Spatial ARMA model for wind speed data

Laura Paulionienė

Department of Statistics, Klaipėda University

H. Manto 84, LT-92294 Klaipėda

E-mail: laura.paulioniene@gmail.com

Abstract. Spatial time series model for wind speed data is proposed. Based on few similar papers, first at each location univariate time series model, containing seasonal component and higher order autoregressive component is fitted. After eliminating time dependence in time series, empirical semivariogram based on time independent residuals is fitted and spatial weights for new location are calculated.

Keywords: spatio-temporal, AR, semivariogram, wind speed, spatial connection.

Introduction

In last year's there are a lot of different spatial-temporal modeling techniques suggested. Some authors analyse Wind Speed (WS) data. One of the first studies describing the spatial-temporal structure of WS has been performed by Haslett and Raftery [3] on Irland's winds. The authors base their inference on the model including deseasonalization, kriging, and fractional ARMA modeling with the main purpose of evaluating the average power output to be expected in the long term from a wind turbine at a given site. Yan et al. [5] apply a generalized linear model for modeling daily WS time series in Northwestern Europe. Cripps et al. [2] use a Bayesian hierarchical model to predict the surface wind field 1 day ahead within Sydney Harbour. Ailliot et al. [1] propose a Markov switching AR model for modeling the spatial-temporal evolution of WS. Similar results are described in Šaltytė-Benth and Šaltytė [4]. Modeling technique in this paper is based on the same idea. Data about daily WS data in 20 meteorological stations in Lithuania were analyzed. Lithuanian Hidrometeorological Service has provided data in pdf format starting from 1967 year. For modeling 18 stations where selected, and in each of them seasonal time series model was fitted. After eliminating dependence in time, spatial dependence was analyzed. Spatial dependence in Šaltytė-Benth and Šaltytė [4] was based on spatial exponential correlation function. In this paper spatial connection, based on semivariogram is presented.

1 Time series modeling

We consider spatial-temporal data to be a realization of a random field

$$\big\{ Z(\mathbf{s}; t) \colon \mathbf{s} \in D \subset R^2, \ t \in [0; \infty) \big\}, \tag{1}$$

where $\mathbf{s}$ defines spatial coordinates and t indexes time. As it is usual, we decompose a spatial-temporal model $Z(\mathbf{s}; t)$ into a mean component $\mu(\mathbf{s}; t)$ modeling the trend of

the field, and residual component $\varepsilon(\mathbf{s}; t)$ modeling the variations around the trend in both space and time. Then, the spatial-temporal random field $Z(\mathbf{s}; t)$ may be written as

$$Z(\mathbf{s}; t) = \mu(\mathbf{s}; t) + \varepsilon(\mathbf{s}; t), \tag{2}$$

where $\mu(\mathbf{s}; t)$ is a deterministic function of the space and time coordinates and is defined by

$$\mu(\mathbf{s}; t) = S(\mathbf{s}; t) + \sum_{i=1}^{p_s} \phi_i(\mathbf{s})\big(Z(\mathbf{s}; t - 1) - S(\mathbf{s}; t - 1)\big), \tag{3}$$

here $S(\mathbf{s}; t)$ is a seasonal function of space and time, and $\phi_i(\mathbf{s})$ are space-dependent parameters of an $AR(p_s)$ process.

Next, we decompose the residual field into

$$\varepsilon(\mathbf{s}; t) = \sum (\mathbf{s}; t) \epsilon(\mathbf{s}; t), \tag{4}$$

where $\sigma(\mathbf{s}; t)$ is a non-random function, satisfying the condition $\sigma(\mathbf{s}; t) = \sigma(\mathbf{s}; t + 365)$ for any time t.

We also assume that

$$\big\{\epsilon(\mathbf{s}; t) \colon \mathbf{s} \in D \subset R^2, \ t \in [0; \infty)\big\}, \tag{5}$$

is a zero-mean stationary Gaussian spatial-temporal random field which is independent in time and has a spatial correlation function.

We therefore assume that the cross-correlation in time and space is equal to zero. We also note that the residuals $\varepsilon(\mathbf{s}; t)$ are uncorrelated but dependent in time and correlated in space. Therefore, the model of the spatial-temporal covariance function for $\varepsilon(\mathbf{s}; t)$:

$$C(s, s + h_s; \theta) = \rho(h_s; \theta_s) \sum_s (\theta_s) \sum_{s+h_s} (\theta_t). \tag{6}$$

Where $\rho(h_s; \theta_s)$ is spatial correlation function and $\sum_s(\theta_t)$ is a diagonal variance matrix at the location $\mathbf{s} \in D$.

First time series model at a single spatial location independently of the spatial information is considered, and then model spatial correlation function on time-independent residuals. All parameters from time series analysis into a spatial model are combined.

Let $Z_k(t)$ is a time series in the spatial location $s_k \in D$, $k = 1, \dots, n$; $t = 1, \dots, T$. Then

$$Z_k(t) = \mu_k(t) + \varepsilon_k(t), \tag{7}$$

where $\mu_k(t)$ and $\varepsilon_k(t)$ denote the mean and the residual process at the moment $t = 1, \dots, T$ at the spatial location $s_k \in D$, $k = 1, \dots, n$ respectively.

Here

$$\mu_k(t) = S_k(t) + \sum_{i=1}^{p_k} \phi_i^k(s)\big(Z_k(t - 1) - S_k(t - 1)\big), \tag{8}$$

with ϕ_i^k being parameters of a $AR(p_k)$ process and

$$S_k(t) = a_1^k + \sum_{l=1}^{L} a_{2l+1}^k \cos\left(\frac{(2l + 2)\pi t}{365}\right) + a_{2l+1}^k \sin\left(\frac{(2l + 2)\pi t}{365}\right) \tag{9}$$

describing the seasonality of WS in spatial location $s_k \in D$.
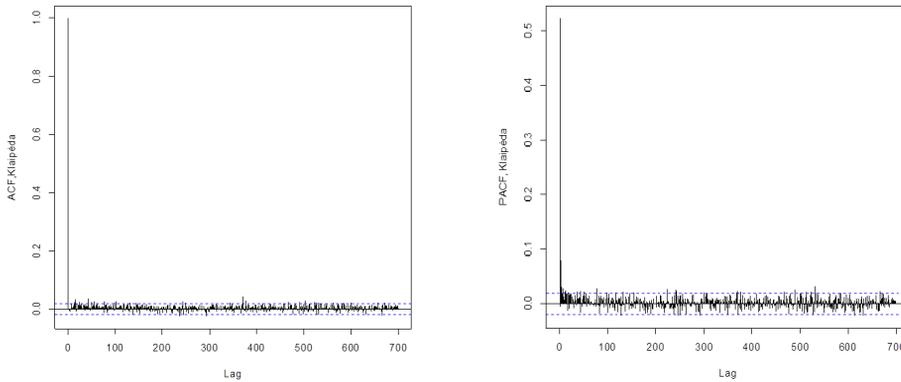
Fig. 1. The ACF and PACF functions after eliminating seasonality in time series.

We assume that the residual process $\varepsilon_k(t)$ is of the following form

$$\varepsilon_k(t) = \sigma_k(t)\epsilon_k(t) \tag{10}$$

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

The model was estimated on in-sample data from the period 1 January 1977 to 31 December 2006, consisting of 10,950 observations recorded in 18 stations. First seasonal function (9) for WS data at each station separately was estimated. The function with only three parameters was sufficient to remove seasonal fluctuations in data.

In Fig. 1, the ACF and partial ACF (PACF) of residuals obtained after eliminating seasonal effects from data in Klaipėda are plotted. Seasonal variations were removed in the data; however, the residuals still exhibited strong autocorrelations. As indicated by the PACF plot, a higher-order AR process is needed to explain the autocorrelations.

As AR order selection criteria AIC together with Box-Ljung test and histograms of the residuals were used. We started with a low-order AR process and proceeded with estimating the processes with an increasing order. The order was bounded to 10. For most of the stations, $AR(3)$ or $AR(4)$ was sufficient to capture the variations in the residuals. However, in four stations an $AR(6)$ or $AR(8)$ in Nida was needed. An ARMA process with the combination of AR and MA components did not provide a better fit.

The residuals after fitting AR became not autocorrelated according to the Box-Ljung statistics. The autocorrelations of residuals seemed to vary randomly around zero; however, the ACF of the squared residuals revealed time dependency in the variance (Fig. 2).

The residuals were first normalized by dividing them by the empirical variance of residuals.

For model validation, we use 365 out-of-sample WS observations from the year 2007 available in all 18 stations. To validate the temporal model, one-step-ahead predictions were generated for out-of-sample observations and the prediction errors (PE)
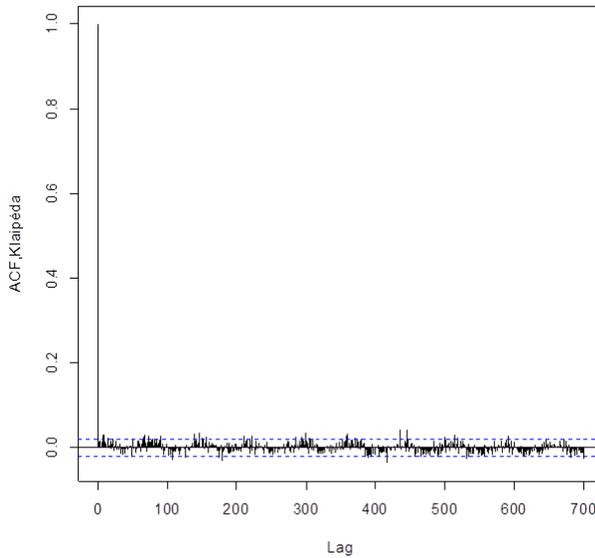
Fig. 2. ACF of squared residuals after seasonal and AR(p) components were eliminated from data in Klaipėda.

defined as the differences between the observations and predictions were calculated. The PEs were normally distributed for all but two stations Even though the normality assumption was not verified at the 5% significance level in Šiauliai and Utena, the histograms demonstrated a reasonable symmetry in these stations.

2   Spatial modeling

In spatial connection method for fitted time series parameters semivariogram is used. Semivariogram is fitted to time independent and normally distributed residuals. As in each location there was 10950 observations (residuals), before fitting empirical semivariogram model, residuals where averaged for each day, resulting in 365 days.

Exponential anisotropic semivariogram was fitted:

$$\gamma(\mathbf{h};\theta) = \begin{cases} 0, & \text{when } \mathbf{h} = 0, \\ \theta_0 + \theta_1\big(1 - \exp\big(-\frac{1}{\theta_2}\sqrt{t^2 h_1^2 + h_2^2}\big)\big), & \text{when } \mathbf{h} > 0 \end{cases} \tag{11}$$

with parameters $\theta_0 = 0.03$; $\theta_1 = 0.002$; $\theta_2 = 38472.13$ and anisotropy parameter $t = 1.15$.

Time series model for new k-th location is based on spatial weights:

$$\delta_{kj} = \frac{\frac{1}{\gamma(s_{ij})}}{\sum_{j=1}^{N} \frac{1}{\gamma(s_{ij})}} \tag{12}$$

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

Table 1. Estimated time series parameters for Palanga and Trakų Vokė.

	Palanga	Trakų Vokė
a_0	1.268	0.971
a_1	0.151	0.721
a_2	−0.016	0.437
ϕ_1	0.465	0.485
ϕ_2	−0.053	−0.031
ϕ_3	0.057	0.063
ϕ_4	0.051	
δ^2	0.198	0.232

Inverse dimension is used, while semivariogram is measure of dissimilarity and inverse dimension let us consider that stronger dependence is between nearest locations.

Then parameters for new station can be calculated by:

$$\alpha_k = \alpha \cdot \delta'_k, \tag{13}$$

here $\alpha = (\alpha_1, \ldots, \alpha_n)$, when α_i is vector of estimated parameters in i-th location, i.e. $\alpha_i = (a_{0i}, a_{1i}, a_{2i}, \phi_{1i}, \phi_{2i}, \phi_{3i}, \phi_{4i}, \phi_{5i}, \phi_{6i}, \phi_{7i}, \phi_{8i}, \sigma_i^2)'$ and $\delta_k = (\delta_{1k}, \ldots, \delta_{2k})$ – weight vector.

In the estimated temporal model, there are 12 parameters in each station. The $AR(3)$ or $AR(4)$ process was sufficient to capture the autocorrelations in the deseasonalized data. In four stations, however, a higher-order AR was needed. We therefore assumed that the $AR(8)$ process was fitted in all stations and set unestimated parameters equal to zero.

In Table 1 estimated parameters for two out of sample stations are shown.

By using estimated parameters, spatio-temporal predictions were calculated for out of sample data (2007 January 1st), predictions were compared to real data in Palanga and Trakų Vokė stations. Prediction in Palanga was 7.9 m/s, while observed WS is 8.8. Prediction in Trakų Vokė was 6.1 m/s, while observed WS for this day was 6.6 m/s.

Results shows, that spatial-temporal model for WS data fit quite well.

References

[1] P. Ailliot, V. Monbet and M. Prevosto. An autoregressive model with time-varying coefficients for wind fields. *Environmetrics*, **17**:107–117, 2006.

[2] E. Cripps, D. Nott, W.T.M. Dunsmuir and Ch. Wikle. Space-time modelling of Sydney Harbour winds. *Aust. NZ J. Stat.*, **47**(1):3–17, 2005.

[3] J. Haslett and A.E. Raftery. Space-time modelling with long-memory dependence: Assessing Ireland's wind power resource. *Appl. Stat.*, **38**(1):1–50, 1989.

[4] J. Šaltytė-Benth and L. Šaltytė. Spatial-temporal model for wind speed in Lithuania. *J. Appl. Stat.*, **38**(6):1151–1168, 2011.

[5] Z. Yan, S. Bateand R.E. Chandler, V. Isham and H. Wheater. An analysis of daily maximum wind speed in northwestern europe using generalized linear models. *J. Climate*, **15**:2073–2088, 2002.

REZIUMĖ

Erdvinis ARMA modelis vėjų greičių duomenims

L. Paulionienė

Straipsnyje pateikiamas erdvės – laiko modelis, skirtas vėjų greičių, matuotų kasdien nuo 1967 m. sausio 1 d. analizei. Pirmiausia, kiekviename mieste sudarytas laiko eilutės modelis, kadangi vėjų greičiams būdingas sezoniškumas, pirmiausia įvertinta sezoninė funkcija, tada $AR(p)$ modelis. Laiko eilučių liekanos normalizuotos padalinus jas iš standartinio nuokrypio. Tokiu būdu, gavus nepriklausomus laike stebėjimus, semivariogramos pagalba, įvertinta erdvinė priklausomybė. Tikrinant modelio validumą, apskaičiuoti erdviniai svoriai Trakų Vokės bei Palangos stotims, kurios nebuvo įtrauktos į modeliavimą dėl per mažo stebėjimų skaičiaus bei apskaičiuota vieno žingsnio į priekį prognozė 2007 metų sausio 1 dienai.

Raktiniai žodžiai: erdvės – laiko, AR, semivariograma, vėjo greitis, erdvinis sujungimas.