuoti, prognozuoti bei įvertinti reikšmių taškuose, nesančiuose imtyje. Bendru atveju reikia atsižvelgti į koreliacijas ir erdvėje, ir laike bei nustatyti ryšius tarp jų.

Darbo tikslas. Sudaryti laiko eilutės modelį naujame, nestebėtame taške, erdvinę sąveiką (angl. spatial interacion) inkorporuojant per įvertintų laiko eilučių parametrų erdvinį interpoliavimą.

Darbo uždaviniai. Sudaryti erdvės – laiko modelių realizavimo algoritmus.

Patikrinti siūlomus modelius erdvės – laiko duomenų realizacijai.

Atlikti išsamų siūlomų modelių palyginimą su realiais duomenimis R programos pagalba.

Tyrimų metodika. Darbe naudojami šie tyrimų metodai: mokslinės literatūros analizė disertacijos tema, klasikiniai statistiniai metodai, erdvinės statistikos metodai. Siūlomi metodai realizuojami atvirojo kodo sistema R.

Darbo naujumas ir jo reikšmė

1. Pasiūlyta keletas originalių erdvės – laiko duomenų modeliavimo metodų, pagrįstų laiko eilučių parametrų interpoliavimu ir ištirtos jų savybės.

2. Išanalizuotos siūlomų erdvės – laiko modelių prognozės galimybės.

3. Atliktas erdvės – laiko modelių patikrinimas bei palyginimas, remiantis kryžminio patikrinimo metodu, modeliuojant LR vėjų greičių duomenis.

Darbo rezultatų praktinė reikšmė

1. Pateikta gana paparsta erdvės – laiko duomenų modeliavimo metodika, kurią nesudėtinga realizuoti kompiuterinių programų pagalba.

2. Sudarytas erdvinių laiko eilučių modelis, leidžiantis apskaičiuoti prognozę naujame erdvės taške s_0 nauju laiko momentu $T + k$.

Ginamieji teiginiai

1. Dažnai erdvės – laiko atsitiktinį lauką $Z(\boldsymbol{s}; t)$ galima sumodeliuoti atskiriant erdvės ir laiko komponentes.

2. Pasiūlyti erdvinių svorių bei ordinaraus krigingo metodai, pagrįsti semivariogramomis leidžia pakankamai tiksliai aprašyti analizuojamus erdvės – laiko duomenis. Šie metodai tikslesni už trendo paviršiaus bei svertinių atvirkštinių atstumų metodus.

3. Pasiūlyta erdvės – laiko modeliavimo technika gali būti plačiai taikoma įvairiuose tyrimuose.

Darbo rezultatų aprobavimas

Disertacijos tema atspausdinti 5 straipsniai. Disertacijos tema skaityti 2 pranešimai respublikinėse konferencijose bei 6 pranešimai tarptautinėse:

Tarptautinės konferencijos:

Modified STARIMA model. 9th International Vilnius Conference on Probability Theory and Mathematical Statistics, June 25-30, 2006, Vilnius, Lithuania.

Spatio-temporal modeling in case of anisotropic variogram structure. Nordic Conference on Mathematical Statistics 2008 (NORDSTAT), 16-19 June, 2008, Vilnius, Lithuania.

Adequacy Testing of Two Methods for Modeling Spatial Time Series. The XIIIth International Conference "Applied Stochastic Models and Data Analysis" (ASMDA-2009), Vilnius, Lithuania.

Spatial-temporal modeling of Baltic Sea coastal zone parameters. 2nd Nordic-Baltic Biometric conference, 10-12 June, 2009, Tartu, Estonia.

Spatio-temporal model based on separable characteristics. 4th Nordic-Baltic Biometric Conference, June 10-12, 2013, Stockholm, Sweden.

Comparison of two Spatio-temporal modelling techniques for Wind Speed data. Spatial Statistics Conference, June 4-6, 2013, Columbus, Ohio, USA.

Respublikinės konferencijos:

Modifikuotas STARIMA modelis erdvės – laiko duomenims. 46 LMD konferencija, birželio 15–16 d., 2005, VU, Vilnius.

Erdvinis ARIMA modelis vėjų greičių duomenims LR. 53 LMD konferencija, 11-12 birželio, 2012, Klaipėda, KU.

Straipsniai disertacijos tema recenzuojamuose mokslo žurnaluose:

J. Šaltytė – Benth, L. Šaltytė. (2011). Spatial-temporal model for wind speed in Lithuania. *Journal of Applied Statistics*. Vol 38. No 6, p. 1151-1168, ISSN 0266-4763 (print), ISSN 1360-0532 (online)

Šaltytė L., Dučinskas K. (2004). Spatial time series modeling with system R. *Lietuvos Matematikos Rinkinys*. *LMD darbai*, Vol. 44, 770-773, ISSN 0132-2818

Šaltytė L. (2005). Modified STARIMA model for space – time data. *Lietuvos Matematikos Rinkinys. LMD darbai*, Vol. 45, 375-379, ISSN 0132-2818

Paulionienė L. (2012). Spatial ARMA model for wind speed data. *Lietuvos Matematikos Rinkinys*. *LMD darbai*, Vol. 53, 102–107, ISSN 0132-2818

Straipsniai disertacijos tema konferencijų pranešimų recenzuojamuose mokslo leidiniuose:

Šaltytė L**.** (2009). Adequacy Testing of Two Methods for Modeling Spatial Time Series**.** *In Proceedings of the 13th International Conference „Applied Stochastic Models and Data Analysis" (ASMDA-2009): selected papers*, 33-36, Vilnius: Technika. ISBN 978-9955-28-463-5.

Disertacijos struktūra.

Disertaciją sudaro įvadas, trys skyriai, išvados, literatūros sąrašas bei autorės publikacijų disertacijos tema sąrašas.

Bendra disertacijos apimtis 84 puslapiai, numeruotų formulių 66, 13 lentelių, 20 paveikslų. Literatūros sąrašą sudaro 79 šaltiniai.

Pirmame disertacijos skyriuje aprašytos bendros erdvės – laiko duomenų charakteristikos, jų savybės. Išanalizuoti erdvės – laiko duomenų tipai, aptarti populiariausi žinomi modeliai VARMA, STARMA bei atskiri jų atvejai.

Antrame disertacijos skyriuje aprašyta keletas originalių erdvinių laiko eilučių modeliavimo metodų. Siūlomi metodai pirmiausia analizuoja vienmates laiko eilutes, o pašalinus laikinę priklausomybę jose, laiko eilučių liekanoms, vertinama erdvinė priklausomybė. Tikslas – sudaryti modelį, leidžiantį prognozuoti požymio reikšmę naujame, nestebėtame taške, nauju laiko momentu. Tokio modelio sudarymas remiasi laiko eilučių parametrų erdviniu interpoliavimu. Darbe analizuojami du vienmačių laiko eilučių modeliai: klasikinis $ARIMA$ modelis bei modelis, susidedantis iš trendo, sezoninės komponenčių bei $ARMA$ proceso. Siūkomi modeliai remiasi prielaida, kad kiekviename erdvės taške stebimą informaciją galima aprašyti tokiu pačiu (tos pačios eilės) laiko eilutės modeliu. Erdviniam parametrų interpoliavimui siūlomi dviejų tipų metodai. Pirmo tipo metodai atsižvelgia į erdvinę koreliaciją, darbe nagrinėjami du šio tipo metodai: sudaryta erdvinių svorių, kurie apskaičiuojami semivariogramų pagalba, formulė (ES) bei pritaikytas formalus ordinarus krigingas (OK). Antram tipui priklauso metodai, ignoruojantys erdvinę koreliaciją, t.y. trendo paviršiaus metodas, kurio parametrai vertinami mažiausių kvadratų metodu (TPM); atvirkštinių atstumų svorių metodas (IDW).

Trečiame skyriuje visi šie metodai realizuoti interpoliuojant įvertintus laiko eilučių parametrus vėjų greičių duomenims Lietuvos respublikoje (LR) bei atliktas visų keturių interpoliavimo metodų palyginimas, taikant kryžminio matikrinimo metodą.

Bendrosios išvados

1. Darbo tikslas buvo sudaryti laiko eilutės modelį naujame, nestebėtame taške, erdvinę sąveiką (angl. spatial interacion) inkorporuojant per įvertintų laiko eilučių parametrų erdvinį interpoliavimą.

2. Pasiūlyti gana paparasti metodai, erdvėje interpoliuojantys vienmačių laiko eilučių parametrus, kurie gana tiksliai aprašo vėjo greičio dinamiką erdvėje ir laike. Pirmiausia, nagrinėjamos vienmatės laiko eilutės kiekvienoje stotyje. Laiko eilučių modeliavimui pasiūlyti du modeliai

3. Turint laike nepriklausomus stebėjimus, gana paprasta įvertinti erdvinę struktūrą, remiantis empirine semivariograma. Emipirinė semivariograma įvertinta nepriklausomiems ir vienodais pasiskirsčiusiems stebėjimams (replikacijoms), šio įverčio savybėms suformuluota lema.

4. Erdviniam modeliavimui pasiūlyti keturi metodai, du iš jų inkorporuoja erdvinę koreliaciją, t.y. erdvinių svorių bei formalaus ordinaraus krigingo metodai , o kiti du atsižvelgia tik į taškų tarpusavio išsidėstymą, t.y. trendo paviršiaus bei svertinio atvirkštinio atstumo metodai.

5. Vėjo greitis yra gana triukšmingas meteorologinis parametras, kas labai apsunkina jo modeliavimą. Dėl šios priežasties siūlomi modeliai turi gana daug parametrų, tačiau iš kitos pusės modeliai yra gana aiškūs ir juos lengva taikyti erdvės – laiko prognozei, kas yra labai svarbu įvairiuose taikymuose. Nustačius modelio komponentes galima būtų taikyti maksimalaus tikėtinumo metodą normalumo prielaidai patikrinti. Tačiau labai sunku parametrizuoti tokį triukšmingą parametrą kaip vėjo greitis.

6. Modeliuose galima atskirti erdvę ir laiką, taikant gerai žinomą $ARIMA$ procesą laike ir Gauso atsitiktinį lauką erdvėje. $ARIMA$ parametrai yra gana stabilūs erdvėje, visas triukšmas paliktas liekanose. Tokie modeliai gali būti taikomi tuo atveju, kai analizuojamas dydis visoje teritorijoje yra homogeniškas. Didesnėje ar labiau komplikuotoje geografinėje vietoje, modeliai būtų daug sudėtingesni ir su dar daugiau komponenčių.

7. Siūlomiems modeliams pritaikius kryžminio patikrinimo metodą buvo nustatyta, kad analizuojamą vėjo greitį Lietuvos respublikoje geriausiai

aprašo sezoninis $AR(p)$ modelis, kurio parametrų interpoliavimui buvo pritaikytas erdvinių svorių metodas.

Trumpos žinios apie autorių

Laura Paulionienė gimė 1980 m. rugsėjo 10 d. Kretingalėje, Klaipėdos r.

2002 m. įgijo matematikos bakalauro laispsnį Klaipėdos universiteto Gamtos ir matematikos mokslų fakultete. 2004 m. įgijo matematikos magistro laipsnį Klaipėdos universiteto Gamtos ir matematikos mokslų fakultete. Nuo 2004 m. dirba Klaipėdos universiteto statistikos katedros asistente, o nuo 2007 m. Klaipėdos universiteto statistikos katedros lektore. 2012 – 2013 m. doktorantė eksternė Vilniaus universiteto matematikos ir informatikos fakultete.