

# ÉGHAJLATVÁLTOZÁS: HOMOGENIZÁLT VAGY NYERS ADATSOROKAT VIZSGÁLJAK?

## CLIMATE CHANGE: SHOULD I STUDY HOMOGENIZED OR RAW DATASETS?

Izsák Beatrix, Bihari Zita, Szentes Olivér

Országos Meteorológiai Szolgálat, 1024 Budapest, Kitaibel Pál utca 1., [izsak.b@met.hu](mailto:izsak.b@met.hu), [bihari.z@met.hu](mailto:bihari.z@met.hu), [szentes.o@met.hu](mailto:szentes.o@met.hu)

**Összefoglaló:** A 2021-ben életbe lépett nyílt OMSZ adatpolitika lehetőséget teremt arra, hogy bárki elemezze, statisztikai vizsgálatoknak vesse alá a meteorológiai méréseket. Ahhoz azonban, hogy a jelen és az elmúlt időszakok éghajlatát vizsgálni tudjunk, elkerülhetetlen a nyers műszeres mérések feldolgozása, az úgynevezett adatszerzés, ami az adatellenőrzést, a hiányzó adatok pótlását, a homogenizálást és az adatok térbeli interpolálását jelenti. Bemutatjuk az alkalmazott matematikai statisztikai módszereket, hangsúlyt fektetve az elmúlt évtizedekben hazánkban, az Országos Meteorológiai Szolgálat Éghajlati Osztályán kifejlesztett eljárásokra. Végezetül rámutatunk arra, hogy mekkora hibát vétünk, ha a könnyebb utat választjuk, és nyers adatsorokon detektálunk trendeket.

**Abstract:** The open data policy of OMSZ, which came into force in 2021, provides an opportunity for anyone to analyze and subject the meteorological measurements to statistical studies. However, in order to study the climate of the present and recent periods, it is inevitable to process raw instrumental measurements, the so-called data management, which means quality control, missing data completion, homogenization, and spatial interpolation of the data. We present the applied mathematical statistical methods, with an emphasis on the procedures developed in Hungary in recent decades in the Climate Department of Hungarian Meteorological Service. Finally, we show how much error we make when we choose the easier way and detect trends on raw data sets.

### Bevezetés

Azt már általános iskolában megtanuljuk, hogy míg az időjárás a légkör pillanatnyi állapota az adott helyen, az adott időpontban, addig az éghajlat az adott hely hosszú időszakra vonatkozó időjárási viszonyainak összessége. A Meteorológiai Világszervezet ajánlása alapján 30 éves időszak alatt mutatott statisztikai tulajdonságaival jellemezzük egy adott földrajzi hely éghajlatát. Ha a matematikai statisztika oldaláról közelítjük meg az éghajlatot, akkor úgy fogalmazhatunk, hogy ha az adott helyen ismerjük a meteorológiai (vektor) változók eloszlását, akkor ismerjük az éghajlatát. Amennyiben ezek a valószínűségi eloszlások változnak, akkor beszélünk éghajlatváltozásról (*Szentimrey*, 2011). Tehát a statisztikus megközelítésben az éghajlatváltozást a meteorológiai adatokból, vagyis a mérésekből tudjuk becsülni, ezekből lesz a statisztikai minta a vizsgálatokhoz. Ebből következik, hogy csak olyan matematikai statisztikai modellek, módszerek alkalmazhatók, melyek képesek figyelembe venni a valószínűségi eloszlás, azaz az éghajlat változását. Továbbá az is következik, hogy ezekből az adatokból minden hibás információt el kell távolítani, mielőtt vizsgálatoknak vetjük alá.

### Homogenizálás

Elsősorban tehát jó minőségű adatokra van szükség: mind térben, mind időben úgynevezett reprezentatív adatbázist kell létrehozni. Ez a gyakorlatban annyit jelent, hogy a lehető legtöbb mérőállomás adatait, az elérhető leg-hosszabb időszakra használjuk fel a vizsgálatainkhoz. A mérésekkel azonban gond van: inhomogenitással terheltek a nyers adatsoraink. Ugyanis a rendszeres műszeres mérések kezdete óta a mérőeszközök folyamatosan cserélődtek, a meteorológiai állomásokat sokszor költöztették,

és történtek módszertani váltások is. Ezek a külső változások inhomogenitást okoznak a mérési sorokban. Az adatokat homogenizálni kell, ez annyit jelent, hogy a jelen mérési körülményekhez igazítjuk a múltbeli méréseket (*WMO*, 2020). Gondoljunk csak például arra, hogy ha a nappali három méréshez hozzávesszük a negyedik éjszakai mérést, akkor erősen csökkentjük a napi átlaghőmérsékleteket, ami nem éghajlatváltozás, hanem módszertani váltás. Mi van, ha egy melegebb klímájú, belterületi mérőállomást külterületre költöztetnek? Ez valóban éghajlatváltozás? Ezek és ezekhez hasonló problémák felismerése és ennek kezelése indította útjára a homogenizálást a meteorológiában. Emellett az adatok hibával is terheltek, tehát egy jó adatellenőrző program is szükséges, továbbá sok esetben találkozunk adathiánnyal, melyeket a legjobb tudásunk szerint pótolni kell. Ennek a három problémának az együttes kezelésére készült el a MASH (Multiple Analysis of Series of Homogenization) szoftver (*Szentimrey*, 2014). Miért kellett szoftvert is készíteni, miért nem volt elég kidolgozni az elméletet? A homogenizálás matematikája az 1990-es évekre még nem volt kidolgozva, akadtak ugyan próbálkozások külföldi meteorológiai szolgálatoknál, készítettek homogenizáló szoftvereket, de ezek nem szolgáltattak megfelelő eredményeket, vagy nem voltak elég hatékonyak. A MASH szoftver kiválóan alkalmazható erre a feladatra, mely annak köszönhető, hogy az elméleti háttere klasszikus matematikai statisztikai tételeken, összefüggéseken alapul. Relatív homogenizálási elven működik, azaz a környező állomások idősoraival veti össze az adott állomás idősorát, ezen kívül adekvát matematikai modell biztosítja annak feltételét, hogy csak olyan töréspontokat detektáljunk az adatsorokban, amelyek nem az éghajlat változékonyságából erednek. Hosszan lehetne még taglalni a MASH tulajdonságait. Az egyik legfontosabb,

hogy hipotézisvizsgálaton alapul, ez teszi lehetővé, hogy – korábbi eredményekre alapozva – térben és időben is frissíteni tudjuk a homogenizált adatbázisunkat. Érdeemes megvizsgálni, hogy mennyire más képet kapunk, ha a nyers adatsorokból becsüljük az éghajlat megváltozását (Izsák, Szentimrey, 2020), ami igazolja, hogy fontos kiszűrni az inhomogenitásokat a mérésekből.

**Tapasztalatok a MASH szoftver eredményei alapján**

A napi adatok ellenőrzése során rengeteg olyan hibát találtunk, ami arra utal, hogy a rögzítés során került rossz érték az adatbázisba. Az archív adatok ellenőrzésére az OMSZ-ban kizárólag a MASH rendszert használjuk. Természetesen vannak olyan helyzetek, melyekben a MASH gyanús értéket jelez, de további információ hiányában nem tudjuk eldönteni, hogy valóban hibás-e az érték. Sok esetben az évkönyvekben megtalálható egyéb információk egybevetésével már tudunk dönteni. Ilyen például a felhőzet, csapadék, minimum és maximum hőmérséklet összevetése a napi átlaghőmérséklettel. Például egy nagyon alacsony napi középhőmérséklet érték lehet valós, ha egész nap borult, csapadékos idő volt, és a maximum érték is alacsony. Mivel a múltban sokkal ritkább volt az állomáshálózat, így csak a korabeli feljegyzésekre hagyatkozhatunk. A közelmúltban viszont rendelkezésünkre állnak további információk is, például radar és műholdas adatok. Arra, hogy mennyire hatékony a MASH rendszer, bizonyíték, hogy a folyamatosan rögzített archív adatok ellenőrzése során szinte minden esetben olyan értéket jelzett gyanús-nak, amelyre később, az évkönyveket fellapozva magyarázatot találtunk. Például hőmérséklet esetén elfelejtették rögzíteni az előjelet, vagy nem azt a hónapot rögzítették, amelyiknek a dátuma szerepel az eredeti táblázat fejlécében.

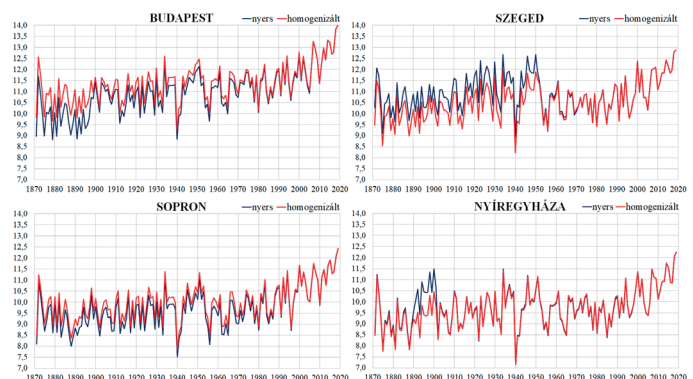
A homogenizálás során igen érdekes dolgokra bukkantunk a MASH szoftver futtatásával. Miskolcon 1901-től 1908-ig az évi középhőmérsékleti idősorokat tekintve alacsonyabb értékek szerepelnek, mint az azt következő időszakban. A MASH erre az állomásra nagy inhomogenitást jelez a különbség sorozatok vizsgálata alapján – ugyanis ez egy relatív homogenizálási elven működő szoftver, azaz nem önmagában keresi a töréspontokat az idősorokban, hanem a környező állomásokkal összevetve. Ha nem végezzük el a homogenizálást, akkor azt kapnánk, hogy az ország egyetlen pontján egyetlen év alatt szokatlan mértékben, kb. 2 °C-ot nött az átlaghőmérséklet. Fellapozva azonban a napi jelentéseket az OMSZ archívumában, meg is van a magyarázat: Miskolcon akkoriban Réaumur-skálán rögzítették a méréseket, és az észlelési könyvben sajnos nem mindenhol húzta át az észlelő a °C-ot. Tehát a digitalizálás során az adatrögzítőnek elkerülte a figyelmét, hogy a papíron szereplő értéket még át kell váltani.

Egy másik példa arra, hogy miért nem szabad nyers adatokat vizsgálni: Nyíregyháza állomás napi középhőmérséklet értékei 1890-1901-ig. Nyomozást folytatva az archív adatok között azt találtuk, hogy ebben az időszakban az észlelő nem reggel 7h-kor, délután 2h-kor és este 9h-kor észlelt, hanem 8h-kor, 13h-kor és 20h-kor, így kiugróan magas a napi és ezáltal az évi átlaghőmérséklet ezekben

az években. Természetesen ez az eltérés sem az éghajlatváltozásnak tudható be, hanem az észlelési időpontok megváltozásából adódik. Jelen esetben a nyers adatsorban lévő eltolódás megegyező nagyságú a mérések kezdetétől bekövetkezett melegedéssel (Izsák, 2021)!

**Trendbecslés nyers és homogenizált adatsorok alapján**

Az Országos Meteorológiai Szolgálat alapításának 150. évfordulója alkalmából közzétettük négy magyarországi nagyváros adatsorát 1870-től kezdődően. Ezen magyarországi állomások éghajlatát és annak megváltozását vizsgáltuk. Az eredeti és homogenizált adatsorokat összehasonlítva a grafikonokon jól látszik, hogy mennyire inhomogének az eredeti mérési sorok (1. ábra). Budapest esetén a városhatás rajzolódik ki, Szegeden egyértelműen az 1951-es külterületre való telepítés okozza a legnagyobb törést, hasonló a helyzet Sopronban is. Nyíregyházán pedig a mérési időpontokban történt váltás eredményez a detektált éghajlatváltozással azonos nagyságrendű inhomogenitást.



1. ábra. Homogenizált és nyers évi átlaghőmérséklet idősorok 1871-2019-ig (°C).

Természetesen a költöztetés, mérési időpont megváltoztatása, műszercsere, módszertani váltás és a városhatás együttes hatása mutatkozik meg az inhomogenitásokban. Amennyiben a homogenizált adatsorok helyett a nyers adatsorokat választjuk ahhoz, hogy éghajlati értékelőt készítsünk, nagyon félrevezető és hibás eredményeket kapunk. Ennek bemutatására elkészítettük az évi átlaghőmérséklet becsült megváltozását nyers adatsorok és homogenizált sorok alapján is (1. táblázat). Budapest és Sopron esetében a nyers adatsorok alapján igen nagyfokú melegedést kaptunk, míg a homogenizált sorok mérsékelt melegedést mutatnak. A másik két állomáson fordított a helyzet, ott a nyers adatsorok kisebb mértékű felmelegedésre engednek következtetni.

	SOPRON	BUDAPEST	SZEGED	NYÍREGYHÁZA
<b>Homogenizált</b>	1,43 °C	1,69 °C	1,48 °C	1,50 °C
<b>Nyers</b>	1,82 °C	2,53 °C	0,60 °C	1,10 °C
<b>Különbség</b>	-0,39 °C	-0,84 °C	0,88 °C	0,40 °C

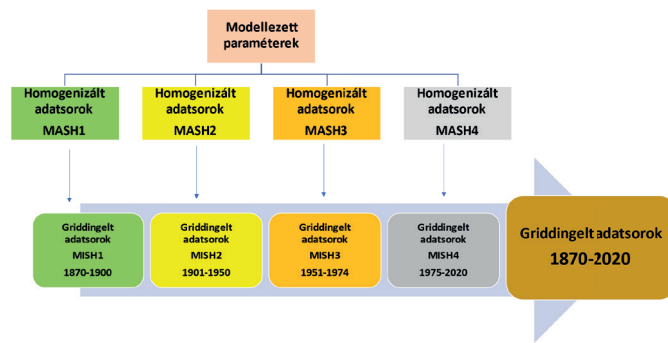
1. táblázat. Az évi középhőmérséklet értékek teljes időszak alatt változása (1870-2019) lineáris trendbecsléssel kapott értékei homogenizált és nyers állomási adatsorok alapján (az értékek minden esetben szignifikánsak a 0,1-es szignifikancia szinten).

## Interpoláció

A MASH használatával tehát már van egy időben reprezentatív adatbázisunk, de térben még nem az. Ugyanis vegyük azt a gyakorlati példát, hogy mivel sok meteorológiai mérőállomásunk van a Dunántúlon, kevesebb az Alföldön, és a hegyeinkben is kevés a mérés, akkor ezek átlaga nem kellő mértékben képviseli a teljes magyarországi területre vonatkozó értékeket, ráadásul például a tiszántúli vagy a borsónyi viszonyokról semmit nem tudunk. A következő lépés tehát a térbeli reprezentativitás biztosítása: sűrű, szabályos rácshálózatra interpoláljuk az értékeket, és így az ország teljes területén tudunk az éghajlatról és annak esetleges megváltozásáról nyilatkozni. A térbeli interpoláció egy olyan matematikai eljárás, amely egy állapotjelző (meteorológiai elem) értékét határozza meg egy tetszőleges pontban a szomszédos helyeken megfigyelt, ismert adatok alapján (Cressie, 1991; Szentimrey et al., 2011). Erre a feladatra a MISH (Meteorological Interpolation based on Surface Homogenized Data Basis) szoftver készült el az OMSZ Éghajlati Osztályán (Szentimrey és Bihari, 2014). Miért nem jók a meteorológiai adatokra a már meglévő interpolációs szoftverek? Sajnos még napjainkban is sokan használják a legközelebbi társ és az inverztávolság módszereket, annak ellenére, hogy ezek nagyobb térségre vagy változó domborzati viszonyokra egyáltalán nem használhatók. Az interpolációs módszerek közül a földtudományokban legelterjedtebbek, legismertebbek a geostatistikai módszerek (Cressie, 1991; Szentimrey et al., 2011). Alkalmazásuk a meteorológia területén nem szolgáltat megfelelő eredményeket. A meteorológiában nem élhetünk azzal a feltételezéssel, hogy nincsen térbeli trend, hiszen nem lehet ugyanaz a várható érték egy meteorológiai elemnek például Szegeden és a Kékestetőn, pedig ez sok módszer alapja (közönséges kriging). Az univerzális kriging már feltételezi, hogy van térbeli trend, de még nem használja fel ennek modellezésére az időbeli mintát. Egy minőségi ugrás a reziduális kriging, hiszen itt már időbeli mintából becsüljük a térbeli trendértékeket. Ugyanakkor az eddig felsorolt kriging módszerek közös jellemzője, hogy két mérőállomás közötti sztochasztikus kapcsolat modellezéséhez mindössze egyetlen időbeli realizációt alkalmaznak. A földtudományokban sok esetben ez az egyetlen járható út. A meteorológia viszont abban a szerencsés helyzetben van, hogy rendelkezik hosszú adatsorokkal, ezáltal sokkal több információ áll rendelkezésre egy-egy adott térségre, mint egyéb tudományokban. Ezeket a MASH-sel készült, hosszú adatsorokban rejlő információkat használja fel a MISH szoftver a modellezéshez, ezáltal sokkal jobb becslést tudunk adni egy olyan pontba, ahol korábban nem volt mérés, mintha csak egy pillanatkép alapján adnánk becslést a sztochasztikus kapcsolatokra vagy a térbeli trendekre. A MASH és MISH használatával tehát már valóban, térben és időben reprezentatív adatbázist tudunk készíteni. A két szoftver közös tulajdonsága, hogy az adott elemnél figyelembe veszi a valószínűségi eloszlást, ez alapján választhatunk a multiplikatív vagy additív modell között (Szentimrey, 2014; Szentimrey, Bihari, 2014). Ez sajnos nem jellemző a legtöbb homogenizáló és interpoláló szoftverre, melyek valószínűségi eloszlástól függetlenül, a normális eloszlás esetén alkalmazott, additív modellel számolnak (Venema et al., 2012).

## Reprezentatív adatbázis készítése az OMSZ Éghajlati Osztályán

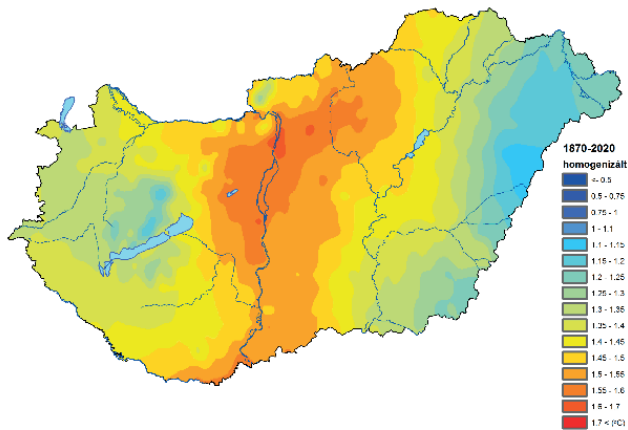
A múlt és jelen éghajlatának megismeréséhez, az éghajlatváltozás tanulmányozásához térben és időben reprezentatív adatbázis szükséges. Ma már az Országos Meteorológiai Szolgálat állomáshálózata kellően sűrű ahhoz, hogy az egész ország klimatológiai viszonyairól pontos képet kapjunk. Azonban nem volt ez mindig így. Ugyan már a 19. század közepétől vannak feljegyzések, de lényegesen kevesebb meteorológiai elemre és nagyon kevés állomásra. Azt is figyelembe kell venni, hogy az archív adatokat folyamatosan digitalizálják, évről évre több állomás adatsorát rögzítik, így szükséges ezek ellenőrzése, homogenizálása és pótlása is, és csak ezután lehet felhasználni ezeket klimatológiai célú vizsgálatokhoz. Az OMSZ Éghajlati Osztályán minden évben frissítjük az elmúlt év adataival az éghajlati adatbázisunkat. Alapelvünk, hogy a lehető legtöbb meteorológiai állomás adatsorát, a mérésekben rejlő legtöbb információt használjuk fel. Például középhőmérséklet idősorhoz 1870-től 11, 1901-től 33, 1951-től 55 és 1975-től 110 állomás adatsora áll rendelkezésünkre az adatbázis elkészítéséhez. Ehhez a négy különböző állomáshálózaton alapuló homogenizált rendszert egymáshoz harmonizáljuk, és így biztosítjuk, hogy mind a közös rész, mind önmagukban a különböző állomásrendszerek homogének legyenek. Az állomási adatsorokat a MISH szoftverrel interpoláljuk, és a végső griddingelt adatbázis úgy áll elő, hogy minden időpontban a legtöbb állomási adatsorból interpolált értékek kerülnek be az adatbázisunkba (2. ábra).



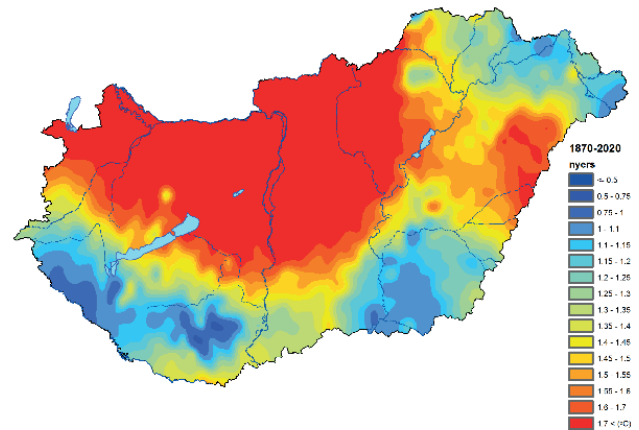
2. ábra. Klimatológiai adatbázis készítésének folyamata.

Az így elkészült adatbázis nemcsak az éghajlatváltozás tanulmányozására alkalmas, hanem inputként szolgál a jövőbeli projekciók elkészítéséhez, alkalmas a klímamodellek eredmények verifikálására. Ezekre alapulnak a legújabb, 1991–2020-as klímanormálok, ezáltal Magyarország éghajlati atlaszának megújításához előállt egy reprezentatív éghajlati adatbázis. Azt is meg kell jegyeznünk, hogy a különböző állomáshálózatok reprezentativitás értékei, melyek a MISH eljárás alkalmazása során előállnak, hasznos információkat hordoznak a jövőbeli telepítésekhez, azaz optimalizálni tudjuk a tervezéskor a teljes magyarországi állomáshálózatot.





3. ábra. Homogenizált állomási adatsorokból interpolált évi középhőmérséklet értékek alapján, lineáris trendbecsléssel kapott hőmérsékletváltozás az 1870-2020-as időszakra vonatkozóan (°C).



4. ábra. Nyers állomási adatsorokból interpolált évi középhőmérséklet értékek alapján, lineáris trendbecsléssel kapott hőmérsékletváltozás az 1870-2020-as időszakra vonatkozóan (°C).

## Konklúzió

A címben feltett kérdésre egyértelmű a válaszunk: homogenizált adatsorok alapján tudjuk pontosabban becsülni az éghajlat változásait. Ahhoz, hogy lássuk, hogy mekkora ez a különbség, érdemes megtekinteni a 3. és 4. ábrát. Ha a nyers adatsorokból szeretnénk következtetéseket levonni, akkor nagyon könnyen tévútra juthatunk. Nem csak azzal van a gond, ha valamely magyarországi területen nem kapunk statisztikai értelemben szignifikáns trendet a nyers adatokból, noha a homogenizált adatok szignifikáns változást mutatnak. Az is aggályos, ha a világhálón elérhető, bizonytalan minőségű adatforrásokon alapuló becslés jócskán meghaladja a valós értéket. Gondoljunk arra, hogy a klímamodell eredményeket jellemzően az elmúlt 100 év alatti változások tükrében értelmezzük. Ha nem homogenizált, kétes forrásokból származó információkat idéz egy szakember, például azt, hogy 4 °C-ot melegedett Budapest az elmúlt 100 évben, akkor egy laikus, de akár egy döntéshozó sem érzi veszélyesnek a klímamodell által 2050-re becsült 2 °C-os hőmérséklet növekedést. Amennyiben a szakember azt kommunikálja, hogy 1,69 °C-ot melegedett Budapest, és ez jóval az országos és globális átlag fölötti érték, akkor a további 2 °C-os melegedés már komoly és egyben elkerülendő melegedés. Sajnos a mai napig sokan használják klimatológiai vizsgálatokhoz a nyers adatsorokat, amik hamis eredményre vezetnek. Ugyanígy elterjedt a reanalízis adatok használata, melyeknél sokkal pontosabb a felszíni méréseken alapuló adatbázisok vizsgálata (Bandhauer et al., 2021). Persze ez esetben sem mindegy, hogy milyen szoftvert használunk. A MISH és a MASH szoftverek nagy előnye, hogy adekvát matematikai alapon nyugszanak. Ez teszi lehetővé, hogy ezek használatával valós képet kapjunk a jelen és az elmúlt időszakok éghajlati viszonyairól.

## Irodalom

Bandhauer, M., Isotta, F., Lakatos, M., Lussana, C., Bäse-  
rud, L., Izsák, B., Szentés, O., Tveito, O. E., and Frei, C.,  
2021: Evaluation of daily precipitation analyses in E-OBS  
(v19.0e) and ERA5 by comparison to regional high-re-

solution datasets in European regions. *International Journal of Climatology*, 1–21.

<https://doi.org/10.1002/joc.7269>

Cressie, N., 1991: *Statistics for Spatial Data*. Wiley, New York  
Izsák, B., Szentimrey, T., 2020: To what extent does the  
detection of climate change in Hungary depend on the  
choice of statistical methods? *Int. J. Geomath* 11, 17  
<https://doi.org/10.1007/s13137-020-00159-7>

Izsák, B., 2021: Homogenizálás a meteorológiában:  
Vissza a jövőbe. *ÉLET ÉS TUDOMÁNY* (0013-6077):  
LXXVI. évfolyam 6. szám

Szentimrey, T., 2014.: *Manual of homogenization software  
MASHv3.03*. Hungarian Meteorological Service,  
Budapest, Hungary

Szentimrey, T. and Bihari, Z., 2014: *Manual of interpolation  
software MISHv1.03*. p. 89. Hungarian Meteorological  
Service, Budapest, Hungary

Szentimrey, T., Bihari, Z., Lakatos, M., and Szalai, S., 2011:  
Mathematical, methodological questions concerning the  
spatial interpolation of climate elements. *Proceedings  
from the Second Conference on Spatial Interpolation  
in Climatology and Meteorology*, Budapest, Hungary,  
2009, *Időjárás* 115, 1–2,

Szentimrey, T., 2011: Éghajlati adatsorok elemzése, homo-  
genizálása In (szerk.: Bartholy Judit, Bozó László,  
Haszpra László): *Klímaváltozás – 2011*, MTA-ELTE  
Meteorológia Tanszék – Budapest – 2011), 38-43 p.,

Venema, V., Mestre, O., Aguilar, E., Auer, I., Guijarro,  
J. A., Domonkos, P., Vertacnik, G., Szentimrey, T.,  
Štěpánek, P., Zahradnicek, P., Viarre, J., Müller-West-  
ermeier, G., Lakatos, M., Williams, CN., Menne, M.,  
Lindau, R., Rasol, D., Rustemeier, E., Kolokythas,  
K., Marinova, T., Andresen, L., Acquafredda, F., Fra-  
tianni, S., Cheval, S., Klancar, M., Brunetti, M.,  
Gruber, C., Duran, MP., Likso, T., Esteban, P. and  
Brandsma, T., 2012: Benchmarking monthly homo-  
genization algorithms. *Clim ate of the Past* 8:89–115.  
<https://doi.org/10.5194/cp-8-89-2012>

*World Meteorological Organization (WMO)*, 2020: *Guidelines  
on Homogenization*, WMO-No. 1245, p. 63