

Speciális meteorológiai adathalmaz statisztikai vizsgálata

Csanády Viktória

Soproni Egyetem

Informatikai és Matematikai Intézet

csanady.viktoria@uni-sopron.hu

ÖSSZEFOGLALÓ. A különböző meteorológiai adatok - így az ég-föld közötti villámok - időbeli és területi változása, adott területen a változás tendenciája különböző statisztikai módszerek alkalmazásával jellemezhető. Az alábbiakban 9 terület kerül összehasonlításra, valamint ezek közül egy terület 32 évi adatsorának modellezése kerül bemutatásra, Gauss függvények alkalmazásával.

ABSTRACT. The temporal and spatial changes of different meteorological data – such as lightning between the sky and the ground – and the tendency of the change in a given area can be characterized by using different statistical methods. In the following, 9 areas are compared, of which the modeling the 32-year data series of one area is presented using Gaussian functions.

1. Bevezetés

A különböző meteorológiai adatok, leggyakrabban hőmérséklet, csapadékmennyiség már számtalan esetben kerültek bele különböző biológiai vizsgálatokba, mint a vizsgálat tárgyának befolyásoló tényezői, vagy mint az alkalmazott modell független változója vagy változói. A klímaváltozás óta azonban, ma már nagy jelentőséggel bíró tényezők mellett számtalan más egyéb meteorológiai paraméter kutatása is elkerülhetetlen. A szélsőséges időjárási történések indokolják ezeket a kutatásokat, és felvetik annak a kérdését, miként lehet statisztikailag jellemezni ezeket a speciális adathalmazokat. Legtöbb jellemző esetén nem megismételhető mérési adatról van szó, hanem egyetlen adatról, ami lehet természetesen idő- vagy helyfüggő, akár mindkettő. Emiatt a legáltalánosabb közismert statisztikai jellemzők és vizsgálatok nem alkalmazhatók, illetve nem adnak számunkra érdemleges sokszor értelmes információt. Az alábbiakban egy a már említett tulajdonságú meteorológiai adathalmaz kerül terítékre, melynek vizsgálata különböző megközelítésből történik.

A vizsgálat tárgyát az ég-föld közti villámok száma képezi, Ausztria szövetségi tartományaiban 1992-től 2023-ig, az idősoros vizsgálat egy kiemelt szövetségi tartomány adatsorát érinti. Az adatsorok a <https://www.aldis.at> internetes oldal szabadon hozzáférhető adatbankjából származnak, ahonnan ingyenesen letölthetők.

Az alábbiakban először villámsűrűségi értékek kerülnek elemzésre szövetségi tartomány összehasonlításban, klaszteranalízissel. A villámsűrűség a villámcsapások átlagos száma km^2 -ként évente. Ezt követi, Stájerország, mint kiválasztott szövetségi tartomány időfüggő vizsgálata, a már említett időintervallumban, melynek során az évenként, havonta észlelt ég-föld villámok előfordulási gyakoriságai kerülnek elemzésre. A vizsgált 32 év, évenkénti adatsorának tendenciája haranggörbével jellemezhető, így alkalmasnak bizonyult a

transzformált Gauss görbe illesztése. A kapott eredményekből, az évenként illesztett modellek eltolási paraméterei ezek után összevethetők. Az illesztések pontosságának ellenőrzését a mindenkori korrelációs együttható, (nem lineáris korrelációs együttható) értéke szolgálja, figyelembe véve, hogy a vizsgált meteorológiai adatsor esetén a vizsgált érték alacsonyabb lehet, nem várható el olyan magas korreláció, mint pontosan mérhető fizikai jellemzőknél.

A vizsgálat célja a szövetségi tartományok esetében, hogy a villámsűrűségi adatok alapján csoportosítsa azokat, igazolt statisztikai módszerrel. A 32 év esetében a cél pedig az, hogy kimutassa, hogy a villámgyakoriságok éves tendenciája megfelel a feltételezett tendenciának, illetve a modellek paramétereinek elemzésével a szélsőséges évek statisztikailag kimutathatók legyenek.

Az adathalmaz vizsgálata során alkalmazott software a STATISTICA, a modellek illesztésénél a szükséges kezdőértékek az adatsorból jól becsülhetők.

A vizsgálat tárgya:

- az ausztriai szövetségi tartományok átlagos villámsűrűsége (2012-2021 évek átlaga) (www.aldis.at)
- ausztriai ég-föld villámlási adatok szövetségi tartományonként, 1992. január – 2023. július, havonkénti összesített esetszám. (www.aldis.at)

A vizsgált adathalmazok és az alkalmazott modellek:

Szövetségi tartomány	Villámsűrűség
Burgenland	1,27
Kärnten/Karintia	1,7
Niederösterreich/Alsó-Ausztria	1,05
Oberösterreich/Felső-Ausztria	1,2
Salzburg	1,28
Steiermark/Stájerország	1,82
Tirol	1,21
Vorarlberg	1,04
Wien/Bécs	0,97

1. táblázat. Villámsűrűség szövetségi tartományonként

Az alábbi adathalmaz esetében eltekintünk a teljes részletességtől az adatok megtekinthetők a fentiekben megadott internetcím alatt.

Év	Jan.	Febr.	Marc.	Ápr.	Máj.	Jún.	Júl.	Aug.	Szept.	Okt.	Nov.	Dec.
1992	0	0	0	560	1037	6584	7757	8828	3635	592	13	1
1993	0	0	8	518	5329	15648	21283	15363	1166	281	9	38
1994	19	5	10	494	3233	3365	13897	19855	930	64	2	0
1995	14	3	15	1460	3121	5594	15696	7207	1172	0	114	20
.
.
2020	27	11	5	115	428	1890	6649	7423	895	190	0	0
2021	0	0	2	31	294	3021	6404	4667	297	3	1	0
2022	0	0	0	47	3626	10658	4128	2567	725	3	2	0
2023	1	1	59	230	964	6438	5490					

2. táblázat. Regisztrált ég-föld villámok száma Stájerországban

Az alkalmazott regressziós modell:

- Transzformált Gauss függvény

- matematikai alakja:

$$y = \frac{b_3}{e^{(b_2(x-b_1))^2}} + b_0$$

- számítógépes alak:

$$\text{Var } n = b_3 / \exp(b_2 * (\text{NewVar} - 1 * b_1))^2 + b_0.$$

A modell rövid jellemzése:

Az alkalmazott Gauss függvény négy paraméteres általános modell, ami rendelkezik a szükséges eltolási és nyújtási paraméterekkel. A paraméterek értékeiből pontosan megadhatók a modell fő matematikai jellemzői, így a lokális szélsőérték helye és értéke. A négy paraméter kezdőértékének megadása az adathalmazból becsülhető szükség esetén. A változók jelentése az illesztések során:

- NewVar = Naptári hónapok sorszáma
- Var n = 1, 2, ..., 32 Az évek 1992-től 2023-ig (1992 esetén n=1).

2. Számított eredmények, kiértékelés

2.1. Villámsűrűségi átlagértékek klaszter eredményei

Az alábbi két táblázatban a klaszteranalízis eredményei olvashatók 3 klaszter esetén, elsőként a weboldal által megadott villámsűrűségi értékekre, ami tíz év (2012-2021) átlaga, majd pedig 32 év átlagai (szerző által számított) esetén az egyes szövetségi tartományokra nézve.

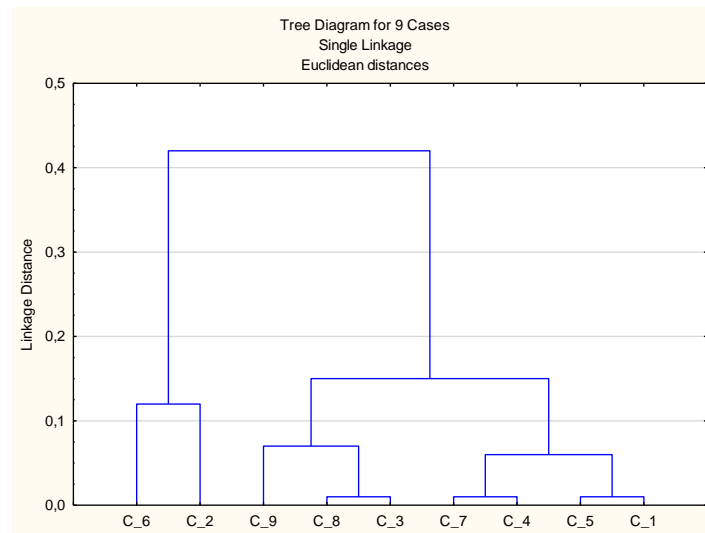
	Villámsűrűségi átlagok 10 év esetén			
	Átlag	Sorszám	Klaszter	Távolság
Burgenland	1,27	1	2	0,03
Kärnten/Karintia	1,7	2	3	0,06
Niederösterreich/Alsó-A.	1,05	3	1	0,03
Oberösterreich/Felső-A.	1,2	4	2	0,04
Salzburg	1,28	5	2	0,04
Steiermark/Stájerország	1,82	6	3	0,06
Tirol	1,21	7	2	0,03
Voralberg	1,04	8	1	0,02
Wien/Bécs	0,97	9	1	0,05

3. táblázat. Villámsűrűségi átlagok klasztere 10 év esetén

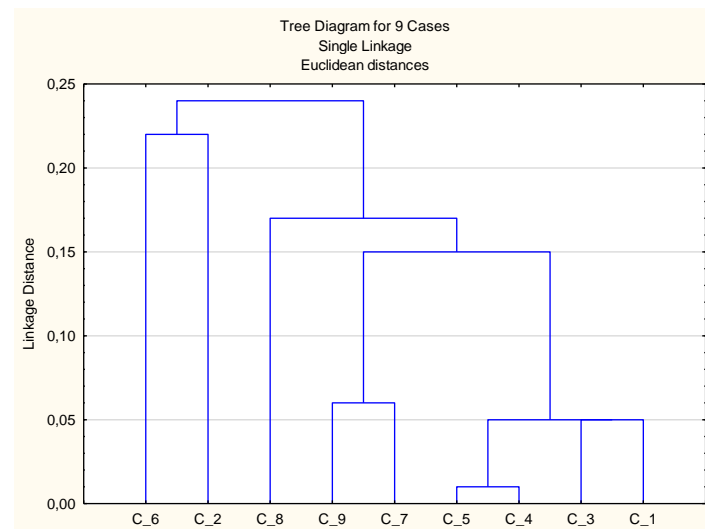
	Villámsűrűségi átlagok 32 év esetén			
	Átlag	Sorszám	Klaszter	Távolság
Burgenland	1,06	1	2	0,03
Kärnten/Karintia	1,36	2	3	0,06
Niederösterreich/Alsó-A.	1,01	3	2	0,03
Oberösterreich/Felső-A.	1,12	4	2	0,04
Salzburg	1,11	5	2	0,04
Steiermark/Stájerország	1,58	6	3	0,06
Tirol	0,86	7	1	0,03
Voralberg	0,63	8	1	0,02
Wien/Bécs	0,8	9	1	0,05

4. táblázat. Villámsűrűségi átlagok klasztere 32 év esetén

Mindkét adatsorra látjuk az alábbiakban a klaszterfát.

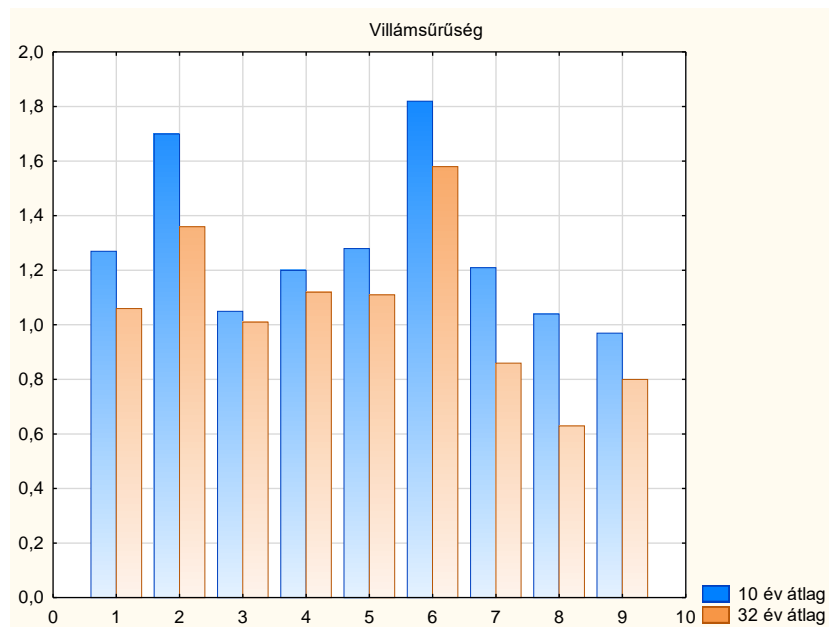


1. ábra. Villámsűrűségi átlagok klaszterfája 10 év esetén



2. ábra. Villámsűrűségi átlagok klasztere 32 év esetén

Az alábbi ábrán a klaszteranalízisben megadott szövetségi tartományi sorrendben szerepel két villámsűrűségi érték, késsel jelölve a tíz év átlagértékeit, mely értékek az Aldis oldalán nyilvánosak, valamint a megadott 32 év adataiból szerző által számítottak.



2. ábra. Villámsűrűségi átlagok

2.2. 32 év éves adatsorának regressziós vizsgálata Stájerország esetén

Elsőként feltüntetésre kerül a 32 év leíróstatistikai táblázata havi bontásban.

	Havi leíróstatistika a 32 évre				
	Évek száma	Átlagos esetszám	Minimum	Maximum	Szórás
Jan.	32	6,91	0,000	61,00	12,193
Febr.	32	5,66	0,000	47,00	10,114
Márc.	32	48,25	0,000	363,00	74,617
Ápr.	32	516,38	19,000	2668,00	612,743
Máj.	32	4882,84	294,000	26998,00	4937,903
Jún.	32	10189,75	1890,000	33784,00	6849,792
Júl.	32	13728,47	3529,000	33260,00	8791,817
Aug.	31	9537,58	2207,000	23173,00	5329,991
Szept.	31	1908,29	20,000	8407,00	2116,534
Okt.	31	220,29	0,000	1176,00	296,030
Nov.	31	15,90	0,000	114,00	27,800
Dec.	31	9,32	0,000	68,00	16,851

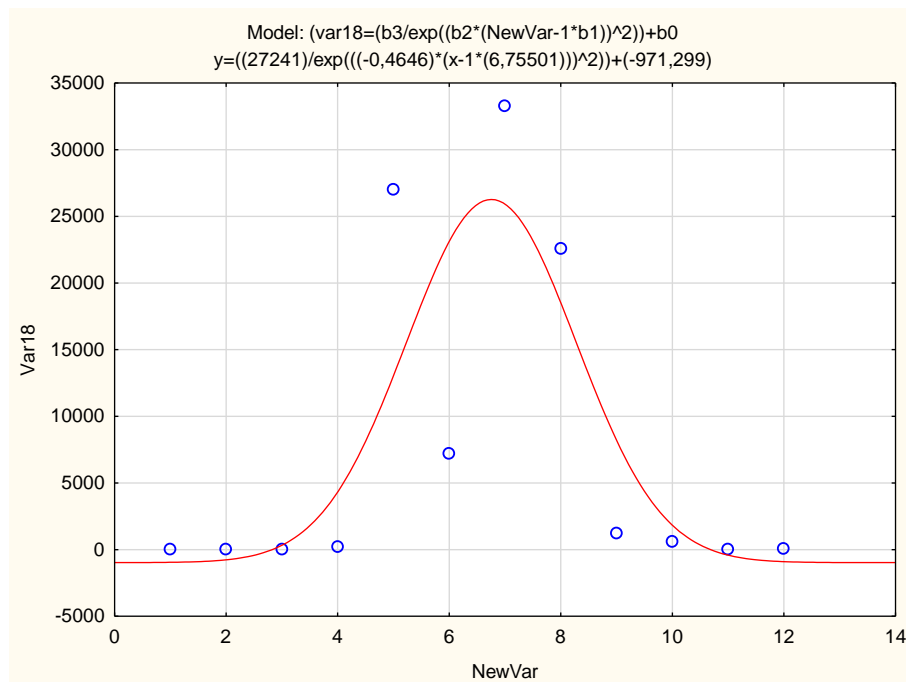
5. táblázat. Villámgyakoróságok statisztikai adatai havi bontásban

Az egyes illesztések eredményeinek részletes feltüntetésétől, az illesztések ábráinak bemutatásától a 32 évre nézve eltekintünk, a legkedvezőbb és legkedvezőtlenebb eset viszont az alábbiakban bemutatásra kerülnek.

A kiválasztás a korrelációs együttható értéke alapján történik, leszámítva az utolsó csonka évet, ahol az R értéke 0,9997, ami arra utal, hogy az adatsor illeszkedése nagyon jó, ezt viszont nagyban befolyásolja az utolsó 5 hónap adatának hiánya. Elsőként a 2009-es év illesztési eredményét prezentáljuk, ami szélsőségesen gyenge korrelációs értéket mutat $R=0,8076$. Ezt követően 2019-es év kerül bemutatásra $R=0,9975$ értékkel.

- A 2009-es év eredménye:

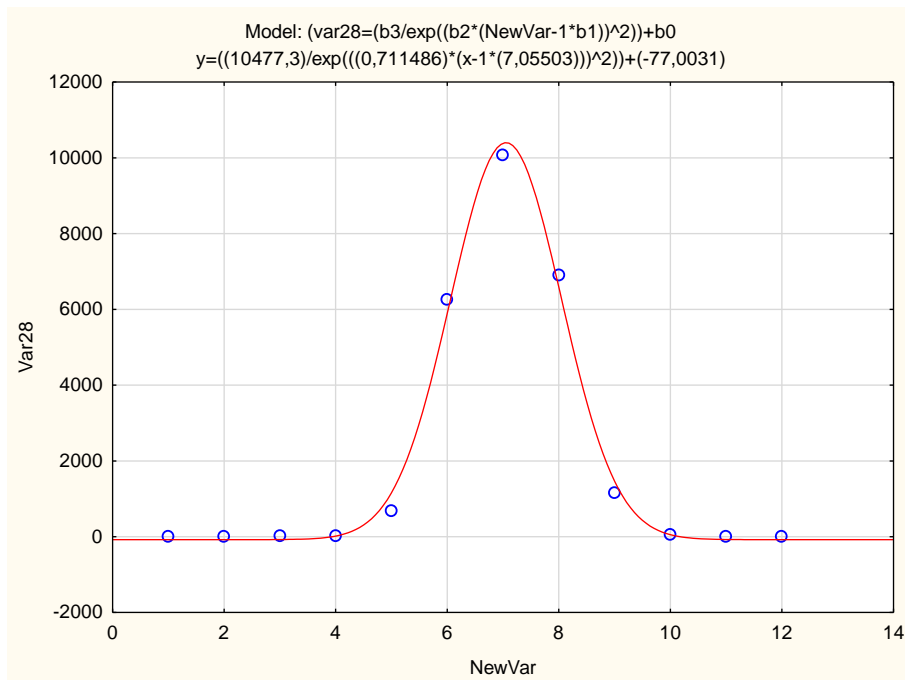
Model: $\text{var18}=(b3/\exp((b2*(\text{NewVar}-1*b1))^2))+b0$ (STtábla fordítva)				
Dep. var: Var18 Loss: (OBS-PRED)**2				
Final loss: 587530179,78 R= ,80761 Variance explained: 65,224%				
N=12	b3	b2	b1	b0
Estimate	27240,91	-0,464601	6,755001	-971,299



3. ábra. Havi villámgyakoróság 2009

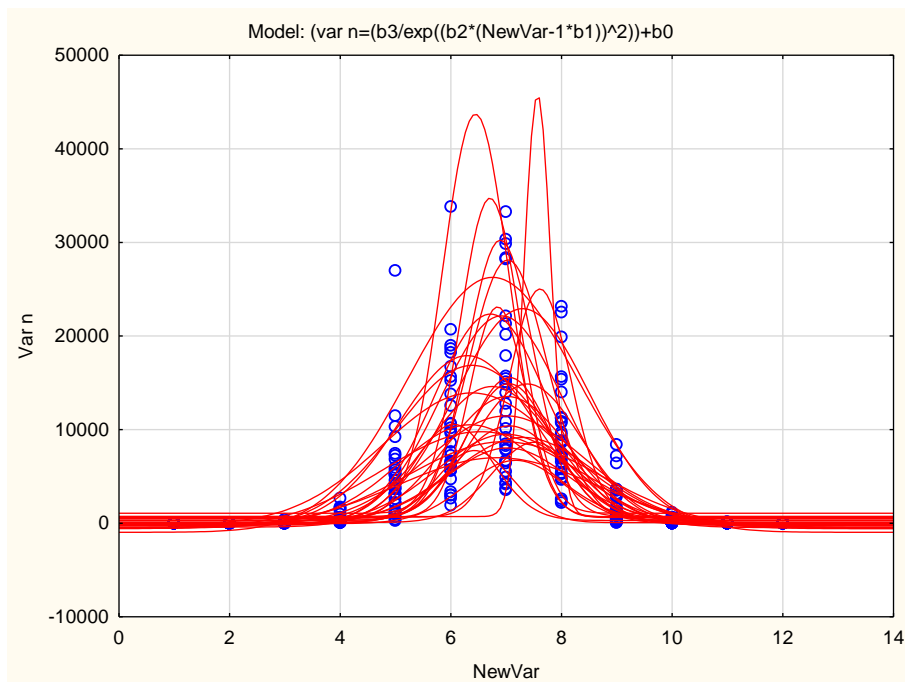
- A 2019-es év eredménye:

Model: $\text{var28}=(b3/\exp((b2*(\text{NewVar}-1*b1))^2))+b0$ (STtábla fordítva)				
Dep. var: Var28 Loss: (OBS-PRED)**2				
Final loss: 687022,44568 R= ,99750 Variance explained: 99,500%				
N=12	b3	b2	b1	b0
Estimate	10477,25	0,711481	7,055021	-77,0031



5. ábra. Havi villámgyakoriság 2019

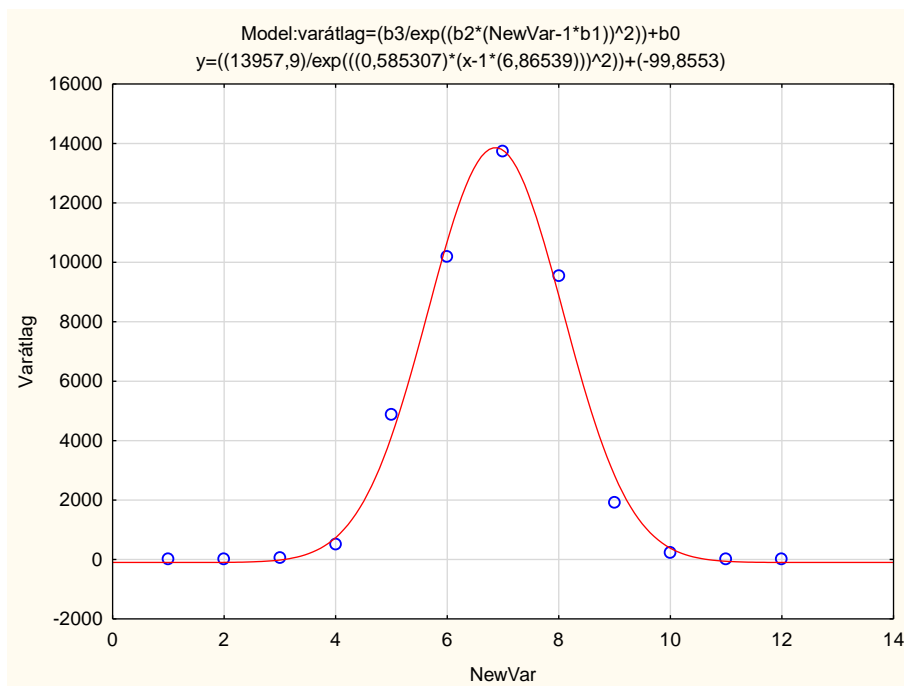
- 32 év illesztéseinek egyesített ábrája és az átlagolt adatokra illesztett modell eredményei:



6. ábra. 32 év egyesített ábrája

A 32 év átlagadataira történt illesztés korrelációs értéke magas, szoros illeszkedésre utal, a modell paraméterei lefedik a szélsőségeket.

Model: $\text{várátlag} = (b3/\exp((b2*(\text{NewVar}-1*b1))^2)) + b0$ (sthaviátlag)				
Dep. var: Var2 Loss: (OBS-PRED)**2				
Final loss: 2217498,3701 R= ,99589 Variance explained: 99,180%				
N=12	b3	b2	b1	b0
Estimate	13957,90	0,585307	6,865395	-99,8553



7. ábra. Illesztés az átlagadatokra

A 6. táblázat a kiválasztott szövetségi tartomány adataira történő 32 illesztés egy paraméterét, az eltolási paramétert (Maximumérték hely becsült) valamint a mindenkori tényleges értéket mutatja. R az egyes illesztések nem lineáris korrelációs együtthatója. Eltekintünk a modell többi paraméterének megadásától, ami a szerzőnél rendelkezésre áll.

Évek	Maximumérték helye becült (hónap)	Maximumérték helye tényleges (hónap)	Korrelációs együttható R
1992	7,31	8	0,9829
1993	6,92	7	0,9925
1994	7,61	8	0,9834
1995	7,06	7	0,9841
1996	6,35	6	0,8555
1997	6,99	7	0,9033
1998	6,72	7	0,993
1999	6,83	8	0,8362
2000	6,38	6	0,8869
2001	6,99	7	0,9492
2002	6,74	7	0,9881
2003	6,29	6	0,9718
2004	7,02	7	0,9746
2005	6,71	7	0,976
2006	6,44	6	0,9952
2007	6,39	6	0,8894
2008	7,29	8	0,9618
2009	6,76	7	0,8076
2010	7,03	7	0,9938
2011	6,96	8	0,8874
2012	6,89	7	0,9927
2013	7,57	8	0,899
2014	7,02	7	0,9833
2015	6,84	7	0,9745
2016	7,05	7	0,9911
2017	7,38	8	0,9336
2018	6,53	6	0,8432
2019	7,06	7	0,9975
2020	7,51	8	0,9942
2021	7,17	7	0,996
2022	6,05	6	0,9765
2023	6,42	6	0,9997

6. táblázat. Illesztett függvények maximum helyei és korrelációs együtthatói

A 32 évre az illesztés során kapott eltolási paraméter és korrelációs együttható néhány leíróstatisztikai jellemzői a következők:

	Leíróstatisztika (max. hely és R)						
	Elemzés	Átlag	Konf. -95%	Konf. 95%	Minimum	Maximum	Szórás
Max. hely	32	6,88	6,74	7,02	6,05	7,61	0,389
R	32	0,9498	0,9294	0,9702	0,8076	0,9997	0,05658

7. táblázat. A maximum helyek és R értékek statisztikái

2.3. Elemzés, értékelés

Villámsűrűségi adatok klaszter elemzésének értékelése:

A vizsgálat első lépésben 2012-2021 évek átlagainak felhasználásával indult, majd pedig 1992-2023 évek átlagaival folytatódott, mindkét esetben a 8 szövetségi tartomány és Bécs területére vonatkozóan. A számítások során alkalmazásra került a K-means clustering, aminél a választott klaszterszám mindkét esetben 3 volt. Ezt követően pedig a Joining (tree clustering), ugyancsak mindkét adathalmazra, ami a dendrogramot, klaszterfát adja.

A K-means clustering igazolja azt az adatokból is szembetűnő feltételezést, hogy a szövetségi tartományok közül két tartomány magas villámsűrűségi értékével eltér a többitől. A három klaszter csoport viszont indokolt. Összehasonlítva a vizsgált 10 és 32 éves átlagadatsort nem mutatkozik a klaszter csoportokban jelentős eltérés, mindössze két tartomány cserél helyet az 1 és 2 csoportokra nézve. Ez arra utal, hogy az átlagértékek változása nem jelentős az évek számát (10 illetve 32 év) illetően, az egyes klaszter csoportokba sorolt tartományok kettőt leszámítva fix helyükön maradtak. A kiugró legmagasabb értékekkel bíró tartományok tehát mindkét esetben Karintia és Stájerország maradt, szignifikánsan eltérve a többi tartomány értékétől.

A Gauss görbe illesztésének eredményei és értékelése:

A kiválasztott szövetségi tartomány, Stájerország 32 éves adatsorára történő illesztések a maximális villámgyakoriság értékekhez tartozó, időpont értékekre adnak becslési lehetőséget. A becsült értékek a transzformált Gauss görbe paraméteréből adóttak. A maximumérték helyek (6. táblázat) statisztikai elemzése azt mutatja, hogy a legmagasabb átlagos villámsűrűségi érték 95%-os konfidencia szinten a hatodik hónap második felétől a hetedik hónap elejéig várható.

Az egyes illesztések korrelációs együtthatójának értéke, az adatsor várható ingadozása ellenére magas, a legkisebb $R=0,8076$ a legnagyobb $R=0,9975$. Itt eltekintünk a csonka 2023 év eredményétől, ami $R=0,9997$. 95%-os konfidencia szinten az R alsó határa 0,9294. Figyelembe véve a felsorolt értékeket kijelenthető, hogy a választott regressziós modell jól jellemezi a vizsgált trendet.

Az alábbi táblázat összefoglalja azon éveket, amelyek esetében az R értéke alacsonyabb az R alsó konfidencia határánál.

Évek	R
1996	0,8555
1997	0,9033
1999	0,8362
2000	0,8869
2007	0,8894
2009	0,8076
2011	0,8874
2013	0,8990
2018	0,8432

8. táblázat. Gauss modell R összefoglaló

A felsorolt 9 év bizonytalanságot mutat, azonban illesztési pontossága jónak mondható, ha figyelembe vesszük az adatsor jellegét.

Az illesztett modell X irányú eltolási paramétere a 32 évre nézve 13 évben mutatott a tényleges maximumérték helytől 5%-os hibánál nagyobb hibaértéket, két esetben haladta meg a 10%-t 14,6% és 13% értékkel, 1999 és 2011 esetén, amit igazol a korrelációs együttható gyengébb értéke is.

A 32 illesztés ábrája nem került egyenként bemutatásra, de a leggyengébb korrelációs együttható esete 2009 $R=0,8076$ jól mutatja, hogy az anomáliát, a nyári hónapok villámgyakoriságának véletlenszerű ingadozása okozza. A 8. táblázatban felsorolt évek esetén is ez volt jellemző, ami egyértelműen utal az adott év időjárási szélsőségesére. Természetesen szerepet játszhat a vizsgálat hibájában az is, hogy a 32 év során alkalmazott megfigyelő állomások technikai adottságai jelentős mértékben fejlődtek, az észlelések száma az idő múltával pontosabban regisztrált.

Összességében azonban kijelenthető, hogy a villámgyakoriság éves trendje jól jellemezhető az alkalmazott modell segítségével, a maximumérték helye a vizsgált 32 éves adatsor alapján jól becsülhető, a szélsőséges évek kimutathatók.

Összefoglaló

A klímaváltozás napjainkban egyre indokoltabbá teszi, hogy speciális meteorológiai adatsorok statisztikai jellemzését is elvégezzük. Ezen adatsorok nem reprodukálhatók, egyedi mérések, melyek lehetnek hely és időfüggők. A vizsgálat tárgya a villámsűrűség, adott régiókban illetve a villámgyakoriság, adott területen idő függvényében. A felhasznált statisztikai elemzés a klaszteranalízis, aminek segítségével az egyes régiók csoportosíthatók, illetve a regresszióanalízis, melynek révén alkalmas modell segítségével jellemezhetjük a változó trendjét, ezt követően, pedig az alkalmazott modell paraméteréből megfelelő becslést végezhetünk.

Az eredmények azt mutatják, a villámsűrűsége vonatkozólag, hogy a vizsgált szövetségi tartományok csoportosíthatók, 3 csoportba, mind a 10, mind a 32 évi adatsor átlagai esetén. Nincs jelentős eltérés a két adatsor vizsgálata között, mindkét esetben a két legnagyobb értékkel jellemezhető tartomány szignifikánsan eltér a többitől.

A villámgyakorisági adatok 32 évi vizsgálatának illesztései azt igazolják, hogy a trend jól jellemezhető egy transzformált Gauss függvényvel. A magas korrelációs együtthatók ellenére a maximumértékek helyének becslésében előfordulnak kisebb eltérések a tényleges értékektől. Az említett magas R érték viszont alkalmas arra, hogy általa kiszűrjük a szélsőséges éveket.

Irodalomjegyzék

- [1] **Csanády, V., Horváth-Szováti, E., Szalay, L.**, Alkalmazott statisztika, Sopron, Nyugat-Magyarországi Egyetem Kiadó (2013), 175p..
- [2] AGES - Österreichische Agentur für Gesundheit und Ernährungssicherheit GmbH, <https://www.aldis.at>
- [3] **Csanády, V.**, Időjárás elemzés regressziós eljárás alkalmazásával, Dimenziók, Matematikai Közlemények III. (2015) 25-34. doi:10.20312/dim.2015.04