Modelling selected climate parameters in the *ISATIS* environment

Tomáš Hlásny

Department of Ecology and Biodiversity of Forest Ecosystems, National Forest Centre, T. G. Masaryka 22

960 01 Zvolen, Slovakia

hlasny@nlcsk.org

Abstract. In the paper we described the main features of ISATIS system, data organisation and modelling workflow. We introduced an approach to modelling July average air temperature and July average precipitation totals in Slovakia by means of ISATIS capabilities. Ordinary kriging and external drift kriging were used to produce surfaces of these variables. The elevation was used as the predictor variable in the latter method. The accuracy of developed models was evaluated by comparing true vs. estimated values, as produced by the crossvalidation procedure. As was expected, supporting the calculation by elevation data reduced the amount of uncertainty in the models and errors of both variables were lower in this case. The final maps are given in the paper as well.

Keywords: geostatistics, GIS, ISATIS, climate, external drift kriging, ordinary kriging

1 Introduction

Geographic information systems along with data modelling techniques originating in many fields of science play a crucial rule in modelling of natural phenomena. Recently users can find tools for time series analysis, hydrological modelling techniques, multidimensional statistical methods for classification/regression of spatial data, as well as comprehensive geostatistical methods integrated in most of commercial or freely available GIS. On the other side, professional managing of complex problems often requires much advanced tools than offer these systems. To fill this demand, there are several softwares narrowly specialized in a given topic. The example of open source solution is package SWIM (*Soil and Water Integrated Model*) for hydrological modelling integrated in GRASS. Geostatistical software ISATIS or products such as ModFlow and GMS stand for commercial ones. However, these often lack database capabilities and tools for final map making, thus their combination with standard GIS is often necessary.

In the paper, we paid attention to geostatistical modelling of selected climate parameters in the frame of Slovakia using ISATIS capabilities. General system description and modelling workflow are described. Practically, we focus on the modelling of July average air temperature and July precipitation totals, both per period of 1951-1980. The two approaches to modelling were employed – ordinary kriging and external drift kriging. The effectiveness of both methods was evaluated by comparing true and estimated values produced by crossvalidation procedure. The reason for using external drift kriging was the structure of source data exactly matching the nature of this method, as will be described below. The ordinary kriging was used as standard method for univariate prediction. This might be also used as a comparative base to multivariate prediction of external drift kriging. Particular steps of variographic analysis and kriging as well as final map products are introduced. All the modelling has been carried out within the frame of project supported by the Ministry of Agriculture of the Slovak Republic „Global Climate Change Impacts on the Forest of Slovakia". Since only a small part of produced results will be given in the paper, the web page www.climate.sk was launched to make the results available for a broader public. Produced grids, charts and animations can be found there.

2 ISATIS – description of selected features

ISATIS belongs to cutting edge products in the field of geostatistics, however not widely used by the GIS community. The reasons of this are mainly its narrow specialisation on geostatistics (although it comprises also some standard GIS tools), demands on strong theoretical background of users in specialized algorithms, and user interface not allowing its intuitive use. Besides, „standard" GIS products often offer basic tools of geostatistics, thus it is meant for advanced users mainly. ISATIS allows us working with 2 and 3 dimensional grids, integrating various data sources and importing/exporting from/to a spectrum of formats. Mining, oil & gas and environmental applications are of the primary interest.
Strong feature of ISATIS interface is a dynamic graphical linkage between the windows – this means, once a user opens a map of spatial distribution of source data, related histogram, correlation chart and variogram, highlighting/masking a value in one window changes the content of all the remaining windows and all the statistics are being dynamically recalculated. This allows very effective look into the data structure, identifying outliers, checking for extreme values, etc. ISATIS provides comprehensive tools for variographic analysis, with many interactive features. A spectrum of theoretical variograms might be fitted to empirical data, using manual interactive methods, as well as automated sill fitting procedures. Zonal and geometrical anisotropy might be modeled flexibly. As far as kriging and other interpolation methods are concerned, two sets of tools are available. *Interpolation* tool contains a set of interpolators not requiring a variogram defined by a user. Except distance weighted methods (inverse distances, linear kriging model), also least square approaches (trend surface modelling) or discrete splines are available. *Prediction* tool contains simple, ordinary and universal (co)kriging, kriging with variance measurements errors, external drift kriging, collocated cokriging, indicator kriging and disjunction kriging. A specific tool for modelling indicator variogram is available as well. Besides, strong simulation capabilities might be found in the system. These belong to less known fields of geostatistics, although with high importance for the environmental modelling. ISATIS offers both conditional and non-conditional

simulations, specifically Gaussian Sequential Simulation, Turning Bands Simulation, Plurigaussian Simulation, Boolean Simulations and others. Interactive simulation postprocessing tool might be found as well. The latest release also offers effective interface for visualisation of 3-D data.

3 Data description and organization

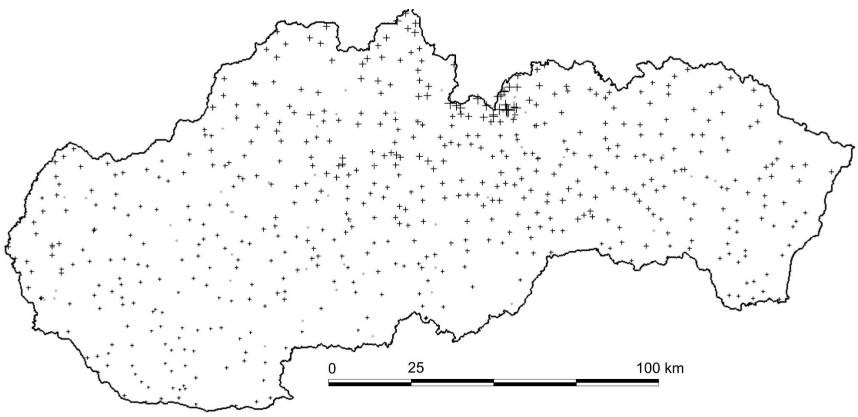
As stated, July precipitation totals and July average air temperature (temperature and precipitation thereinafter) per period of 1951-1980 are of the interest. Primary data were gathered at meteorological stations distributed over the entire Slovakia. As many of climate parameters exhibits strong correlation with orography, elevation of Slovakia (digital elevation model with 30 meters resolution), was used as predictor (auxiliary) variable. Spatial distribution of meteorological stations of both target variables might be seen in the Fig. 1. Descriptive statistics are given in the Table 1 below.

Tab. 1. Descriptive statistics of temperature, precipitation and elevation data used for the modelling

Temperature	n	Mean	St. Dev.	Skewness	Range	Min	Max
(°C)	170	17.2	2.3	-1.29	13.8	6.9	20.7

Precipitation	n	Mean	St. Dev.	Skewness	Range	Min	Max
(mm / month)	552	92	31.2	2.4	231	50	281

DEM	n	Mean	St. Dev.	Skewness	Range	Min	Max
(m. a. s. l.)	grid	448.7	309.1	1.22	2545	94	2654



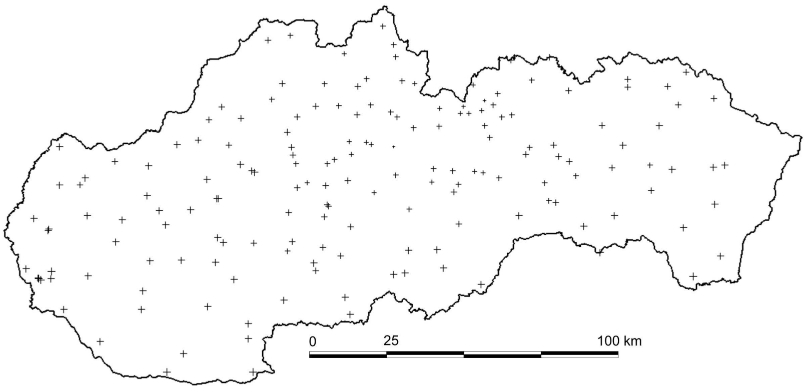


Fig. 1. Distribution of precipitation (up) and temperature (bottom) data

All the source data are stored in ArcInfo geodatabase (MDB), along with other comprehensive records on climate characteristics of Slovakia. A complex set of geographical features (networks, borders, elevation, etc.) is stored in the database, as well. The point-distributed data (meteorological stations) were imported into the ISATIS as Excel spreadsheet, having two columns dedicated for latitude/longitude records. DEM data were imported as ArcInfo grids (ASC). Besides, the border of Slovakia was used to restrict all the calculation only into its frame.

In ISATIS, the data are organized in so-called *Studies*, each of them having set the parameters on working units, graphical display of data, memory allocation issues, etc. Each study contains several *Directories* (in a standard sense). *Files* are located in the *Directories*, each of them comprising a set of *Variables*. To exemplify, a *File* might be the grid of Slovakia, while the *Variables* are data values associated to each grid cell (elevation, temperature, etc.) (Fig. 2).

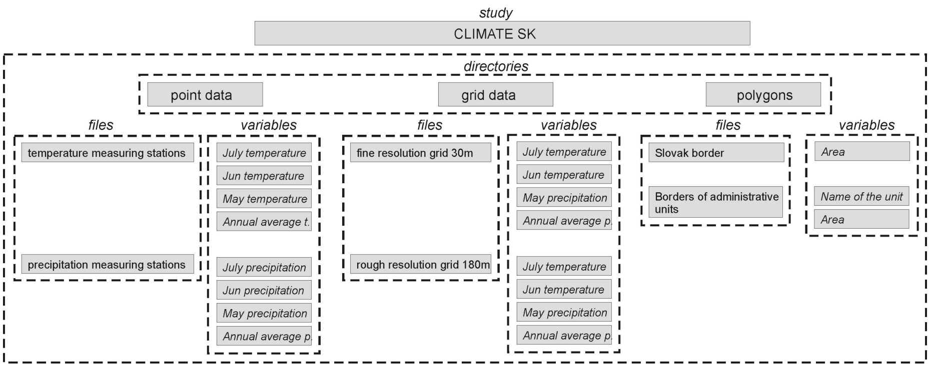


Fig. 2. Data organisation used in this paper. Only the selected *Variables* are given.

The same organisation pays for point distributed data – a *File* expresses the location of meteorological stations, while *Variables* comprise all the data measured on. Thus one physical structure has attached multiple attributes. A specific kind of variable is a *Mask Variable*, several of which might be associated to each *File*. This contains information restricting the calculations – by geographical definition of certain area, giving a threshold up to which or beyond which the data are not considered, etc. In the paper, the border of Slovakia was used as *Mask Variable*, to restrict all the calculations within its extent. Several other specific file types are recognized as well, however not used in this paper.

4 Methods

Description of theoretical background of geostatistical modelling goes beyond the scope of this paper. The books by Goovaerts [2], Isaaks and Srivastava [5] or Wackernagel [8] provide a fundamental information on this topic. In Slovak, the paper on variogram modelling was elaborated by Hlásny [3]. Furthermore we describe in brief, leaving aside mathematical notations, the essential methods employed in the paper.

Data transformation

As might be seen in the Table 1, temperature and precipitation data are strongly skewed. As ordinary kriging requires the normality of data, these must be transformed. Keeping the transformation parameters, the back transformation might be done after kriging (or other analysis) straightforwardly. ISATIS facilitates so called Gaussian Anamorphoses modelling (Normal Score Transformation), which enables transforming to normal and back transforming any kind of frequency distribution. Methodologically, this firstly ranks the data and than assigns to every *i*th observation the value of *i*th percentile of standardized normal distribution N(0,1). Thus the percentiles express the transformed values of the observation data. The back transformation is carried out as linear interpolation using the reference points provided by initial transformation of the data. Only the histogram tails require a specific treatment, e.g. using Hermite Polynomials. More detailed discussion on this topic might be found in Olea [6] or Rivoirard [7].

Variogram modelling

Variogram modelling is a fundamental component of geostatistics. It focuses on the evaluation how spatial variability develops by specific distances and direction. The variogram is recognized as a key quantity of this approach. It is constructed by plotting mean squared differences of data values separated by a range of distances against separation distance. Produced chart is fitted by a function, in order to produce a variogram model, which is a key component of any kriging system. Detailed discussion on this topic was elaborated for example by Clark [1].

Ordinary kriging (OK)

Ordinary kriging is perhaps the best recognized of kriging methods, also because it is implemented in many GIS, often as default option. OK uses the results of variographic analysis (variogram model) to compute the weights for observed data, in order to predict values at unknown locations. Important features are the minimisation of so-called error variance and its property of unbiased estimator, in the sense the systematic error of resultant model (bias) is equal to zero. To the contrary to standard distance weighted methods (IDW and its variants), it considers, except distances between estimated location and known data values, also (statistical) distances between known data values themselves. This information is used for example to solve the redundancy in data and related screening effect (see e.g. [8]). The references on this procedure might be found in the literature above. In Slovak, Hlásny [4] discusses some of these points.

Kriging with External Drift (KED)

Kriging with external drift allows us combining point distributed target variable with one or more grid distributed predictor variables, under the assumption these are linearly correlated. The produced surface takes information after predictor variable, depending on the intensity of correlation. To the contrary to simple spatial regression, the produced surface passes through all the values of target variable. KED uses the results of variographic analyses in the same manner as OK. For further information see for example [2] or [8].

Acuraccy assesment

Accuracy assessment has been carried out using selected results of crossvalidation procedure. The methods might be used for assessing the accuracy of models produced by kriging (or other method) and for checking the compatibility between source data and variogram model. Technically, the target variable is temporarily discarded at one data point and prediction is calculated for it. This is repeated for all data points. The estimated values are then compared to true values, in order to estimate mean error and variance of errors. The crossplot of true vs. estimated values and residuals histogram are often used for the visualization (Fig. 6).

Standardized forms of these quantities can be obtained by dividing them by theoretical (kriging) variance. Mean (standardized) error is the measure of unbiasedness of the prediction. The ratio of experimental and theoretical variance (variance of standardized errors), says about the suitability of variogram model – optimally this should be 1. If it differs significantly, this quantity might be used to adjust the sill of variogram model. This part was not discussed in the paper.

5 Results

5.1 Variographic analysis of temperature and precipitation data

The variograms of temperature and precipitation data might be seen in the Fig. 3. As the variances were calculated for distances up to more than 100 km, their increase indicates strong elevation driven drift in the data. However, “local” stationarity might be observed within the shorter distances, which were used to fit the variogram. No anisotropy was observed within this range. Spherical variogram with nugget effect was used for both variables. Automatic sill fitting procedure was used to set the sill value, taking into account a number of pairs used to calculate the variogram values. Description of variogram models is given in the Table 2 below.

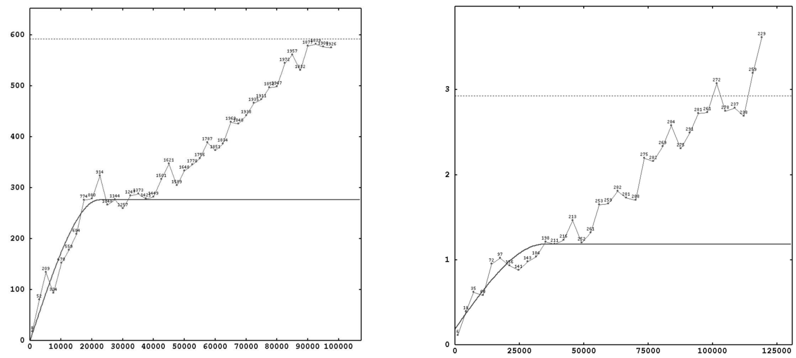


Fig. 3. Empirical and model variograms of precipitation (left) and air temperature (right) data.

Tab. 2. Parameters of model variograms

Temperature	Range	Sill	Nugget	Nugget (proportional to sill)
	33 994 m	1.7	0.08	4.7 %
Precipitation	Range	Sill	Nugget	Nugget (proportional to sill)
	22 869 m	276	0.138	0.05 %

5.2 Ordinary Kriging and External Drift Kriging of temperature and precipitation data

Standard ISATIS (Co)kriging module was used to carry out both ordinary and external drift kriging. The task to be solved in this step is a proper design of the neighbourhood. ISATIS facilitates flexible definition of mowing neighbourhoods, inclusive division of space around an active point to sectors, proposing minimal/optimal number of sample points per sector, setting minimum distance between the samples, maximum number of consecutively empty sectors or maximum distance without any sample. In fact, proposing a proper neighbourhood definition is a tricky task, often based on the crossvalidation procedure. To produce the surfaces of temperature and precipitation data, the space around each predicted location was divided into six angular sectors, asking for at least 15 points to be equally distributed in this space. The maximum search distance was set equal to respective variogram range. The produced maps might be seen in the figures below.

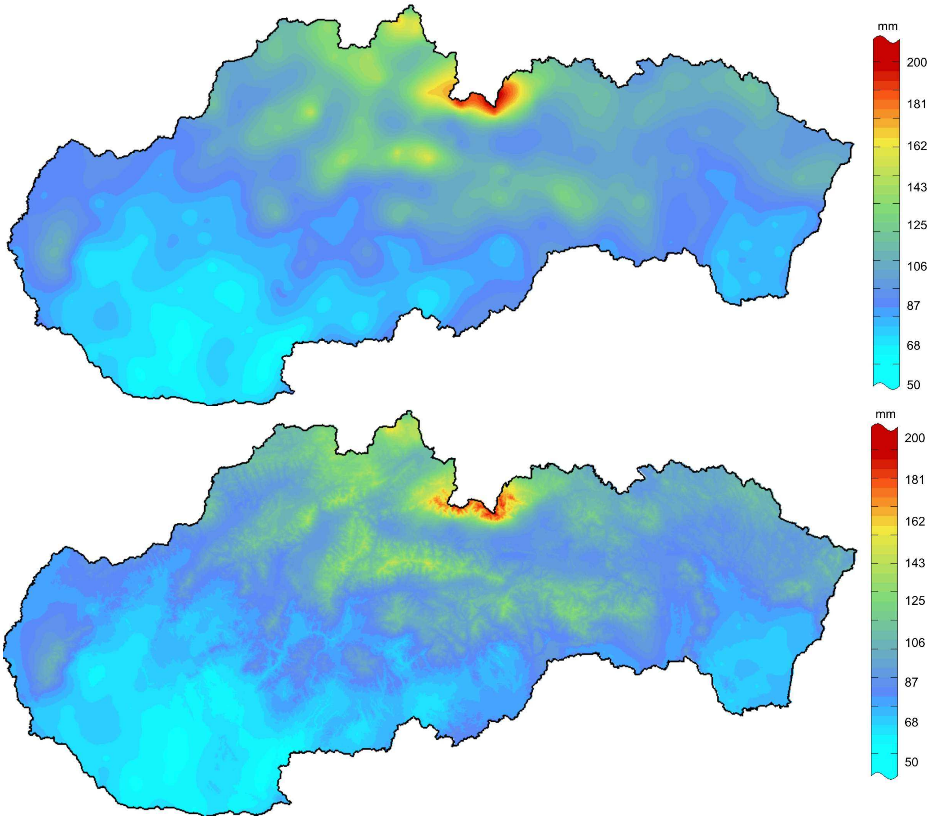


Fig. 4. Maps of July precipitation totals produced by means of ordinary kriging (top) and external drift kriging (bottom).

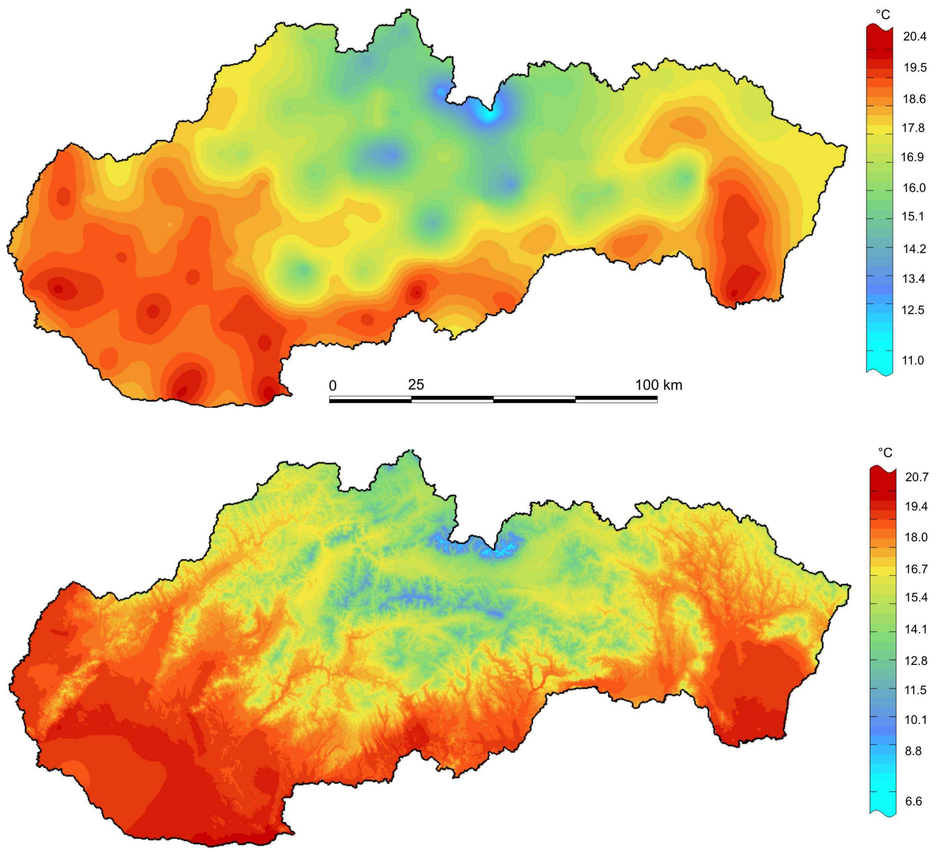


Fig. 5. Maps of July average temperature produced by means of ordinary kriging (top) and external drift kriging (bottom).

5.3 Accuracy assessment

Differences between true and estimated values were used to analyse the accuracy of produced models. As might be seen in the charts below, the residuals have zero approaching mean (the result is unbiased) and more-less normal distribution. The distribution of values around the regression line also indicates relatively high accuracy of the prediction. If we compare the variances of the residuals produced by OK and KED, the latter method provided more accurate results in both cases. Further discussion this topic is given in the Summary and Conclusion.

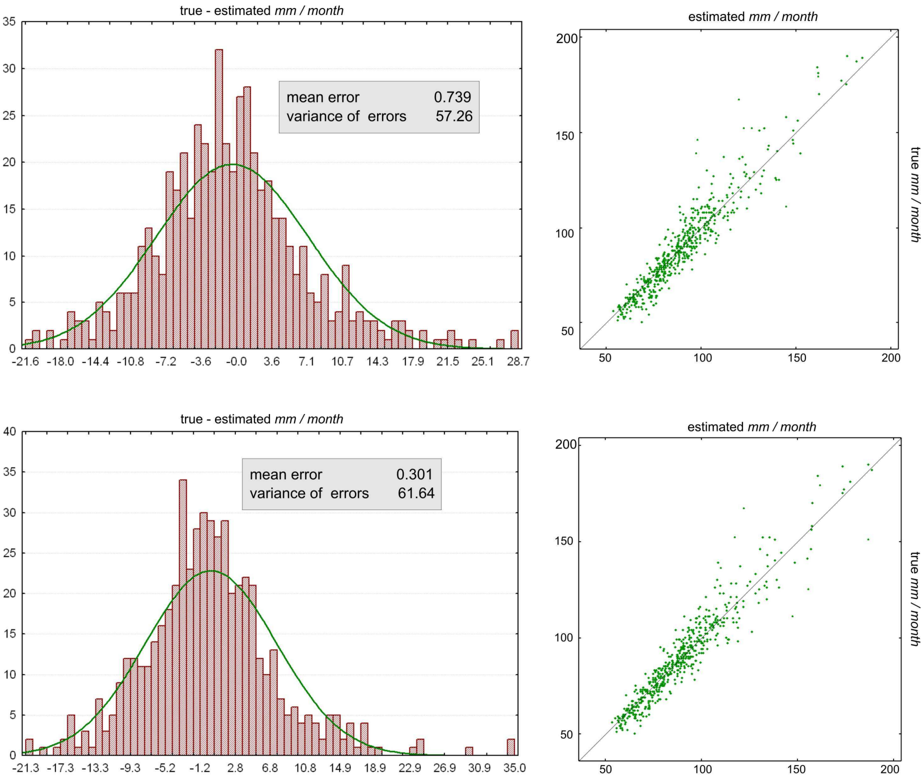
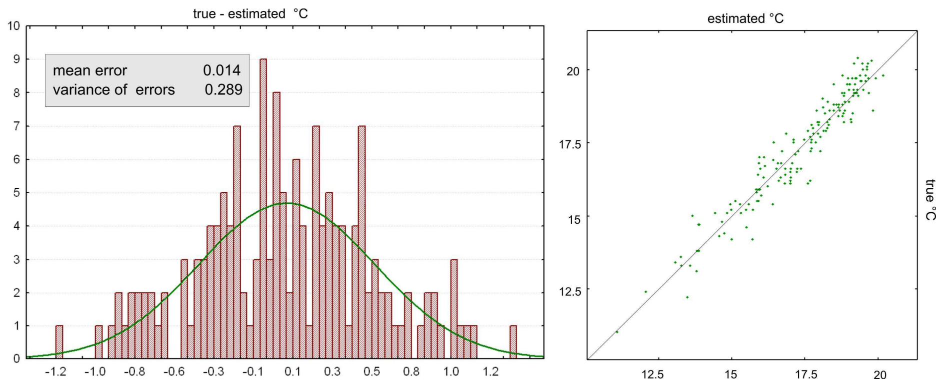


Fig. 6. Frequency distribution of true – estimated values and crossplot of true vs. estimated values produced by External Drift Kriging (up) and by Ordinary Kriging (bottom) of precipitation data.



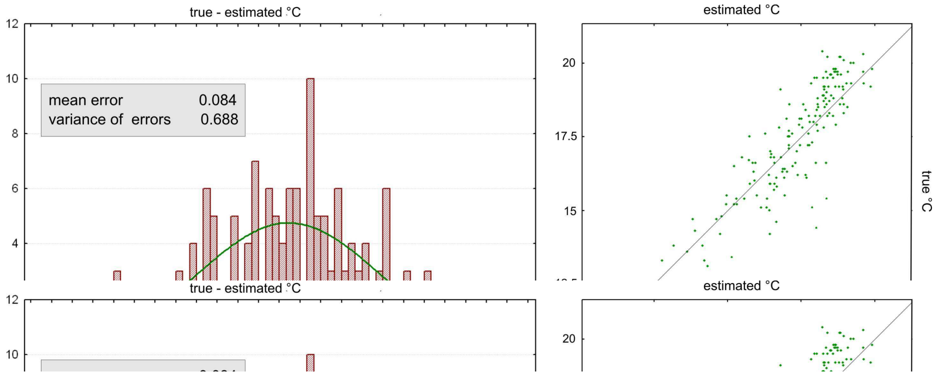


Fig. 6. Frequency distribution of true – estimated values and crossplot of true vs. estimated values produced by External Drift Kriging (up) and by Ordinary Kriging (bottom) of temperature data.

6 Summary and conclusion

In the paper, we introduced the approach to modelling July average air temperature and July average precipitation totals by means of ISATIS capabilities. Ordinary kriging and external drift kriging were used to produce the surfaces of these variables. As we expected, accuracy analyses based on the comparision of true vs. estimated values, proved higher accuracy of models produced by the latter method. The use of elevation, as a predictor variable strongly correlated with both target variables, reduced the uncertainty of produced models. However, its impact significantly differed between temperature and precipitation data. While in the case of temperature data standard deviation of residuals decreased from 0.82 to 0.59 °C (28% increase of accuracy), in the case of precipitation data only the 5.21% improvement has been noticed. This is mainly because the correlation of July temperature and elevation is much stronger ($r = 0.97$) than the case of precipitation ($r = 0.82$). Thus the elevation brought much important piece of information into temperature data. In fact, the estimated errors reflect the “game” between numbers of sample points, correlation of target variable with predictor variable as well as the variograms structure.

Introduced approach might be effectively used for modelling of many other climatic parameters – numbers of days the temperature of which exceeds certain threshold, numbers of days with snow cover, occurrences of various extremes, etc. Produced models might be effectively used in practical management of natural resources, as well as in further research activities. Mainly global climate change related research, as recent „hot“ topic, benefits from these results and several case studies have already been carried out. In addition, developed models of essential climatic parameters allows us modelling much complex climatic phenomena – evapotranspiration, definition of bioclimatic zones of plant and animal species, evaluating their changes due to various factors (natural or human interventions), etc. These topics are a subject of further research and will be reported in next papers.

References

1. Clark, I., Harper, W. V. *Practical Geostatistics 2000*. Ecosse North Amer Llc., 2000. ISBN: 0970331703
2. Goovaerts, P. *Geostatistics for natural resources evaluation.* Oxford University Press, 1997, New York. ISBN 9780195115383
3. Hlásny, T. Geoštatistický koncept priestorovej závislosti pre geografické aplikácie. *Geografický časopis 54, 2, pp. 97-116.* Bratislava, 2005. ISSN 1335-1257
4. Hlásny, T. GIS v lesníckom výskume, *In: M. Koreň (Ed.) Aplikácie GIS v lesníctve, pp. 83-126.* Technická univerzita vo Zvolene, 2006, Zvolen. in press
5. Isaaks, H. E., Srivastava, R. M. *Introduction to Applied Geostatistics.* Oxford University Press, 1989, New York. ISBN 9780195050134
6. Olea, R. A. *Geostatistics for Engineers & Earth Scientists.* Kluwer Publishers, 1999, Boston. ISBN 0792385233
7. Rivoirard, J. Introduction to Disjunctive Kriging and Non-linear Geostatistics., Clarendon, 1994, Oxford. ISBN 0198741804
8. Wackernagel, H. *Multivariate Geostatistics.* Springer – Verlag, 1998, Berlin. ISBN 354064721X