Interpolation of weather generator parameters using GIS (*... and 2 other methods*)

M. Dubrovsky[(1)], D. Semeradova[(2)], L. Metelka[(3)], O. Prosova[(3)], M. Trnka[(2)]

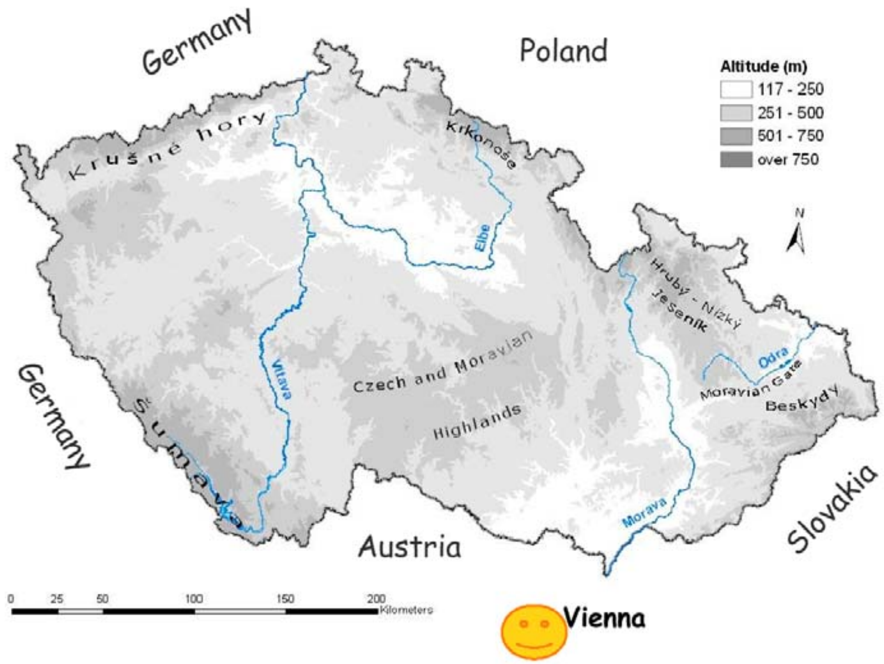
[(1)] Institute of Atmospheric Physics ASCR, Prague, Czechia (dub@ufa.cas.cz),
[(2)] Institute of Landscape Ecology MAFU, Brno, Czechia,
[(3)] Czech Hydrometeorological Institute, Hradec Kralove, Czechia

Acknowledgement: sponsored by the Grant Agency of the Czech Republic, project 205/05/2265

www.ufa.cas.cz/dub/calimaro/calimaro.htm

Abstract

Weather generators (WG) are often used in agricultural and hydrological studies to supply crop models or raifall-runoff models with synthetic weather series representing present or future climate. Typically, WG parameters are derived from the observed series in a given station. CaliM&Ro project is focused on calibrating a stochastic single-site daily weather generator Met&Roll for sites with non-existent or insufficient historical daily weather series. Interpolation of WG parameters from the surrounding stations is used here to calibrate the generator for the ungauged station. Selected WG parameters are interpolated using GIS (ArcView), neural network and nearest-neighbours-based interpolation technique. The tests are based on observational data from 125 stations in Czechia.



Met&Roll weather generator

Met&Roll is a single-site 4-variate stochastic daily weather generator.

variables:
- precipitation amount (PREC)
- daily temperature maximum (TMAX)
- daily temperature minimum (TMIN)
- global solar radiation (SRAD)

model:
- precipitation occurrence ~ Markov chain of 1st to 3rd order (parameters: transition probabilities)
- precipitation amount ~ Gamma distribution (parameters: α, β)
- (SRAD, TMAX, TMIN) ~ 3-variate 1st order AR model (parameters: means and std's /separately for wet and dry days/ of the three variables; two 3×3 matrices representing lag-0 and lag-1 covariances among the 3 variables)
- daily weather generator is linked to the 1st order AR model-based monthly generator (parameters: two 4×4 matrices)
- most parameters of the generator account for the annual cycle

Met&Roll has been mainly used in agricultural studies [see references] dealing with sensitivity of crop yields to changes in climatic characteristics, including impact of the future climate change. It is also implemented in the **PERUN** system designed for climate change impact studies and probabilistic crop yields forecasting. Within the present project, a new system (**M&Rwin**) is being developed. This system links Met&Roll with crop models (now available are WOFOST and CERES models) and hydrological model Sacramento. Climate change scenarios and interpolation of WG parameters is implemented in this system, too.

GIS

GIS is a system designed to capture, store, update, manipulate, analyze, and display the geographic information. ArcGIS is an integrated collection of GIS software products for handling spatial data, developed by Environmental Systems Research Institute (ESRI). For interpolation of selected WG parameters the Geostatistical Analyst (GA) extension of the version ArcView 9.1. were used, which provides a set of tools to create a continuous surface using deterministic and geostatistical methods.

For this study geostatistical techniques (Kriging, Cokriging) based on both statistical and mathematical methods were selected. Kriging methods rely on the notion of autocorrelation as a function of distance. Predicted value depends on two factors: a trend and a fluctuation from the trend, called spatially-autocorrelated random error. Cokriging improves surface prediction of a primary variable by taking into account secondary variables, provided that the primary and secondary variables are spatially correlated.

In first step the spatial data variability, spatial data dependence, and global trends were explored using Exploratory spatial data analysis tools included with GA. This phase helps to select an appropriate model to create a surface. In the next step diagnostics were performed to assess the uncertainty of the predictions. Subsequently various analytical tools and calculated statistics allow to compare models if more than one surface is produced.

Neural Networks

Neural networks (NNs) are inspired by the signal processing in the biological neural systems. Artificial NN is composed from a network of individual neurons. Each neuron receives a number of inputs (either from original data, or from the output of other neurons. Each input comes via a connection (synapse) that has a strength (or weight, which corresponds to synaptic efficacy in a biological neuron) and a single threshold value. The weighted sum of the inputs minus threshold composes the activation of the neuron (= post-synaptic potential of the neuron). The activation signal is then passed through a non-linear activation function (= transfer function) to produce the output of the neuron. A simple network has a feedforward structure: signals flow from inputs (which carry the signal of independent variables), forwards through hidden units, eventually reaching the output units (which form predictions or dependent variables values). Quasi-optimal values of network free parameters (synaptic weights, neuron thresholds) are fixed during iterative “training” from randomly generated initial state of the network, several training algorithms are available. Trained neural network should mirror general features of training data (relations between independent and dependent variables, classification, etc.) and may be used for processing new cases.

The maps shown here were made using the ensemble generalization: The network was repeatedly (50-times) trained with different network initialization and different data division into training and selection subsets. Final interpolated values are the mean values of all 50 individual models (ensemble members). [network type = 3-3-1, 19 degrees of freedom]

[StatSoft, Inc. (2006). Electronic Statistics Textbook. Tulsa, OK: StatSoft. http://www.statsoft.com/textbook/stathome.html]

Nearest Neighbours

In this approach, the interpolated value X_i^ is defined as a weighted average of observed values of parameter X related to the nearest stations with applying a correction for differences in latitude, longitude and altitude:

$$X_i\hat{}(\lambda,\varphi,h) = \sum_{j=1,..,K} w_{ij}\times[X(\lambda_j,\varphi_j,h_j) + b\times(\varphi-\varphi_j) + c\times(\lambda-\lambda_j) + d\times(z-z_j)]$$

where

- w_{ij} is weight accounting for the distance d_{ij} between the two stations:

$$w_{ij} = [1 - (d_{ij}/D)^3]^3] \text{ for } d_{ij} < D, \quad w_{ij} = 0 \text{ for } d_{ij} > D$$

- b, c, d are parameters of the tri-variate linear regression model ($X = a + b\times\varphi + c\times\lambda + d\times z$), which are estimated using data from all available stations (except for the one, for which we interpolate – during the cross-validation test)
- D defines the surroundings of the location for which we interpolate. The value of D is a subject of optimisation. We found, that D~100 km is optimal for the present experiment.

This technique requires an access to the database of the station-specific values of all WG parameters.

Data

- Weather data: 125 stations from Czechia with data coverage greater or equal to 75% during 1961-1990 (= at least 75% of terms have observed values of *TMAX*, *TMIN*, *PREC*, and *CLOUD* or *SUNSHINE* used to estimate *SRAD*) [location of the stations and their altitude is displayed in Figure 1]

- Topography of Czechia is derived from the global digital elevation model GTOPO30 (horizontal grid spacing = 30 arc seconds; i.e. approximately 600×900 m resolution within the territory of Czechia) [http://edc.usgs.gov/products/elevation/gtopo30/gtopo30.html]

Experiment

A complete set of the Met&Roll's parameters for a single station includes a large number of characteristics, mostly accounting for the annual cycle. Of these, 26 parameters were selected for the present experiment (Table I).

1) Data from all 125 stations were used for **mapping** individual WG parameters into 30′×30′ grid using the GTOPO30 topography.

2) In **cross-validation** test (station-specific errors are shown in the Figures, summary statistics are given in **Table I**), WG parameters for each station are interpolated using WG data from the remaining 124 station and the station-specific altitude.

Conclusions

The poster presents the **first results** of experiments with **interpolating weather generator parameters**. **3 interpolation techniques** were tested: **(i) co-kriging** (using GIS), **(ii) neural networks**, **(iii) nearest neighbours**. Following the validation scores given in Table I, one could deduce that:

- **GIS** appears to provide the best results, **but**: the RV and RMSE scores given in the table for this technique are overestimated (RV) and underestimated (RMSE), as the settings of the interpolation procedure for each single WG parameter was optimised using all 125 stations.
- **nearest neighbours** provide slightly better results compared to **neural network**. **But**: **(i)** to use the nearest neighbours, all learning data must be available during interpolation into ungauged stations (in contrast with neural network, which do not need access to learning data after being trained). **(ii)** Both interpolation techniques may be subject of further improvements. **Let's not be hasty!**
- **some WG parameters** are successfully interpolated [e.g.mean values of TMAX, TMIN and mean precipitation amount on wet days (= Γsh×Γsc)], which follows from their close relationships with latitude+longitude+altitude as inditated by r^2 value.

Plans for future (within the frame of *CaliM&Ro* project):

- **more detailed validation** of the interpolation techniques **[(i)** with using unbiased skill scores for co-kriging, **(ii)** trying to further improve neural networks and nearest neighbours interpolators, **(iii)** involve other WG parameters]
- to test **how the interpolated weather generator reproduces various climatic characteristics** (means, variability, extremes, heat/cold/dry/wet spells, ...)
- to use synthetic series generated by "interpolated" weather generator as an input to **crop models and hydrological rainfall-runoff model**. Output from these models will be compared with the output obtained with using generator calibrated by the site-specific observational data.

References

[1] Dubrovský, M., 1997: 'Creating Daily Weather Series with Use of the Weather Generator', *Environmetrics* **8**, 409-424.

[2] Buchtele, J., Buchtelová, M., Fořtová, M., and Dubrovský, M., 1999: 'Runoff Changes in Czech River Basins - the Outputs of Rainfall-Runoff Simulations Using Different Climate Change Scenarios', *J. Hydrol. Hydromech.*, **47,** 180-194.

[3] Dubrovský M., Zalud Z. and Stastna M., 2000: "Sensitivity of CERES-Maize yields to statistical structure of daily weather series". *Climatic Change* **46**, 447- 472.

[4] Huth R., Kyselý J., Dubrovský M., 2001: "Time structure of observed, GCM-simulated, downscaled, and stochastically generated daily temperature series". *Journal of Climate*, **14**, 4047-4061.

[5] Žalud Z., Dubrovský M., 2002: "Modelling climate change impacts on maize growth and development in the Czech republic". *Theoretical and Applied Climatology*, **72**, 85-102.

[6] Huth R., Kyselý J., Dubrovský M., 2003: Simulation of Surface Air Temperature by GCMs, Statistical Downscaling and Weather Generator: Higher-Order Statistical Moments. *Studia Geophysica et Geodaetica* **47**, 203-216

[7] Hejzlar J., Dubrovský M., Buchtele J. and Růžička M., 2003: "The effect of climate change on the concentration of dissolved organic matter in a temperate stream (the Malše River, South Bohemia)". *Science of Total Environment* **310** *143-152.*

[8] Dubrovský M., Buchtele J., Žalud Z., 2004: High-Frequency and Low-Frequency Variability in Stochastic Daily Weather Generator and Its Effect on Agricultural and Hydrologic Modelling. *Climatic Change* **63** (No.1-2), 145-179.

[9] Trnka M., Dubrovský M., Semerádova D., Žalud Z., 2004: Projections of uncertainties in climate change scenarios into expected winter wheat yields. *Theoretical and Applied Climatology*, **77**, 229-249.

[10] Trnka M., Dubrovský M., Žalud Z., 2004: Climate Change Impacts and Adaptation Strategies in Spring Barley Production in the Czech Republic. *Climatic Change* **64** (No. 1-2), 227-255.

[11] Kyselý J., Dubrovský M., 2005: Simulation of extreme temperature events by a stochastic weather generator: effects of interdiurnal and interannual variability reproduction. *Int.J.Climatol.* **25**, 251-269.

www.ufa.cas.cz/dub/dub.htm

www.ufa.cas.cz/dub/calimaro/calimaro.htm

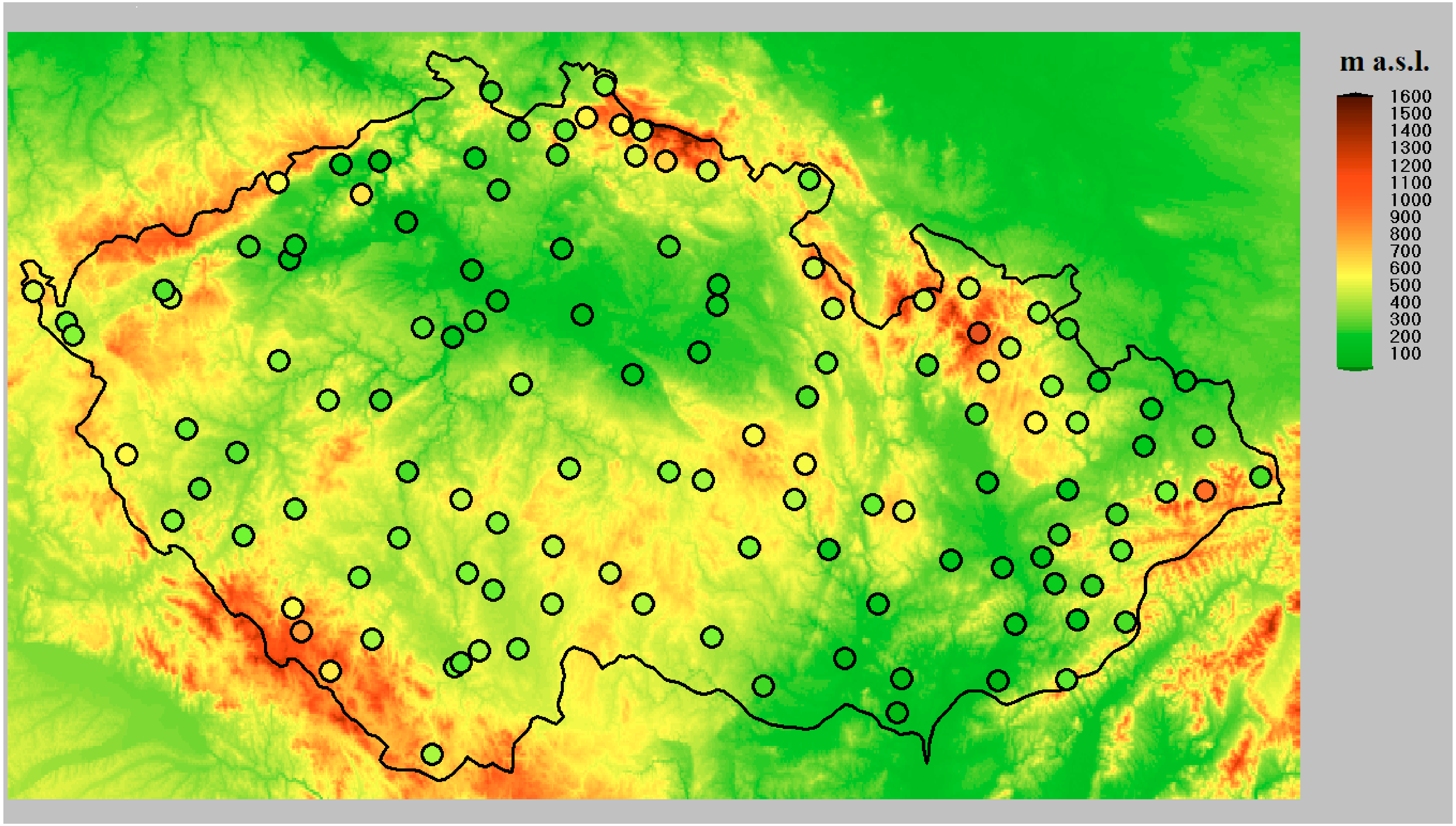


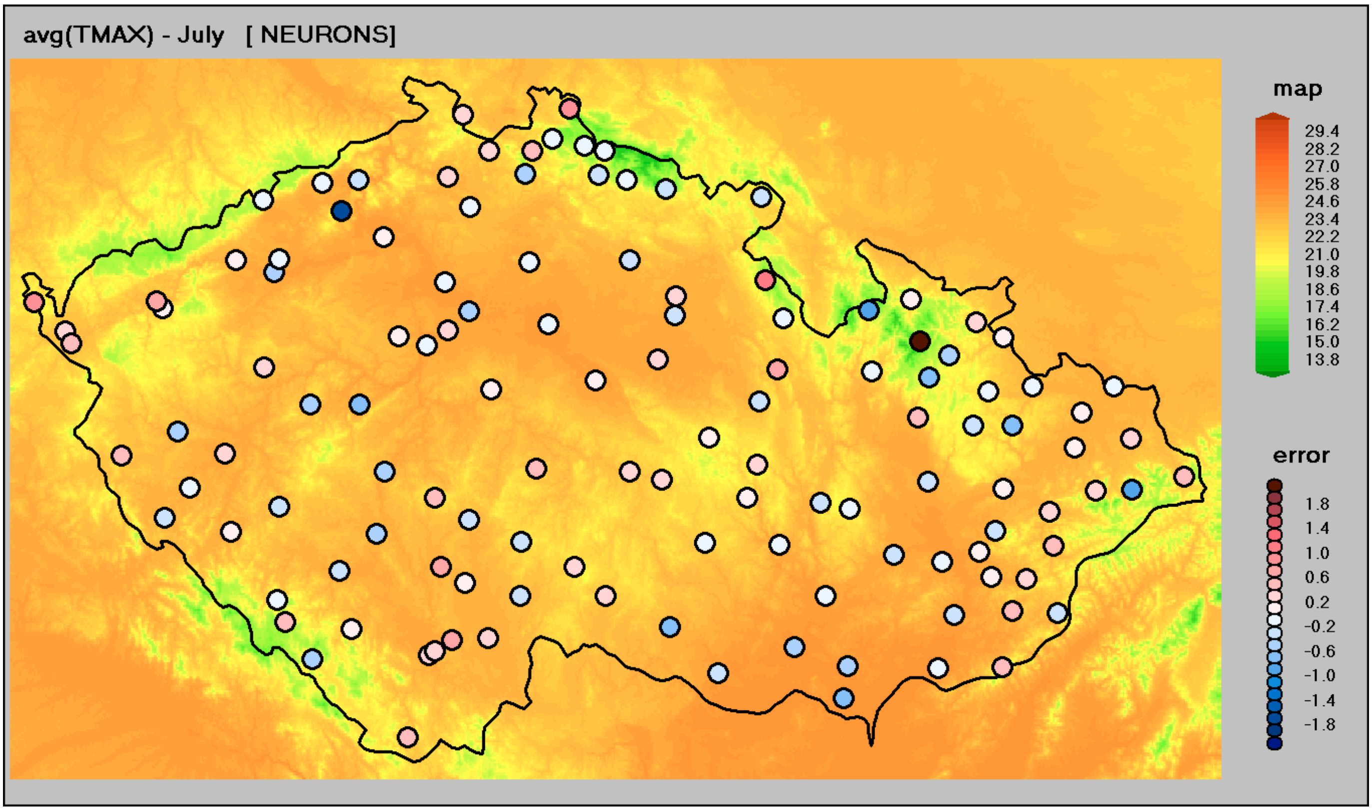
Figure 1 Topography of Czechia and location of the 125 stations used for interpolation. Altitude of the stations is represented by a colour according to the scale

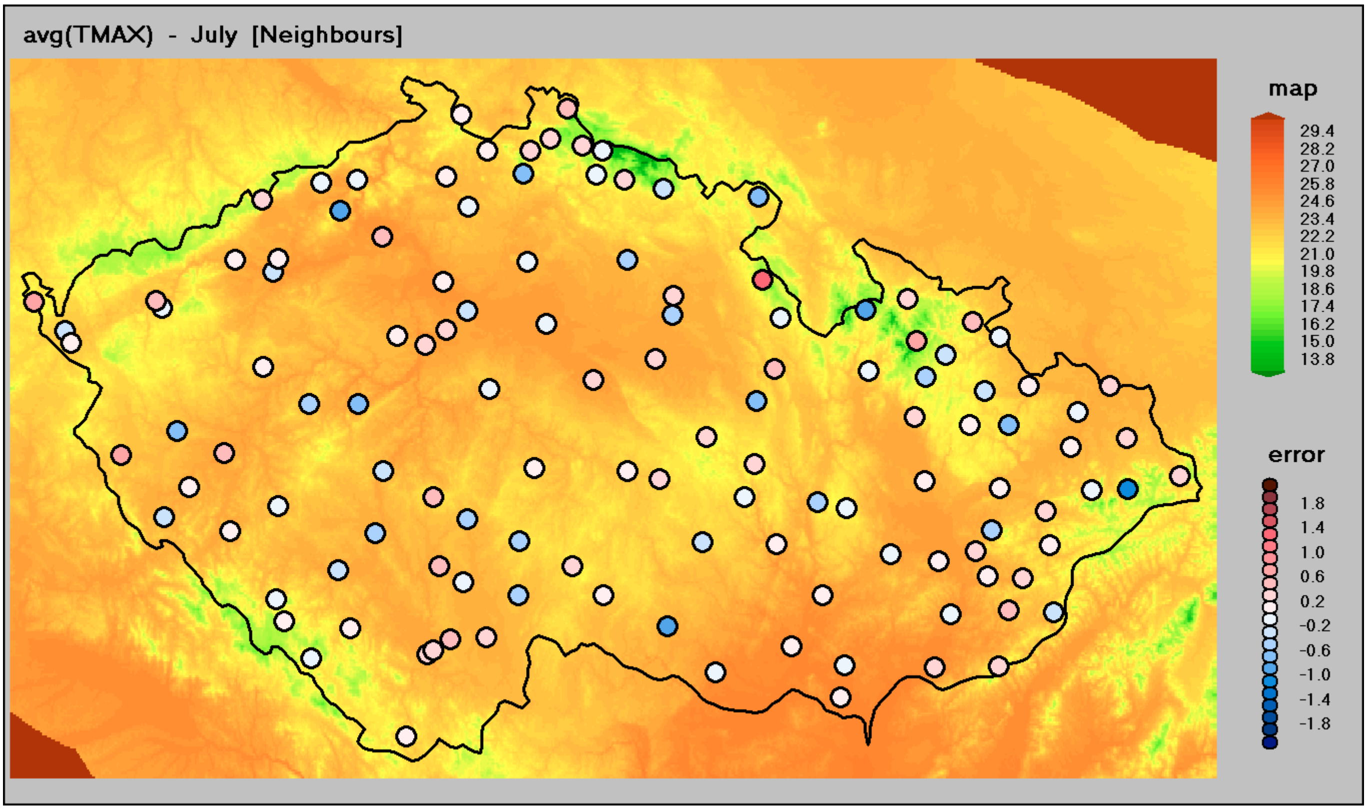
Table I. Summary of WG parameters' characteristics (average, standard deviation, minimum and maximum from the set of 125 stations) and **results of the cross-validation** of the three interpolation techniques

Fig.		avg	std	min	max	r2 (λ,φ,z)	Neural Network RMSE	Neural Network RV[%]	Nearest Neighbours RMSE	Nearest Neighbours RV[%]	co-kriging RMSE	co-kriging RV[%]
	a) Solar radiation											
	avg, January	2.54	0.35	1.72	4.38	0.47	**0.26**	**43**	**0.26**	**45**		
	avg, July	18.30	1.13	15.70	21.44	0.20	**1.09**	**8**	**1.06**	**12**		
	std, January	1.31	0.19	0.58	1.72	0.32	**0.16**	**23**	**0.16**	**29**		
	std, July	5.93	0.38	4.87	6.85	0.16	**0.35**	**15**	**0.35**	**17**		
	b) Daily temperature maximum											
	avg, January	0.0	1.1	-4.9	1.4	0.75	**0.48**	**81**	**0.49**	**81**	**0.67**	**63**
2	avg, July	22.6	2.0	12.8	25.4	0.96	**0.54**	**92**	**0.38**	**96**	**0.92**	**78**
	std*, January	4.63	0.34	3.80	5.39	0.20	**0.26**	**40**	**0.27**	**35**		
3	std*, July	4.58	0.23	4.07	5.17	0.60	**0.14**	**61**	**0.14**	**62**	**0.13**	**68**
	c) Daily temperature minimum											
4	avg, January	-5.8	1.0	-9.6	-3.5	0.69	**0.61**	**66**	**0.60**	**67**	**0.591**	**68**
	avg, July	11.1	1.2	6.6	13.4	0.64	**0.78**	**58**	**0.72**	**63**	**0.597**	**75**
	std, January	5.87	0.57	4.51	7.12	0.24	**0.54**	**11**	**0.48**	**29**		
	std, July	2.92	0.19	2.57	3.49	0.38	**0.16**	**31**	**0.15**	**38**		
	d) Precipitation											
	Γsh, January	0.83	0.10	0.62	1.27	0.03	**0.099**	**4**	**0.101**	**-1**		
	Γsh, July	0.72	0.08	0.45	0.97	0.10	**0.081**	**-8**	**0.075**	**7**		
5	Γsh×Γsc, Jan.	2.69	1.00	1.48	7.27	0.42	**0.812**	**33**	**0.684**	**53**	**0.271**	**93**
	Γsh×Γsc, July	5.80	1.14	3.83	11.18	0.55	**0.760**	**55**	**0.734**	**58**		
6	Pwet, January	0.49	0.07	0.32	0.68	0.47	**0.056**	**41**	**0.054**	**46**	**0.024**	**89**
	Pwet, July	0.46	0.05	0.36	0.69	0.37	**0.043**	**32**	**0.038**	**46**		
	Pwd, January	0.33	0.04	0.24	0.47	0.22	**0.038**	**14**	**0.037**	**19**		
	Pwd, July	0.32	0.03	0.27	0.44	0.27	**0.024**	**24**	**0.023**	**31**		
	e) lag-0 and lag-1 covariances among SRAD*, TMAX* and TMIN*											
	cov(SRAD*,TMAX*)	0.38	0.05	0.26	0.51	0.25	**0.037**	**30**	**0.034**	**41**		
	cov(SRAD*,TMIN*)	-0.16	0.11	-0.47	0.16	0.22	**0.104**	**12**	**0.099**	**19**		
7	cov(TMAX*,TMIN*)	0.62	0.11	0.32	0.86	0.34	**0.131**	**-50**	**0.085**	**37**	**0.071**	**56**
	lag-1-cov(SRAD*)	0.28	0.06	0.17	0.43	0.03	**0.064**	**-5**	**0.064**	**-6**		
8	lag-1-cov(TMAX*)	0.68	0.02	0.63	0.71	0.29	**0.015**	**26**	**0.013**	**42**	**0.012**	**53**
9	lag-1-cov(TMIN*)	0.66	0.05	0.52	0.77	0.33	**0.052**	**-2**	**0.043**	**29**	**0.039**	**41**

legend:
- r2(λ,φ,z) = multiple correlation coefficient of WG parameter with λ, φ, z
- RMSE = Root mean square error
- RV = reduction of variance
- std* = standard deviation of the deviations from a mean annual cycle
- Γsh, Γsc = shape and scale parameters of Gamma distribution (product of the two parameters equals the mean of the Gamma-distributed variable
- Pwet = probability of wet day occurrence
- Pwd = transition probability of wet day occurrence given the previous day was dry
- lag-1-cov = lag-1day covariance; e.g. lag-1-cov(TMAX*) = cov(TMAX*(d), TMAX*(d-1))
- SRAD*, TMAX*, TMIN* = standardised values of SRAD, TMAX, and TMIN. For standardisation, the averages and std.deviations relate to a given Julian day and for either wet or dry day

Fig. 2. avg(TMAX) – July





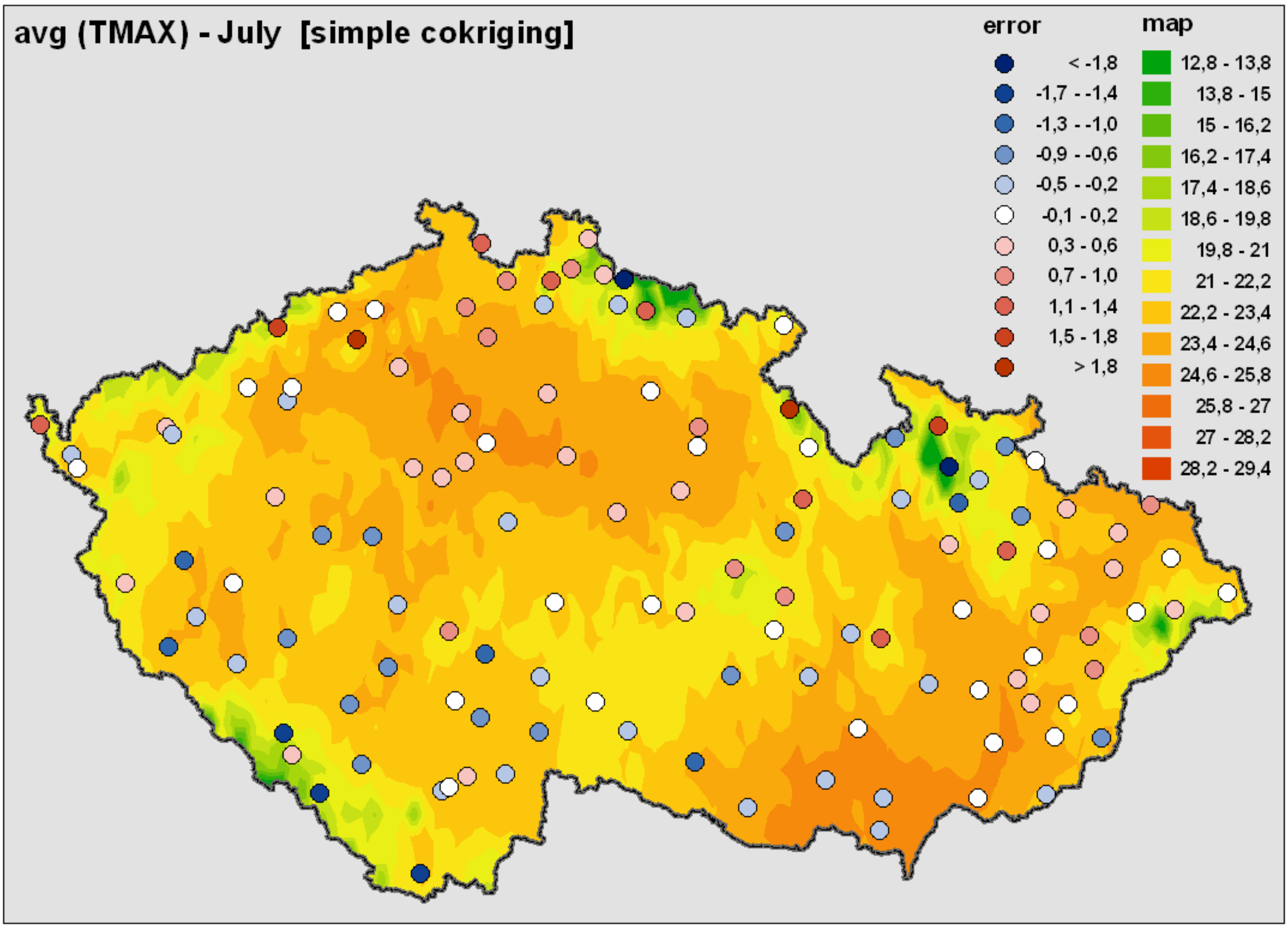
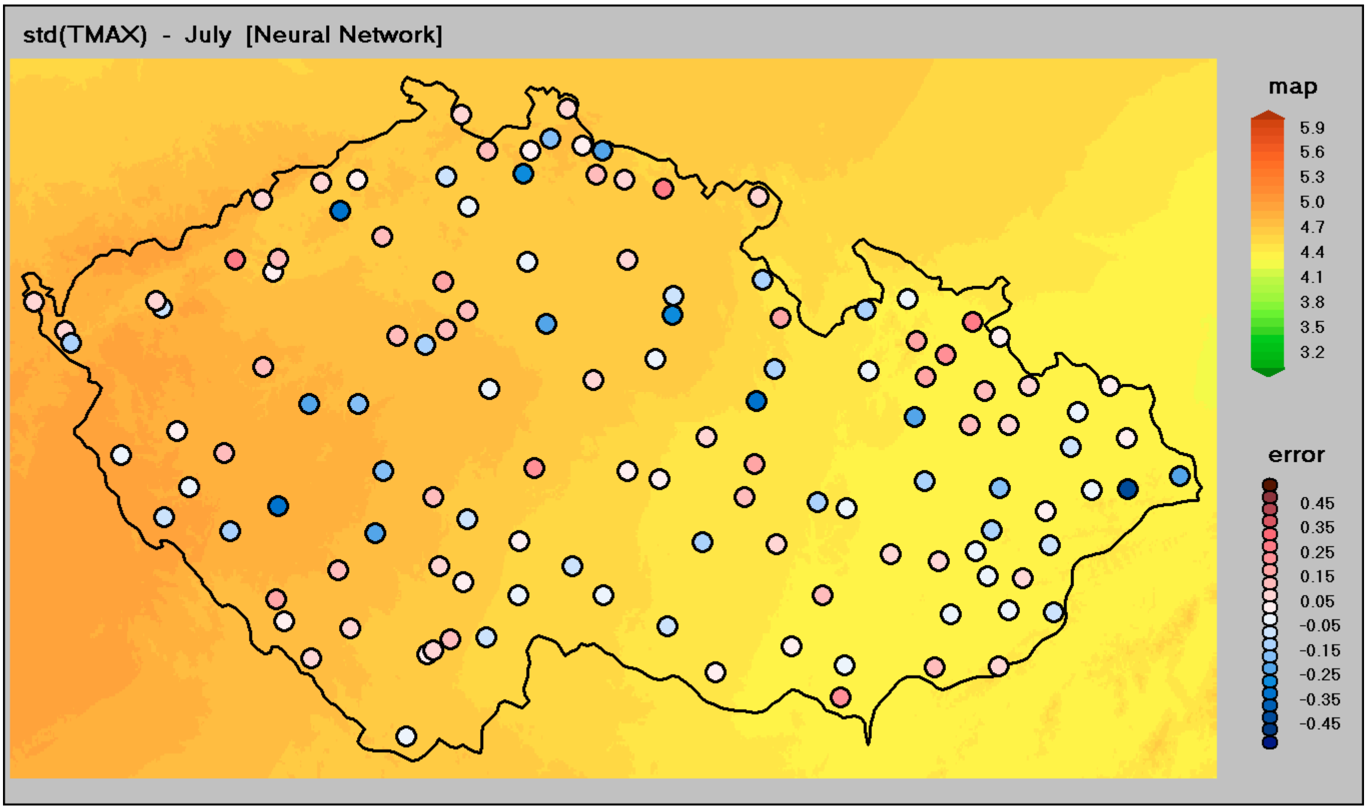
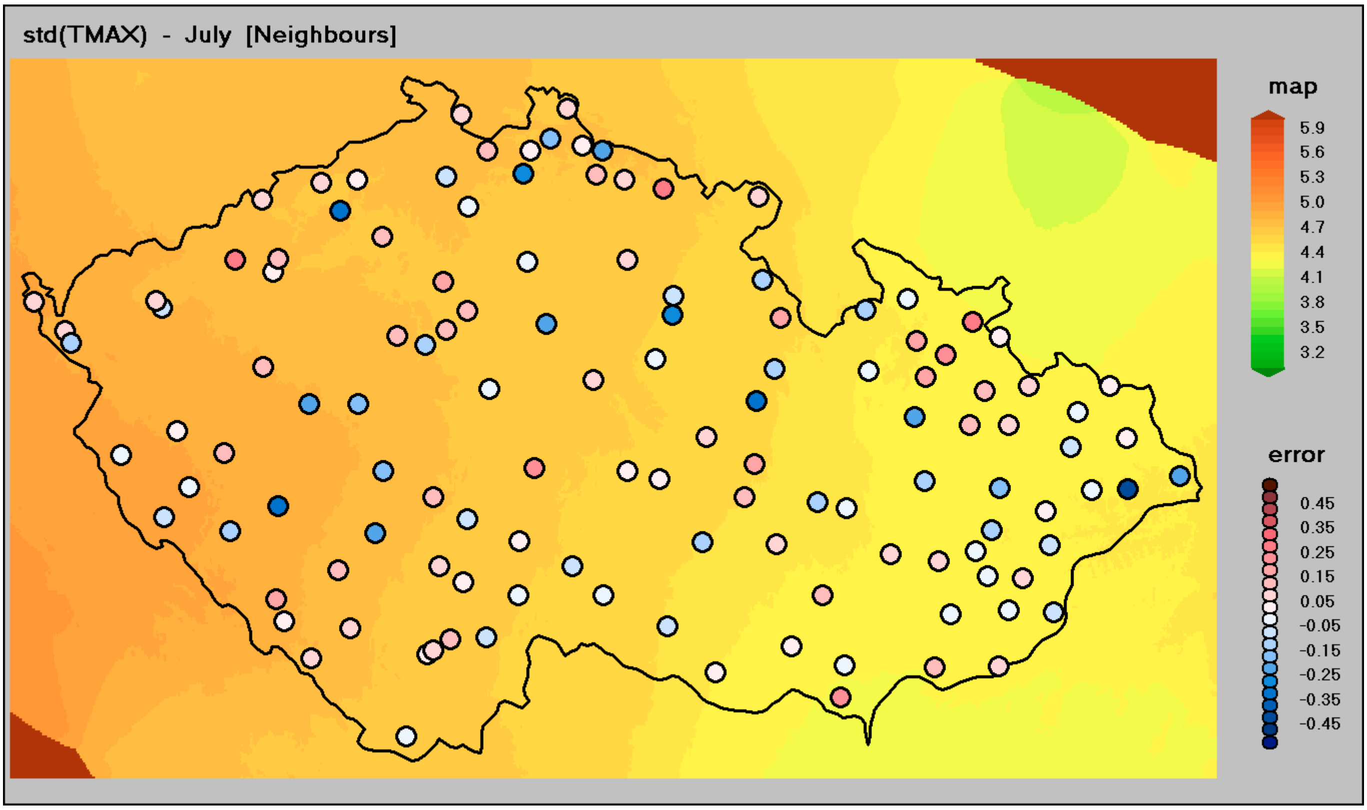

avg (TMAX) - July [simple cokriging]
error
< -1,8
-1,7 - -1,4
-1,3 - -1,0
-0,9 - -0,6
-0,5 - -0,2
-0,1 - 0,2
0,3 - 0,6
0,7 - 1,0
1,1 - 1,4
1,5 - 1,8
> 1,8
map
12,8 - 13,8
13,8 - 15
15 - 16,2
16,2 - 17,4
17,4 - 18,6
18,6 - 19,8
19,8 - 21
21 - 22,2
22,2 - 23,4
23,4 - 24,6
24,6 - 25,8
25,8 - 27
27 - 28,2
28,2 - 29,4

Fig. 3. std(TMAX) – July





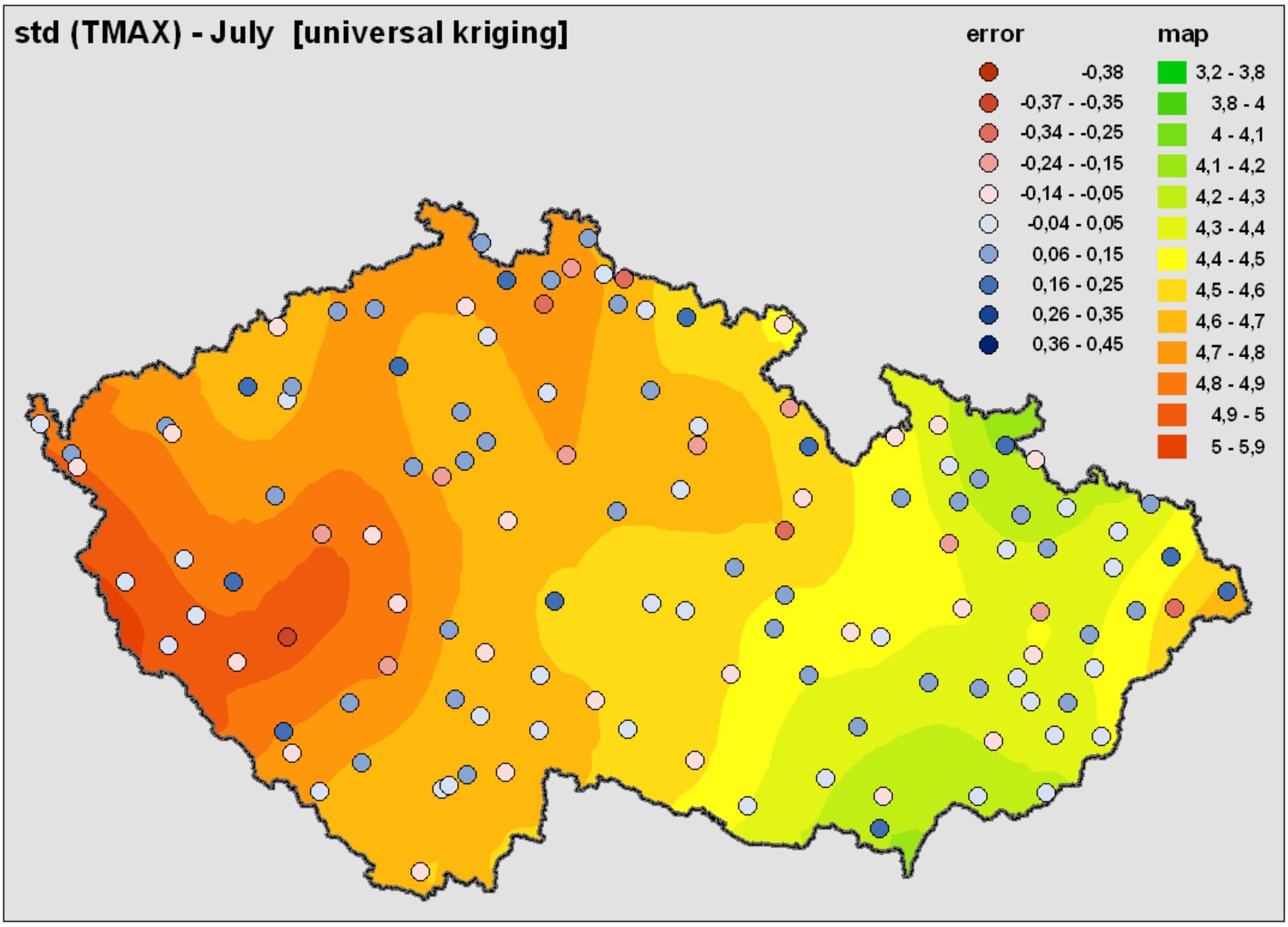
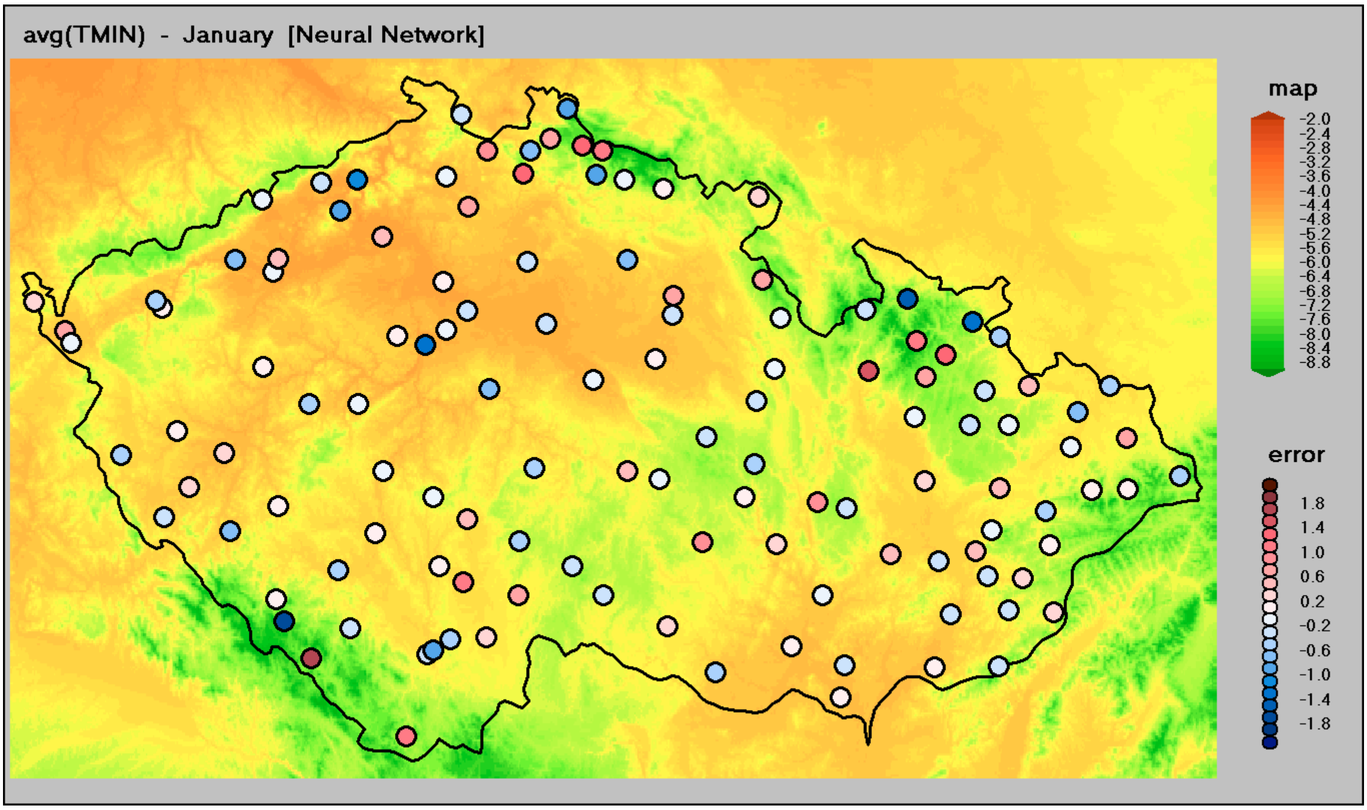
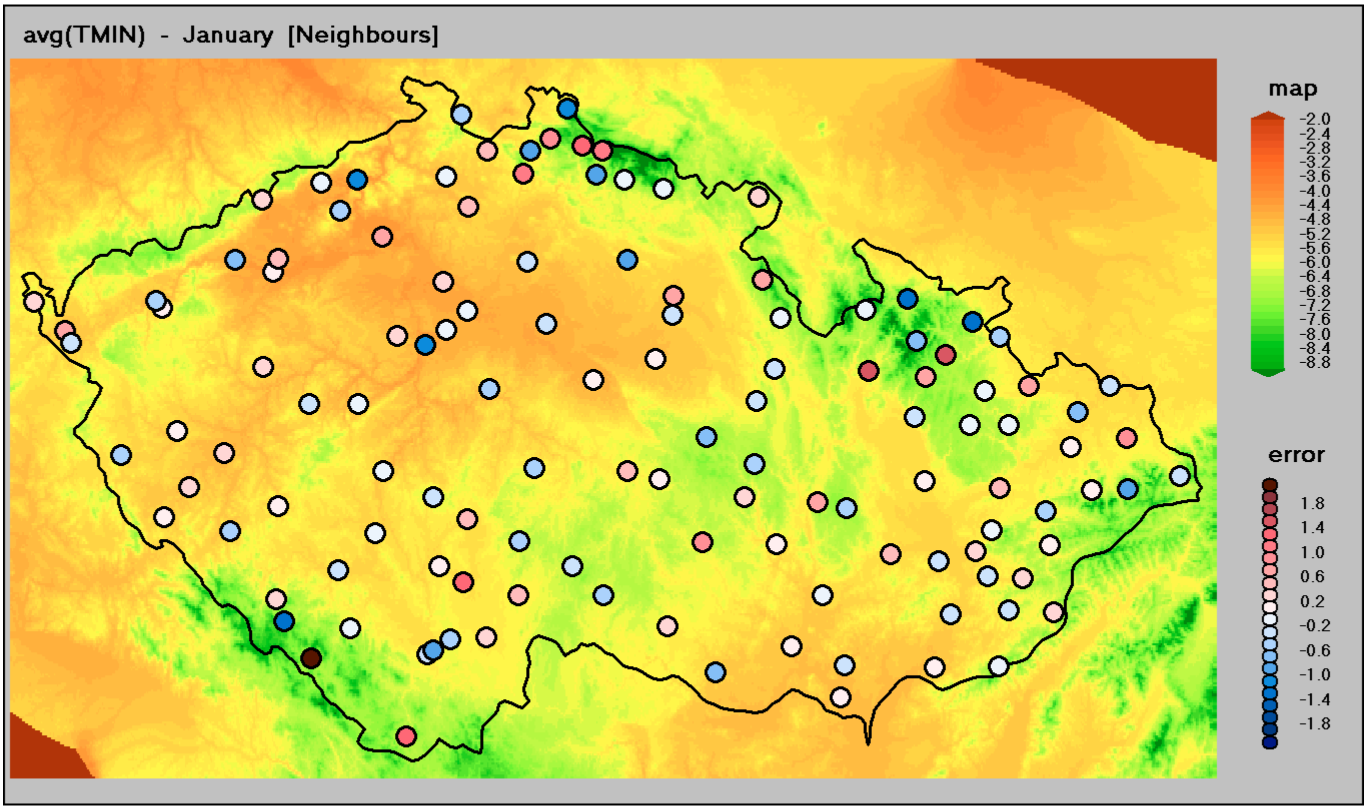

std (TMAX) - July [universal kriging]
error
-0,38
-0,37 - -0,35
-0,34 - -0,25
-0,24 - -0,15
-0,14 - -0,05
-0,04 - 0,05
0,06 - 0,15
0,16 - 0,25
0,26 - 0,35
0,36 - 0,45
map
3,2 - 3,8
3,8 - 4
4 - 4,1
4,1 - 4,2
4,2 - 4,3
4,3 - 4,4
4,4 - 4,5
4,5 - 4,6
4,6 - 4,7
4,7 - 4,8
4,8 - 4,9
4,9 - 5
5 - 5,9

Fig. 4. avg(TMIN) – January





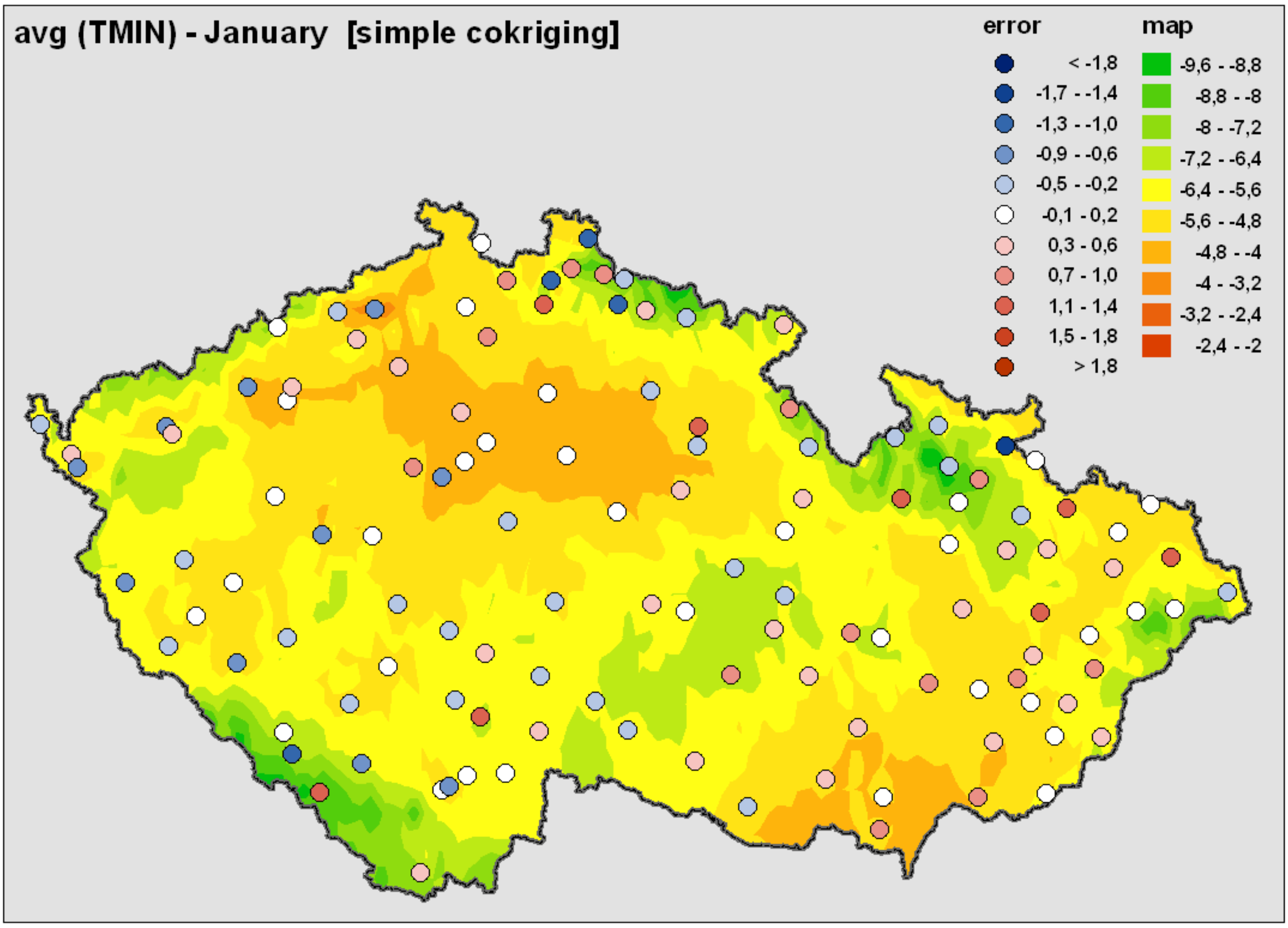
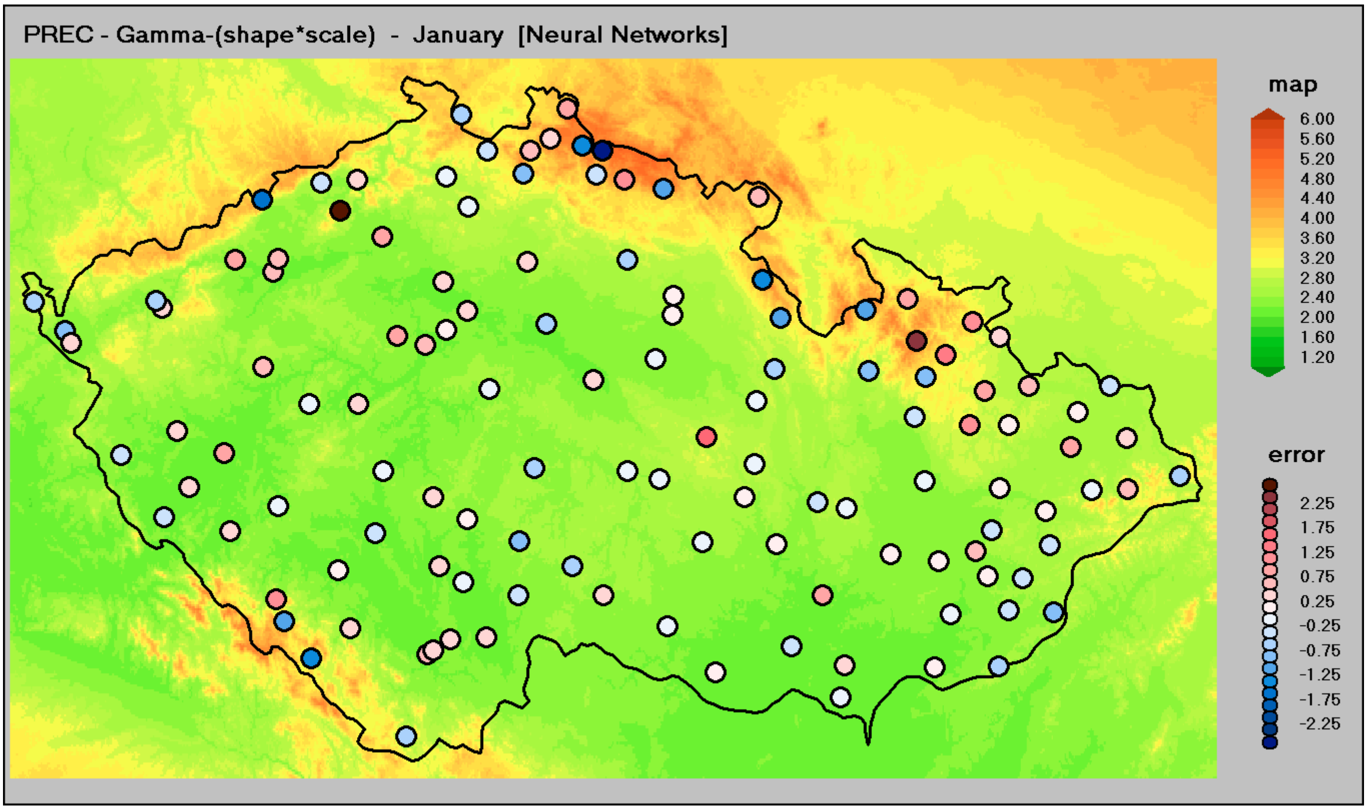
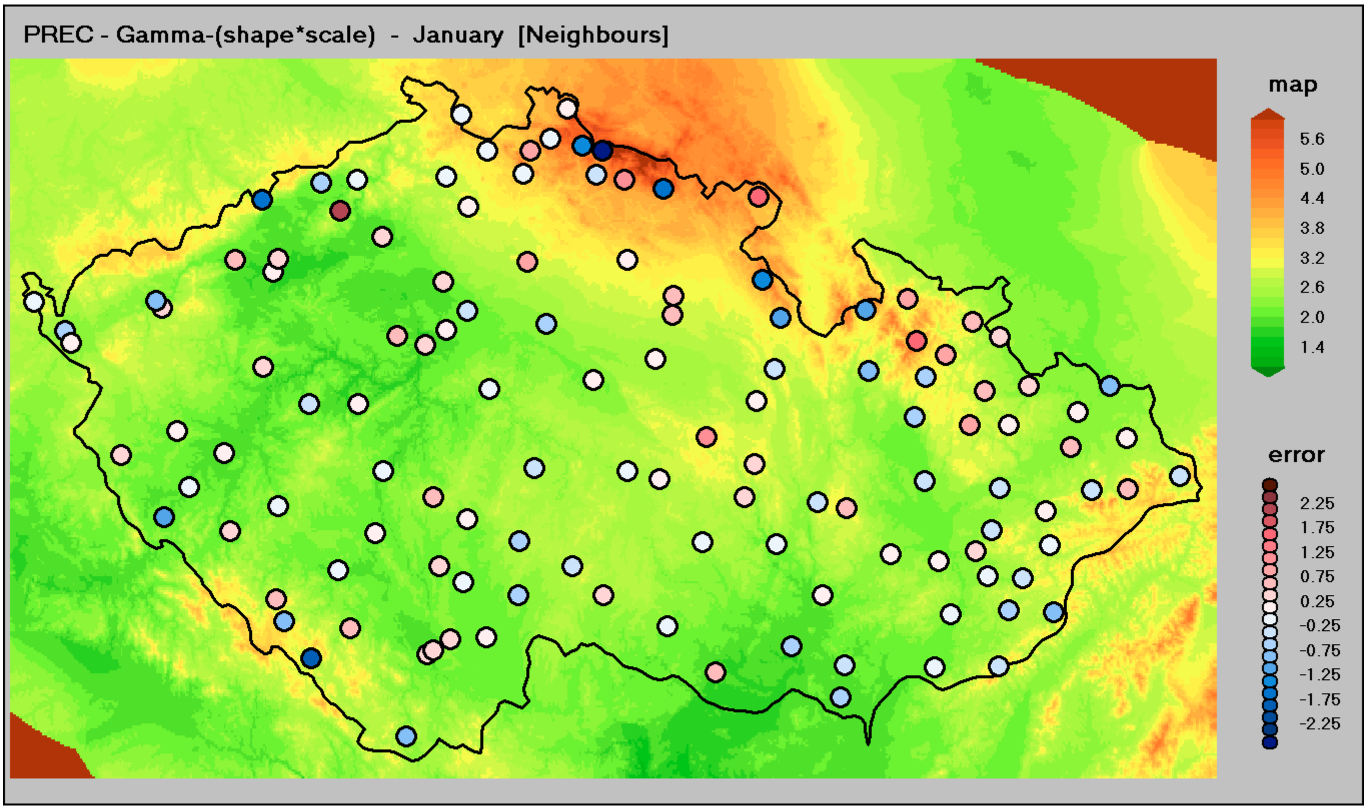

avg (TMIN) - January [simple cokriging]
error
< -1,8
-1,7 - -1,4
-1,3 - -1,0
-0,9 - -0,6
-0,5 - -0,2
-0,1 - 0,2
0,3 - 0,6
0,7 - 1,0
1,1 - 1,4
1,5 - 1,8
> 1,8
map
-9,6 - -8,8
-8,8 - -8
-8 - -7,2
-7,2 - -6,4
-6,4 - -5,6
-5,6 - -4,8
-4,8 - -4
-4 - -3,2
-3,2 - -2,4
-2,4 - -2

Fig. 5. PREC – Γshape×Γscale – January





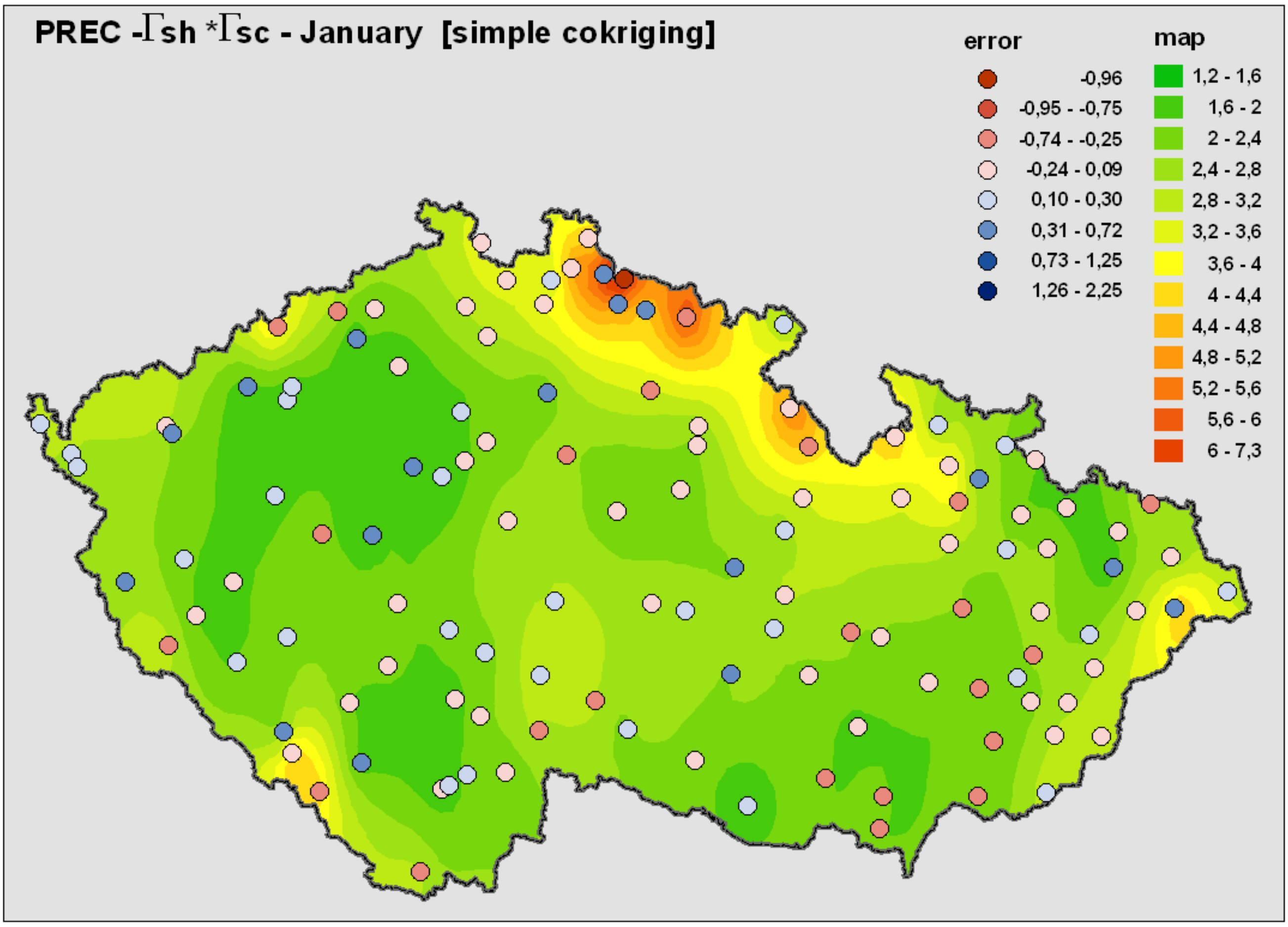
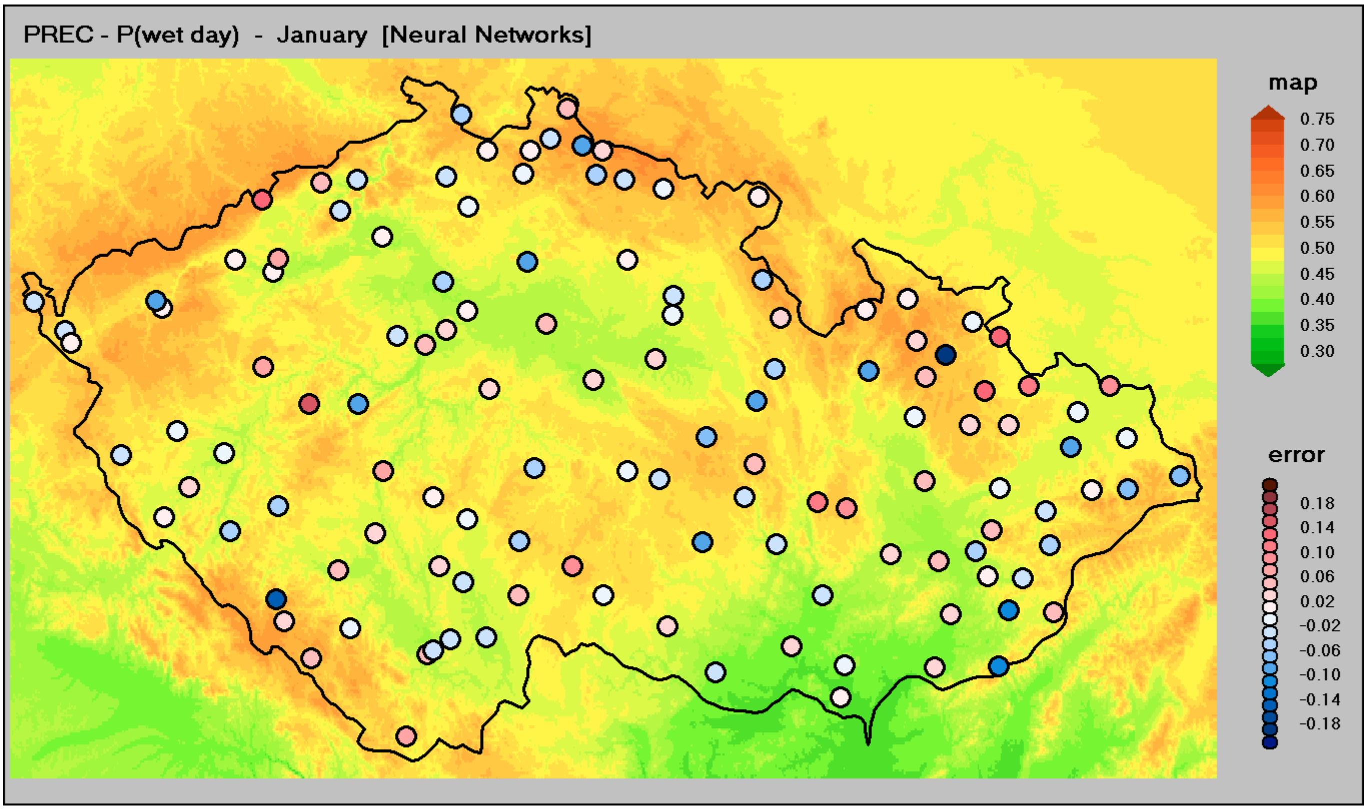
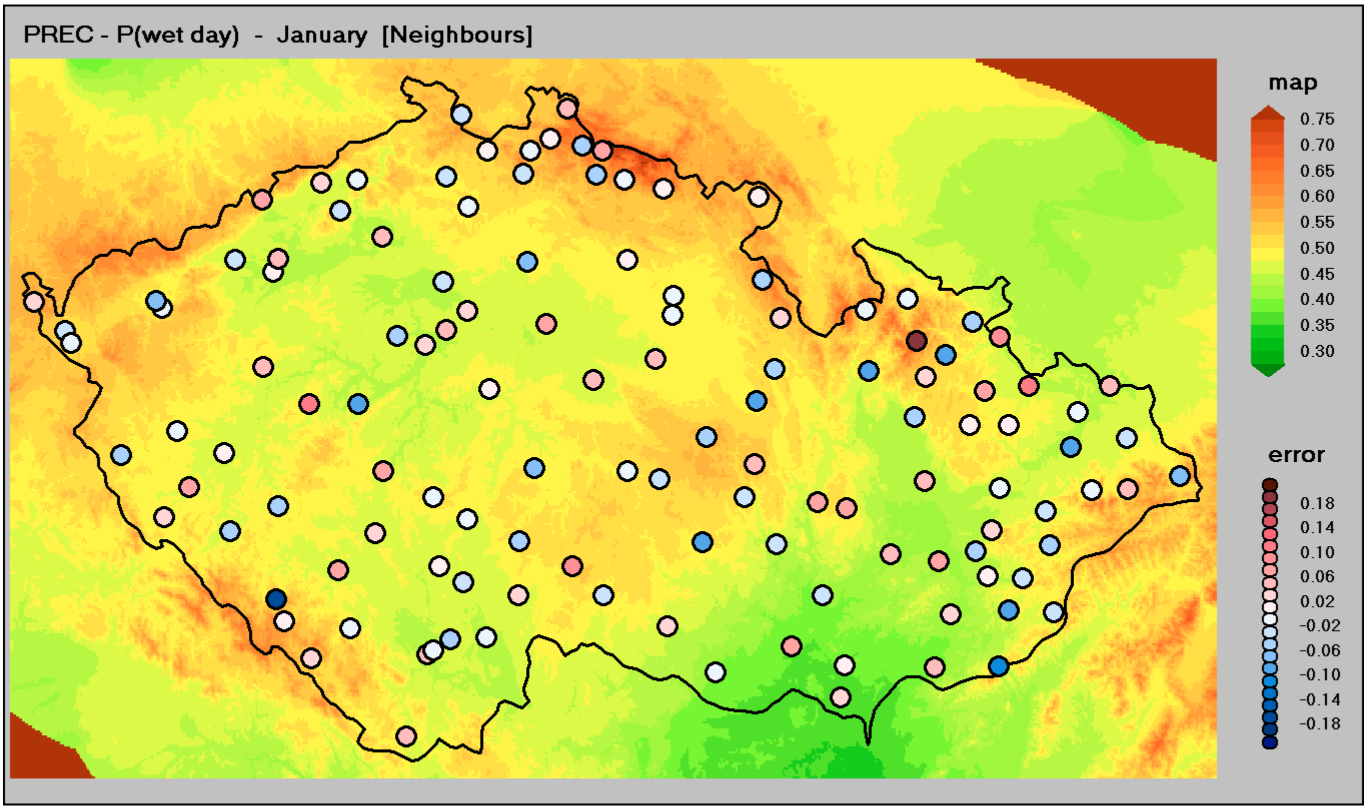

PREC -Γsh *Γsc - January [simple cokriging]
error
-0,96
-0,95 - -0,75
-0,74 - -0,25
-0,24 - 0,09
0,10 - 0,30
0,31 - 0,72
0,73 - 1,25
1,26 - 2,25
map
1,2 - 1,6
1,6 - 2
2 - 2,4
2,4 - 2,8
2,8 - 3,2
3,2 - 3,6
3,6 - 4
4 - 4,4
4,4 - 4,8
4,8 - 5,2
5,2 - 5,6
5,6 - 6
6 - 7,3

Fig. 6. PREC: Prob(wet day) – January





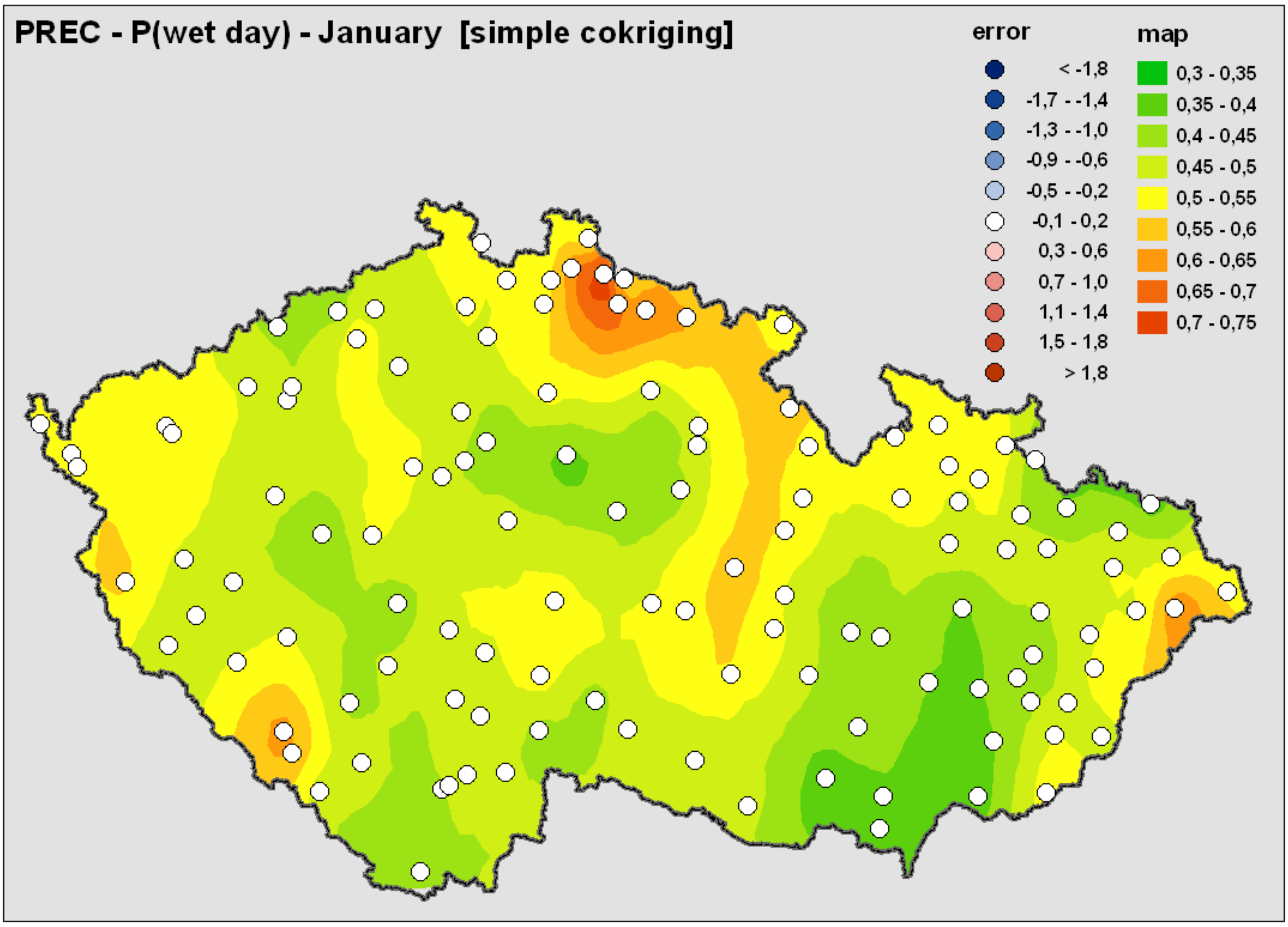
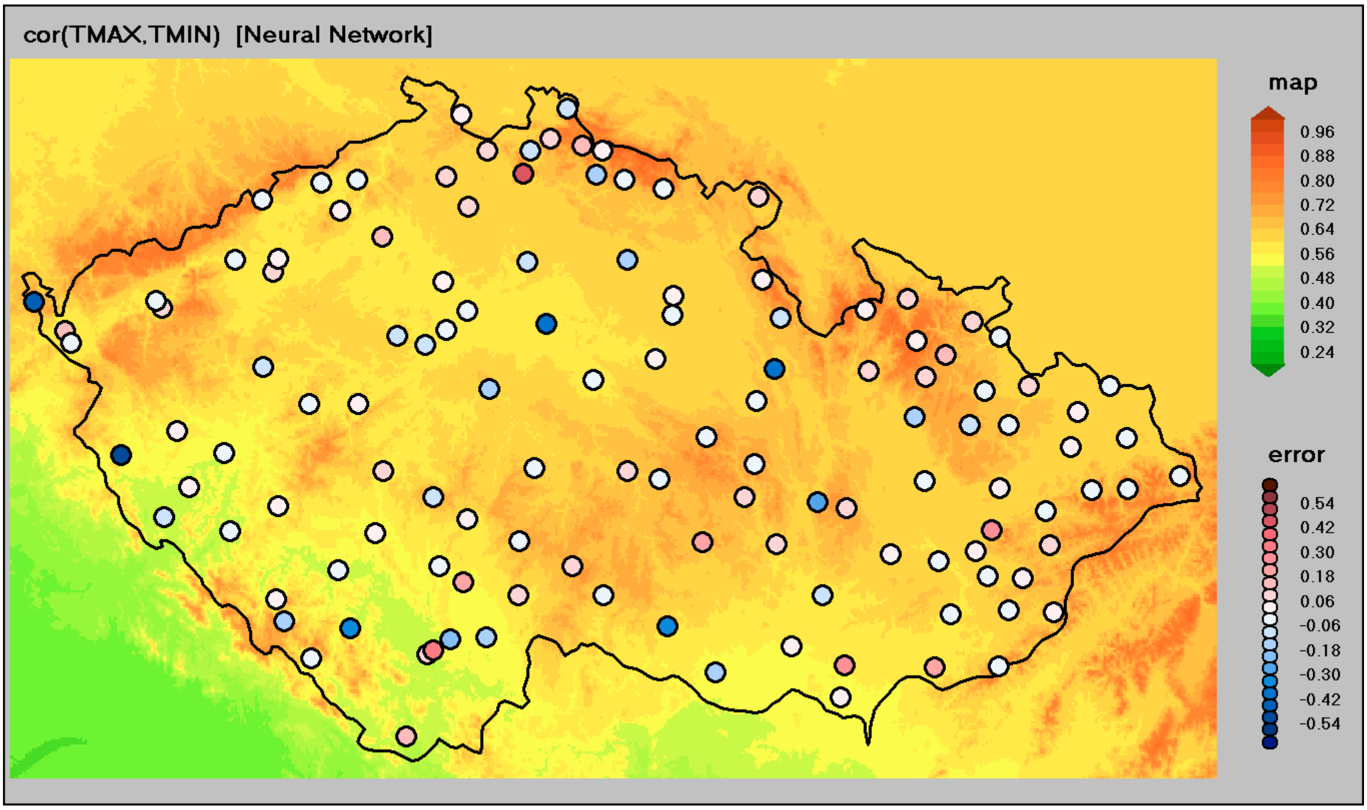
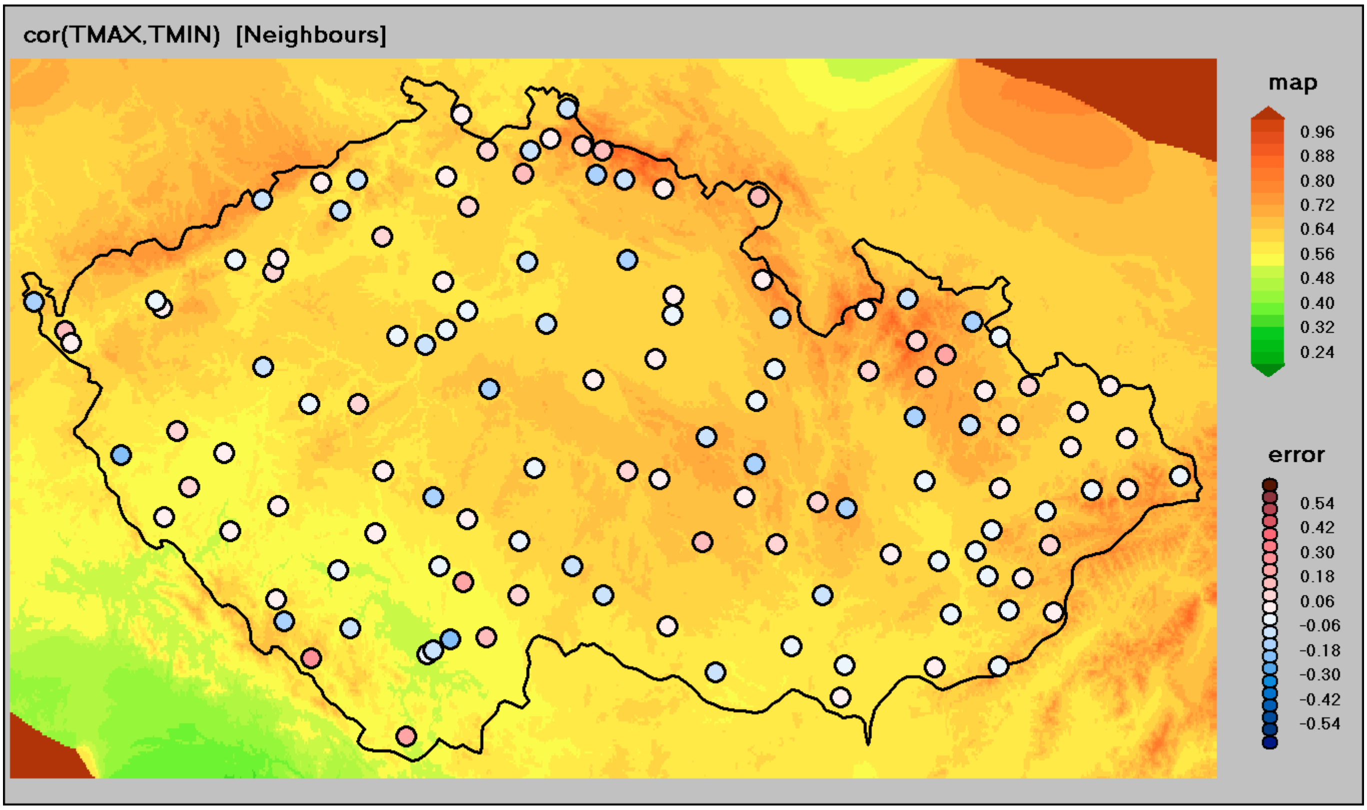

PREC - P(wet day) - January [simple cokriging]
error
< -1,8
-1,7 - -1,4
-1,3 - -1,0
-0,9 - -0,6
-0,5 - -0,2
-0,1 - 0,2
0,3 - 0,6
0,7 - 1,0
1,1 - 1,4
1,5 - 1,8
> 1,8
map
0,3 - 0,35
0,35 - 0,4
0,4 - 0,45
0,45 - 0,5
0,5 - 0,55
0,55 - 0,6
0,6 - 0,65
0,65 - 0,7
0,7 - 0,75

Fig. 7. cov(TMAX*,TMIN*)





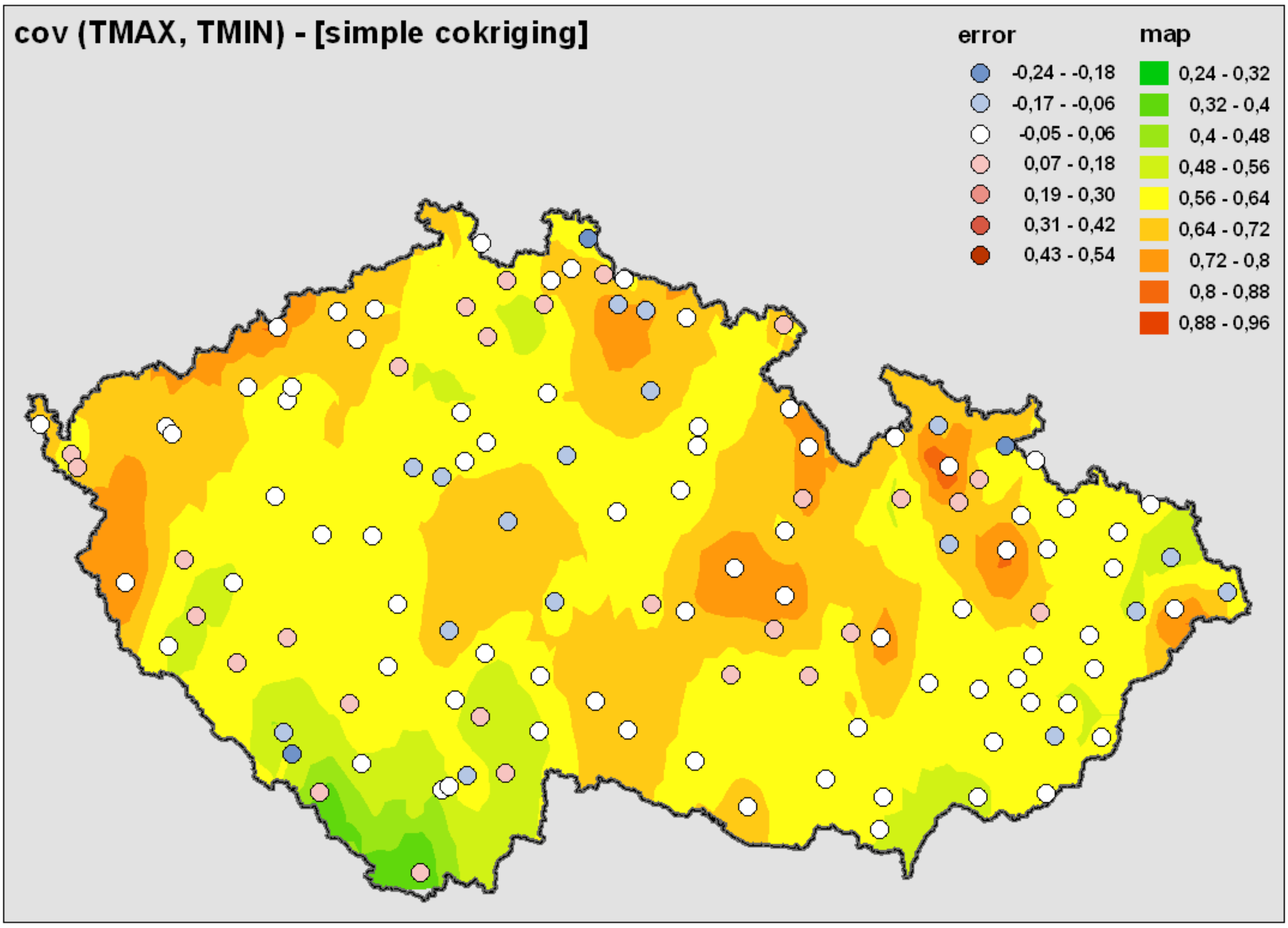
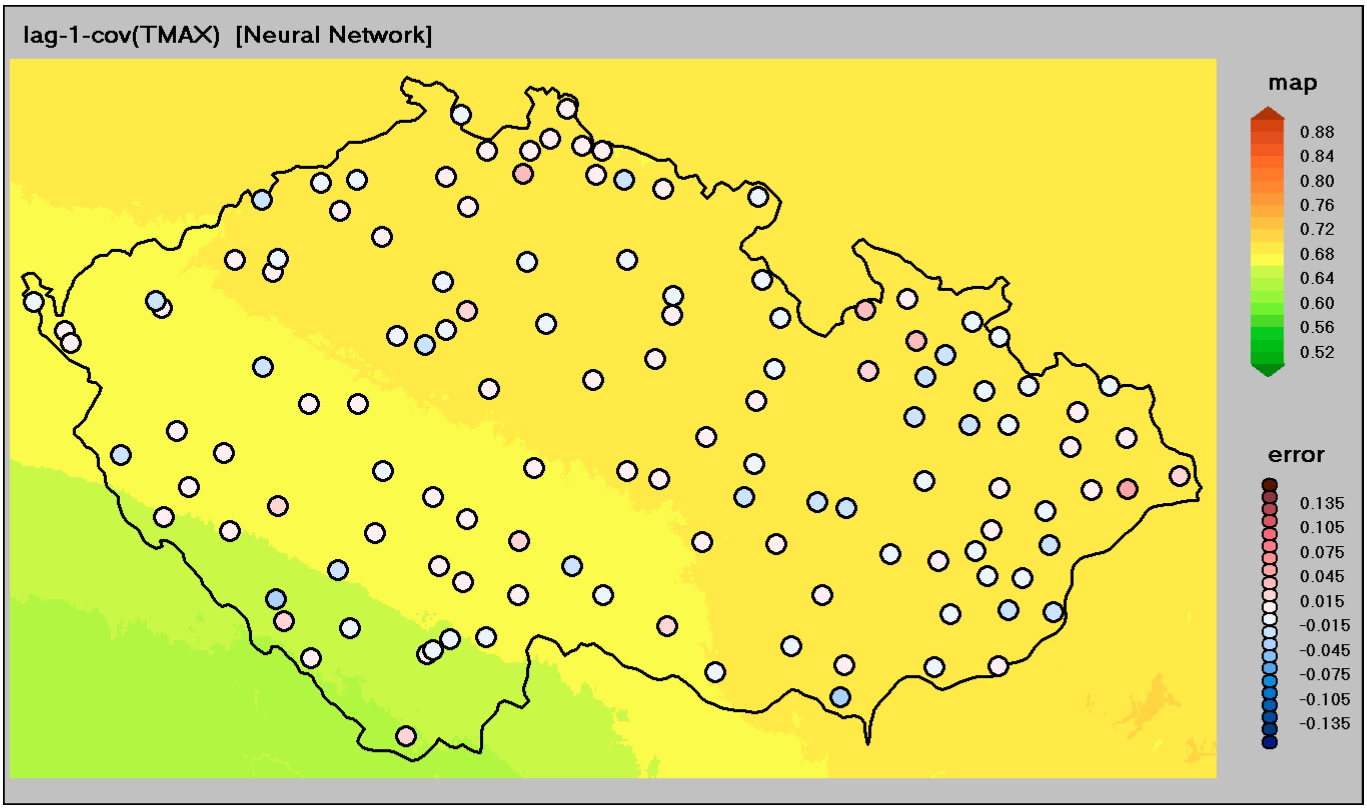
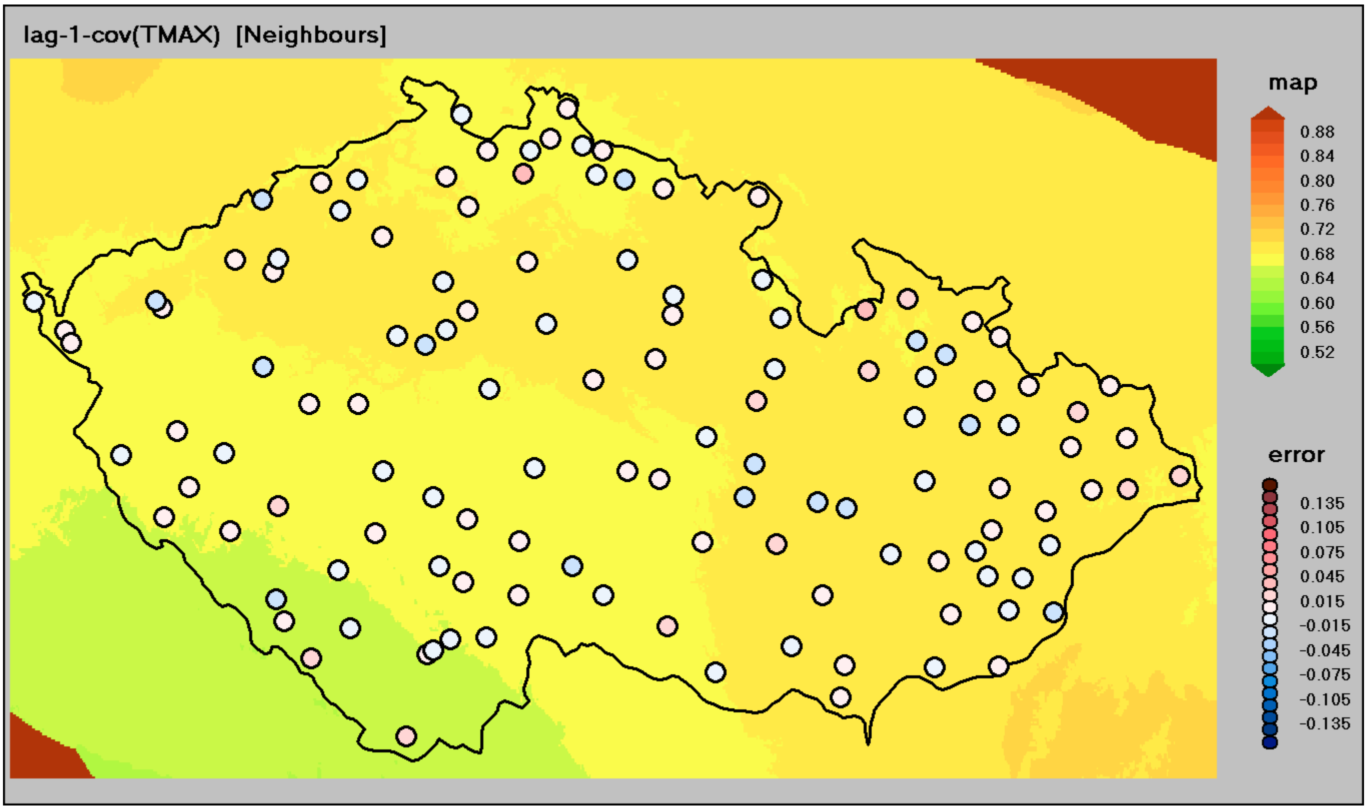

cov (TMAX, TMIN) - [simple cokriging]
error
-0,24 - -0,18
-0,17 - -0,06
-0,05 - 0,06
0,07 - 0,18
0,19 - 0,30
0,31 - 0,42
0,43 - 0,54
map
0,24 - 0,32
0,32 - 0,4
0,4 - 0,48
0,48 - 0,56
0,56 - 0,64
0,64 - 0,72
0,72 - 0,8
0,8 - 0,88
0,88 - 0,96

Fig. 8. cov(TMAX*(d), TMAX*(d-1))





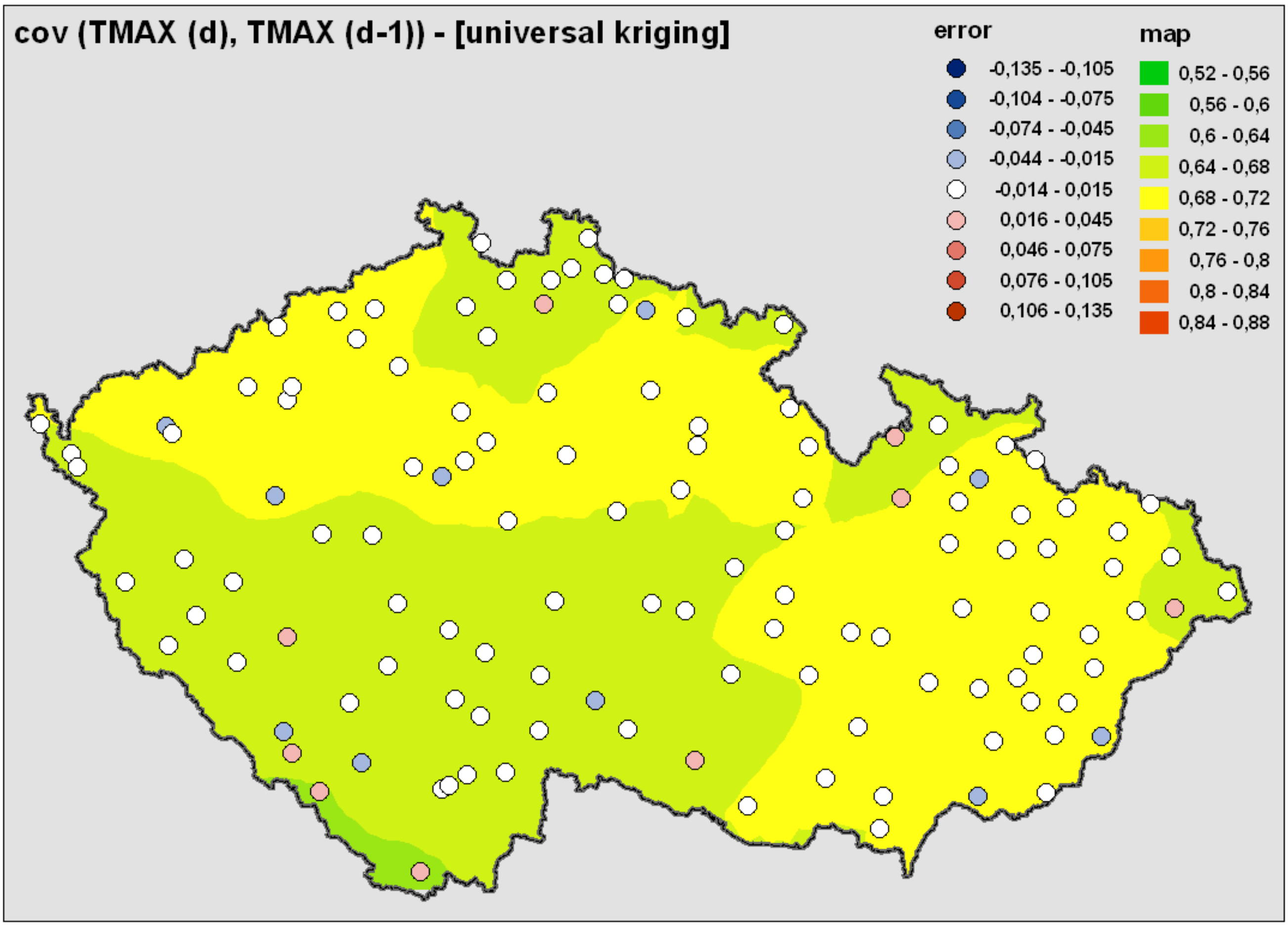
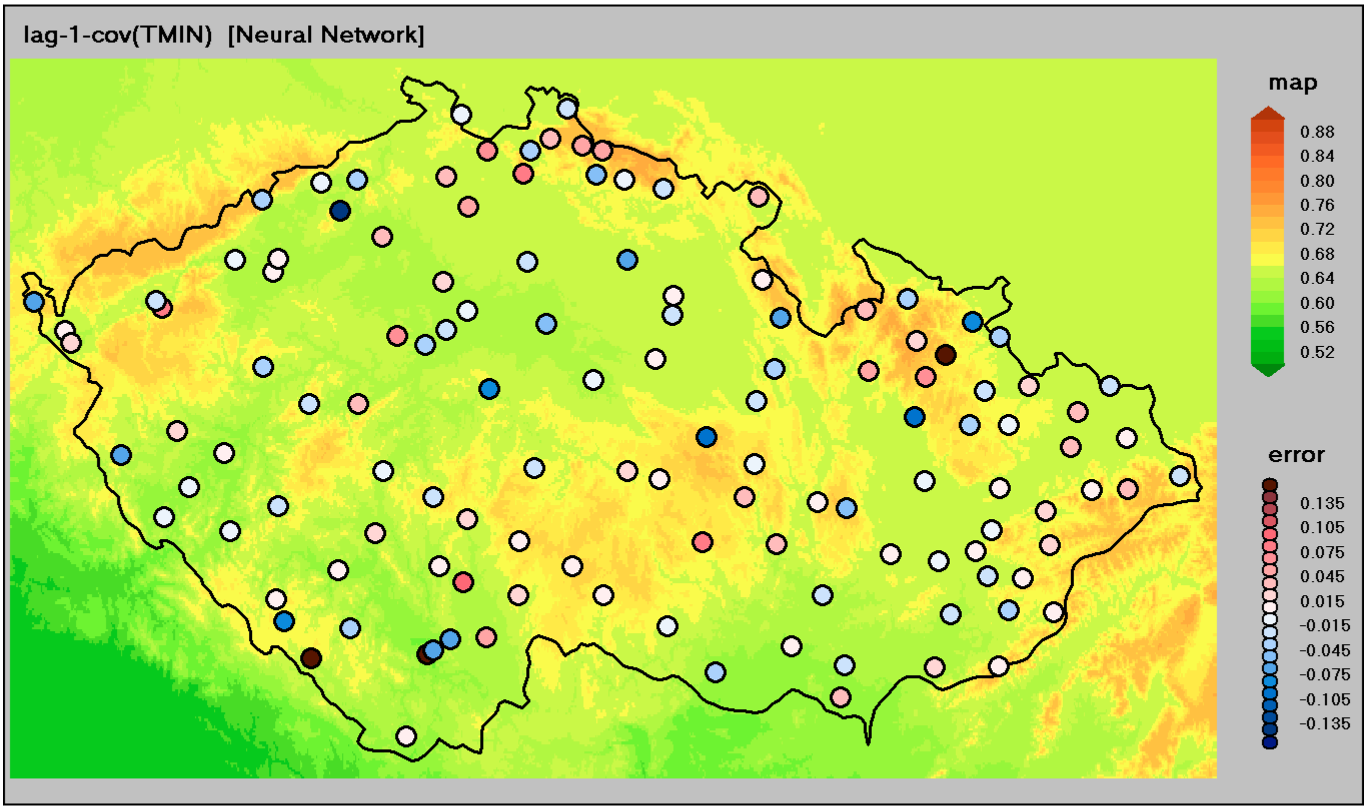
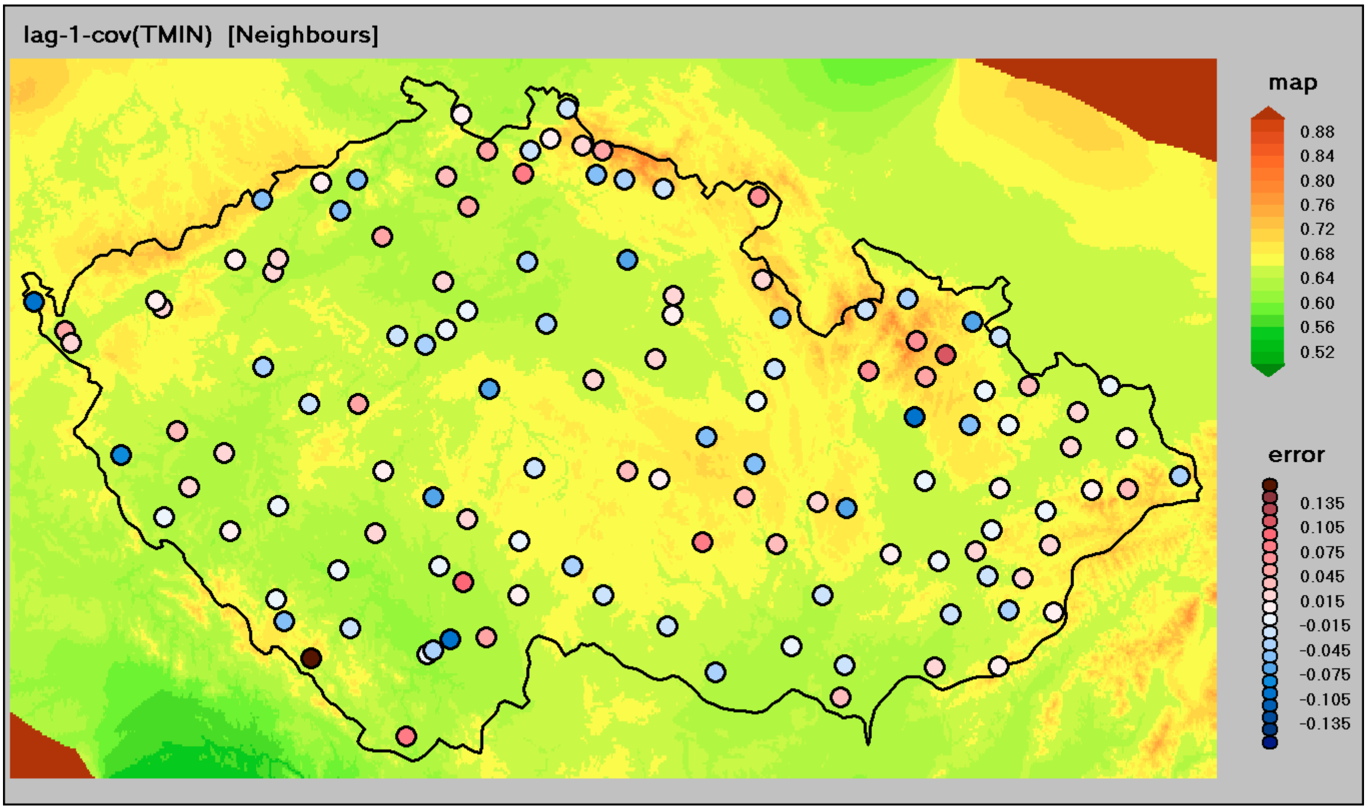
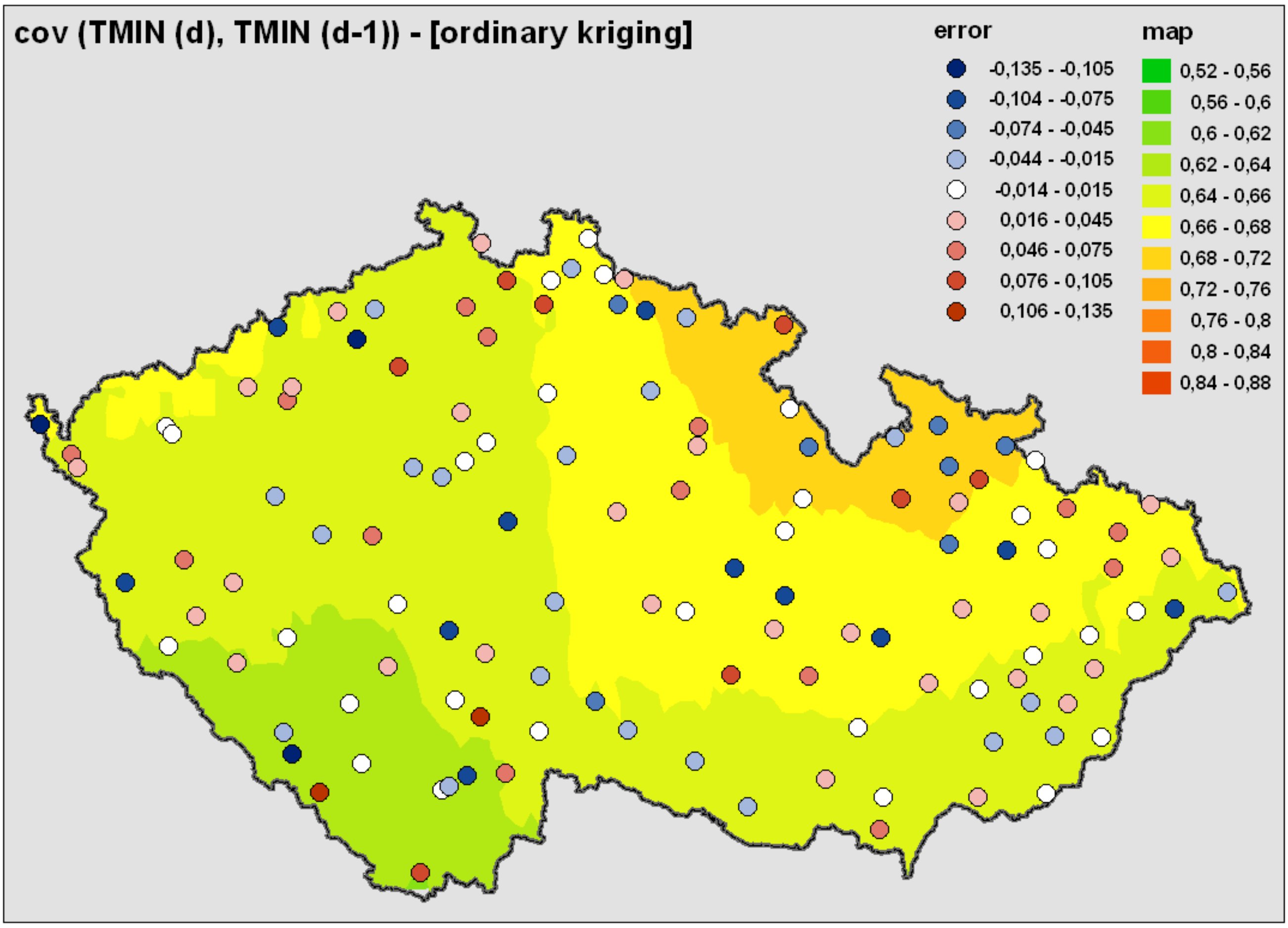

cov (TMAX (d), TMAX (d-1)) - [universal kriging]
error
-0,135 - -0,105
-0,104 - -0,075
-0,074 - -0,045
-0,044 - -0,015
-0,014 - 0,015
0,016 - 0,045
0,046 - 0,075
0,076 - 0,105
0,106 - 0,135
map
0,52 - 0,56
0,56 - 0,6
0,6 - 0,64
0,64 - 0,68
0,68 - 0,72
0,72 - 0,76
0,76 - 0,8
0,8 - 0,84
0,84 - 0,88

Fig. 9. cov(TMIN*(d), TMIN*(d-1))






cov (TMIN (d), TMIN (d-1)) - [ordinary kriging]
error
-0,135 - -0,105
-0,104 - -0,075
-0,074 - -0,045
-0,044 - -0,015
-0,014 - 0,015
0,016 - 0,045
0,046 - 0,075
0,076 - 0,105
0,106 - 0,135
map
0,52 - 0,56
0,56 - 0,6
0,6 - 0,62
0,62 - 0,64
0,64 - 0,66
0,66 - 0,68
0,68 - 0,72
0,72 - 0,76
0,76 - 0,8
0,8 - 0,84
0,84 - 0,88