Detecting inhomogeneities caused by methodological changes using MASH software

Beatrix Izsák[1], Olivér Szentes[1,2] , Kinga Bokros[1,2]

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

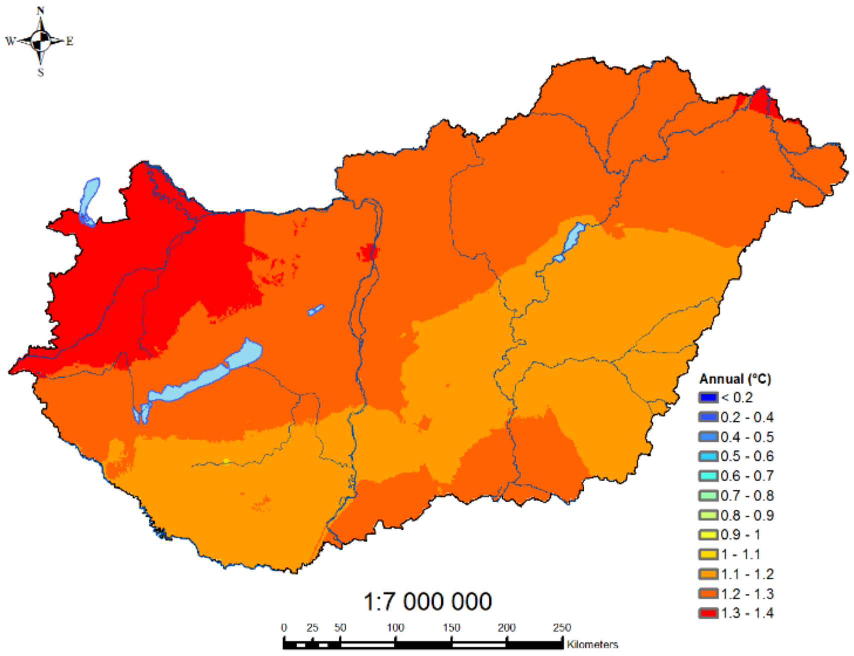
[2]ELTE Eötvös Loránd University, Faculty of Science, Doctoral School of Earth Sciences, Budapest, Hungary

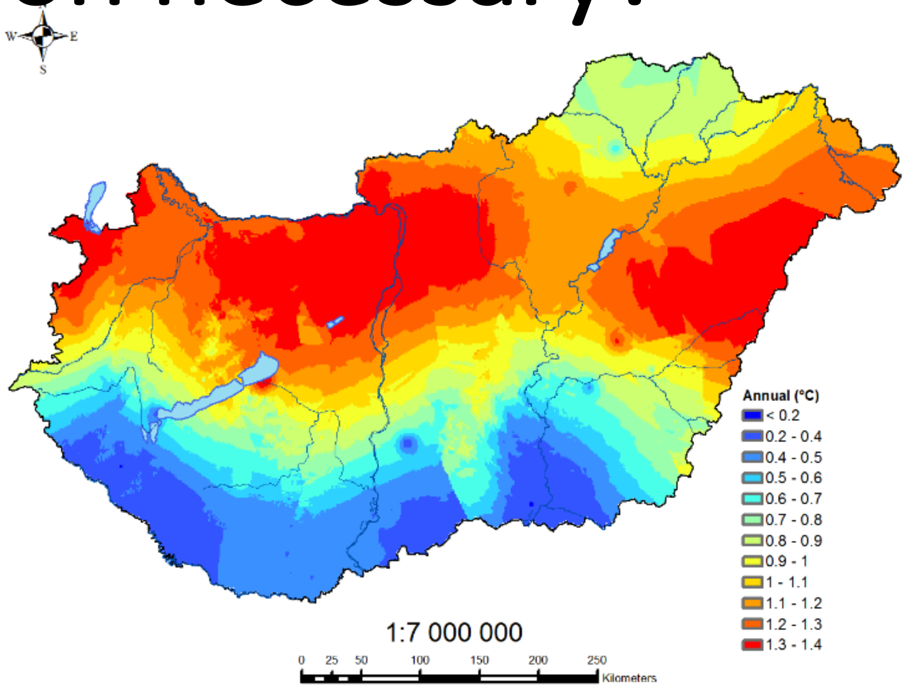
Data Management Workshop 2025. Oslo

HungaroMet

Hungarian Meteorological Service

Is homogenization necessary?





Annual mean temperature, homogenized series:
Estimation of change over the total period 1901-2018 (°C)

Annual mean temperature, raw series:
Estimation of change over the total period 1901-2018 (°C)

MASHv3.03

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

For homogenization, quality control and missing value completion of station daily data series

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

MASHv4.01

For homogenization in mean and standard deviation, quality control and missing value completion of station daily data series

Methodological changes

- Information about methodological changes is not always available. Nowadays, a large amount of paper-based measurements are being digitized in Hungary.
- In the case of temperature, there are 10-12 different methods alone. These newly recorded data must be verified.
- The MASH software also has a very good data check for daily data.
- Station relocation, instrument change and methodological change often occur simultaneously. Often, we cannot find an explanation for large inhomogeneities in the META data.
- In today's presentation, we will present cases where the inhomogeneity can be attributed to a methodological change. We show examples of the inhomogeneities detected by MASH that clearly result from methodological changes.

Statistical spatiotemporal modelling of monthly series in practice
(Monthly series for a given month in a small region)

Relative Additive Model (normal distribution, e.g. temperature)

$$X_j(t) = \mu(t) + E_j + IH_j(t) + \varepsilon_j(t) \qquad \left(j = 1,2,\ldots,N;\ t = 1,2,\ldots.,n \right)$$

Relative Multiplicative Model (e.g. precipitation)

$$X_j(t) = \mu(t) \cdot E_j \cdot IH_j(t) \cdot \exp\left(\varepsilon_j(t)\right)$$

μ : unknown climate change signal; E : spatial trend;
IH : inhomogeneity signal; ε : normal noise

Type of inhomogeneity $IH(t)$ in general: 'step-like function'

Noise $\boldsymbol{\varepsilon}(t) = \left[\varepsilon_1(t),\ldots..,\varepsilon_N(t)\right]^{\mathrm{T}} \in N(\mathbf{0},\mathbf{C})$ $(t = 1,\ldots,n)$ are independent

Difference series constitution (Relative model)

Aim: to filter out the unknown climate signal $\mu(t)$.

$X_j(t)$: candidate series ; $X_i(t)\ (i \neq j)$: reference series

Difference series for reference series $X_i(t)\ (i \neq j)$:

$$Z_j(t) = X_j(t) - \sum_{i \neq j} \lambda_{ji} X_i(t) = IH_j(t) - \sum_{i \neq j} \lambda_{ji} IH_i(t) + \varepsilon_{Z_j}(t)$$

where sum of weighting factors: $\sum_{i \neq j} \lambda_{ji} = 1$

Difference series: $Z(t) = IH_Z(t) + \varepsilon_Z(t) \qquad (t = 1,..,n)$,

$IH_Z(t)$: inhomogeneity with K break points, $T_1 < T_2 < ..... < T_K$

$\varepsilon_Z(t) \in N(E_Z, \sigma_Z^2)\ (t = 1,...,n)$ are independent

Optimal difference series constitution (MASH)
(interpolation between series)

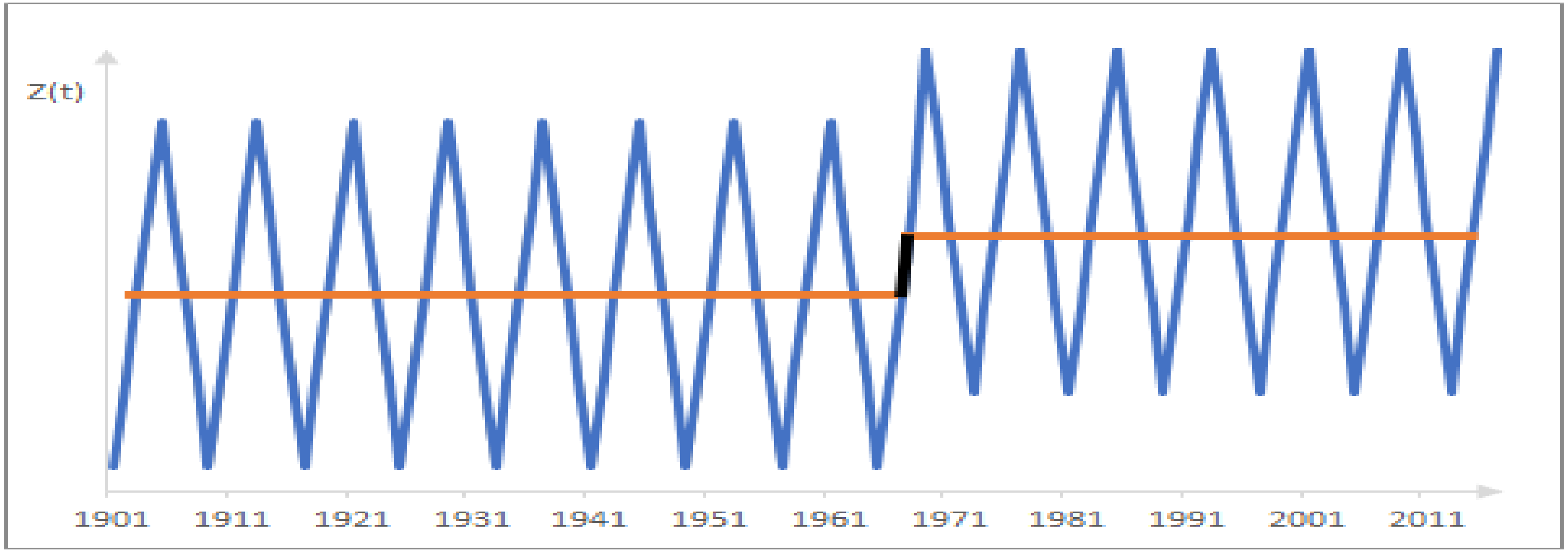
Minimization of the variance of noise $\mathrm{V}\left(\varepsilon_{Z_j}\right) = \mathrm{V}\left(Z_j\right)$.

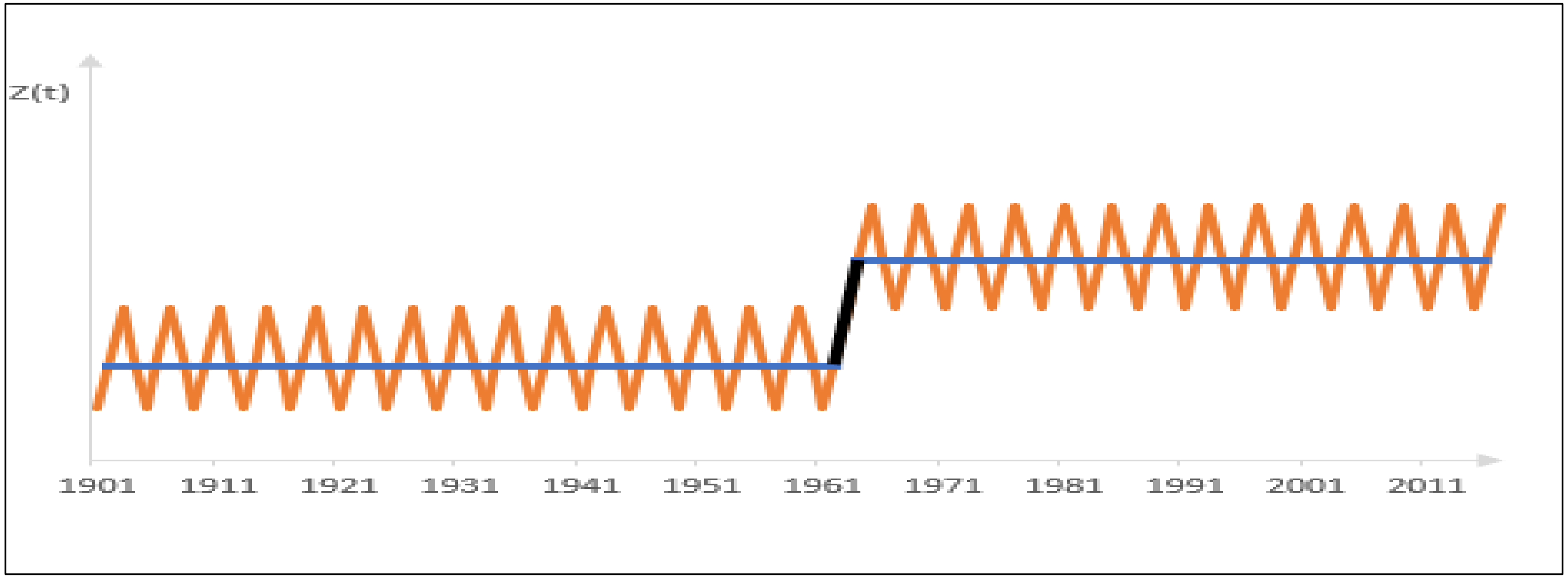
The optimal weighting factors $\lambda_{ji}\ (i \neq j)$ in vector form:

$$\boldsymbol{\lambda}_j = \mathbf{C}_{j,ref}^{-1}\left(\mathbf{c}_{j,ref} + \frac{\left(1 - \mathbf{1}^{\mathrm{T}}\mathbf{C}_{j,ref}^{-1}\mathbf{c}_{j,ref}\right)}{\mathbf{1}^{\mathrm{T}}\mathbf{C}_{j,ref}^{-1}\mathbf{1}}\mathbf{1}\right)$$

$\mathbf{c}_{j,ref}$: candidate-reference covariance vector,

$\mathbf{C}_{j,ref}$: reference-reference covariance matrix


Z(t)
1901
1911
1921
1931
1941
1951
1961
1971
1981
1991
2001
2011


Z(t)
1901
1911
1921
1931
1941
1951
1961
1971
1981
1991
2001
2011

MASH: Multiple break points detection

- **Break point (changepoint) detection for monthly series**

Examination (more) difference series to detect the break points and to attribute (separate) for the candidate series.

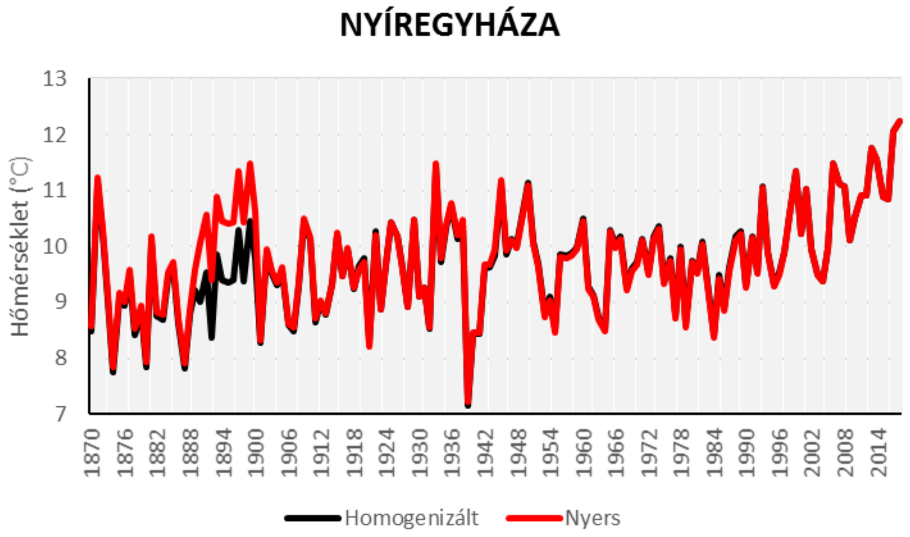
- Multiple break points detection based on **Test of Hypothesis,** confidence intervals for the break points, that make possible automatic use of metadata: MASH (*Szentimrey*)

Algorithm of MASH iteration procedure

1. To choose the candidate series
2. Series comparison:
   constitution of optimal difference series system
3. Break points detection for difference series:
   hypothesis testing, point estimation, confidence intervals
4. Estimation of shifts for difference series:
   point estimation, confidence intervals
5. Analysis of results: separation of break points
   and shifts for candidate series
6. Correction of candidate series if it is possible

Iteration of steps 1-6! Each series is examined many times!

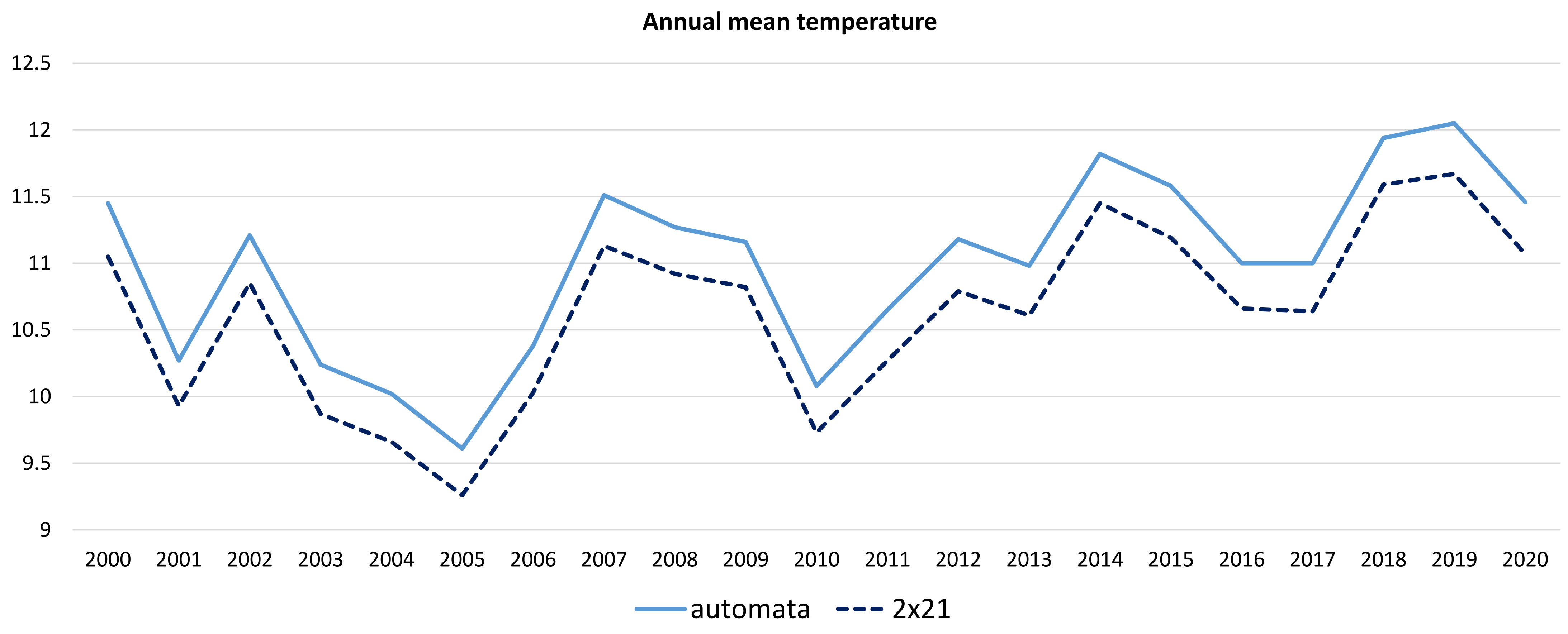
Change of measurement times



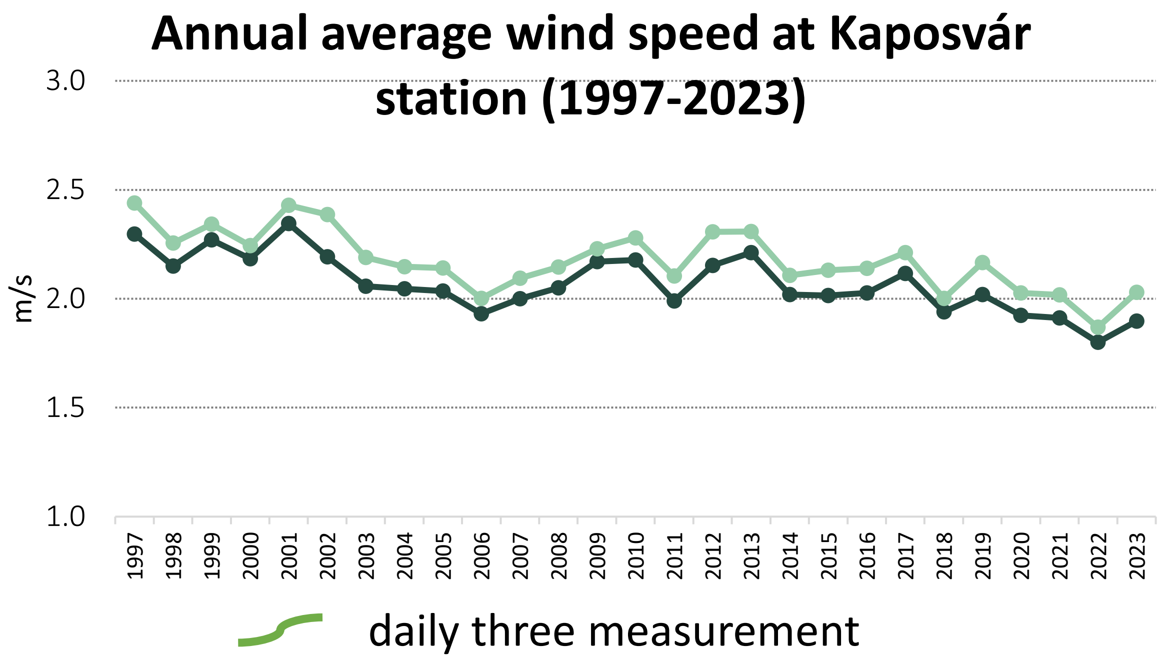
In Nyíregyháza, the change in measurement times causes inhomogeneity of the same magnitude as the detected climate change.

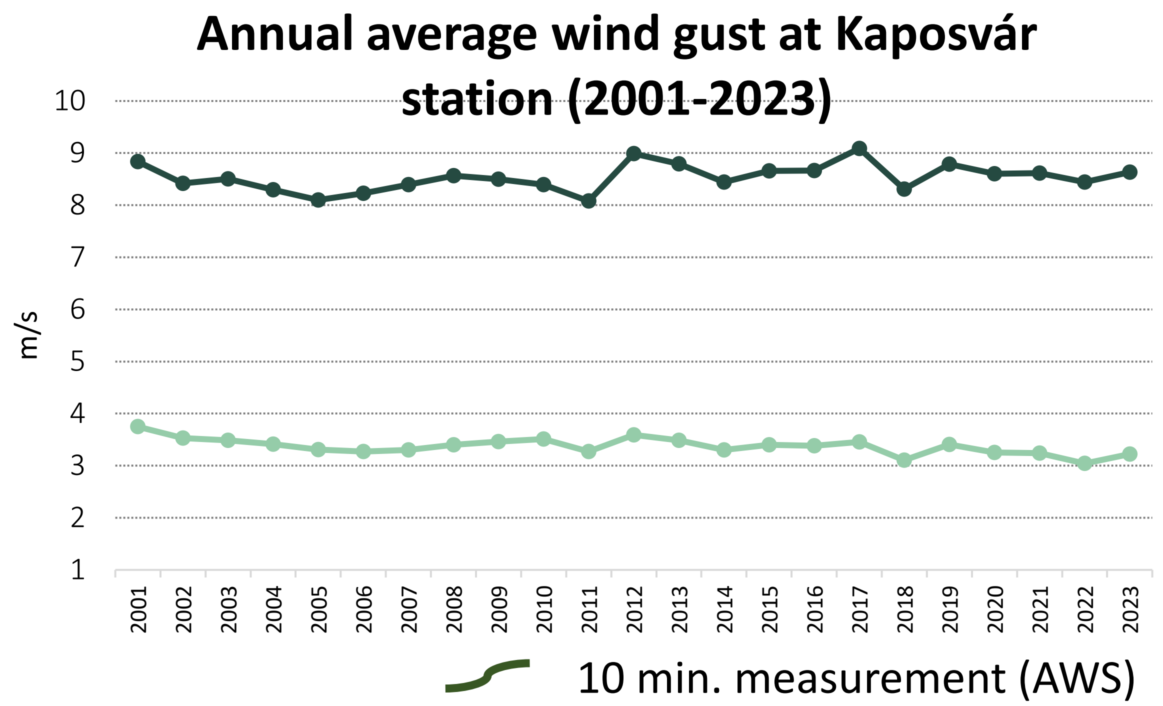
Annual mean temperature values calculated from raw and homogenized data series, at Nyíregyháza station. The graph clearly shows that from 1890 to 1901 the homogenized series is well below the raw data series, which is caused by the fact that the observations were made at different times. The morning measurement is 1 hour later, the evening measurement is 1 hour earlier.

Automatic measurements and average of 06,12 and 2x21 h measurements (°C)



Annual average wind and gusts calculated from three daily measurements and automatic measurements



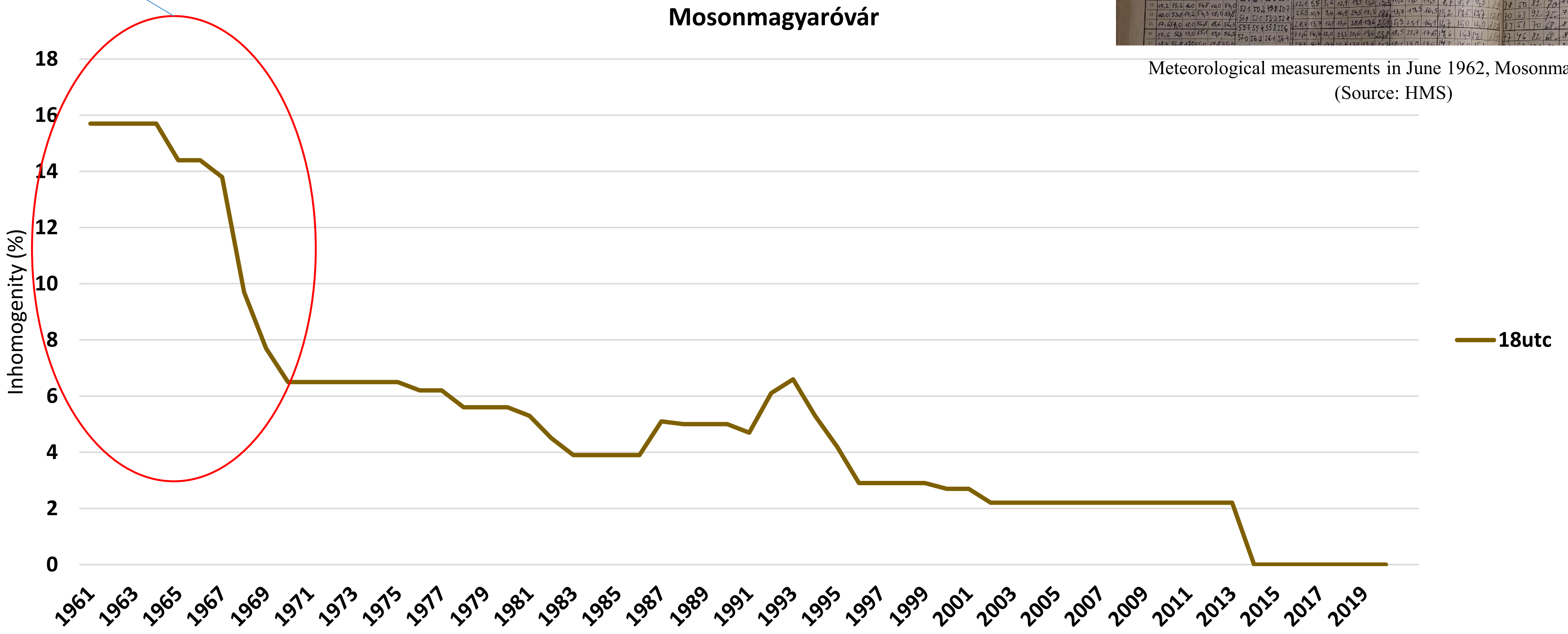


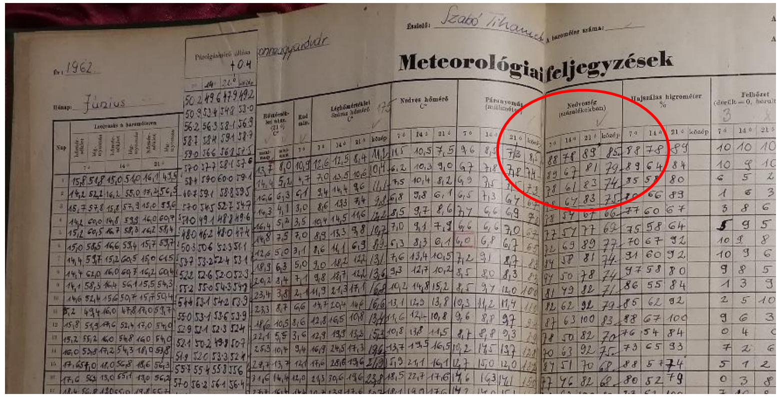
In the case of wind measurements, the methodological change always occurred together with the instrument change and relocation. Here we can only estimate how much of the detected inhomogeneity is due to the methodological change.

HungaroMet Nonprofit Zrt.

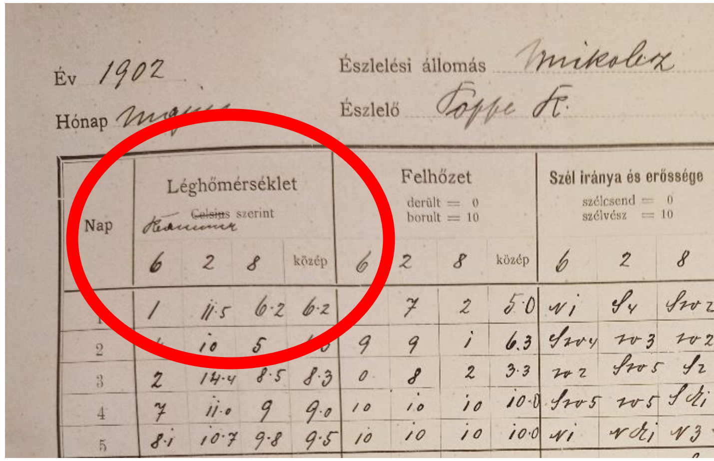
Relative humidity

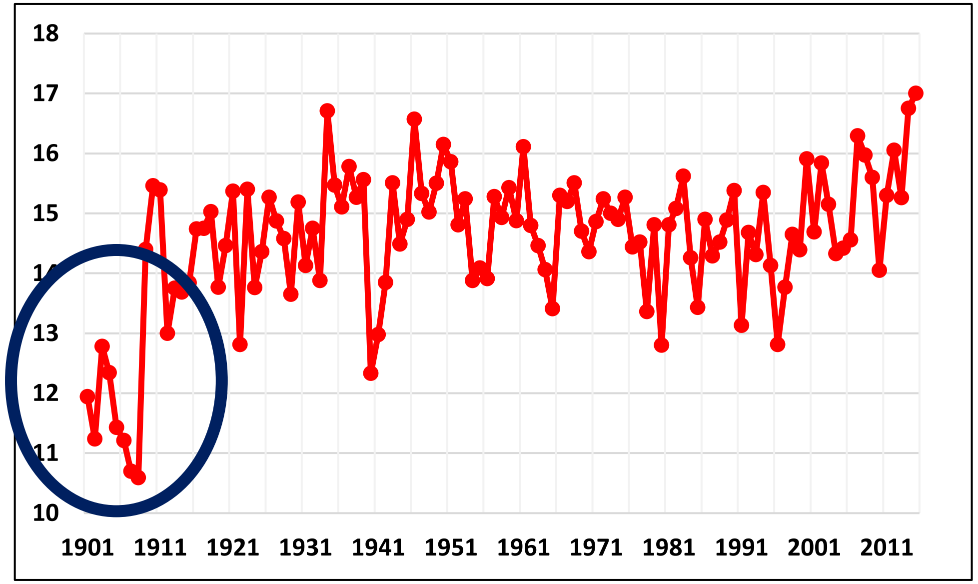
Evening measurement taken 3 hours later



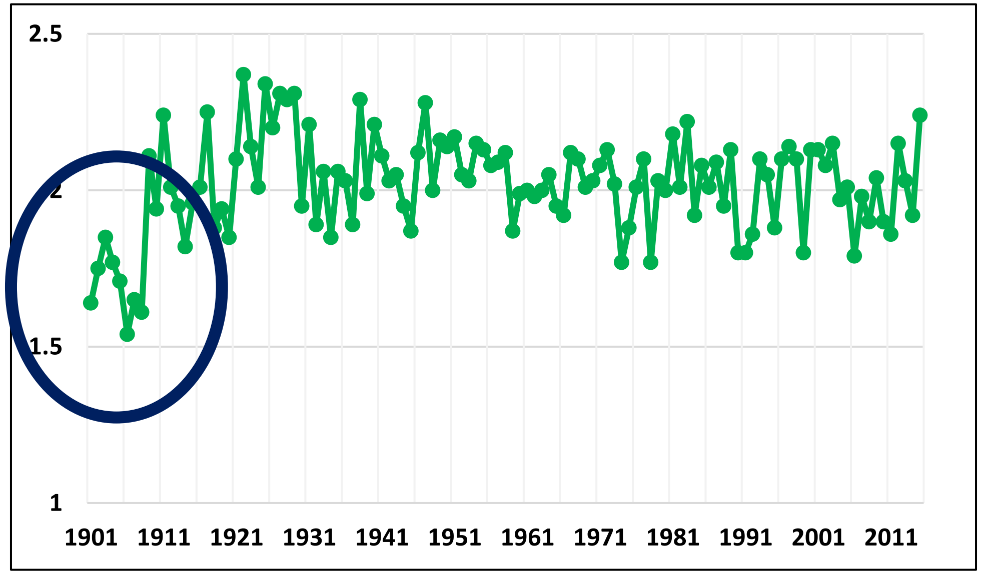


Meteorological measurements in June 1962, Mosonmagyaróvár (Source: HMS)





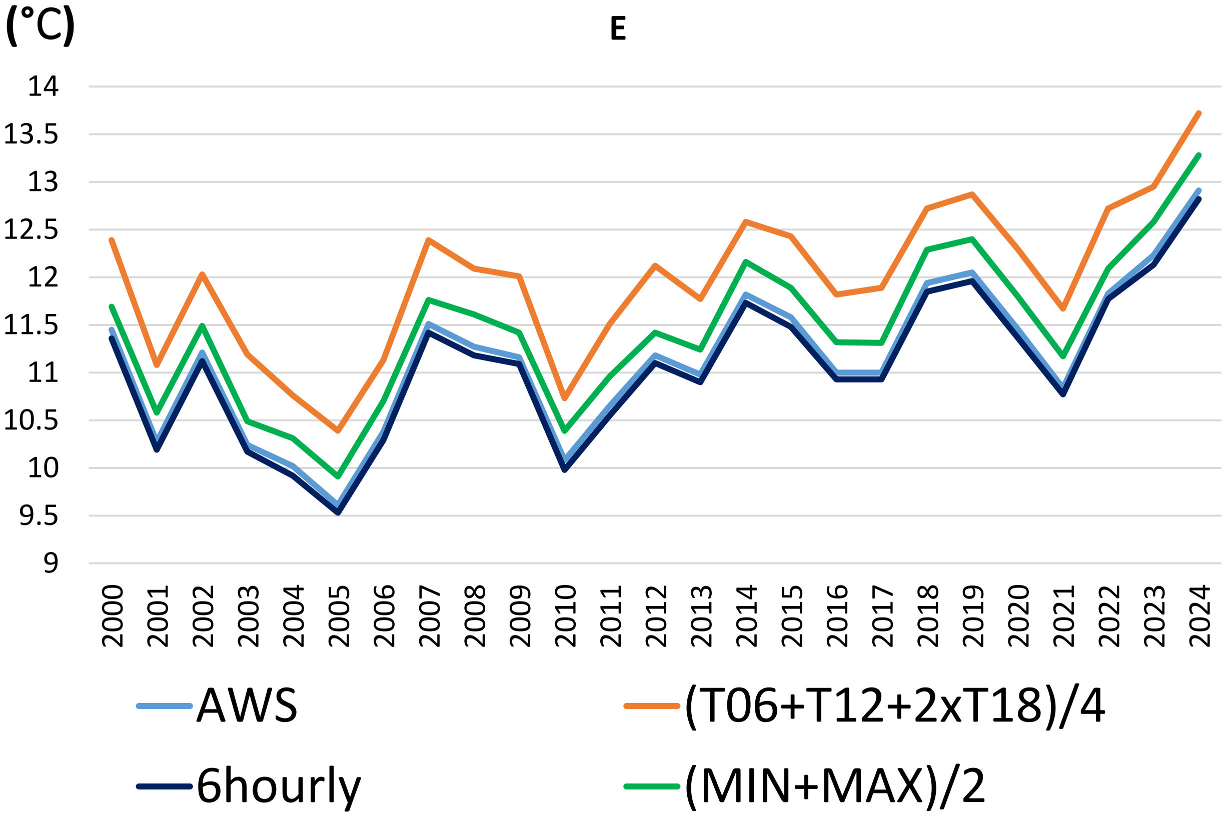
Annual average series of monthly series $\bar{Y}(t)$ for characterization of annual mean (*E*).



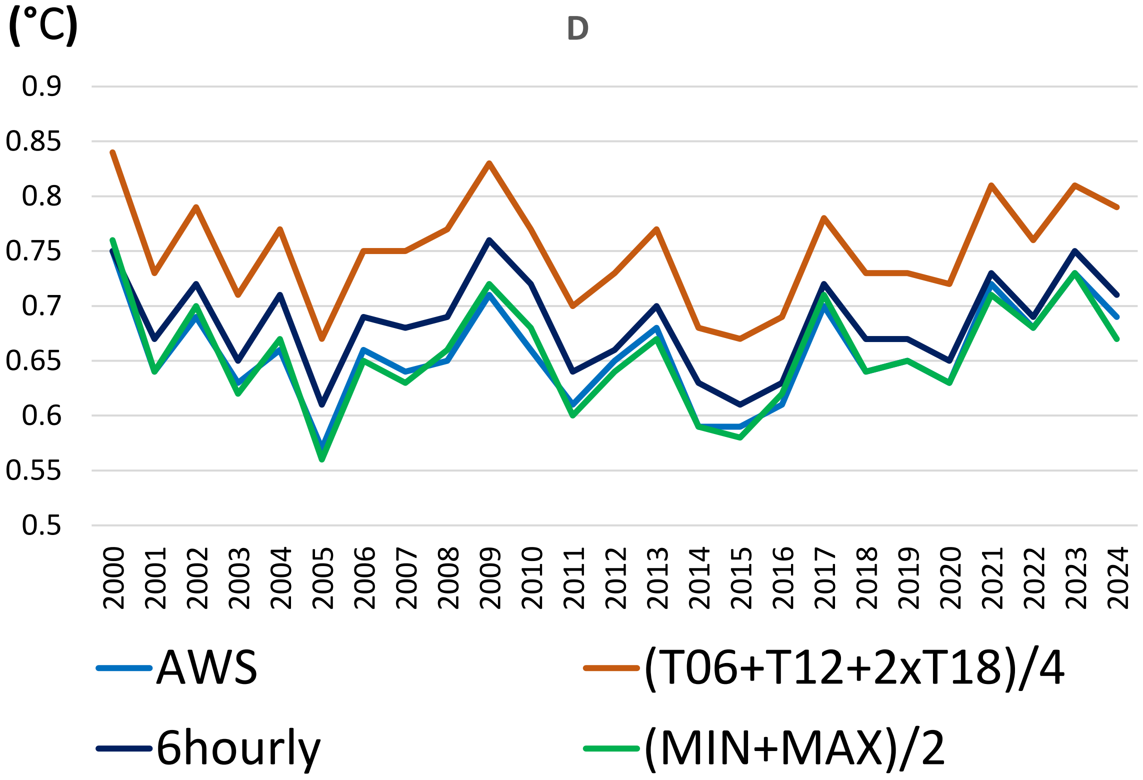
Series for characterization of annual average of the monthly standard deviations (*D*) based on series *S(t).*

Inhomogeneity in 1901-1908, measured in Réaumur: 1 °C=0.8 Re

Not only does the expected value change with the change in methodology, but so does the standard deviation.



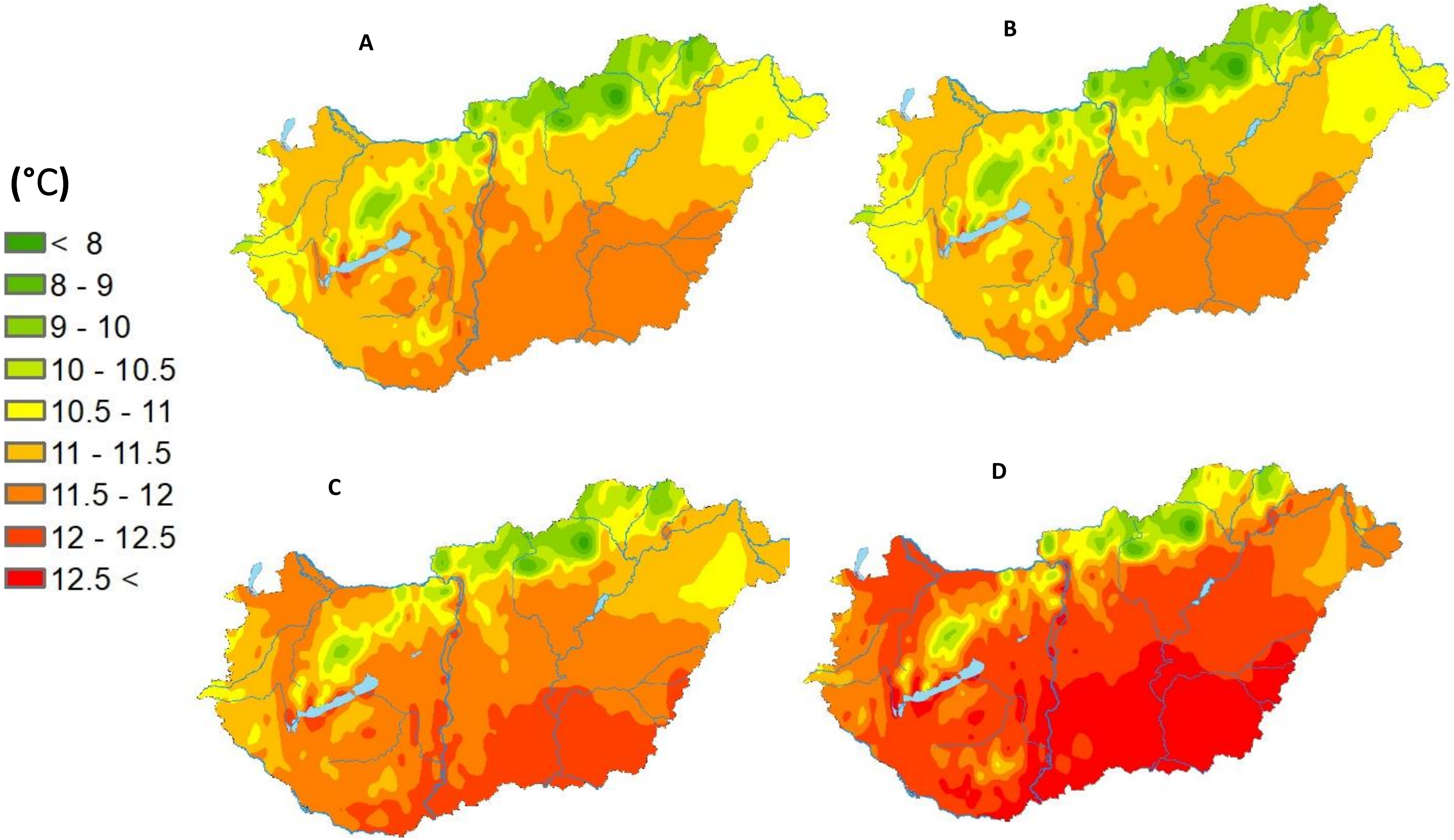
Annual average temperature, 2000-2024



Spatial standard deviation of annual mean temperature (empirical standard deviation), national average, 2000–2024

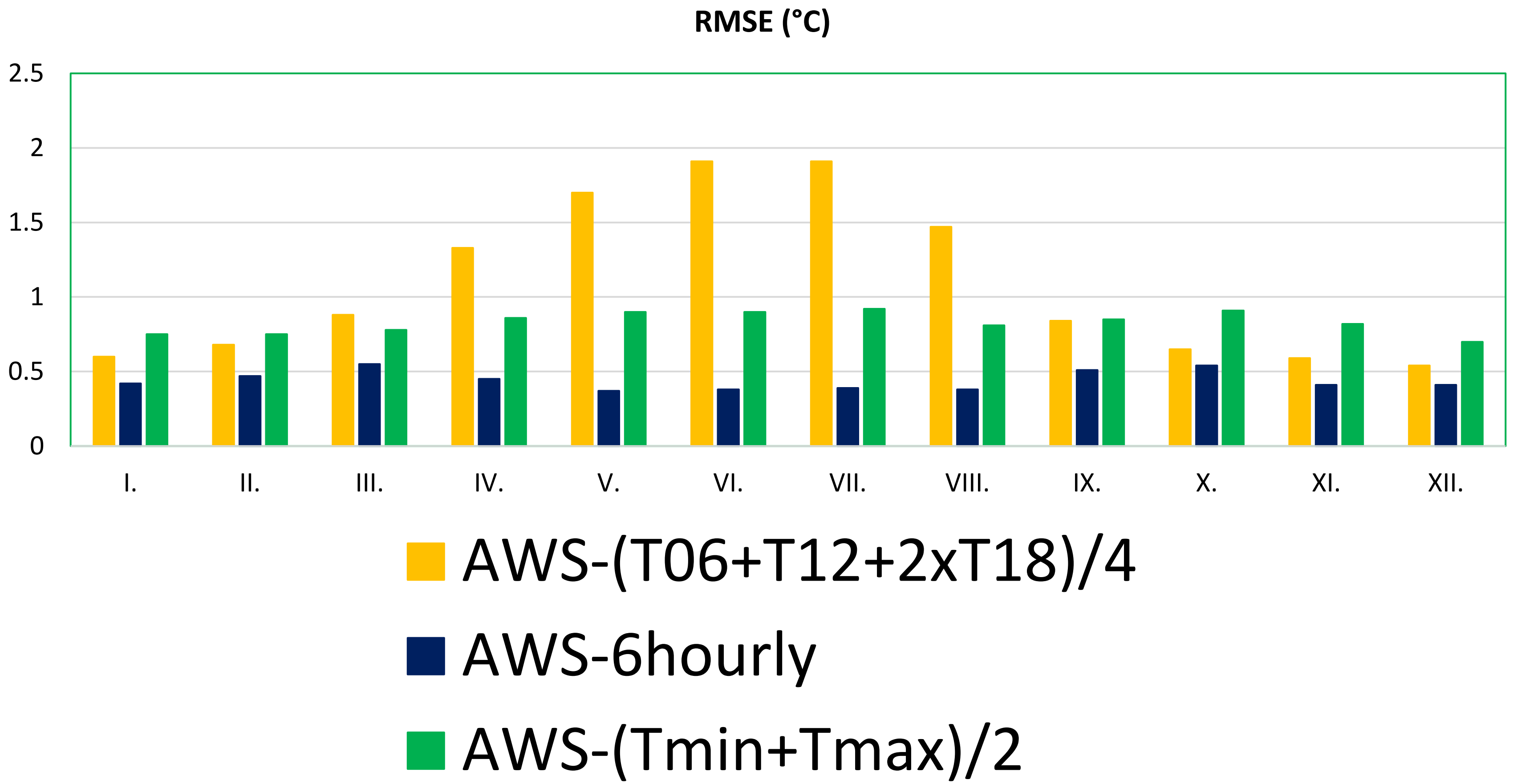
Today, with automatic measurements, we can simulate the change in measurement time in a simple way.

The impact of methodological changes is also different in terms of space.



Average annual temperature 2000-2025, A: automatic measurements, B: 4 measurements per day, C: (Tmin+Tmax)/2, D: evening measurements doubled

The methodological change has an annual cycle.



Software MASHv4.01 (Multiple Analysis of Series for Homogenization)

(**2023**, *T. Szentimrey*)

The MASH system is based on homogenization of monthly series derived from daily series. The procedures depend on the distribution of climate elements.

Quasi normal distribution (e.g. temperature)

Beside the monthly mean series another type of monthly series is also derived to estimate the inhomogeneity of standard deviation (D). These series are homogenized in standard deviation (D) by multiplicative model. The monthly mean series adjusted with the estimated inhomogeneity of standard deviation (D) are homogenized by additive model in mean (E).

Quasi lognormal distribution (e.g. precipitation)

Monthly mean or sum series are homogenized by multiplicative model.

Theorem **(Problem of inhomogeneity of the standard deviation)**

Daily data: $Y(t)$ $(t = 1,..,30)$, monthly mean: $\bar{Y} = \frac{1}{30}\sum_{t=1}^{30} Y(t)$

Monthly variable for examination of standard deviation (D): $S = \sqrt{\frac{1}{29}\sum_{t=2}^{30}(Y(t) - Y(t-1))^2}$

Daily data with inhomogeneity in mean (*E*) and standard deviation (*D*):

$E(Y_{ih}(t)) = E(Y(t)) + \beta$,

$D(Y_{ih}(t)) = \alpha \cdot D(Y(t))$ $(t = 1,..,30)$

The appropriate monthly variables: $\bar{Y}_{ih} = \frac{1}{30}\sum_{t=1}^{30} Y_{ih}(t)$,

$S_{ih} = \sqrt{\frac{1}{29}\sum_{t=2}^{30}(Y_{ih}(t) - Y_{ih}(t-1))^2}$

i, **Then the monthly mean is also inhomogeneous in mean (*E*) and st. deviation (*D*):**

$E(\bar{Y}_{ih}) = E(\bar{Y}) + \beta$ and $D(\bar{Y}_{ih}) = \alpha \cdot D(\bar{Y})$

ii, **Moreover variable S_{ih} can be used to estimate the inhomogeneity of st. deviation (*D*):**

$E(S_{ih}) = \alpha \cdot E(S)$

Steps (Homogenization of monthly series $\mathbf{S(t)}$, $\bar{\mathbf{Y}}\mathbf{(t)}$):

- Homogenization of series $S(t)$ by multiplicative model: break points detection, estimation of inhomogeneity of standard deviation (D).
- Adjustment of standard deviation of series $\bar{Y}(t)$.
- Homogenization of adjusted series $\bar{Y}(t)$ by additive model: break points detection, estimation of the inhomogeneity of mean (E).
- Adjustment of mean of series $\bar{Y}(t)$.

Assumption: homogeneity in higher order (>2) moments. This assumption is always right in case of normal distribution!

HOMOGENIZATION OF DAILY SERIES

- Homogenization of mean (E) and standard deviation (D) on the basis of the monthly results.
- The used monthly information are the break points and the monthly adjustments of the mean (E) and standard deviation (D).
- If the daily data are normally distributed then there is no inhomogeneity in the higher order moments according to *Theorem 2* in the papers of *Szentimrey*, 2021, 2023.

- Szentimrey, T., 2021: Mathematical questions of homogenization and summary of MASH, Proceedings of the 10**th** Seminar for Homogenization and Quality Control in Climatological Databases and 5th Conference on Spatial Interpolation Techniques in Climatology and Meteorology (Ed. Lakatos M, Hoffmann L, Kircsi A, Szentimrey T), Budapest, Hungary, 2020, WCDMP-No. 86, pp. 4-17
- Szentimrey, T., 2023: Overview of mathematical background of homogenization, summary of method MASH and comments on benchmark validation, International Journal of Climatology, 43(13), 6314–6329 https://doi.org/10.1002/joc.8207

META data PÉCS

Kezdő dátum: 1856 1 1

Befejező dátum: 2024 5 10

Sorozat típusa:
- (•) Hőmérséklet
- ( ) Csapadék
- ( ) Napfény (észlelős elemek)

Típus részletezés:
- ( ) sor pótlás nélkül
- (•) sor pótlással
- ( ) pótlás és típusrészletezés

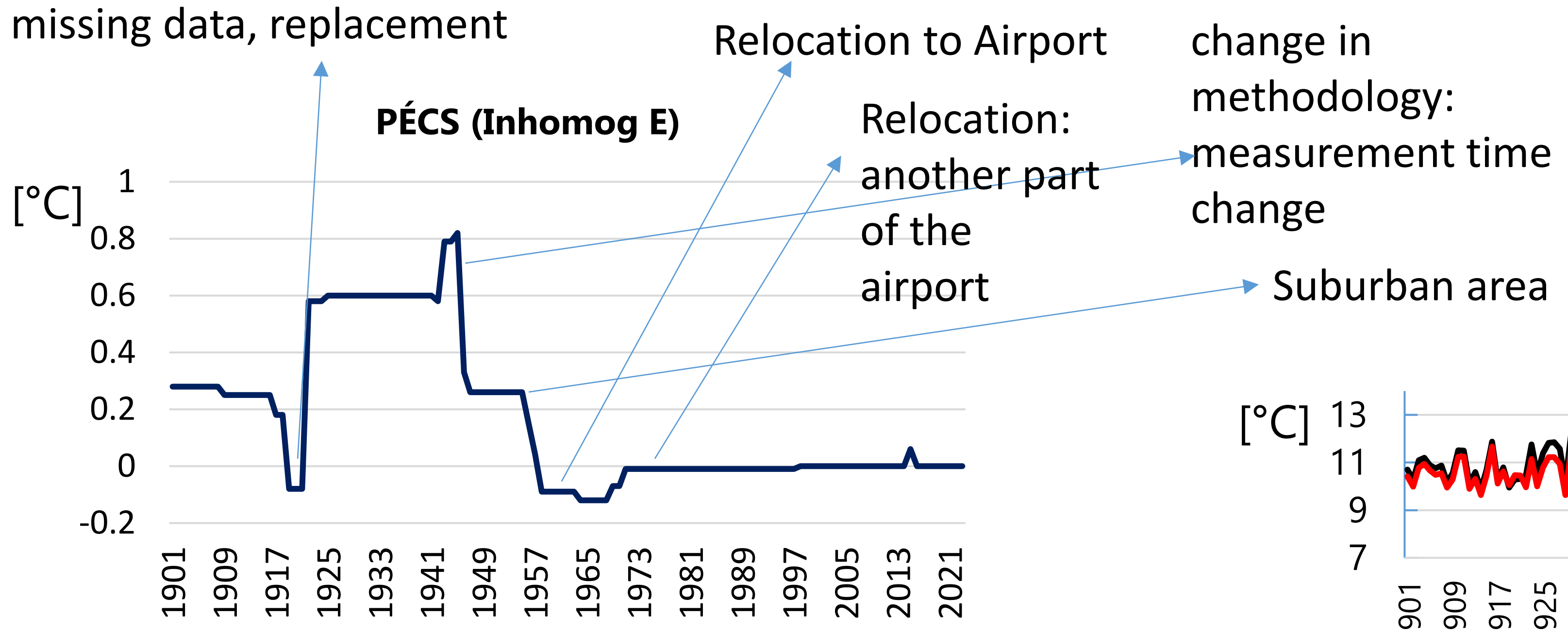
Sorozat

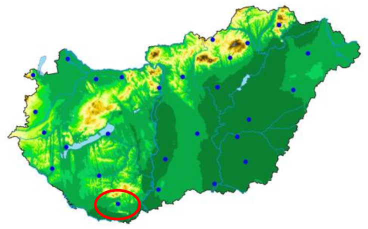
Lekérés típusa: Adott állomás

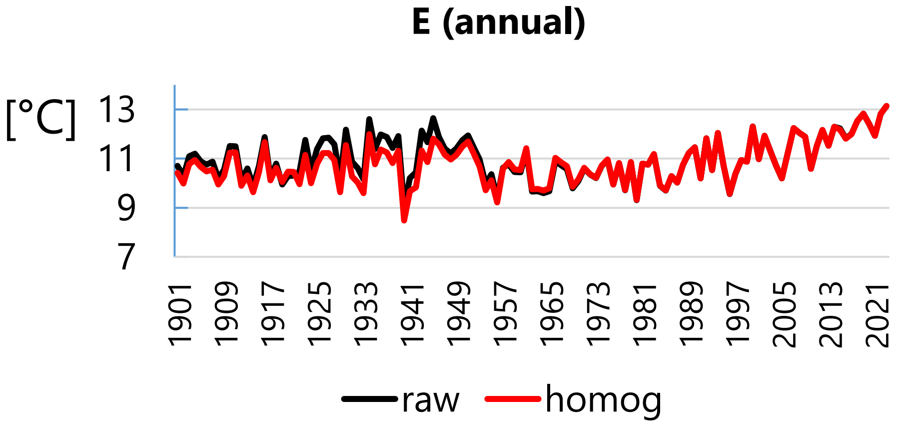
Állomás száma: 39115

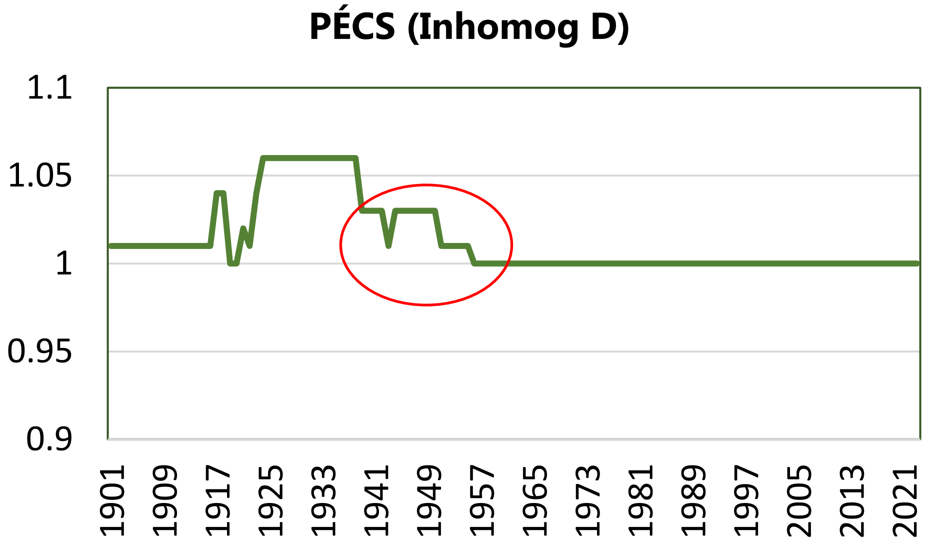
Állomás neve: Pogány repülőtér Repülőtér

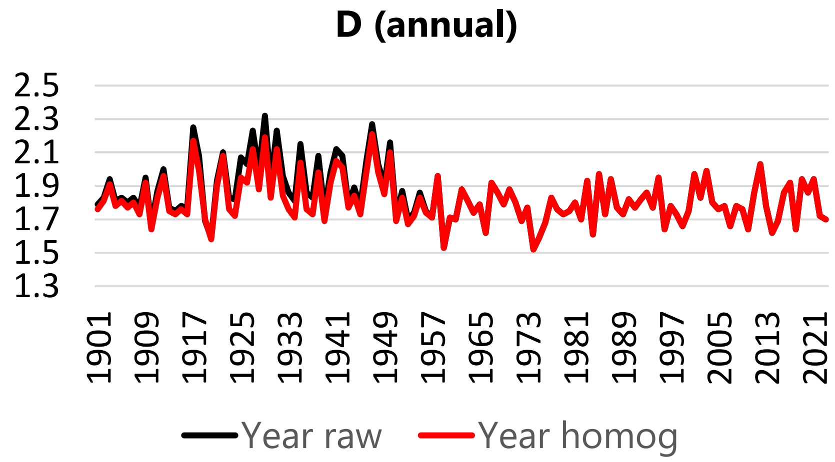
Állomásszám	Állomásnév	Időszak kezdete	Időszak vége	Állomás típusa	Pótló állomás
38509	Pécs Fertőtlenítő Intézet	1916. 09. 01 00:00	1918. 11. 30 23:59	K3	
38304	Szálka Montenuovo hercegi erdőbirtok erdőgondnoksága	1918. 12. 01 00:00	1919. 01. 07 23:59	K3	F0
38509	Pécs Fertőtlenítő Intézet	1919. 01. 08 00:00	1919. 05. 20 23:59	K3	
38304	Szálka Montenuovo hercegi erdőbirtok erdőgondnoksága	1919. 05. 21 00:00	1921. 11. 21 23:59	K3	F0
38512	Pécs Erzsébet Egyetem Főreáliskola / Erzsébet Egyetem	1921. 11. 22 00:00	1922. 06. 25 23:59	K3	
38510	Pécs Notre Dame zárda	1922. 06. 26 00:00	1924. 10. 21 23:59	K3	F0
38512	Pécs Erzsébet Egyetem Főreáliskola / Erzsébet Egyetem	1924. 10. 22 00:00	1943. 06. 30 23:59	K3	
38531	Pécs Pedagógiai Főiskola Pedagógiai Főiskola	1943. 07. 01 00:00	1943. 09. 05 23:59	K3	F0
38512	Pécs Erzsébet Egyetem Főreáliskola / Erzsébet Egyetem	1943. 09. 06 00:00	1943. 09. 30 23:59	K3	
38531	Pécs Pedagógiai Főiskola Pedagógiai Főiskola	1943. 10. 01 00:00	1943. 10. 13 23:59	K3	F0
38512	Pécs Erzsébet Egyetem Főreáliskola / Erzsébet Egyetem	1943. 10. 14 00:00	1943. 10. 31 23:59	K3	
38531	Pécs Pedagógiai Főiskola Pedagógiai Főiskola	1943. 11. 01 00:00	1944. 02. 23 23:59	K3	F0
38512	Pécs Erzsébet Egyetem Főreáliskola / Erzsébet Egyetem	1944. 02. 24 00:00	1944. 07. 31 23:59	K3	
38531	Pécs Pedagógiai Főiskola Pedagógiai Főiskola	1944. 08. 01 00:00	1944. 08. 31 23:59	K3	F0
38512	Pécs Erzsébet Egyetem Főreáliskola / Erzsébet Egyetem	1944. 09. 01 00:00	1944. 10. 31 23:59	K3	
38531	Pécs Pedagógiai Főiskola Pedagógiai Főiskola	1944. 11. 01 00:00	1944. 11. 30 23:59	K3	F0
37706	Szekszárd	1944. 12. 01 00:00	1945. 04. 30 23:59	K3	F0
38512	Pécs Erzsébet Egyetem Főreáliskola / Erzsébet Egyetem	1945. 05. 01 00:00	1946. 02. 18 23:59	K3	
38506	Pécs Mecsekalji Repülőtér	1946. 02. 19 00:00	1956. 07. 06 23:59	K3S2K8	
38507	Pécs Pogány Repülőtér	1956. 07. 07 00:00	1969. 03. 03 23:59	S2K8Z1	
38527	Pogány Repülőtér adóház	1969. 03. 04 00:00	1998. 03. 24 14:59	S2Z1K8S1	
38522	Pogány Repülőtér	1998. 03. 24 15:00	2004. 03. 31 08:59	S2Z1S1	
39113	Pogány Repülőtér	2004. 03. 31 09:00	2021. 08. 02 08:59	S1S2Z1MS	
39115	Pogány repülőtér Repülőtér	2021. 08. 02 09:00	2024. 05. 10 09:01	S1MA	


missing data, replacement
Relocation to Airport
Relocation: another part of the airport
change in methodology: measurement time change
Suburban area
PÉCS (Inhomog E)
[°C]
1
0.8
0.6
0.4
0.2
0
-0.2
1901 1909 1917 1925 1933 1941 1949 1957 1965 1973 1981 1989 1997 2005 2013 2021

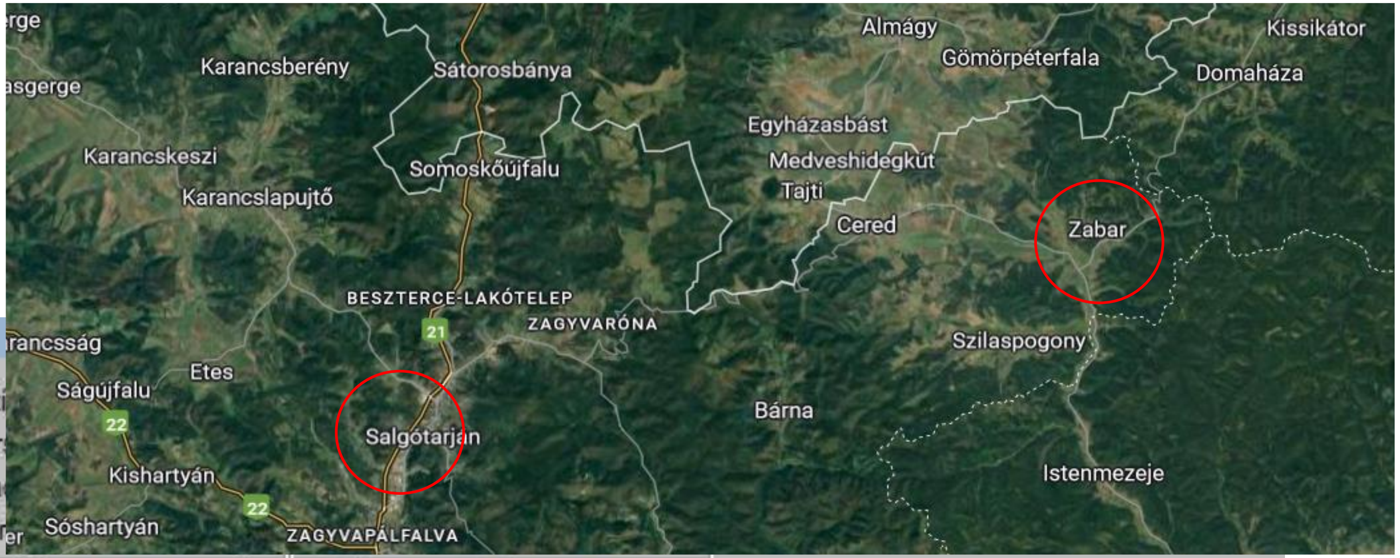


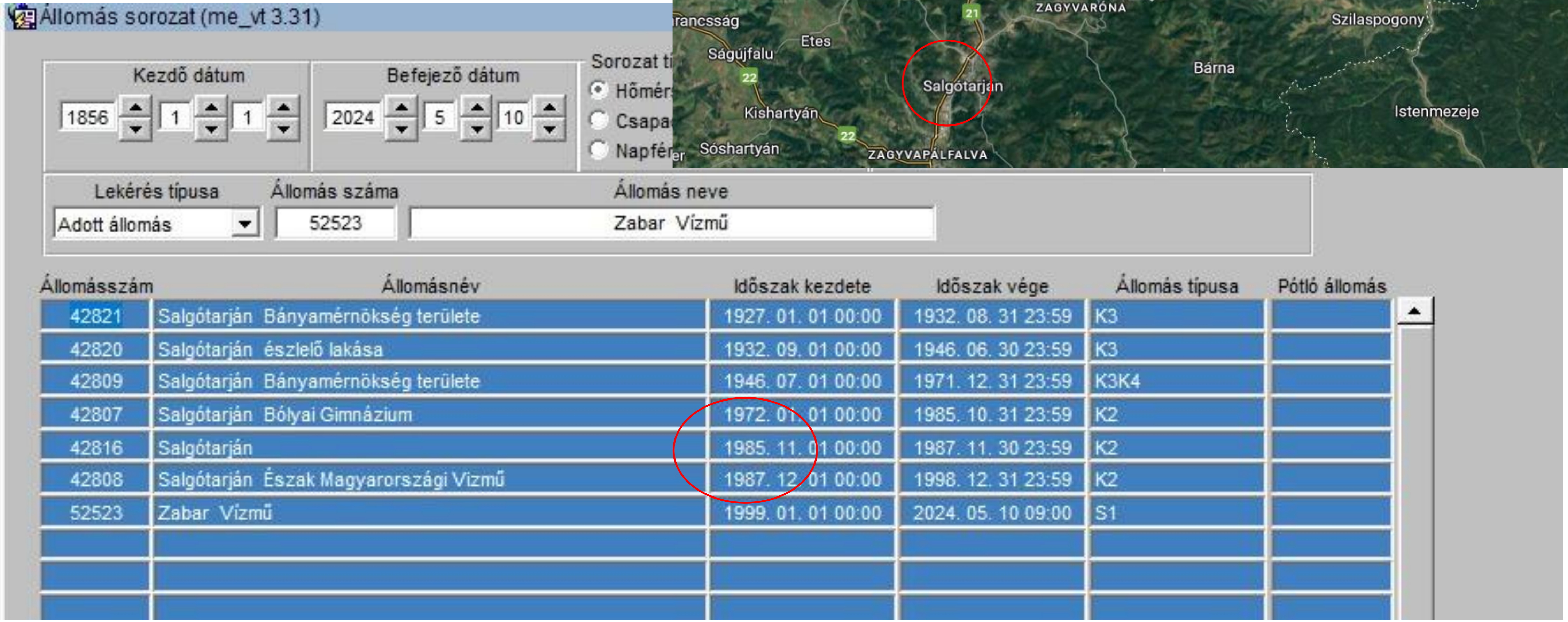

E (annual)
[°C]
13
11
9
7
1901 1909 1917 1925 1933 1941 1949 1957 1965 1973 1981 1989 1997 2005 2013 2021
raw
homog


PÉCS (Inhomog D)
1.1
1.05
1
0.95
0.9
1901 1909 1917 1925 1933 1941 1949 1957 1965 1973 1981 1989 1997 2005 2013 2021


D (annual)
2.5
2.3
2.1
1.9
1.7
1.5
1.3
1901 1909 1917 1925 1933 1941 1949 1957 1965 1973 1981 1989 1997 2005 2013 2021
Year raw
Year homog

META data ZABAR





Állomás sorozat (me_vt 3.31)

Kezdő dátum: 1856 1 1

Befejező dátum: 2024 5 10

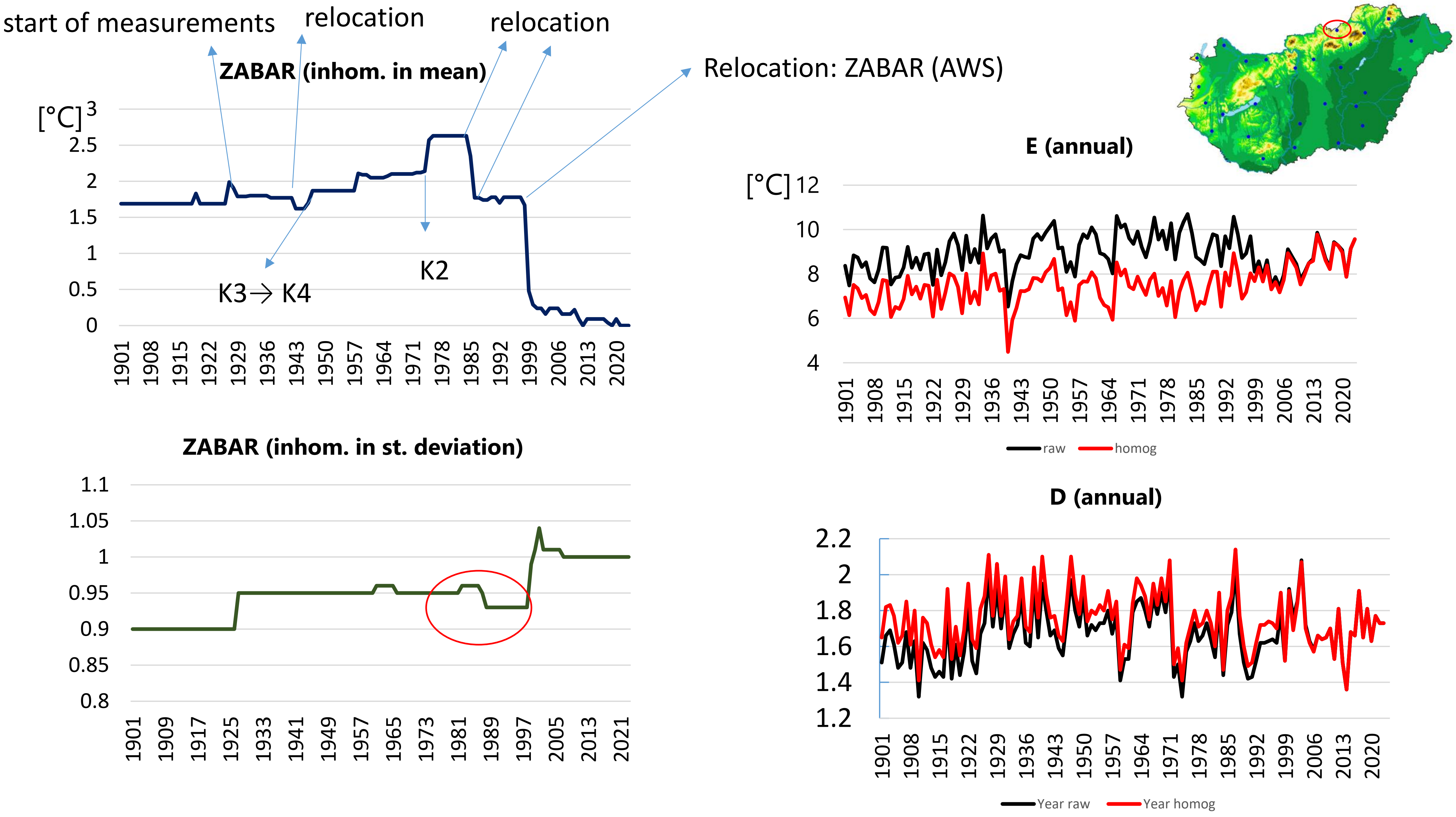
Sorozat ti…

Lekérés típusa: Adott állomás

Állomás száma: 52523

Állomás neve: Zabar Vízmű

Állomásszám	Állomásnév	Időszak kezdete	Időszak vége	Állomás típusa	Pótló állomás
42821	Salgótarján Bányamérnökség területe	1927. 01. 01 00:00	1932. 08. 31 23:59	K3	
42820	Salgótarján észlelő lakása	1932. 09. 01 00:00	1946. 06. 30 23:59	K3	
42809	Salgótarján Bányamérnökség területe	1946. 07. 01 00:00	1971. 12. 31 23:59	K3K4	
42807	Salgótarján Bólyai Gimnázium	1972. 01. 01 00:00	1985. 10. 31 23:59	K2	
42816	Salgótarján	1985. 11. 01 00:00	1987. 11. 30 23:59	K2	
42808	Salgótarján Észak Magyarországi Vízmű	1987. 12. 01 00:00	1998. 12. 31 23:59	K2	
52523	Zabar Vízmű	1999. 01. 01 00:00	2024. 05. 10 09:00	S1	

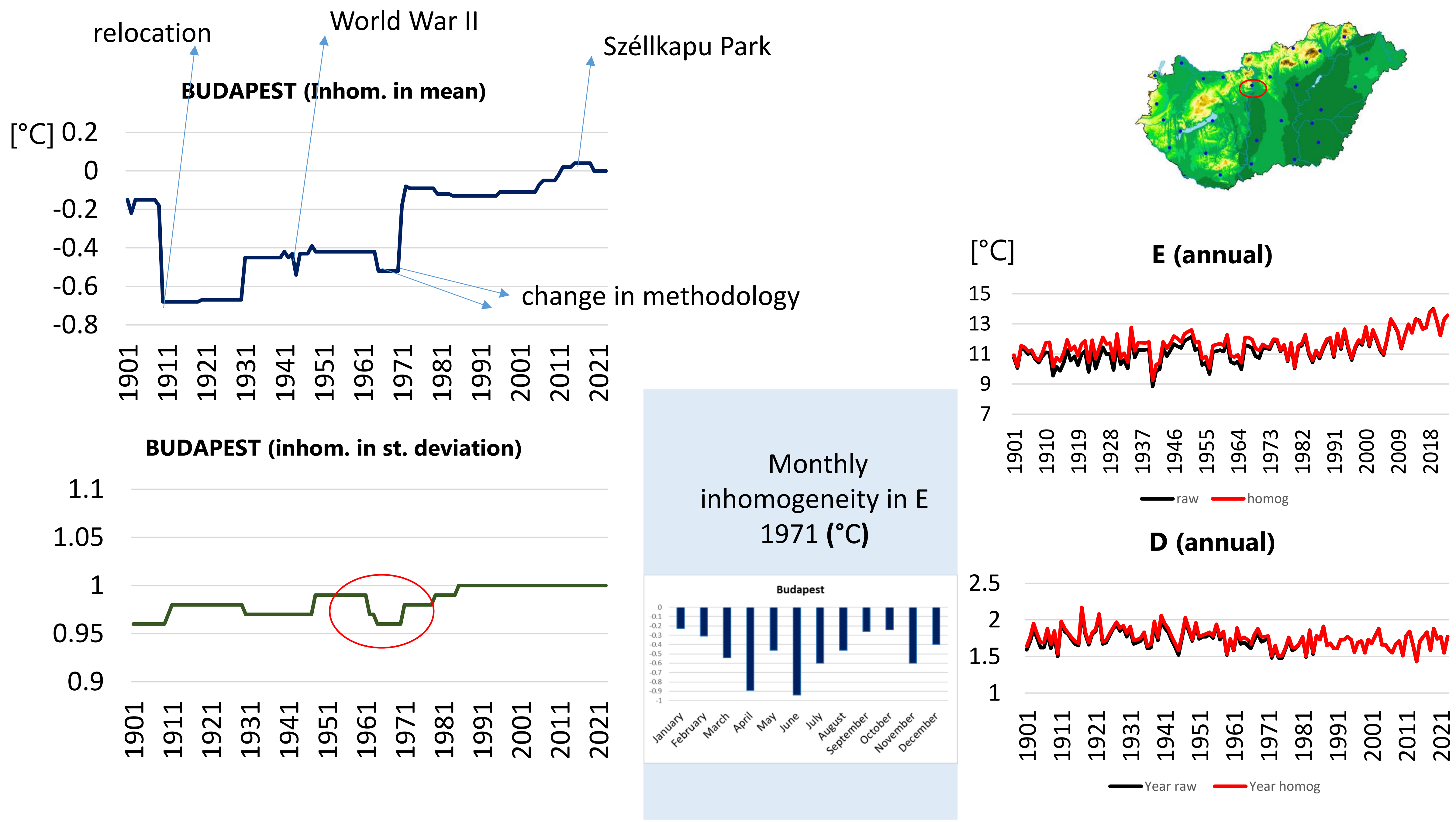

start of measurements
relocation
relocation
ZABAR (inhom. in mean)
Relocation: ZABAR (AWS)
[°C]
K3→ K4
K2
E (annual)
[°C]
raw
homog
ZABAR (inhom. in st. deviation)
D (annual)
Year raw
Year homog

Széllkapu Park

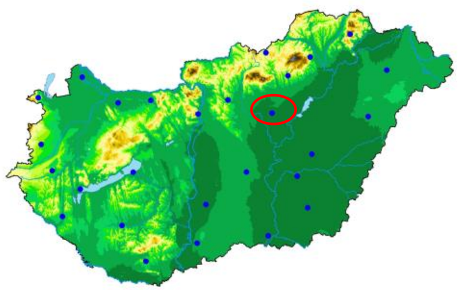
Lekérés típusa	Állomás száma	Állomás neve
Adott állomás	44121	Budapest belterület OMSZ Torony

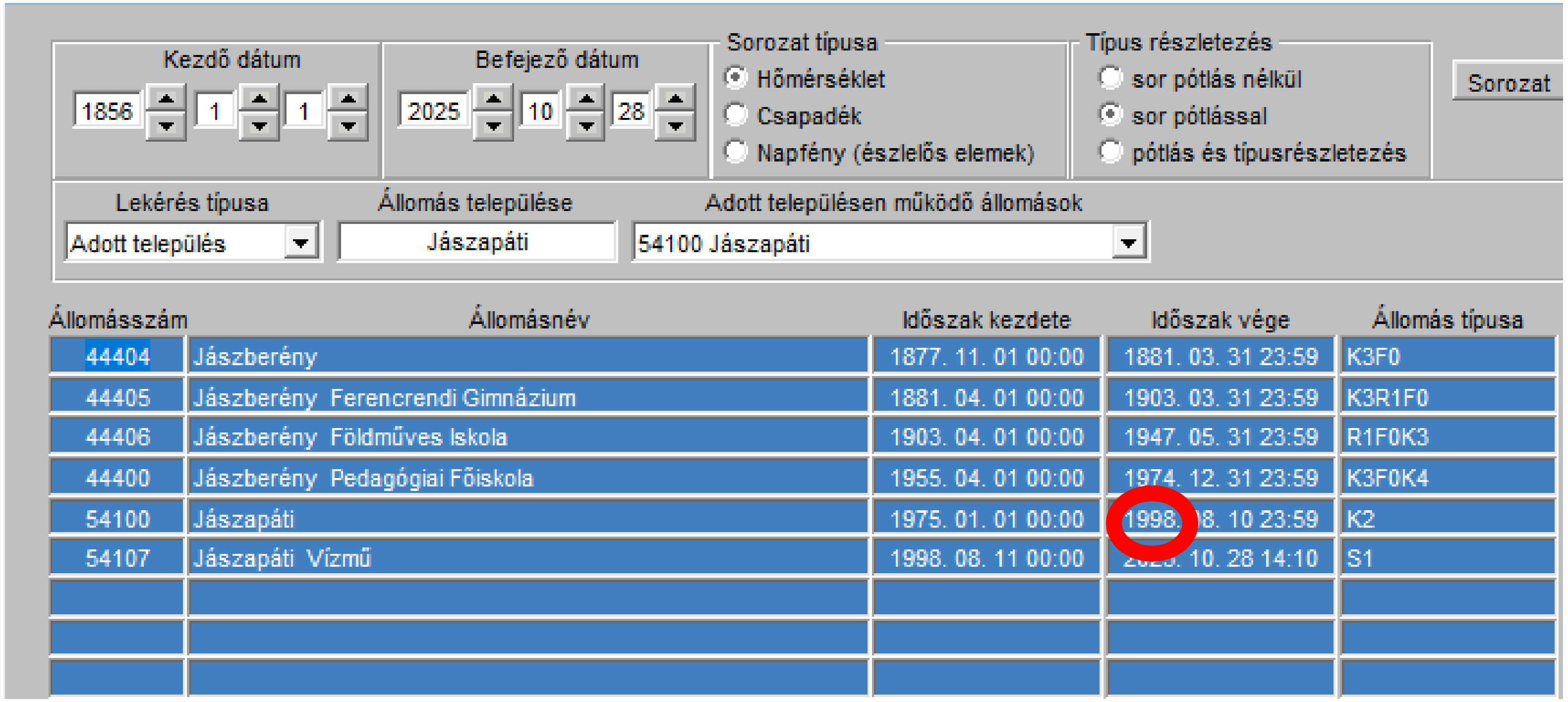
Állomásszám	Állomásnév	Időszak kezdete	Időszak vége	Állomás típusa
44101	Budapest Budai vár Csillagvizsgáló	1856. 01. 01 00:00	1860. 12. 31 23:59	K3
44116	Budapest Víziváros Reáliskola (később:Toldy Ferenc Gimnázium	1861. 03. 14 00:00	1870. 12. 19 23:59	K3
44117	Budapest Várnegyed Hofhauser Casino	1870. 12. 20 00:00	1872. 09. 30 23:59	K3
44118	Budapest Krisztinaváros Novák villa - felső állomás	1872. 10. 01 00:00	1890. 12. 31 23:59	K3
44172	Budapest Vár alsó állomás	1891. 01. 01 00:00	1900. 01. 31 23:59	K3
44119	Budapest Víziváros	1900. 02. 01 00:00	1910. 02. 28 23:59	K3
44120	Budapest Országút KMI	1910. 03. 01 00:00	1985. 03. 31 23:59	K3K4
44120	Budapest Országút KMI	1910. 03. 01 00:00	1960. 12. 31 23:59	K3
44120	Budapest Országút KMI	1961. 01. 01 00:00	1985. 03. 31 23:59	K4
44121	Budapest belterület OMSZ Torony	1985. 04. 01 00:00	2025. 10. 20 13:21	K4K8S2S1
44121	Budapest belterület OMSZ Torony	1985. 04. 01 00:00	1992. 03. 31 23:59	K4
44121	Budapest belterület OMSZ Torony	1992. 04. 01 00:00	1993. 08. 31 23:59	K8
44121	Budapest belterület OMSZ Torony	1992. 04. 01 00:00	1993. 08. 31 23:59	S2
44121	Budapest belterület OMSZ Torony	1993. 09. 01 00:00	2013. 04. 07 19:59	K4
44121	Budapest belterület OMSZ Torony	1997. 11. 30 12:00	2025. 10. 20 13:21	S1

META data BUDAPEST

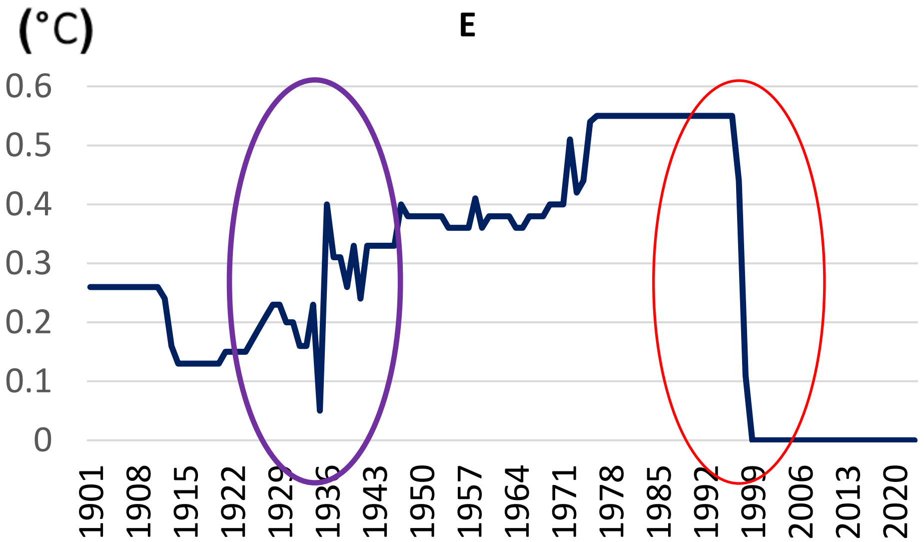

relocation
World War II
Széllkapu Park
BUDAPEST (Inhom. in mean)
[°C] 0.2
0
-0.2
-0.4
-0.6
-0.8
1901 1911 1921 1931 1941 1951 1961 1971 1981 1991 2001 2011 2021
change in methodology
BUDAPEST (inhom. in st. deviation)
1.1
1.05
1
0.95
0.9
1901 1911 1921 1931 1941 1951 1961 1971 1981 1991 2001 2011 2021
Monthly inhomogeneity in E 1971 (°C)
Budapest
January February March April May June July August September October November December
[°C]
E (annual)
15
13
11
9
7
1901 1910 1919 1928 1937 1946 1955 1964 1973 1982 1991 2000 2009 2018
raw
homog
D (annual)
2.5
2
1.5
1
1901 1911 1921 1931 1941 1951 1961 1971 1981 1991 2001 2011 2021
Year raw
Year homog

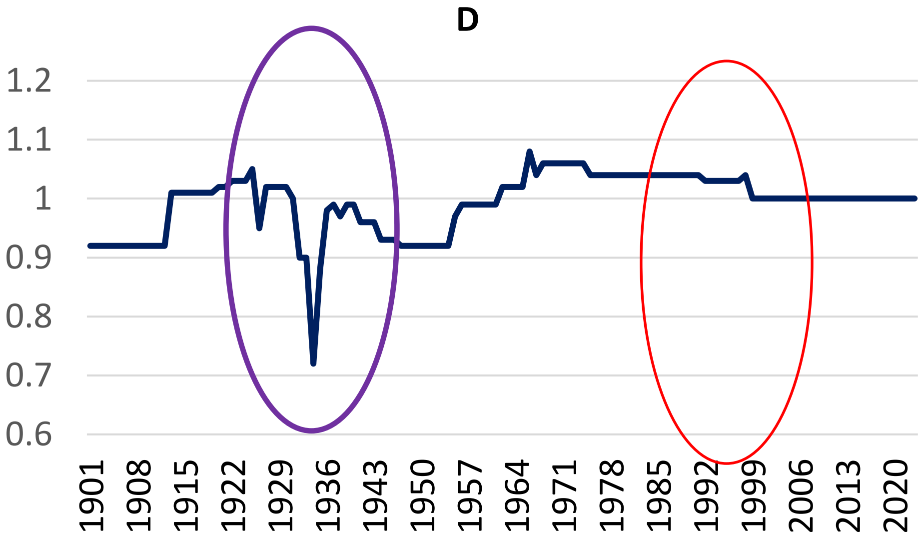
Jászapáti

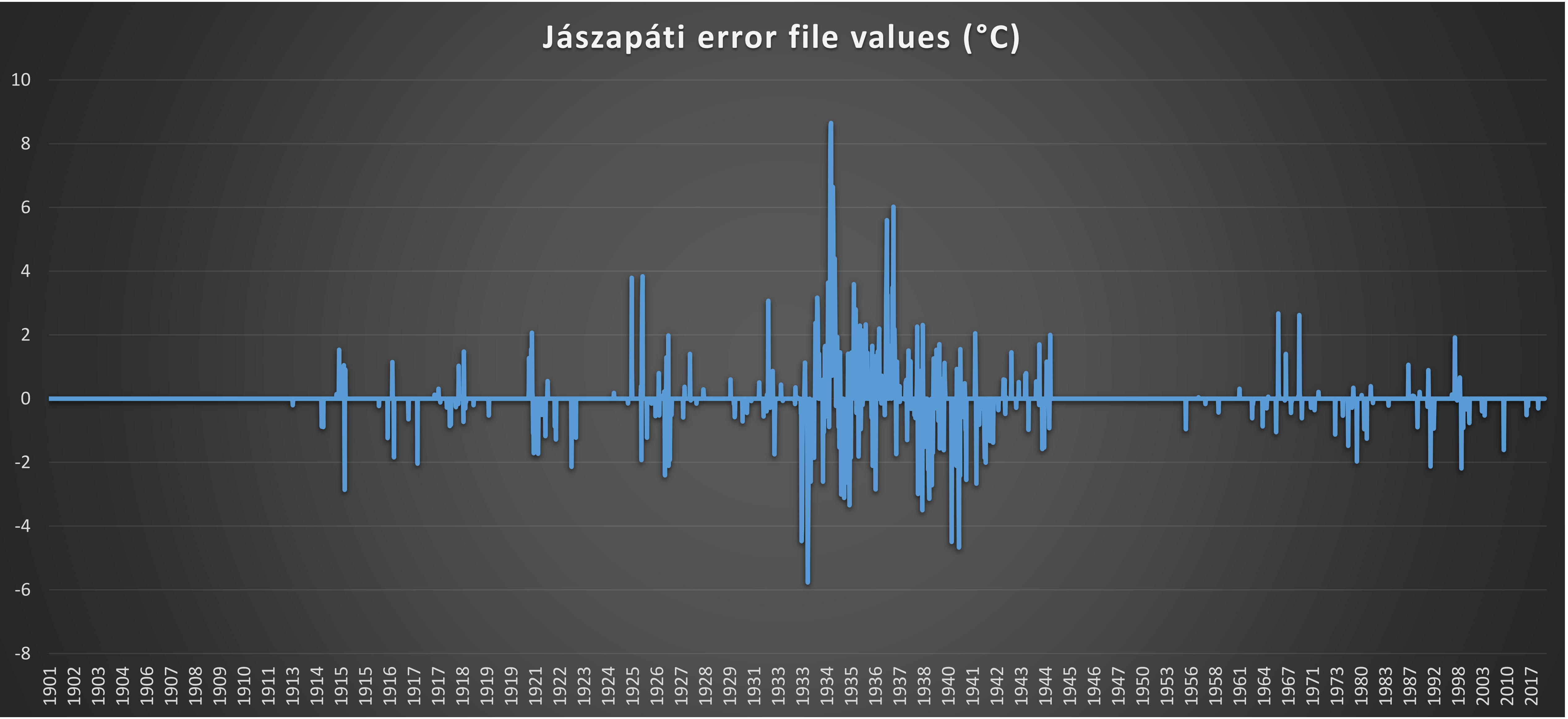




Állomásszám	Állomásnév	Időszak kezdete	Időszak vége	Állomás típusa
44404	Jászberény	1877. 11. 01 00:00	1881. 03. 31 23:59	K3F0
44405	Jászberény Ferencrendi Gimnázium	1881. 04. 01 00:00	1903. 03. 31 23:59	K3R1F0
44406	Jászberény Földműves Iskola	1903. 04. 01 00:00	1947. 05. 31 23:59	R1F0K3
44400	Jászberény Pedagógiai Főiskola	1955. 04. 01 00:00	1974. 12. 31 23:59	K3F0K4
54100	Jászapáti	1975. 01. 01 00:00	1998. 08. 10 23:59	K2
54107	Jászapáti Vízmű	1998. 08. 11 00:00	[illegible]. 10. 28 14:10	S1







Jászapáti error file values (°C)
10
8
6
4
2
0
-2
-4
-6
-8
1901 1902 1903 1904 1906 1907 1908 1909 1910 1911 1913 1914 1915 1915 1916 1917 1917 1918 1919 1919 1921 1922 1923 1924 1925 1926 1927 1928 1929 1931 1933 1933 1934 1935 1936 1937 1938 1940 1941 1942 1943 1944 1945 1946 1947 1950 1953 1956 1958 1961 1964 1967 1971 1973 1980 1983 1987 1992 1998 2003 2010 2017

Conclusion

- The inhomogeneities detected by the MASHv3.03 and MASHv4.01 software in the archive data can in many cases be explained by a change in methodology. Their magnitude and direction are consistent with the simulated data.
- In the case of temperature, the change in methodology can cause inhomogeneity of up to 2 degrees, while in the case of relative humidity, the largest detected and calculated inhomogeneity is approximately 16%.
- In the case of precipitation, the change in methodology did not cause much inhomogeneity in Hungary, given that precipitation was measured in the morning hours in all cases. (However, automatic precipitation gauges provide different precipitation data than the traditional method.)
- In the case of wind measurements, we cannot separate the methodological change from the other factors, because in all cases, relocation (including measurement height) and instrument replacement took place at the same time.

With MASHv4.01 software (in the case of an additive model), the inhomogeneity caused by the methodological change can also be detected in the second moment.

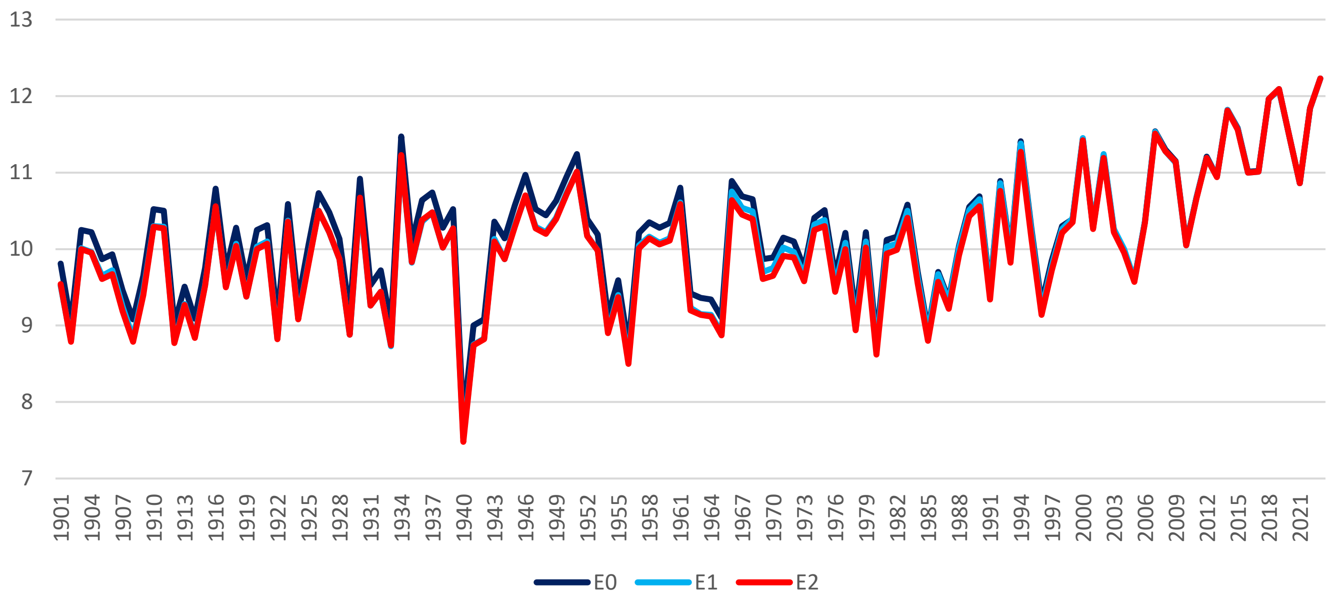
Thank you for your attention!

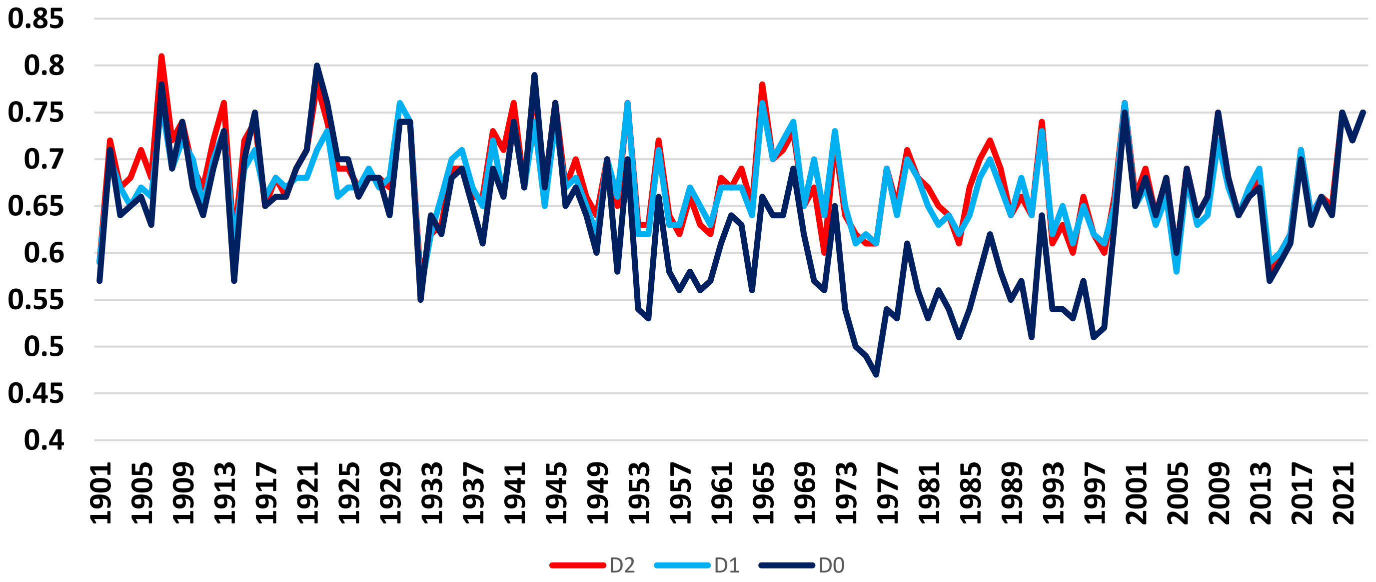
Save the date!

The **12th Seminar for Homogenization** and Quality Control in Climatological Databases and the 7th Interpolation Conference will be organized in **Budapest**, at the headquarters of the HungaroMet Hungarian Meteorological Service, and online, from **5 to 7 May 2026**.

Contact: seminar@met.hu

HungaroMet


13
12
11
10
9
8
7
1901
1904
1907
1910
1913
1916
1919
1922
1925
1928
1931
1934
1937
1940
1943
1946
1949
1952
1955
1958
1961
1964
1967
1970
1973
1976
1979
1982
1985
1988
1991
1994
1997
2000
2003
2006
2009
2012
2015
2018
2021
E0
E1
E2


0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
1901
1905
1909
1913
1917
1921
1925
1929
1933
1937
1941
1945
1949
1953
1957
1961
1965
1969
1973
1977
1981
1985
1989
1993
1997
2001
2005
2009
2013
2017
2021
D2
D1
D0