

## ÉSZAK-ATLANTI TÁVKAPCSOLATI JELENSÉGEK AZONOSÍTÁSA AZ R PROGRAMNYELV FELHASZNÁLÁSÁVAL

Kristóf Erzsébet <sup>(1)</sup> , Barcza Zoltán <sup>(1,2)</sup> 

<sup>(1)</sup> ELTE TTK Kiválósági Tudásközpont, 2462 Martonvásár, Brunszvik u. 2.

<sup>(2)</sup> ELTE Meteorológiai Tanszék, 1117 Budapest, Pázmány Péter sétány 1/A  
e-mail: [ekristof86@caesar.elte.hu](mailto:ekristof86@caesar.elte.hu); [zoltan.barcza@ttk.elte.hu](mailto:zoltan.barcza@ttk.elte.hu)

### Bevezetés

Az R egy nyílt forráskódú objektumorientált programnyelv, amelyet az 1990-es évek közepén fejlesztettek ki (R Core Team, 2020). Létrehozásának alapvető célja az volt, hogy olyan eszközt biztosítson a felhasználók számára, amellyel statisztikai vizsgálatok egyszerűen végezhetőek el, eredményeik pedig könnyen jeleníthetők meg (Ihaka & Gentleman, 1996).

Az R programnyelv előnye, hogy programozói tapasztalattal rendelkezve bárki bővítheti új funkciókkal. A telepítését követően azonnal elérhető függvényeken – a továbbiakban: beépített függvényeken – kívül fejlesztett új függvények R csomagokban (R packages) érhetőek el a felhasználók számára, amelyek az R konzolfelületén vagy az integrált fejlesztői környezetében, az RStudio-ban egyetlen paranccsal tölthetők le és telepíthetők. Az R közösség által hivatalosan jóváhagyott R csomagok a Comprehensive R Archive Network (CRAN) honlapján [1] érhetőek el részletes dokumentációjukkal együtt, azonban számos, nem hivatalos csomag, függvény is megtalálható olyan internetes platformokon, mint a GitHub.

Az R széleskörű használatát elősegítette, hogy Windows, Linux és Macintosh operációs rendszerek alatt is használható. Alkalmazási lehetőségeit pedig kiterjeszti, hogy interfészek segítségével más programnyelven, például Python és C++ nyelven írt kódok is integrálhatók, futtathatók az R-ben (Eddelbuettel & Francois, 2011; Ushey et al., 2020). Népszerűségét mutatja, hogy az elmúlt évtizedben közel tízezerrel, 2 000 körüliről 12 000 fölé nőtt a le-tölthető R csomagok száma (Epskamp, 2019).

Az R-ben lehetőségünk nyílik olyan kódok megírására, amelyek az adatok letöltésétől azok feldolgozásán, majd statisztikai vizsgálatán át az eredmények megjelenítéséig minden lépést elvégeznek, így az adatok különböző szoftveres platformok közötti hordozása elkerülhető. A kódokból és a futtatási eredményekből dinamikus dokumentumok is szerkeszthetők az R-ben (Xie, 2014; Xie et al., 2018). Egyszerű szintaktikája miatt a 2010-es évek végére az R-beli adatfeldolgozással foglalkozó, ingyenesen elérhető elektronikus könyvek sorát írták meg, amely tovább növelte a programnyelv népszerűségét [2]. Az elérhető megoldások könnyen adaptálhatók, ami elősegítette az R földtudományi, így meteorológiai célú alkalmazásának elterjedését. Napjainkban rendelkezésre áll kifejezetten meteorológiai célú R-beli mező-elemzési módszereket bemutató könyv is, elméleti összefoglalókkal és részletes példákkal (Wikle et al., 2019).

Az alábbiakban három reanalízis adatbázis 500 hPa-os geopotenciális magasság mezőinek észak-atlanti távkapcsolatokra vonatkozó összevetésén keresztül mutatjuk be az adatok letöltésének, statisztikai vizsgálatra való előkészítésének, elemzésének és az eredmények megjelenítésének R-beli példáját. A három reanalízist a rácspontonkénti idősorok korreláltatása útján előállított legerősebb negatív korrelációmezők felhasználásával – teleconnectivity method, más néven: negative extrema method (Horel, 1981; Wallace & Gutzler, 1981) – és főkomponens-elemzéssel (pl. Hurrell, 1995; Hurrell & Deser, 2009), az első empirikus ortogonális függvény (a továbbiakban: EOF<sup>1</sup>) és az ahhoz tartozó főkomponens idősor

<sup>1</sup> empirical orthogonal function, EOF

értékelése alapján hasonlítjuk össze. A kódok megírásakor törekedtünk az egyszerű és gyors megoldásokra, ahol csak lehetett az R beépített függvényeire szorítva. Előbbiek alkalmazásával az adatok letöltésére, feldolgozására, a korrelációanalízis és a főkomponens-elemzés elvégzésére saját függvényeket írtunk. Hasonló adatelemzési és vizualizációs megoldások korábbi tanulmányainkban is alkalmazásra kerültek (pl. Kristóf et al., 2018; 2020; Ács et al., 2020a; 2020b).

A tanulmányban bemutatott vizsgálathoz írt R kódok a GitHub tárhelyén érhetők el [3].

### Felhasznált adatbázisok, az adatok letöltése

Az R-beli statisztikai vizsgálat bemutatásához az ECMWF<sup>2</sup> ERA5 reanalízis adatbázisát (Hersbach et al., 2020), az NCEP/NCAR<sup>3</sup> Reanalysis 1 (a továbbiakban: NCEP R1) adatbázisát (Kalnay et al., 1996) és az NCEP/DOE<sup>4</sup> Reanalysis 2 (a továbbiakban: NCEP R2) adatbázisát (Kanamitsu et al., 2002) használtuk fel. Az ERA5 és az NCEP R2 adatbázist a legmodernebb reanalízisek között tartják számon [4].

Jelen tanulmány írásakor az ERA5-ben és az NCEP R2-ben 1979-től, az NCEP R1-ben 1948-tól érhetők el az adatok. Mindhárom reanalízis előnye, hogy folyamatosan bővülnek, adataik pár napos késéssel tölthetők le. Közös jellemzőjük, hogy globálisan tartalmaznak adatokat, amelyek előállításához műholdas méréseket is asszimilálnak. Az ERA5 1950-ig való kiterjesztése napjainkban zajlik. Térbeli és időbeli felbontását tekintve az ERA5 a legfinomabb felbontású a három reanalízis adatbázis közül. Órás geopotenciális magasság adatai 0,25°-os rácson érhetők el. Hersbach et al. (2020) az ERA5 erősségei között említi a troposzférikus folyamatok jobb reprezentációját a korábbi ECMWF reanalízis adatbázisokhoz képest. Az NCEP R1 és az NCEP R2 adatai durvább, 2,5°-os rácson, hat óránként tölthetők le. Az NCEP R1 vizsgálata történetileg érdekes, hiszen adatasszimilációs rendszere az 1990-es évek közepe óta változatlan módon működik. Az NCEP R2 reanalízist modernebb fizikai parametrizációkkal látták el, javították az NCEP R1 ismert hibáit, amelyek például a hó-takaró, a hóolvadás és az óceáni albedó reprezentációjához kötődnek. Kanamitsu et al. (2002) a két reanalízis adatbázis legnagyobb eltérései között említi, hogy az előbb felsorolt korrekciók következtében az NCEP R2-beli felszínhőmérséklet és két méteres léghőmérséklet magasabb, a globális felhőborítottság pedig nagyobb, mint a NCEP R1-ben.

A három reanalízis adatbázis összevethetősége érdekében az R-ben januári hat óras, 500 hPa-os geopotenciális magasság adatokat töltöttünk le 1981–2010, 1986–2015 és 1991–2020 közötti harminc éves időszakokra 2,5° felbontású rácson, az északi félgömbön. Az ERA5-beli adatok Pythonban futtatható scripttel érhetők el, de az R-beli letöltésre lehetőséget biztosít az *ecmwf* csomag (Hufkens et al., 2019). A csomag függvényeivel az adatok netCDF vagy GRIB<sup>5</sup> formátumban tölthetők le a felhasználó által választott térbeli és időbeli tartományra, illetve felbontással. Az adatokhoz a Copernicus Climate Data Store honlapján [5] való regisztrációt követően férhetünk hozzá. Az NCEP R1 és NCEP R2 adatok regisztráció nélkül is elérhetők az NCEP honlapján [6, 7]. R-beli letöltésük az RNCEP csomag függvényeivel lehetséges (Kemp et al., 2012). Az eredetileg netCDF formátumban hozzáférhető adatokat a felhasználó által kiválasztott térbeli és időbeli tartományra, valamint időbeli felbontással az R saját állományformátumában, *rds* típusú állományban tölthetjük le.

---

<sup>2</sup> European Centre for Medium-Range Weather Forecasts, ECMWF

<sup>3</sup> National Centers for Environmental Prediction, NCEP; National Center for Atmospheric Research, NCAR

<sup>4</sup> Department of Energy, DOE

<sup>5</sup> Network Common Data Form, NetCDF; Gridded Binary Form, GRIB

## Az adatmezők előkészítése és statisztikai vizsgálata

Az ERA5 adatokat tartalmazó netCDF állományt a *ncdf4* csomag (Pierce, 2019) függvényeivel hatékonyan kezelhetjük az R-ben. Pár sornyi kóddal menthetjük az állományban tárolt meta adatokat, a változók és a dimenziók értékeit. Egymás után több állomány automatizált beolvasását megkönnyíti, hogy az R-ben olyan változók is rendelkezésre állnak, amelyek dimenziói, elemeinek száma előzetesen nem deklarálandó. Többek között a lista (*list*) típusú objektumok alkalmasak arra, hogy tetszőleges típusú és dimenziójú adatokat tároljunk ezek részlistáiként (*sublists*). Listákba menthetjük a netCDF állományok meta adatait. Ezek olyan összetett objektumok, amelyek egyszerre tartalmaznak különböző elemszámú szövegszerű, numerikus és dátum formátumú változókat. Az *rds* típusú állományokban egyetlen objektum, a reanalízis adatok többdimenziós tömbje kerül tárolásra, amely alapján azonban a tömb dimenziói, azaz a földrajzi hosszúsági és szélességi körök értékei, valamint az időpontok meghatározhatók az R beépített függvényeivel.

A reanalízisek összehasonlításához az észak-atlanti távkapcsolati mintázat (EA<sup>6</sup>) és az Észak-atlanti Oszcilláció (NAO<sup>7</sup>) azonosítását tűztük ki célul, amelyek térbeli és időbeli különbségeit vizuálisan, továbbá néhány alapstatisztika felhasználásával vizsgáljuk meg. Az elemzéshez az é. sz. 20°–80° és a ny. h. 75°–k. h. 40° közötti atlanti-európai területet választottuk, amely régió alkalmas mind az EA, mind a NAO azonosítására (Wallace & Gutzler, 1981; IPCC, 2013). A hat óránkénti geopotenciális magasság adatsorokból havi átlagokat számoltunk, amelyeket standardizáltunk. A vizsgálatokat így 1175 rácspont 30 elemű idősorain hajtottuk végre.

Az R-ben számos mezőelemzési módszer bemeneteként táblázatba rendezett adatok szolgálnak. A jelenségek detektálása érdekében az R-ben végrehajtandó korreláció-analízishez és főkomponens-elemzéshez a rácsponti idősorok háromdimenziós tömbjei egyszerűen alakíthatók kétdimenzióssá az R beépített függvényeinek felhasználásával. Az átalakítás eredményeképpen a táblázat minden egyes oszlopába egy rácspontra vonatkozó idősor kerül, ezáltal az R a továbbiakban *vektorizált objektumon*, vagyis kétdimenziós mátrixba rendezett oszlopvektorokon hajt végre műveleteket. Ez a lépés nélkülözhetetlen az R programok futási idejének csökkentéséhez. A háromdimenziós tömbök adatain egymásba ágyazott for ciklussal végighaladva nagyságrendekkel nő a futási idő.

A korrelációanalízis során Pearson-korrelációkat számítottunk ki. Minden egyes rácsponti idősor összes többivel való korreláltatását a *Hmisc* csomag *rcorr* függvényével végeztük el (Harrell et al., 2020). A rácspontonkénti legerősebb negatív korrelációk meghatározásához és a szignifikáns korrelációs szint megállapításához az R beépített függvényeit használtuk fel. Utóbbi érdekében permutációtesztet alkalmaztunk, amelyet ezerszer ismételtünk meg. (A módszer részletes leírását Kristóf et al. (2020) tartalmazza). A főkomponens-elemzést az R-ben alapértelmezetten elérhető *prcomp* függvénnyel hajtottuk végre, amely lehetővé teszi az eljárást abban az esetben is, ha a megfigyelések, vagyis az időpontok száma kisebb, mint a változók, azaz a rácspontok száma.

A rácsponti idősorok korrelációinak meghatározását követően – szintén az R beépített függvényeinek segítségével – kerestük ki az egymással kölcsönösen egyértelműen legerősebb negatív korrelációban álló rácspontpárokat. Ezeket tekintjük a távkapcsolati jelenségek potenciális akciócentrumainak (Kristóf et al., 2020). A reanalízisenként és időszakonként eltérő számú rácspontpár, valamint az *rcorr* és *prcomp* függvénnyel végrehajtott vizsgálat eredményeképpen előállított összetett objektumok R-beli tárolására szintén a lista típusú változót

<sup>6</sup> Eastern-Atlantic pattern, EA

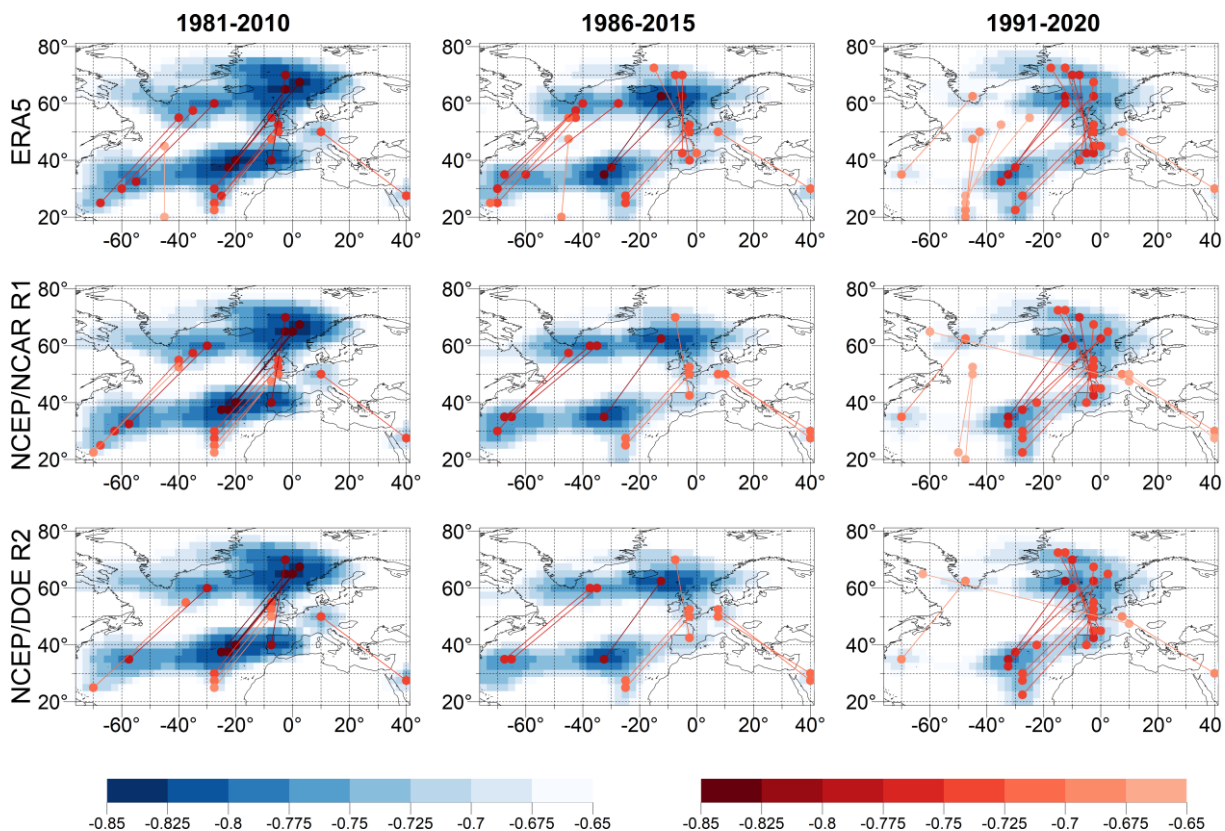
<sup>7</sup> North-Atlantic Oscillation, NAO

választottuk. Az első EOF-hoz tartozó főkomponens idősorok hasonlóságát lineáris korreláció és az átlagos négyzetes hiba (RMSE) négyzetgyökének kiszámításával értékeltük.

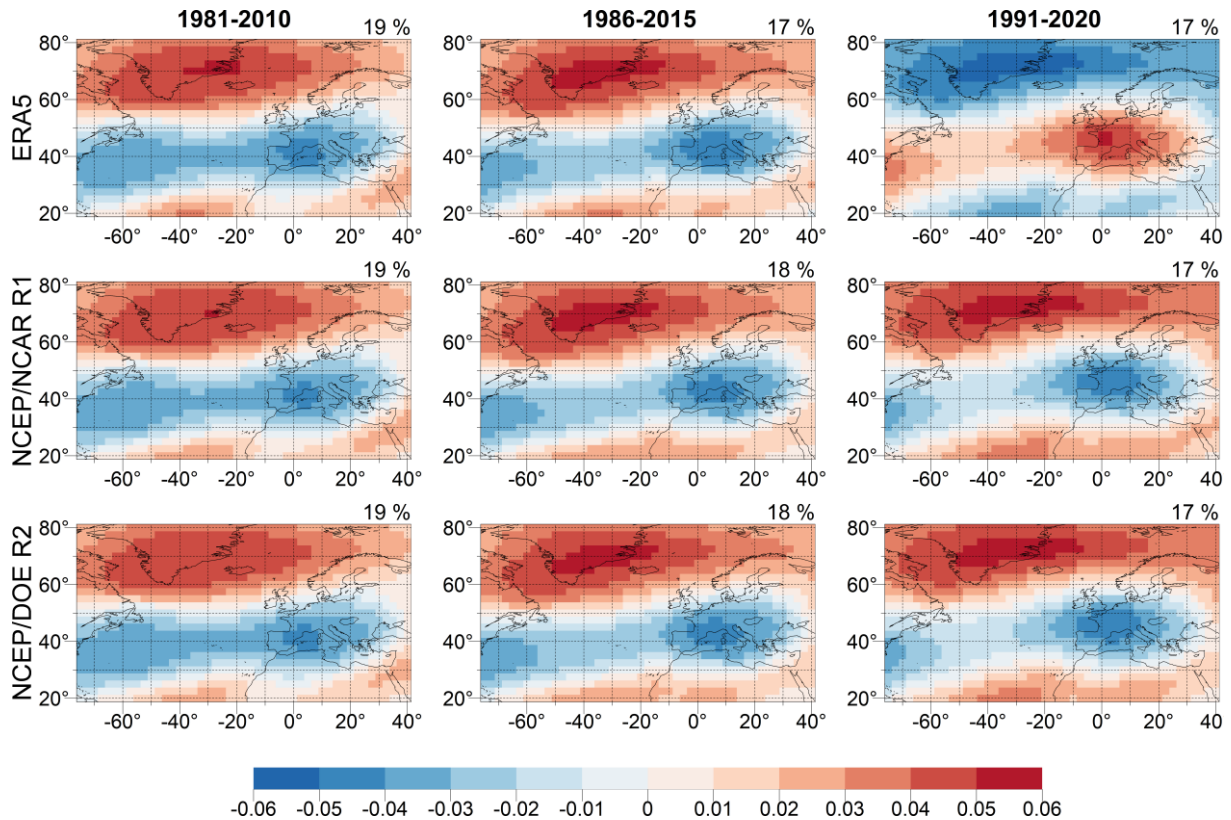
A legerősebb negatív korrelációmezők és a főkomponens-elemzés alapján kapott első EOF-mezők ábrázolására térképen került sor (1–2. ábra), míg az első EOF-hoz tartozó főkomponens idősorokat oszlopdiagramokon jelenítettük meg (3. ábra). A beépített *image* függvényvel négyzetes hengervetületű térkép készíthető az ábrázolandó értékek táblázatának, az ezekhez tartozó hosszúsági körök és szélességi körök vektorainak felsorolásával. Ország-határok illesztésére a *maps* csomag *map* függvénye alkalmazható (Brownrigg et al., 2018). Az R-ben alapértelmezetten elérhető számos színskála, amelyek manipulálásra, így a rendezésre álló színárnyalatok közé új színárnyalatok interpolálására az *RColorBrewer* csomagot használtuk (Neuwirth, 2014). A színskála térképekre illesztésére és a térképek egy ábrára rendezésére kizárólag az R beépített függvényeit alkalmaztuk. Az oszlopdiagramokat a leg-egyszerűbb módon, az R beépített *barplot* függvényével készítettük el, amely lehetővé teszi az oszlopok feltételes színezését.

## Eredmények

A három reanalízis adatbázis alapján meghatározott legerősebb negatív korrelációmezők (1. ábra) és az első empirikus ortogonális függvények (2. ábra) mintázatai nagyon hasonlóak minden egyes időszakban. Bár az NCEP R1 nem tartozik a legfejlettebb reanalízisek közé, mégis figyelemre méltó hasonlóságot mutat az ERA5 és NCEP R2 reanalízissel.



1. ábra: 1 %-os szignifikancia szinten szignifikánsnak tekinthető, rácspontonkénti legerősebb negatív korrelációk (kék árnyalattal) az ERA5, az NCEP/NCAR R1 és az NCEP/DOE R2 reanalízisek januári geopotenciális magasság idősorai alapján, három időszakban. A potenciális akciócentrumok rácspontpárjait vonalakkal összekötött körök jelölik (piros árnyalattal). A sötétebb színárnyalatok erősebb, a világosabb színárnyalatok gyengébb korrelációkat jelentenek.



2. ábra: Rácsponyi súlyok az első empirikus ortogonális függvény alapján az ERA5, az NCEP/NCAR R1 és az NCEP/DOE R2 reanalízisek januári geopotenciális magasság idősorai alapján, három időszakban. A térképek jobb felső sarkában a teljes adattömb változékonyságának (szórásnégyzetének) az első empirikus ortogonális függvény által kifejezett hányada található.

Az észak-atlanti szektoron végrehajtott korrelációanalízis eredményeképpen a Wallace & Gutzler (1981) által azonosított három pólusú észak-atlanti mintázat nyugati pólusaihoz – é. sz. 25° – ny. h. 25° és é. sz. 55° – ny. h. 20° – hasonló pozícióban található távkapcsolati mintázat. A főkomponens-elemzés során meghatározott első EOF alapján pedig északnyugat-délkelet tengelyre szimmetrikus dipólus szerkezet rajzolódik ki, Izlandhoz és Délnyugat-Európához közeli központtal. Az általunk azonosított észak-atlanti mintázat nem tévesztendő össze a szintén észak-atlanti mintázatként számon tartott, észak-dél tengelyre szimmetrikus dipólussal az atlanti-európai térség fölött, amely a második empirikus ortogonális függvény alapján definiálható (Barnston & Livezey, 1987).

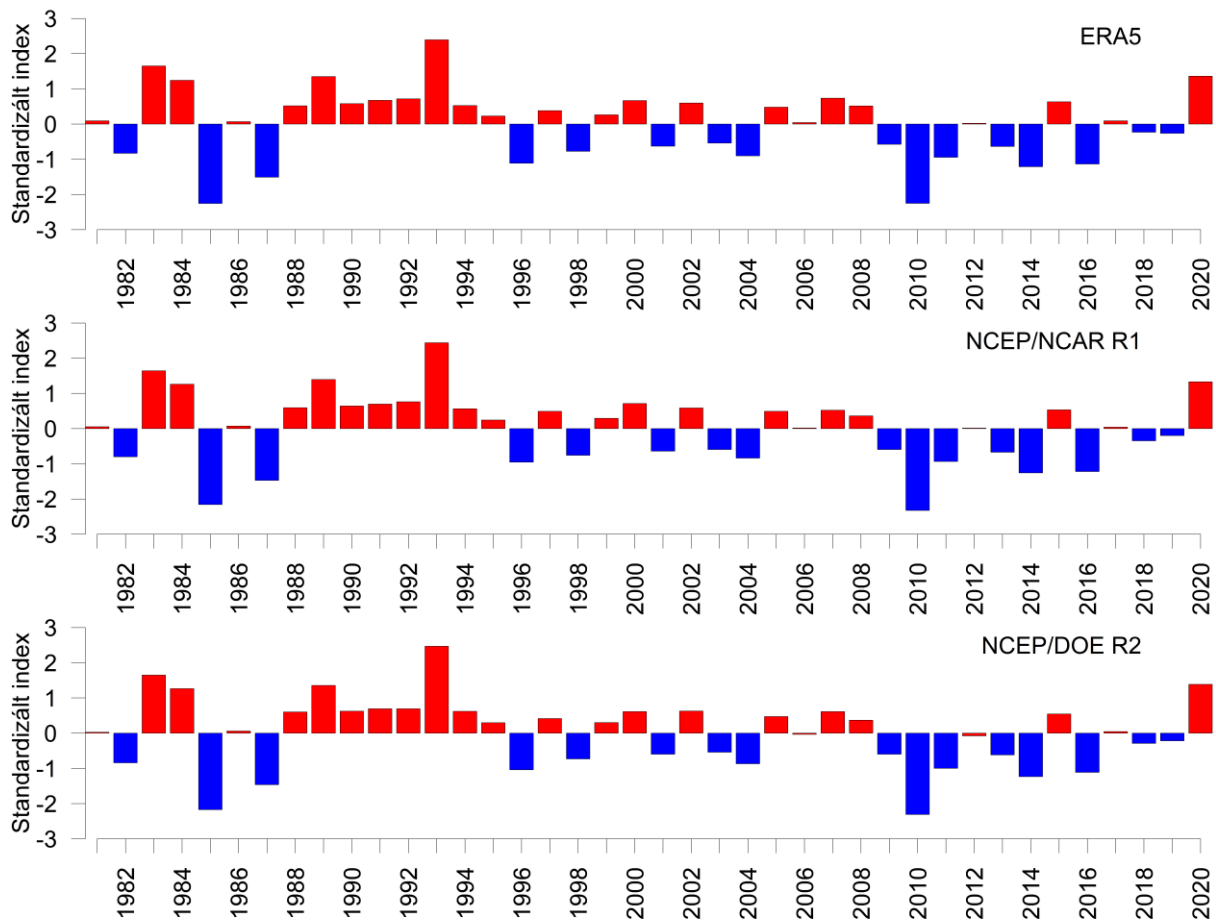
A legerősebb negatív korrelációmezők alapján meghatározott potenciális akciócentrumok közül a legerősebbek az Atlanti-óceán északkeleti medencéje felett találhatóak. A legerősebb negatív korrelációjú potenciális akciócentrum földrajzi koordinátái 1986–2015 és 1991–2020 esetén mindhárom reanalízisben megegyeznek, míg 1981–2010 között minimális, két rácspontnyi eltérés figyelhető meg (1. táblázat). A távkapcsolati jelenség legintenzívebb területét illetően tehát a reanalízisek szinte teljesen megegyeznek. Az időben előrehaladva gyengülés figyelhető meg a korrelációmezőben, amely leginkább az Atlanti-óceán északnyugati részét érinti. Wallace & Gutzler (1981) ebben a térségben detektálta a nyugat-atlanti mintázatot<sup>8</sup>, é. sz. 55° – ny. h. 55° és é. sz. 55° – ny. h. 30° akciócentrumokkal.

<sup>8</sup> Western-Atlantic pattern

1. táblázat: A legerősebb negatív korrelációjú potenciális akciócentrumokat reprezentáló rácspontpár földrajzi koordinátái és a kapcsolódó korreláció érték.

Adatbázis	Időszak	Déli centrum (szélesség, hosszúság)		Északi centrum (szélesség, hosszúság)		Korreláció
ERA5	1981–2010	37,5°	-22,5°	65°	-2,5°	-0,83
NCEP R1		37,5°	-22,5°	65°	0°	-0,83
NCEP R2		37,5°	-25°	65°	-2,5°	-0,83
ERA5	1986–2015	35°	-32,5°	62,5°	-12,5°	-0,83
NCEP R1		35°	-32,5°	62,5°	-12,5°	-0,82
NCEP R2		35°	-32,5°	62,5°	-12,5°	-0,82
ERA5	1991–2020	35°	-32,5°	62,5°	-12,5°	-0,78
NCEP R1		35°	-32,5°	62,5°	-12,5°	-0,78
NCEP R2		35°	-32,5°	62,5°	-12,5°	-0,78

Az ERA5, az NCEP R1 és az NCEP R2 reanalízisek első empirikus ortogonális függvényéhez tartozó főkomponens idősorából az Észak-atlanti Oszcilláció intenzitását mérő, standardizált index állítható elő. A három adatbázis alapján az 1981 és 2020 közötti időszakra előállított indexek rendkívül hasonlóak (3. ábra). Lineáris korrelációjuk mértéke minden esetben 0,99 fölötti, az RMSE négyzetgyöke pedig 0,1 alatti.



3. ábra: Az első empirikus ortogonális függvényhez tartozó főkomponens idősor 1981 és 2020 között az ERA5, az NCEP/NCAR R1 és az NCEP/DOE R2 reanalízisek januári geopotenciális magasság idősorai alapján.

## Diszkusszió és konklúzió

Tanulmányunk célja egy teljesen R programnyelvben végrehajtott vizsgálat bemutatása volt, amely során könnyen kezelhető függvényekkel, vizuálisan és alapstatisztikák meghatározásával hasonlítottunk össze három reanalízis adatbázist. A három reanalízis adatbázis korrelációmezői, az első empirikus ortogonális függvények mezői és az utóbbihoz tartozó főkomponens idősorok időszakonként jelentős hasonlóságot mutatnak.

Megmagyarázandó az ERA5 1991–2020 közötti időszakra kapott mintázata a 2. ábrán, amelyben az első EOF-beli értékek, vagyis a rácspontok súlyai a geopotenciális magasságmező változékonyságának kifejezésében a többihez képest ellentétes előjellel találhatók. Belátható a főkomponens-elemzés módszerének ismertetése alapján, hogy az előjel megváltása önkényes (Horel, 1981). Főkomponens-elemzéssel a többdimenziós ponthalmazt, vagyis a megfigyelt vagy mért értékek pontfelhőjét alacsonyabb dimenziójú térre vetítjük le ortogonálisan. Az alacsonyabb dimenziójú tér meghatározásához a levetítendő pontok merőleges vetületei hossznegyzeteinek átlagát maximalizáljuk. A maximalizálandó kifejezésben pedig a ponthalmaz kovariancia mátrixának sajátvektorai találhatók (Peixoto & Oort, 1992). Sajátérték-sajátvektor feladat oldandó meg tehát a sajátvektorok meghatározása érdekében, amely során hosszúságot maximalizálunk. A sajátvektorok és mínusz egyszereseik hosszának egyezése miatt pedig ugyanazt az eredményt kapjuk, függetlenül a sajátvektor előjelétől. A 2. ábra EOF-mintázataiban a rácsponti súlyok éppen a sajátvektorok komponensei. Az előjelváltás tehát azon múlik, hogy a pozitív vagy a negatív előjelű sajátvektort választjuk. A sajátvektorok előjelváltásával együtt az empirikus ortogonális függvényekhez tartozó főkomponens idősorok elemeinek előjelei is megváltoznak. A főkomponens idősorok ugyanis az eredeti megfigyelések új koordinátái az egyes sajátvektorok által meghatározott egyenesre való merőleges vetítésüket követően.

Az első empirikus ortogonális függvények alapján kirajzolódó mintázatok, valamint az első főkomponens idősorok hasonlósága különösen figyelemre méltó, hiszen a főkomponens-elemzés nagyon érzékeny a bemeneti adatokra, kis eltérés is jelentős változásokhoz vezethet (Horel, 1981). A reanalízisek nagyfokú hasonlósága azt jelenti, hogyha a vizsgálatot elvégezzük például általános cirkulációs modellek (GCM<sup>9</sup>) historikus szimulációin, akkor referenciaként bármelyik reanalízist használhatjuk, hiszen szinte egyező eredményeket fogunk kapni. Ezáltal a modellek validálhatók, azaz kijelölhetők azok a GCM-ek, amelyek a legpontosabban reprodukálják az azonosított távkapcsolati jelenségeket.

A főkomponens-elemzés eredményeképpen a 2. ábrán megjelenített mintázatok alapján meghatározhatók azok a térségek, amelyek a legnagyobb súllyal vesznek részt a teljes adattömb adott empirikus ortogonális függvény által kifejezett változékonyságának kialakításában, továbbá a különböző előjelű térségekről megállapítható, hogy ellentétesen viselkednek. A módszer azonban arról nem ad információt, hogy ezek a területek egymással negatív korrelálnak-e (Dommenges & Latif, 2002). A legerősebb negatív Pearson-korrelációk vizsgálata viszont alkalmas a lineáris kapcsolat szorosságának felderítésére. Horel (1981) viszont ezen módszer főkomponens-elemzéshez képesti legnagyobb hátrányaként a számításigényét említi. Az R-ben vektorizált objektumok létrehozatalával, majd olyan függvények alkalmazásával, mint az *rcorr* és a *prcomp*, a többszörösen egymásba ágyazott ciklusok használata elkerülhető, a számítási igény nagyságrendekkel csökkenthető. A letöltött adattömbök statisztikai vizsgálatára írt kódok hamar, pár perc alatt lefutnak napjaink személyi számítógépein – míg vektorizáció nélkül a futási idő elérheti az egy órát. Az adatok letöltésének időtartama,

---

<sup>9</sup> general circulation model, GCM

viszont változó, az azokat tároló szerver leterheltségének és a saját adatátviteli kapacitásunk függvénye.

Mind az *rcorr*, mind a *prcomp* függvény összetett objektumokat hoz létre. A korrelációanalízis esetén többek között a kiszámított korrelációk, a szignifikanciájukat jellemző p-értékek, a főkomponens-elemzés esetén az empirikus ortogonális függvények, a főkomponens idősorok és a kifejezett varianciahányadok találhatóak egy objektumban. Az R tehát egyszerűen és hatékonyan alkalmazható, akár mélyebb programozói és statisztikai ismeretek nélkül is eredményekhez juthatunk. Éppen emiatt felhasználói szintű használata veszélyeket hordoz magában, hiszen az átgondolatlanul elvégzett műveletek útján kapott eredményeket hibásan értelmezhetjük, tévesen tulajdoníthatunk azoknak fizikai jelentőséget. Az R-beli adatfeldolgozás megkezdése előtt ezért célszerű alaposan utánanézni az alkalmazandó függvények által végrehajtandó számításoknak, a jobb megértés érdekében akár "manuálisan" reprodukálandók a függvényekkel kapott eredmények.

Napjainkban az R programnyelv a meteorológiai adatfeldolgozás népszerű eszköze (Charalampopoulos, 2020). Népszerűsége a viszonylag könnyű tanulhatóságából és a számos rendelkezésre álló adatfeldolgozási, elemzési és ábrázolási eljárásból fakad. Fejlődése folyamatos, egyre több olyan alkalmazás érhető el az R-ben, amelyek korábban csak más platformon voltak használhatók. A fent említett, adatletöltésre és adatelemzésre lehetőséget biztosító csomagokon túl példaként említhető a *magick* csomag (Ooms, 2020), amellyel a Linux platform alatt népszerű ImageMagick nyílt forráskódú képszerkesztő válik R-ben alkalmazhatóvá. Magyarországi fejlesztést tekintve, az AgroMo Projekt keretében olyan biogeokémiai modell válik elérhetővé az R programnyelvben, amely lehetővé teszi az agro-ökoszisztéma legfontosabb folyamatainak széleskörű vizsgálatát, számos új modellezési és adatelemzési funkciót, valamint R-ben szerkesztett grafikus felhasználói felületet biztosítva (Pokovai et al., 2020). A fent leírtak az R programnyelv meteorológiai adatfeldolgozásra, adatelemzésre és vizualizációra való használatát rendkívül előnyössé teszik.

## Köszönetnyilvánítás

A kutatást a Széchenyi 2020 program, Magyarország Kormánya és az Európai Regionális Fejlesztési Alap támogatja (GINOP-2.3.2-15-2016-00028).

## Hivatkozások

- Ács, F., Zsákai, A., Kristóf, E., Szabó, A.I., Breuer, H., 2020a: Carpathian Basin climate according to Köppen and a clothing resistance scheme. *Theoretical and Applied Climatology*, 141: 299–307. <https://doi.org/10.1007/s00704-020-03199-z>
- Ács, F., Zsákai, A., Kristóf, E., Szabó, A.I., Breuer, H., 2020b: Human thermal climate of the Carpathian Basin. *International Journal of Climatology*, 1–14. <https://doi.org/10.1002/joc.6816>
- Barnston, A.G., Livezey, R.E., 1987: Classification, Seasonality and Persistence of Low-Frequency Atmospheric Circulation Patterns. *Monthly Weather Review*, 115(6): 1083–1126. [https://doi.org/10.1175/1520-0493\(1987\)115<1083:CSAPOL>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2)
- Brownrigg, R., Minka, T.P., Deckmyn, A., 2018: maps: Draw Geographical Maps. R package version 3.3.0. <https://CRAN.R-project.org/package=maps>
- Charalampopoulos, I., 2020: The R Language as a Tool for Biometeorological Research. *Atmosphere*, 11: 682. <https://doi.org/10.3390/atmos11070682>
- Dommenget, D., Latif, M.A., 2002: Cautionary Note on the Interpretation of EOFs. *Journal of Climate*, 15(2): 216–225. [https://doi.org/10.1175/1520-0442\(2002\)015<0216:ACNOTI>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<0216:ACNOTI>2.0.CO;2)
- Eddelbuettel, D., Francois, R., 2011: Rcpp: Seamless R and C++ Integration. *Journal of Statistical Software*, 40(8): 1–18. <https://doi.org/10.18637/jss.v040.i08>





- Epskamp, S.*, 2019: Reproducibility and Replicability in a Fast-Paced Methodological World. *Advances in Methods and Practices in Psychological Science*, 1–11. <https://doi.org/10.1177/2515245919847421>
- Harrell, F.E., Dupont, C., et al.*, 2020: Hmisc: Harrell Miscellaneous. R package version 4.4-1. <https://CRAN.R-project.org/package=Hmisc>
- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., et al.*, 2020: The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146: 1999–2049. <https://doi.org/10.1002/qj.3803>
- Horel, J.D.*, 1981: A Rotated Principal Component Analysis of the Interannual Variability of the Northern Hemisphere 500 mb Height Field. *Monthly Weather Review*, 109(10): 2080–2092. [https://doi.org/10.1175/1520-0493\(1981\)109<2080:ARPCAO>2.0.CO;2](https://doi.org/10.1175/1520-0493(1981)109<2080:ARPCAO>2.0.CO;2)
- Hufkens, K., Stauffer, R., Campitelli, E.*, 2019: The ecwmfr package: an interface to ECMWF API endpoints. <https://doi.org/10.5281/zenodo.2647541>
- Hurrell, J.W.*, 1995: Decadal Trends in the North Atlantic Oscillation: Regional Temperatures and Precipitation. *Science*, 269: 676–679. <https://doi.org/10.1126/science.269.5224.676>
- Hurrell, J.W., Deser, C.*, 2009: North Atlantic climate variability: The role of the North Atlantic Oscillation. *Journal of Marine Systems*, 78(1): 28–41. <https://doi.org/10.1016/j.jmarsys.2008.11.026>
- Ihaka, R., Gentleman, R.*, 1996: R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, 5(3): 299–314. <https://doi.org/10.2307/1390807>
- IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change; Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M., Eds. Cambridge University Press, Cambridge, UK, New York, NY, USA. 233.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., et al.*, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society*, 77(3): 437–472. [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2)
- Kanamitsu, M., Ebisuzaki, W., Woollen, J., Yang S-K., Hnilo, J.J., Fiorino, M., Potter, G.L.*, 2002: NCEP-DOE AMIP-II Reanalysis (R-2). *Bulletin of the American Meteorological Society*, 83(11): 1631–1644. <https://doi.org/10.1175/BAMS-83-11-1631>
- Kemp, M.U., van Loon, E.E., Shamoun-Baranes, J., Bouten, W.*, 2012: RNCEP: global weather and climate data at your fingertips. *Methods in Ecology and Evolution*, 3(1): 65–70. <https://doi.org/10.1111/j.2041-210X.2011.00138.x>
- Kristóf, E., Bartholy, J., Pongrácz, R.*, 2018: Statistical relationships between large-scale oscillation phenomena and meso-scale atmospheric conditions over Central/Eastern Europe and their representation in CMIP5 General Circulation Models. In: 98th Annual Meeting of the American Meteorological Society. Konferencia helye, ideje: Austin, USA, 2018.01.07-2018.01.20. American Meteorological Society, Paper 331148.
- Kristóf, E., Barcza, Z., Hollós, R., Bartholy, J., Pongrácz, R.*, 2020: Evaluation of Historical CMIP5 GCM Simulation Results Based on Detected Atmospheric Teleconnections. *Atmosphere*, 11(7): 723. <https://doi.org/10.3390/atmos11070723>
- Neuwirth, E.*, 2014: RColorBrewer: ColorBrewer Palettes. R package version 1.1-2. <https://CRAN.R-project.org/package=RColorBrewer>
- Ooms, J.*, 2020: magick: Advanced Graphics and Image-Processing in R. R package version 2.5.1. <https://CRAN.R-project.org/package=magick>
- Peixoto, J.P., Oort, A.H.*, 1992: Physics of Climate. AIP-Press, USA, 492–497, ISBN: 978-0883187128.
- Pierce, D.*, 2019: ncd4: Interface to Unidata netCDF (Version 4 or Earlier) Format Data Files. R package version 1.17. <https://CRAN.R-project.org/package=ncdf4>
- Pokovai, K., Hollós, R., Bottyán, E., Kis, A., Marton, T., Pongrácz, R., Pásztor, L., Hidy, D., Barcza, Z., Fodor, N.*, 2020: Estimation of agro-ecosystem services using biogeochemical models. *Időjárás / Quarterly Journal of the Hungarian Meteorological Service*, 124(2): 209–225. <https://doi.org/10.28974/idojaras.2020.2.4>

- R Core Team*, 2020: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Ushey, K., Allaire, J.J., Tang, Y.*, 2020: reticulate: Interface to 'Python'. R package version 1.18. <https://CRAN.R-project.org/package=reticulate>
- Wallace, J.M., Gutzler, D.S.*, 1981: Teleconnections in the Geopotential Height Field during the Northern Hemisphere Winter. *Monthly Weather Review*, 109(4): 784–812. [https://doi.org/10.1175/1520-0493\(1981\)109<0784:TITGHF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2)
- Wikle, C. K., Zammit-Mangion, A., Cressie, N.*, 2019: Spatio-Temporal Statistics with R, Chapman & Hall/CRC, Boca Raton, USA. ISBN: 978-1138711136. <https://doi.org/10.1201/9781351769723>
- Xie, Y.*, 2014: knitr: A Comprehensive Tool for Reproducible Research in R. In: Stodden, V., Leisch, F., Peng, R.D., Eds. Implementing Reproducible Computational Research. Chapman and Hall/CRC, Boca Raton, USA. ISBN 978-1466561595
- Xie, Y., Allaire, J.J., Grolemund, G.*, 2018: R Markdown: The Definitive Guide. Chapman and Hall/CRC, Boca Raton, USA. ISBN 978-1138359338. <https://doi.org/10.1201/9781138359444>

### Internetes hivatkozások

- [1 – Comprehensive R Archive Network] <https://cran.r-project.org>
- [2 – bookdown] <https://bookdown.org/>
- [3 – GitHub repository] <https://github.com/Erzsi86/Met2020-data-processing-example>
- [4 – Advancing Reanalysis] <https://reanalyses.org/atmosphere/overview-current-atmospheric-reanalyses>
- [5 – Copernicus Climate Data Store] <https://cds.climate.copernicus.eu/#!/home>
- [6 – NCEP/NCAR Reanalysis 1] <https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html>
- [7 – NCEP/DOE Reanalysis 2] <https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.html>
- 

### ORCID

- Kristóf E.*  <https://orcid.org/0000-0001-9892-9552>
- Barcza Z.*  <https://orcid.org/0000-0002-1278-0636>