Interpolation methods of weather phenomena

KAREL DEJMAL, VLASTIMIL KRATOCHVÍL
Department of Military Geography and Meteorology
University of Defense
Kounicova 65, 662 10 Brno
CZECH REPUBLIC
karel.dejmal@unob.cz http://www.unob.cz

Abstract: The military operations and activities are affected by weather conditions. Relevant information from area of interest is necessary. Because of irregularly distributed information the interpolation must be used. Two methods of interpolation were developed and they are introduced in this article. The methods have considered not only horizontal dependence but also vertical dependence of values. Their fruitfulness for interpolation of SYNOP observations in Central Europe was tested by cross-validation and backward interpolation. The usage of the methods for interpolation of climate data is also discussed.

Key-Words: - Interpolation, kriging, inverse distance weighting, verification, SYNOP report, least-square method.

1 Introduction

The weather affects people and human activities. The aim of the research was an impact of weather on military activities, such as traffic, health, usage and functionality of electronic equipment etc. ([4]-[7]). It is important to be able to obtain relevant information for area of interest from available weather data. Two different types of data are considered. The first type are real observed data, that are mostly very important in cases of military activity, e.g. temperature, precipitation etc. The second type are climate data that are important for planning and preparation phase of the activity. The main difference between these types is spatial variability (larger for current values in comparison with mean values).

The difference between interpolated quantities could be very significant because of their distribution and above mentioned spatial variability. We tried to find a relatively simple, fast and universal method for interpolation of surface meteorological quantities. The vertical dependence is obvious for most of quantities. Two interpolation methods were developed considering not only horizontal dependence but also possible vertical dependence. The vertical profile can be obtained by employing of theoretically derived relations, radiosonde measurements, climate profiles etc. These approaches cannot be used universally for different quantities. Vertical relationship was derived directly from analyzed data in the considered methods.

The final product was an interpolation to regular grid with horizontal resolution of approximately 5 km. The cross validation was used for verification of interpolation methods. The value for each station was interpolated from all other stations and this value was compared to the real value. The considered quantities were temperature, wind, relative humidity, visibility and amount of precipitations from synoptical station (SYNOP report)

The backward interpolation was tested as well. The values were interpolated into regular grid with horizontal resolution about 30 km. Values for station positions were back interpolated from grid values. The effect of grid distance was tested for selected quantities on both the 10 km and 15 km grids.

2 Interpolation methods

The geographic projection was defined by the following equations:

$$x = r \cdot (lat - lat0)$$
$$y = r \cdot \cos(lat) \cdot (lon - lon0)$$

where *lat* is latitude, *lon* longitude, *r* radius of the Earth (6371.1 km), *lat0* and *lon0* reference latitude (49.5 N) and longitude (15.0 E).

The area of interest covering the Czech Republic and adjacent territory was selected with respect to data availability and aspiration to analyze the accuracy for different types of relief. In Fig. 1 white represents the lowlands and black represents highest places (High Tatras, Alps). The altitudes of stations varied from 116 m to 2635 m above MSL and average distance between stations were approx. 50 km.

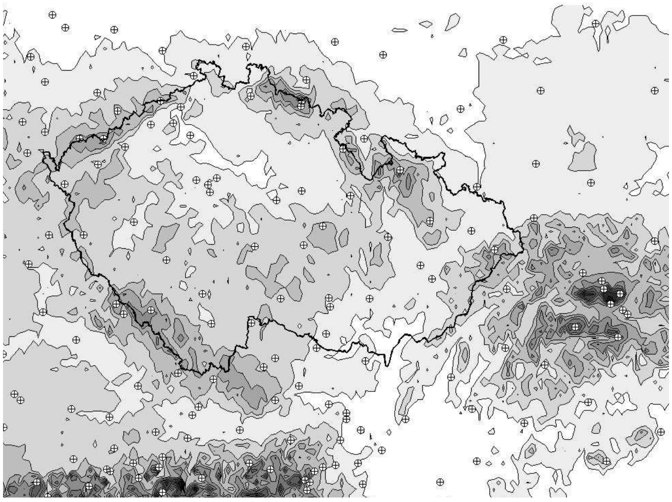

Fig. 1: Area of interest - orography

The first method (M1) was inspired by approach described in [3]. The method consists of the following steps:

1) The vertical gradient was calculated by means of three-dimensional first-order polynomial approximation using the least-square method. The gradient was calculated for stations with minimum of 5 other stations in the vicinity of 300 km. The elimination of gradient that differed more then 4-σ from mean value was found useful.
2) The vertical gradient was interpolated by inverse distance weighting (IDW) for those stations where it could not be derived. The weight decreases with square of distance.
3) The vertical recalculation of analyzed values to the selected referential level was performed by using these vertical gradients. The median of considered station altitudes (mostly about 360 m above MSL) was used as a reference value.
4) Vertical gradient was interpolated into points of interest (using IDW). It was also possible to reduce the interpolated gradient values to defined range.
5) Values in reference level were interpolated into points of interest (using IDW).
6) Afterwards the values corresponding to the real (model) orography were calculated. The analyzed values can be reduced to prescribed interval (e.g. 0 % ÷ 100 % for relative humidity).

The second method (M2) is described in following paragraphs. Position of each point in space of dimension *D* is given by vector

$\boldsymbol{x}_i^{\mathrm{T}} = \left({}^1x_i \quad {}^2x_i \quad \ldots \quad {}^Dx_i\right)$ for $i = 1, \ldots, n$ (meteorological stations),

$\boldsymbol{x}_j^{\mathrm{T}} = \left({}^1x_j \quad {}^2x_j \quad \ldots \quad {}^Dx_j\right)$ for $j = 1, \ldots, m$ (general points, grid points).

The measured values z_i are known for each station. The values z_i are considered to be a function of position ***x***. The measured (functional) values are organized into the vector $\boldsymbol{z}^{\mathrm{T}} = [z_1 \quad \ldots \quad z_n]$. The goal is to determine the vector of estimate values $\hat{\boldsymbol{z}}^{\mathrm{T}} = [\hat{z}_1 \quad \ldots \quad \hat{z}_m]$ for general points.

Linear regression in form following form is used at the first step

$$z = a_0 + a_1{}^1x + a_2{}^2x + a_3{}^3x + r_z .$$

The equation could be expressed in matrix convention for *n* known points (meteorological stations).

$$\boldsymbol{z} = \boldsymbol{X}\,\boldsymbol{a} + \boldsymbol{r}_z ,$$

where

$$\boldsymbol{X} = \begin{bmatrix} 1 & {}^1x_1 & {}^2x_1 & {}^3x_1 \\ 1 & {}^1x_2 & {}^2x_2 & {}^3x_2 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & {}^1x_n & {}^2x_n & {}^3x_n \end{bmatrix}.$$

Vector of parameters ***a*** of linear regression is derived by use of least square method solving the equation

$$\boldsymbol{a} = \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{z} ,$$

which fulfils condition $\boldsymbol{r}_z^{\mathrm{T}}\boldsymbol{r}_z \overset{def}{=} \min$.

Using the linear regression function in the second step the values $\boldsymbol{z}_0$, $\left(\boldsymbol{z}_0^{\mathrm{T}} = [z_{01} \quad \ldots \quad z_{0n}]\right)$ are computed for known meteorological stations and for general points of interpolation $\overline{\boldsymbol{z}}_0$, $\left(\overline{\boldsymbol{z}}_0^{\mathrm{T}} = [\overline{z}_{01} \quad \ldots \quad \overline{z}_{0m}]\right)$.

The purpose of linear regression is to reduce dependence of the functional values *z* on the coordinates. Difference of function values and values derived by linear regression on known points is denoted $\Delta z_i = z_i - z_{0i}$.

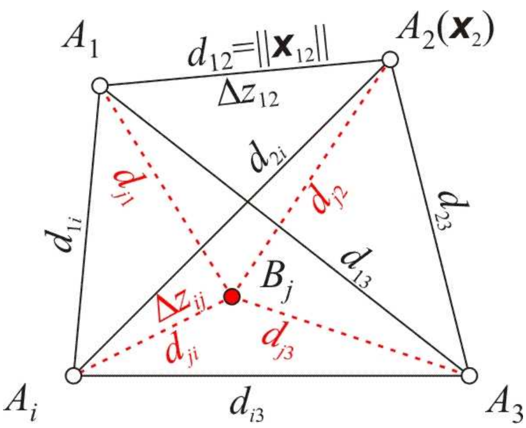

Fig. 2: The principle of selection of distances and differences of functional values

Estimation of functional values on the general point is then calculated by

$\hat{z}_j = \hat{z}_{0j} + \Delta\hat{z}_{0j}$,

where

$\Delta\hat{z}_{0j}$ is the correction of values,

$\hat{z}_{0j}$ is the first approximation computed by use of linear regression.

Correction is computed using ordinary kriging algorithm. Kriging is generally based on the assumption that the functional value of a phenomenon is a function of the same type of phenomena in the vicinity and can be calculated using the weighted average. In our case

$$\Delta\hat{z}(\boldsymbol{x}_j) = \sum_{i=1}^{n} p_{ij}\, \Delta z(\|\boldsymbol{x}_{ij}\|) \text{ for } j = 1, \ldots, m,$$

where

$\boldsymbol{x}_j \ldots$ location of general point;

$\|\boldsymbol{x}_{ij}\| = d_{ij} \ldots$ distance between the meteorological station and general point, $p_{ij} \ldots$ weight of difference Δz and

$$d_{ij} = \|\boldsymbol{x}_{ij}\| = \sqrt{\sum_{K=1}^{D} ({}^{K}x_j - {}^{K}x_i)^2}\ .$$

Let us suppose that the weight of difference is inversely proportional to the variance of function (measured) values, ie.

$$p_{ik} = \frac{const.}{\mathrm{var}(\Delta z_{ik})} \text{ with requirement that } \sum p_{ik} = 1$$

The first step is to calculate the values Δz_{ik} for each combination of meteorological stations.

For the same combinations of stations the distances d_{ik} are calculated. The difference Δz_{ik} can thus be assigned to any distance. Lengths and associated differences are sorted into non-descending sequence and divided to the appropriate number (*q)* of groups. For each group we calculate the average length $\bar{d}$ and average variance $\bar{s}_{ik}$;

$$\bar{d} = \frac{1}{c}\sum d_{ik}\ ;$$

$$\bar{s} = \mathrm{var}(\Delta z_{ik}) = \frac{1}{c}\sum \left(\Delta z_{ik} - \tfrac{1}{c}\sum \Delta z_{ik}\right)^2,$$

where

c is the number of lengths (differences) in a total of *q* groups.

A total of *q* average lengths and variances is received. Assuming that the variance is a function Φ of distance, it is possible to write for each group of sorted lengths

$(\bar{s}_{ik})_c = (\Phi(\bar{d}_{ik}))_c$ for $c = 1, \ldots, q$,

but also $\upsilon_{ij} = \Phi(d_{ij})$ for $i, k = 1, \ldots, n;\ j = 1, \ldots, m$,

where

υ_{ij} is variance computed as a function of the distance between the general and known points.

A linear dependence was considered for our solution. Parameters of linear regression for variance in the shape $\bar{s} = b_0 + b_1 \cdot \bar{d}$ were calculated using least-squares while assuming the following conditions:

1) variance must be non-negative,

2) variance is constant or increasing with distance.

The next step is calculation of variance estimates (from derived equation of linear function) for each combination of lengths, i.e.:

$s_{ij} = \Phi(d_{ij})$, respectively $\upsilon_{ik} = \Phi(d_{ik})$.

After this step all the values for calculations are known.

Ordinary kriging interpolation requires complying with the following matrix equation

$$\begin{bmatrix} s_{11} & \cdots & s_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ s_{n1} & \cdots & s_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} p_1 \\ \vdots \\ p_n \\ m_1 \end{bmatrix} = \begin{bmatrix} \upsilon_1 \\ \vdots \\ \upsilon_n \\ 1 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{e} \\ \boldsymbol{e}^{\mathrm{T}} & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{p} \\ m \end{bmatrix} = \begin{bmatrix} \boldsymbol{v} \\ 1 \end{bmatrix}$$

For *m* interpolated values on the general points the previous equation can be written as

$$\begin{bmatrix} s_{11} & \cdots & s_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ s_{n1} & \cdots & s_{nn} & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nm} \\ m_1 & m_2 & \cdots & m_m \end{bmatrix} = \begin{bmatrix} \upsilon_{11} & \upsilon_{12} & \cdots & \upsilon_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ \upsilon_{n1} & \upsilon_{n2} & \cdots & \upsilon_{nm} \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

or

$$\underbrace{\begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{e}_n \\ \boldsymbol{e}_n^{\mathrm{T}} & 0 \end{bmatrix}}_{\boldsymbol{\Gamma}}{}_{(n+1,n+1)} \underbrace{\begin{bmatrix} \boldsymbol{P} \\ \boldsymbol{m}^{\mathrm{T}} \end{bmatrix}}_{\boldsymbol{Q}}{}_{(n+1,m)} = \underbrace{\begin{bmatrix} \boldsymbol{V} \\ \boldsymbol{e}_m^{\mathrm{T}} \end{bmatrix}}_{\boldsymbol{G}}{}_{(n+1,m)}$$

where

s … variances calculated from the known (measured) values at meteorological stations,

υ … variances in the general points calculated from the linear regression function,

m … Lagrange multipliers.

Weight matrix is calculated by formula

$$\boldsymbol{Q} = \boldsymbol{\Gamma}^{-1}\boldsymbol{G}\ .$$

We obtain a total of *m* weight vectors $\boldsymbol{p}$, that consist from *n* elements. Matrix $\boldsymbol{P}$ can be written also in the form

$$\boldsymbol{P} = \lfloor \boldsymbol{p}_1 \quad \dots \quad \boldsymbol{p}_j \quad \dots \quad \boldsymbol{p}_m \rfloor,$$

where

$$\boldsymbol{p}_j^{\mathrm{T}} = \lfloor p_{j1} \quad \dots \quad p_{jn} \rfloor.$$

Corrections of functional values are then calculated from the formula

$$\Delta \hat{z}_j = \boldsymbol{p}_j^{\mathrm{T}} \Delta \boldsymbol{z} .$$

Interpolated value of the function in the general point equals to

$$\hat{z}_j = z_{0j} + \Delta \hat{z}_j .$$

Estimated standard deviation of correction $\sigma_{\Delta\hat{z}}$ is computed according to

$$\boldsymbol{\Sigma}_{\Delta\hat{z}} = \boldsymbol{QG} .$$

The main diagonal of matrix $\boldsymbol{\Sigma}_{\Delta\hat{z}}$ contains the estimates of variances $\sigma^2_{\Delta\hat{z}_j}$ of interpolated differences Δz_j in general points B_j.

Described interpolation method is suitable for interpolating variables whose change is continuous, without anomalous values.

The normalization of parameters (coordinates) improves using of the method in a number of factors:

- ✓ improves numerical stability of interpolation algorithm (regardless of the selected type of interpolation);
- ✓ parameters (coordinates), from which the euclidean distance *d* is calculated, have not the physical dimension and their number may be changed (space dimension *D* is not limited);
- ✓ normalized parameters partly respect importance of particular parameter for the interpolated value.

Along with ordinary kriging method the same functional values were calculated using algorithm based on radial basis functions (RBF interpolation). Both methods provide almost identical results. The method of Kriging was selected for possibility to calculate the accuracy estimates of interpolated values.

Both described methods were compared with IDW method which did not consider vertical relation (M1_0g). Some modifications of described methods were also used in tests. The list and explanation of tested methods is described below:

M1	described method No. 1
M1_L	M1 results are modified to be in interval <D;H>. All values smaller than D are set to D and values greater than H to H. Rest of values are unchanged.
M1_0g	application of IDW in 2D (without consideration of vertical dependence) with the same weight function as for M1
M1_Q	M1, where vertical gradients are derived only from the vicinity of 100 km
M1_Q_L	M1_Q results are modified to be in interval <D;H>.
M2	described method No. 2
M2_L	M2 results are modified to be in interval <D;H>.

3 Verification of methods

The methods were tested for directly observed data from main synoptic terms (0, 6, 12 and 18 UTC) from August 2008 till May 2009 Data from several stations were for some cases missing. The interpolation was performed for cases with minimum of 50 available values (from 150 possible). The considered quantities were:

- Temperature
- Precipitation
- Wind
- Relative humidity
- Visibility
- Cloud base.

The following characteristics were used in verification:

- the root mean square error (RMSE)

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(A_i - O_i)^2}$$

- systematic error (AVG),

$$AVG = \frac{1}{N}\sum_{i=1}^{N}(A_i - O_i)$$

- maximum overestimate (MAE), $MAE = \max(A_i - O_i)$, for all *i*,
- maximum underestimate (MIE), $MIE = \min(A_i - O_i)$, for all *i*,
- 2.5% quantile of difference (Q1),
- 97.5% quantile of difference (Q2).

The contingency table for threshold value and derived statistical characteristics were considered as criteria of dangerous phenomena detection. The contingency table contains number of real observed/unobserved phenomena (first index in matrix element notation, i.e. 1/0) in case of analyzed/unanalyzed phenomena (second index in matrix element notation, i.e. 1/0). Following

derived characteristics are considered:
- equitable score (EQS),
  $EQS = a00 / (a00+a01) + a11 / (a11+a10) - 1$;
- false alarm (FAR),
  $FAR = a01 / (a01 + a11)$;
- probability of detection (POD)
  $POD= a11 / (a11 + a10)$;
- bias (BIAS)
  $BIAS= (a01 + a11) / (a11 + a10)$;

The characteristics were always derived from all available measurement. Therefore stations with smaller data amount affected results less then stations having larger data set.
In order to study vertical dependence of interpolation accuracy the stations were divided into 6 groups. The groups with their notation are in Table 1. The seasonal difference of accuracy was also the object of study.

Table 1.: Group of stations

Elevation [m]	**Notation**	**Number**
< 200	D200	20
201-350	D350	41
351-500	D500	27
500-700	D700	29
701-1000	D1000	17
> 1000	N1000	15

4 Results

The statistical characteristics for temperature analysis are stated in Table 2. The most accurate method is M2 followed by M1. The resulting error is less then 3 °C for 95% cases. The RMSE values are about 1.3 °C and it is an evident improvement in comparison with M1_0g. Vertical gradients from method M1_Q vary in space more than M1 and it can lead to bad forecast (maximum error greater than 56 °C) and worse RMSE values.
The accuracy decreases with elevation. The big error is evident in event of temperature inversions. The significant error can be found for M1 in lowland (for example in Hungary) where the elevation of station does not differ too much and when we use determined gradient for station whose elevation differs significantly.
The analyses are better at 18 and 12 UTC in comparison with night or morning (0 and 6 UTC). However, the accuracy also varies in different months. The most accurate results come from March, the worst results from April (RMSE for M2 about 2.0 °C). This could be connected to higher variability or inversion presence in this month, but the further study is needed.

Table 2. : Temperature interpolation accuracy [K]

	RMSE	MAE	MIE	AVG	Q1	Q2
M1_Q	1.7	56.4	-29.5	0.0	-3.0	3.2
M1_0g	2.7	16.5	-15.8	-0.2	-5.4	6.8
M1	1.4	13.8	-13.3	0.0	-3.0	3.0
M2	1.3	11.5	-12.9	0.0	-2.7	2.9

As regards to relative humidity the interpolation was made using the same methods. The interpolated values should be inside of interval 0% - 100%. Therefore the results of methods were also limited (0% for all negative values, 100% for all values greater than 100%).
The RMSE values are presented in Table 3 for selected methods and considered group of stations. The best results were provided again by M2 followed by M1. The interpolation is better for stations with lower elevation again.
The analyzed values from method M1_Q often lay outside a physically possible range mainly for the highest stations (compare M1_Q and M1_Q_L). The analyzed values differ more then about 100% from possible values. From this point of view the method M1 is better, however in some cases producing the value 140% or negative values. The M2 never produced negative value and very rarely values significantly greater than 100%. The method M2 does not interpolate outside of observed values very often, which is the main difference in comparison with M1. When we want to develop universal method this is convenient property.

Table 3. : Relative humidity interpolation RMSE [%]

	M1	M2	M1_0g	M1_Q	M1_Q_L
D200	6.4	6.5	6.6	6.3	6.3
D350	7.2	7.1	8.2	7.2	7.1
D500	7.3	7.0	8.5	6.9	6.9
D700	8.1	8.5	9.3	7.9	7.9
D1000	9.8	9.8	11.7	9.4	9.4
N1000	14.4	12.4	18.6	21.8	17.9
ALL	8.6	8.3	10.3	10.0	9.1

The wind velocity was interpolated by two techniques. The first technique was direct velocity interpolation. The second technique was a separate interpolation of meridional and zonal part of velocity and the final velocity was computed from these results. Not only velocity but also direction is obtained using the second technique. However, the disadvantage is exclusion of measurements without defined direction (for example variable wind). For enabling of comparison of these two techniques these data were discarded. The results for considered group of stations are presented in Table 4 and Table 5.

Table 4. : Wind velocity RMSE [m/s] – direct velocity interpolation

	M1	M2	M1_0g	M1_Q
D200	1.7	1.8	1.9	1.7
D350	1.8	1.7	2.1	1.7
D500	1.5	1.4	1.7	1.4
D700	1.8	1.8	2.3	1.7
D1000	3.1	2.9	3.8	2.9
N1000	4.4	4.4	5.8	5.8
ALL	2.3	2.3	2.9	2.6

Table 5. : Wind velocity RMSE [m/s] – composition of meridional and zonal wind

	M1	M2	M1_0g	M1_Q
D200	1.8	1.8	1.9	1.8
D350	1.8	1.7	2.0	1.7
D500	1.5	1.5	1.6	1.4
D700	1.9	1.7	2.2	1.7
D1000	3.1	2.9	3.7	3.0
N1000	4.7	4.2	6.1	5.8
ALL	2.4	2.3	2.9	2.6

No significant difference between used techniques was found. The methods M1 and M2 are comparable and it is evident improvement in comparison with M1_0g. The improvement of method M1_Q is not evident.

The RMSE value is significantly larger for stations with higher elevation. This is caused by frequent values underestimation. The RMSE values are larger in winter than in summer months.

The zonal and meridional part of wind vector interpolation accuracy is stated in Table 6. There is not significant difference between zonal and meridional part of velocity. The best results came from M1 and M2 again.

Table 6. : RMSE for velocity

	Zonal velocity [m/s]	Meridional velocity [m/s]
M1_Q	2.7	2.5
M1_0g	2.6	2.7
M1	2.3	2.3
M2	2.2	2.3

The accuracy of visibility interpolation was assessed by use of statistical characteristics EQS, FAR, BIAS and POD that were determined for four limits of visibility – 0.5, 1, 3 and 5 km.

Interpolated values were often overestimated. The EQS (Table 7.) increases with threshold value. The best results are provided by method M2, whilst the worst method is M1_0g which is mainly caused by small values of POD of this method. Nevertheless the results of all methods are worse than desired, FAR (0.3; 0.5), POD (0.2; 0.5). Large space and time variability of visibility appeared to be a major problem. The variability can not be affected by available measurement density. The additional information is probably needed for significant improvement.

Table 7. : Statistical characteristics for visibility

Char.	Thre. val.	M1	M2	M1_Q	M1_0g
EQS	0.5 km	0.19	0.24	0.24	0.02
	1 km	0.21	0.28	0.25	0.05
	3 km	0.31	0.41	0.34	0.21
	5 km	0.42	0.50	0.45	0.34
FAR	0.5 km	0.55	0.38	0.51	0.55
	1 km	0.51	0.35	0.46	0.53
	3 km	0.39	0.33	0.37	0.41
	5 km	0.29	0.28	0.29	0.34
POD	0.5 km	0.21	0.25	0.26	0.02
	1 km	0.23	0.30	0.27	0.05
	3 km	0.35	0.45	0.38	0.23
	5 km	0.47	0.56	0.50	0.39
BIAS	0.5 km	0.47	0.41	0.52	0.05
	1 km	0.47	0.46	0.51	0.11
	3 km	0.57	0.66	0.61	0.40
	5 km	0.67	0.79	0.71	0.59

The 24 hour precipitation interpolation was verified. The method M1 according to RMSE gives the best results (Table 8). But the difference between all considered methods is not significant. Methods fail mainly for highest stations.

The ability to analyze limit values of precipitation was studied. As the limit values were 1, 10 and 20 mm. The computed characteristics are again very similar for all considered methods. The EQS values decrease from 0.8 for 1 mm threshold to 0.4 for 20 mm threshold, the FAR values from 0.2 to 0.4 and POD from 0.9 to 0.4. The precipitation analyses are better than visibility analyses but the additional information from other sources (radar measurement, etc.) could improve analyses significantly.

Table 8. : Daily precipitation RMSE [mm]

	M1	M2	M1_Q	M1_0g
D200	2.0	2.1	2.0	2.1
D350	2.2	2.1	2.2	2.4
D500	2.2	2.3	2.2	2.3
D700	2.3	2.4	2.2	2.5
D1000	2.8	3.1	2.8	2.8
N1000	5.4	6.2	7.4	5.6
ALL	2.8	3.0	3.3	2.9

To simplify the problem the cloud base was processed without considering the amount of clouds. Computed characteristics for two threshold values, 300 m and 600 m above surface, are in Table 9. Comparing to visibility, these results are better. The best results are provided by method M2 but other methods are comparable and a little better than method M1_0vg. The method M1_0vg fail for groups of station N1000 and D1000 and threshold value 300 m. In this case the EQS is lower than 0.15. Generally for all methods, results for stations with lower elevations are better.

Table 9 : Statistical characteristics for cloud base analysis

		M1	M2	M1_Q	M1_0g
EQS	**300 m**	0.40	0.46	0.42	0.34
	600 m	0.56	0.56	0.56	0.51
FAR	**300 m**	0.39	0.44	0.41	0.42
	600 m	0.31	0.33	0.32	0.34
POD	**300 m**	0.46	0.54	0.49	0.40
	600 m	0.71	0.73	0.71	0.67
BIAS	**300 m**	0.75	0.97	0.83	0.69
	600 m	1.02	1.09	1.04	1.02

But the errors could be very large. The overestimation of cloud base, which is common for all methods, evidently grows with station elevation. Great error for highest stations in winter period could be caused by presence of low inversion cloudiness. The value interpolated from surrounding of stations with lower elevation (below inversion) is low but in reality there are only higher clouds or there is no cloudiness at all. The cloud base analysis can be affected by observation accuracy (similar as for visibility). Most observations are provided by equipment with possibility of human correction but for some stations and time the raw data from equipment are reported. The problem is also with accuracy and representativity of measurement. The amount of cloudiness could be included into analysis to receive more usable results.

5 Backward interpolation

Interpolation error is dependent on the density of stations. The distance to which must be interpolated is larger in cross-validation with comparison to real density station because of omitting station. The resulting errors are therefore larger than would correspond to an interpolation error using all stations. The following approach was chosen for estimate of this effect:

1) Values of all stations were interpolated to a regular network.
2) The interpolated values were interpolated back to point of observation.

The grid resolution 30 km was considered (corresponds to the order of the average distance between stations). Further tests were carried out with finer grid resolution 10 km and 15 km. The analyses were computed only for two months. The first was January 2009, which represents cold period and the second August 2008, which represent warm part of year.

The influence of grid distance is discussed firstly. This experiment is carried out only for the M2 for the temperature range in the months of August and January. Test results for temperature are in the Table 10 and for the visibility in Table 11. For two months we get a similar increase in the RMSE values with grid distance increase. Considering all the stations irrespective of altitude, the values increase with the square root of grid distance (for temperatures slightly faster than for visibility). The accuracy obtained in cross-validation is comparable with grid resolution approximately 80 km if this dependence was considered. This roughly corresponds to twice the distance of the considered stations.

For other methods and variables was performed backward interpolation only for the grid distance 30 km. But the backward interpolation for this net resolution was computed for all methods as cross validation. The results of interpolation are, as expected, better than of cross-validation. Improvement for each station is strongly influenced by distance from the grid point. For all methods except for M2 (and its variants) is the same observed and analyzed value (for all points with observation), therefore accurate.

For the best cross-validation methods M2 and M1 backward interpolation is usually worse than the other methods. This fact may be due to the fact that the resulting fields are strongly influenced by nearest stations. The resulting field is then smoother and absents in him great extremes and therefore the better results are obtained for cross-validation. The nearest station is crucial for other methods. Therefore, in cross validation tests are better, but again failed to interpolation to points farther from observation points.

Table 10: RMSE for backward interpolation and cross validation for temperature and method M2 [°C]

January 2009				
Group\grid	**10 km**	**15 km**	**30 km**	**Cross-val.**
D200	0.4	0.5	0.7	1.4
D350	0.4	0.5	0.9	1.5
D500	0.3	0.5	0.8	1.3
D700	0.6	0.7	1	1.6
D1000	0.7	0.9	1.4	1.8
N1000	1.2	1.7	2.2	1.8
All	0.6	0.8	1.1	1.6
August 2008				
Group\grid	**10 km**	**15 km**	**30 km**	**Cross-val.**
D200	0.3	0.4	0.6	1.2
D350	0.3	0.4	0.7	1.1
D500	0.2	0.4	0.6	1,0
D700	0.4	0.5	0.8	1.4
D1000	0.6	0.7	1.1	1.5
N1000	0.9	1.2	1.6	1.7
All	0.5	0.6	0.9	1.3

Table 11: RMSE for backward interpolation and cross validation for visibility and method M2 [km]

January 2009				
Group\grid	**10 km**	**15 km**	**30 km**	**Cross-val.**
D200	1.7	2.3	3.4	7.0
D350	1.7	2.2	3.5	7.1
D500	2.1	2.9	4.5	8.6
D700	2.3	3.3	5.1	10.2
D1000	5.3	6.3	9.7	13.1
N1000	14.1	16.8	22.5	22.3
All	5.2	6.3	8.7	11.1
August 2008				
Group\grid	**10 km**	**15 km**	**30 km**	**Cross-val.**
D200	3.9	5.0	7.6	1.7
D350	4.5	5.8	8.8	1.7
D500	3.4	5.1	8.1	2.1
D700	3.8	5.4	8.3	2.3
D1000	6.8	8.1	11.6	5.3
N1000	13.8	15.1	19.0	14.1
All	6.1	7.3	10.3	5.2

6 Other usage and examples

Some examples of final output of described methods will be introduced in this section. The methods have not been used only for quantities described in test but they have been used also for interpolation of climate data. These fields are usually slighter in comparison with term measurements and therefore the methods are relatively successful. The results of quantities were visually compared with results in climate atlas [8]. There was used bigger amount of stations and longer period (mostly all 30-year period). Structure of our results mostly is comparable with structure of "real" fields.

The example of temperature analysis is shown. The results of methods M1 and M2 are depicted in Fig. 2 and Fig. 3. The analysis produced by numerical weather prediction (NWP) model ALADIN is in Fig. 4. All tree fields are similar. Larger differences are in north east part of the area, where are no observations. The distribution of uncertainty as a product of method M2 is in Fig. 5. The dependence on stations distribution is obvious.

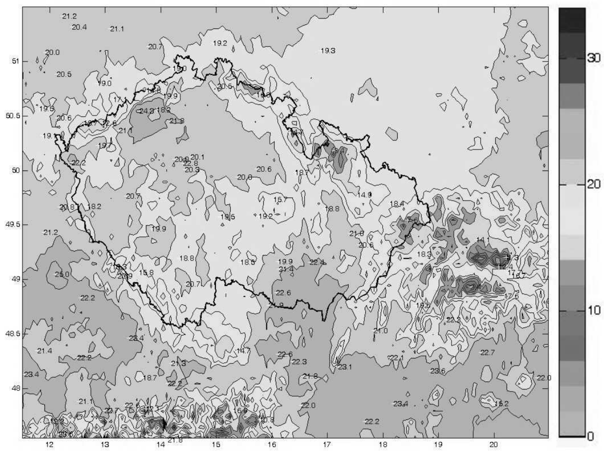

Fig. 2: Temperature analysis by method M1 (30th August 2008, 12 UTC) in °C.

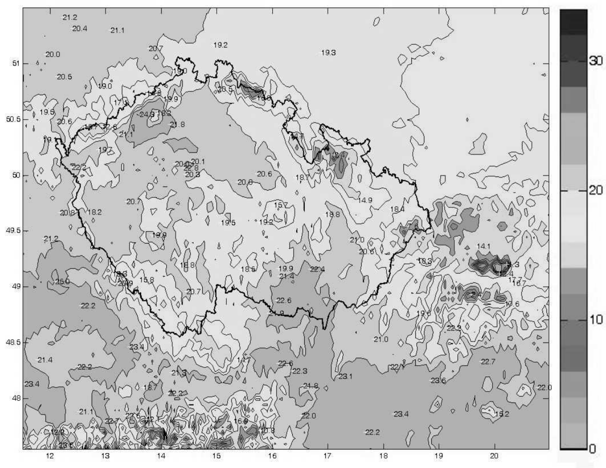

Fig. 3: Temperature analysis by method M2 (30th August 2008, 12 UTC) in [C].

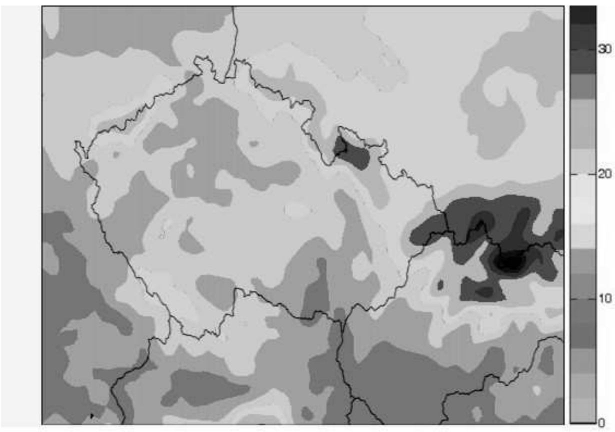

Fig. 4: Temperature analysis by NWP model ALADIN (30th August 2008, 12 UTC) in °C.

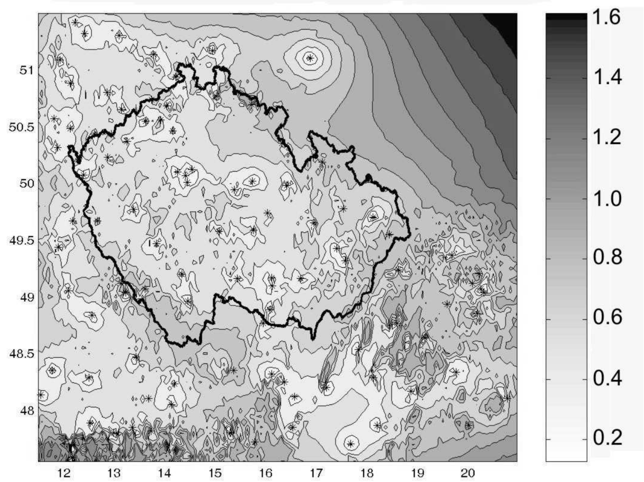

Fig. 5: Distribution of temperature analysis uncertainty by method M2 (30th August 2008, 12 UTC) in °C.

7 Conclusion

The tested methods M1 (based on inverse distance weight) and M2 (based on ordinary kriging) were used for interpolation of selected measured quantities. Not only horizontal dependence but also vertical dependence was considered in these methods. Both of them offer significant interpolation improvement for most of selected quantities in comparison with inverse distance method without consideration of vertical dependence. The method M2 does not extrapolate values too much outside of measured interval. That is not obvious for M1 and it can lead to big errors and physically nonsensical values.

The results of immediate temperature, wind, precipitation and relative humidity seem to be comparable with local NWP outputs. Unfortunately, the visibility interpolation is not sufficient. The methods seem to be usable for interpolation of climate data, which are smoother.

The correction of analyses could be made only manually by user. This possibility will be offered in software which use methods described in this article and which is currently being developed. The methods based on assimilation of the additional information are promising way for improvement. The NWP model outputs, radar data and model data could be included in future.

The significant motivation for interpolation was short range visibility forecast. Visibility is crucial quantity which affects military operations and which is not routinely forecasted by NWP models. Actual visibility is a good predictor. It seems to be better way to derive relation between visibility change and changes of others weather quantities rather than to forecast directly visibility from them. The persistent forecast could be often better then such forecast. The interpolation of time values of visibility has been shown to be unusable. The interpolation could be sometimes improved with satellite data enabling detection of areas with fogs. But there is limitation because these areas could be detected only in case without cloudiness. The other possibility, more promising, is combination with climate data. The field of selected monthly characteristic is smoother in comparison with measured data. The question is if average or median of values could be considered because of visibility distribution. The analyzed values would be climate fields corrected using measured values. The advantage is that the evident nonsense, e.g. for highest stations with high elevations, could be eliminated on climate data more easily than using time data.

8 Acknowledgement
This paper is a particular result of the defense research project "Meteor – geographic and meteorological battle-filed factors, their dynamic visualization a localization in command and control systems" based on statement Nr. 801 8 6020 R.

References:

[1] HEBÁK, P, HUSTOPECKÝ, J. *Vícerozměrné statistické metody s aplikacemi*. Praha : SNTL/ALFA, 1987. 456 p.

[2] JOHNSTON, K., et al. *Using ArcGIS Geostatical Analyst*. Redlands, CA : ESRI, 2001. 300 p. GIS by ESRI. ISBN 1-58948-006-6.,

[3] SOKOL, Z. and ŠTEKL, J., *3-D mesoscale objective analysis of selected elements from SYNOP and SYRED reports.* Meteorologische Zeitschrift, 1994, vol. 3, no. 4, p. 242-246.

[4] MAZILU, M., MARINESCU, R. C. *The Tourism and Climate Changes. Interferences*. WSEAS TRANSACTIONS on BUSINESS and ECONOMICS, 2008, vol. 5, no. 7, p. 393-402. ISSN 1109-9526.

[5] KU, M. Y., et. al. *Visual Motorcycle Detection and Tracking Algorithms.* WSEAS TRANSACTIONS on ELECTRONICS, 2008, vol. 5, no. 4, p. 121-131. ISSN 1109-9445.

[6] GHAZANFARI, S., et. al. *Evaluating the effects of UHI on climate parameters (A case study for Mashhad, Khorrasan)*. WSEAS TRANSACTIO NS on ENVIRONMENT and DEVELOPMENT, 2009, vol. 5, no. 7, p. 508-517. ISSN 1109-9526.

[7] MALLESHAM, G., JHA, A. N. *A fine load forecasting using neural networks and fuzzy neutral networks*. Proceedings of the 5th WSEAS Int. Conf. on Power Systems and Electromagnetic Compatibility, 2005 vol. 5, no. 7, p. 224-229.

[8] TOLASZ, R., et al. *Climate atlas of Czechia.* 1st edition. Praha : Český hydrometeorologický ústav, 2007. 255 p.

[9] PODEROUX, et. al. *Adaptive Hierarchical RBF Interpolation for Creating Smooth Digital Elevation Models*. IPARLA (LaBRI-INRIA) Project Université Bordeaux, 9 p., 2004 [cit. 2009-12-30] In <http://iparla.labri.fr/publications/2004/PTGG04/>

[10] TEN-LIN L., KAI-WEI J., DAR-YUAN L. *Interpolating Soil Properties Using Kriging Combined with Categorical Information of Soil Maps.* Published in Soil Sci Soc Am J 70:1200-1209. 2006 [cit. 2009-12-24] In http://soil.scijournals.org/cgi/content/abstract/67/6/1848.