Interreg
Danube Region

Co-funded by
the European Union

Danube-ADAPT

Homogenization with MASH – the climatological database of the Danube region

Beatrix Izsák[1], Olivér Szentes[1], Tamás Szentimrey[2], Mónika Lakatos[1], Zita Bihari[1]

[1]Department of Climate Research, HungaroMet Hungarian Meteorological Service, Budapest, Hungary

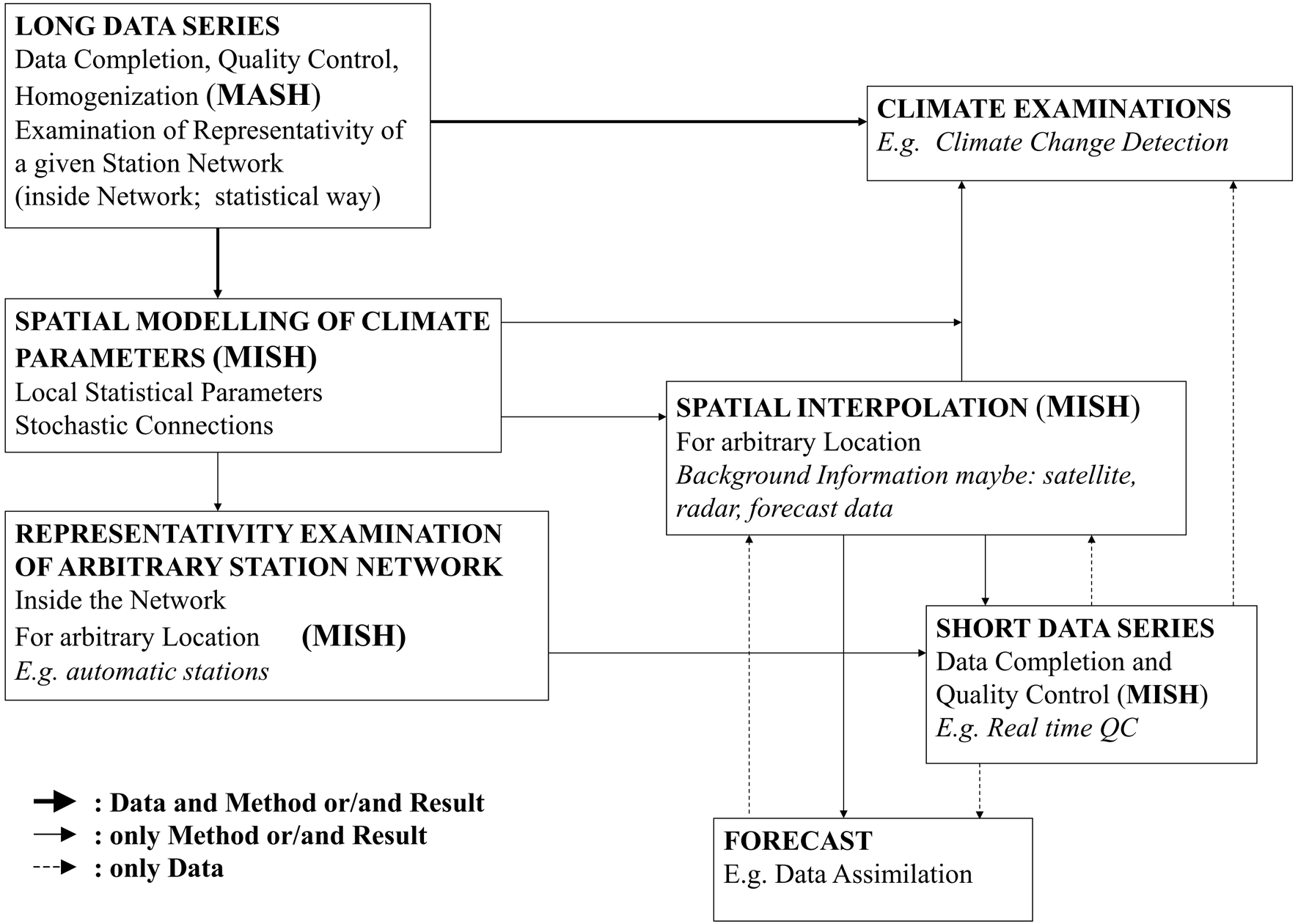
[2]Varimax Limited Partnership, Budapest, Hungary

Budapest, 7th May 2026

1. Joint MASH homogenization methodology for the Danube Region climate observation database (Szentimrey, 2025)
2. MASHv3.03
3. Steps
4. Verification statistics
5. Joint MISH modelling and gridding methodology for the Danube Region climate observation database (Szentimrey, 2025)

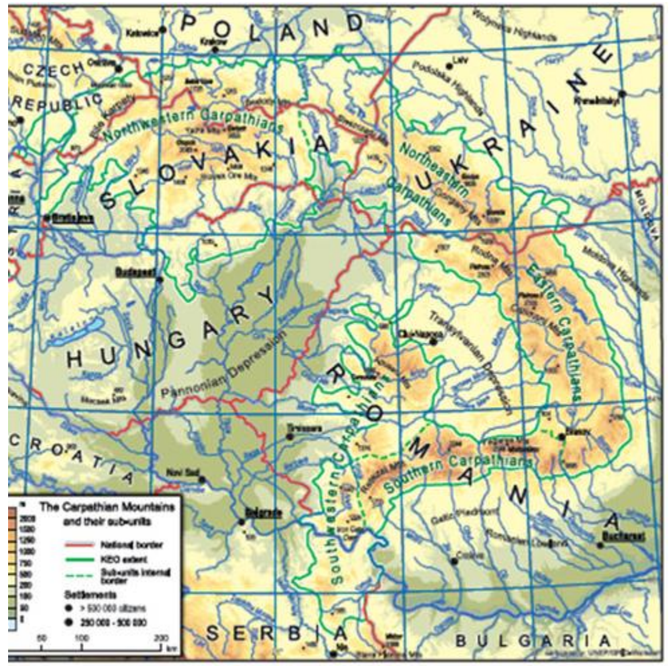
Interreg Danube Region
Co-funded by the European Union
Danube-ADAPT

Possible Connection of Topics and Systems (Szentimrey)



In DA project, the homogenization, data quality control, data completion and gridding will be implemented by these common software MASH and MISH at national or sub-region level, based on experiences from the CarpatClim project. Data series from near border stations will be exchanged between neighboring countries, sub-regions for cross-border harmonization.

https://www.carpatclim-eu.org/pages/about



Interreg Danube Region

Co-funded by the European Union

Danube-ADAPT

MASHv3.03 and MISHv1.03 download

The latest MASH (version 4) is already available here, but we are still using MASHv3.03.

https://www.met.hu/en/rolunk/rendezvenyek/homogenization_and_interpolation/software/index.php

We will apply the version MASHv3.03 that includes some automation.

The intention was to develop a flexible, interactive automatic, artificial intelligence (AI) system. These automatic 'user friendly' procedures make the homogenization easier for the users.

The main properties of the version MASHv3.03

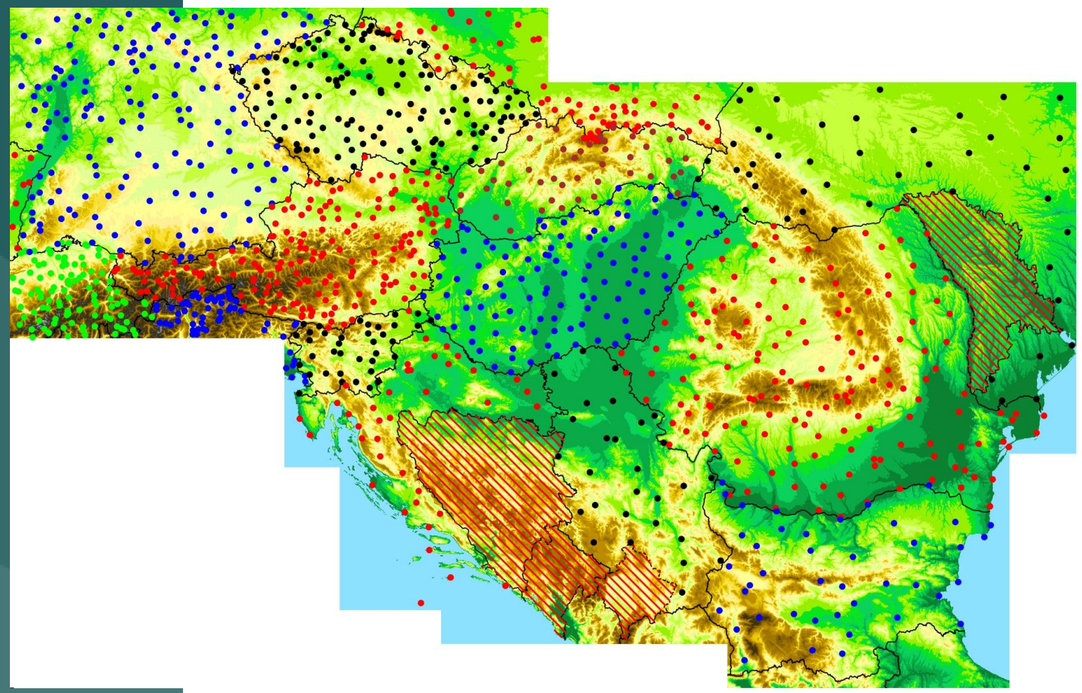
Advantages of MASHv3.03 in the homogenization of monthly series:

- **It is a relative homogeneity test procedure.**
- **It is a step-by-step iteration procedure: the role of series (candidate, reference) changes step by step in the course of the procedure.**
- **Interactive automatic, artificial intelligence (AI) system (Szentimrey, 2023).**
- **An additive (e.g. temperature) or multiplicative (e.g. precipitation) model can be used depending on the distribution.**
- **It includes quality control and missing data completion.**
- **It provides the homogeneity of the seasonal and annual series as well.**
- **Metadata (probable dates of break points) can be used automatically.**
- **The homogenization results can be evaluated on the basis of verification tables generated automatically during the procedure.**

In the homogenization of daily series:

- **The procedure is based on the detected monthly inhomogeneities.**
- **It includes quality control and the completion of missing data in daily data.**

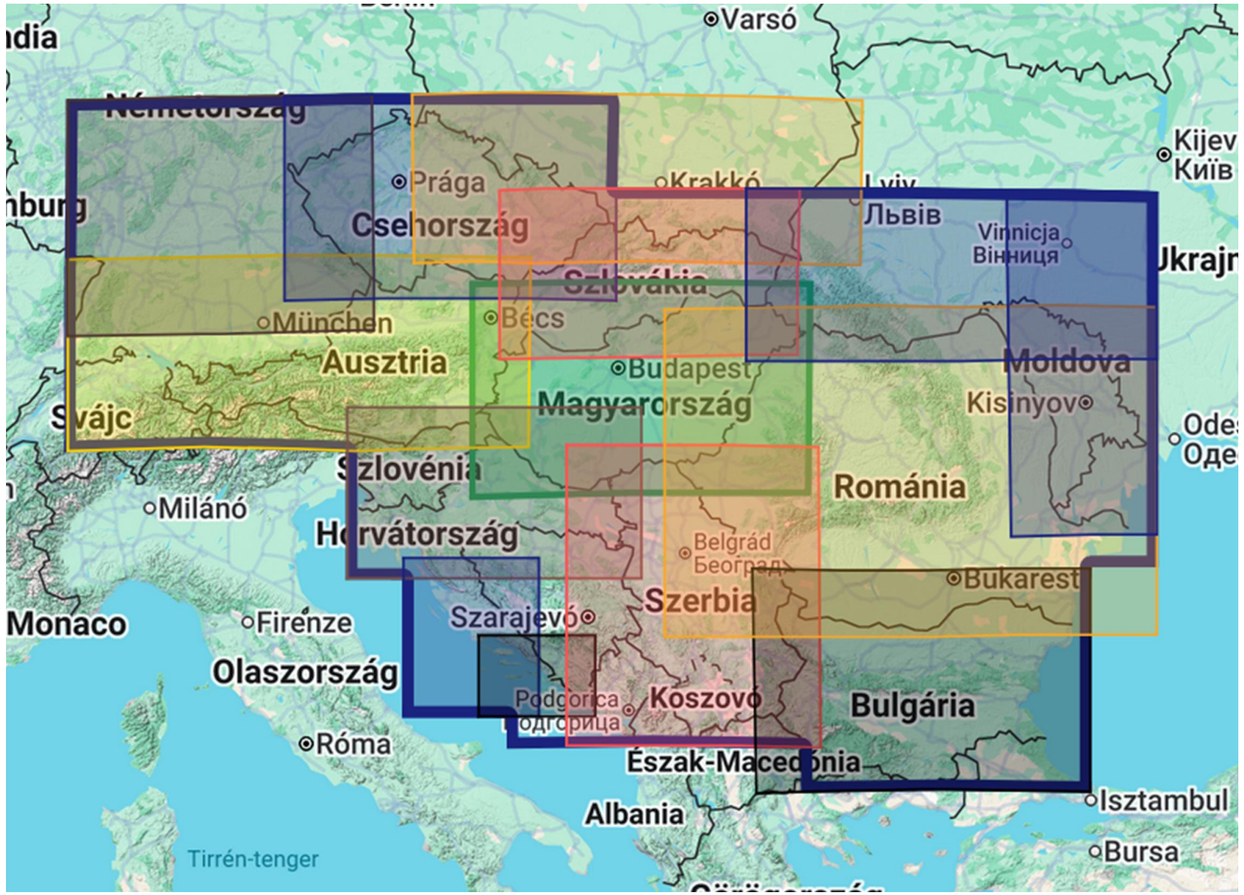
The meteorological elements have been selected so that all working groups (impact assessment, vulnerability) can use them.



Variable	*Description*	*Units*	*Time period*
Ta	*2 m mean daily air temperature*	*°C*	*1970-2024*
Tmin	*Minimum air temperature*	*°C*	*1970-2024*
Tmax	*Maximum air temperature*	*°C*	*1970-2024*
p	*Accumulated total precipitation*	*mm*	*1970-2024*
VV	*10 m horizontal wind speed*	*m/s*	*2000-2024*
Vmax	*10 m horizontal max. wind speed*	*m/s*	*2000-2024*
Rglobal	*Global radiation*	*J/cm^2*	*2000-2024*
RH	*Relative humidity*	*%*	*1970-2024*
pair	*Surface air pressure*	*hPa*	*1970-2024*

Overview of the main steps for gridded datasets:

1. Near border data exchange before homogenization
2. Homogenization (MASH)
3. Data exchange after homogenization
4. Controlling of the cross-border harmonization (MASH)
5. Gridding, interpolation (MISH) per countries, sub-regions with exchanged data
6. Compilation of gridded series from countries in one file per variable



Slovakia, Ukraine, and Romania prepare the database for their own countries, while HungaroMet prepares the database for the other areas.

I. Compilation of the raw station data series.

1. Selection of the stations (with the coordinates: φ, λ), determination of the time period.
2. Collecting the daily station data series (missing data are allowed) and the metadata per countries or sub-regions. Exchange of the near border station data series and the metadata between the neighbouring countries, sub-regions.

II. Homogenization, quality control, data completion for the station data series by MASHv3.03 at national, sub-region level, with using the near border data.

1. Derivation of monthly station data series from the daily station data series. Homogenization, quality control, data completion of the monthly station data series.
2. Continuation for daily station data series: homogenization, quality control, data completion. This procedure is based on the results and continuation of II.1.
3. Exchange of the near border homogenized daily data series for cross-border harmonization.
4. Evaluation of the results of the homogenization and quality control. Controlling of the cross-border harmonization of the homogenized data series. The cross-border harmonization will be continued after modelling procedure of MISH.

Some remarks on using MASHv3.03

MASH is an *interactive automatic* procedure, and it is worth considering if the station network is dense, or numerous stations have to be examined.

Applying manual supervision after/in addition to the automatic procedure, assuming mathematical knowledge, gives better results.

The principle of the quality control and the missing data completion procedures built in MASH is certain multiple comparisons of the data by spatial interpolation technique. In the case of quality control, the analysis of interpolation errors for detection of the wrong data is based on confidence intervals. The necessary climate statistical parameters, such as expected values, spatial and temporal covariance structure, are estimated on the basis of examined data series. The quality control and the missing data completion procedures are based on strong mathematical fundament and fully automated in the MASH software.

During the execution of the quality control and homogenization the test results e.g. detected errors, degree of inhomogeneity, number of break points, estimated corrections and certain verification results are documented in automatically generated tables which make the evaluation of the results efficient.

-If we have the final result, i.e. DailyHomQC.dat, we suggest also some control procedure! (the description is in the project document)

The verification statistics in MASH

The test statistics generated automatically during the procedure:

TEST STATISTICS FOR SERIES INHOMOGENEITY
- Test statistics after homogenization
- Test statistics before homogenization
- Statistics for estimated inhomogeneities

CHARACTERIZATION OF INHOMOGENEITY
- Relative estimated inhomogeneities
- Relative modification of series
- Lower confidence limit for relative residual inhomogeneities

REPRESENTATIVITY OF STATION NETWORK

EVALUATION OF META DATA
- Test statistics
- Representativity of META data

Szentimrey, T., 1999: „Multiple Analysis of Series for Homogenization (MASH)", Proceedings of the Second Seminar for Homogenization of Surface Climatological Data, Budapest, Hungary; WMO, WCDMP-No. 41, pp. 27-46.

Szentimrey, T., 2017: „Manual of homogenization software MASHv3.01", Hungarian Meteorological Service, p. 61.

Interreg Danube Region

Co-funded by the European Union

Danube-ADAPT

VERIFICATION OF HOMOGENIZATION (Ordered Statistics)

I. TEST STATISTICS FOR SERIES INHOMOGENEITY

Null hypothesis: the examined series are homogeneous.
Critical value (significance level 0.05): 20.91
Test statistics (TS) can be compared to the critical value.
The larger TS values are more suspicious!

1. Test Statistics After Homogenization

Series	Index	TSA	Series	Index	TSA	Series	Index	TSA
B2DUKO	308	98.93	ABO	182	95.14	100	86	87.79
02708	269	86.65	05111	299	75.71	133	113	69.29
15310	325	63.10	199	159	63.02	161	135	62.60
150	125	57.97	C1CHUR	313	56.17	BUF	189	52.18
WAE	236	51.79	98	85	49.24	SAM	224	49.17
03147	273	48.63	89034	386	47.93	COM	192	46.22
89009	362	46.18	B2MBUD	311	45.90	01550	260	45.12
68	59	45.00	INT	207	44.69	99011	240	43.84
88033	336	43.27	24	23	43.05	02638	267	43.00
20105	181	42.27	LET	213	42.07	36	31	41.37

.....

05440	302	11.08	50	43	11.00	CHO	191	10.79
16414	328	10.75	44	38	10.62	05404	301	10.39
05792	305	10.24	SCU	226	10.05	01346	257	9.97
115	99	8.88	19720	179	8.64	89025	377	8.55
SMM	230	8.26	ULR	233	7.78	C1HKVI	314	6.88
75	65	6.72	17	16	6.17	05664	303	5.87

AVERAGE: 25.42

2. Test Statistics Before Homogenization

Series	Index	TSB	Series	Index	TSB	Series	Index	TSB
01443	258	3199.68	03307	276	1727.06	00232	247	1672.36
88	76	1602.07	108	94	1515.88	02812	271	1515.69
89042	393	1430.68	00073	244	1420.48	89020	372	1415.24
16400	177	1337.01	89034	386	1261.47	87	75	1248.56
89036	387	1194.31	40	35	1158.91	89204	399	1144.27
03147	273	1124.17	03366	277	1066.39	03485	280	1018.55
04300	289	1002.93	89019	371	981.28	C2BYNO	318	937.95

.....

01468	259	28.00	68	59	27.73	138	117	26.26
348	348	24.67	97	84	24.50	ROB	221	23.94
205	164	21.90	SAE	223	21.37	01346	257	20.61
71	62	20.00	213	167	19.12	JUN	208	18.00
ULR	233	17.27	103	89	16.66	16414	328	12.76

AVERAGE: 278.39

Error.res

20241111	0.00	0.00	0.00	0.00
20241125	0.00	0.00	0.00	0.00
20241126	0.00	0.00	0.00	0.00
20241214	0.00	0.00	0.00	0.00
20241228	0.00	0.00	0.00	0.00
20241230	0.00	0.00	0.00	0.00
20241231	0.00	0.00	0.00	0.00

Number of days with error(s): 3112
Total number of errors: 4726
Maximal positive error: 22.33
Minimal negative error: -29.58

III. REPRESENTATIVITY OF STATION NETWORK
(1-relative interpolation error)

Series	Index	RS	Series	Index	RS	Series	Index	RS
01LYSA	86	0.79	C2BORK	33	0.83	02PASE	97	0.83
02SUMP	98	0.84	01358	146	0.87	010PAV	89	0.87
01832	148	0.87	98032	172	0.87	H1PECS	45	0.88
98088	179	0.88	L3KVAZ	76	0.88	98033	173	0.88
02JEVI	94	0.89	03166	151	0.89	C2VBRO	43	0.89
L2PRIM	70	0.89	01CERV	80	0.89	C2TREB	42	0.90
89	196	0.90	C1HKVI	25	0.90	98026	168	0.90
U2VARN	137	0.90	01JAVO	83	0.90	03946	154	0.90
L3AS00	73	0.90	11841	184	0.90	U1DOKS	119	0.90
02OLOM	96	0.90	U1ZATE	129	0.91	H2DEST	48	0.91
L2KONL	66	0.91	00991	142	0.91	H1LBOU	44	0.91
C1CHUR	24	0.91	11866	185	0.91	L2ZBIB	72	0.91

.....

U2DOKY	133	0.95	P1PKBE	104	0.95	B1KROM	4	
C2HLAS	37	0.95	98021	165	0.95	03VALM	100	
P1PLIB	106	0.96	152	198	0.96	P3NRYC	116	
U1KOPI	121	0.96	B2TROU	22	0.96	04911	158	
P1PKLE	105	0.96	06314	163	0.96	B2BZAB	13	
P2BRAN	108	0.97						

AVERAGE: 0.92

Overview of the main steps for gridded datasets:

Near border data exchange before homogenization

Homogenization (MASH)

Data exchange after homogenization

Controlling of the cross-border harmonization (MASH)

Gridding, interpolation (MISH) per countries, sub-regions with exchanged data

Compilation of gridded series from countries in one file per variable

As it can be seen, the implementation of the harmonization is planned in more steps during the homogenization and gridding procedures.

Interreg Danube Region

Co-funded by the European Union

Danube-ADAPT

The main features of MISHv1.03

The software MISHv1.03 consists of two units that are the modelling and the interpolation systems. The interpolation system can be operated on the results of the modelling system.

Modelling system for climate statistical parameters in space:

Based on long homogenized data series and supplementary deterministic model variables e.g. topography. Neighborhood modelling of climate statistical parameters.

cross-validation test for interpolation error or representativity.

Modelling procedure must be executed only once before the interpolation applications!

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 values and many years' means can be interpolated.

Few predictors are also sufficient for the interpolation.

The interpolation error or representativity is modelled too.

Capability for application of supplementary background information (stochastic variables) e.g. satellite, radar, forecast data.

Capability for gridding of data series.

As regards the problem of cross-border harmonization, in this respect the neighbourhood modelling of the necessary climate statistical parameters, e.g. spatial trend and covariance (or variogram) structure, is a very useful feature of MISH.
Similarly to the modelling part, at the interpolation system is used also neighbourhood technique, that is the applied predictors are chosen from certain neighbourhood of the predictand.
This property can facilitate the harmonization of the gridded series across the border if we use the near border homogenized station data series as predictors from both sides.

The statistical result files of the modelling and gridding procedure in order to give a chance for evaluation of the quality of modelling and interpolation (p.31 in Manual).
These files in subdirectory MISH\MODEL\Modelres12 are as follows.

- Monthly modelling results for the spatial trend (p.31, 47 in Manual):

detmodstat01.res,....., detmodstat12.res

- Monthly benchmark results (p.31, 48 in Manual):

benchmark01.res,..., benchmark12.res

- Maps of modelled many years monthly means (p.31, 48 in Manual):

meangrid01.res,..., meangrid12.res

- Statistical results of gridding per months (pp.6-7, 40 in Manual):

gridstat01.res,...., gridstat12.res

Interreg Danube Region
Co-funded by the European Union
Danube-ADAPT

BENCHMARK STUDY: cross-validation test, interpolation between the stations
Number of stations: 68
MEAN REPRESENTATIVITY VALUES (REP) FOR THE MONTHS
REP=1-RMSE/(Standard Deviation)
REPop: interpolation with optimum parameters
REPmp: interpolation with modelled parameters

MONTH	REPop	REPmp
1	0.891	0.820
2	0.913	0.848
3	0.919	0.827
4	0.892	0.779
5	0.891	0.767
6	0.864	0.721
7	0.857	0.713
8	0.879	0.756
9	0.889	0.773
10	0.877	0.763
11	0.899	0.807
12	0.879	0.781
MEAN	0.887	0.780

Maximum air temperature from CarpatClim

Table 1. Monthly modelling results for the spatial trend of the series system of Hungary, Croatia and near border series from Austria, Serbia, Romania, Ukraine, Slovakia

.....

MODELLING OF SPATIAL TREND (linear regression)

```
 1: number of model variables:  6    correlation: 0.727    percent: 31.4%
variables:            h        1        3        5        7       10
coefficients:   -0.0029   0.0202   0.0259  -0.0458   0.0826   0.0518

 2: number of model variables:  5    correlation: 0.815    percent: 42.0%
variables:            h        1        3        5        9
coefficients:   -0.0025   0.0278   0.0357  -0.0424  -0.0460

 3: number of model variables:  4    correlation: 0.881    percent: 52.7%
variables:            h        1        3        9
coefficients:   -0.0027   0.0377   0.0443  -0.0887
```

Interreg Danube Region

Co-funded by the European Union

Danube-ADAPT

Thank you for your attention!

Interreg
Danube Region

Co-funded by
the European Union

Danube-ADAPT