IDŐJÁRÁS

Quarterly Journal of the Hungarian Meteorological Service
Vol. 115, No. 1–2, January – June 2011, pp. 71–85

Geostatistical modeling of high resolution climate change scenario data

Ladislav Vizi[1], **Tomáš Hlásny**[2,3*], **Aleš Farda**[4], **Petr Štepánek**[4], **Petr Skalák**[4], and **Zuzana Sitková**[2]

[1]*Technical University Košice, Faculty of BERG,*
Park Komenského 19, Košice 04200, Slovakia; E-mail: ladislav.vizi@tuke.sk

[2]*National Forest Centre – Forest Research Institute,*
T. G. Masaryka 22, Zvolen – 96001, Slovakia
E-mail: sitkova@nlcsk.org

[3]*Czech University of Life Sciences, Faculty of Forestry and Environment,*
Kamýcká 1176, Praha 6 – Suchdol 165 21, Czech Republic

[4]*Czech Hydrometeorological Institute,*
Na Šabatce 17, 14306, Prague, Czech Republic
E-mails: ales.farda@chmi.cz; skalak@chmi.cz; stepanek@chmi.cz

[*]*Corresponding author; E-mail: hlasny@nlcsk.org*

(Manuscript received in final form November 22, 2010)

Abstract—Air temperature and precipitation data organized within a 10×10 km grid covering the whole of Slovakia were subject to analysis. The source data are produced by the ALADIN Climate/CZ regional climatic model. The output of the global climatic model ARPEGE-Climat (Meteo-France) provided the driving data for the regional model. The IPCC A1B scenario provides the information on the future development of greenhouse gas emissions. Such scenario was developed within the 6th Framework Programme project CECILIA (Central and Eastern Europe Climate Change Impacts and Vulnerability Assessment).

Geostatistical prediction of annual mean air temperature and precipitation data was carried out for the reference (1961–1990) and distant future (2071–2100) climates. The experimental data were non-stationary and significantly correlated with elevation. Therefore, we used non-stationary multivariate geostatistical techniques allowing for the integration of such information. In particular, we used kriging of residuals, universal kriging with external drift, and external drift kriging in the scope of IRF-*k* (intrinsic random functions of order *k*). Prediction based on linear regression of elevation data was used as a complementary technique. Accuracy assessment was based upon the mean square errors produced by cross-validation in case of kriging-based predictions and upon the mean square residual in case of linear regression-based prediction.

We found that all kriging-based techniques outperformed the linear regression-based approach, yielding mean square error lower by 53–75%. External drift kriging in the scope of IRF-*k* produced slightly better results for most of the climate variables analyzed. The poorest results were achieved in the case of annual mean air temperature for the period 1961–1990, where the variogram of residuals was very erratic.

External drift kriging-based techniques were found to be very efficient for interpolating annual mean air temperature and annual precipitation data organized in regular grids. Accuracy assessment indicated that the three predictors used yielded almost identical results for a single variable, while significant differences in mean square error were observed in a between-variables comparison.

Key-words: mean annual air temperature, annual precipitation totals, ALADIN regional climatic model, external drift kriging, non-stationary modeling, Slovakia

1. Introduction

Maps of various climate elements produced by spatial interpolation of point-distributed data are frequently used to improve understanding of climate's spatio-temporal variability as well as for various studies of climate impacts on society and ecosystems (*Haines et al.*, 2006; *Trnka et al.*, 2004; *Hlásny* and *Turčáni*, 2009).

Recent availability of large amounts of climate data produced by global and regional climate models (GCMs / RCMs) has drawn attention to the need for optimizing the spatial interpolation of such data (e.g., *Haylock et al.*, 2008). The data are primarily organized in regular grids with spacing depending on the respective GCM / RCM. However, follow-up studies on agriculture, forestry, air pollution, and other areas often ask for seamless information on climate rather than point-distributed data. Therefore, the search for optimal interpolation techniques is a timely task (*Mulugeta*, 1996; *Dobesch et al.*, 2007). A growing number of recent works comparing interpolation techniques and identifying optimal data- or region-specific methods is testimony to this issue's importance (*Goovaerts*, 2000a; *Haberland*, 2007).

In addition to the frequently used non-model-based techniques (not using a variogram, such as inverse distance weighting or spline interpolation, e.g., *Hancock* and *Hutchinson*, 2006), there exists a range of geostatistical techniques allowing for specific improvements of spatial interpolation, mainly by integrating heterogeneous data (*Isaak* and *Srivastava*, 1989). In this paper, we demonstrate the use of several external drift kriging-based techniques (EDK, hereinafter) (*Matheron*, 1973) for interpolating high resolution climate change scenario data. These techniques allow for flexible integration of point-distributed climatic data with correlated grid-distributed predictor variables, such as elevation and solar insolation.

An early paper on EDK's use for predicting air temperature and precipitation in Scotland was published by *Hudson* and *Wackernagel* (1994). Later, EDK's ability to integrate heterogeneous data prompted many other

climatology studies. *Carera-Hernandez* and *Gaskins* (2007) found that the use of elevation as a secondary variable improves the prediction, even if the correlation is low. The influence of such other terrain-related parameters as relief slope and aspect was investigated by *Attore et al.* (2007), who found that universal kriging with external drift performed the best for 17 out of 21 climatic variables analyzed. *Goovaerts* (2000a) tested the efficiency of several approaches to spatial interpolation of rainfall data (linear regression, ordinary cokriging, kriging with external drift, simple kriging with local means) and stressed the benefits of incorporating the elevation data. That author found that the latter two named techniques yield slightly better results than did the others.

The purpose of this paper is to analyze the climatic data produced by the ALADIN-Climate/CZ regional climate model (*Farda et al.*, 2010) for the whole of Slovakia in 10×10 km spatial resolution. Maps of mean annual air temperature (hereinafter just air temperature) and mean annual precipitation totals (hereinafter just precipitation) for the reference (1961–1990) and distant future (2071–2100) climates were to be produced. In particular, we focused on:

(1) describing and preprocessing the data,

(2) using several EDK-based techniques and a linear regression-based approach for spatial prediction of air temperature and precipitation data for the reference and distant future climates, and

(3) assessing the accuracy of the maps produced and discussing the results.

2. Data

The reference and future climate data were originally calculated using the GCM ARPEGE–Climat V4 (*Déqué*, 2007) in an experiment performed by CNRM/-Météo-France. Because of rather coarse resolution of the GCM (~50 km over Central Europe), the RCM ALADIN-Climate/CZ (*Farda et al.*, 2010) was used for additional downscaling of the GCM data. The IPCC A1B emission scenario was adopted to provide information on future development of greenhouse gas emissions. The data were developed as part of the CECILIA (Central and Eastern Europe Climate Change Impacts and Vulnerability Assessment, www.cecilia-eu.org) project under the European Union's 6th Framework Programme. The RCM covers Central Europe with a resolution of 10 km. Such resolution allows for better representation of the driving physical processes (e.g., more accurate resolution of geographical features and thus, various interactions with the surface), thus, leading to better description of local climate and positively affecting the quality of the simulations.

The data used in this study comprise a subset of ALADIN's entire integration domain covering the Slovak Republic. The 10 km resolution grid, with rotation 6^{o} azimuth, is extended beyond the country's borderline by

approximately one grid point in order to reduce interpolation errors in the edge locations (*Fig. 1*). In total, 644 grid points are used for the analysis. Source data statistics are given in *Table 1*.

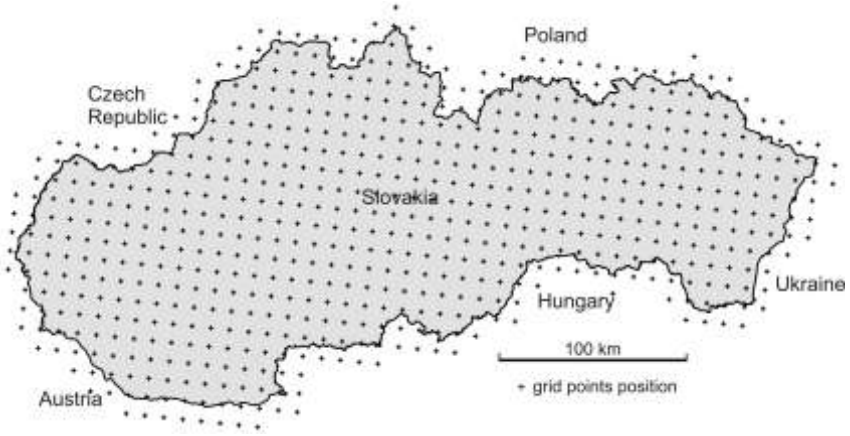

Fig. 1. Spatial arrangement of 10×10 km grid of the ALADIN-Climate/CZ regional climate model in Slovakia.

Table 1. Source data statistics. Abbreviations: N – number of observations, Min – minimum, Max – maximum, Avg – average, Med – median, SD – standard deviation, IQR – inter-quartile range, Skew – skewness, Kurt – kurtosis. Variables: T – mean annual air temperature for the given period, P – mean annual precipitation totals for the given period

Variable	N	Min	Max	Avg	Med	SD	IQR	Skew	Kurt
T 1961–1990	644	0.6	10.7	7.3	7.4	2.0	3.1	–0.32	–0.42
T 2071–2100	644	5.2	13.4	10.7	10.8	1.8	3.2	–0.40	–0.90
P 1961–1990	644	416.5	1206.7	697.3	675.6	145.1	207.8	0.83	0.12
P 2071–2100	644	472.5	1175.0	671.3	629.7	154.5	214.9	0.94	0.05

Elevation of the study area is used as a supportive variable (*Fig. 1*). It is organized in a 180 m resolution grid, which is more than 55 times denser than the ALADIN grid.

3. Methods

The climate data used are clearly non-stationary, as they have a global elevation-controlled trend in the south-north direction. Therefore, we describe here the concepts for multivariate non-stationary geostatistical modeling that are used. All steps of the geostatistical analysis were carried out in the ISATIS v.9 environment (*Geovariances, Centre de Géostatistique in Fontainebleau*). For regression modeling, STATISTICA v.7 (*StatSoft, Inc.,* 2004) was used.

3.1. Stationary spatial models

Stationarity of spatial data, i.e., the presence of a stable mean for an analyzed variable, is the simplest and most frequently documented case of geostatistical analysis. This allows for straightforward modeling of the variogram, which measures the spatial correlation of the studied variable, as well as for an optimal estimation using kriging. In a stationary case, where drift $m(\mathbf{x})$ is a constant, the variogram γ for distance **h** is estimated as:

$$2\gamma_Z(\mathbf{h}) = \mathrm{E}[Z(\mathbf{x}+\mathbf{h}) - Z(\mathbf{x})]^2. \quad (1)$$

For a regionalized variable, as one realization of a random function, the variogram is estimated by forming the average dissimilarities for all $N(\mathbf{h})$ pairs of data $z(\mathbf{x}_\alpha)$ and $z(\mathbf{x}_\alpha + \mathbf{h})$ available at sample points $\mathbf{x}_\alpha$ that are linked by the vector $\mathbf{h} = \mathbf{x}_\alpha - \mathbf{x}_\beta$ (*Hudson* and *Wackernagel*, 1994):

$$2\gamma_Z(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} (z(\mathbf{x}_\alpha) - z(\mathbf{x}_\beta))^2. \quad (2)$$

Usually, we observe that the average dissimilarities between the couples of values increase when **h** is increased, up to a value of the variable autocorrelation (range of influence). Beyond this value, the dissimilarities become more or less constant around an upper asymptote (sill of the variogram) that is approximately equal to the data variance.

3.2. Non-stationary spatial models

In a non-stationary case, there is a definite trend in the data, being a gradient in a given direction (*Hayet et al*., 2000). The non-stationary approach to spatial modeling considers the phenomenon under study as a sum of two terms:

$$Z(\mathbf{x}) = Y(\mathbf{x}) + m(\mathbf{x}), \quad (3)$$

where $Y(\mathbf{x})$ describes the local variation of $Z(\mathbf{x})$, and it is assumed to be stationary with constant mean. The term $m(\mathbf{x})$ describes a large-scale variation of $Z(\mathbf{x})$ (drift). It is assumed that the drift can be represented by a polynomial of order *L*:

$$m(\mathbf{x}) = \sum_{l=0}^{L} a_l f_l(\mathbf{x}), \quad (4)$$

where a_l are unknown coefficients of known functions $f_l(\mathbf{x})$ of the spatial coordinates. Note that for $L=1$, Eq. (4) reduces to a constant term, a_0, which

indicates no trend in the spatial coordinates. The term $Y(\mathbf{x})$ in Eq. (3) represents the residual, i.e., the amount of variability remaining after the drift has been removed. The residuals have a stationary covariance (variogram) function between any pairs of random variables $\{Y(\mathbf{x}),Y(\mathbf{x}+\mathbf{h})\}$. The drift is essentially the mean value of the variable as a function of the location at which the variable is measured. In a non-stationary case, we can rewrite Eq. (1) as follows:

$$\begin{aligned} 2\gamma_Z(\mathbf{h}) &= \mathrm{E}[Z(\mathbf{x}+\mathbf{h})-Z(\mathbf{x})]^2 \\ &= \mathrm{E}[Y(\mathbf{x}+\mathbf{h})+m(\mathbf{x}+\mathbf{h})-Y(\mathbf{x})-m(\mathbf{x})]^2 \\ &= \mathrm{E}[Y(\mathbf{x}+\mathbf{h})-Y(\mathbf{x})]^2-[m(\mathbf{x}+\mathbf{h})-m(\mathbf{x})]^2 \\ &= 2\gamma_Y(\mathbf{h})-[m(\mathbf{x}+\mathbf{h})-m(\mathbf{x})]^2, \end{aligned} \tag{5}$$

where the trend values $m(\mathbf{x})$ and $m(\mathbf{x}+\mathbf{h})$ are unknown. The second term on the right side of Eq. (5) provides the drift estimate in a particular direction. The most straightforward approach to non-stationary modeling is based on computation of a residual variogram $2\gamma_Y(\mathbf{h})$. A proper use of this technique is documented by *Dowd* (1984) and *Goovaerts* (2000b), who suggested several ways for coping with certain shortcomings of this technique, as discussed, for example, by *Hayet et al.* (2000).

Another approach to non-stationary modeling used in this paper is the method of increments based on the theory of intrinsic random function of order *k* (IRF-*k*) (*Matheron*, 1973). It defines a linear combination of *Z* data that filters out the drift component $m(\mathbf{x})$. In a stationary case, the first order difference, or increment $[Z(\mathbf{x}+\mathbf{h})-Z(\mathbf{x})]$, filters out the constant drift *m*. In a non-stationary case, higher order differentiation is required to filter out the higher orders of the polynomial drift. This approach leads to a so-called generalized covariance model $K(\mathbf{h})$ instead of a variogram $\gamma(\mathbf{h})$. The most widely used models for generalized covariances are polynomial in form (*Matheron*, 1973):

$$K(\mathbf{h})=a_0+\sum_{k=0}^{K}(-1)^{k+1}a_{k+1}|\mathbf{h}|^{2k+1}. \tag{6}$$

More information on this technique can be found in *Dowd* (1984) or *Chiles* and *Delfiner* (1999).

3.2.1. Case of external drift(s)

In case of a non-stationary spatial model, we consider the trend $m(\mathbf{x})$ of the variable $Z(\mathbf{x})$ to be a function of spatial coordinates. For some applications, exhaustive data for one or more regionalized variables $s_j(\mathbf{x})$ may be available

in the studied domain (representing, e.g., elevation). If such data are available, it is worthwhile to use them as additional constraints to the interpolation.

If we assume that $Z(\mathbf{x})$ is on average equal to $s_j(\mathbf{x})$ up to linear way and with coefficients a_0 and b_1, then:

$$\mathrm{E}[Z(\mathbf{x})] = m(\mathbf{x}) = a_0 + \sum_{j=1}^{J} b_j s_j(\mathbf{x}). \tag{7}$$

Because variables $s_j(\mathbf{x})$ are exhaustively available, they reflect the average shape of $Z(\mathbf{x})$, where just the scaling is different (*Hudson* and *Wackernagel*, 1994).

3.3. Interpolation techniques

In this study, we include several techniques under the term external drift kriging. Their common feature is that the elevation acts as an external drift correlated with the primary climatic variables. In addition, a linear regression of the elevation data was used to predict the climate data.

Residual kriging, known also as regression kriging (*Odeh et al.*, 1994) or kriging after detrending (*Goovaerts*, 2000b), predicts the residuals at all nodes of the interpolated grid $\mathbf{x}_0$, $Y^*(\mathbf{x}_0)$. Residual kriging uses the drift $m^*(\mathbf{x})$ calculated by a polynomial of a selected degree by the least squares method. Residuals $Y(\mathbf{x}_\alpha)$ are calculated as the differences between $Z(\mathbf{x}_\alpha)$ and $m^*(\mathbf{x}_\alpha)$ at all sample points. Using the variogram of residuals, the kriging system for weights $\omega(\mathbf{x}_\alpha)$, $\alpha = 1,\ldots,n$ includes $n+1$ linear equations:

$$\begin{cases} \sum_{\beta=1}^{n} \omega(\mathbf{x}_\beta)\gamma_Y(\mathbf{x}_\alpha,\mathbf{x}_\beta) + \lambda_0 = \gamma_Y(\mathbf{x}_0,\mathbf{x}_\alpha) \quad \text{for} \quad \alpha = 1,\ldots,n, \\ \sum_{\alpha=1}^{n} \omega(\mathbf{x}_\alpha) = 1. \end{cases} \tag{8}$$

The residual kriging estimator is a linear combination of available n data $y(\mathbf{x}_\alpha)$ for only n random variables $Z(\mathbf{x}_\alpha)$:

$$y^*(\mathbf{x}_0) = \sum_{\alpha=1}^{n} \omega(\mathbf{x}_\alpha) y(\mathbf{x}_\alpha). \tag{9}$$

Finally, the estimated drift, Eq. (4) and kriged residuals, Eq. (9) are added together.

Universal kriging provides an unbiased estimation, which considers drift $m(\mathbf{x})$ as a continuous and regular function (Eq. (4)), usually restricted to polynomials up to the order of 2. It uses a model representing both local and global variability of the variable in space. It determines the underlying variogram of $Y(\mathbf{x})$ and estimates the degree of drift. We modeled the drift by Eq. (4), including elevation as the external drift. The simultaneous system of equations for the universal kriging estimator, considering both internal and external drift, is as follows:

$$\begin{cases} \sum_{\beta=1}^{n} \omega(\mathbf{x}_\beta)\gamma_Z(\mathbf{x}_\alpha,\mathbf{x}_\beta) + \lambda_0 + \sum_{l=1}^{L}\lambda_l f_l(\mathbf{x}_\alpha) + \sum_{j=1}^{J}\lambda_j s_j(\mathbf{x}_\alpha) = \gamma_Z(\mathbf{x}_o,\mathbf{x}_\alpha) \\ \qquad\qquad \text{for } \alpha=1,\ldots,n, \\ \sum_{\alpha=1}^{n} \omega(\mathbf{x}_\alpha) = 1, \\ \sum_{\alpha=1}^{n} \omega(\mathbf{x}_\alpha) s_j(\mathbf{x}_\alpha) = s(\mathbf{x}_o) \qquad \text{for } j=1,\ldots,J, \\ \sum_{\alpha=1}^{n} \omega(\mathbf{x}_\alpha) f_l(\mathbf{x}_\alpha) = f_l(\mathbf{x}_o) \qquad \text{for } l=1,\ldots,L. \end{cases} \tag{10}$$

The kriging system for IRF-*k* is similar to the universal kriging system, Eq. (10). The only difference is that it uses the generalized covariance model $K(\mathbf{h})$ (Eq. (6)) instead of the variogram $\gamma(\mathbf{h})$ (Eq. (2)). More details about IRF-*k* can be found in *Dowd* (1984) or *Chiles* and *Delfiner* (1999).

3.4. Linear regression-based estimation

The generally recognized relationship between the climate variables addressed and elevation allows for a simple prediction of climate data at all positions for which elevation data are available. There exists a set of collocated climate $z(\mathbf{x}_\alpha)$ and elevation $s(\mathbf{x}_\alpha)$ data $[z(\mathbf{x}_\alpha), s(\mathbf{x}_\alpha)]$; $\alpha=1,\ldots,n$, where n is the number of observations. The prediction $z^*(\mathbf{x}_o)$ is based on a linear relationship:

$$z^*(\mathbf{x}_o) = a_0^* + b_1^* s(\mathbf{x}_\alpha), \tag{11}$$

where coefficients a_0^* and b_1^* are estimated from the collocated climate and elevation data. A major shortcoming of this type of prediction is that the climate data at a particular grid node are derived only from the collocated elevation, regardless of the surrounding observed climate data (*Goovaerts*, 2000a).

3.5. Accuracy assessment

Two techniques were used to assess the accuracy of the maps produced and the performance of the predictors used. A cross-validation procedure was used in case of geostatistical predictions (*Isaaks* and *Srivastava*, 1989; *Clark*, 1986). The technique temporarily removes one observation at a time from the data set and "re-estimates" this value from the remaining data using a given predictor. Such procedure produces couples of values, the differences between which yield cross-validation residuals. The main criterion for assessing accuracy is mean square error (MSE), which measures the average squared difference between the observed $z(\mathbf{x}_\alpha)$ and predicted $z^*(\mathbf{x}_\alpha)$ values:

$$\mathrm{MSE} = \frac{1}{n} \sum_{\alpha=1}^{n} [z(\mathbf{x}_\alpha) - z^*(\mathbf{x}_\alpha)]^2 , \tag{12}$$

where *n* is the number of observations.

Correlation coefficients of observed versus predicted values, normality of residuals distribution, mean value of residuals (criterion that the mean is approaching zero), and degree of randomness of spatial distribution of residuals can also be used.

Another approach was used in the case of linear regression-based prediction. The MSE was computed as the average square residual value for the linear model fitted using all observations:

$$\mathrm{MSE} = \frac{1}{n} \sum_{\alpha=1}^{n} [z(\mathbf{x}_\alpha) - (a_0^* + b_1^* s(\mathbf{x}_\alpha))]^2 . \tag{13}$$

4. Results

4.1. Drift identification

To identify an optimal global trend, the polynomials of order one (linear) and two (quadratic) plus one external drift (elevation) were tested by the cross-validation procedure for the lowest mean square error. Other criteria, such as mean of residuals approaching zero, minimal variance, normal distribution, and well-structured directional experimental variograms, were used as well. We found that the linear drift along the x and y coordinates (internal drift), together with the elevation (external drift),

$$m^*(\mathbf{x}) = a_0 + a_1 \mathrm{x} + a_2 \mathrm{y} + b_1 s(\mathbf{x}), \tag{14}$$

performed the best for all climate variables.

4.2. Kriging-based predictions

To perform the residuals kriging, we used the trend functions described above to filter out the residuals $y(\mathbf{x}_\alpha)$ from the regionalized variable $z(\mathbf{x}_\alpha)$, then estimated the residual variogram models $\gamma_Y(\mathbf{h})$ for all the climate variables analyzed (*Fig. 2*). Estimation of the variogram model for the variable T 1961–1990 was problematic, because there were erratic directional experimental variograms without clear spatial structure. Therefore, an omnidirectional model was fitted to the experimental variogram values in this case. The variogram's origin was estimated from the directional variogram constructed in the azimuth 6^o that rises from the variogram's value at about 0.1 $(^oC)^2$. Directional experimental variograms are presented in *Fig. 2* to demonstrate that no anisotropy can be modeled in this case.

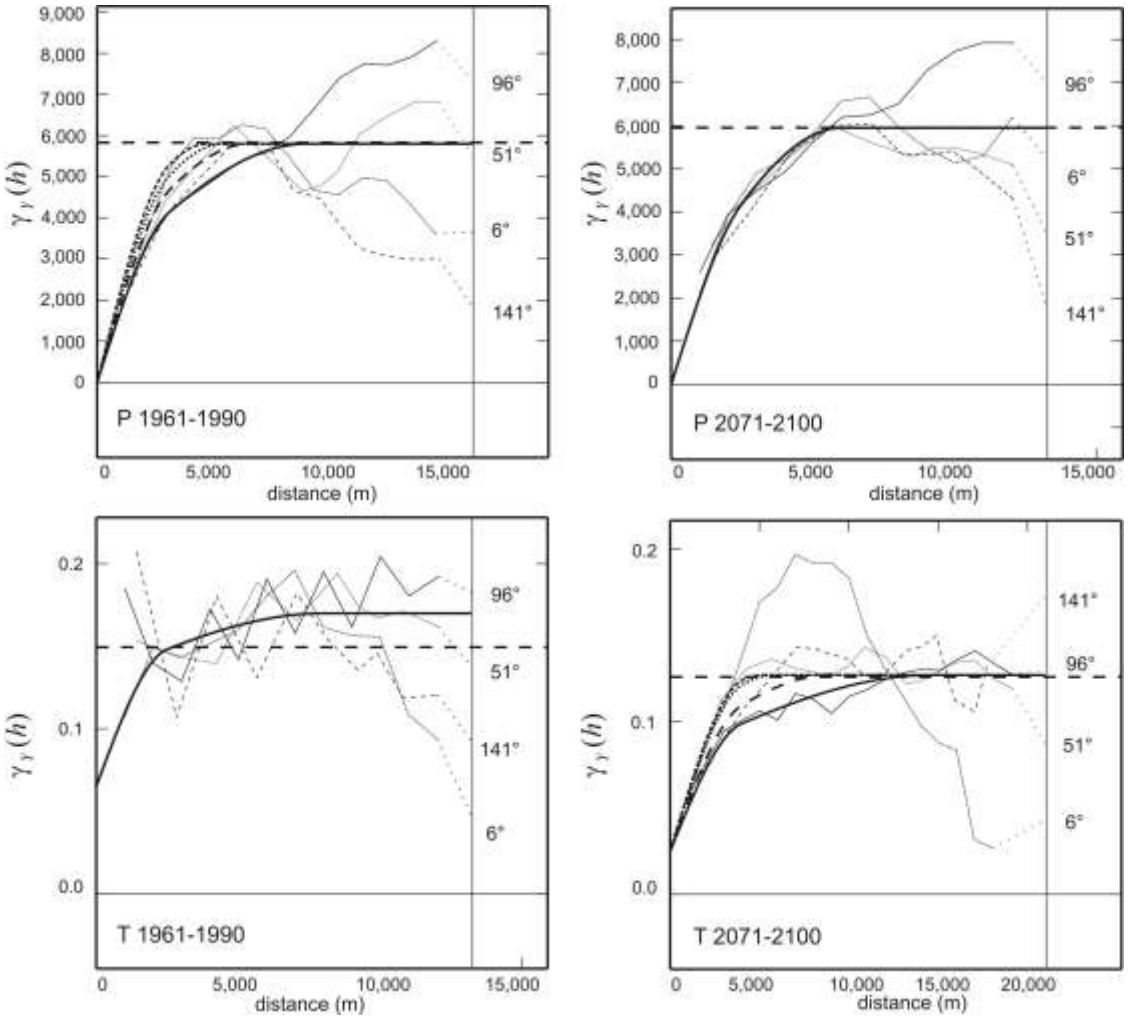

Fig. 2. Directional experimental residual variograms (thin lines) and respective variogram models (thick lines). The numbers on the right side indicate the angles at which the variograms were calculated. Abbreviations: P 1961–1990 – mean precipitation totals during the period 1961–1990, P 2071–2100 – mean precipitation totals during the period 2071–2100, T 1961–1990 – mean annual air temperature during the period 1961–1990, T 2071–2100 – mean annual air temperature during the period 2071–2100.

The estimation of kriging weights $\omega(\mathbf{x}_\alpha)$ was based on Eq. (8). Estimation of residuals $y^*(\mathbf{x}_o)$ was based on the linear combination of available data according to Eq. (9). Finally, the kriged residuals were summed with the trend model according to Eq. (3).

In case of universal kriging, the trend component is directly included into the kriging system according to Eq. (10) for the drift estimation (Eq. (14)). The final estimation was performed directly using the raw variable $Z(\mathbf{x})$.

In case of IRF-*k*, the automatic fitting procedure of the ISATIS environment was used to determine both the degree *k* of the drift and the generalized covariance. For all variables, the degree of the drift was 1 (linear in X and Y directions) plus the external drift represented by the elevation. The generalized covariance of order 1 (similar to the linear model of the variogram) without nugget effect was used for all climate variables.

For interpolation neighborhood definition (*Isaaks* and *Srivastava*, 1989), we used a so-called unique neighborhood, i.e., all available data were used to estimate a value at a particular grid node. We also tested several designs for a moving neighborhood, such as a first ring neighborhood (4 adjacent samples), second ring neighborhood (16 adjacent samples), and third ring neighborhood (36 adjacent samples). The cross-validation tests indicated that the unique neighborhood was performing the best for all climate variables. In addition, the use of moving neighborhoods resulted in "radial" artefacts in the maps produced, due to the resolution of the estimated grid which is more than 55 times higher than that of the ALADIN grid.

The maps of both variables for both time slices produced by EDK in the scope of IRF-*k* can be seen in *Fig. 3*. We can see that the elevation pattern is much stronger in the case of temperature than in that of precipitation data due to the different correlation of climate variables with elevation (*Table 2*).

4.3. Linear regression-based prediction

Linear regression-based prediction was used to provide the reference value for assessing the accuracy of the kriging-based techniques. Regression parameters from elevation and the respective climatic variables are based on all 644 observed values (*Table 2*). Mean square error was calculated using Eq. (13).

4.4. Accuracy assessment

Accuracy assessment was based on comparison of the MSE yielded by kriging-based predictions (*Table 3*) with that from the linear regression-based prediction (*Table 2*) (*Goovaerts*, 2000a). The latter technique provided the MSE reference value for evaluating the performance of kriging techniques. Proportional values of MSE are illustrated in *Fig. 4*. Such an approach allows for evaluating the performance of respective predictors for a single variable as well as for between-variable comparison. The results are discussed below.

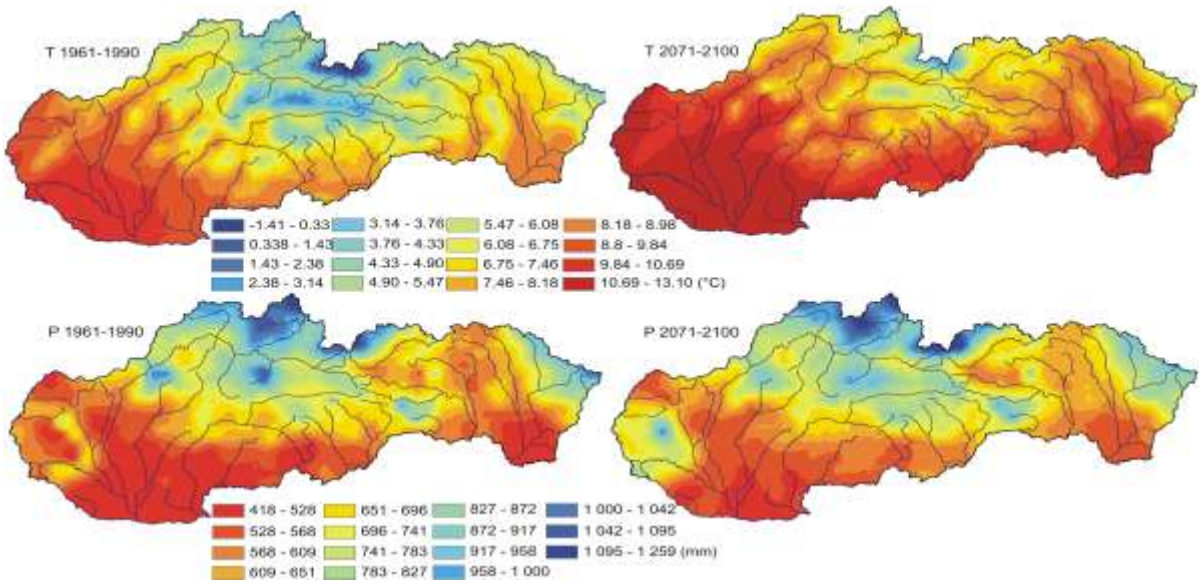

Fig. 3. Maps of mean annual air temperature and mean annual precipitation totals for the reference (1961–1990) and distant future (2071–2100) climates produced using external drift kriging in the scope of IRF-*k*.

Table 2. Linear regression parameters between elevation (X) and respective climatic variables (Y). Abbreviations: a_0^* – intercept, b_1^* – slope, R – correlation coeficient, R^2 – coefficient of determination, MSE – mean square error

X	Y	a_0^*	b_1^*	R	R^2	MSE
Elevation	P 1961–1990	526.86	0.39180	0.75	0.563	9141
Elevation	P 2071–2100	488.27	0.42060	0.76	0.578	10123
Elevation	T 1961–1990	10.26	−0.00688	−0.95	0.903	0.394
Elevation	T 2071–2100	13.34	−0.00606	−0.94	0.887	0.359

Table 3. Results of the cross-validation based accuracy assessment. Abbreviations: KR – residuals kriging, UK – universal kriging with external drift, IRF-*k* – external drift kriging in the scope of IRF-*k*, R – correlation coefficient between observed and predicted values, R^2 – coefficient of determination, MSE – mean square error of prediction, AVG – average value of residuals

Variable	Interpolator	R	R^2	MSE	AVG
T 1961–1990	KR	0.978	0.956	0.17457	0.000286
T 1961–1990	UK	0.978	0.957	0.17165	0.000286
T 1961–1990	IRF-*k*	0.976	0.953	0.184628	0.000410
T 2071–2100	KR	0.990	0.981	0.056014	−0.000449
T 2071–2100	UK	0.991	0.981	0.055269	−0.000327
T 2071–2100	IRF-*k*	0.990	0.981	0.054161	−0.000306
P 1961–1990	KR	0.954	0.909	1588.9	−0.1220
P 1961–1990	UK	0.954	0.910	1564.6	−0.0280
P 1961–1990	IRF-*k*	0.955	0.912	1536.9	−0.0099
P 2071–2100	KR	0.939	0.881	2329.5	−0.0180
P 2071–2100	UK	0.940	0.884	2273.9	0.0470
P 2071–2100	IRF-*k*	0.943	0.889	2163.9	0.0660

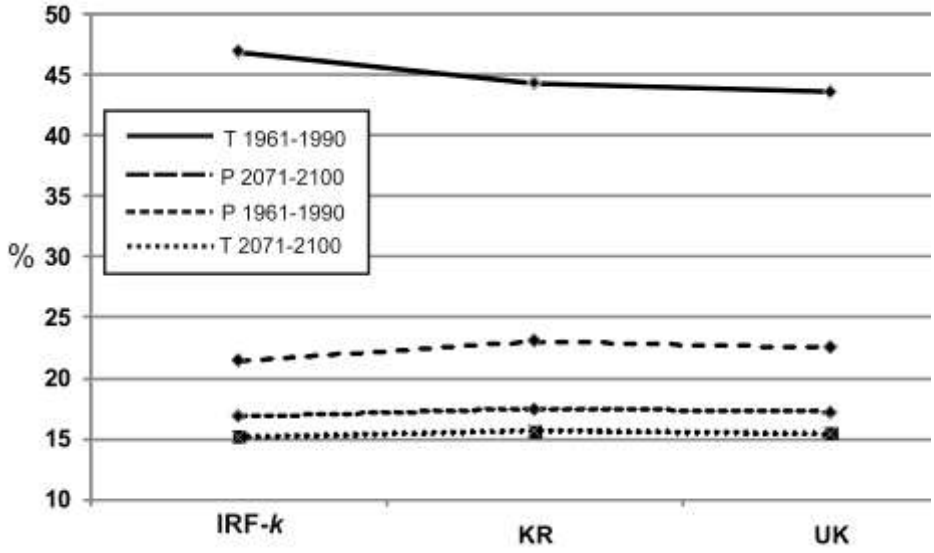

Fig. 4. Results of accuracy assessment for predictions produced by three interpolation techniques for mean annual air temperature and mean annual precipitation totals for the periods 1961–1990 and 2071–2100. The figures indicate the proportions of the MSE (mean square error) yield by cross-validation in the case of EDK-based techniques to MSE yield using the linear regression approach.

5. Conclusions and discussion

We performed a series of analyses of high resolution RCM data covering Slovakia. Both air temperature and precipitation data are well correlated with elevation (*Table 2*), and thus, we focused on the integration of that variable into the interpolation. Such supportive variable is presumed to reduce the amount of uncertainty in the maps produced. We used three external drift kriging-based techniques: residuals kriging, universal kriging with external drift, and external drift kriging in the scope of IRF-*k*. We described in details the particular steps of the geostatistical analysis to allow for a deeper understanding of those techniques used.

All kriging-based techniques produced comparable results for a single climate variable. The reason for this evidently lies in the high correlation of climate data with elevation, which covers the impact of different interpolation algorithms. In the cases of the variables T 2071–2100, P 1961–1990, and P 2071–2100, EDK in the scope of IRF-*k* yields slightly better results than do the remaining kriging-based techniques. This can reflect the benefit of using an automatized procedure in generalized covariance calculation for regularly distributed data in comparison to manual variogram fitting (the case of residuals kriging). The poorest results were reached in the case of the variable T 1961–1990, where the residuals were very erratic, and thus, they influenced the shape of the respective variograms (*Fig. 2*). This applies also for the remaining techniques, because drift parameters remain more or less stable.

Subsequently, we tested the ratio of mean square errors produced by kriging-based techniques to those from linear regression-based estimation. All kriging techniques significantly outperformed the linear regression-based estimation, which yields a mean square error 15 – 47% higher (depending on the variable). This means that, despite high correlation between climate data and elevation, information about the configuration of the surrounding data significantly improved the estimation. The accuracy assessment indicated that the three predictors used yielded almost identical results for a single variable, while significant differences in mean square error were observed by between-variables comparison.

Geostatistical techniques, in general, require a certain extent of user intervention and cannot be fully automatized. In any case, large amounts of climate data produced by various instruments require at least a semi-automatized approach when producing series of climate maps for various time slices. External drift kriging in the scope of IRF-*k* is a candidate technique for this. It yielded slightly better results than did the remaining EDK-based techniques for three out of four variables analyzed, and the underlying generalized covariance may be calculated automatically (see implementation in the ISATIS environment used in this paper). By contrast, residuals kriging requires a series of user interventions, which were not, however, compensated by improved accuracy of the prediction.

Acknowledgements **—**This research was conducted as part of the 6th Framework Programme project CECILIA (Central and Eastern Europe Climate Change Impacts and Vulnerability Assessment), and within projects of the Ministry of Agriculture of the Czech Republic, No. QH91097 "Analysis of the climate change impacts on the distribution and voltinism of *Ips typographus* in spruce forests of the Czech Republic as underlying information for their sustainable management", and the Scientific Grant Agency of the Ministry of Education of the Slovak Republic, VEGA 1/0222/08 "Application of geostatistical tools for multivariate analysis and data integration of regionalized variables".

References

Attore, F., Alfo, M., De Sanctis, M., Francesconi, F., Bruno, F., 2007: Comparison of interpolation methods for mapping climatic and bioclimatic variables at regional scale. *Int. J. Climatol. 27*, 1825-1843.

Carrea-Hernandez, J.J., Gaskin, S.J., 2007: Spatio temporal analysis of daily precipitation and temperature in the Basin of Mexico. *J. Hydrol. 336*, 231-249.

Chiles, J.P., Delfiner, R., 1999: *Geostatistics – Modeling Spatial Uncertainty*. Oxford University Press.

Clark, I., 1986: The art of cross validation in geostatistical applications. In *19*th *Application of Computers and Operational Research in the Mineral Industry* (ed.: *V. Ramani*), 211-220.

Déqué, M., 2007: Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: model results and statistical correction according to observed values. *Global Planet. Change 57*, 16-26.

Dobesch, H., Dumolard, P., Dyras, I. (eds.), 2007: *Spatial Interpolation for Climate Data: The Use of GIS in Climatology and Meteorology. Geographical Information Systems Series*. Wiley-ISTE, New York.

Dowd, P.A., 1984: *MINE5260: Non-Stationarity. MSc notes in Mineral Resources and Environmental Geostatistics.* Dpt. of Mining and Mineral Engineering, University of Leeds, United Kingdom.

Farda A., Déqué, M., Somot S., Horányi A., Spiridonov V., Tóth. H., 2010: Model ALADIN as a Regional Climate Model for Central and Eastern Europe. *Studia Geophysica et Geodaetica, Journal of the Czech Academy of Sciences 54*, 313-332.

Geovariances: ISATIS v.9, Centre de Géostatistique in Fontainebleau.

Goovaerts, P., 2000a: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *J. Hydrol. 228*, 113-129.

Goovaerts, P., 2000b: Using elevation to aid the geostatistical mapping of rainfall erosivity. *Catena 34*, 227-242.

Haberland, U., 2007: Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event. *J. Hydrol. 332*, 144-157.

Haines, A., Kovats, R. S., Campbell-Lendrum, D., Corvalan, C., 2006: Climate change and human health: impacts, vulnerability, and mitigation. *Public Health 367*, 2101-9.

Hancock, P.A., Hutchinson, M.F., 2006: Spatial interpolation of large climate data sets using bivariate thin plate smoothing splines. *Environ.Modell. Softw. 21*, 1684-1694.

Hayet, Ch., Galli, A., Ravenne, C., Tesson, M., de Marsily, G., 2000: Estimating the Depth of Stratigraphic Units from Marine Seismic Profiles Using Nonstationary Geostatistics. *Nat. Resour. Res. 9*, 77-95.

Haylock, M.R., Hofstra, N., Klein Tank, A.M.G., Klok, E.J., Jones, P.D., New, M., 2008: A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *J. Geophys. Res. 113*, D20119.

Hlásny, T., Turčáni, M., 2009: Insect pests as climate change driven disturbances in forest ecosystems. In Bioclimatoloy and Natural Hazards (eds.: *K. Střelcová, C. Mátyás, A. Kleidon, M. Lapin, F. Matejka, M. Blaženec, J. Škvarenina, J. Holécy*). Springer, Netherlands, 165–178.

Hudson, G., Wackernagel, H., 1994: Mapping temperature using kriging with external drift: Theory and example from Scotland. *Int. J. Climatol. 14*, 77-91.

Isaaks, H. E., Srivastava, R. M., 1989: *Introduction to Applied Geostatistics*. Oxford University Press, New York.

Matheron, G., 1973: The intrinsic random function and their applications. *Adv. Appl. Probab. 5*, 439-469.

Mulugeta, G., 1996: Manual and automated interpolation of climatic and geomorphic statistical surfaces: An evaluation. *Annals of the Association of American Geographers 86*, 324-342.

Odeh, I., McBratney, A., Chittleborough, D., 1994: Spatial prediction of soil properties from land-form attributes derived from a digital elevation model. *Geoderma 63*, 197-214.

StatSoft, Inc., (2004). STATISTICA Cz (Software system for statistical analyses), Version 7. www.StatSoft.Cz

Trnka, M., Dubrovský, M., Semerádova, D., Žalud, Z., 2004: Projections of uncertainties in climate change scenarios into expected winter wheat yields. *Theor. Appl. Climatol. 77*, 229-249.