Hungarian development of observation based temperature dataset

Beatrix Izsák, Tamás Szentimrey, Zita Bihari and Zsófia Barna

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

CLIMATE CHANGE
National Laboratory

HUNGARIAN METEOROLOGICAL SERVICE

2023 BUDAPEST

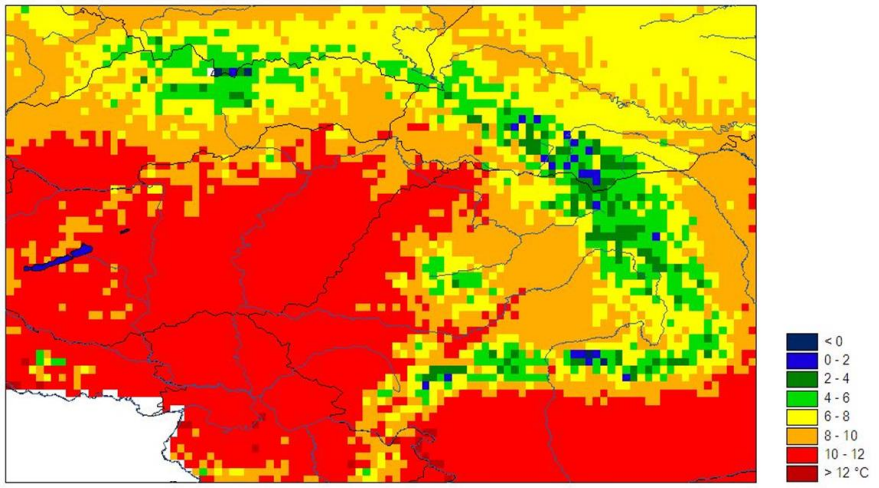
Daily mean temperature database

- CarpatClim 1961-2010, the western part is missing (Alpine region)
- HuClim 1901-2022 for the whole territory of Hungary, updated annually
- 1971-2022 odp.met.hu

ORSZÁGOS METEOROLÓGIAI SZOLGÁLAT Meteorológiai Adattár

Name	Last modified	Size	Description
Parent Directory		-	Homogenizált adatsorok
gridded_data_series/	2023-04-03 11:37	-	Rácsponti adatsorok
station_data_series/	2023-02-22 07:38	-	Állomási adatsorok

CARPATCLIM
Climate of the Carpathian Region



Annual mean temperature, 50 years means (1961-2010) on the grids (5895)

The creation of representative climatological database with MASH and MISH software:

http://www.met.hu/en/omsz/rendezvenyek/homogenization_and_interpolation/software/

MASHv3.03

(Multiple Analysis of Series for Homogenization; *Szentimrey, T.*)

For temporally representativity (homogenization, quality control and missing value completion of station data series)

MISHv1.03

(Meteorological Interpolation based on Surface Homogenized Data Basis; *Szentimrey, T. and Bihari, Z.*)

For spatially representativity (interpolation of homogenized data series)

HUNGARIAN METEOROLOGICAL SERVICE

References:
Szentimrey, T., Bihari, Z., 2014: Manual of interpolation software MISHv1.03, Hungarian Meteorological Service, p. 60.
Szentimrey, T. 2017: Manual of homogenization software MASHv3.03, Hungarian Meteorological Service, p.71.

INTERPOLATION SOFTWARE: MISHv1.03

I. Modelling system for climate statistical parameters in space

(expected values, standard deviations, spatiotemporal correlations)

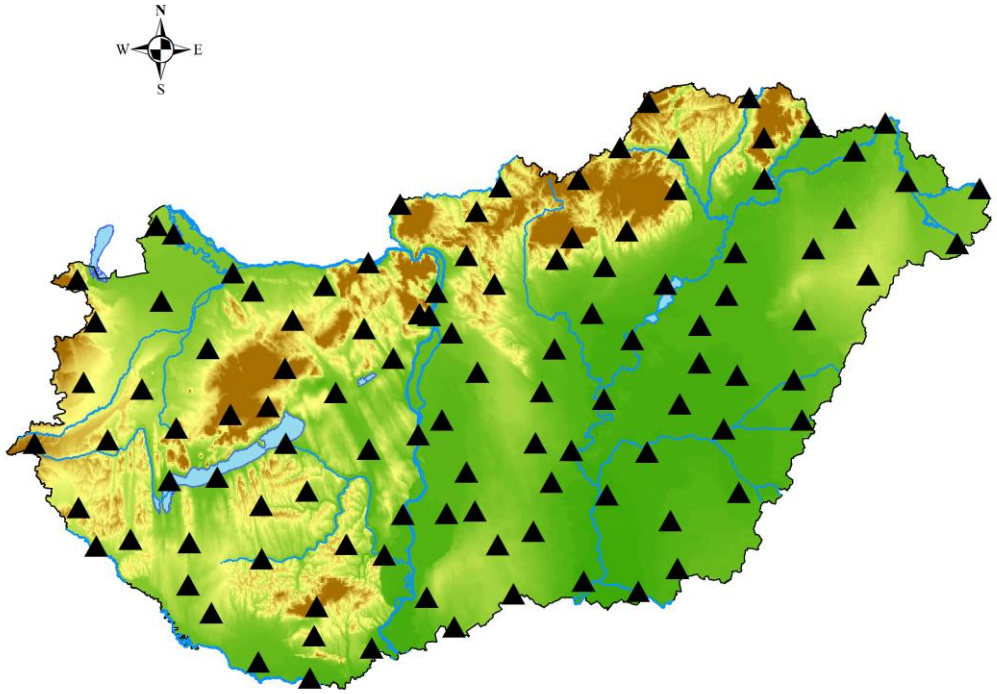
- **Based on long homogenized data series and model variables.**
- **Modelling procedure must be executed only once before the interpolation applications.**

II. Spatial interpolation system

- Additive (e.g. temperature) or multiplicative (e.g. precipitation) model and interpolation formula can be used depending on the climate elements.
- Daily, monthly, annual values and many years' means can be interpolated.
- Capability for application of background information such as satellite, radar, forecast data.
- Capability for gridding of data series.

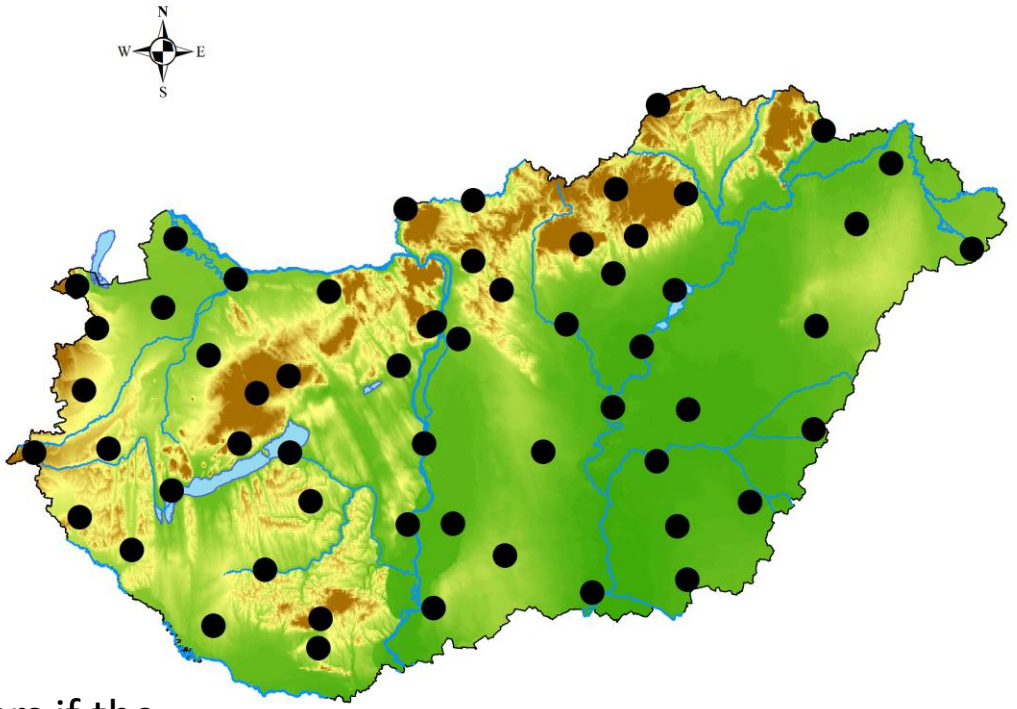
HUNGARIAN METEOROLOGICAL SERVICE

Geographical location of stations used in modelling climate statistical parameters



new station system (112)

TEMPERATURE (mean, maximum, minimum)

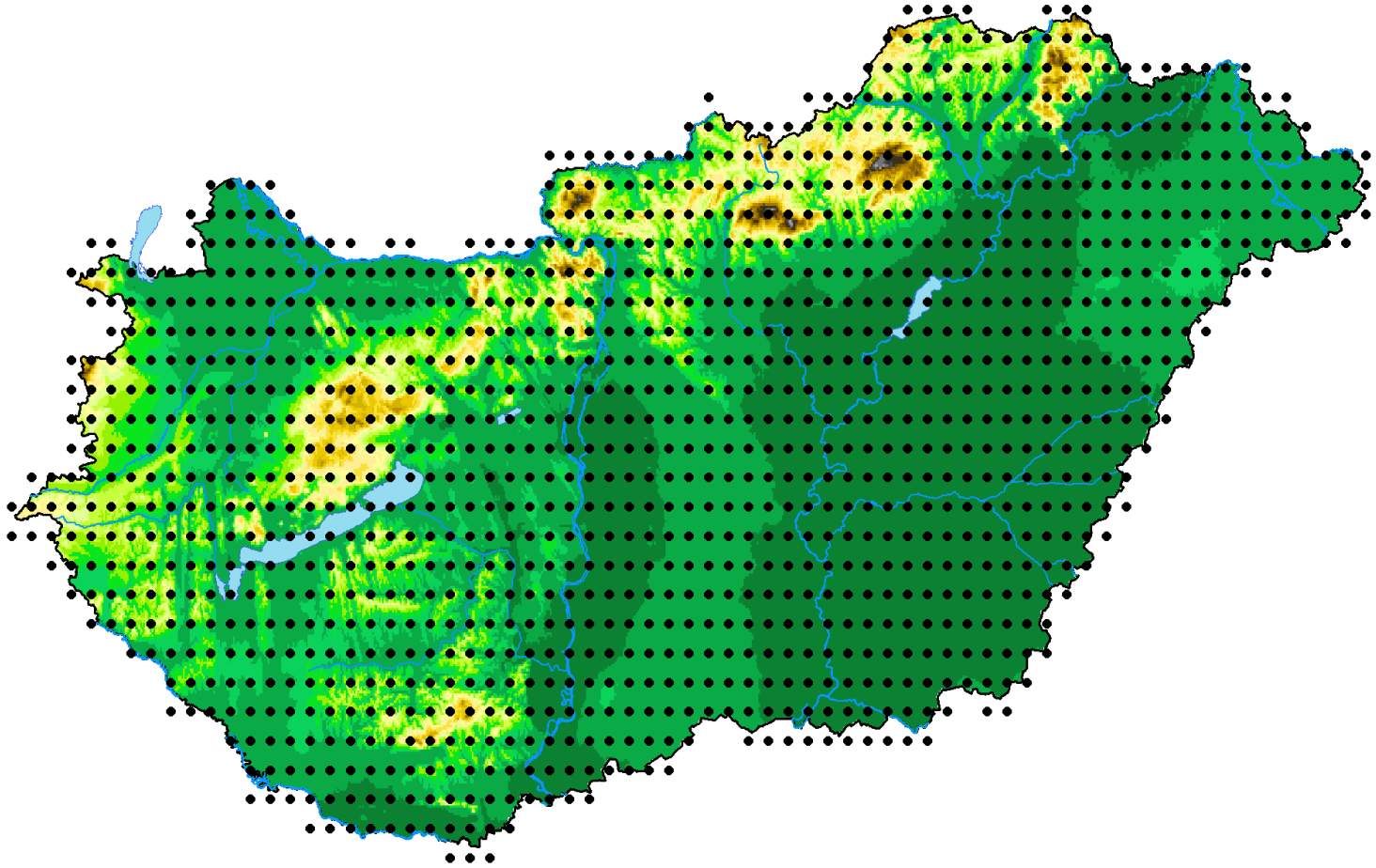


previous station system (58)

It is worth re-modelling the climate statistical parameters if the station network has expanded significantly.

Gridded datasets for Hungary (HuClim)

After homoganization we interpolate the homogenized, controlled and completed series with MISH to a grid with resolution of 0.1x0.1°, which means 1233 grid points in case of Hungary.



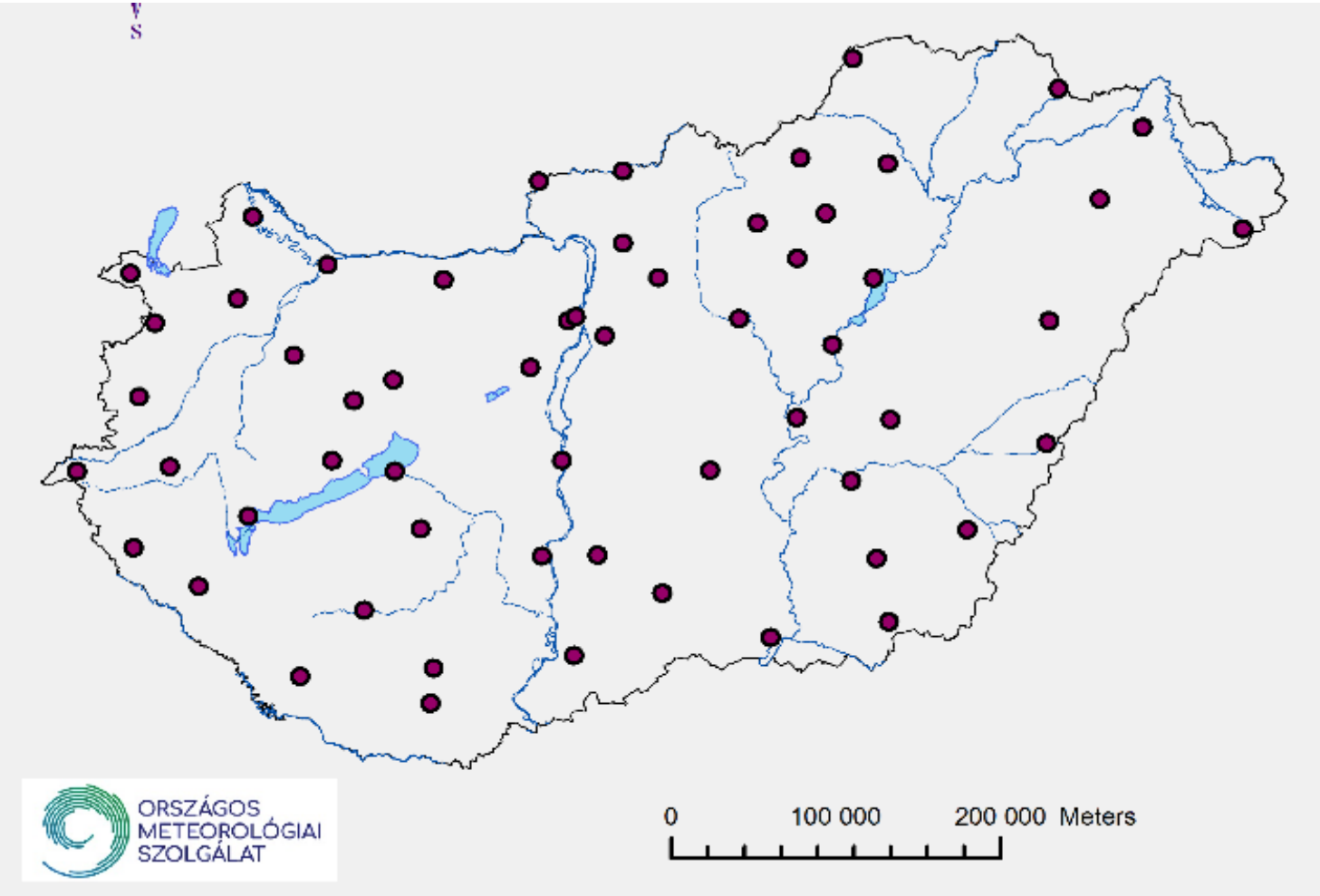
HUNGARIAN METEOROLOGICAL SERVICE

KLIMADAT PROJECT, NATIONAL LABORATORY

Assessment of climate change impacts in Hungary with regional climate model simulations and development of a representative climate database

- With the development of climate models, there has been an increasing need to study the intraday characteristics of temperature and precipitation, and this has required the refinement of the daily grid point observation database over time.
- This task has been carried out for 00 UTC, 06 UTC, 12 UTC and 18 UTC for temperature and precipitation sum.

DATA: station daily data series 1971-2022 and the six-hourly values



Temperature

- 00 UTC
- 06 UTC
- 12 UTC
- 18 UTC

58 stations

Homogenisation of hourly temperature values with MASH

Before homogenizing the series, we have to calculate the test statistics with MASH. On the basis of an examination of the verification statistics and this further step can then be decided on.

The test statistics generated automatically during the procedure:

Test statistics after homogenization (TSA): expresses the degree of inhomogeneity after homogenisation.

Test statistics before homogenization (TSB): expresses the degree of inhomogeneity before homogenisation.

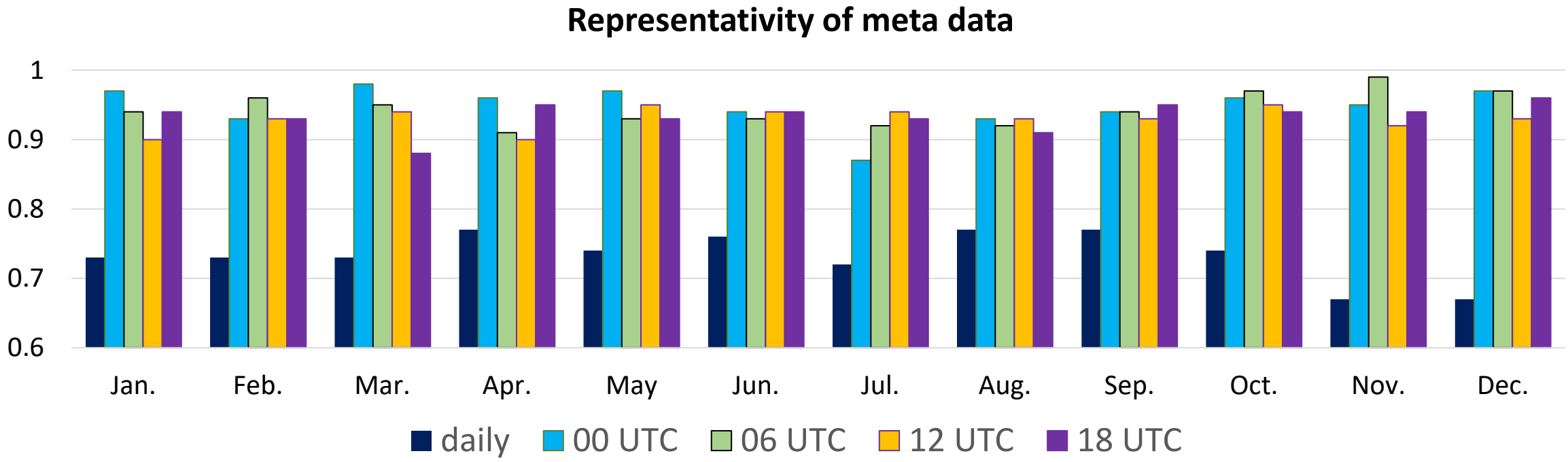
Representativity of meta data (RM): RM indicates whether the relative part of the estimated inhomogeneity can be explained by the meta data, the closer to 1 the more useful the meta data.

Homogenisation of hourly temperature values with MASH

	Jan.	*Feb.*	*Mar.*	*Apr.*	*May*	*Jun.*	*Jul.*	*Aug.*	*Sep.*	*Oct.*	*Nov.*	*Dec.*
00 UTC	44.0	43.0	81.5	170.6	103.4	83.8	155.7	95.0	175.5	71.7	52.7	49.7
06 UTC	41.3	43.2	59.0	104.7	106.2	156.7	318.7	140.6	119.4	98.9	43.0	42.0
12 UTC	69.1	80.9	133.3	112.7	197.9	101.1	136.8	101.6	115.2	116.5	77.1	281.4
18 UTC	53.1	66.8	114.5	133.9	108.0	126.9	124.7	103.6	302.5	89.3	122.2	91.9

Test statistics before homogenization (TSB): expresses the degree of inhomogeneity before homogenisation.

- Temperature has a marked diurnal cycle.
- The six-hourly temperature series are highly inhomogeneous, and their inhomogeneities are not the same as the daily series, i.e. we cannot depart from the daily trend for inhomogeneities.

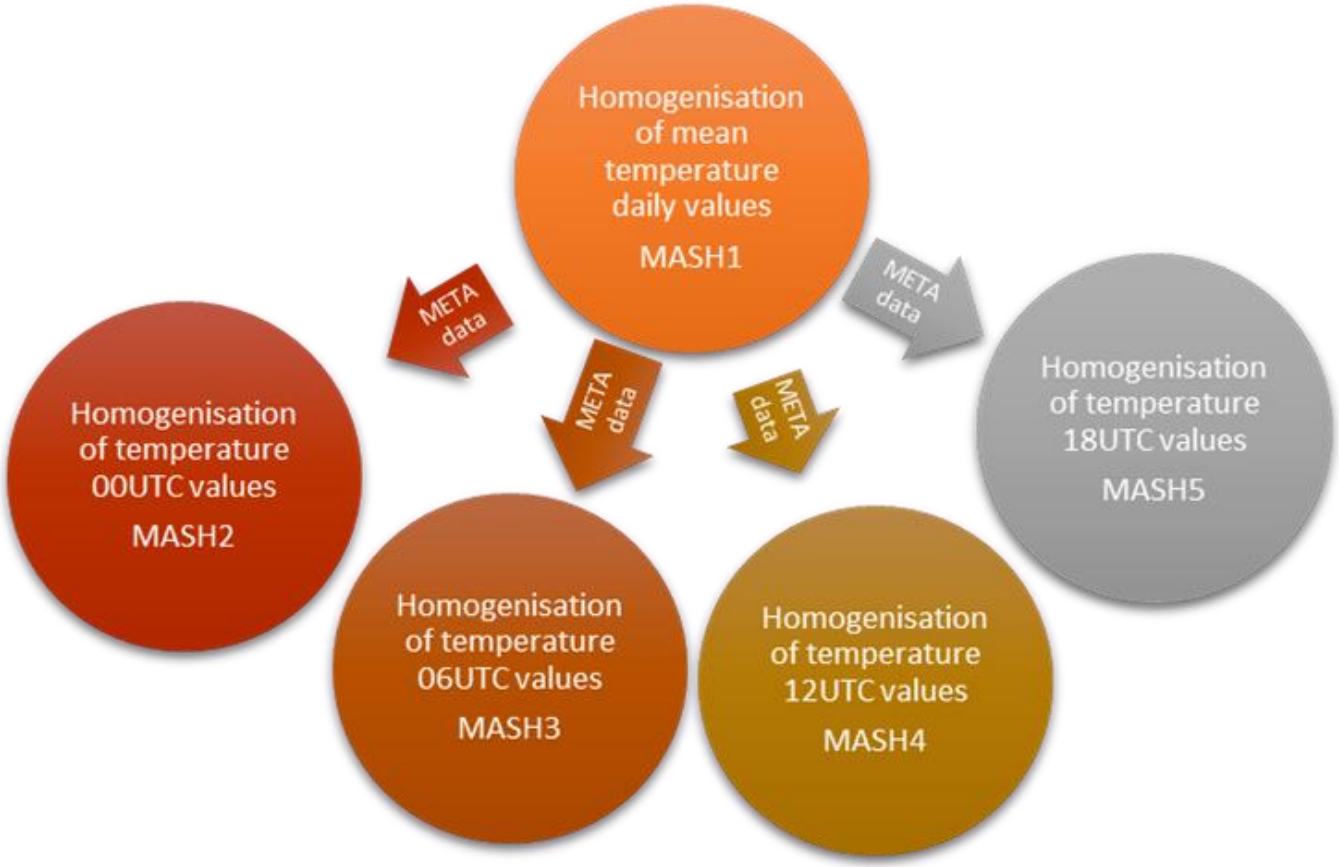


Representativity of metadata (RM): indicates whether the relative part of the estimated inhomogeneity can be explained by the metadata, the closer to 1 the more useful the meta data.

Statistics show that the time of the breakpoints obtained during the homogenization of daily data can be well used as META data when homogenizing hourly values.

Homogenization of hourly temperature values with MASH

- **Consequently, we are forced to homogenise the hourly data series separately, using the standard MASH procedure.**
- ***However, when homogenising the hourly data series with the MASH system, the breakpoints detected in the daily data series can be automatically used as META data.***



Interpolating daily values with MISH

Assuming the linear model – in case of normal distribution - the appropriate additive meteorological interpolation formula is as follows,

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

where Z($\mathbf{s}_0$) predictand, $\mathbf{s}$ are the element of the given space domain $\mathcal{D}$

Z($\mathbf{s}_i$) (i=1,…,M) predictors

Where $\sum_{i=1}^{M} \lambda_i = 1$ and the λ_i (i=1,…,M) minimize the root-mean-square error and these are known functions of some climate statistical parameters.

$E(\mathbf{s}_i)$ (i=0,…,M) are the expected values or spatial trend values. (In case of normal distribution the expected value equals median.)

Interpolating hourly values with MISH

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

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

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

Interpolating hourly values with MISH

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:

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

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

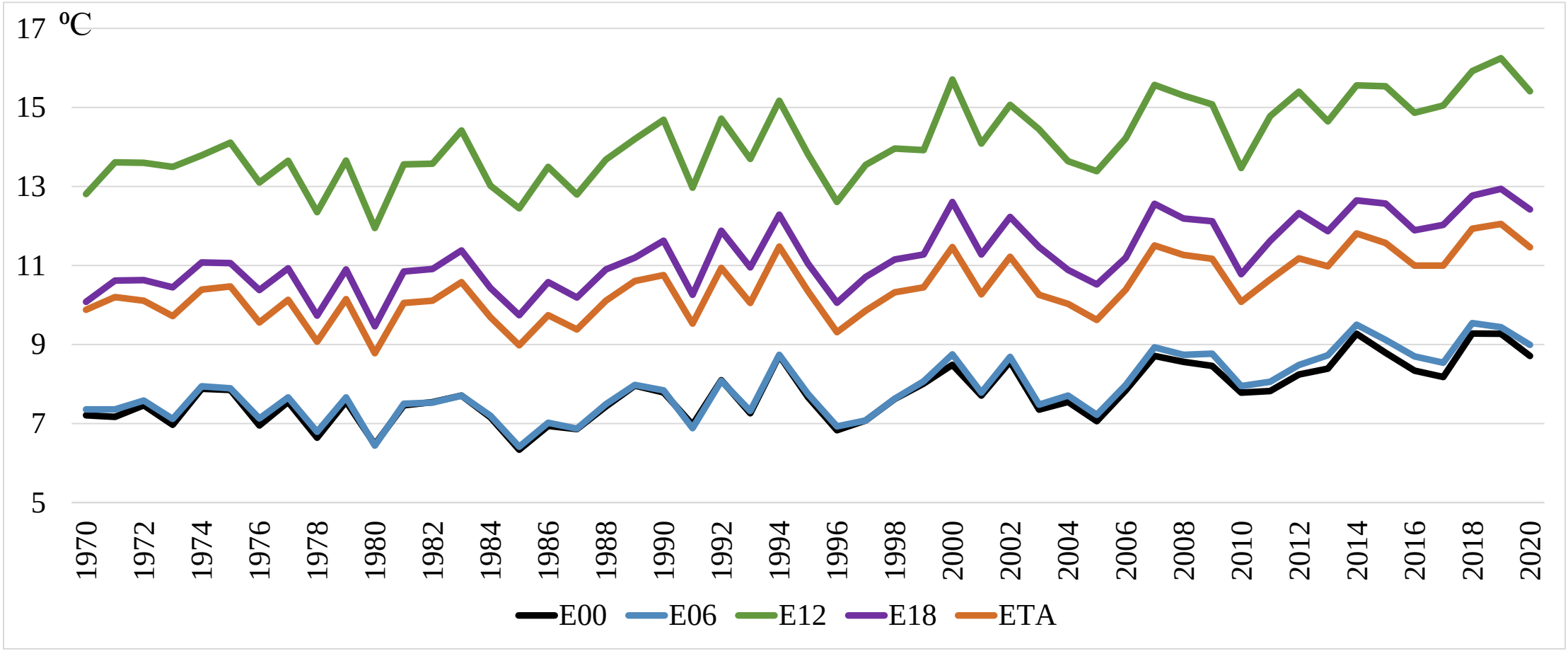
The alpha, beta regression parameters for 12 months, and together with the 4 hourly values

	T00		T06		T12		T18	
MONTH	*alfa*	*beta*	*alfa*	*beta*	*alfa*	*beta*	*alfa*	*beta*
1	-0.97	**0.97**	-1.52	**0.88**	2.15	**1.06**	0.32	**1.11**
2	-1.26	**0.85**	-2.08	**0.79**	2.76	**1.08**	0.45	**1.13**
3	-1.38	**0.84**	-1.61	**0.70**	2.97	**1.13**	0.02	**1.19**
4	-0.38	**0.75**	-0.69	**0.79**	2.28	**1.17**	-0.16	**1.15**
5	1.27	**0.69**	0.23	**0.85**	0.66	**1.23**	-1.79	**1.22**
6	1.21	**0.73**	0.21	**0.89**	0.22	**1.21**	-2.78	**1.25**
7	2.51	**0.69**	0.57	**0.86**	1.10	**1.16**	-3.01	**1.24**
8	0.97	**0.77**	0.70	**0.82**	1.84	**1.15**	-3.68	**1.25**
9	0.06	**0.81**	1.27	**0.71**	1.67	**1.20**	-2.87	**1.22**
10	-0.57	**0.84**	-0.07	**0.70**	2.32	**1.20**	-1.42	**1.16**
11	-0.46	**0.84**	-0.80	**0.75**	2.03	**1.17**	-0.36	**1.10**
12	-0.73	**0.95**	-1.15	**0.89**	1.90	**1.07**	0.07	**1.08**

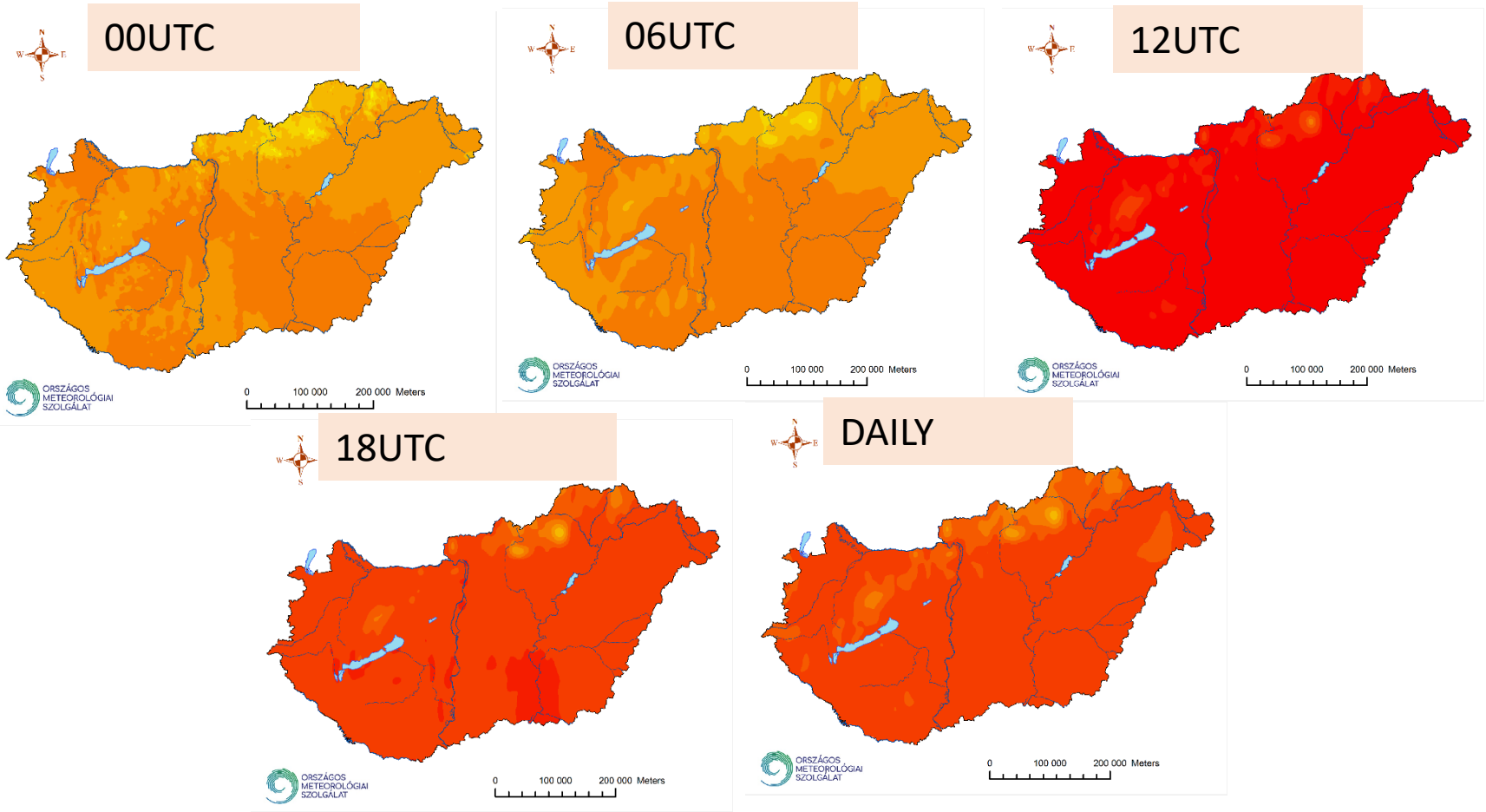
Correlations (corr) between daily and hourly data and t-test statistics (tstat) for 12 months, and together for hourly values

	T00		T06		T12		T18	
MONTH	corr	tstat	corr	tstat	corr	tstat	corr	tstat
1	**0.95**	23.02	**0.95**	22.78	**0.89**	14.32	**0.98**	40.44
2	**0.89**	14.41	**0.88**	13.64	**0.86**	12.46	**0.98**	34.25
3	**0.87**	12.97	**0.84**	11.42	**0.83**	11.23	**0.97**	31.82
4	**0.80**	9.87	**0.91**	16.19	**0.86**	12.54	**0.94**	20.80
5	**0.78**	9.19	**0.94**	21.08	**0.91**	15.94	**0.94**	20.91
6	**0.78**	9.18	**0.96**	25.07	**0.91**	16.08	**0.95**	22.00
7	**0.69**	7.19	**0.95**	22.54	**0.88**	14.04	**0.93**	19.70
8	**0.73**	8.04	**0.89**	14.53	**0.83**	11.13	**0.93**	19.32
9	**0.81**	10.30	**0.83**	11.03	**0.80**	10.15	**0.97**	28.24
10	**0.83**	11.10	**0.80**	10.00	**0.78**	9.29	**0.96**	24.57
11	**0.90**	15.16	**0.85**	12.02	**0.83**	11.05	**0.98**	39.47
12	**0.95**	21.98	**0.94**	20.40	**0.89**	14.49	**0.98**	40.03

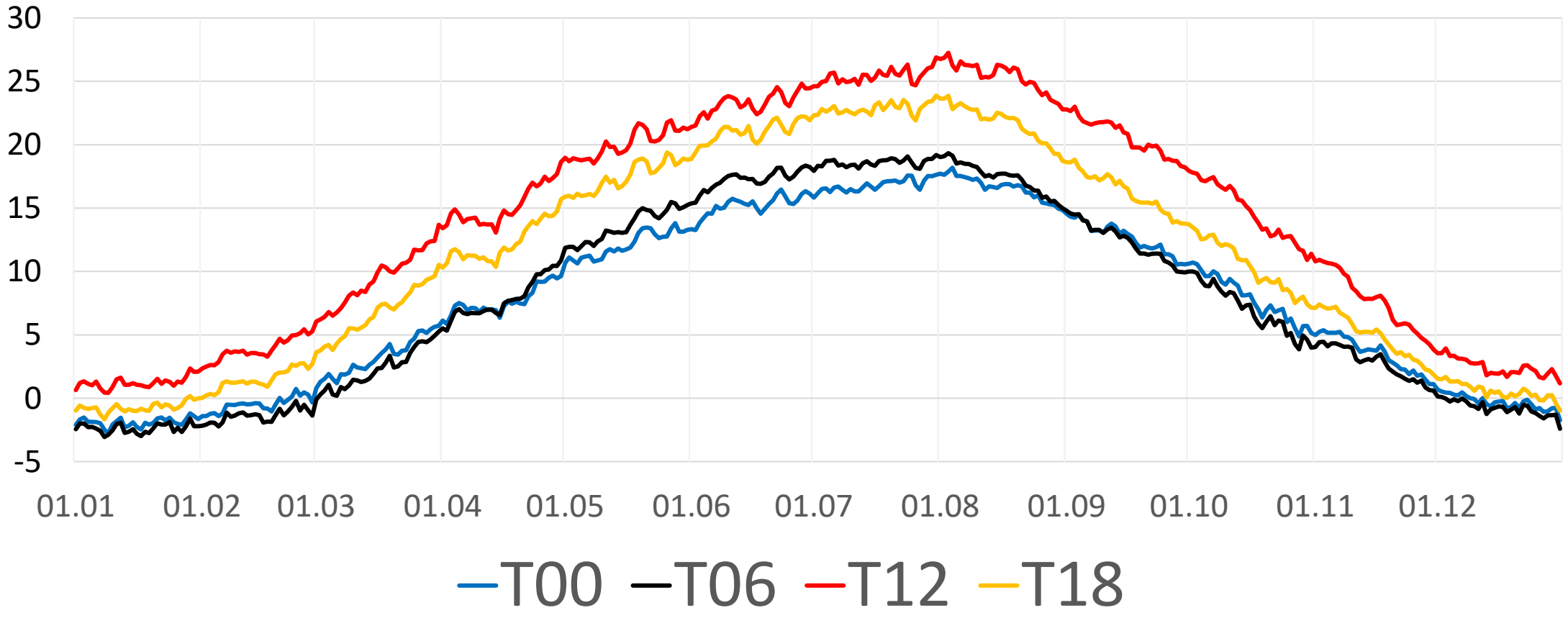
Spatial mean, 1971-2020,hourly values (E00, E06, E12, E18) and daily values (ETA)



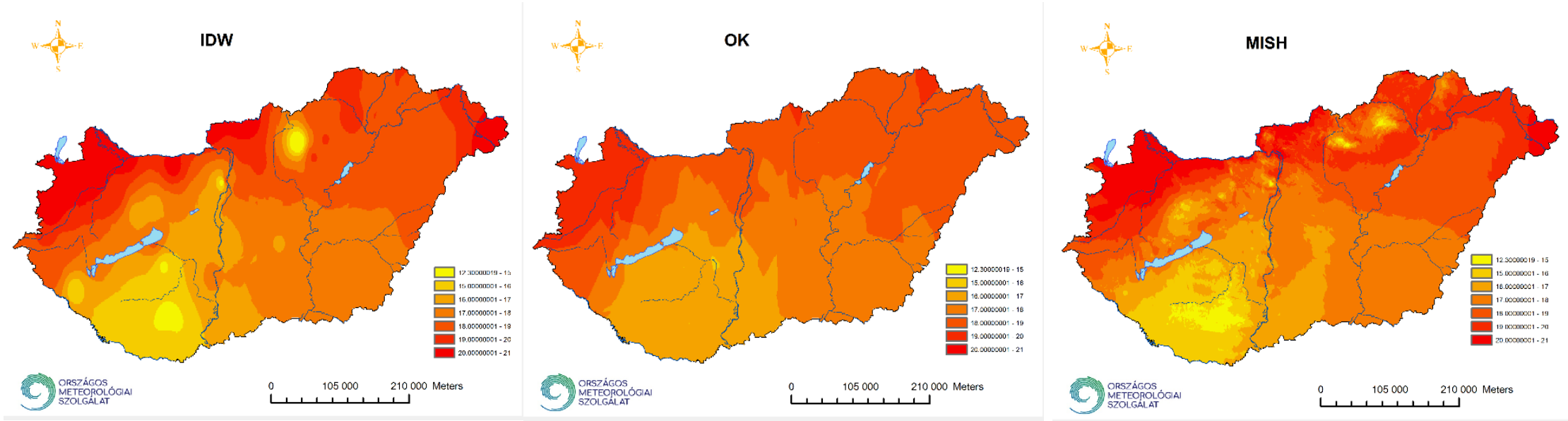
Temporal mean, 1971-2020,hourly values and daily values



Annual cycle of hourly values (°C)



Interpolated values of 20 May 2020 12 UTC measurements using IDW (A), OK (B) and MISH (C)



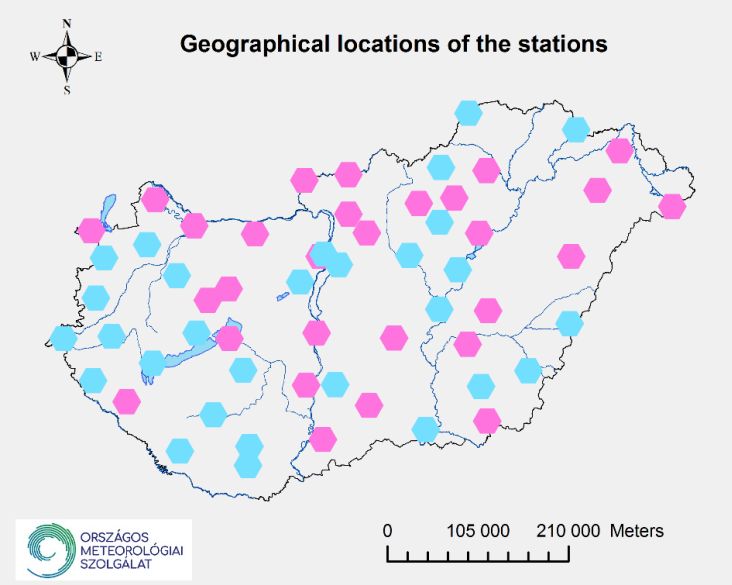
A *B* *C*

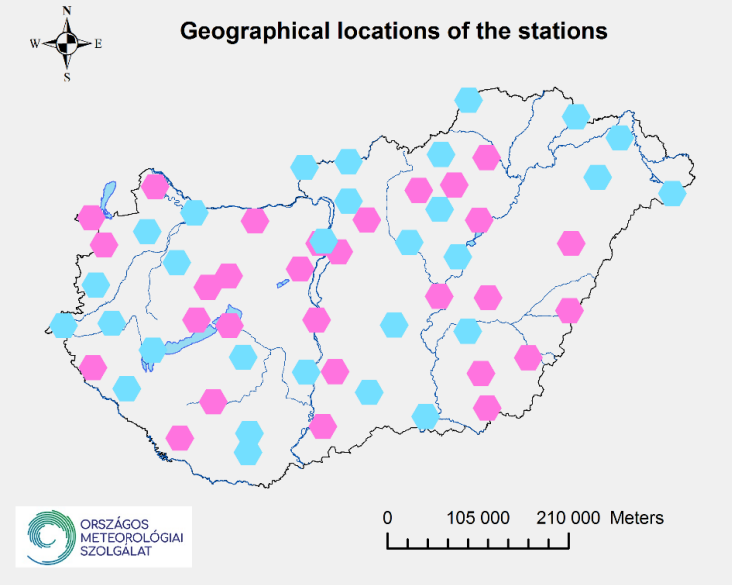
The good interpolation method is the one with the lowest RMSE. It is not the "beauty" of the map that counts.

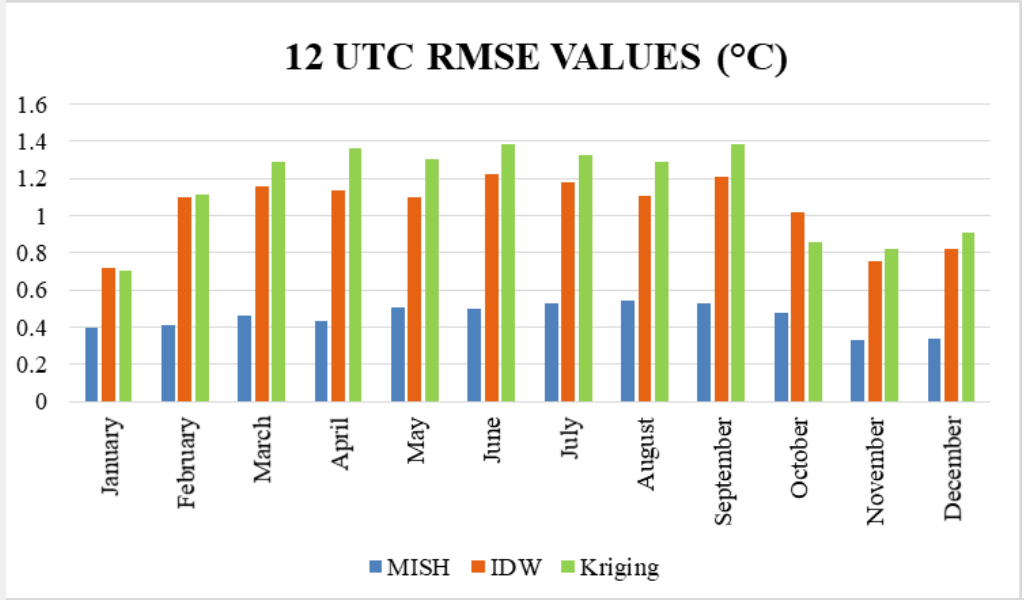
Comparison of interpolation methods for six-hour temperature data series

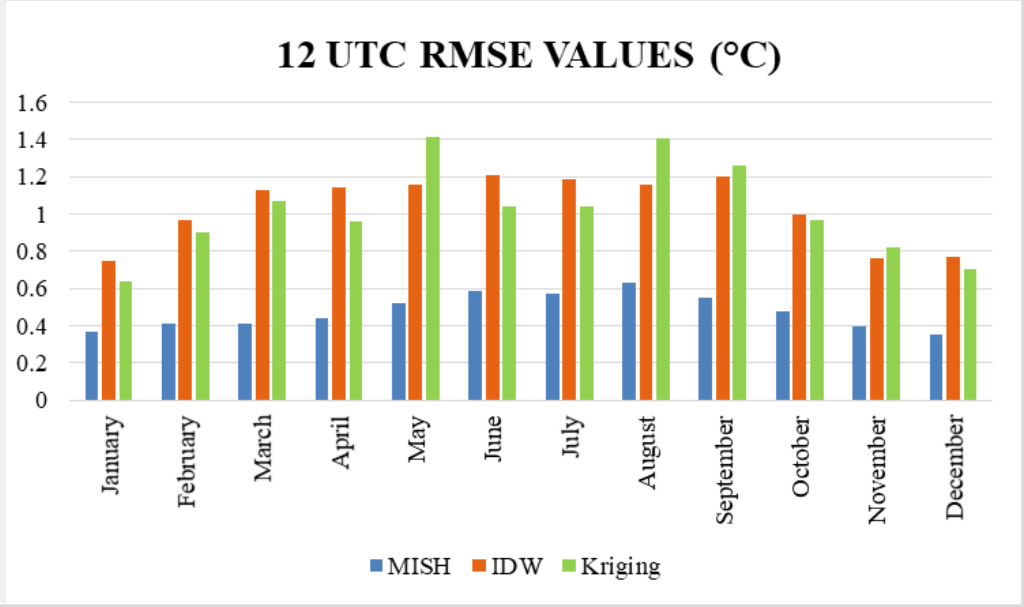
- Cross validation technique was used to compare different interpolation methods which is available in the Geostatistical Analyst toolbox of the ArcMap software. This software randomly divides the station system into training and test datasets. For each subset and each hourly value 150 interpolations were performed. After that we calculated the RMSE (root mean square error) values for all three methods.
- Barna Zs., Izsák, B., Pieczka, I., 2023: Comparison of interpolation methods for six-hour temperature 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,* Budapest, 2023.05. 9-11.
- Data availability:

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









In general, the errors are higher in the summer semester than in the winter semester. The smallest errors occur with MISH interpolation.

Groups of stations used for validation of 12 UTC values, pink for 'training' and blue for 'test' stations.

Summary

- The MASH software is capable of homogenizing subdaily data, in line with the use of breakpoints obtained when homogenizing daily data.
- The temperature values have a marked diurnal trend, and we cannot ignore the diurnal trend of inhomogeneity.
- The MISH system can also be used to interpolate hourly values with good quality.
- The modelled daily spatial trend values and the estimated hourly regression coefficients can be used to interpolate the hourly values.

Thank you for your kind attention!