

Annual Mean Air Temperature and Precipitation Layers for the Czech Republic Created in GIS and Visualized Dynamically on the Web

MULLER A.

Czech Technical University in Prague, PRAHA 6 - DEJVICE, CZECH REPUBLIC

BACKGROUND AND OBJECTIVES

This study presents mapping of two climatological factors: mean air temperature (MT) and precipitation (P) over the area of the Czech Republic (CR) using spatial interpolation of climatological data. Spatial modeling of climate is of wide interest since many other environmental variables depend on climate. Climate maps are needed in many disciplines related to Earth Sciences, for example in hydrology, forest management, agriculture, ecology, urban environments, energy, or climate change.

The research topic of this study was specified during author's exchange studies at the Kansas State University, USA in 2008 and the University of Adelaide, Australia in 2010. The study became part of author's final thesis [4] defended at the Department of Mapping and Cartography at the Czech Technical University in Prague in June, 2010.

The primary objective of this study was to explore the utility of regression analysis, which allows the studying of topographic influence on spatial climate patterns. The sub-objective derived from the interpolation technique applied, was to investigate and select meaningful geographic or topographic factors, which play an important role in annual MT and P modeling in the area of CR. The last objective was to publish the results in an interactive manner on the Internet using and open source technologies.

APPROACH AND METHODS

Geographic Information System ArcGIS provided the means to produce climatological surfaces for the entire CR between the years 1998–2009. The temporal resolution of the resulting climatological layers is annual and the spatial resolution is 90 m. The main limitation of this study was a sparse density of meteorological stations. Climatological data – annual mean air temperatures and precipitation values were obtained for only 22 meteorological stations available on-line at the Czech Hydrometeorological Institute (CHMI) websites. Only 5 out of 22 stations lie higher than 500 m above the sea level and the area per station is approximately 3.6 km². Spatially denser data from 132 stations (including those 22 stations discussed earlier) – were obtained only for one year and were therefore used for validation.

Accurate climatological data are collected at meteorological stations, which are discrete point locations in space. Values at any other point must be derived from neighboring stations or from relationships with other variables. In complex terrain, climatic patterns are delineated by topographic slopes and barriers [3]. The underlying hypothesis is that climate at any location is influenced by the environmental attributes of the surroundings. The geographical variables of the weather station (typically coordinates - latitude and longitude) or topographic variables (e.g. elevation, aspect, and slope) are used as independent variables. The geographic or topographic information often comes from digital elevation model (DEM).

DEM used in this work (fig. 1) originates from the NASA Shuttle Radar Topography Mission (SRTM). For latitudes of central Europe each pixel represents a rectangle of 60 × 90 m. The vertical accuracy of the DEM characterized by an RMSE (calculated from differences in real altitudes of 132 meteorological stations and altitudes derived from the DEM) is 7 m. Several variables which appear throughout the scientific literature to be significant predictors of climate were derived from the DEM in ArcGIS: altitude, slope, aspect, curvature, maximum altitude in a wedge (Z_x) and UTM coordinates X and Y, representing longitude and latitude.

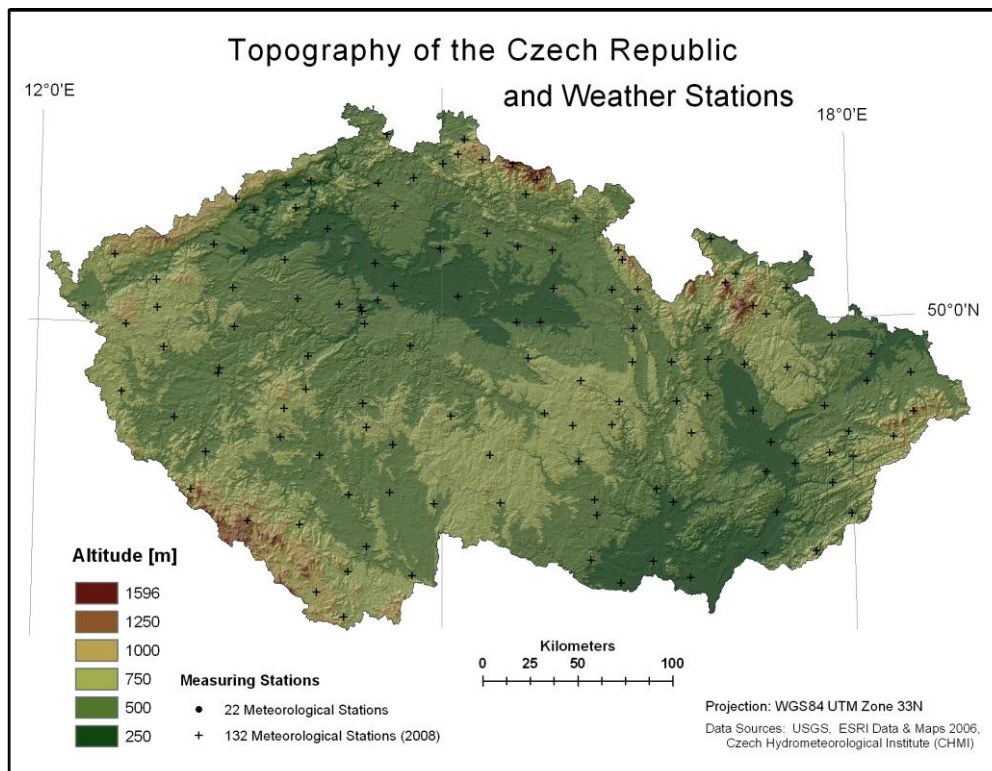


Fig. 1: Study Area with the hypsometry derived from SRTM DEM, black dots represent locations of the 22 meteorological stations, and plus signs represent locations of the 132 stations

The method used – multivariate regression approach combined with a local interpolation of residuals follows the work of Ninyerola et al., who have developed a methodology combining global and local interpolators with the advantage of using terrain variables [6-8]. Combined methods generally use a global method for prediction and a local method for interpolation of residuals. The later method is in fact a correction used to obtain measured values at the particular station [10]. This combined approach based on regression analysis is scientifically interesting because, in addition to the interpolation process, it gives information about the relationship between the geographic reality of the land and climate. Common interpolation techniques such as IDW, splines or kriging are not suitable for such a sparse density of measured locations [5].

The linear multivariate regression can be described by the following function:

$$z(x) = \beta_0 + \beta_1 P_1 + \beta_2 P_2 + \dots + \beta_n P_n$$

where z is the predicted value at point x , β_0 - β_n are the regression coefficients and P_1 - P_n are the values of independent variables at point x . Regression as a global method is an inexact interpolator, since the predicted value does not equal the value recorded at the station. The known error or difference between the two values is a residual:

$$\text{residual} = \text{observed data} - \text{predicted data}$$

Multivariate regression models were calculated in GeoDa software, which is able to read SHP files. Backward stepwise approach was used for selecting significant independent variables. Backward stepwise selection begins with the examination of the combined effect of all of the independent variables on the dependent variable. One by one, independent variables (starting with the weakest predictor) are removed based on some outset criteria (t-statistics), and a new analysis is performed.

RESULTS

Based on the backward selection process, final regression models look as following:

- Regression model for MT with two independent variables:

$$MT = \beta_0 + \beta_1 \text{ALTITUDE} + \beta_2 Y \quad (1)$$

- Regression model for P containing four independent variables:

$$P = \beta_0 + \beta_1 \text{ALTITUDE}_{1\text{km}} + \beta_2 Y + \beta_3 X + \beta_4 Z_{\text{xE25}} \quad (2)$$

Significant independent variables for predicting MT were found altitude and latitude. Four out of 26 independent variables studied (some variables like altitude or curvature were resampled to 1 or 5 km and Z_x variable was defined in four directions and within four radii of 2.5, 5, 10, and 25 km) were proven to be

significant for P: altitude resampled to 1 km, longitude, latitude and the variable ZxE25 depicting topographical barriers in wedge of 45° facing the East in the radius of 25 km.

Putting the equations above into the Raster Calculator in ArcGIS ‘potential maps’ [8] were calculated. In order to have the measured values at meteorological stations, residual values were interpolated into a raster image called ‘corrector map’ using IDW interpolation. For both MT and P, IDW with the power of 2 and the maximum of 10 possible neighboring points was used to calculate the value at each cell. The power of 2 was recommended by Ninyerola [6], because the higher the power the less influence from distant points. The number of nearest input sample points was tested for 3, 5, 8, 10 and 12 points and the number 10 was selected arbitrary according to visual consideration of corrector maps. Corrector maps (example in fig. 2) can be seen as anomaly maps, of great interest to reveal the singularities of the climate at the local scale [8].

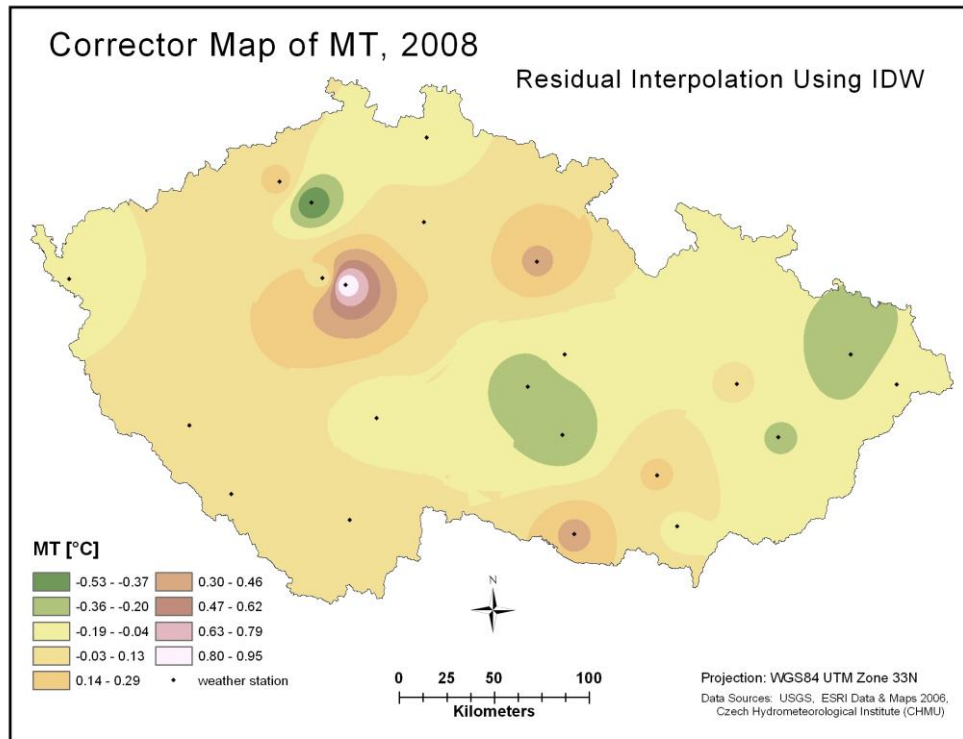


Fig. 2: Example of a ‘corrector map’ created by interpolating residuals of MT

By adding the ‘corrector map’ to the ‘potential map’ final climate surface is created. Once the final climate surface is created, low-pass filter (focal mean of 3 surrounding cells in case of MT and 10 surrounding cells in case of P) is then applied to smooth the surface, particularly the edges of class boundaries, and to get rid of ‘lonely’ cells. Below in fig. 3 and 4 are two examples of final thematic maps.

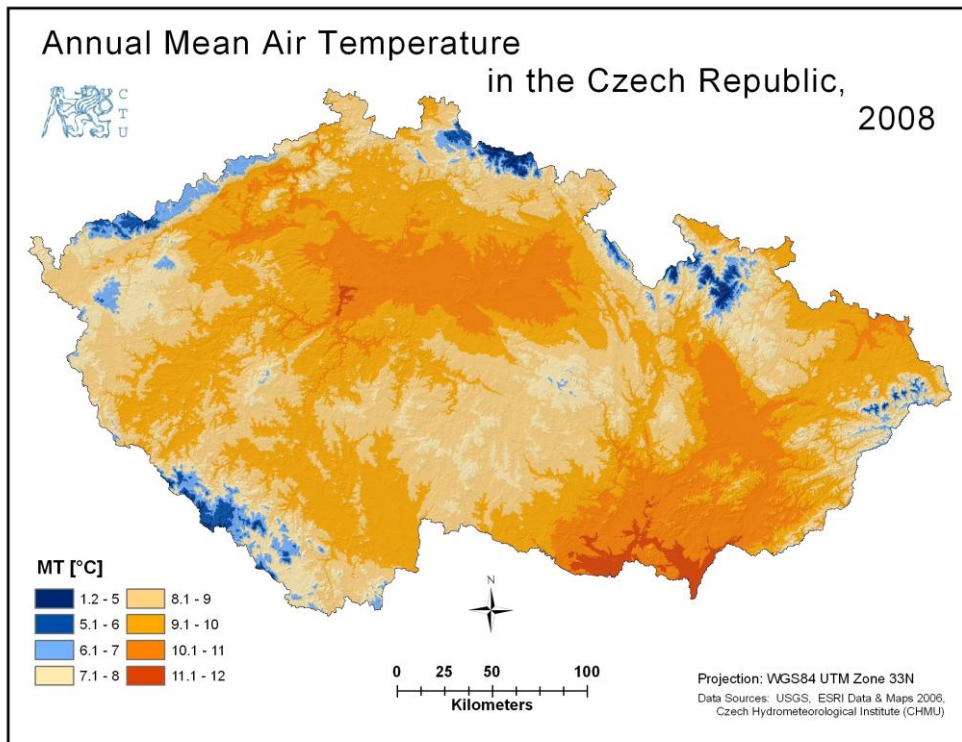


Fig. 3: One example of final MT map

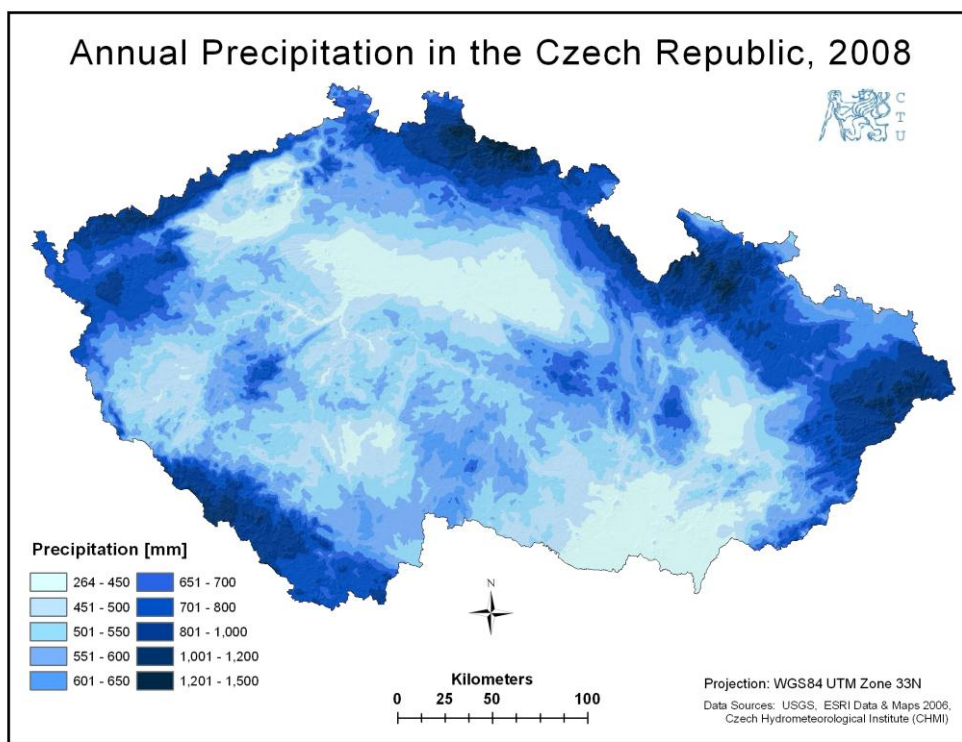


Fig. 4: One example of final P map

Spatial pattern of MT copies the topography. Since temperature is highly dependent on altitude, the warmest areas are river valleys in central Bohemia and south Moravia, while the coldest areas are situated along the borders in the mountains. There are only two focus points around Prague and in southern Moravia where annual MT reaches over 11 °C (only in the years 2000, 2007, and 2008). The spatial pattern of P is similar to the MT. The rainiest places are along the borders in the mountains, while the driest areas lie in the mountain shadow in the west Bohemia and in the south Moravia.

Annual climatological maps were compared to the long-term 30-year normal (period 1961–1990). Long-term normal layers were interpolated using the same methodology, the same number of stations, and the same variables. Interpolated 30-years normal values were subtracted, cell by cell, from annual rasters in order to create anomaly (deviation from normal) maps. Below are two anomaly maps, which were calculated as a mean from all annual anomaly maps. Fig. 5 implies that in the last 12 years MT was higher than the 30-year normal, overall by 1.0 °C. Last 12 years have also been slightly wetter, in average by 40 mm (calculated as a mean value of the layer in fig. 6).

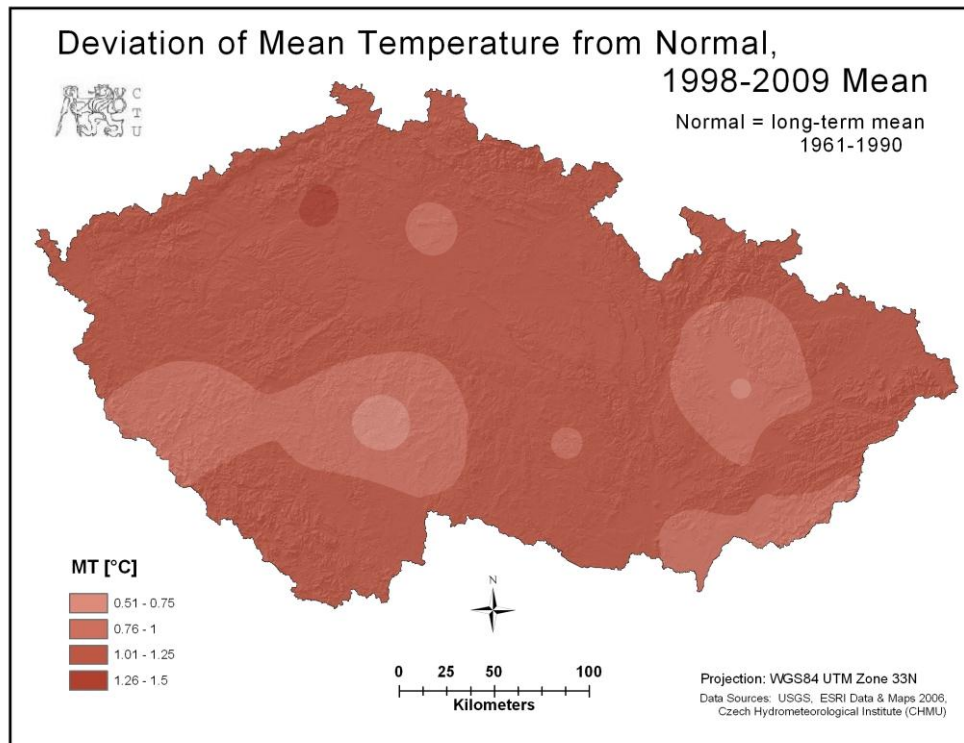


Fig. 5: 12-year mean of deviation from normal of MT

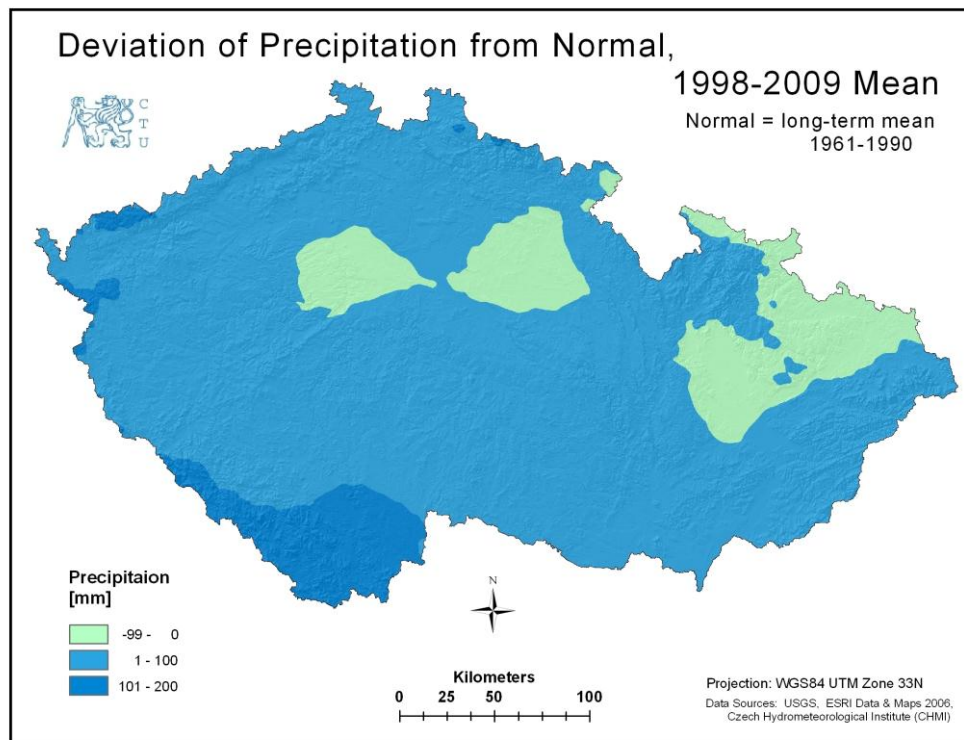


Figure 6: 12-year mean of deviation from normal of P

ACCURACY

When there are many observations available, typically 60% of them are used for estimating the regression model and the remaining 40%, often randomly selected from the whole dataset, are used for independent validation. If the model is considered reliable and the results are satisfying, the remaining 40% of observations often enter the regression as well to improve the model by including all of the stations.

The case of this study is somehow the opposite. Since there is only a small subset of observations available (22 stations) for each year, the denser dataset with 110 observations (132 minus 22 stations which are the same as in the first dataset) is used for validation of the one year (2008). So there are only 18% of observations used in prediction and 72% for validation. This way the interpolated climatological values of only one year were independently validated, nevertheless this validation can provide a good measure of the overall performance of both regression models.

The validation of the 2008 layers (in tab. 1) expressed by RMSE (standard deviation) is 0.5 °C for MT. In the case of P, the accuracy of the 2008 layer given by RMSE is approximately 90 mm.

Table 1: Validation of MT and P for the year 2008

MT2008		P2008	
MBE =	-0.20 °C	MBE =	-1.8 mm
MAE =	0.34 °C	MAE =	67.4 mm
RMSE =	0.45 °C	RMSE =	88.7 mm
MAX =	1.55 °C	MAX =	280.2 mm
MED =	0.25 °C	MED =	53.1 mm

Table 2.: Statistical Criteria

Definitions	<p>N... number of observations</p> <p>F... degrees of freedom</p> <p>O... observed value</p> <p>\bar{O}... mean of observed values</p> <p>P... predicted value</p>
Mean bias error (MBE)	$MBE = \frac{1}{N} \sum_{i=1}^N (P_i - O_i)$
Mean absolute error (MAE)	$MAE = \frac{1}{N} \sum_{i=1}^N P_i - O_i $
Root mean square error (RMSE)	$RMSE = \sqrt{\frac{1}{F} \sum_{i=1}^N (P_i - O_i)^2}$

The overall accuracy of the final climate maps is characterized by the root mean square (RMS), calculated as a quadratic mean of RMSE of the regression models built in GeoDa (shown in tab. 3).

Table 3: R^2 , standard deviations from protocols of regression & quadratic mean (RMS)

Year	R ²		RMSE	
	MT	P	MT [°C]	P [mm]
1998	0.95	0.90	0.40	94.5
1999	0.96	0.92	0.37	72.3
2000	0.97	0.86	0.33	109.6
2001	0.96	0.82	0.36	141.2
2002	0.96	0.78	0.36	145.4
2003	0.95	0.87	0.41	72.7
2004	0.95	0.92	0.41	70.4
2005	0.93	0.89	0.49	102.4
2006	0.90	0.85	0.52	105.4
2007	0.97	0.85	0.34	105.8
2008	0.97	0.82	0.34	101.0
2009	0.95	0.83	0.40	120.5
RMS =			0.40	106.1

The fit of the MT model is excellent (R^2 between 0.90 and 0.97), while the fit of P model is better than expected (R^2 between 0.78 and 0.92). Altitude and latitude can explain more than 90% of the spatial variability of MT. P is in general more difficult to predict and therefore more variables are needed in order to explain more than 78% of the variability of P. The accuracy characterized by RMS is 0.4 °C for MT and 106 mm for P.

A 'more accurate' map of P was interpolated for the year 2008, using all of the 136 observations available, and same method as above (same topographical and locational variables described in equation 2). The resulting surface was, cell by cell, subtracted from the previous one interpolated from only 22 observations in order to uncover areas with significant deviations to the model. The subtracted surface, shown in fig. 7, reveals red (drier) areas where the model over predicts P and blue (wetter) areas where the model under predicts P. The red areas, where the model should predict more precipitation, spread out mostly along the borders and in central Bohemia and east of Prague. This analysis was done only for P, since MT model predicts well.

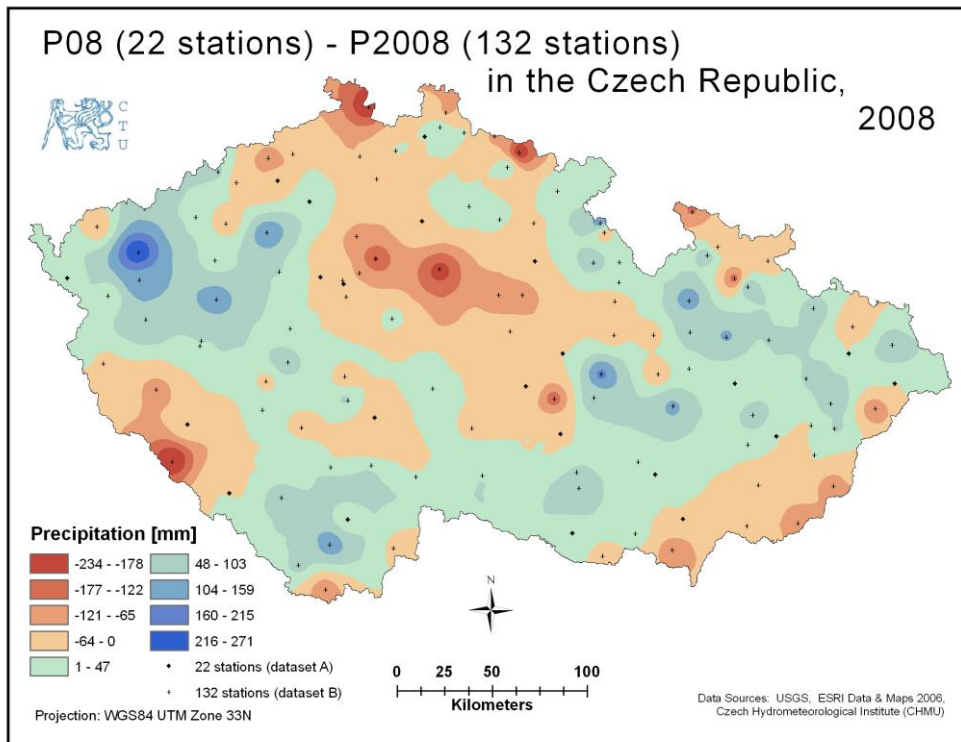


Fig. 7: Deviations to the model of precipitation

A visual comparison with a map from a different source can serve a cartographer as an independent assessment of correctness as well. Spatial patterns of long-term MT and P were compared with large scale maps published by CHMI in the Climate Atlas of Czechia. The confrontation confirmed that the combined regression method yields excellent results of MT as well as satisfying spatial distribution of P.

VISUALIZATION AND WEB APPLICATION

Final rasters containing climatological information were published on the Internet, firstly as animated images and secondly, in an interactive manner using a MapServer. It was crucial for the animation to maintain the same classes and layout so that the only features changing was the map itself and the year. The same classification is essential as it makes comparison among the twelve years possible. A minor disadvantage is that there are more classes included in the legend than displayed in the map.

Final climatological surfaces are laid over a shaded relief with some transparency (10% for P and 15% for MT) in order to emphasize the climate-topography relationship. Animations were created in the free software Gimp as GIF files with the rate of 3s. Firstly, a web page following the W3C standards for XHTML 1.0 was created displaying animations of both final climatological surfaces and deviations next to each other. In this layout, one can not only see the final climatological surface, but can also follow the deviation from the long-term normal at the same time.

Secondly, a map server was set up in order to display climatological layers dynamically. A map server is operated by text parameters. Map servers are running 'under' a web server such as Apache, which makes a request by handing over the parameters. Map server uses information passed in the request and a mapfile to create an image of the requested map.

A variety of map servers exist. The commercial branch is certainly represented by ArcGIS Server from ESRI, or TopoL Internet Server and T-MapServer from Czech firms, while the open source branch is quickly approaching the commercial one. The most known open source map servers are GeoServer and the UMN MapServer from the University of Minnesota. The UMN MapServer had been running at maps.fsv.cvut.cz at CTU and was therefore used also for this project. The purpose of the MapServer is to dynamically display spatial data over the internet. But MapServer alone does not provide the high level of interactivity, like pre-rendering, caching of tile images, etc.

The interactive interface can be set up through a template. There are a number of open source templates, which can be used and modified in order to provide the user with the high level of interactivity, and without the need of hard programming them from the beginning. The choices considered were ka-Map, p.mapper, and OpenLayers. After exploring existing sample applications, ka-Map was selected by the author since it reflected his requirements.

Ka-Map is an open source java script template for developing highly interactive web-mapping interfaces (fig. 8). The ka-map package is in fact a folder structure consisting of many HTML, PHP and CSS files. The whole structure occupies less than 4MB memory. ka-Map is not being installed, it is possible to start 'out of the box' with minimal configuration. However, layers and their attributes need to be defined in a mapfile. Two separate mapfiles were created, one for MT and second for P layers, which the user can choose from in a drop-down box. Ka-Map supports the usual array of user GIS interface elements such as: continuous panning without reloading the page, zooming to pre-set scales, scalebar, legend, and keymap. One useful tool of the application is the ability for the user to search by the names of cities, towns and villages.

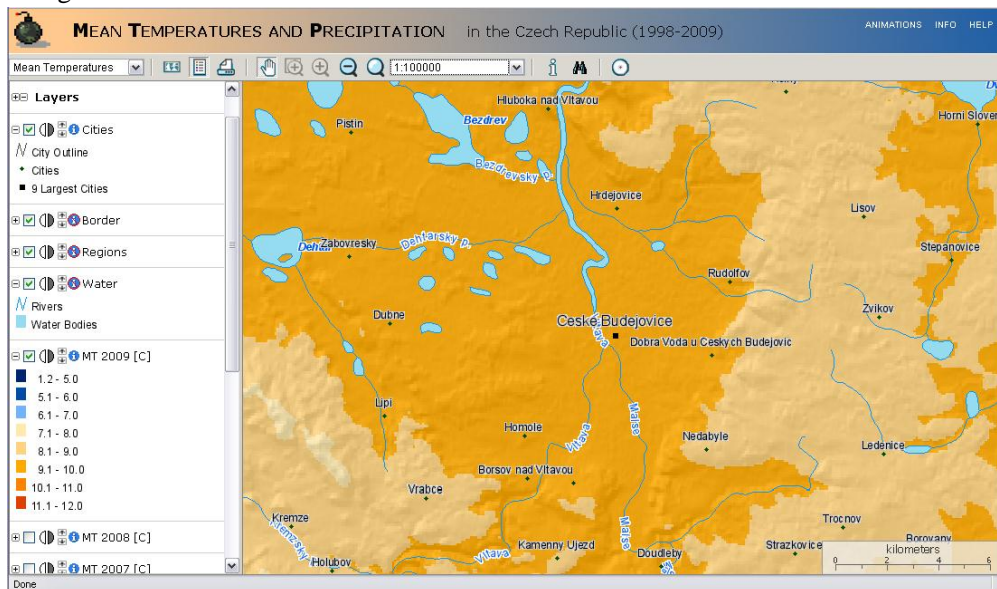


Figure 8: ka-Map Web application

CONCLUSIONS AND FUTURE PLANS

The combined regression approach has proven to be a powerful method for interpolating climatological factors from a sparse density of meteorological stations. The method involves regression using geographical and terrain variables as predictors followed by local interpolation of residuals. 6 predictors (altitude, latitude, longitude, curvature, aspect, solar radiation and slope) were tested for significance in the MT regression model. For precipitation, 26 predictors were tested in the regression analysis, though some were resampled. Meaningful predictors were selected using the backward selection process. Altitude has proven to be overall the best predictor of both climatological variables.

Final regression model of annual mean air temperature enables us to describe 90–97% of spatial variability with the standard deviation $RMSE = 0.4 \text{ }^{\circ}\text{C}$ using altitude and latitude as independent variables. Final precipitation model shows a moderate degree of explained spatial variance 78–92% with four significant terrain variables: altitude, longitude, latitude and the variable $ZxE25$ depicting topographical barriers in a specific easterly direction. The accuracy of final P layers is characterized by the $RMSE = 106 \text{ mm}$. The key success of this study is in obtaining good results with only a subset of meteorological stations by using the topography-climate relationship. Climatological surfaces can be easily updated with up-to-date climatological data. The accuracy and overall performance of the regression models corresponds to the results of similar studies [1, 2, 7, 9]. Visual interpretation of the results as well as the numerical assessment of accuracy approves the methodology, which is especially suitable for mountainous regions with a sparse density of stations.

Annual variation in mean air temperature and precipitation was visualized using animated maps published at <http://maps.fsv.cvut.cz/~muller/>. Animations of resulting climatological maps show the annual change in mean air temperature and rainfall patterns over the past 12 years. Annual MT and P maps were compared to the long-term 30-year mean (normal). Animations of deviations from normal run simultaneously with the final climatological surfaces at the website presented above. Animated maps of deviations of MT from normals show that the last 12 years have been warmer than the 30-year normal, overall by $1 \text{ }^{\circ}\text{C}$. Final climatological rasters were published interactively along with various planimetric layers at <http://maps.fsv.cvut.cz/ka-map/> using open source applications UMN MapServer and ka-Map template.

Obtaining all meteorological stations for the years studied would improve accuracy and would better account for local anomalies in regression models. It would be interesting to examine amplitudes rather than just average values of both climatological variables. Least but not last, monthly data, which has been also available at the CHMI websites, could be interpolated and studied in the future.

ACKNOWLEDGEMENTS

This work has been supported by the CTU grant SGS10/151/OHK1/2T/11. Thanks to Kelly Arbon for her language improvements.

REFERENCES

- [1] BROWN, D.P., COMRIE, A.C. Spatial modeling of winter temperature and precipitation in Arizona and New Mexico, USA. *Climate Research*, 2002, vol. 22, no. 2, p. 115-128. ISSN: 0936-577X.
- [2] CLAPS, P., et al. Spatial distribution of the average air temperatures in Italy: Quantitative analysis. *Journal of Hydrologic Engineering*, 2008, vol. 13, no. 4, p. 242-249. ISSN: 1084-0699.
- [3] DALY, C., et al. A knowledge-based approach to the statistical mapping of climate. *Climate Research*, 2002, vol. 22, no. 2, p. 99-113. ISSN: 0936-577X.
- [4] MULLER, A. *Spatial Modeling of Climate*, in *Faculty of Civil Engineering, Department of Mapping and Cartography*. 2010, The Czech Technical University: Prague. p. 74.
- [5] MULLER, A. *Climate Characteristics of the Czech Republic using ArcGIS*, Department of Mapping and Cartography. 2008, The Czech Technical University: Prague. p. 42.
- [6] NINYEROLA, M., et al. A methodological approach of climatological modelling of air temperature and precipitation through GIS techniques. *International Journal of Climatology*, 2000, vol. 20, no. 14, p. 1823-1841. ISSN: 0899-8418.
- [7] NINYEROLA, M., et al. Monthly precipitation mapping of the Iberian Peninsula using spatial interpolation tools implemented in a Geographic Information System. *Theoretical and Applied Climatology*, 2007, vol. 89, no. 3-4, p. 195-209. ISSN: 0177-798X.
- [8] NINYEROLA, M., et al. Objective air temperature mapping for the Iberian Peninsula using spatial interpolation and GIS. *International Journal of Climatology*, 2007, vol. 27, no. 9, p. 1231-1242. ISSN: 0899-8418.
- [9] SUN, R., et al. A Multivariate Regression Model for Predicting Precipitation in the Daqing Mountains. *Mountain Research and Development*, 2008, vol. 28, no. 3-4, p. 318-325. ISSN: 0276-4741.
- [10] VICENTE-SERRANO, S.M., et al. Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): application to annual precipitation and temperature. *Climate Research*, 2003, vol. 24, no. 2, p. 161-180. ISSN: 0936-577X.