

Rácsponti adatbázis hatórási adatokra

Szentimrey Tamás
VARIMAX Bt.
szentimrey.t@gmail.com

1. Adatok, eljárások

Állomás adatsorok hőmérsékletre és csapadékösszegre.

Hőmérséklet: órás értékek a 0, 6, 12, 18 időpontokhoz.

Csapadékösszeg: 6 órás összegek a 6, 12, 18, 24 időpontokhoz.

Rácsrendszer: 0,1 fokos rács Magyarországra, 1233 rácspontra.

Az eljárás főbb lépései az állomás adatsorokra:

- Homogenizálás, adatellenőrzés, adatpótlás (MASH)
- Térbeli interpoláció, gridding a rácsrendszerre (MISH)

A MASH és a MISH szoftverek rövid leírása a következő pontokban található. Ezek a rendszerek napi, havi adatsorokra készültek, tehát órás adatokra való alkalmazásuk esetén, bizonyos módosításokkal vagyunk kénytelenek élni, illetve bizonyos fejlesztések is szükségesek.

1.1. Homogenizálás, adatellenőrzés, adatpótlás (MASH)

A MASH rendszert (Multiple Analysis of Series for Homogenization) hosszú havi és napi adatsorok pótlására, ellenőrzésére és homogenizálására dolgoztuk ki (Szentimrey, 1999, 2008). Lényegében ez a többi rendszerünk alapja is, hiszen az éghajlatra vonatkozó információ szempontjából a hosszú adatsorok jelentik a térbeli és időbeli mintát. A legújabb MASHv3.03 verzió (Szentimrey, 2014a) módszertani alapelvei az alábbiak.

Havi adatsorok homogenizálása, ellenőrzése, pótlása:

- A rendszer az ún. relatív homogenitás vizsgálati elv alapján működik. Ennek lényege, hogy egy adott éghajlati elem, különböző megfigyelési állomásokhoz tartozó, azonos időszakra vonatkozó, ugyanazon havi – vagy ugyanazon évszakos, esetleg éves – adatsorait hasonlítjuk össze, és az esetleges ellentmondásokat keressük, vizsgáljuk. Az ezen elv alapján történő homogenizálás célja pontosan a felfedezett ellentmondások megszüntetése, vagyis az ezeket okozó inhomogenitások kiszűrése az adatsorokból.
- Havi adatsorokat homogenizálunk, és a homogenizált évszakos, éves sorokat a homogenizált havi sorokból származtatjuk.
- A rendszer alkalmas additív (pl. hőmérséklet) és kumulatív (pl. csapadék) jellegű éghajlati elemek havi, évszakos, éves adatsorainak homogenizálására.
- Lépésről lépésre működtethető interaktív, fél-automatikus iterációs eljárás.
- Az adattörténeti információkat (ún. meta adatokat) a rendszer képes automatikusan használni.
- A homogenizálás eredménye kiértékelhető, verifikálható. Ez annyit jelent, hogy az eredeti adatsorok és az eredményül kapott sorok inhomogenitásának mértékét, valamint a változtatás mértékét az eljárás során kvantitatíve jellemezzük.

Napi adatsorok homogenizálása, ellenőrzése, pótlása:

- A napi adatok homogenizálása a becsült havi inhomogenitások felhasználásával történik.
- A rendszer automatikus adatellenőrző és adatpótló eljárást is tartalmaz.

1.2. Térbeli interpoláció (MISH)

A MISH rendszert (Meteorological Interpolation based on Surface Homogenized Data Basis) meteorológiai elemek földfelszíni értékeinek interpolációjára dolgoztuk ki (Szentimrey és Bihari, 2007, 2011). Az alábbiakban röviden összefoglaljuk a MISHv1.03 programrendszerrel (Szentimrey és Bihari, 2014b) kapcsolatos legfontosabb tudnivalókat. A rendszer tulajdonképpen két részre bontható, nevezetesen az éghajlati modellező és az interpolációs részre. Az interpolációs rendszer a modellező rendszer eredményeinek felhasználásával működik.

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.
- Lehetséges mind a napi, havi értékek, mind a sokévi átlagok interpolálása.
- Kevés prediktor is elegendő, tekintettel a korábbi modellezésre.
- Becslés az interpolációs hibákra, mégpedig a reprezentativitás értékekre.
- Lehetőség háttérinformáció (pl. műholdas információk, radaradatok, illetve előrejelzések) használatára.
- A programrendszer képes adatsorok rácspontokba való interpolációjára (gridding).

2. Az eredetileg tervezett módosított eljárások lényege

2.1. A MASH rendszer alkalmazása

A kiindulást mind a hőmérséklet, mind a csapadék esetén a négy időponthoz tartozó óraérték adatsorok jelentik. Lényegében a négy időpontra és a két elemre alkalmazzuk a MASH rendszert, ami tehát összesen nyolc MASH vizsgálatot jelentene. Ez esetben a havi sorokat az azonos óraértékekhez tartozó adatsorokból képezzük.

Tekintettel azonban az óraértékek várhatóan kicsi jel-zaj arányára, továbbá az óraértékekhez tartozó eredmények összehangolásának szükségességére, ezért a napi átlagok (összegek) adatsoraira is alkalmazzuk a MASH eljárást.

Terveink szerint ezzel kezdenénk, és az eredményül kapott napi inhomogenitásokat próbálnánk felhasználni az óraérték adatsorok homogenizálására. Amennyiben az inhomogenitásoknál eltekinthetünk a napi menettől, ami a csapadék esetében feltételezhető, akkor ilyen módon jobb minőségű eredményeket kaphatunk. A feltételezés helyességéről, elfogadhatóságáról, a MASH-be beépített verifikációs, hipotézisvizsgálati eljárás alapján dönthetünk. Amennyiben nem fogadható el, ami esetleg a hőmérsékletnél elképzelhető, akkor is hasznosak lehetnek a napi adatsorok vizsgálati eredményei, hiszen például a detektált töréspontok meta adatként automatikusan felhasználhatók az óraérték adatsorok homogenizálásánál.

2.2. A MISH rendszer alkalmazása

Mint a MISH rendszer leírásából látható, a rendszer tulajdonképpen két részre bontható, nevezetesen az éghajlati modellező és az interpolációs részre. Az interpolációs rendszer a modellező rendszer eredményeinek felhasználásával működik. Továbbá ezek a rendszerek napi, havi adatsorokra készültek, tehát órás adatokra való alkalmazásuk esetén, bizonyos fejlesztésekre van szükség. A probléma megint csak az esetleges napi menet kérdéssel kapcsolatos. Amennyiben az eloszlásban eltekinthetünk a napi menettől, ami a csapadék esetében feltételezhető, akkor az eddigi modellezési eredmények felhasználásával, az interpolációs rendszer griddingelő eljárása módosítás nélkül alkalmazható a négy időponthoz tartozó óraérték adatsorokra. Ez a feltételezés azonban a hőmérsékletre valószínűleg nem fogadható el, ezért szükséges a napi menet modellezése is, továbbá a griddingelő algoritmust is fejleszteni kell, hogy képes legyen figyelembe venni a napi menetre vonatkozó modellezési eredményeket.

3. A megvalósított eljárások és eredmények hőmérsékletre

Adatok: órás értékek a 0, 6, 12, 18 időpontokhoz, T00, T06, T12, T18, 1970-2018, 58 állomásra. Tehát viszonylag hosszú adatsorok álltak rendelkezésünkre, azonban igen sok hiánnyal.

A vizsgálatokhoz az 58 állomás napi középhőmérsékleti (T_a) adatsorainak (1971-2018) homogenizálási eredményeit (Szentimrey, Izsák, 2017a, 2017b) kívántuk felhasználni.

Adatok előkészítése: néhány program, az Adatok könyvtárban.

1. Az órás adatsorok MASH-formátumra való konvertálása.

Adatok\Programok\Convert\ConvertH.exe

Input: Data1.dat (adatbázisból)

Output: Data2.dat (MASH-formátum)

2. Az órás adatsorok sorrendjének megváltoztatása a nap közép (T_a) adatsorai szerint.

Adatok\Programok\FilaStatSorrend\FilaStatSor.exe

Input: Data1.dat (órás adatsorok 1970-től), filastat.par (T_a)

Output: Data2.dat

3. Az órás adatsorok 1971-től.

Adatok\Programok\Cut\Cut.exe

Input: Data1.dat (órás adatsorok 1970-től)

Output: Data2.dat (órás adatsorok 1971-től)

Ezek alapján az órás adatsorok MASH inputjainak előállítását, 1970-től, illetve 1971-től: daily.dat, filastat.par

3.1. Homogenizálás a Homogenizalas könyvtárban

Eredeti tervünk (2.1.) szerint a napi középhőmérsékletnél (T_a) detektált inhomogenitásokat próbálnánk felhasználni az óraérték adatsorok T00, T06, T12, T18 homogenizálására.

Az azonos inhomogenitás feltételezésének helyességéről, elfogadhatóságáról, a MASH-be beépített verifikációs, hipotézisvizsgálati eljárás alapján dönthetünk.

3.1.1. Hipotézisvizsgálat az eredeti órás adatsorokra T00, T06, T12, T18, és a napi adatsorokkal (Ta) való azonos inhomogenitásra (1971-től)

A hipotézisvizsgálathoz a MASHv3.03 programrendszert (Szentimrey, 2014a) használtuk. Az eljárást a felhasználói kézikönyv (Manual) 62. oldalán található algoritmus alapján hajtottuk végre.

Input: daily.dat (órás adatok 1971-től), filastat.par (58 állomás)

Lépések:

- 1-4.1
- a Ta havi $m\{j\}$ i.-k bemásolása a Homogenizalas\Ta\Samend könyvtárból
- 4.2 MASHMONTHLY\SAM\SAMMISS\Sammiss.bat;
- 4.3 MASHMONTHLY\SAM\SAMVERI\Samveri.bat
(Output files in directory SAM: $v\{j\}$. ($j=1, \dots, 17$) and verisum)

Az órás adatsorokra vonatkozó eredményeket, a T00\MashTest\Samend, T06\MashTest\Samend, T012\MashTest\Samend, T18\MashTest\Samend alkönyvtárak tartalmazzák. A SAMEND könyvtár tartalmának leírása a felhasználói kézikönyv (Manual) 33. oldalán található.

A számunkra fontos hipotézisvizsgálati eredményeket a $v\{j\}$. ($j=1, \dots, 17$) verifikációs fájlok tartalmazzák. Nevezetesen, az eredeti órás adatsorokra vonatkozó homogenizálás előtti próbastatisztikákat (Test Statistics Before Homogenization (TSB)), illetve a napi inhomogenitások felhasználásával történő homogenizálás után próbastatisztikákat (Test Statistics Before Homogenization (TSA)). Mindkét esetben, és valamennyi órás érték esetén, ezek a próbastatisztikák igen nagyok, több százas nagyságrendűek, míg a kritikus érték csupán 20 körül van.

Tehát levonhatjuk azt a következtetést, hogy

- egyrészt az órás adatsorok igen inhomogének,
- másrészt inhomogenitásaik nem azonosak a napi sorokéval, azaz az inhomogenitásoknál nem tekinthetünk el a napi menettől.

Következésképpen az órás adatsorokat kénytelenek vagyunk külön-külön homogenizálni, a szokásos MASH eljárással.

3.1.2. Meta adatok

Visszatérve az eredeti tervünkhöz (2.1.), akkor is hasznosak lehetnek a napi adatsorok homogenizálási eredményei, ha az órás adatsorokat kénytelenek vagyunk külön-külön homogenizálni. Ugyanis az óraérték adatsorok MASH rendszerrel történő homogenizálásánál, a napi adatsoroknál detektált töréspontok meta adatként automatikusan felhasználhatók.

Meta adat fájlok:

MetaNap.dat: a napi adatsorok homogenizálásánál használt, az állomástörténetek alapján, 1971-től

MetaOra.dat: a MetaNap.dat kiegészítve a napi adatsorok inhomogenitásai alapján, 1971-től

MASH eljárás a MetaOra.dat előállítására, megint a felhasználói kézikönyv 62. oldalán található algoritmus alapján.

Input: daily.dat (napi adatok (Ta) 1971-től), filastat.par, MetaNap.dat

Lépések:

- 1-4.1
 - MASHMONTHLY\SAM\SAMMISS\Sammiss.bat
 - 4.4, 4.5
 - a Ta éves inhomogenitások YEARI, YEAREI bemásolása a Homogenizalas\Ta\Samend könyvtárból a MASHMONTHLY\SAM\MASH könyvtárba, MASHINH.SER, illetve MASHEINH.SER néven
 - MASHMONTHLY\SAM\MASH\Mashcor.bat (válaszok: 1,n,0 vagy 1,0)
 - 4.11
- Output: SAM\MashMeta.dat

Az órás adatsorok homogenizálásánál ezt a MetaOra.dat=MashMeta.dat fájlt használjuk.

3.1.3. Futtatások, eredmények a T00, T06, T12, T18 órás adatsorokra (MASH)

Ezek után mind a négy időpontra homogenizáltuk az órás adatokat (1970-2018) a szokásos MASH eljárással, a 62-63. lévő oldalakon lévő algoritmus szerint, az előbbi MetaOra.dat meta adatok automatikus felhasználásával. Lényegében a napi adatsoroknál detektált törések elegendőnek bizonyultak az órás inhomogenitások becsléséhez. Erre már a meta adatokra vonatkozó próbastatisztikák is utaltak a $v\{j\}$. ($j=1, \dots, 17$) $v\{j\}$. ($j=1, \dots, 17$) verifikációs fájlokban.

Az órás adatsorok homogenizálási eredményeit a T00, T06, T12, T18 könyvtárak MashHomogEred alkönyvtárai tartalmazzák. Ezek az eredmények részben a SAMEND alkönyvtárban található, továbbá az alábbi, szokásos MASH output adatsorok:

Daily.dat: eredeti adatsorok

DailyC.dat: csak pótolts adatsorok

DailyHom.dat: pótolts, homogenizált adatsorok

DailyHomQC.dat: pótolts, homogenizált és ellenőrzött adatsorok

DailyInhom.dat: inhomogenitások sorai

Error.res: becsült adathibák

Ami a jövőt illeti, az órás adatsorok évenkénti frissítése, a napi adatsorokra kidolgozott algoritmus (Szentimrey, Izsák, 2017a; 7. o.) szerint történhet.

3.2. Az órás adatok rácspontra való interpolációja (gridding)

A homogenizálás után, a pótolts, homogenizált és ellenőrzött órás adatsorokat T00, T06, T12, T18 (DailyHomQC.dat) egy 0,1 fokos, Magyarországot lefedő rácsra interpoláltuk (gridding).

Eredeti tervünk (2.2.) szerint a MISHv1.03 (Szentimrey, 2014b) rendszert kívántuk alkalmazni az óraérték adatsorok interpolálására. A MISH interpolációs rendszere a modellező rendszer eredményeinek felhasználásával működik. Azonban ezek a rendszerek napi, havi adatsorokra készültek, tehát órás hőmérsékleti adatokra való alkalmazásuk esetén, bizonyos fejlesztésekre volt szükség. A probléma a napi menet kérdésével kapcsolatos.

3.2.1. Matematikai modell

A napi középhőmérséklet adatokra alkalmazott interpolációs formula:

$$\hat{Z}(s_0) = \sum_{i=1}^M \lambda_i (E(s_0) - E(s_i)) + \sum_{i=1}^M \lambda_i Z(s_i)$$

ahol $Z(\mathbf{s}_0)$ (\mathbf{s} : hely) a prediktandusz, $Z(\mathbf{s}_i)$ ($i = 1, \dots, M$) a prediktorok (megfigyelések),
 $\sum_{i=1}^M \lambda_i = 1$ és a λ_i ($i = 1, \dots, M$) súlytényezők a sztochasztikus kapcsolatoktól függenek,
továbbá $E(\mathbf{s}_i)$ ($i = 0, \dots, M$) a napi térbeli trendértékek.

Ez alapján a $t = 0, 6, 12, 18$ órás értékekre, a súlytényezők elfogadásával, az alábbi interpolációs formulát kellene alkalmazni:

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

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

Az $E(\mathbf{s}, t)$ ($t = 0, 6, 12, 18$) órás térbeli trendértékek modellezésére, az alábbi lineáris modellt választottuk:

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

Ez esetben az órás értékek interpolációs formulája:

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

Tehát a modellezett $E(\mathbf{s}_i)$ ($i = 0, \dots, M$) napi térbeli trendértékek és a becsült $\beta(t)$ ($t = 0, 6, 12, 18$) órás regressziós együtthatók alapján történhet az órás értékek interpolációja.

3.2.2. Modellezési eljárás az órás regressziós paraméterekre

Az $\alpha(t)$, $\beta(t)$ ($t = 0, 6, 12, 18$) regressziós paraméterekre, a napi térbeli trendhez hasonlóan, 12 hónapra, havi becsléseket adtunk. Ehhez mintaként magukat a pótoltt, homogenizált, ellenőrzött órás T00, T06, T12, T18 és napi Ta adatsorokat (DailyHomQC.dat) használtuk fel az 1971-2018-as időszakra, és 58 állomásra.

Első lépésként meghatároztuk a sokéves (1971-2018) havi átlagokat, 12 hónapra, 58 állomásra.

Napi értékekhez tartozó átlagok: $\hat{E}_k(\mathbf{s}_i)$ ($i = 1, \dots, 58; k = 1, \dots, 12$)

Órás értékekhez tartozó átlagok ($t = 0, 6, 12, 18$): $\hat{E}_k(\mathbf{s}_i, t)$ ($i = 1, \dots, 58; k = 1, \dots, 12$)

A második lépés a lineáris regressziós együtthatók becslése, az előbbi minta alapján, ahol a függő változók ($Y = \hat{E}_k(\mathbf{s}_i, t)$) ($t = 0, 6, 12, 18; k = 1, \dots, 12$), míg ($X = \hat{E}_k(\mathbf{s}_i)$) ($k = 1, \dots, 12$) a független változók, és 58-as nagyságú minták állnak rendelkezésre. Ezek alapján a legkisebb négyzetek módszere szerint becsült lineáris regressziós paraméterek:

$$\hat{\alpha}_k(t), \hat{\beta}_k(t) \quad (t = 0, 6, 12, 18; k = 1, \dots, 12)$$

Azaz $\hat{E}_k(\mathbf{s}, t)$ lineáris becslése $\hat{\alpha}_k(t) + \hat{\beta}_k(t) \cdot \hat{E}_k(\mathbf{s})$.

A modellezés végrehajtása, az alábbi programokkal történt.

Gridding\Modellezes\Ta alkönyvtárban

DailyMonthly.exe

Input: Daily.dat (=DailyHomQC.dat, napi adatsorok 1971-2018)

Output: Monthly.dat (havi közép adatsorok, 1971-2018)

MonthlyE12.exe

Input: Monthly.dat (havi közép adatsorok, 1971-2018)

Output: MonthlyE12.res (1971-2018-as sokéves átlagok 12 hónapra, másolni MonthlyE12X.res néven a T00, T06, T12, T18 alkönyvtárakba, független változó)

Gridding\Modellezes\T00, T06, T12, T18 alkönyvtárakban

DailyMonthly.exe

Input: Daily.dat (=DailyHomQC.dat, órás adatsorok 1970-2018)

Output: Monthly.dat (havi közép adatsorok, 1970-2018)

MonthlyE12.exe

Input: Monthly.dat (havi közép adatsorok, 1971(!)-2018)

Output: MonthlyE12.res (1971-2018-as sokéves átlagok 12 hónapra)

LinReg12.exe

Input: MonthlyE12Y.res (=MonthlyE12.res, függő változó),

MonthlyE12X.res (=MonthlyE12.res a Modellezes\Ta alkönyvtárból, független változó)

Output: AlfaBetaOra.par (az α , β regressziós paraméterek 12 hónapra)

LinReg.res (az α , β regressziós paraméterek, korrelációk (corr) és t-próbastatisztikák (tstat) 12 hónapra)

Az eredmények együtt megtalálhatók az EredményekEgyütt alkönyvtárban.

Gridding\Modellezes\EredményekEgyütt

LinReg124Ora.exe

Input: LinReg00.res, LinReg06.res, LinReg12.res, LinReg18.res, a T00, T06, T12, T18 alkönyvtárakban lévő LinReg.res eredményfájlok.

Output: AlfaBeta4Ora.res (az α , β regressziós paraméterek 12 hónapra, és együtt a 4 órára)

CorrTstat4Ora.res (korrelációk (corr) és t-próbastatisztikák (tstat) 12 hónapra, és együtt a 4 órára)

Ezek a modellezési eredmények az alábbi ábrákon láthatók.

	T00		T06		T12		T18	
	alfa	beta	alfa	beta	alfa	beta	alfa	beta
1	-0.97	0.97	-1.52	0.88	2.15	1.06	0.32	1.11
2	-1.26	0.85	-2.08	0.79	2.76	1.08	0.45	1.13
3	-1.38	0.84	-1.61	0.70	2.97	1.13	0.02	1.19
4	-0.38	0.75	-0.69	0.79	2.28	1.17	-0.16	1.15
5	1.27	0.69	0.23	0.85	0.66	1.23	-1.79	1.22
6	1.21	0.73	0.21	0.89	0.22	1.21	-2.78	1.25
7	2.51	0.69	0.57	0.86	1.10	1.16	-3.01	1.24
8	0.97	0.77	0.70	0.82	1.84	1.15	-3.68	1.25
9	0.06	0.81	1.27	0.71	1.67	1.20	-2.87	1.22
10	-0.57	0.84	-0.07	0.70	2.32	1.20	-1.42	1.16
11	-0.46	0.84	-0.80	0.75	2.03	1.17	-0.36	1.10
12	-0.73	0.95	-1.15	0.89	1.90	1.07	0.07	1.08

1. ábra. AlfaBeta4Ora.res, az alfa, beta regressziós paraméterek 12 hónapra, és együtt a 4 órára

	T00		T06		T12		T18	
	corr	tstat	corr	tstat	corr	tstat	corr	tstat
1	0.95	23.02	0.95	22.78	0.89	14.32	0.98	40.44
2	0.89	14.41	0.88	13.64	0.86	12.46	0.98	34.25
3	0.87	12.97	0.84	11.42	0.83	11.23	0.97	31.82
4	0.80	9.87	0.91	16.19	0.86	12.54	0.94	20.80
5	0.78	9.19	0.94	21.08	0.91	15.94	0.94	20.91
6	0.78	9.18	0.96	25.07	0.91	16.08	0.95	22.00
7	0.69	7.19	0.95	22.54	0.88	14.04	0.93	19.70
8	0.73	8.04	0.89	14.53	0.83	11.13	0.93	19.32
9	0.81	10.30	0.83	11.03	0.80	10.15	0.97	28.24
10	0.83	11.10	0.80	10.00	0.78	9.29	0.96	24.57
11	0.90	15.16	0.85	12.02	0.83	11.05	0.98	39.47
12	0.95	21.98	0.94	20.40	0.89	14.49	0.98	40.03

2. ábra. CorrTstat4Ora.res, a korrelációk (corr) és a t-próbastatisztikák (tstat) 12 hónapra, és együtt a 4 órára

Az 1. ábra szerint meglehetősen markáns különbség van a beta (β) regressziós együtthatók között, a T00, T06, T12, T18 óraértékeknél. Tehát a napi menet figyelembevételével célszerű interpolálni. A 2. ábra szerint viszont az óras térbeli trendek (napi menet) jól modellezhetők a napi térbeli trendre vonatkozó lineáris regresszióval.

3.2.3. Gridding eljárás az óras adatok rácspontra való interpolációjára

A modellezésen kívül a griddingelő algoritmust is fejleszteni kellett, hogy képes legyen figyelembe venni a napi menetre vonatkozó modellezési eredményeket.

A 3.2.1. Matematikai modell rész (1) formulája szerint, az óras értékek interpolációs formulája, 12 hónapra,

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

ahol $\beta_k(t)$ ($t = 0, 6, 12, 18; k = 1, \dots, 12$) a modellezett óras regressziós együtthatók és

$E(\mathbf{s}_i)$ ($i = 0, \dots, M$) a modellezett napi térbeli trendértékek. Tehát ez alapján átalakítható a napi értékekre kidolgozott MishInterSerA12.bat griddingelő eljárás, óras értékekre.

Ily módon az óras értékek griddingelése a MISH rendszerrel, szintén a felhasználói kézikönyv (Szentimrey, 2014b) 39. oldal 2.1. pontja szerint történik. Az egyetlen eltérés, hogy a MishInterSerA12.bat fájl helyett a MishInterSerA12O.bat fájl kell használni, a speciális AlfaBetaOra.par inputtal.

Tehát a módosított batch fájl az alábbi:

MishInterSer\MishInterSerA12O.bat

Speciális input, egy adott órára: AlfaBetaOra.par

Eltérés a MishInterSerA12.bat fájlától:

medpotsa1.exe, medpotsa2.exe helyett medora.exe

medpotsa.par helyett medora.par

interpar4a.exe helyett interpar4ora.exe

A MISHv1.03 rendszerbe beírandó fájlok a Homerseklet\Gridding\MishInterSer könyvtárból, az ottani elrendezés szerint másolhatók!

(AlfaBetaOra.par fájl az óráknak megfelelően cserélni kell.)

3.2.4. Futtatások, eredmények a T00, T06, T12, T18 órás adatsorokra (MISH)

Az órás adatsorok rácsponti (gridding) eredményeit a T00, T06, T12, T18 könyvtárak MishGridding\Ered alkönyvtárai tartalmazzák. Ezek az eredmények az alábbi, szokásos MISH gridding output adatsorok.

Az Interpol\MishInterSer\ResultSeries12 alkönyvtár inputjai:
Statisztikai eredmények havonta: gridstat01.res, ..., gridstat12.res
Griddingelt sorok: DailyGrid.ser

4. A megvalósított eljárások és eredmények csapadéokra

Adatok: 6 órás összegek a 6, 12, 18, 24 időpontokhoz, R06, R12, R18, R24, 1997-2018, 89 állomásra. Tehát meglehetősen rövid adatsorok álltak rendelkezésünkre, sőt igen sok hiánnyal.

A vizsgálatokhoz 461 állomás napi csapadékösszeg (RR) adatsorainak (1951-2018) homogenizálási eredményeit (Szentimrey, Izsák, 2017a, 2017b) kívántuk felhasználni. Azonban sajnos a 89 állomás közül csak 55 állomás része a 461 állomásnak, 34 nem. Így csak 55 állomásra tudnánk felhasználni a napi összegek homogenizálási eredményeit.

Adatok előkészítése: néhány program, az Adatok könyvtárban.

1. Az órás adatsorok MASH-formátumra való konvertálása.

Adatok\Programok\Convert\ConvertCs.exe

Input: Data1.dat (adatbázisból)

Output: Data2.dat (MASH-formátum)

2. A koordináta, állomásszám fájl MASH-formátumra való konvertálása.

Adatok\Programok\FilaStat\FilaStat.exe

Input: fila89.txt (adatbázisból)

Output: filastat89.par (MASH-formátum)

3. A 461 állomásból az 55 kiválasztása

Adatok\Programok>Select461>Select461.exe

Input: filastat89.par, filastat461.par

Output: select.par (a 89 állomáshoz tartozó 461-es sorszámok)

4. A 461 állomás havi inhomogenitásainak kiválasztása a 89 állomás számára

Adatok\Programok\Inhom461\MonthlyIH12.exe

Input: $m_{j,i,461}$ ($j=1, \dots, 12$), filastat89.par, select.par

Output: $m_{j,i}$ ($j=1, \dots, 12$) (havi feltételezett inhomogenitások 55 állomásra, 34-nél homogenitás feltételezés)

Ezek alapján az órás adatsorok MASH inputjainak előállítása, 1997-től:
daily.dat, filastat.par

4.1. Homogenizálás a Homogenizalas könyvtárban

Eredeti tervünk (2.1.) szerint a napi csapadékösszegeknél (RR) detektált inhomogenitásokat próbálnánk felhasználni az óras adatsorok R06, R12, R18, R24 homogenizálására.

Az azonos inhomogenitás feltételezésének helyességéről, elfogadhatóságáról, a MASH-be beépített verifikációs, hipotézisvizsgálati eljárás alapján dönthetünk.

4.1.1. Hipotézisvizsgálat az eredeti óras adatsorokra R06, R12, R18, R24, és a napi adatsorokkal (RR) való azonos inhomogenitásra (1997-től)

A hipotézisvizsgálathoz a MASHv3.03 programrendszert (Szentimrey, 2014a) használtuk. Az eljárást a felhasználói kézikönyv (Manual) 62. oldalán található algoritmus alapján hajtottuk végre.

Input: daily.dat (óras adatok 1997-től), filastat.par (89 állomás)

Lépések:

- 1-4.1
- az RR havi $m\{j\}i$ -k bemásolása az Adatok\Programok\Inhom461könyvtárból
- 4.2 MASHMONTHLY\SAM\SAMMISS\Sammiss.bat;
- 4.3 MASHMONTHLY\SAM\SAMVERI\Samveri.bat
(Output files in directory SAM: $v\{j\}$. ($j=1, \dots, 17$) and verisum)

Az óras adatsorokra vonatkozó eredményeket, az R06\MashTest\Samend, R012\MashTest\Samend, R18\MashTest\Samend, R24\MashTest\Samend alkönyvtárak tartalmazzák. A SAMEND könyvtár tartalmának leírása a felhasználói kézikönyv (Manual) 33. oldalán található.

A számunkra fontos hipotézisvizsgálati eredményeket a $v\{j\}$. ($j=1, \dots, 17$) verifikációs fájlok tartalmazzák. Nevezetesen, az eredeti óras adatsorokra vonatkozó homogenizálás előtti próbastatisztikákat (Test Statistics Before Homogenization (TSB)), illetve a napi inhomogenitások felhasználásával történő homogenizálás után próbastatisztikákat (Test Statistics Before Homogenization (TSA)).

Mindkét esetben, és valamennyi óras érték esetén, ezek a próbastatisztikák általában elfogadhatóan kicsik. Vannak nagyobbak is, de ezeket a jelentős adathiány okozza.

Ami a felhasznált napi inhomogenitásokat illeti, azok nem túl jelentősek, az RI2 (Relative Modification of Series) statisztikák szerint mindössze néhány százalékosan változtatják csak az adatsorokat. Ez érthető is, hiszen egyrészt, rövid időszakról (22 év) van szó, másrészt napi inhomogenitás információnk mindössze csak kb. az állomások 60%-ra volt.

Tehát az alábbi következtetéseket vonhatjuk le:

- az adott rövid időszakban az óras, napi adatsorok elég homogének,
- nem érdemes a detektált csekély napi inhomogenitásokat az óras adatsorok homogenizálásához felhasználni.

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

4.1.2. Futtatások, eredmények az R06, R12, R18, R24 óras adatsorokra (MASH)

Ezek után mind a négy időpontra homogenizáltuk az óras adatokat (1997-2018) a MASH eljárással, a 62-63. lévő oldalakon lévő algoritmus szerint.

Az óras adatsorok homogenizálási eredményeit az R06, R12, R18, R24 könyvtárak MashHomogEred alkönyvtárai tartalmazzák. Ezek az eredmények részben a SAMEND alkönyvtárban találhatóak, továbbá az alábbi, szokásos MASH output napi adatsorok:

Daily.dat: eredeti adatsorok

DailyC.dat: csak pótoltt adatsorok

DailyHom.dat: pótoltt, homogenizált adatsorok

DailyHomQC.dat: pótoltt, homogenizált és ellenőrzött adatsorok

DailyInhom.dat: inhomogenitások sorai

Error.res: becsült adathibák

Ami a jövőt illeti, az óras adatsorok évenkénti frissítése, a napi adatsorokra kidolgozott algoritmus (Szentimrey, Izsák, 2017a; 7. o.) szerint történhet.

4.2. Az óras adatok rácspontokba való interpolációja (gridding)

Csapadék esetében nincs okunk változó napi menet feltételezésére, mindemellett egy esetleges modellezéshez, megfelelő mintával sem rendelkezünk. Ezért eredeti tervünk (2.2.) szerint a MISHv1.03 (Szentimrey, 2014b) rendszer napi összegre vonatkozó eljárását (MishInterSerA12.bat) alkalmaztuk az óraérték adatsorok interpolálására.

4.2.1. Futtatások, eredmények a R06, R12, R18, R24 óras adatsorokra (MISH)

Az óras adatsorok rácsponti (gridding) eredményeit a R06, R12, R18, R24 könyvtárak MishGriddingEred alkönyvtárai tartalmazzák. Ezek az eredmények az alábbi, szokásos MISH gridding output adatsorok.

Az Interpol\MishInterSer\ResultSeries12 alkönyvtár inputjai:

Statisztikai eredmények havonta: gridstat01.res, ..., gridstat12.res

Griddingelt sorok: DailyGrid.ser

Hivatkozások

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., Bihari, Z., 2007: „Mathematical background of the spatial interpolation methods and the software MISH (Meteorological Interpolation based on Surface Homogenized Data Basis)”, Proceedings from the Conference on Spatial Interpolation in Climatology and Meteorology, Budapest, Hungary, 2004, COST Action 719, COST Office, 2007, pp. 17-27

Szentimrey, T., 2008: "Development of MASH homogenization procedure for daily data", Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases, Budapest, Hungary, 2006; WCDMP-No. 71, WMO/TD-NO. 1493, 2008, pp. 123-130.

Szentimrey, T., Bihari, Z., Lakatos, M., 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, 1-11

Szentimrey, T., 2014a: “Manual of homogenization software MASHv3.03”, Hungarian Meteorological Service, p. 69.

Szentimrey, T., Bihari, Z., 2014b: Manual of interpolation software MISHv1.03, Hungarian Meteorological Service, p. 60

Szentimrey, T., Izsák, B., 2017a: Eljárások különböző hosszú napi adatsorok homogenizálására, griddingelésére és évenkénti frissítésére. OMSZ tanulmány.

Szentimrey, T., Izsák, B., 2017b: Homogenizált, griddingelt magyarországi adatsorok. OMSZ tanulmány.