

A parlagfűpollen koncentrációjának időbeli és térbeli explicit, adatvezérelt becslése Európára

Makra, L., Matyasovszky, I., Tusnády, G., Ziska, L.H., Hess, J.J., Nyúl, L.G., Chapman, D.S., Coviello, L., Gobbi, A., Jurman, G., Furlanello, C., Brunato, M., Damialis, A., Charalampopoulos, A., Müller-Schärer, H., Schneider, N., Szabó, B., Sümeghy, Z., Páldy, A., Magyar, D., Bergmann, K-C., Deák, Á.J., Mikó, E., Thibaudon, M., Oliver, G., Albertini, R., Bonini, M., Šikoparija, B., Radišić, P., Mitrović Josipović, M., Gehrig, R., Severova, E., Shalaboda, V., Stjepanović, B., Ianovici, N., Berger, U., Kofol Seliger, A., Rybníček, O., Myszkowska, D., Dąbrowska-Zapart, K., Majkowska-Wojciechowska, B., Weryszko-Chmielewska, E., Grewling, Ł., Rapiejko, P., Malkiewicz, M., Šaulienė, I., Prykhodo, O., Maleeva, A., Rodinkova, V., Palamarchuk, O., Ščevková, J., Bullock, J.M., 2023: **A temporally and spatially explicit, data-driven estimation of airborne ragweed pollen concentrations across Europe. Science of the Total Environment**, 905, 167095, pp. 1-16. doi: 10.1016/j.scitotenv.2023.167095

Makra László

emeritus professzor

Szegedi Tudományegyetem,
Mezőgazdasági Kar,
Gazdálkodási és Vidékfejlesztési Intézet
Hódmezővásárhely
2024. április 12.



"Aki pedig versenyben vesz részt, mindenben önmegtartóztató: azok azért, hogy elhervadó koszorút nyerjenek, mi pedig azért, hogy hervadhatatlant. Én tehát úgy futok, mint aki előtt nem bizonytalan a cél, úgy öklözök, mint aki nem a levegőbe vág, hanem megsanyargatom és szolgává teszem a testemet, hogy amíg másoknak prédikálok, magam ne legyek alkalmatlanná a küzdelemre."

(Pál első levele a korinthusiakhoz, 9. fejezet, 24-27)

Motiváció: pollenérzékenység

Tartalomjegyzék

1. Célkitűzés
2. A téma fontossága
3. Globális felmelegedés
4. Állomások és a vizsgált időszak
5. Módszertan
 - 5.1 A hiányzó napi parlagfűpollen adatok helyreállítása
 - 5.1.1 Adat-helyreállítás a gaussi módszerrel (GM)
 - 5.1.2 Adat-helyreállítás mesterséges intelligenciával (DL)
 - 5.1.3 Térképészeti háttér
 - 5.1.4 A térképek pontosságának értékelése az eredeti és a helyreállított adatkészletekkel
6. Állomáshálózat, valamint a parlagfűpollen kvantitatív és kvalitatív jellemzőinek térképei Európára
7. A GM és DL modell adatrekonstrukciója jóságának az összehasonlítása
8. Összegzés, az eredmények jelentősége

1. Célkitűzés

Az aeroallergéneket termelő növényfajok jelenlegi és jövőbeni éghajlatváltozás okozta terjeszkedése komoly humán egészségi problémát jelent Európa-szerte és másutt is (Ziska et al., 2019).

Sürgősen szükség van pontos, időbeli dinamikus térképek elkészítésére kontinentális szinten, különösen a változó éghajlattal összefüggésben.

Ennek a tanulmánynak az volt a célja, hogy

- a) helyreállítsa a hiányzó napi parlagfű pollen adatkészleteket Európára, továbbá
- b) elkészítse a parlagfűpollen kvantitatív és fenológiai térképeit, ami az eddigi legteljesebb és legrészletesebb, nagy felbontású parlagfűpollen jellemzők térképeket eredményezte Európára.

2. A téma fontossága:

— egészségügy

— mezőgazdaság

— turisztika

— természetvédelem

- **Egy** parlagfűpollen-érzékeny szénanáthás **beteg szezonális gyógyszerköltsége kb. 30.000 Ft** (Harsányi, 2009); Teljes (kumulatív) infláció 2022-ig: 65,35%, azaz **a mai költség: 49.600 Ft**;
- A pollenallergiában szenvedők **évi teljes gyógyszerköltsége kb. 27 milliárd Ft** (Mányoki et al., 2011); Teljes (kumulatív) infláció 2022-ig: 51,27%, azaz **a mai költség: 40.8 md Ft**;
- A pollen-allergiában szenvedő betegek **rendelőintézeti, illetve kórházi kezelési költségei** ⇒ további **évi 16-20 milliárd Ft** (Basky, 2009); Teljes (kumulatív) infláció 2022-ig: 65,35%, azaz **a mai költség: 26,5-33,1 md Ft**;
- **A parlagfű tenyészésével járó terméskiesés** ⇒ további **évi kb. 70-90 milliárd Ft** árbevétel-kiesés (Kazinczy et al., 2009; Basky, 2009); Teljes (kumulatív) infláció 2022-ig: 65,35%, azaz **a mai költség: 115,8-148,8 md Ft**;
- Egyéb közvetlen és közvetett hatások: **a** táppénzen töltött napok száma, **turisztikai,** és **természetvédelmi károk,** **a** parlagfű-magvakkal szennyezett vetőmag, stb.
- ⇒ **AZ ÖSSZES OKOZOTT KÁR: évi kb. 120-200 milliárd Ft** (Mányoki et al., 2011); Teljes (kumulatív) infláció 2022-ig: 51,27%, azaz **a mai költség: 181,5-302,5 md Ft**; [TELJES GDP, 2022: 66.075 md Ft (168.9 md EUR)]
- **AZ ÖSSZES OKOZOTT KÁR: Magyarország teljes évi GDP-je 0.4%-a!**
- ⇒ **a parlagfű Magyarországon alapvető természeti-, gazdasági-, illetve humán- és környezet-egészségügyi probléma.**

3. Globális felmelegedés

A globális felmelegedéssel (a) nő a pollenkoncentráció, (b) hosszabbodik a pollenszezon, (c) az allergén taxonok élőhelyei a magasabb szélességek felé kiterjednek \Rightarrow egyre több ember válik érintetté \Rightarrow globálisan nő az allergiás megbetegedések száma \Rightarrow nő a globális közegészségügyi kockázat (Ziska et al., 2019);

A parlagfű pollenjével szembeni érzékenység több mint kétszeresére nő Európában - 33 millió főről (1986-2005) 77 millió főre (2041-2060). Az érzékenység a meglévő parlagfű problémával rendelkező országokban csekély mértékben nő (pl. Magyarországon 27%-kal, Horvátországban 26%-kal, és Szerbiában 21%-kal). Azonban a legnagyobb növekedés ott fog bekövetkezni, ahol parlagfűpollen érzékenység ma még alig, vagy egyáltalán nem tapasztalható (példák a növekedésre: Németország, 235%, Lengyelország, 292%, és Franciaország 231% (Lake et al., 2018).

4. Állomások és a vizsgált időszak

Európa: 625 aerobiológiai állomás méri a napi parlagfűpollen koncentrációt;

Nyugat-Európa: a legtöbb állomás csak az 1980-es évek óta működik;

Ukrajna és Délnyugat-Oroszország: az összes állomás 2010-től vagy kevéssel azt megelőzően kezdte működését;

A kiválasztott közös vizsgált időszak: július 15 – október 15

Csak azon állomásokat választottuk ki, amelyekre a vizsgált 1995-2010 közötti 16 év július 15 – október 15 közötti időszakának legfeljebb 25%-ában (legfeljebb 4 évben) volt az adatlefedettség kisebb, mint 40%.

⇒ A 625 aerobiológiai állomás közül 67 állomás 1995-2010 közötti 16 éves adatsorait választottuk ki;

⇒ Hasonlóan, 162 állomás évi adatsorait választottuk ki a 2010. évből (a legteljesebb adatsorokat tartalmazó év);

5. Módszertan

5.1 A hiányzó napi parlagfűpollen adatok helyreállítása

5.1.1 Adat-helyreállítás a gaussi módszerrel (GM)

Ahhoz, hogy megbecsüljük az átlagos évi pollenkoncentrációt (*MATPC*), az egyedi *ATPC*-ket becsültük minden állomásra és minden évre külön-külön. Ahol elég adat állt rendelkezésre megfelelő időbeli eloszlásban, ott a becslést egy lépésben végeztük el. Egyébként egy további lépést, térbeli interpolációt hajtottunk végre.

Ezután minden egyes állomásra átlagoltuk az évi *ATPC*-ket, s megkaptuk az átlagos évi *ATPC*-ket (*MATPC*).

Az első lépésben gaussi görbét illesztettünk minden egyes állomás minden egyes évének napi pollenkoncentráció adataira, a következő módon.

Legyen c_i az év i -edik napjának a pollenkoncentrációja. Ezt Cp_i -vel becsüljük, ahol C jelöli az *ATPC*-t, és

$$p_i = \Phi\left(\frac{(i + 1/2) - m}{\sigma}\right) - \Phi\left(\frac{(i - 1/2) - m}{\sigma}\right)$$

a Φ standard normális eloszlásfüggvény által definiált. A feladat a C , m és σ paraméterek meghatározása.

Adatrekonstrukció egy lépésben: Ha egy adott állomáson és egy adott évben a napi adathiány <10%, akkor a hiányzó napok pollenkoncentrációit, majd az éves *ATPC*-ket a pollenszezon napjaira illesztett Gauss-görbével becsüljük.

Adatrekonstrukció két lépésben: Ha egy adott állomáson és egy adott évben a napi adathiány $\geq 20\%$, akkor az éves *ATPC*-ket az adott állomáshoz legközelebb eső hat állomás éves *ATPC*-inek interpolációjával becsüljük, amelyekre az éves *ATPC*-ket (1) a július 15 – október 15 közötti időszakra vonatkozó hiánytalan napi adatokból számítottuk ki, illetve (2) egylépéses adatvisszaállítás alapján becsültük meg.

5.1.2 Adat-helyreállítás mesterséges intelligenciával (AI)

A napi parlagfűpollen koncentráció adatok rekonstrukciójának másik módja a mély neurális hálózatok (**deep neural networks**) alkalmazó AI megközelítés. **A mély tanulási (DL)** stratégia egy zajtalanító konvolúciós automatikus kódolón (Denoising Convolutional Auto-encoder) (*ConvAE*) keresztül valósul meg. A *ConvAE* csak konvolúciós rétegekből áll, amelyek célja az idősorok időbeli függőségének modellezése és a bemeneti adatok rekonstrukciója.

Figyelembe véve a jelentős adathiányt, a konvolúciós neurális hálózatok könnyebben betaníthatók, mint a bonyolultabb visszatérő architektúrák, pl. a Long-Short Term Memory components (*LSTM*) vagy a Gated Recurrent Units (*GRU*) (Zhang et al., 2019; Zhao et al., 2020).

A *ConvAE* architektúra bemeneti paraméterei: (1) a június 1. – október 31. közötti napi pollenkoncentrációk, (2) a maximum hőmérséklet, (3) minimum hőmérséklet, (4) átlaghőmérséklet és a (5) csapadék napi értékei.

A modell képzése: a napi bemeneti koncentrációk véletlenszerű eltávolításával és az eredeti idősor rekonstrukciójával történt.

A modelleket különböző adatsérülési sémákkal, beviteli jellemzőkkel, veszteségi függvényekkel és egyéb hiperparaméterekkel képeztük.

A legjobb konfigurációt ötszörös keresztellenőrzéssel (5-fold cross-validation) választottuk ki (Kohavi, 1995).

A dolgozatban nyolc különböző modellkonfiguráció eredményeit közöljük:

(1) időjárásinformációkkal és anélkül, (2) eredeti, vagy normalizált veszteségfüggvénnyel és (3) egyedi adathiányos napokkal, vagy egymást követő adathiányos napokkal (ablakokkal) működő perturbációs sémával.

A modellek teljesítményét két értékelési beállítás esetére mutatjuk be: (1) az egyedi adathiányos napok (azaz pontok) és (2) az egymást követő adathiányos napok (azaz ablakok) adatvisszaállítására.

Az általánosított modellek közül kettőt (időjárásinformációkkal és anélkül) választottunk ki, hogy mindkét beállításhoz tartozó napi pollenkoncentráció adatokat a keresztellenőrzés (cross-validation) eredményei alapján rekonstruáljuk.

A két eredményül kapott modellt ezután összehasonlítottuk a “teszt” és a “teszt103” adatsorokkal, hogy értékeljük viselkedésüket. A “teszt” adatkészletet a 67 állomásból, míg a “teszt103” adatkészletet 103 olyan állomás kiválasztásával állítottuk elő, amelyek a 162 állomásos adatkészlethez tartoznak, de nem szerepelnek a 67 állomásos adatkészletben.

A legjobb *DL*-modell eredményét ezután összehasonlítottuk az egylépéses és kétlépéses *ATPC* helyreállítási modellekkel végrehajtott adatrekonstrukciós eljárás eredményével.

A modell teljesítményét négy metrika segítségével értékeltük:

- (1) átlagos négyzetes hiba négyzetgyökének átlaga (*aRMSE*),
- (2) a normalizált átlagos négyzetes hiba négyzetgyökének átlaga (*aNRMSE*),
- (3) a determinációs együttható (R^2) és a
- (4) Pearson-korrelációs együttható (*PCC*).

5.1.3 Térképészeti háttér

A térképezéshez az *ETRS89 / LAEA Europe* koordináta-referenciarendszert (*CRS*) vettük alapul, amely alkalmazható Európára (Annoni et al., 2003).

Az *ETRS89 / LAEA Europe* az *ETRS89* földrajzi *2D CRS*-t használja alap *CRS*-ként, melynek porjekciója az Europe Equal Area 2001 (Lambert Azimuthal Equal Area).

Az interpolációs helyek magassági adatait az *ETOPO1* Globális Relief Modellből nyertük ki (Amante & Eakins, 2009).

A térképeket az *R* szoftverrel (3.1.2-es verzió) hoztuk létre (*R* Core Team, 2014).

A térképek két lépésben készültek:

- 1) az adatpontok felhasználásával egy **nagy felbontású raszteres térképet számítottunk ki az *spTimer* csomagban található Bayes Gaussian Process Modeling segítségével (Bakar & Sahu, 2015).**
- 2) **a modellt az *spt.Gibbs* funkcióval illesztettük a csomag alapértelmezett beállításával („GP” modell, „geodéziai:km” távolság módszer).**

Az eredmények jobb értelmezésének elősegítése érdekében **a helyreállított értékeket 3x3-as átlagoló szűrővel simítottuk.**

A diagramokat a *ggplot* csomag segítségével állítottuk elő (Kahle & Wickham, 2013).

A pollenkoncentráció térképekhez a mért adatokat az interpoláció előtt az általánosított **log transzformációval** leképeztük, és az interpolált eredményeket az inverz log transzformáció segítségével visszaképeztük az eredeti skálára.

Az extrapoláció miatt előálló hamis információk kiszűrésére a térképeket több szempontból is finomítottuk:

- 1) az egyes állomások közötti maximális távolság 150 km (Skjøth et al., 2012; Smith et al., 2013),
- 2) az egyes állomások 150 km-es körzetén belül az adott állomás tszf. magasságánál +/-100 m-en kívüli magasságú területeket nem térképeztük. E területek eltávolítása lehetővé teszi
 - (a) a pollenszezon nagyon késői kezdési és befejezési időpontjainak előrejelzésében felmerülő bizonytalanságok kezelését, és
 - (b) egy külön színskála kialakítását Kelet- és Nyugat-Európa jobb megkülönböztetésére.
- 3) **A szezonális vizsgálata** jelentős évi összes pollenkoncentrációt igényel. Ennek érdekében szezonális paraméter térképeket készítettünk csak azon aerobiológiai állomások felhasználásával, ahol az átlagos évi pollenkoncentráció elérte a legalább $100 \text{ pollen szem} \cdot \text{m}^{-3}$ értéket. Ez a küszöb a (2) pontban említett nagy magassági problémára is vonatkozik.

5.1.4 A térképek pontosságának értékelése az eredeti és a helyreállított adatkészletekkel

A *GM* és *DL* modellek célja: egy pontosabb *ATPC* érték becslése, az előállított térképek pontosságának javítására.

Összehasonlítottuk a két modell rekonstrukciós képességeit, figyelembe véve a „teszt103” adatkészlet 25 állomását. Ezekon az állomásokon nem volt hiányzó adat a 2010-es szezonban, ezért az *ATPC*-jük helyesnek tekinthető.

A 25 állomás idősorából különböző százalékos adatokat (10%-tól 80%-ig) távolítottunk el, és megkértük a modellt, hogy rekonstruálja a valódi *ATPC*-t.

Két forgatókönyvet értékeltünk:

(1) felmértük a modell pontosságát az *ATPC* értékek rekonstruálásakor az R^2 , *RMSE* és a Mean Absolute Error (*MAE*) segítségével.

(2) összehasonlítottuk az átlagos abszolút százalékos hibát (*MAPE*)

(a) a valós és perturbált adatokkal generált térképek között, s a

(b) a valós és rekonstruált adatokkal generált térképek között.

Mindkét kísérletet 10-szer megismételtük, és az eredményeket átlagoltuk. Megjegyezzük, hogy mivel az *SI* módszer (térbeli interpoláció) csak a legközelebbi 6 állomás *ATPC*-jét veszi figyelembe, egyetlen kísérletet végeztünk, ellentétben a *GM* és *DL* esetében végzett 10 kísérlettel.

6. Állomáshálózat, valamint a parlagfűpollen kvantitatív és kvalitatív jellemzőinek térképei Európára

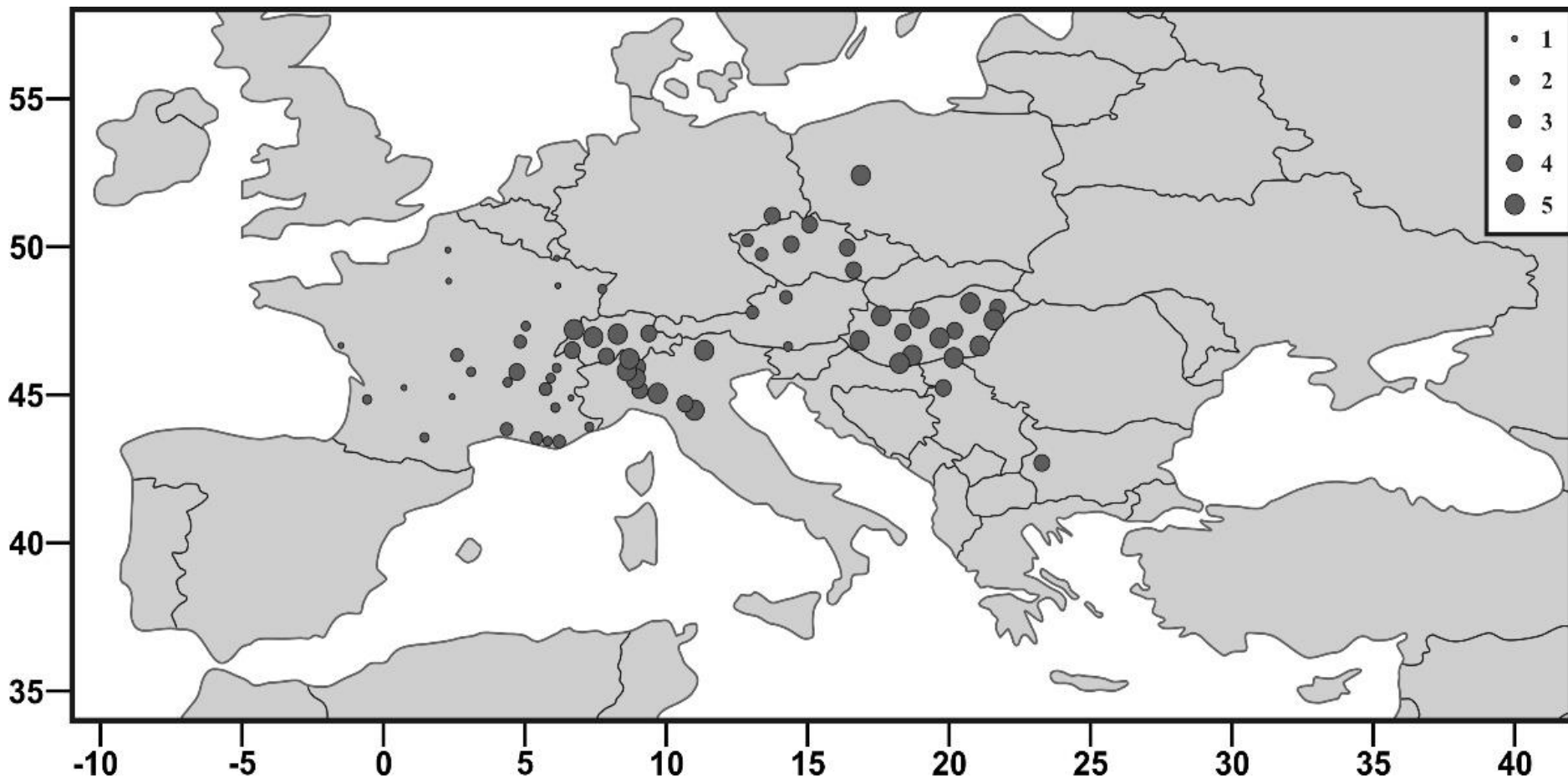
1. táblázat

x (%) adatlefedettséggel rendelkező állomások száma a vizsgált időszakokban¹

*adat- lefedettség, x (%)	*az adatlefedettség grafikus megjelenítése		időszak			
			pollenszezon, 1995-2010		pollenszezon, 2010	
			állomások száma	%	állomások száma	%
	sorszám	méret				
0<x≤ 20	1	.	8	11,9	27	16,7
20<x≤ 40	2	•	13	19,4	17	10,5
40<x≤ 60	3	••	9	13,4	22	13,6
60<x≤ 80	4	•••	17	25,4	22	13,6
80<x≤100	5	••••	20	29,9	74	45,7
Összes	–		67	100,0	162	100,0

*Az állomások adatlefedettsége arányos a fekete pöttyök méretével (lásd: 1a. és 1b. ábrák).

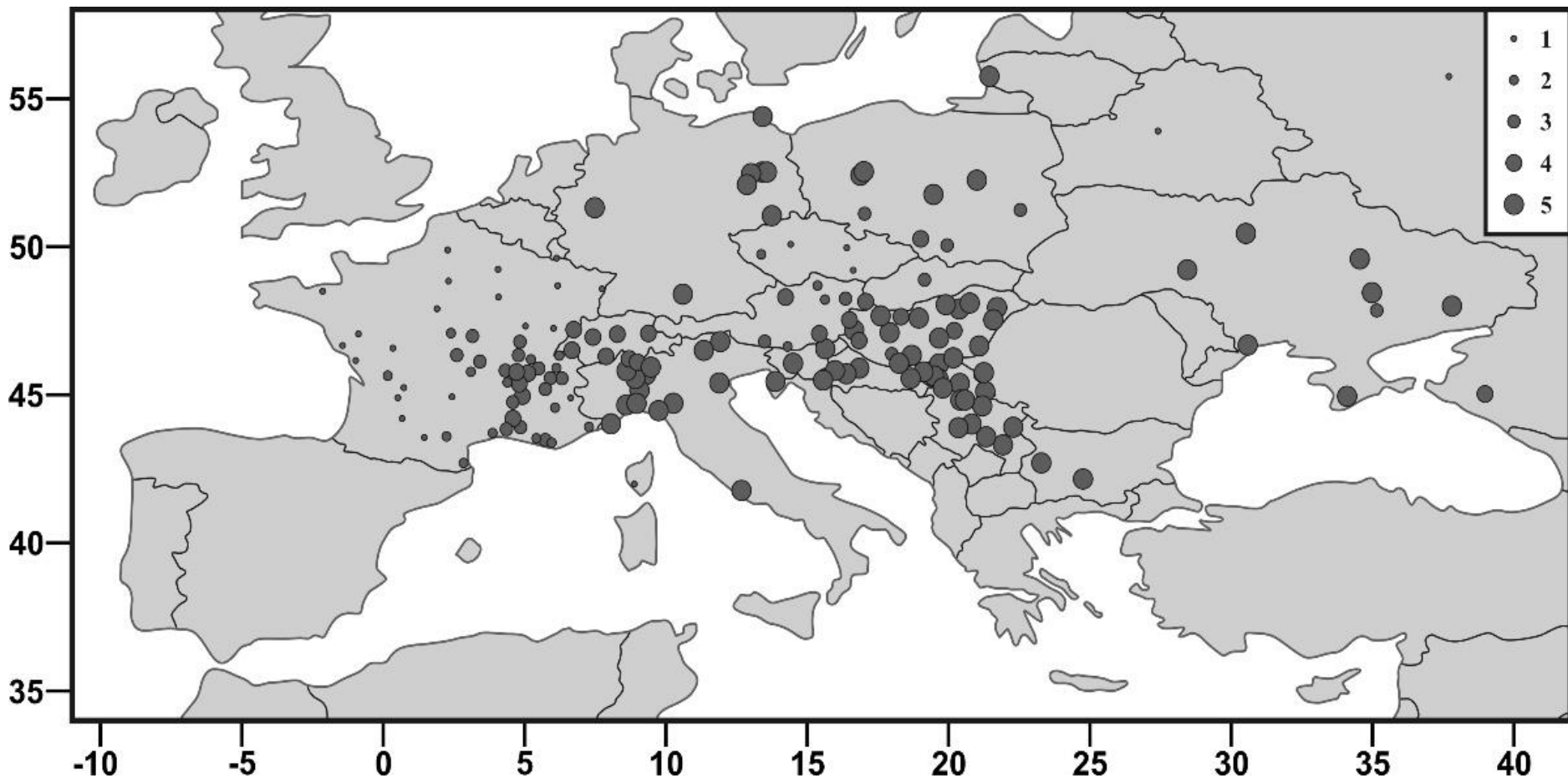
¹Megjegyzés: a táblázat az állomásspecifikus pollenszezonokra vonatkozik, és nem a minden állomásra egységesen július 15-től október 15-ig terjedő kiválasztott vizsgálati időszakra. Mivel a vizsgálati időszak sok állomás esetében nem fedi le a teljes pollenszezont, ezért a dolgozatban meghatározott állomásválasztási kritériumok nem mondanak ellent az 1. táblázat eredményeinek, mely szerint az 1995-2010 közötti évek több mint 25%-ának (valójában az évek 31,3%-ának) az adatlefedettsége kevesebb, mint 40%.



1a. ábra

A vizsgált 67 aerobiológiai állomás földrajzi pozíciói, melyekről rendelkezésre állnak 16 év (1995-2010) parlagfűpollen koncentrációinak átlagos évi összegei. Az aerobiológiai állomások földrajzi pozícióit fekete pöttyök jelzik. Vízszintes tengely: hosszúsági fok; függőleges tengely: szélességi fok. Az állomások adatlefedettsége (x, %) arányos a fekete pöttyök méretével (lásd: 1. táblázat).

Jelmagyarázat (1. táblázat): 1: $0 < x \leq 20$; 2: $20 < x \leq 40$; 3: $40 < x \leq 60$; 4: $60 < x \leq 80$; 5: $80 < x \leq 100$.

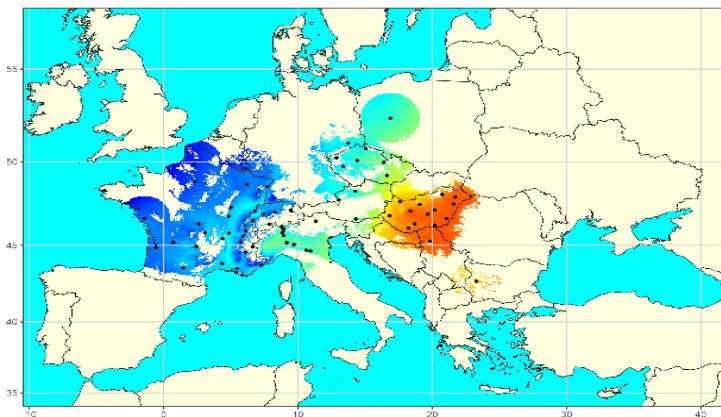
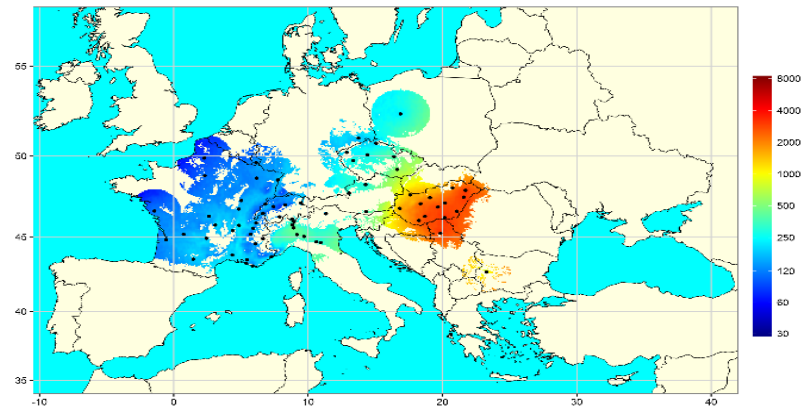
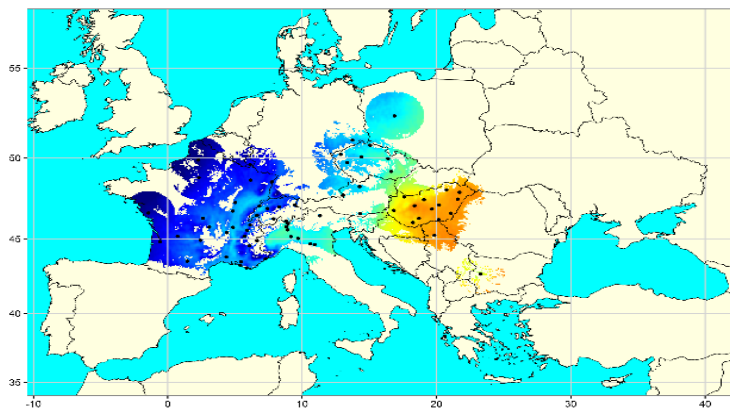


1b. ábra

A vizsgált 162 aerobiológiai állomás földrajzi pozíciói, melyekről rendelkezésre állnak a 2010. év parlagfűpollen koncentrációinak évi összegei. Az aerobiológiai állomások földrajzi pozícióit fekete pöttyök jelzik. Vízszintes tengely: hosszúsági fok; függőleges tengely: szélességi fok.

Az állomások adatlefedettsége (x, %) arányos a fekete pöttyök méretével (lásd: 1. táblázat).

Jelmagyarázat (1. táblázat): 1: $0 < x \leq 20$; 2: $20 < x \leq 40$; 3: $40 < x \leq 60$; 4: $60 < x \leq 80$; 5: $80 < x \leq 100$.



Az eddigi legpontosabb térképek a nemzetközi szakirodalomban:

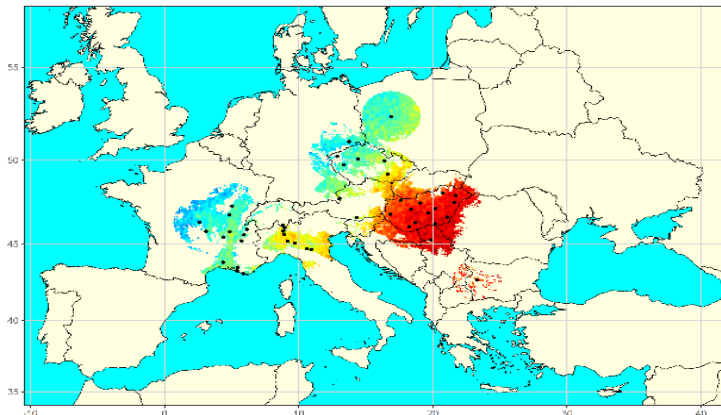
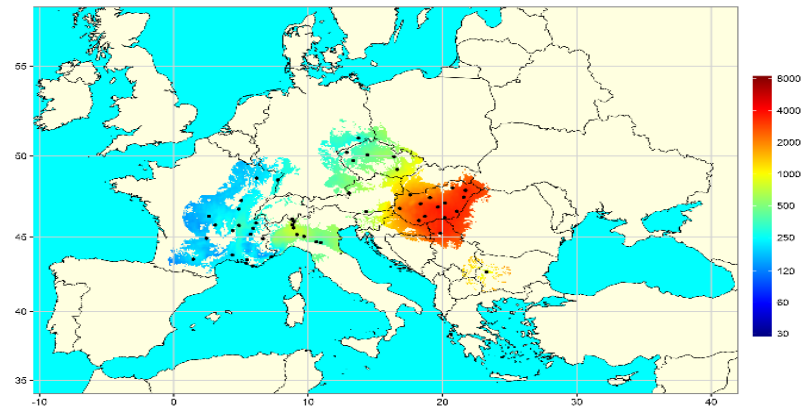
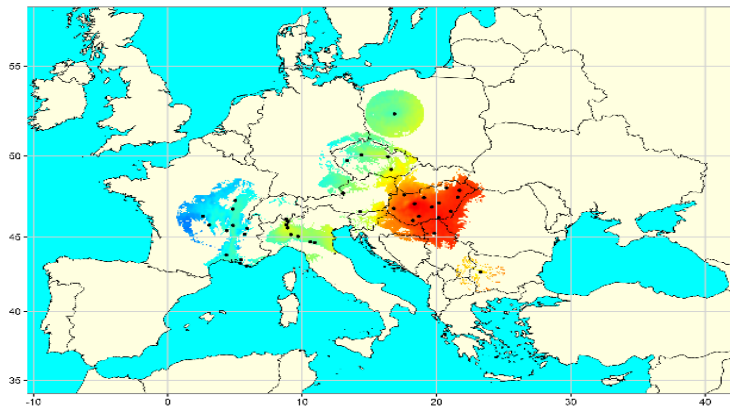
(b) helyreállított(GM) adatokkal
(fent jobbra), és

(c) helyreállított(DL) adatokkal
(lent)

2. ábra

16 éves évi átlagos parlagfűpollen koncentráció Európára 10-es alapú logaritmus színskálán, a vizsgált 67 aerobiológiai állomás földrajzi pozícióival az (a) eredeti adatok (fent, balra), (b) helyreállított(GM) adatok (fent jobbra), és (c) helyreállított(DL) adatok (lent) felhasználásával, 1995-2010, július 15 – október 15, (pollenszem · m⁻³ levegő). A térképek (1) tartalmazzák az összes állomást, (2) az egyes állomások 150 km-es körzetén belül az adott állomás tszf. magasságánál +/-100 m-en kívüli magasságú területeket nem térképeztük, és (3) az állomások közötti maximális távolság 150 km. Vízszintes tengely: hosszúsági fok; függőleges tengely: szélességi fok. Az aerobiológiai állomások földrajzi pozícióit fekete pontok jelzik.

Helyreállított(GM) adatok: a hiányzó napi adatok helyreállítása annak a feltételezésével történt, hogy a napi pollenkoncentrációk szezonális eloszlása normális (Gaussi módszer). **Helyreállított(DL) adatok:** a hiányzó napi adatok helyreállítása konvolúciós auto-encoder modellek felhasználásával és időjárás adatok¹ figyelembe vételével történt (ConvAE + Weather). [Időjárás adatok¹: a maximum hőmérséklet, a minimum hőmérséklet és a középhőmérséklet, valamint a csapadék napi adatai.]



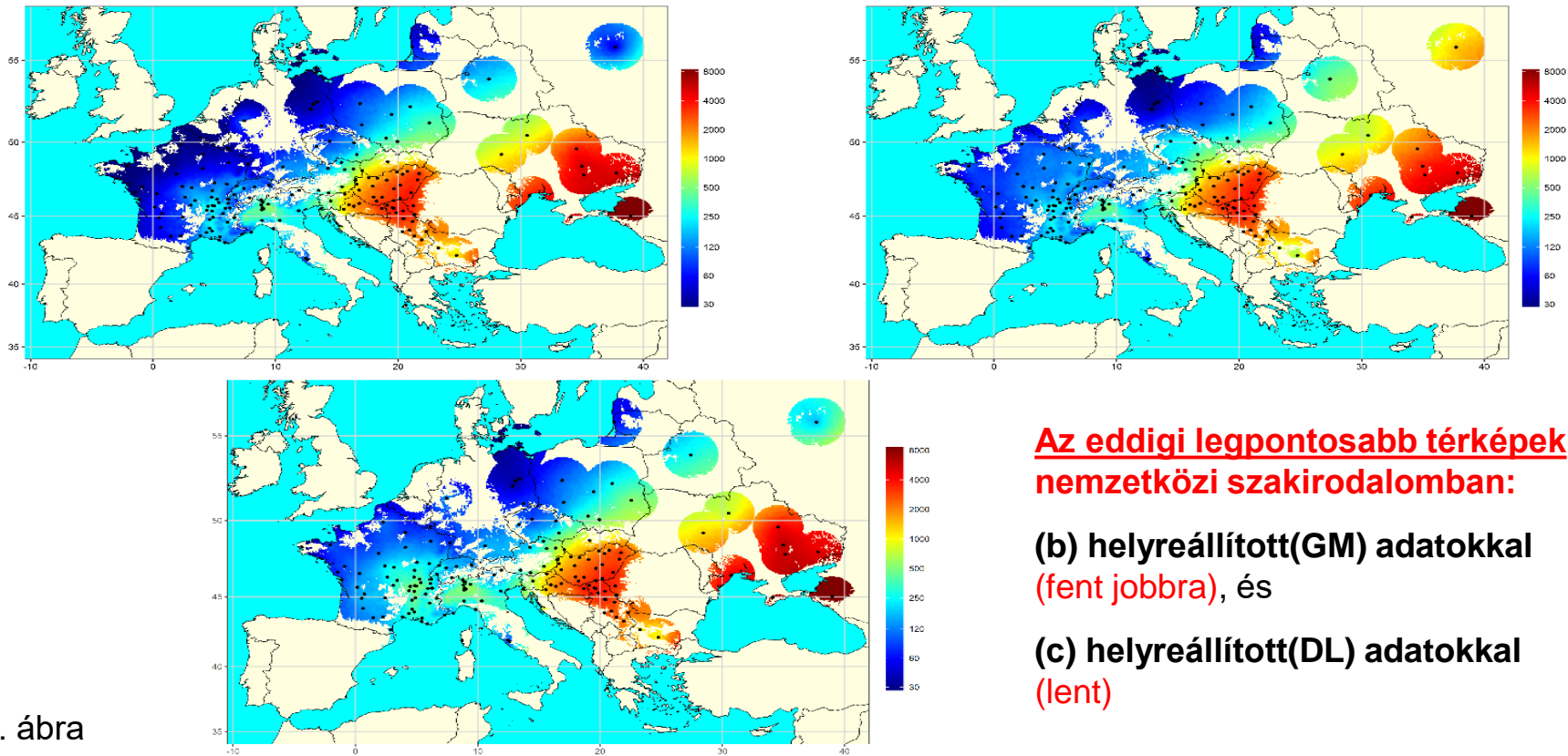
Első térképek a nemzetközi szakirodalomban:

(b) helyreállított(GM) adatokkal (fent jobbra), és

(c) helyreállított(DL) adatokkal (lent)

2'. ábra

16 éves évi átlagos parlagfűpollen koncentráció Európára 10-es alapú logaritmus színskálán, a vizsgált aerobiológiai állomások földrajzi pozícióival az (a') eredeti adatok (fent balra), (b') helyreállított(GM) adatok (fent jobbra), és (c') helyreállított(DL) adatok (lent) felhasználásával, 1995-2010, július 15 – október 15, (pollenszem · m⁻³ levegő). A (1) 100 pollenszem · m⁻³ levegőnél kisebb évi átlagos pollenkoncentrációval rendelkező állomásokat kizártuk a térképezésből, (2) az egyes állomások 150 km-es körzetén belül az adott állomás tszf. magasságánál +/-100 m-en kívüli magasságú területeket nem térképeztük, és (3) az állomások közötti maximális távolság 150 km. Vízszintes tengely: hosszúsági fok; függőleges tengely: szélességi fok. Az aerobiológiai állomások földrajzi pozícióit fekete pontok jelzik. **Helyreállított(GM) adatok:** a hiányzó napi adatok helyreállítása annak a feltételezésével történt, hogy a napi pollenkoncentrációk szezonális eloszlása normális (Gaussi módszer). **Helyreállított(DL) adatok:** a hiányzó napi adatok helyreállítása konvolúciós auto-encoder modellek felhasználásával és időjárás adatok¹ figyelembe vételével történt (ConvAE + Weather). [Időjárás adatok¹: a maximum hőmérséklet, a minimum hőmérséklet és a középhőmérséklet, valamint a csapadék napi adatai.]



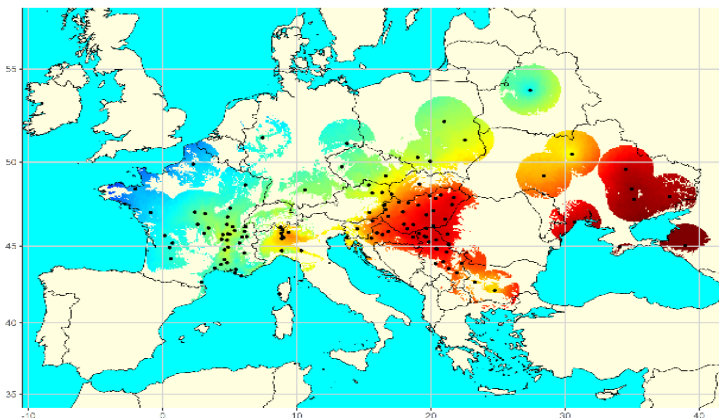
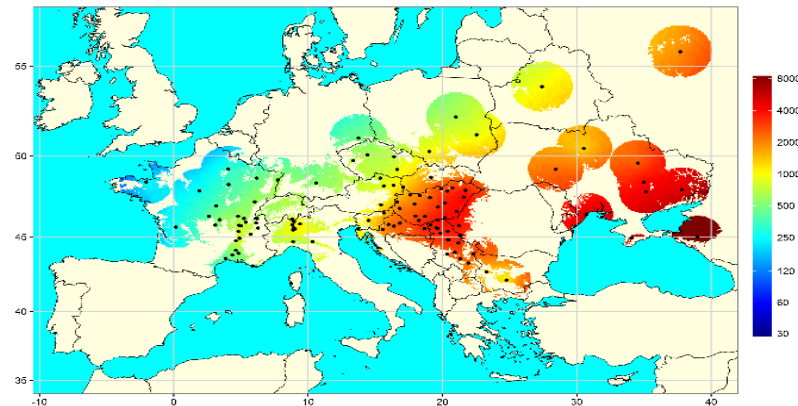
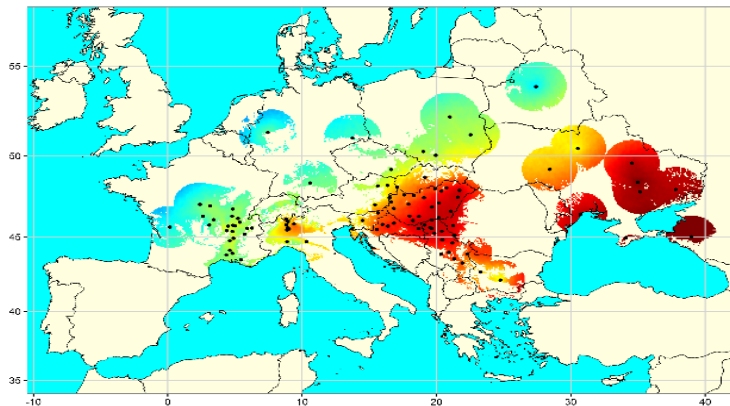
3. ábra

Az eddigi legpontosabb térképek a nemzetközi szakirodalomban:

(b) helyreállított(GM) adatokkal (fent jobbra), és

(c) helyreállított(DL) adatokkal (lent)

Az évi összes parlagfűpollen koncentráció Európára 10-es alapú logaritmus színskálán, a vizsgált 162 aerobiológiai állomás földrajzi pozícióival az (a) eredeti adatok (fent balra), (b) helyreállított(GM) adatok (fent jobbra), és (c) helyreállított(DL) adatok (lent) felhasználásával, 2010 július 15 – október 15, (pollenszem · m⁻³ levegő). A térképek (1) tartalmazzák az összes állomást, (2) az egyes állomások 150 km-es körzetén belül az adott állomás tszf. magasságánál +/-100 m-en kívüli magasságú területeket nem térképeztük, és (3) az állomások közötti maximális távolság 150 km. Vízszintes tengely: hosszúsági fok; függőleges tengely: szélességi fok. Az aerobiológiai állomások földrajzi pozícióit fekete pontok jelzik. **Helyreállított(GM) adatok:** a hiányzó napi adatok helyreállítása annak a feltételezésével történt, hogy a napi pollenkoncentrációk szezonális eloszlása normális (Gaussi módszer). **Helyreállított(DL) adatok:** a hiányzó napi adatok helyreállítása konvolúciós auto-encoder modellek felhasználásával és időjárás adatok¹ figyelembe vételével történt (ConvAE + Weather). [Időjárás adatok¹: a maximum hőmérséklet, a minimum hőmérséklet és a középhőmérséklet, valamint a csapadék napi adatai.]



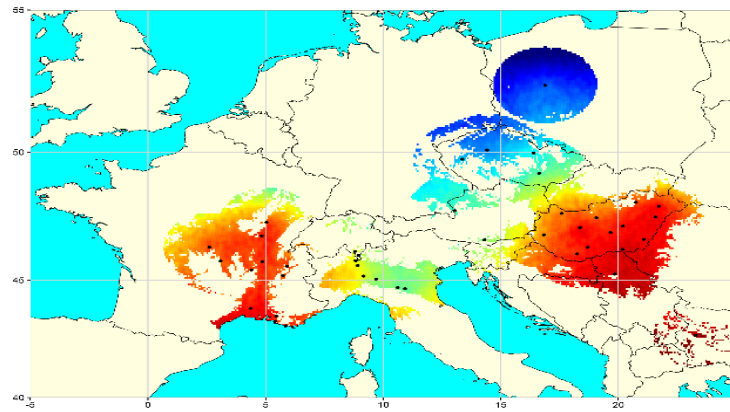
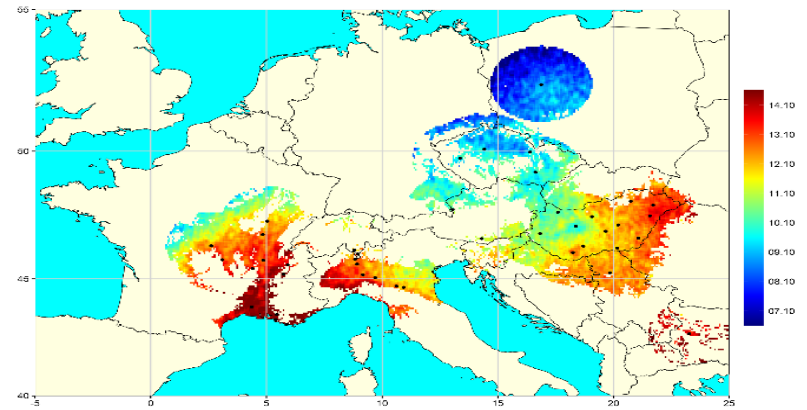
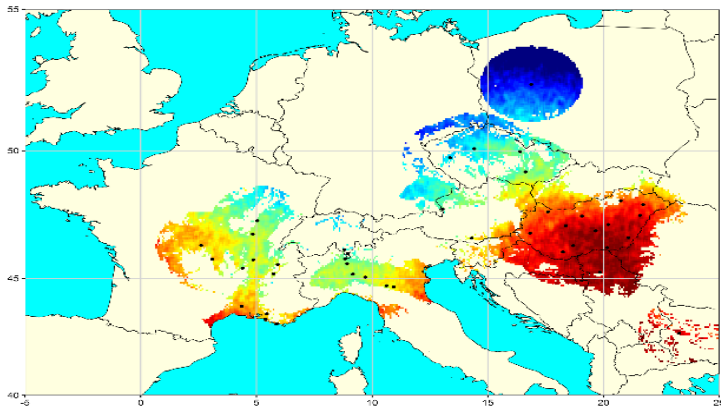
Első térképek a nemzetközi szakirodalomban:

(b) helyreállított(GM) adatokkal (fent jobbra), és

(c) helyreállított(DL) adatokkal (lent)

3'. ábra

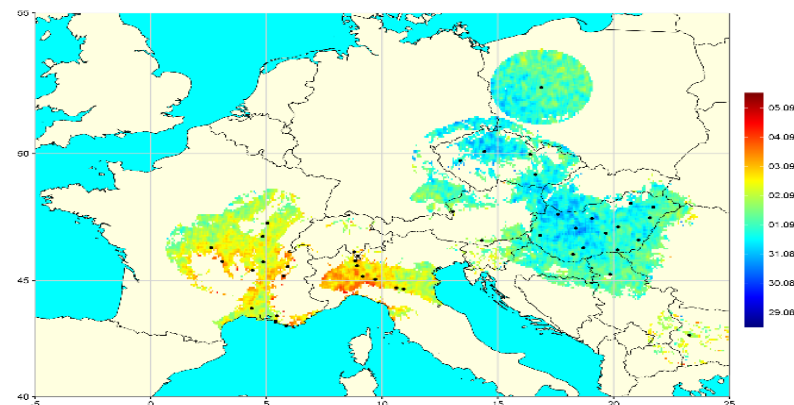
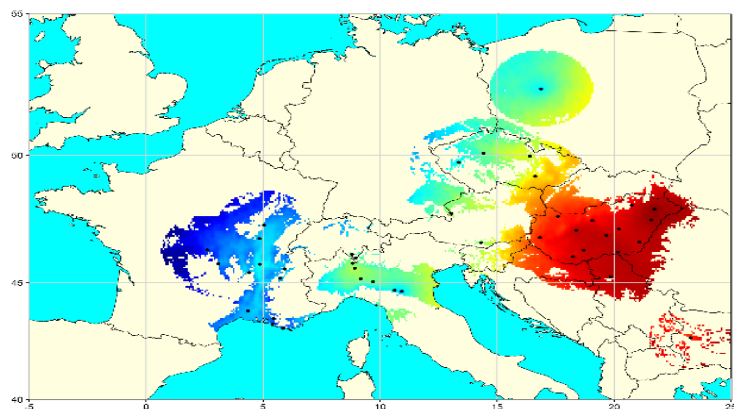
Az évi összes parlagfűpollen koncentráció Európára 10-es alapú logaritmus színskálán, a vizsgált aerobiológiai állomások földrajzi pozícióival az (a) eredeti adatok (fent balra), (b) helyreállított(GM) adatok (fent jobbra), és (c) helyreállított(DL) adatok (lent) felhasználásával, 2010 július 15 – október 15, (pollenszem · m⁻³ levegő). A (1) 100 pollenszem · m⁻³ levegőnél kisebb évi összes pollenkoncentrációval rendelkező állomásokat kizártuk a térképezésből, (2) az egyes állomások 150 km-es körzetén belül az adott állomás tszf. magasságánál +/-100 m-en kívüli magasságú területeket nem térképeztük, és (3) az állomások közötti maximális távolság 150 km. Vízszintes tengely: hosszúsági fok; függőleges tengely: szélességi fok. Az aerobiológiai állomások földrajzi pozícióit fekete pontok jelzik. **Helyreállított(GM) adatok:** a hiányzó napi adatok helyreállítása annak a feltételezésével történt, hogy a napi pollenkoncentrációk szezonális eloszlása normális (Gaussi módszer). **Helyreállított(DL) adatok:** a hiányzó napi adatok helyreállítása konvolúciós auto-encoder modellek felhasználásával és időjárás adatok¹ figyelembe vételével történt (ConvAE + Weather). [Időjárás adatok¹: a maximum hőmérséklet, a minimum hőmérséklet és a középhőmérséklet, valamint a csapadék napi adatai.]



Első térképek a nemzetközi szakirodalomban:

4. ábra

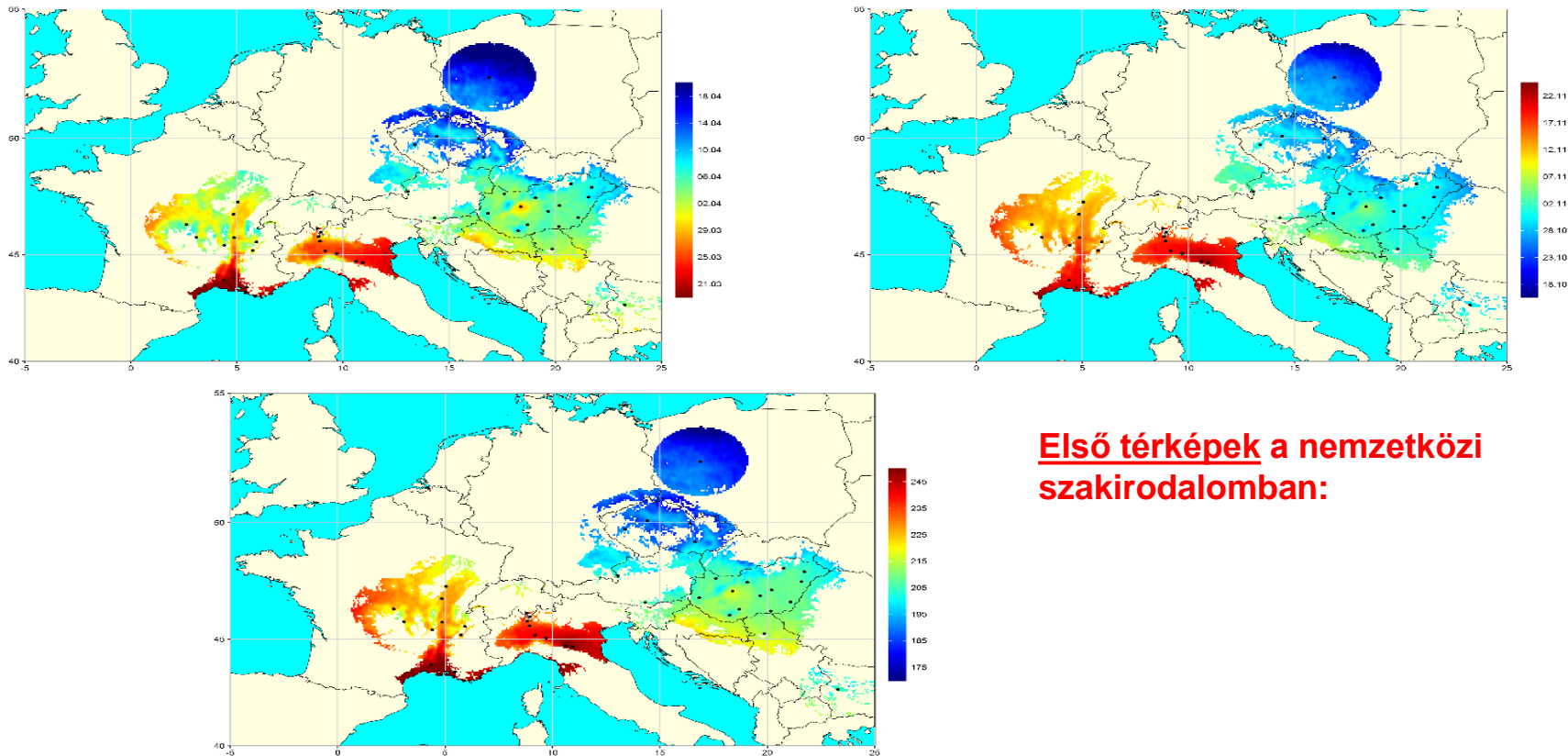
A parlagfűpollen szezon (a) átlagos kezdőnapja (fent balra), (b) átlagos befejező napja (fent jobbra), és (c) átlagos tartama (lent) Európára, a vizsgált aerobiológiai állomások földrajzi pozícióival, eredeti adatok, 16 év, 1995-2010. A (1) 100 pollenszem · m⁻³ levegőnél kisebb évi összes pollenkoncentrációval rendelkező állomásokat kizártuk a térképezésből, (2) az egyes állomások 150 km-es körzetén belül az adott állomás tszf. magasságánál +/-100 m-en kívüli magasságú területeket nem térképeztük, és (3) az állomások közötti maximális távolság 150 km. Vízszintes tengely: hosszúsági fok; függőleges tengely: szélességi fok. Az aerobiológiai állomások földrajzi pozícióit fekete pontok jelzik.



Első térképek a nemzetközi szakirodalomban:

5. ábra

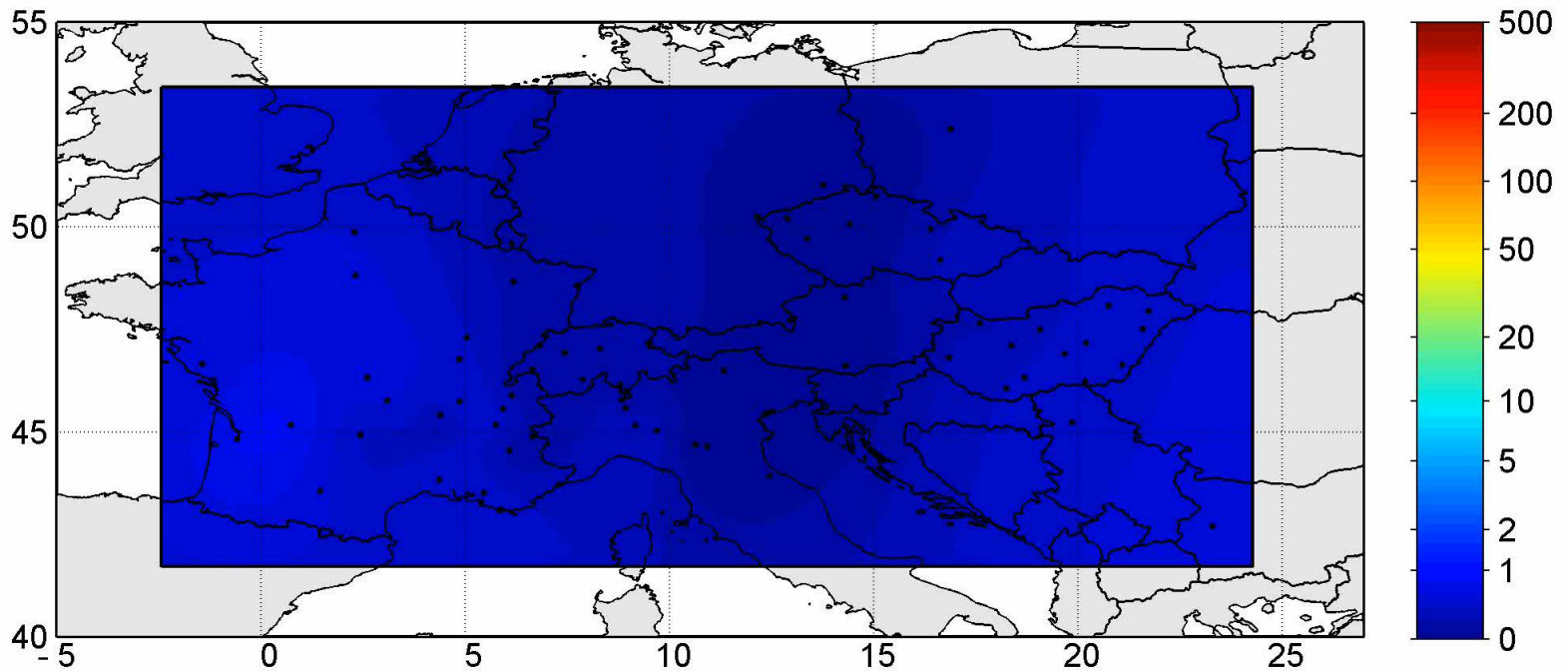
Az átlagos maximális napi parlagfűpollen koncentráció (a) értéke (pollenszem · m⁻³ levegő) 10-es alapú logaritmus színskálán (balra), és (b) napja (jobbra) Európára, a vizsgált aerobiológiai állomások földrajzi pozícióival, eredeti adatok, 16 év, 1995-2010. A (1) 100 pollenszem · m⁻³ levegőnél kisebb évi összes pollenkoncentrációval rendelkező állomásokat kizártuk a térképezésből, (2) az egyes állomások 150 km-es körzetén belül az adott állomás tszf. magasságánál +/-100 m-en kívüli magasságú területeket nem térképeztük, és (3) az állomások közötti maximális távolság 150 km. Vízszintes tengely: hosszúsági fok; függőleges tengely: szélességi fok. Az aerobiológiai állomások földrajzi pozícióit fekete pontok jelzik.



Első térképek a nemzetközi szakirodalomban:

6. ábra

A (a) tavaszi átlagos utolsó fagyos nap (fent balra), az (b) őszi átlagos első fagyos nap (fent jobbra), és a (c) fagymentes időszak átlagos tartama (lent) Európára, a vizsgált aerobiológiai állomások földrajzi pozícióival, hiánytalan eredeti adatok felhasználásával, 16 év, 1995-2010. Minden egyes állomás adatkészlete hiánytalan. A (1) 100 pollenszem · m⁻³ levegőnél kisebb évi összes pollenkoncentrációval rendelkező állomásokat kizártuk a térképezésből, (2) az egyes állomások 150 km-es körzetén belül az adott állomás tszf. magasságánál +/-100 m-en kívüli magasságú területeket nem térképeztük, és (3) az állomások közötti maximális távolság 150 km. Vízszintes tengely: hosszúsági fok; függőleges tengely: szélességi fok. Az aerobiológiai állomások földrajzi pozícióit fekete pontok jelzik.

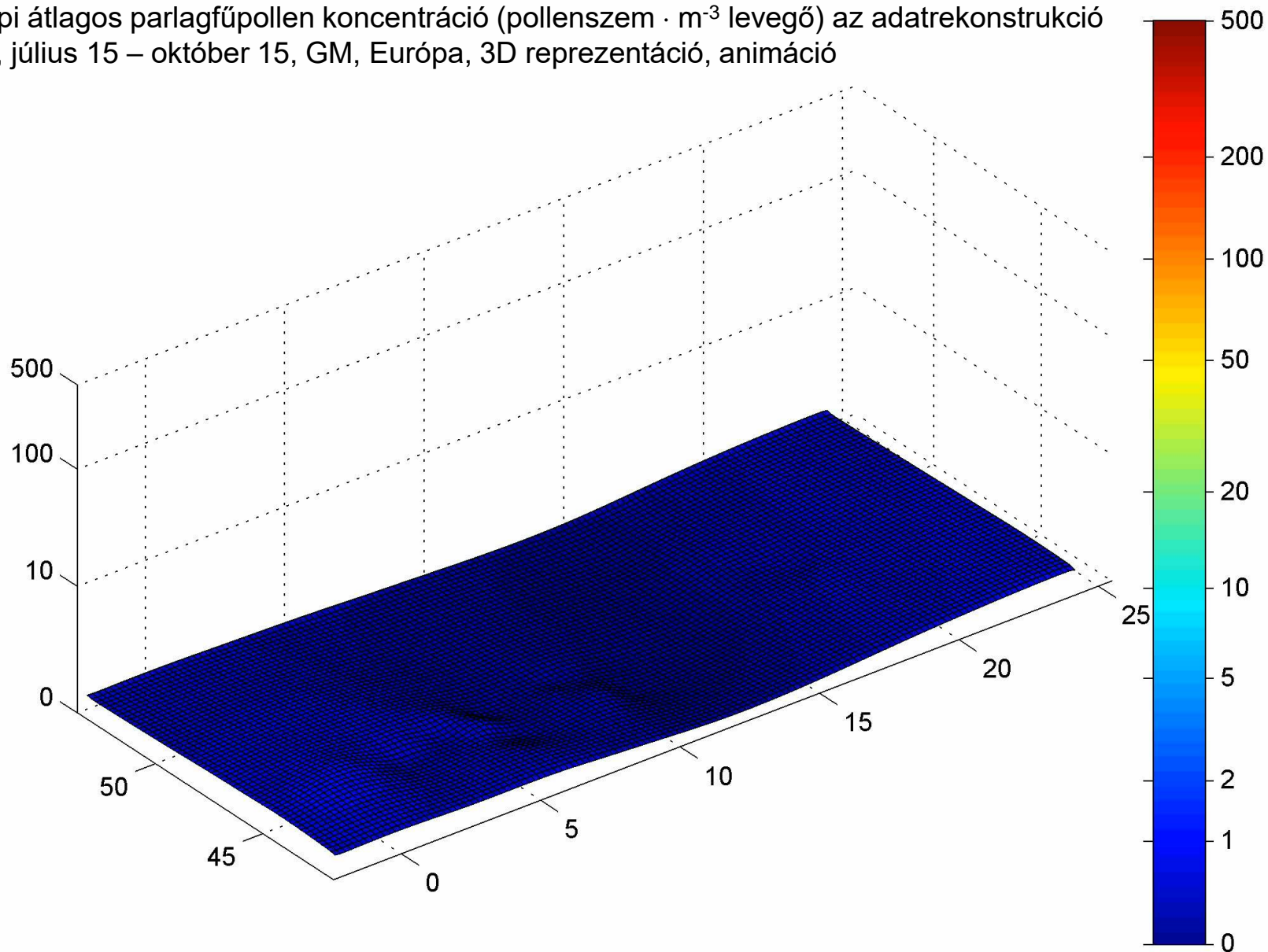


7. ábra

A napi átlagos parlagfűpollen koncentráció (pollenszem · m⁻³ levegő) az adatrekonstrukció után, július 15 – október 15, GM, Európa, 2D reprezentáció, animáció

8. ábra

A napi átlagos parlagfűpollen koncentráció (pollenszem · m⁻³ levegő) az adatrekonstrukció után, július 15 – október 15, GM, Európa, 3D reprezentáció, animáció



7. A GM és DL modell adatrekonstrukciója jóságának az összehasonlítása

2. táblázat

Az ATPC becslésének eredményei az R^2 , $RMSE$ és a MAE alapján. A DL modell bemenetként tartalmazza az időjárési információkat. Minden egyes metrikára az átlag és a szórás kiszámítása a 10 szimuláció alapján történt. [A kétlépéses térbeli interpoláció (SI) eredményei a táblázat utolsó sorában található.]

[**piros/kék aláhúzott**: az adott módszer pontosabban becsüli az ATPC-t.]

Removed data, %	Method	R^2	$RMSE$	MAE
0.1	<u>DL</u>	0.9971 ± 0.00	<u>203.47 ± 44.72</u>	<u>123.03 ± 21.53</u>
	GM	0.9654 ± 0.01	705.25 ± 120.77	460.60 ± 58.38
0.2	<u>DL</u>	0.9912 ± 0.01	<u>343.39 ± 109.26</u>	<u>204.52 ± 50.10</u>
	GM	0.9503 ± 0.05	789.72 ± 333.28	489.66 ± 84.82
0.3	<u>DL</u>	0.9802 ± 0.01	<u>521.27 ± 146.65</u>	<u>323.33 ± 59.59</u>
	GM	0.9130 ± 0.10	1002.28 ± 529.73	545.31 ± 164.75
0.4	<u>DL</u>	0.9678 ± 0.01	<u>680.54 ± 114.38</u>	<u>449.96 ± 66.05</u>
	GM	0.9299 ± 0.05	976.64 ± 286.94	595.34 ± 98.10
0.5	<u>DL</u>	0.9302 ± 0.02	<u>1008.30 ± 118.74</u>	<u>644.69 ± 51.88</u>
	GM	0.8430 ± 0.09	1460.19 ± 433.07	861.59 ± 181.14
0.6	<u>DL</u>	0.8745 ± 0.03	<u>1348.70 ± 187.85</u>	917.29 ± 100.88
	<u>GM</u>	0.8574 ± 0.07	1410.63 ± 341.93	<u>886.81 ± 140.82</u>
0.7	<u>DL</u>	0.8149 ± 0.06	<u>1635.62 ± 244.74</u>	1196.53 ± 168.97
	<u>GM</u>	0.7901 ± 0.08	1732.98 ± 314.95	<u>1042.87 ± 158.84</u>
0.8	DL	0.5262 ± 0.12	2626.26 ± 322.68	1971.28 ± 217.17
	<u>GM</u>	0.6071 ± 0.25	<u>2292.06 ± 742.99</u>	<u>1325.30 ± 271.27</u>
Two-step SI	SI	0.3052	3204.25	2297.74

8. Összegzés, az eredmények jelentősége

Ezek a parlagfűpollen mennyiségi és fenológiai paramétereinek eddigi legpontosabb térképei Európára, mert

- 1) az eddigi legtöbb**, azaz 625 aerobiológiai **állomás adatai alapján készültek**;
- 2) az eddigi leghosszabb**, azaz 16 éves (1995-2010) **adatbázisokat használtunk**;
- 3) az eddigi legpontosabb adatbázisokat állítottuk elő**;
- 4) az eddigi legkorszerűbb statisztikai módszereket** (Gauss módszer, mesterséges intelligencia, „deap learning”) **alkalmaztuk**;
- 5) a legkorszerűbb térképészeti eljárásokat használtuk**;
- 6) a térképek pontosságát statisztikailag értékeltük és összehasonlítottuk**;
- 7) honlapot készítettünk** a nyers és helyreállított adatok használatára (<https://euragweedpollen.gmf.u-szeged.hu/>)

Az egyes térképeken az extrapoláció miatti hamis információk kiszűrésére az alábbi finomítások történtek:

- a) A mérőállomások közötti maximális távolság 150 km. Az ennél távolabb lévő régiókat nem vettük figyelembe a térképezéskor (Skjøth et al., 2012; Smith et al., 2013).
- b) Az egyes állomások 150 km-es körzetén belül az adott állomás tszf. magasságánál +/- 100 m-en kívüli magasságú területeket nem térképeztük. **ÚJ!**
- c) Azon aerobiológiai állomásokat nem térképeztük, ahol az átlagos éves pollenkoncentráció nem érte el a legalább 100 pollenszem · m⁻³ értéket. **ÚJ!**

Köszönöm!