

ARCHÍV NAPI ADATOK ELLENŐRZÉSE MASH ELJÁRÁSSAL: ELMÉLET ÉS A GYAKORLAT TALÁLKOZÁSA

Izsák Beatrix¹, Szentimrey Tamás²

¹ OMSZ ÉEFO Éghajlati Osztály

² Varimax Bt.

MASHv3.03

Multiple Analysis of Series for Homogenization;

Szentimrey, T.

- **Havi adatsorok homogenizálása, ellenőrzése, pótlása**

- Relatív homogenitás vizsgálati elv: különböző állomások azonos időszakra vonatkozó adatsorának összehasonlítása, az ellentmondások vizsgálata
- Cél a felfedezett ellentmondások megszüntetése, az inhomogenitás kiszűrése
- Homogenizált éves, évszakos adatsor előállítása kizárólag a homogenizált havi adatokból
- Alkalmas az additív (pl. hőmérséklet) és a kumulatív jellegű (pl. csapadékösszeg) éghajlati elemek adatsorainak homogenizálására is
- A metaadatokat automatikusan használja
- A homogenizálás eredménye kiértékelhető, verifikálható → az eredeti adatsorok és az eredmény adatsorok inhomogenitását és a változás mértékét kvantitatív módon jellemezzük.

- **Napi adatsorok homogenizálása, ellenőrzése, pótlása**

- A becsült havi inhomogenitások felhasználásával
- **Automatikus adatellenőrző és adatpótló eljárás**

- Bizonyos meteorológiai és matematikai megfontolások alapján a havi statisztikai paramétereket használhatjuk a napi adatokra.
- Adathiány esetén a pótlás:

$$\hat{X}_j(t) = w_{j0}^{\text{opt}} + \sum_{i \neq j} w_{ji}^{\text{opt}} X_i(t)$$

- **Adatellenőrzés a standard hiba sorokon alapul**

$$Z_j(t) = \frac{X_j(t) - \hat{X}_j(t)}{D_j(t)(1 - \text{REP}_j^{\text{opt}})} \in N(0,1)$$

Ahol w_{j0}^{opt} , w_{ji}^{opt} , $\text{REP}_j^{\text{opt}}$ az optimális interpolációs paraméterek havi adatok esetén és $D_j(t)$ a napi szórás.

Hipotézis vizsgálat a standardizált hiba sorokon

$$\underline{(Z(t) \ t=1,\dots,n)}$$

- Ha az adat jó minőségű:

$$Z(t) \in N(0,1) \ t=1,\dots,n$$

- Tegyük fel, hogy $Z(t)$ normáleloszlású AR(1) folyamat pozitív autokorrelációval, ez esetben a kritikus értéket az alábbi összefüggés segítségével tudjuk meghatározni:

$$P \left(\max_t |Z(t)| < z \right) \geq \prod_{t=1}^n P \left(|Z(t)| < z \right) \quad t=1,\dots,n$$

Döntés hipotézisvizsgálattal:

Az adat rossz (gyanús): ha $|Z(t)| > z_p$

ahol z_p a kritikus érték az adott szignifikancia szinten (pl. $p=0.01$), melyet az alábbi módon számítunk:

$$P(\max_t |Z(t)| < z_p) \geq (2\Phi(z_p) - 1)^n = 1 - p,$$

$\Phi(z)$: standard normál eloszlásfüggvény.

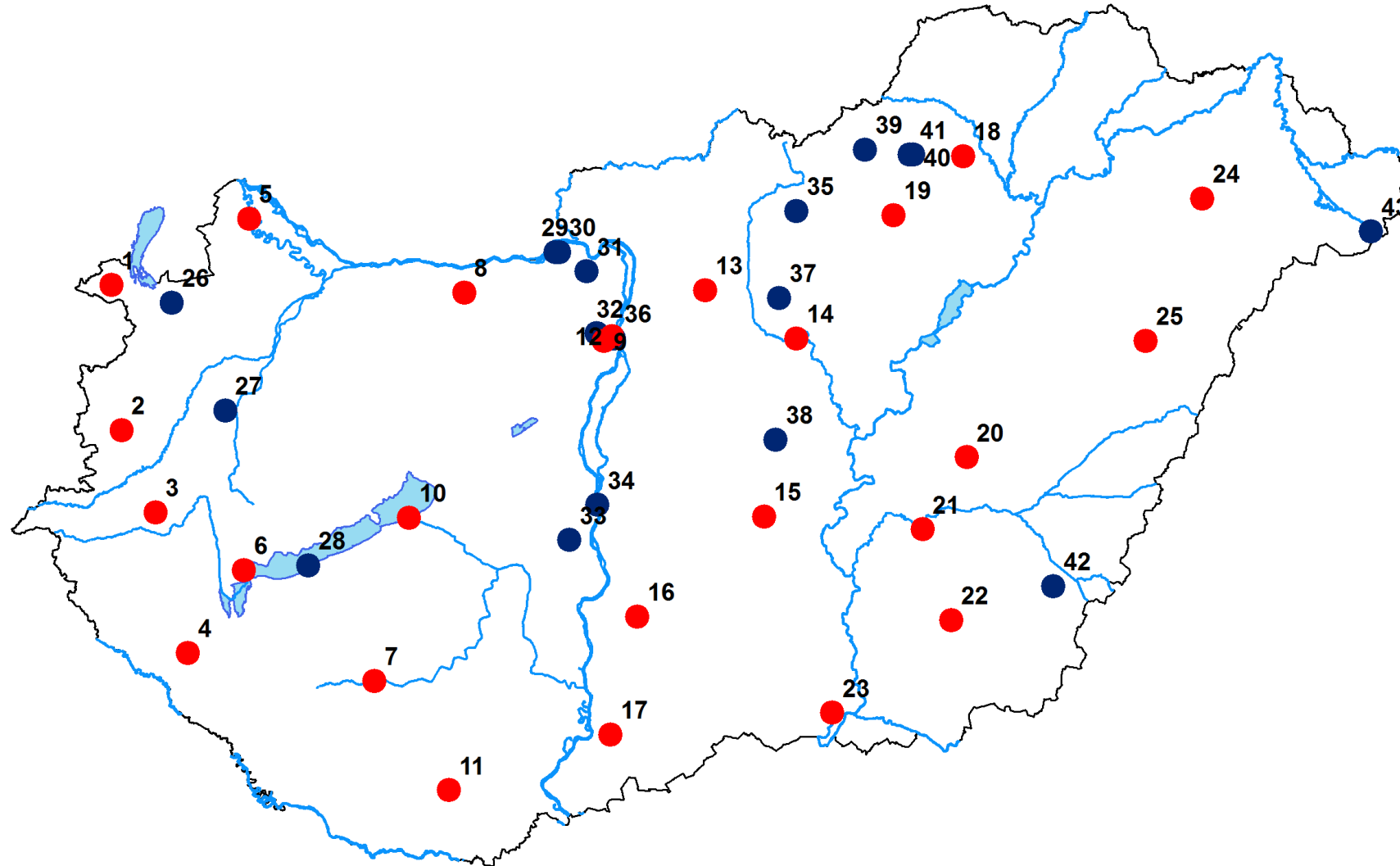
Multiple Quality Control (MASH)

Több standardizált hibasort készítünk az egyes állomásokhoz, melyekben nincs közös referencia állomás, hogy a vizsgált állomáshoz tartozó hibákat egyértelműen szeparálni tudjuk. A hibás adatok korrekciójához konfidencia intervallumot adunk meg.

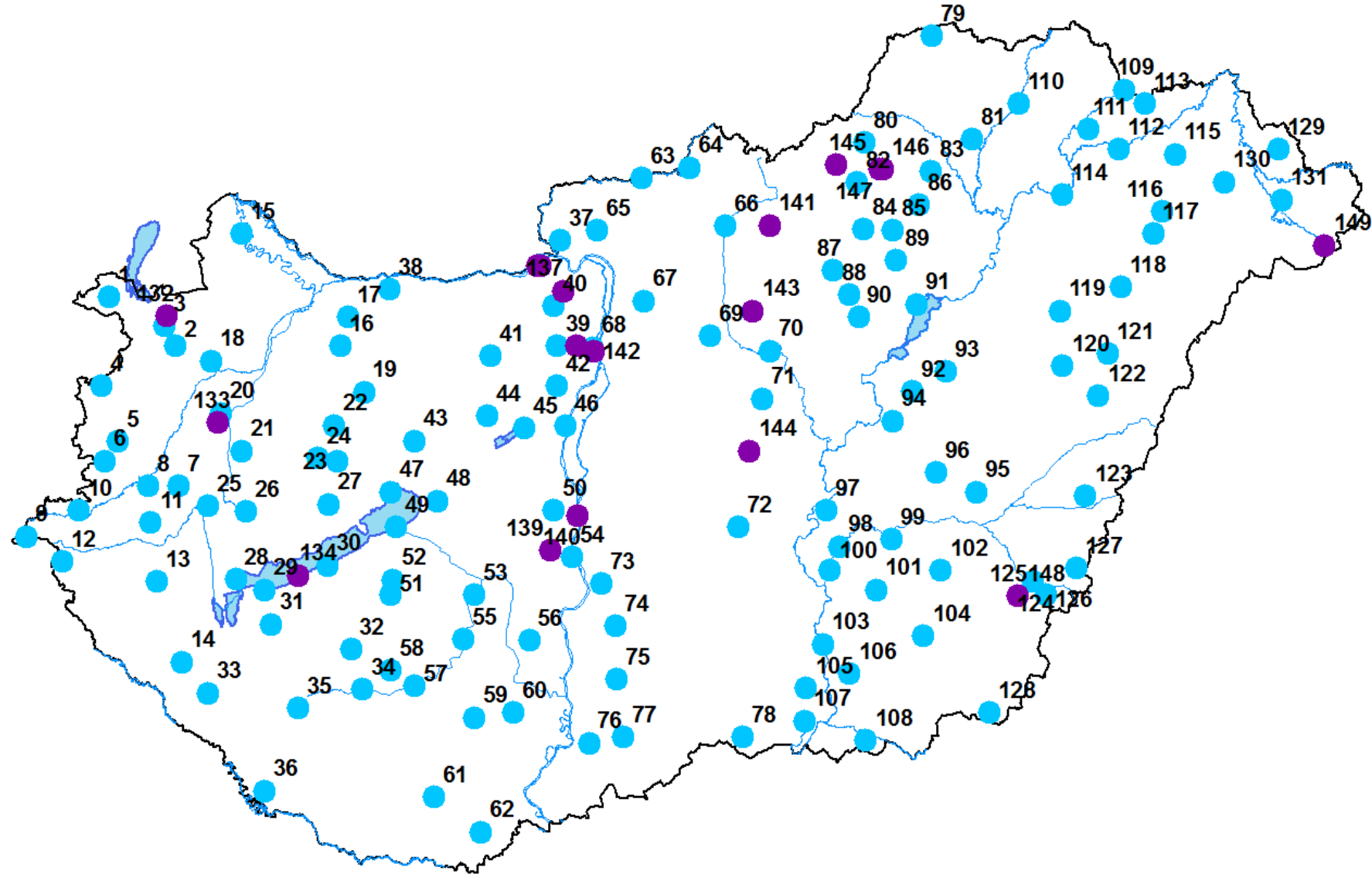
Javasolt eljárás az archív napi adatok ellenőrzésére

- **1. a gyanús adatokat automatikus eljárással keressük meg:**
 - nullhipotézis (H_0): az adat jó minőségű
 - szignifikancia szint, kritikus érték, automatikus eljárás
 - döntés az automatikus eljárás során:
 - ha H_0 elfogadható, az adat jó minőségű,
 - ha H_0 nem elfogadható, az adat gyanús.
- **2. az automatikus eljárás után:**
 - A gyanús adatokat ellenőrizni (évkönyv, makroszinoptikus helyzet stb.)
 - **A hibás és az extrém értékek között statisztikailag nem lehet (más egyéb információ nélkül) különbséget tenni!**

Adatok : Tmin, Tmax 25 régi és 18 új állomás, 1902-2018-ig

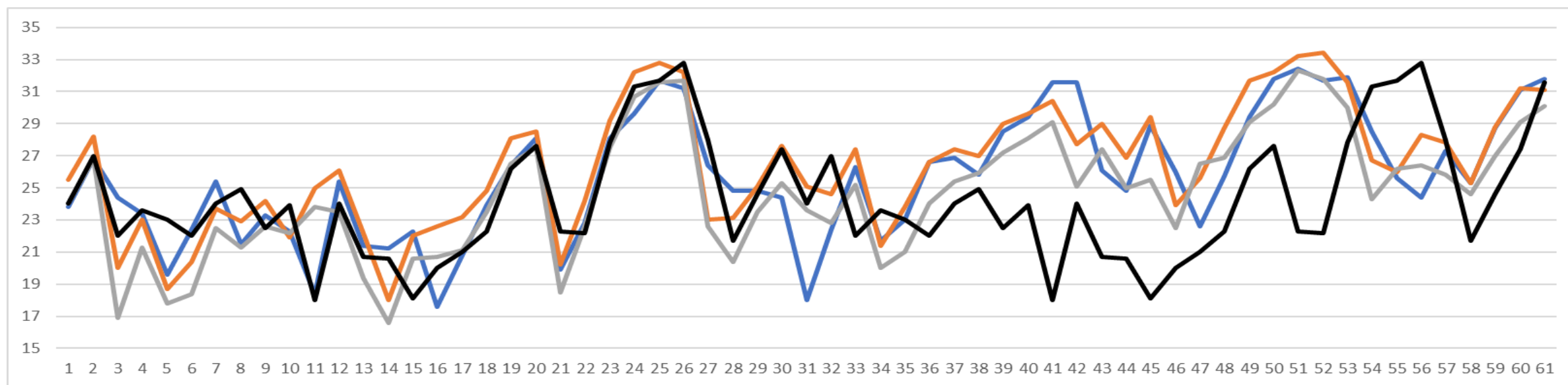
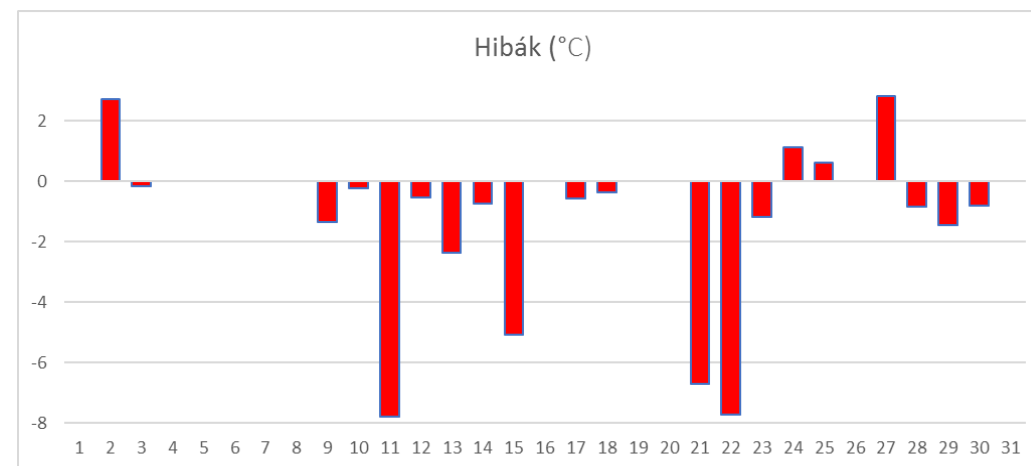


Adatok : R 131 régi és 20 új állomás, 1901-2018-ig



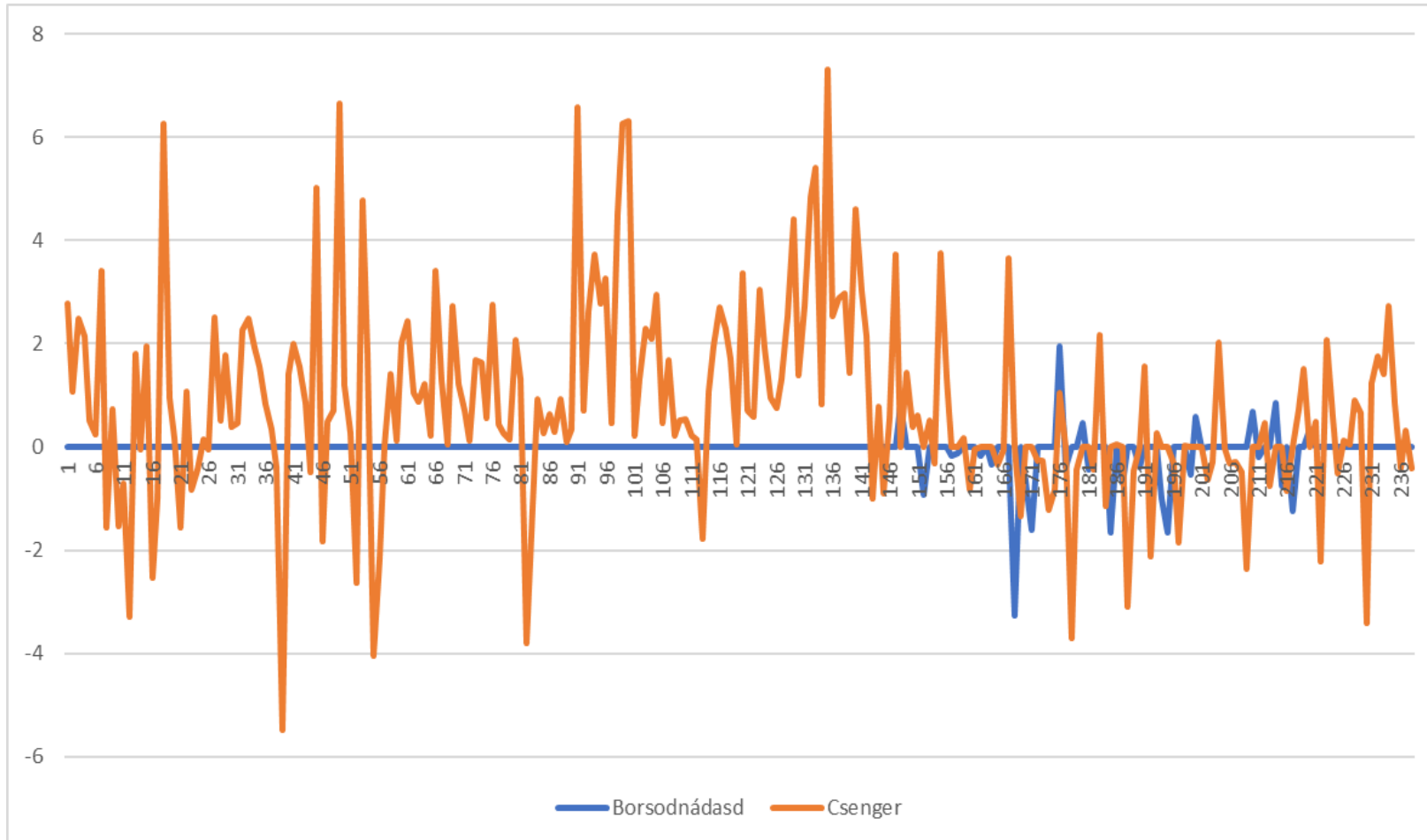
Eredmények - érdekesebb esetek

- Réaumur skála (Mash manual)
- Két egymás utáni hónap értékei megegyeznek (1934 június és július)



Napi maximum hőmérséklet:

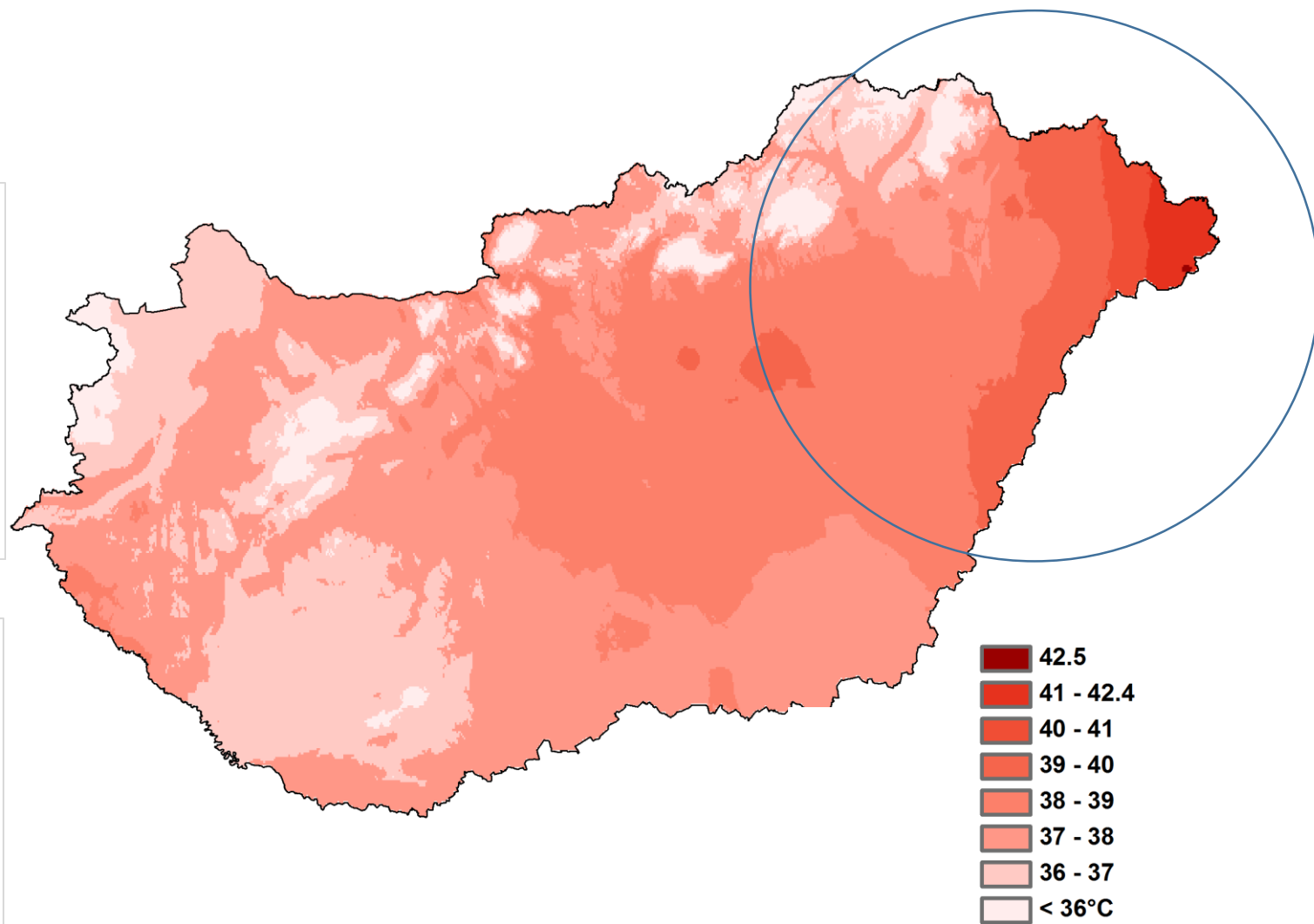
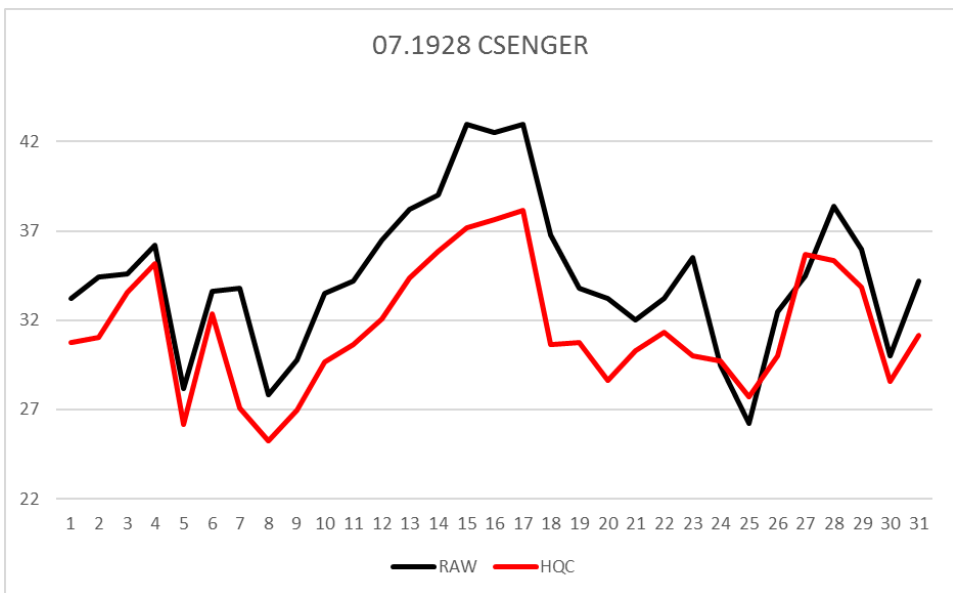
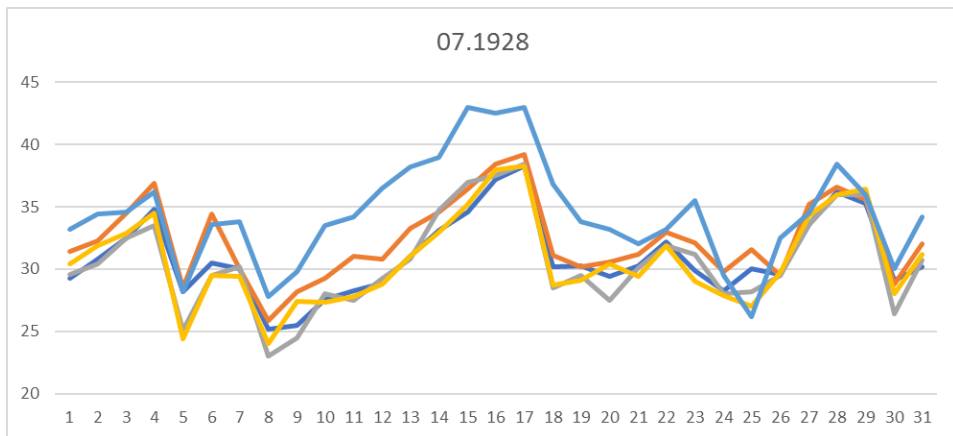
Mi legyen azzal a sorral, ahol minden második adat gyanús? CSENGER
Ráadásul 3 esetben abszolút maximum rekord?



Megegyeztünk abban, hogy a rekordokról egy bizottság dönt (adatellenőrök, méréssel foglalkozó szakemberek, klimatológusok és szinoptikusok), a MASH (és így a matematikai statisztika) eredményét figyelembe véve.

Adathibák 2 ellenőrzött állomáson, az értékek °C-ban megadva

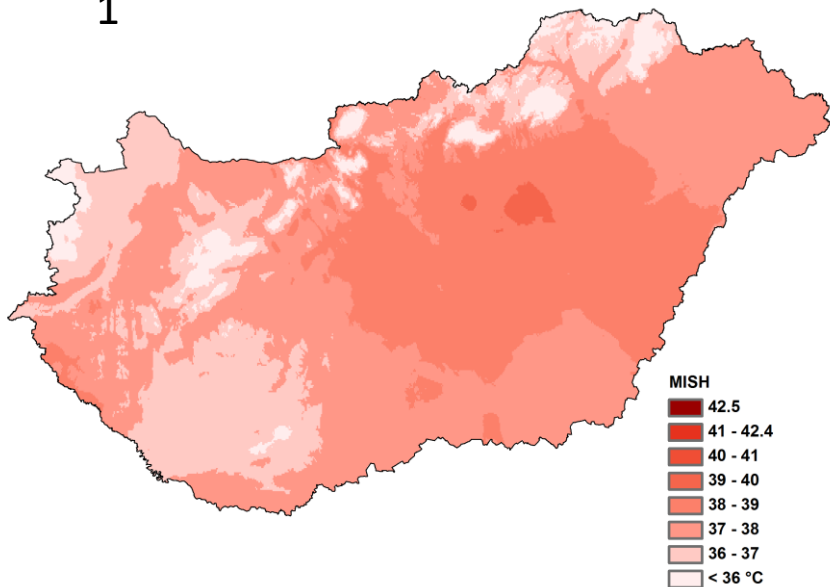
Mérés



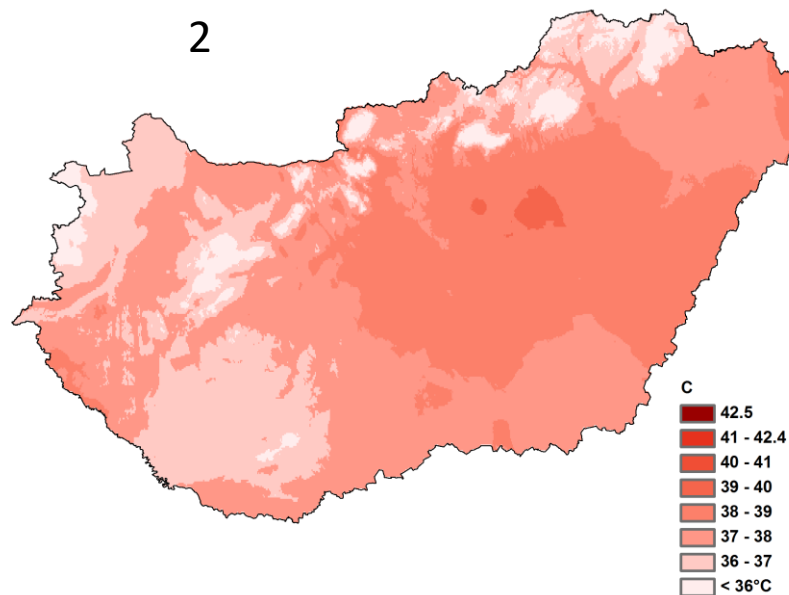
Hibák Csenger állomáson

1928	710	1.08
1928	711	1.95
1928	712	2.71
1928	713	2.29
1928	715	1.7
1928	717	0.05
1928	718	3.37

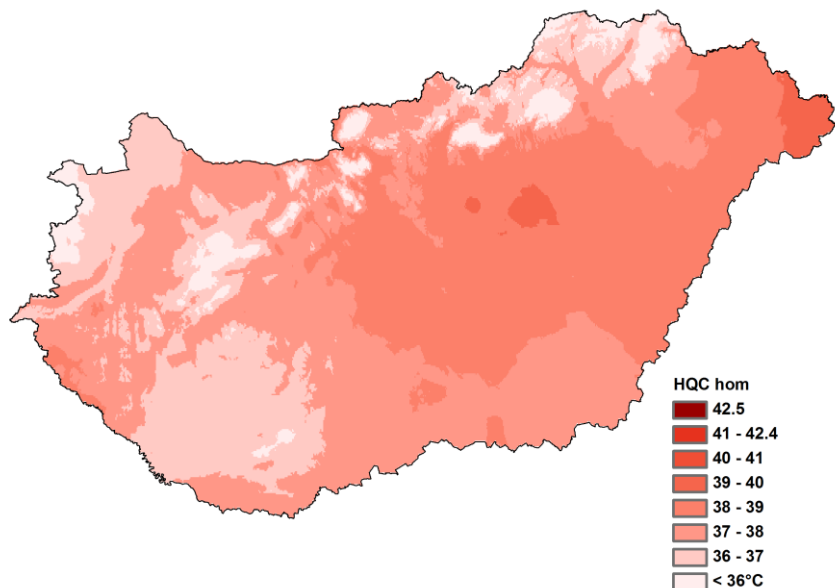
1



2



3



1928.07.16 CSENGER

Mérés: **42.5°C**

Új abszolút maximum rekord?($41,9^{\circ}\text{C}$ a régi rekord)

1. MISH becslés (kihagyva Csenger) **37.6°C**

2. MASH becslés (kihagyva ez a nap, homogenizálás nélkül) **38.1°C**

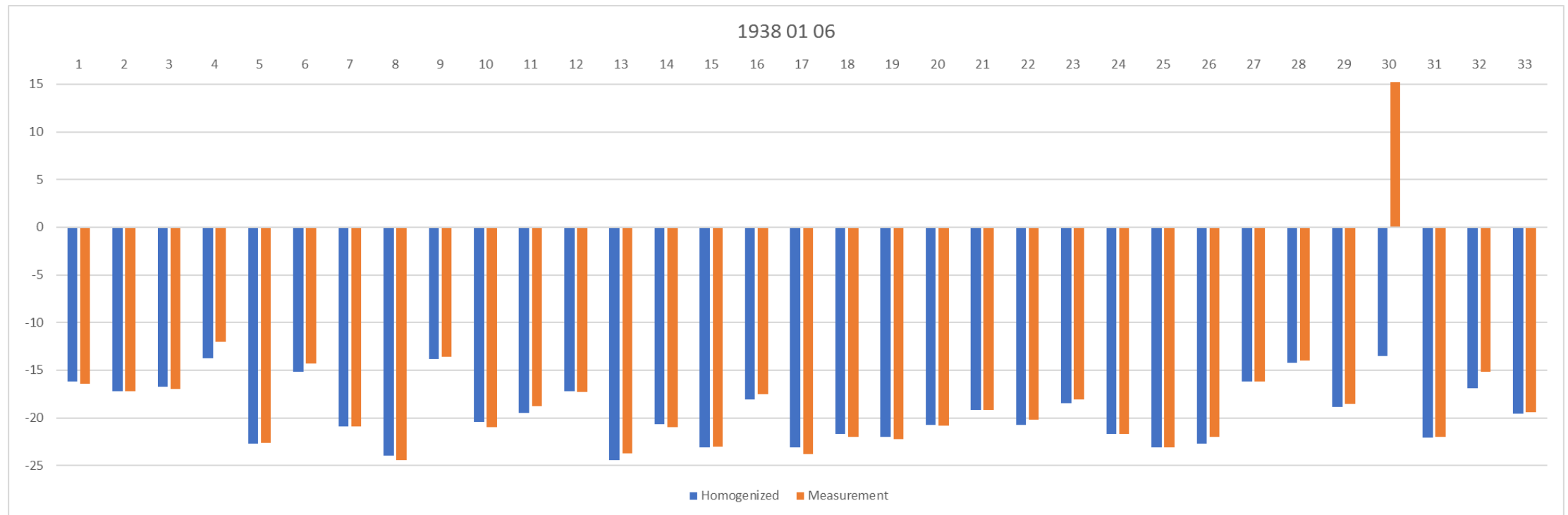
3. MASH homogenizált érték **39.6°C**

havi inhomogenitás: **3.3°C**

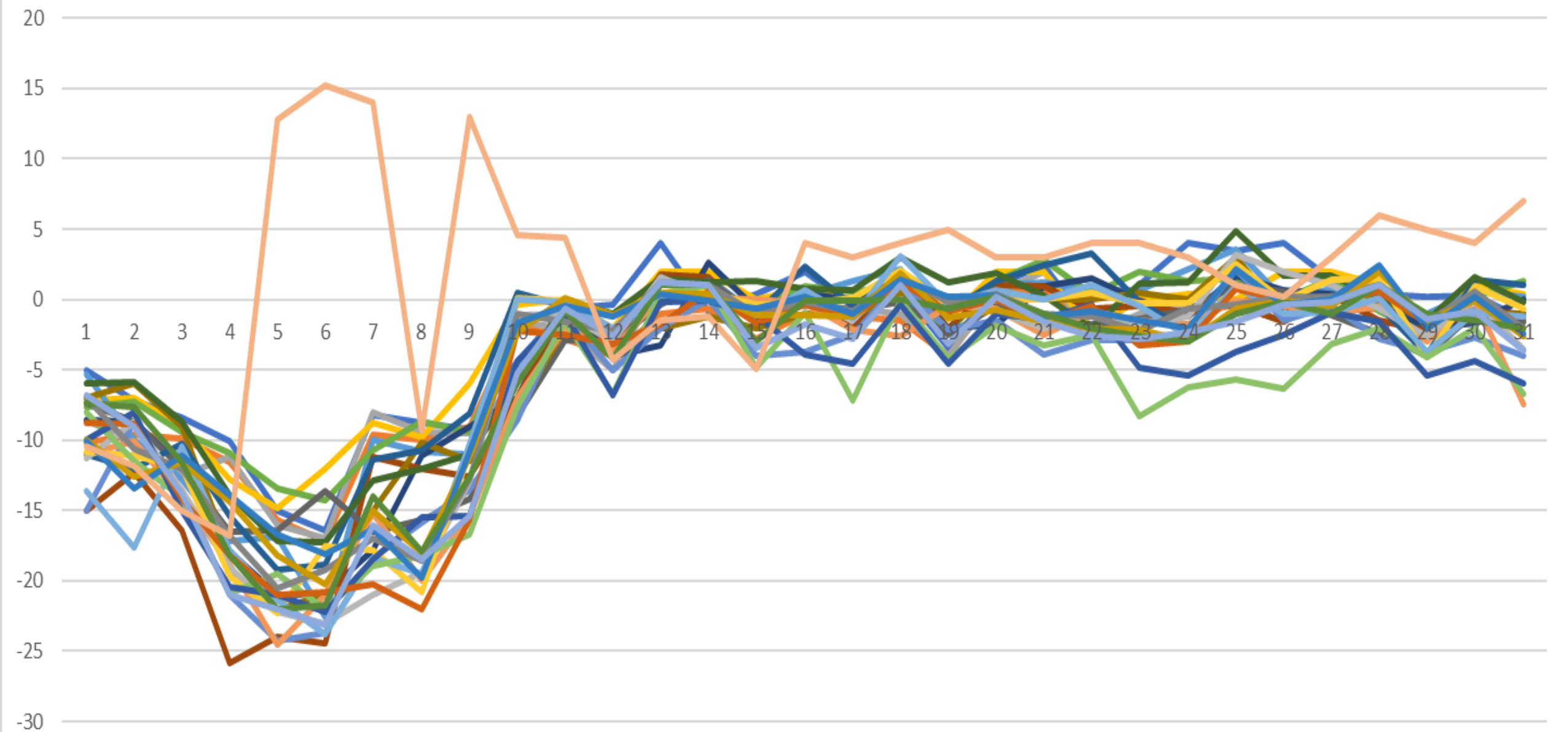
AZ INHOMOGENITÁS NEM HIBA!

Napi minimum hőmérséklet

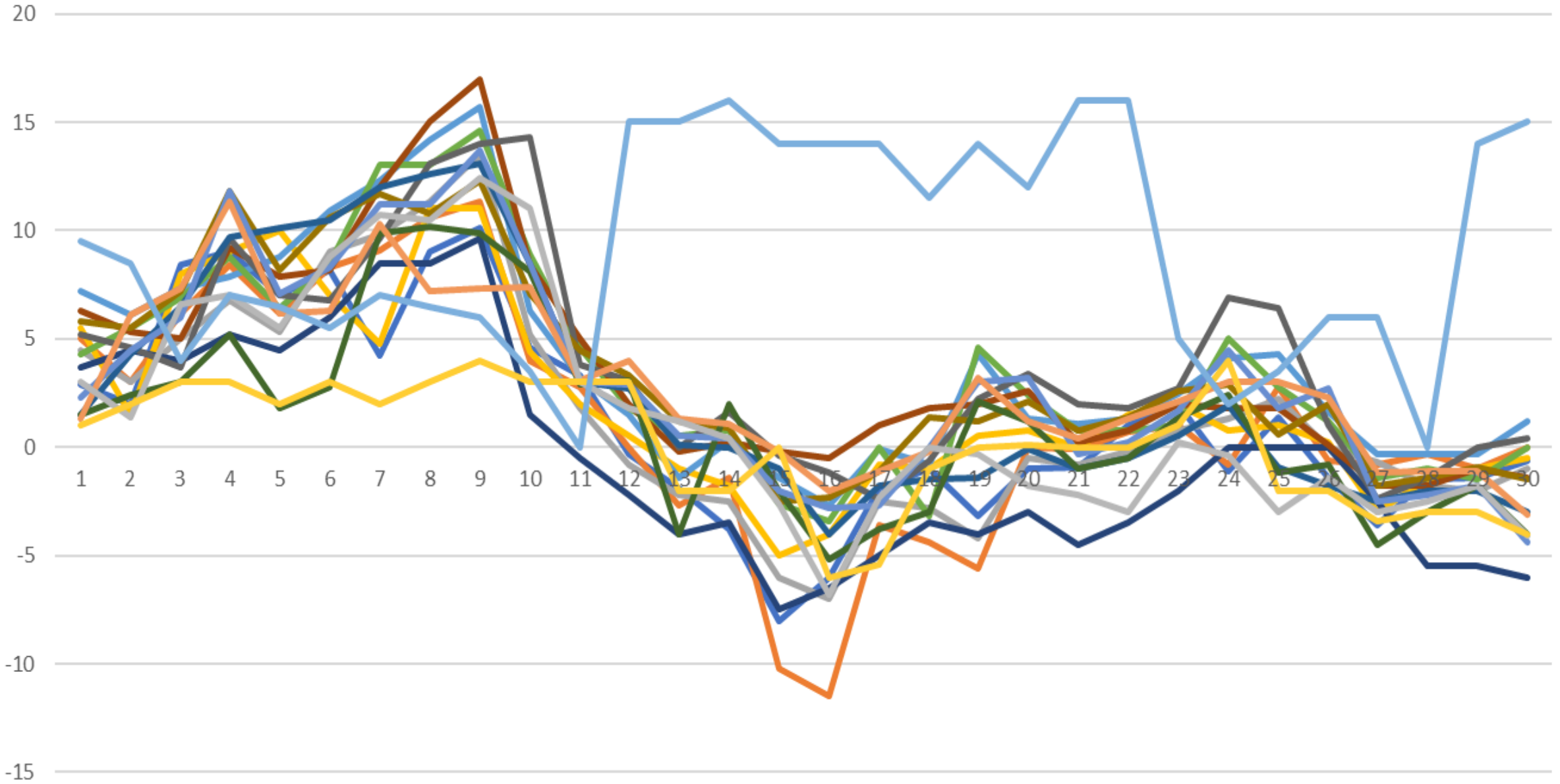
Galyatető, úgy néz ki mintha valaki előjel nélkül rögzített volna..



1938 January Tn

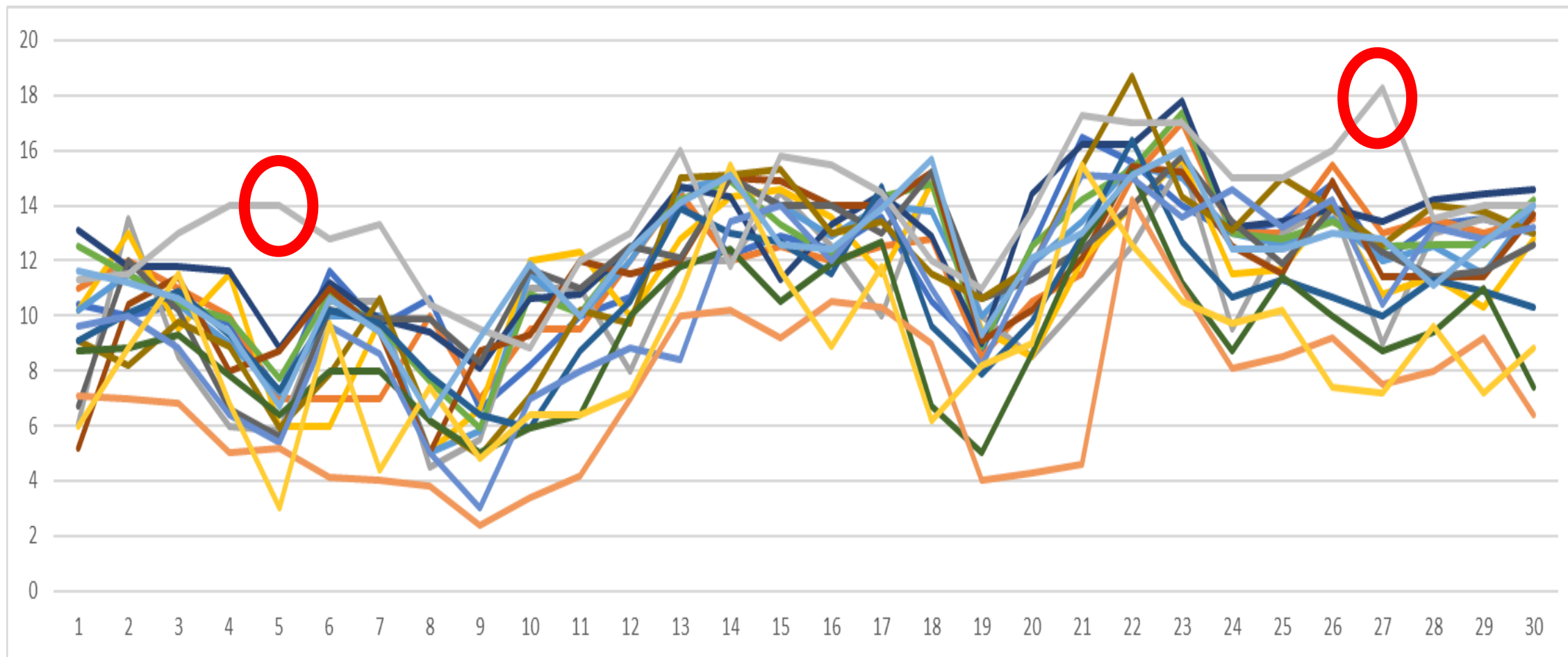


1927 November Tn



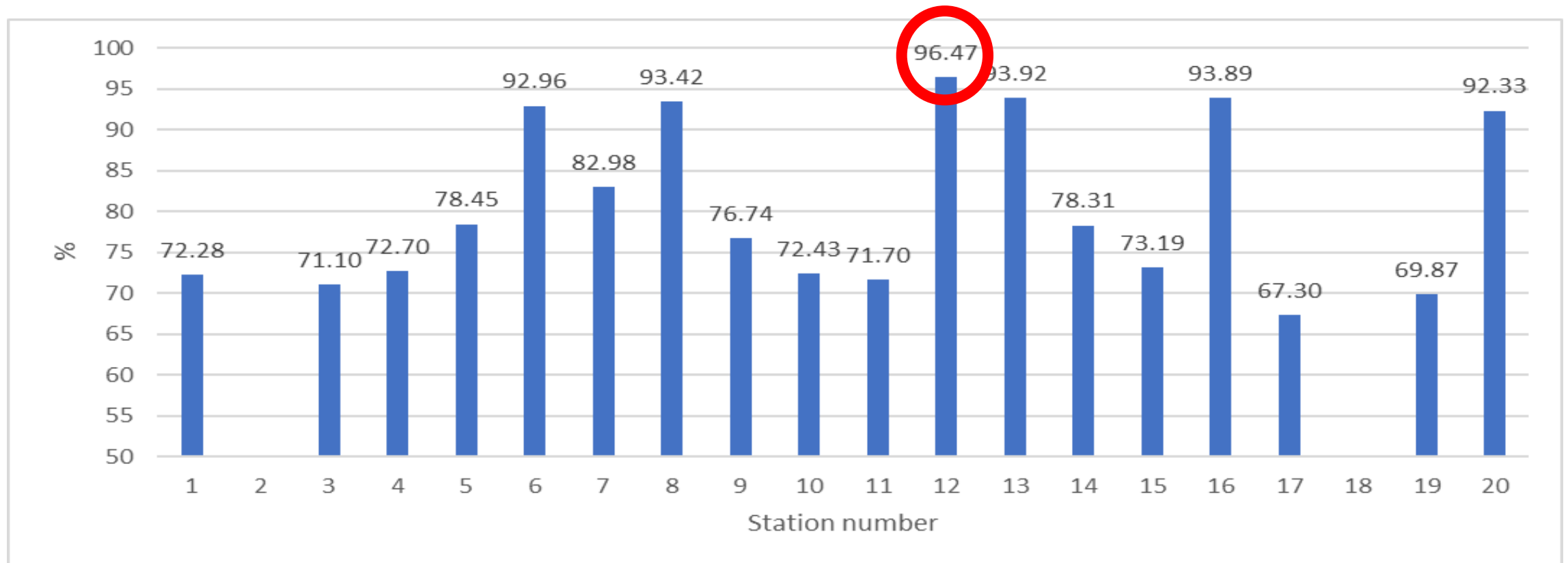
Minor error: 1933.06.05 and 06.27

1933 6 5	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.11	0.00	0.00	0.00
1933 6 27	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.44	0.00	0.00	0.00



Csapadék esetén Bp. Németvölgy 96%-ban 0 értékek (Magyarországon kb. 60-70% az arány)

- 1923.02.09- to 1926.01.31-ig nincs csapadék!!! Sivatatag Budapesten !!!



- Hőmérséklet (Tx)

41*117*365≈1.75 millió adat

Error (41 állomáson):4540, 723 >2°C

Error 16 új állomáson: 1095, 253 >2°C

(Tmin esetén is hasonló az arány)

- Csapadékösszeg (R)

151*118*365 ≈6.5 millió adat

Error (151 állomáson): 2476, 657 >10mm

Error 16 új állomáson: 1095, 42 >10mm

Konklúzió

- Adatellenőrzés MASH eljárással
- A gyanús adatokat manuálisan ellenőrizni (<1%)
- Abszolút vagy napi rekord: bizottság dönt
- A nyers adatokat mindig el kell menteni, csak akkor törölünk, ha bizonyítottan rögzítési vagy mérési hiba történt
- Éghajlati Osztályon minden évben elkészül a homogenizált, pótolta és **ellenőrzött** adatbázis: T_a , T_{min} , T_{max} , r , s_r , f_{max} , f_a , u , p elemekre is.
- Ezekből az ellenőrzött napi adatokból készíti az Éghajlati Osztály a tanulmányait, cikkeit, trendbecsléseit és ezekből szolgáltatunk.

Köszönöm a figyelmet!