# Comparison of interpolation methods for six-hour temperature data series


**Barna Zsófia [1], Izsák Beatrix [1], Pieczka Ildikó [2]**

**[1]Hungarian Meteorological Service, Kitaibel Pál Street 1, H-1024, Budapest, Hungary**
**[2]ELTE Eötvös Loránd University, Institute of Geography and Earth Sciences, Department of Meteorology, Pázmány Péter promenade 1/A, Budapest H-1117, Hungary**



**Abstract**: The application of the MISH (Meteorological Interpolation based on Surface Homogenized Data Basis) interpolation system in meteorological practice and its effectiveness are outstanding. Interpolation with the MISH system compared to ordinary kriging (OK) and inverse distance weighted (IDW) methods used in climate studies has a much smaller error value in the data series of all databases, so its use is much more recommended compared to other interpolation methods. Its reason is the fact that in meteorology we cannot assume that there is no spatial trend since the expected value differs in different landscape units due to the country's complex orography. A major advantage of the MISH system is that it uses homogenized station data series to model climate statistical parameters. When applying the MISH system to hourly values we used the daily model results and modified the spatial trend values using the regression coefficients between daily and hourly values.

Cross validation technique was used to compare different interpolation methods which is available in the Geostatistical Analyst toolbox of the ArcMap software. This software randomly divides the station system into training and test datasets. For each subset and each hourly value 150 interpolations were performed. After that we calculated the RMSE (root mean square error) values for all three methods. The results are shown in this poster.


## Mathematical background

*Meteorological variables:* $Z(s_0,t)$: unknown predictand, $Z(s_i,t)$: known predictors ($i=1,\dots,M$) where the location vectors $s$ are the elements of the given space domain and $t$ is the time. Assuming normal distribution (e.g. temperature) the additive (linear) formula is adequate: $\hat{Z}(\mathbf{s}_0,t) = \alpha_0 + \sum_{i=1}^{M} \alpha_i Z(\mathbf{s}_i,t)$ where $\sum_{i=1}^{M} \alpha_i = 1$ and $\alpha_0, \alpha_i$ interpolation parameters.

*Interpolation error:* The quality of interpolation can be characterized by the Root-Mean-Square Error (RMSE):

$$MSE(s_0) = E\left(\left(Z(s_0,t) - \hat{Z}(s_0,t)\right)^2\right) \Rightarrow RMSE(s_0) = \sqrt{MSE(s_0)}$$

*Inverse distance weighted (IDW) method:* IDW is a very simple method for interpolation, which considers the nearest neighbor points. The general formula is formed as a weighted sum of the data:

$$\hat{Z}_{IDW}(\mathbf{s}_0,t) = \sum_{i=1}^{M} \alpha_i \cdot Z(\mathbf{s}_i,t) \text{ where } \alpha_i = \frac{\left(\frac{1}{d_i}\right)^p}{\sum_{i=1}^{n}\left(\frac{1}{d_i}\right)^p} \text{ and } d_i\text{: distance and } 0 \le p \le 2$$

*Ordinary kriging (OK):* is the most general and widely used kriging method. Ordinary kriging estimates the value of a climatic variable at a point from its values at surrounding stations. The general formula is applied in the Ordinary kriging:

$$\hat{Z}_{OK}(\mathbf{s}_0,t) = \sum_{i=1}^{M} \alpha_i Z(\mathbf{s}_i,t) \text{ where } \sum_{i=1}^{M} \alpha_i = 1.$$

*Interpolating hourly values with MISH:* The interpolation formula (additive model) is applied to the daily mean temperature data:

$$\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^{M} \lambda_i \left(E(\mathbf{s}_0) - E(\mathbf{s}_i)\right) + \sum_{i=1}^{M} \lambda_i Z(\mathbf{s}_i)$$

where $Z(\mathbf{s}_0)$ (s: location) is a predictand, and $Z(\mathbf{s}_i)$ $(i = 1,\dots,M)$ are predictors (observations), and the weighting factors $\sum_{i=1}^{M} \lambda_i = 1$ and $\lambda_i$ $(i = 1,\dots,M)$ depend on the stochastic relations, while $E(\mathbf{s}_i)$ $(i = 0,\dots,M)$ are the daily spatial trend values.

Based on this, the following interpolation formula is applied to the $t = 0,6,12,18$ UTC values, accepting the weighting factors:

$$\hat{Z}(\mathbf{s}_0,t) = \sum_{i=1}^{M} \lambda_i \left(E(\mathbf{s}_0,t) - E(\mathbf{s}_i,t)\right) + \sum_{i=1}^{M} \lambda_i Z(\mathbf{s}_i,t) \qquad (t = 0,6,12,18)$$

where $E(\mathbf{s}_i,t)$ $(i = 0,\dots,M)$ represents the spatial trend values for the given dates.

To model the $E(\mathbf{s},t)$ $(t = 0,6,12,18)$ hourly spatial trend values, the following linear model was chosen:

$$E(\mathbf{s},t) = \alpha(t) + \beta(t) \cdot E(\mathbf{s}) \qquad (t = 0,6,12,18)$$

In this case, the interpolation formula for hourly values is given as:

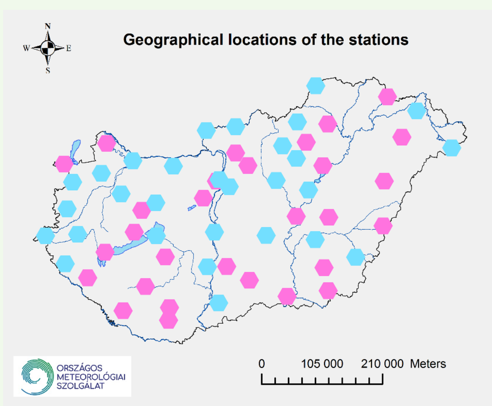
$$\hat{Z}(\mathbf{s}_0,t) = \beta(t) \cdot \left(\sum_{i=1}^{M} \lambda_i \left(E(\mathbf{s}_0) - E(\mathbf{s}_i)\right)\right) + \sum_{i=1}^{M} \lambda_i Z(\mathbf{s}_i,t) \qquad (t = 0,6,12,18)$$

Thus, the modelled daily spatial trend values $E(\mathbf{s}_i)$ $(i = 0,\dots,M)$ and the estimated $\beta(t)$ $(t = 0,6,12,18)$ hourly regression coefficients can be used to interpolate the hourly values (*Szentimrey*, 2019b).

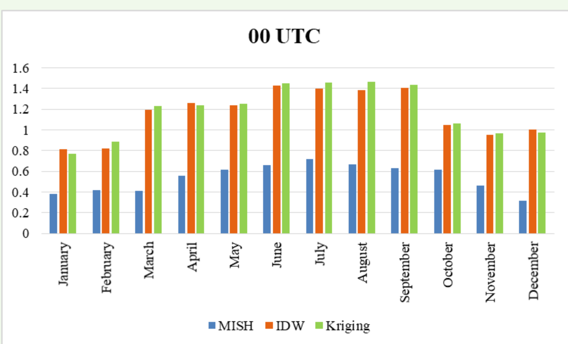
With the help of the ArcMap software, we randomly divided the 58 stations into 29 "training" and 29 "test" stations. The resulting 150 times division were interpolated 150 times using the monthly values of all four databases using ordinary kriging and then IDW method examining the RMSE values in each case. These results were compered to the results of applying the MISH system in general by selecting the two examples for each hourly value shown below for illustration. The decision was made according to the fact that we were looking for a random distribution of the station list during the examination in which the "training" stations - for which we perform the validation of the interpolation process - are mostly located in the Transdanubian region or int he Northern Mountains. This is also shown in all the figures, where pink are the "training" and blue are the "test" stations.

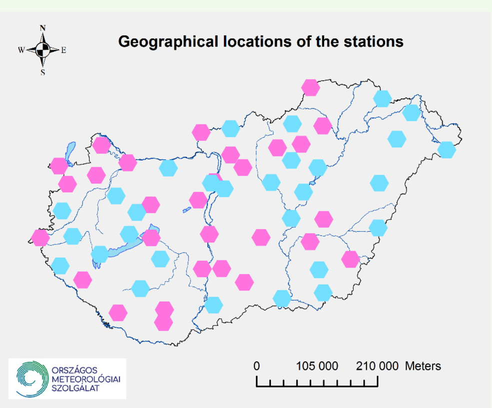
## Results

### 00 UTC

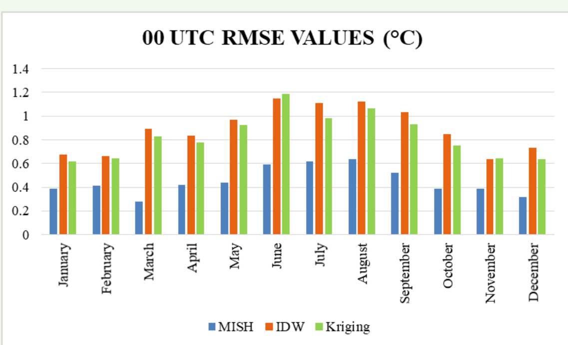



The left figure shows the station division in all cases, while the right side shows the RMSE values occurring during the interpolations. The upper case illustrates a division in relation to a given hourly value, where training stations are mostly found in Transdanubia, and in the lower figure, the training stations are located in the extensive area of the Northern Mountains.

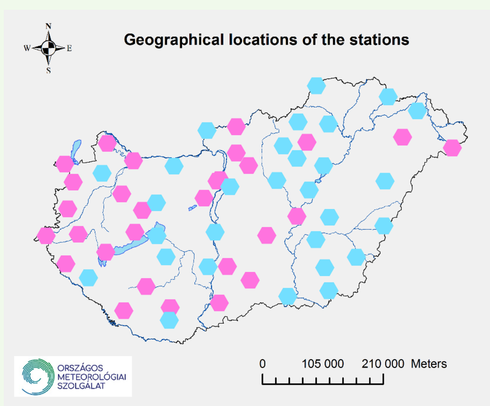


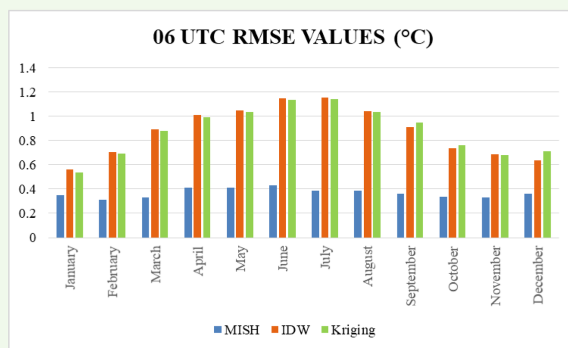

In this case, the extensive area of the Northern Mountains can be characterized by "training" stations. MISH produces the smallest error in March, which is 0.28 °C, and the highest error value is 0.64 °C in August. Compared to this even with the most accurate interpolation in connection with ordinary kriging, the RMSE was 0.62 °C in January, and the highest error value was obtained in June which is 1.18 °C.

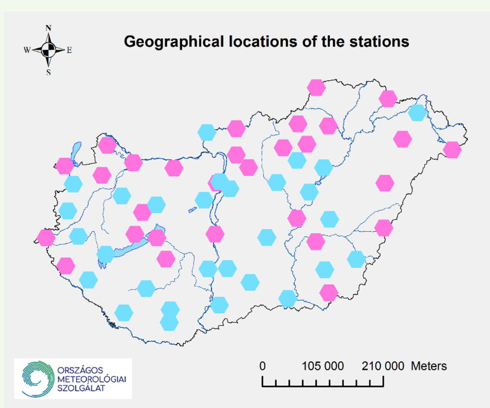



### 06 UTC

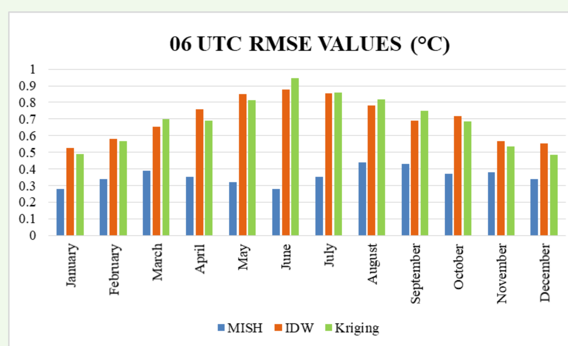



Firstly, we examined the case when most of the "teaching" stations are located in Transdanubia. It is clear for the kriging and inverse methods that the error values are the largest in summer. In comparison, IDW and kriging also produce the lowest errors in January, numerically 0.56 and 0.54 °C in this case. Besides in summer, these errors only increase even more, reaching and then exceeding 1 °C.

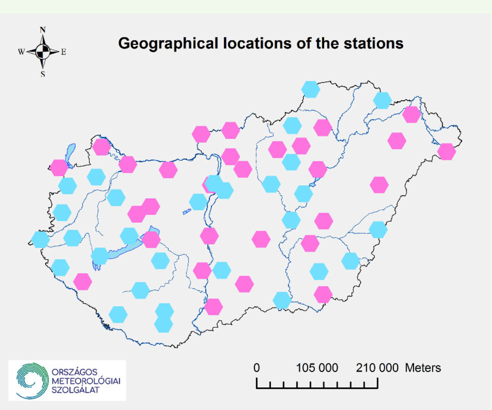



According to this division, the interpolation related to summer is characterized by lower RMSE values, not reaching 1 °C when examining each method. Thus, examining this station assignment, there are no such high appearing error values as if the majority of the stations used for the validation of the interpolation process were located in the Transdanubian region. According to this distribution, kriging has the highest error with a value of 0.95 °C in June.

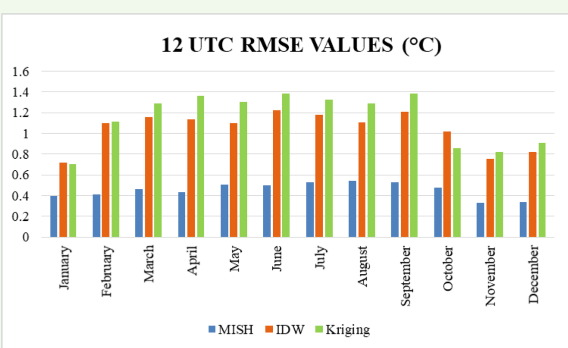


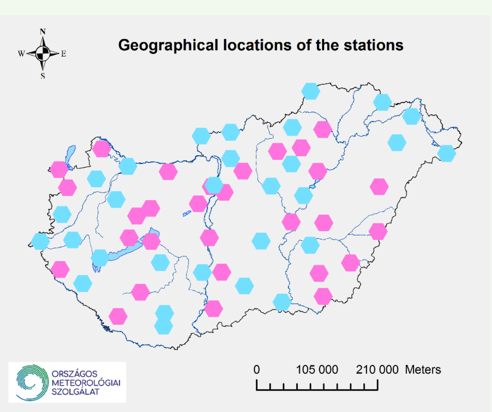

### 12 UTC



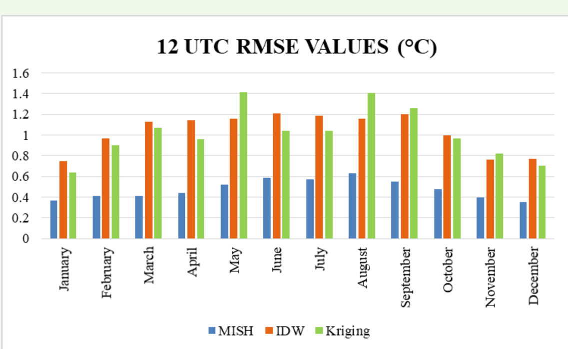

The error of MISH in this simulation is practically 1/3 of the error of IDW and OK. The highest RMSE using MISH over the 12 months is only 0.54 °C. Compared to this, if August is examined further, the RMSE value for this month is 1.29 °C in connection with kriging, and 1.11 °C in the case of IDW.

The common procedures during the two meteorological tasks give errors above 1 °C for a significant part of the year, while with the use of MISH this value is 0.6 °C throughout, but below 0.4 °C during several months.

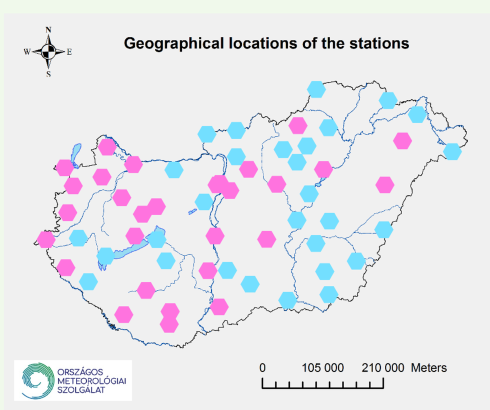


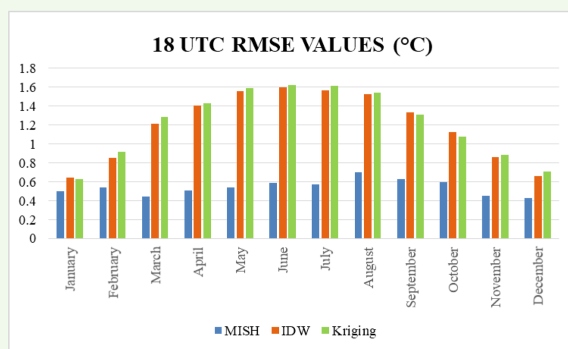

According to this division, lower errors occur in the summer season using the chosen interpolation techniques than in the case of 12 UTC above. The biggest error comes from ordinary kriging in May, which is 1.413 °C.

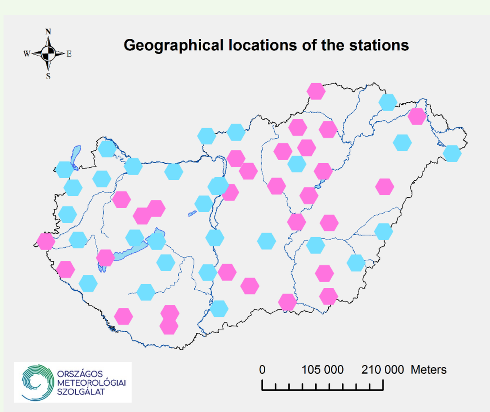



### 18 UTC

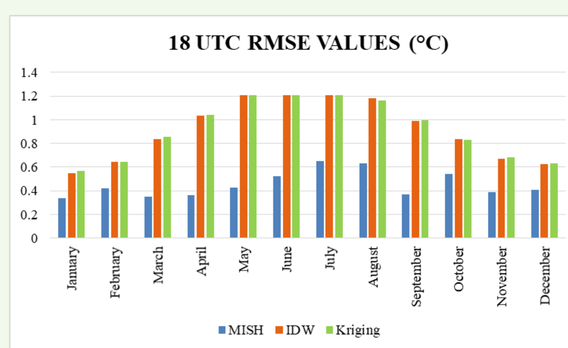



In this case during spring and summer, IDW and OK were interpolated with errors above 1.5 °C, during the examined period. In contrast, MISH's error is only approximately 1/3 of this. The biggest difference between the used methods occurs in summer, but in winter, for example, the MISH proves to be the best choice, the only difference compared to summer's case is that both the OK and the IDW produce better results, but in no case do they come close to the accuracy of MISH.




In relation to 18 UTC, the annual trend of the values is noticeable, the lowest RMSE occurs in winter, both in relation to IDW and kriging. Using the MISH system, the lowest errors for the interpolation of monthly values according to the station distribution are below 0.4 °C, and only in summer and October exceeds the error this threshold. It should be noted that in August, 0.63 °C is the highest RMSE value displayed by MISH interpolation during the year, for the 18 UTC database.




HUNGARIAN METEOROLOGICAL SERVICE — EUMETNET — WORLD METEOROLOGICAL ORGANIZATION


**References:**

Szentimrey, T., Bihari, Z., 2007: Mathematical background of spatial interpolation, meteorological interpolation based on surface homogenized data bases MISH (Meteorological Interpolation based on Surface Homogenized Data Basis). *Proceedings of the Conference on Spatial Interpolation in Climatology and Meteorology*, 2004 Október 24-29, Budapest, 17-27.

Szentimrey,T., Bihari, Z., 2014: Manual of interpolation software MISHv1.03. Budapest, OMSZ, 60p.

Szentimrey T.,2019a: Mathematical methodology and software for comparison of gridded datasets. Budapest, OMSZ . 10p.

Szentimrey, T., 2019b: Rácsponti adatbázis hatórás adatokra. (Gridpoint database for six-hourly data.) Budapest, OMSZ inner documentation.


Data availability: https://odp.met.hu/climate/homogenized_data/gridded_data_series/6_hours_data_series/


11th Seminar for Homogenization and Quality Control in Climatological Databases and 6th Interpolation Conference jointly organized with the 14th EUMETNET Data Management Workshop