

# HUCLIM ADATBÁZIS LÉTREHOZÁSÁNAK MÓDSZERTANA ÉS A LEGÚJABB FEJLESZTÉSEK

Izsák Beatrix, Szentes Olivér,  
Bihari Zita, Lakatos Mónika  
OMSZ Éghajlati Osztály

KLIMADAT PROJEKTZÁRÓ  
Budapest, 2022. március 30.

SZÉCHENYI 2020



MAGYARORSZÁG  
KORMÁNYA

Európai Unió  
Kohéziós Alap



BEFEKTETÉS A JÖVŐBE

# KLIMADAT PROJEKT –ÉGHAJLATI OSZTÁLY

- A rácsponti mérési adatbázis folyamatos frissítése és időbeli felbontásának növelése: az OMSZ az elmúlt években a [CARPATCLIM](#) és a NATÉR programok keretében előállított egy részletes rácsponti adatbázist, mely az 1961–2010 időszakot öleli fel, 10 km-es felbontásban tartalmaz napi adatsorokat a legfontosabb meteorológiai elemekre. Ezt az adatbázist fejlesztenünk kellett: egyrészt folyamatosan frissítjük a 2010 utáni adatokkal, másrészt a napi adatsorokat 6 óránkénti felbontású információkkal egészítjük ki.
- Módszerfejlesztés: az éghajlati adatokon alapuló indikátorok előállítására, rácsponti adatsorok előállítása

# ELŐÁLLÍTOTT ADATSOROK

- **Rácsponti adatbázis:** napi (Tmin, Tmax, Ta, r, globalsugárzás) és 6 óránkénti értékekre (Ta, r)
- **Boxátlagok:** Tmin, Tmax, Ta, r, wind10
- **Homogenizált állomási adatsorok:** Tmin, Tmax, Ta, r
- **Térképekhez:** 22 klímaindex (év, évszak, havi=366fájl)
- **Grafikonokhoz:** 23790 fájl (országos, 19 megye+Bp, 174 járás)

# MÚLTRA VONATKOZÓ ÉGHAJLATI INDEXEK A KLIMADAT TÉRINFORMATIKAI RENDSZERBEN

Változó	Magyarázat	Mértékegység
Napi átlaghőmérséklet	A hőmérséklet napi átlaga (0 és 24 óra közötti)	°C
Napi maximumhőmérséklet	A hőmérséklet napi maximuma (előző nap 18 UTC és aznap 18 UTC közötti)	°C
Napi minimumhőmérséklet	A hőmérséklet napi minimuma (előző nap 18 UTC és aznap 18 UTC közötti)	°C
Napi csapadékösszeg	A csapadék napi összege (előző nap 6 UTC és aznap 6 UTC közötti, ami az előző napra vonatkozik)	mm/hónap
Nyári nap	A napi maximumhőmérséklet meghaladja a 25 °C-ot	nap
Hőségnap	A napi maximumhőmérséklet eléri a 30 °C-ot	nap
Túl meleg éjszaka	A napi minimumhőmérséklet meghaladja a 20 °C-ot	nap
Másodfokú hőhullámos nap	A napi átlaghőmérséklet legalább 3 egymást követő napon keresztül eléri a 25 °C-ot	nap
Fagyos nap	A napi minimumhőmérséklet 0 °C alatt marad	nap
Zord nap	A napi minimumhőmérséklet -10 °C alatt marad	nap
Téli nap	A napi maximumhőmérséklet 0 °C alatt marad	nap
Egymást követő fagyos napok maximális száma	Az a leghosszabb időszak, amikor a napi minimumhőmérséklet 0 °C alatt marad	nap
Vegetációs időszak	Az az időszak, amikor az átlaghőmérséklet legalább 6 egymást követő napon meghaladja az 5 °C-ot, majd (július után) 5 °C alá süllyed	nap
Csapadékos nap	A napi csapadékösszeg eléri az 1 mm-t	nap
Egymást követő száraz napok maximális száma	Az a leghosszabb időszak, amikor a napi csapadékösszeg nem éri el az 1 mm-t	nap
10 mm-t meghaladó csapadékú nap	A napi csapadékösszeg eléri a 10 mm-t	nap
20 mm-t meghaladó csapadékú nap	A napi csapadékösszeg eléri a 20 mm-t	nap
Maximális napi csapadékösszeg	Az 1 nap alatt lehullott legnagyobb csapadékmennyiség	mm
Csapadékintenzitás	A napi csapadékmennyiségek összegének és a csapadékos napok számának hányadosa	mm/nap
Maximális napi csapadék 20-éves visszatérésű értéke	Az a maximális napi csapadékmennyiség, ami átlagosan 20-éves gyakorisággal fordul elő	mm/nap
Maximális napi csapadék 50-éves visszatérésű értéke	Az a maximális napi csapadékmennyiség, ami átlagosan 50-éves gyakorisággal fordul elő	mm/nap
Maximális napi csapadék 100-éves visszatérésű értéke	Az a maximális napi csapadékmennyiség, ami átlagosan 100-éves gyakorisággal fordul elő	mm/nap

# ADATSZERVEZÉS

## Mi a probléma az adatokkal?

A minőség szempontjából: adathiányok, mérési hibák, inhomogenitások (a mérőhálózat változásából következően)

A térbeli reprezentativitás szempontjából: pontonkénti mérések, továbbá ezek és rácspontokra adott háttérinformációk (pl. radar, műhold, előrejelzési adatok) együttes kezelése.



**STATISZTIKUS KLIMATOLÓGIA**

# REPREZENTATÍV ÉGHAJLATI ADATBÁZIS

## MATEMATIKAI SZOFTVEREINK

[http://www.met.hu/en/omsz/rendezvenyek/homogenization\\_and\\_interpolation/software/](http://www.met.hu/en/omsz/rendezvenyek/homogenization_and_interpolation/software/)

### **MASHv3.03**

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

**Állomás adatsorok homogenizálása, ellenőrzése és pótlása**

### **MISHv1.03**

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

**Éghajlati statisztikai paraméterek modellezése, meteorológiai adatok interpolációja és pótlása, *gridding a rácsrendszerre***

**Irodalom:**  
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.

# A MASH RENDSZER FŐBB TULAJDONSÁGAI

## **Havi adatsorok pótlása, ellenőrzése, homogenizálása**

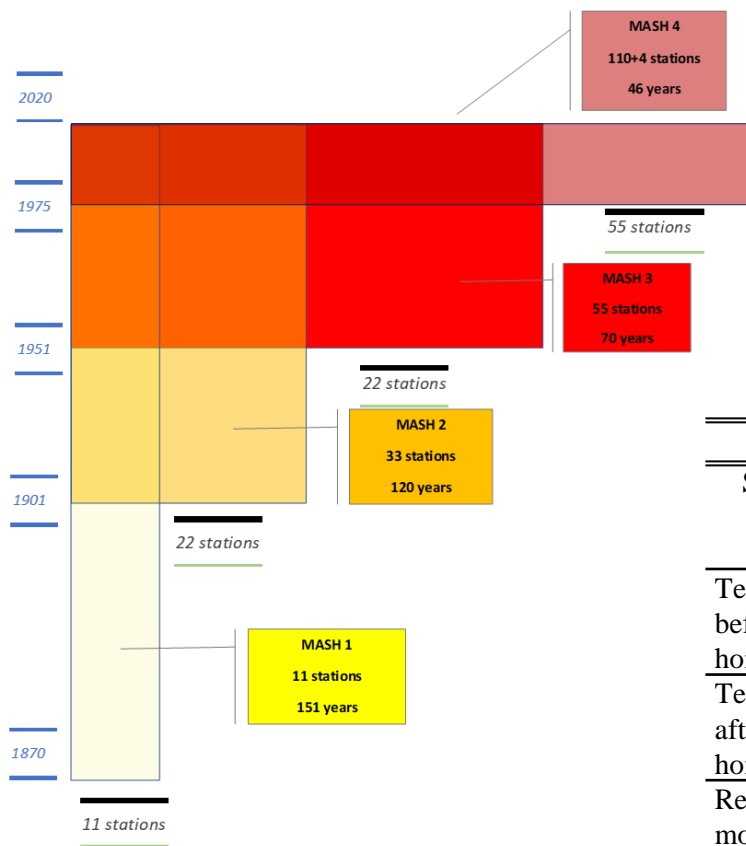
- Relatív homogenitásvizsgálati elv alapján működik.
- Step by step eljárás: a sorok szerepe (jelölt, referencia) lépésről lépésre változik az automatizált statisztikai döntési eljárás során.
- Additív (pl. hőmérséklet) vagy multiplikatív (pl. csapadék) modell alkalmazható, a meteorológiai elem eloszlásától függően.
- Az évszakos, éves sorok homogenitásának biztosítása
- A meta adatok automatikus felhasználása és kiértékelése.
- A homogenizálás eredménye kiértékelhető, verifikálható. (Ez ad alapot a térbeli és időbeli frissítésre!)

## **Napi adatsorok homogenizálása**

- A detektált havi inhomogenitások felhasználásával.
- Automatikus adatpótlás és adatellenőrzés.

Órás értékek? Fejlesztés szükséges!

# HOMOGENIZÁLÁS: NAPI ADATOK



Állomás rendszer a napi középhőmérséklet adatsorokhoz

MASH	MASH1	MASH2	MASH3	MASH4
Significance level: 0.05	Critical value: 22.05	Critical value: 21.76	Critical value: 21.31	Critical value: 20.86
Test statistics before homogenization	1253.08	877.95	357.2	257.49
Test statistics after homogenization	44.89	29.1	24.83	21.81
Relative modification of series	0.43	0.48	0.39	0.38
Representativity of station network	0.83	0.87	0.89	0.89

Izsák, B., Szentimrey, T., Lakatos, M., Pongrácz, R., Szentes, O.: Creation of a representative climatological database for Hungary from 1870 to 2020, *Időjárás* 126, 1-26.  
doi:10.28974/idojaras.2022.1.1

A legfontosabb teszt statisztikák.



# HOMOGENIZÁLÁS: ÓRAÉRTÉKEK

- **Hőmérséklet**

- az órás adatsorok igen inhomogének,

- másképpen inhomogenitásaik nem azonosak a napi sorokéval, azaz az inhomogenitásoknál nem tekinthetünk el a napi menettől.

- a napi adatsoroknál detektált töréspontok META adatként automatikusan felhasználhatók

- **Csapadék**

- az adott rövid időszakban az órás, napi adatsorok elég homogének,

- nem érdemes a detektált csekély napi inhomogenitásokat az órás adatsorok homogenizálásához felhasználni.

Következésképpen az órás adatsorokon külön-külön hajtunk végre egy enyhe homogenizálást, a szokásos MASH eljárással.

Projekt beszámoló:

*Szentimrey Tamás :Rácsponi adatbázis hatóras adatokra, VARIMAX Bt., (2019)*

# A MISH v1.03 RENDSZER FŐBB TULAJDONSÁGAI

## **Modellező programrendszer (az éghajlati statisztikai paraméterekre)**

- Hosszú homogenizált adatsorok és determinisztikus modellváltozók (pl. topográfia) alapján működik.
- A modellezést csak egyszer kell elvégezni az interpolációs alkalmazások előtt.

## **Interpolációs programrendszer**

- Additív (pl. hőmérséklet) vagy multiplikatív (pl. csapadék) modell és interpolációs formula alkalmazható, a meteorológiai elem eloszlásától függően.
- Napi, havi értékek és sokévi átlagok interpolálhatók. **Órás értékek? Fejlesztés szükséges!**
- Kevés prediktor is elegendő, tekintettel a korábbi modellezésre.
- Becslés az interpolációs hibákra, reprezentativitás értékekre.
- Lehetőség háttérinformáció használatára, pl. műhold, radar, előrejelzés.
- Képesség adatsorok rácspontokba való interpolációjára (gridding)

# INTERPOLÁCIÓ

## Hőmérséklet

- A napi középhőmérséklet adatokra alkalmazott interpolációs formula (normál eloszlású elemekre):

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

ahol  $Z(\mathbf{s}_0)$  (s: hely) prediktandusz,

$Z(\mathbf{s}_i)$  ( $i=1, \dots, M$ ) prediktorok

$\sum_{i=1}^M \lambda_i = 1$  és a  $\lambda_i$  ( $i=1, \dots, M$ ) súlytényezők a sztochasztikus kapcsolatoktól függenek,

továbbá  $E(\mathbf{s}_i)$  ( $i=0, \dots, M$ ) a térbeli trendértékek.

Ez alapján az órás értékekre, az alábbi interpolációs formulát kell alkalmazni

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

ahol  $E(\mathbf{s}_i, t)$  ( $i = 0, \dots, M$ ) az adott időpontokhoz tartozó **térbeli trendértékek**.

Az órás térbeli trendértékek modellezésére, az alábbi lineáris modellt választottuk:

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

ahol  $\alpha(t)$  és  $\beta(t)$  az óraértékekhez tartozó regressziós együttható.

$E(\mathbf{s}, t)$  ( $t = 0,6,12,18$ ) az adott időpontokhoz tartozó **térbeli trendértékek**.

$E(\mathbf{s})$  napi térbeli trendértékek **térbeli trendértékek**.

**Ez esetben az órás értékek interpolációs formulája:**

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

Hónapok	T00		T06		T12		T18	
	<i>alfa</i>	<i>beta</i>	<i>alfa</i>	<i>beta</i>	<i>alfa</i>	<i>beta</i>	<i>alfa</i>	<i>beta</i>
1	-0,97	<b>0,97</b>	-1,52	<b>0,88</b>	2,15	<b>1,06</b>	0,32	<b>1,11</b>
2	-1,26	<b>0,85</b>	-2,08	<b>0,79</b>	2,76	<b>1,08</b>	0,45	<b>1,13</b>
3	-1,38	<b>0,84</b>	-1,61	<b>0,70</b>	2,97	<b>1,13</b>	0,02	<b>1,19</b>
4	-0,38	<b>0,75</b>	-0,69	<b>0,79</b>	2,28	<b>1,17</b>	-0,16	<b>1,15</b>
5	1,27	<b>0,69</b>	0,23	<b>0,85</b>	0,66	<b>1,23</b>	-1,79	<b>1,22</b>
6	1,21	<b>0,73</b>	0,21	<b>0,89</b>	0,22	<b>1,21</b>	-2,78	<b>1,25</b>
7	2,51	<b>0,69</b>	0,57	<b>0,86</b>	1,10	<b>1,16</b>	-3,01	<b>1,24</b>
8	0,97	<b>0,77</b>	0,70	<b>0,82</b>	1,84	<b>1,15</b>	-3,68	<b>1,25</b>
9	0,06	<b>0,81</b>	1,27	<b>0,71</b>	1,67	<b>1,20</b>	-2,87	<b>1,22</b>
10	-0,57	<b>0,84</b>	-0,07	<b>0,70</b>	2,32	<b>1,20</b>	-1,42	<b>1,16</b>
11	-0,46	<b>0,84</b>	-0,80	<b>0,75</b>	2,03	<b>1,17</b>	-0,36	<b>1,10</b>
12	-0,73	<b>0,95</b>	-1,15	<b>0,89</b>	1,90	<b>1,07</b>	0,07	<b>1,08</b>

**Az alfa, beta regressziós paraméterek 12 hónapra, és együtt a 4 órára**

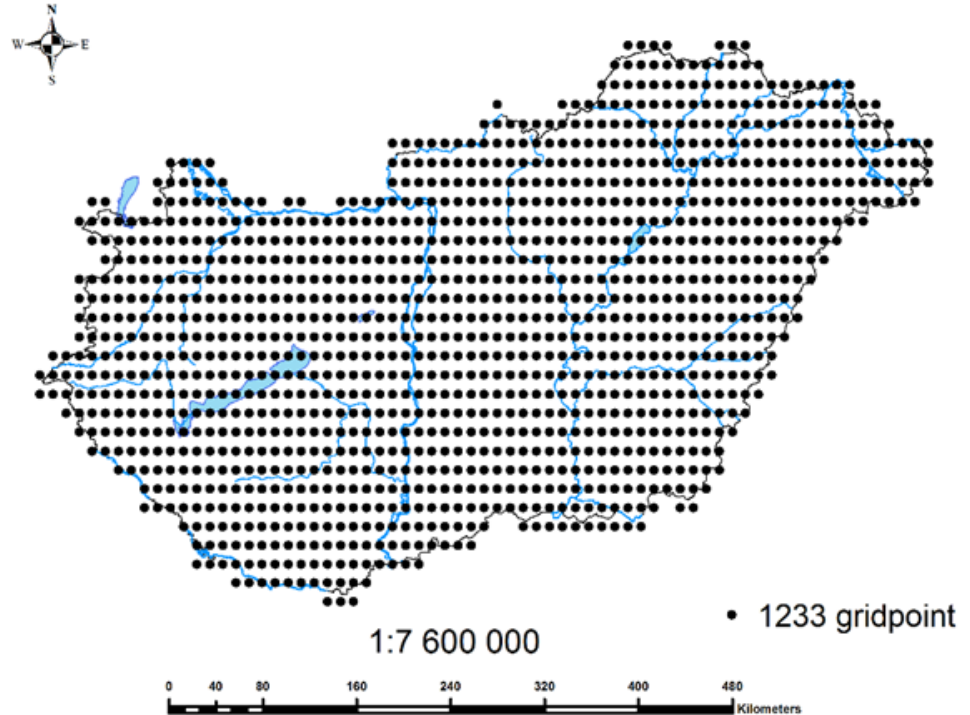
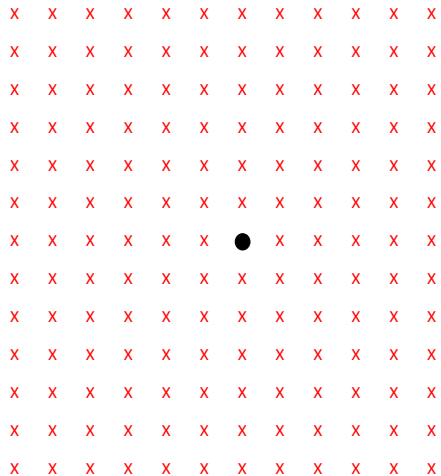
<b>Hónapok</b>	<b>T00</b>		<b>T06</b>		<b>T12</b>		<b>T18</b>	
	<b>corr</b>	<b>tstat</b>	<b>corr</b>	<b>tstat</b>	<b>corr</b>	<b>tstat</b>	<b>corr</b>	<b>tstat</b>
1	<b>0,95</b>	23,02	<b>0,95</b>	22,78	<b>0,89</b>	14,32	<b>0,98</b>	40,44
2	<b>0,89</b>	14,41	<b>0,88</b>	13,64	<b>0,86</b>	12,46	<b>0,98</b>	34,25
3	<b>0,87</b>	12,97	<b>0,84</b>	11,42	<b>0,83</b>	11,23	<b>0,97</b>	31,82
4	<b>0,80</b>	9,87	<b>0,91</b>	16,19	<b>0,86</b>	12,54	<b>0,94</b>	20,80
5	<b>0,78</b>	9,19	<b>0,94</b>	21,08	<b>0,91</b>	15,94	<b>0,94</b>	20,91
6	<b>0,78</b>	9,18	<b>0,96</b>	25,07	<b>0,91</b>	16,08	<b>0,95</b>	22,00
7	<b>0,69</b>	7,19	<b>0,95</b>	22,54	<b>0,88</b>	14,04	<b>0,93</b>	19,70
8	<b>0,73</b>	8,04	<b>0,89</b>	14,53	<b>0,83</b>	11,13	<b>0,93</b>	19,32
9	<b>0,81</b>	10,30	<b>0,83</b>	11,03	<b>0,80</b>	10,15	<b>0,97</b>	28,24
10	<b>0,83</b>	11,10	<b>0,80</b>	10,00	<b>0,78</b>	9,29	<b>0,96</b>	24,57
11	<b>0,90</b>	15,16	<b>0,85</b>	12,02	<b>0,83</b>	11,05	<b>0,98</b>	39,47
12	<b>0,95</b>	21,98	<b>0,94</b>	20,40	<b>0,89</b>	14,49	<b>0,98</b>	40,03

**A korrelációk (corr) és a t-próbastatisztikák (tstat) 12 hónapra, és együtt a 4 órára**

# GRIDDING

0.1°-os felbontás a KLIMADAT számára, ugyanakkor a modellezés 0.5'-es felbontással történik.

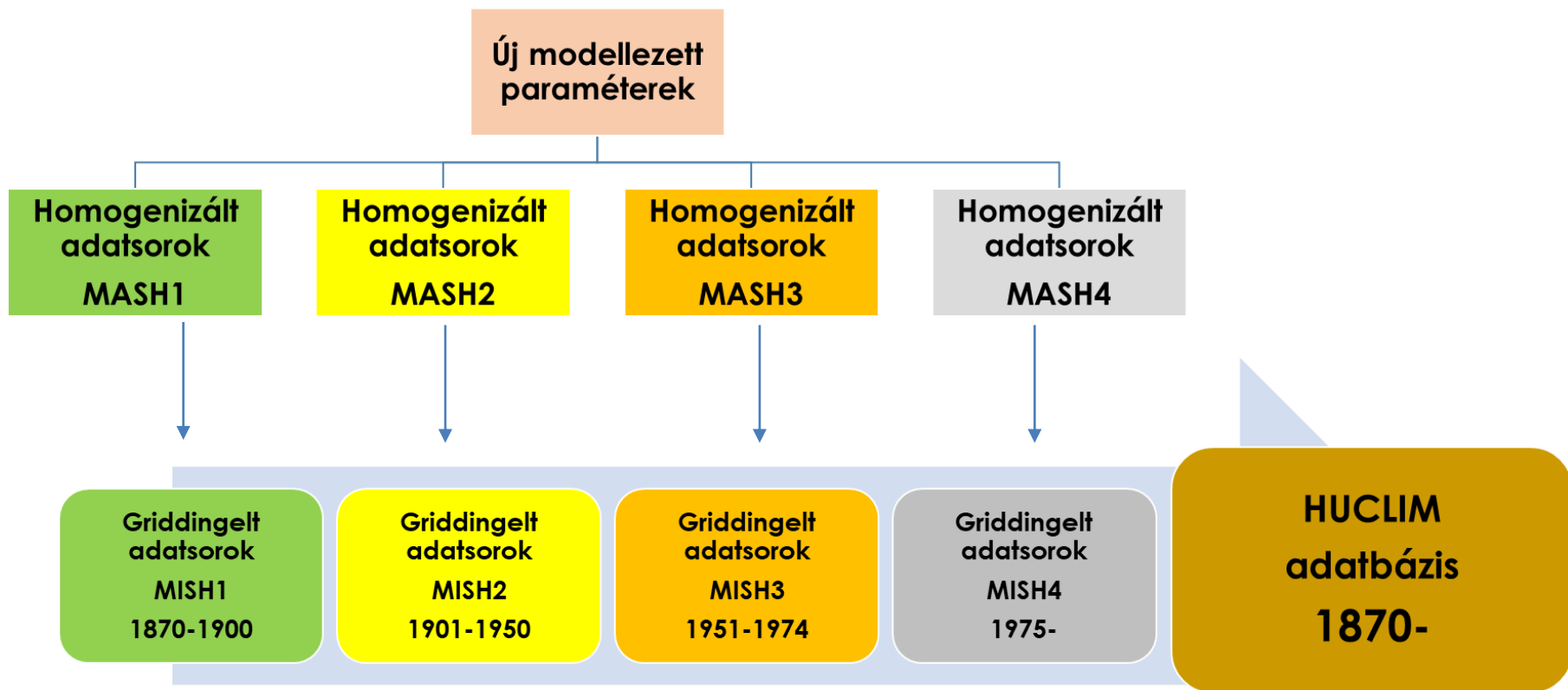
(Ez lehetőséget ad arra, hogy a későbbiekben a boxátlagot meghatározzuk.)



Projekt beszámoló:

Szentimrey Tamás : Transformation of CarpatClim datasets for grid-box average datasets, VARIMAX Bt., (2019)

# HUCLIM ADATBÁZIS ELŐÁLLÍTÁSÁNAK FOLYAMATA NAPI KÖZÉPHŐMÉRSÉKLET ESETÉN



*Izsák, B., Szentimrey, T., Lakatos, M., Pongrácz, R., Szentes, O.: Creation of a representative climatological database for Hungary from 1870 to 2020, *Időjárás* 126, 1-26.*

*doi:10.28974/idojaras.2022.1.1*



# Csapadék

- nincs okunk változó napi menet feltételezésére, mindemellett egy esetleges modellezéshez, megfelelő mintával sem rendelkezünk.
- Multiplikatív formula (lognormál eloszlású elemekre használható):

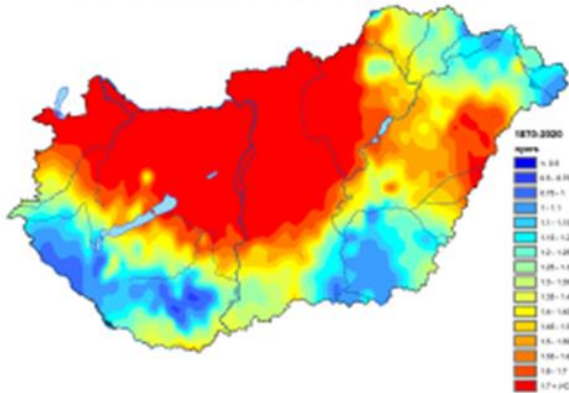
$$\hat{Z}(\mathbf{s}_0) = \vartheta \cdot \left( \prod_{q_i \cdot Z(\mathbf{s}_i) \geq \vartheta} \left( \frac{q_i \cdot Z(\mathbf{s}_i)}{\vartheta} \right)^{\lambda_i} \right) \cdot \left( \sum_{q_i \cdot Z(\mathbf{s}_i) \geq \vartheta} \lambda_i + \sum_{q_i \cdot Z(\mathbf{s}_i) < \vartheta} \lambda_i \cdot \left( \frac{q_i \cdot Z(\mathbf{s}_i)}{\vartheta} \right) \right)$$

ahol  $\vartheta > 0$ ,  $q_i > 0$   $\lambda_i \geq 0$  ( $i = 1, \dots, M$ )  $\sum_{i=1}^M \lambda_i = 1$  ( $i = 1, \dots, M$ ) súlytényezők

# Mennyire tér el egymástól a nyers adatsorok és a homogenizált sorok trendje?

## Éghajlatváltozás?

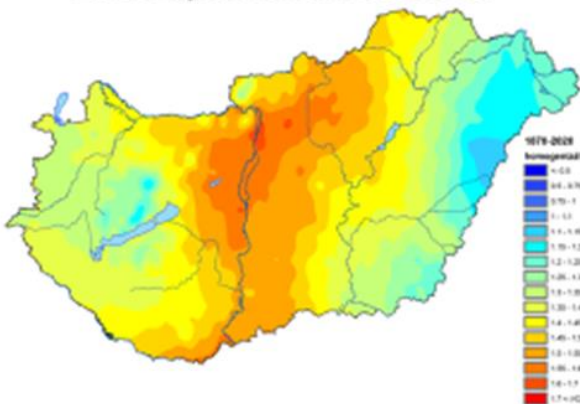
Évi középhőmérséklet változás



- *Nyers adatsorokkal dolgozom*
- *Nem baj, hogy változtak a mérési körülmények*
- *Nem kell adatellenőrzés*
- *Gyorsan publikálom az eredményeket*



Évi középhőmérséklet változás



- *Adatellenőrzést végzek*
- *Pótlom a hiányzó értékeket*
- *Homogenizált adatsorokat elemzek*
- *Adekvát matematikai módszereket és szoftvereket használok*
- *Eredményeket csak ezután publikálom*



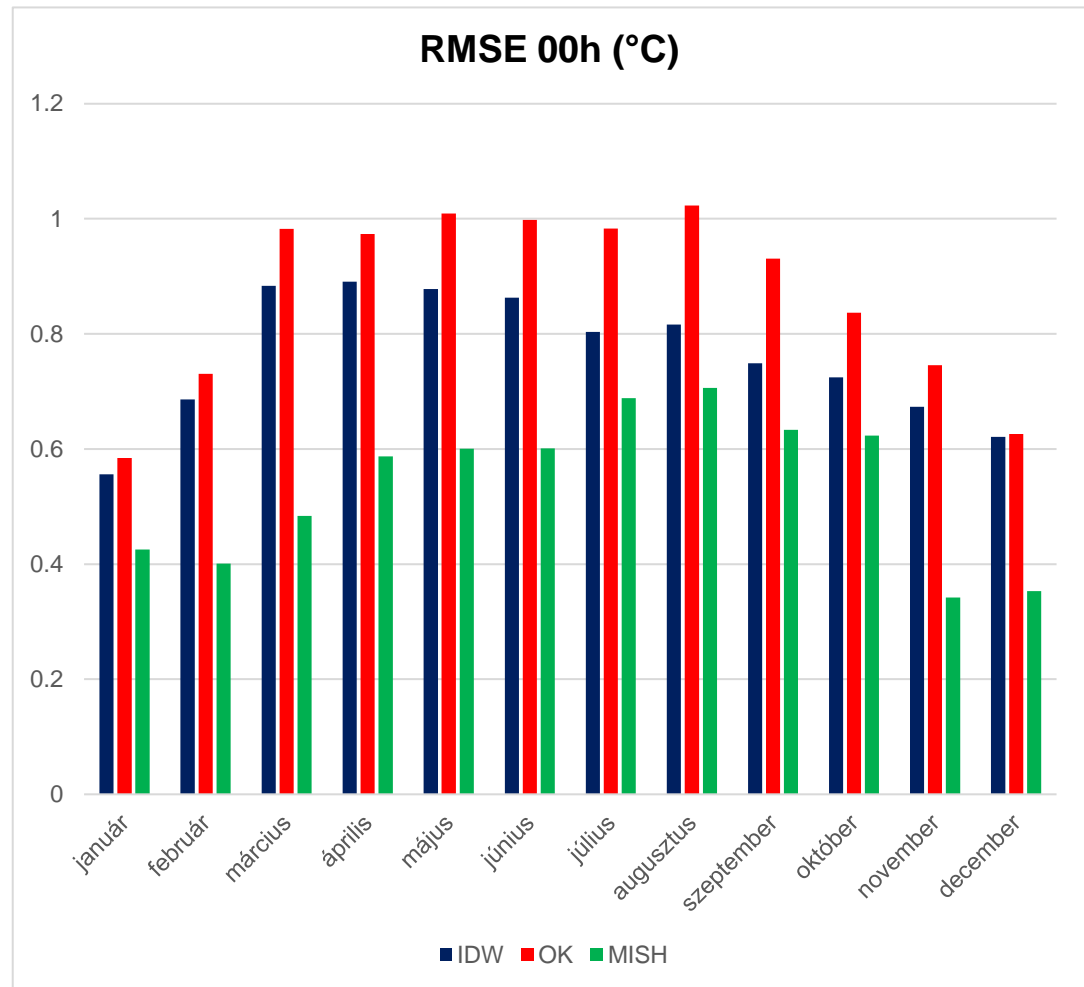
# MISH, KÖZÖNSÉGES KRIGING, INVERZ TÁVOLSÁG MÓDSZERE RMSE ÉRTÉKEI, HŐMÉRSÉKLET ESETÉN

## Az interpolációs hiba jellemzése

$$MSE(\mathbf{s}_0) = E\left(\left(Z(\mathbf{s}_0, t) - \hat{Z}(\mathbf{s}_0, t)\right)^2\right)$$

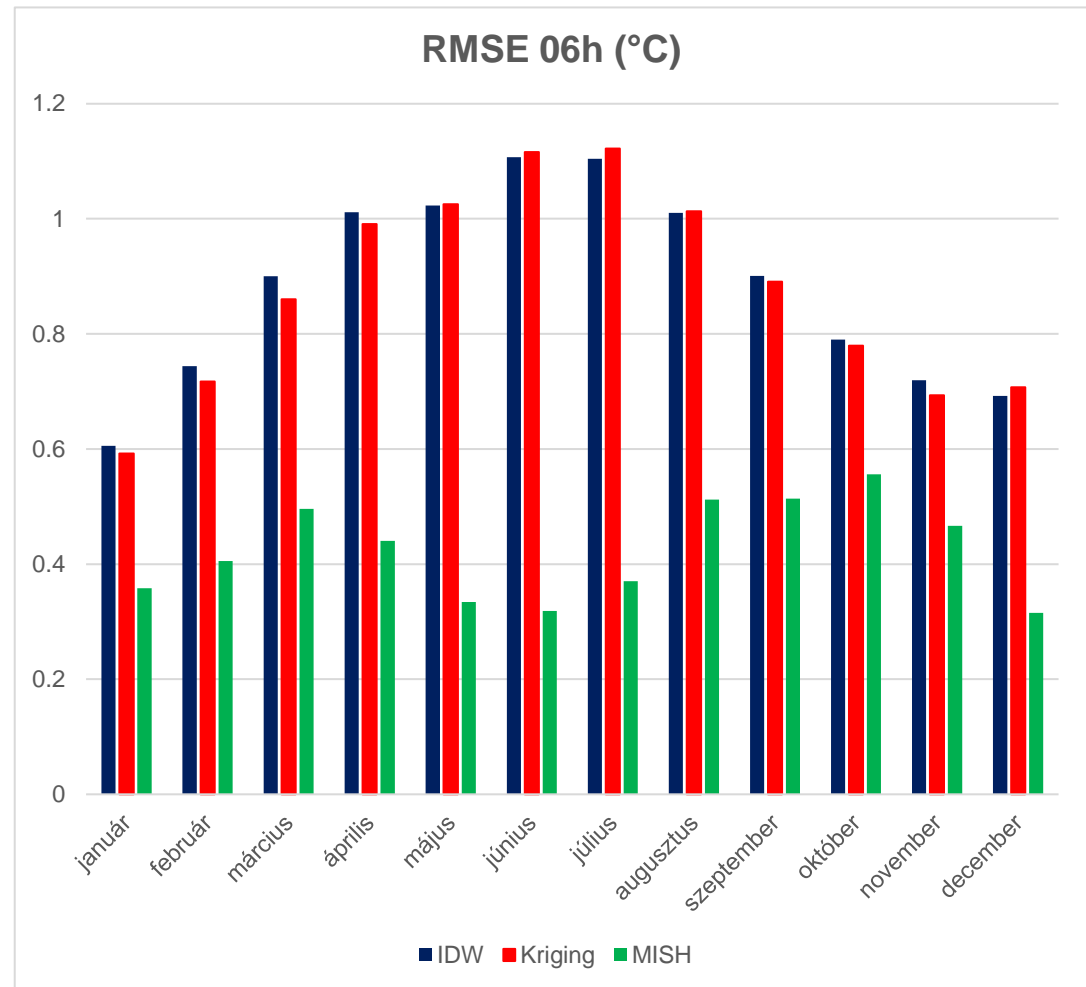
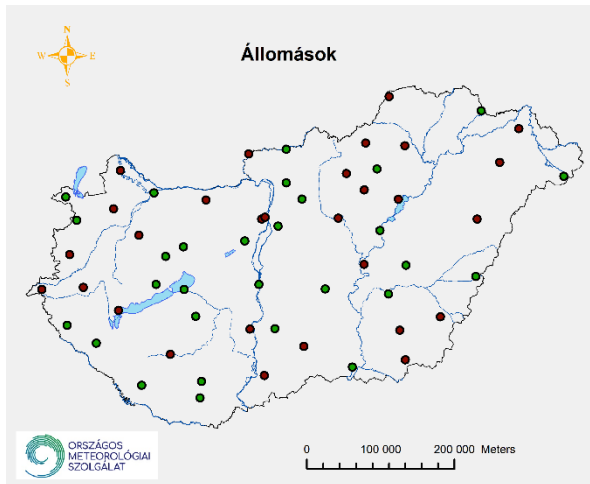
$$RMSE(\mathbf{s}_0) = \sqrt{MSE(\mathbf{s}_0)}$$

Azok a matematikailag optimális interpolációs paraméterek, amelyekre az interpolációs hiba minimális.



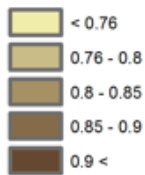
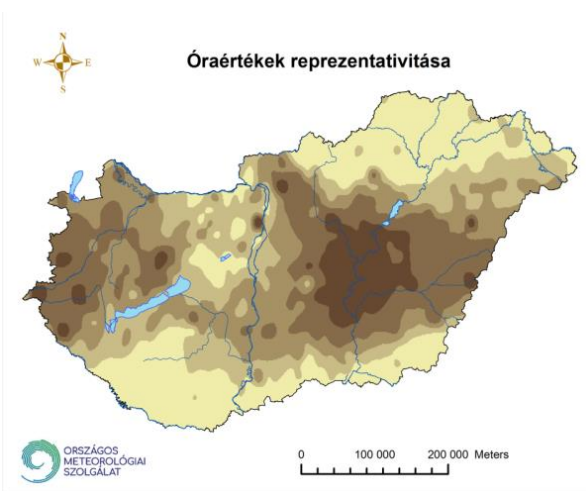
# MISH, KÖZÖNSÉGES KRIGING, INVERZ TÁVOLSÁG MÓDSZERE RMSE ÉRTÉKEI

A *training* és *teszt* állomások földrajzi elhelyezkedése



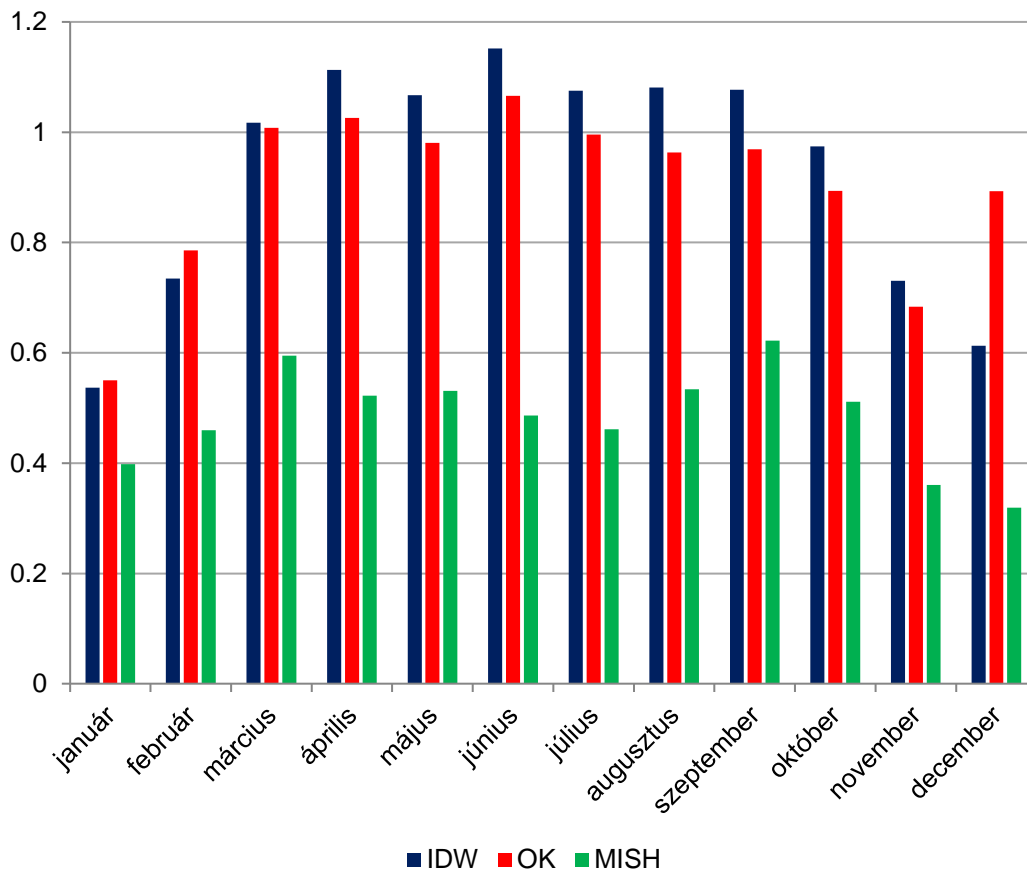
# MISH, KÖZÖNSÉGES KRIGING, INVERZ TÁVOLSÁG MÓDSZERE RMSE ÉRTÉKEI

Átlagos éves reprezentativitás értékek napi középhőmérséklet esetén.



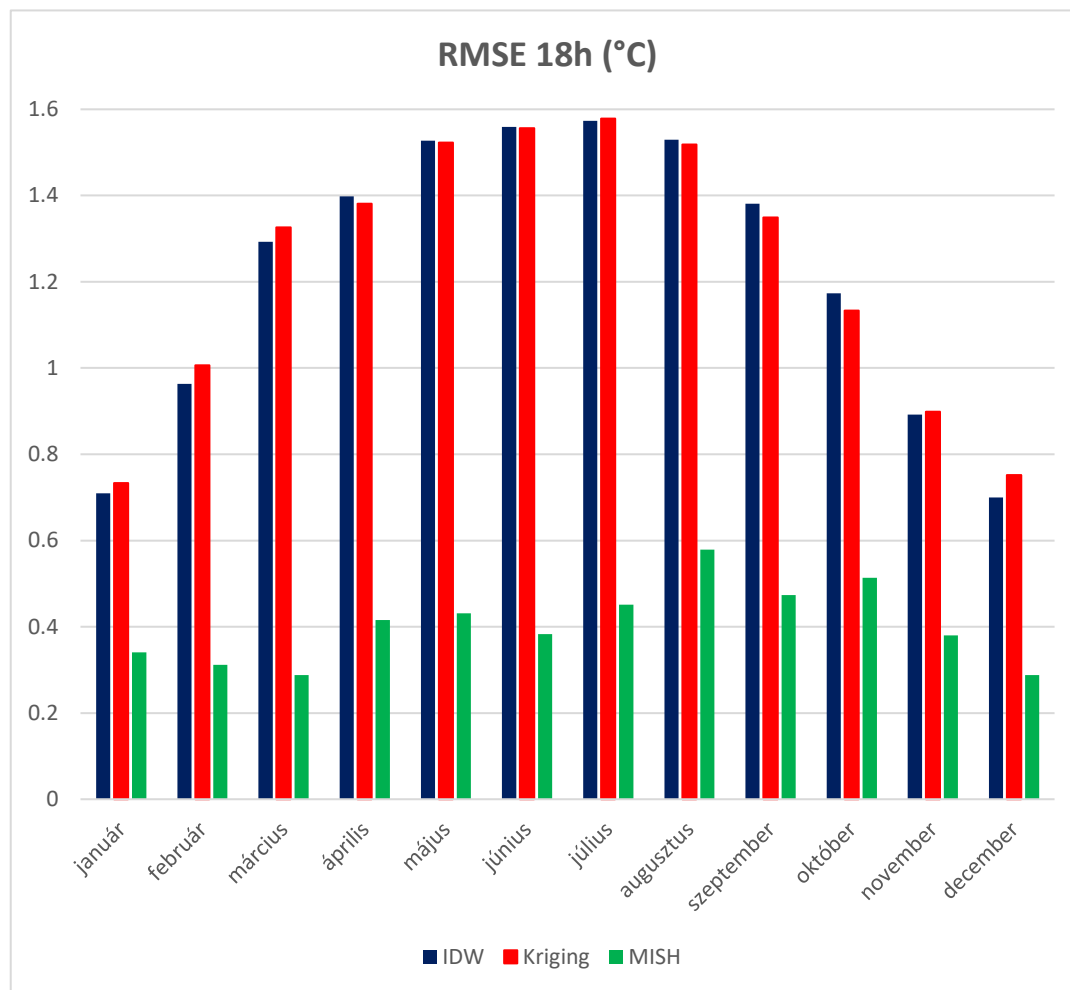
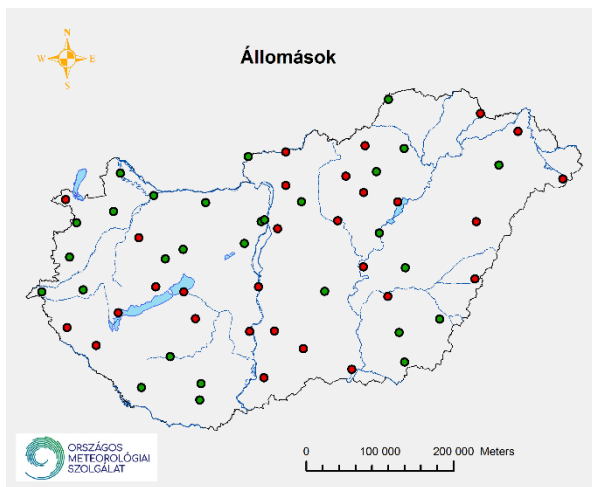
$$REP(s_0) = 1 - \frac{RMSE(s_0)}{D(s_0)}$$

## RMSE 12 UTC (°C)



# MISH, KÖZÖNSÉGES KRIGING, INVERZ TÁVOLSÁG MÓDSZERE RMSE ÉRTÉKEI

A *training* és *teszt* állomások földrajzi elhelyezkedése



# KONKLÚZIÓ

- Regionális klímamodellezés
- Méréseken alapuló klimatológiai célú adatbázis
- Térben és időben reprezentatív adatbázis feltétele az adekvát matematikai alapokon nyugvó szoftverek használata!

PONTOSABB,  
RÉSZLETESEBB  
eredmények,  
megbízhatóbb  
hatásvizsgálat, döntés  
stb...

PI. MISH és  
MASH szoftverek  
használata

# KÖSZÖNÖM A FIGYELMET!

**SZÉCHENYI**  2020



MAGYARORSZÁG  
KORMÁNYA

**Európai Unió**  
Kohéziós Alap



**BEFEKTETÉS A JÖVŐBE**