

**UNIVERSITATEA TEHNICĂ DE CONSTRUCȚII DIN
BUCUREȘTI
ȘCOALA DOCTORALĂ**

Raport 1

REGIUNEA DE STUDIU ȘI ANALIZA DATELOR

Doctorand Youssef Saliba

Conducători doctorat

Conf. Dr. Hab. Ing. Cristian-Ștefan Dumitriu

Prof. Dr. Hab. Alina Bărbulescu

Decembrie 2024

Cuprins

1. Regiunea de studiu	2
2. Metodologie	3
2.1 Setul de date	3
2.2 Statistici de bază	3
2.3. Transformarea datelor	4
2.4. Date originale și transformate	5
2.5 Teste parametrice și neparametrice	5
2.6 Analiza statistică a seriilor de precipitații medii lunare.....	5
2.7 Analiza statistică a varianței.....	7
2.8 Analiza punctelor de schimbare	7
2.9 MDL.....	8
2.10 Simulated annealing	9
3. Rezultate și discuții	10
3.1 Seriile de precipitații transformate	10
3.2 Analiza valorilor medii.....	11
3.3 Analiza de varianță	14
3.4 Comparison.....	15
3.4 Detectarea punctelo de schimbare cu MDL	17
4. Concluzii	19
Bibliografie.....	20

1. Regiunea de studiu

Regiunea Dobrogea este predispusă la evenimente meteorologice severe din cauza topografiei sale și a proximității față de Marea Neagră și cursul inferior al Dunării, în sud-estul României. În cercetările anterioare au fost utilizate diverse metodologii statistice și de modelare, inclusiv analize de tendință, modele probabilistice, tehnici de interpolare spațială și teste de omogenitate, pentru a studia dinamica climatică regională, tendințele temperaturii, variabilitatea precipitațiilor, tiparele secetei și modificările hidrologice. Aceste studii au reșit să evidențieze variabilitatea și schimbările pe termen lung.

Prin analiza evoluției climatice utilizând date de temperatură de la mai multe stații meteorologice din Dobrogea, Maței și Bărbulescu (2008) au adus o contribuție semnificativă. Ca urmare a tendințelor mai largi de încălzire din Europa, cercetarea lor a arătat că temperaturile medii anuale au crescut cu 0,8 °C în intervalul 1997–2004. Aceștia au verificat modificările din tiparele meteorologice folosind testele de amplitudine Pettitt și Buishand. În plus, interpolarea spațială GIS a arătat că temperaturile regionale erau în creștere, demonstrând cât de semnificative au fost aceste schimbări climatice.

Bărbulescu (2016) a utilizat date din perioada 1965–2005 pentru a implementa Metoda Precipitațiilor Celei Mai Probabile (MPPM) în Dobrogea, introducând o tehnică nouă pentru estimarea regională a precipitațiilor. Pentru a aborda factorii climatici diverși din regiune și problema distribuției neuniforme a pluviometrelor, MPPM ia în considerare atât dimensiunea spațială, cât și cea temporală a datelor de precipitații și folosește o metodologie bazată pe frecvență pentru obținerea valorilor reprezentative de precipitații. Această metodă a redus erorile MAE și MSE în comparație cu alte metode tradiționale (cum ar fi IDW, TP, OK), dovedind că poate înregistra fidel tiparele regionale din această zonă complex.

În alt studiu, Bandoc și Prăvălie (2015) au clarificat modul în care schimbările hidrometeorologice afectează managementul resurselor de apă. Aceștia au examinat modificările bilanșului climatic al apei (CWB) la nivel anual și sezonier în nouă amplasamente, din 1961 până în 2009, utilizând testul Mann–Kendall, panta lui Sen și modele de distribuție Pearson, Gamma, Chi-pătrat și Wakeby. Ei au concluzionat că Sulina prezintă cel mai nefavorabil rezultat (-3,2 mm/an) din nordul Dobrogei, unde deficitul de apă se agravează.

În plus, pentru a analiza intervalele aride din regiune, Chelcea ș.a. (2016) au utilizat indici de secetă precum Indicele Standardizat de Precipitații (SPI) și Indicele Standardizat de Precipitații-Evapotranspirație (SPEI). Datorită dependenței sale exclusive de precipitații pentru măsurarea secetei, SPI este bine adaptat pentru secetele meteorologice de scurtă durată. SPEI integrează precipitațiile și evapotranspirația potențială pentru a lua în considerare bilanșul apei, ceea ce face evaluarea secetei mai cuprinzătoare. Ei au concluzionat că partea nordică a Dobrogei este vulnerabilă din cauza precipitațiilor reduse și a evapotranspirației ridicate, secetele meteorologice moderate până la severe fiind cele mai frecvente.

Bărbulescu și Popescu-Bodorin (2019) au demonstrat fiabilitatea înregistrărilor de precipitații pe perioade lungi de timp utilizând independența statistică a datelor istorice și actuale de precipitații din Dobrogea. În acest scop, datele lunare de precipitații de la 49 de stații meteorologice pe o perioadă de 41 de ani au fost analizate folosind testul Székely–Rizzo–Bakirov (SRB) și testul de independență Daugman. Rezultatele au fost inconsecvente între estimările de precipitații trecute și viitoare, ceea ce sugerează că Fluviul Dunărea, Marea Neagră și curenții de aer regionali joacă un rol în modelarea climatului Dobrogei. Descoperirile arată că prognozarea pe baza precipitațiilor istorice nu este de încredere în acest climat complex.

Mai mult, Bărbulescu (2023) a utilizat două variante MPPM pentru a analiza seriile de temperatură din Dobrogea, România, din 1965 până în 2005, și pentru a genera tendințele regionale de

temperatură. O analiză a temperaturilor anuale minime, medii și maxime înregistrate la zece stații meteorologice a reieșit tendințe de încălzire în toate regiunile, în special pentru temperaturile minime înregistrate după 1990, arătând în cele din urmă modul în care Delta Dunării și Marea Neagră influențează variabilitatea climatică din Dobrogea.

Combinând metoda Hershfield (PMP_H) cu metoda spațio-temporală Diaconu (PMP_D) și utilizând date de la 10 stații meteorologice, Bărbulescu și Maftei (2023) au evaluat Precipitația Maximă Probabilă (PMP) în Dobrogea, România, pe o perioadă de 41 de ani. Studiul a abordat provocările asociate cu prognozarea precipitațiilor extreme pentru proiectarea infrastructurii hidrotehnice și managementul riscului la inundații, subliniind climatul semiarid distinct al Dobrogei, caracterizat prin secete prelungite și episoade intermitente de ploi torențiale.

Bărbulescu ș.a. (2021) au dezvoltat *Climate Analyzer*, un set de instrumente bazat pe Matlab pentru estimarea precipitațiilor regionale folosind două variante ale MPPM, acesta fiind utilizat pentru a analiza 41 de ani de date anuale de precipitații de la 10 stații meteorologice din Dobrogea. Analizorul folosește o analiză a subintervalelor bazată pe frecvență (Metoda I) și clusterizarea *k*-means (Metoda II), care a depășit Metoda I prin maximizarea omogenității clusterelor. Acest instrument se adaptează la variabilitatea spațială și temporală a precipitațiilor din Dobrogea, fiind important pentru managementul resurselor de apă și modelarea climatică în zone cu climat complex, cum este Dobrogea.

Utilizând date de precipitații de la 46 de stații meteorologice din Dobrogea, acoperind perioada 1965–2005, acest studiu își propune să realizeze o analiză comparativă a mediilor și varianțelor între stații, pentru a identifica în final tiparele spațiale și temporale ale variabilității precipitațiilor în regiunea Dobrogea. Non-normalitatea din serii a fost abordată prin aplicarea transformării Yeo–Johnson, permițând utilizarea testelor parametrice pentru majoritatea datelor. Diferențele dintre valorile medii au fost testate folosind ANOVA și testul post-hoc Student–Newman–Keuls (SNK). Testul Levene și echivalentele sale non-parametrice au fost utilizate pentru a verifica omogenitatea varianței. În plus, un cadru bazat pe Lungimea Minimă de Descriere (MDL), optimizat prin călire simulată Simulated Annealing (SA), este utilizat pentru detecția punctelor de schimbare (change points), identificând numărul și pozițiile decalajelor semnificative din cadrul seriilor.

2. Metodologie

2.1 Setul de date

Setul de date utilizat în acest studiu include date lunare de precipitații de la 46 de stații meteorologice distribuite în întreaga Dobroge (Figura 1). Acesta acoperă o durată de patru decenii, din 1965 până în 2005. Având Munții Carpați drept barieră orografică la nord-vest, stațiile prezintă o varietate de influențe climatice generate de interacțiunea sistemelor continentale, mediteraneene, oceanice și scandinavo-baltice. Climatul semiarid al Dobrogei se reflectă în setul de date printr-o precipitație medie anuală de mai puțin de 400 mm și o variabilitate regională ridicată, unde zona mai înaltă a Munților Măcin din nord primește mai multe precipitații decât zonele mai joase din sud.

2.2 Statistici de bază

Analiza datelor evidențiază o variabilitate spațială accentuată. Precipitația medie lunară variază între 21,8 mm la Sulina și 49,2 mm la Niculițel, indicând variații climatice locale. În plus, valorile mediane indică distribuții asimetrice la dreapta (pozitiv), în care evenimentele extreme de precipitații influențează indicatorii de medie. De asemenea, se detectează o variație semnificativă a varianței precipitațiilor, cuprinsă între 382,3 și 1444,3 mm². Stațiile Niculițel, Albești și Dăeni prezintă valori mai ridicate, ceea ce indică o variabilitate mai mare a precipitațiilor. Niculițel

înregistrează cel mai mare interval intercuantic (IQR), de 44,75, sugerând o dispersie mai mare a valorilor lunare.

Fiecare stație prezintă o asimetrie (skewness) pozitivă, cele mai mari valori fiind înregistrate la Mangalia, Pecineaga și Albești, ceea ce indică o frecvență mai mare a evenimentelor extreme de precipitații. Deoarece majoritatea stațiilor au valori ale boltirii (kurtosis) mai mari de 3, distribuțiile leptocurtice sunt mai predispuse la valori extreme. Testul global Kolmogorov–Smirnov indică o distribuție non-normală, cu o asimetrie globală de 1,852 și o boltire de 6,058, în timp ce testul Shapiro–Wilk respinge ipoteza de normalitate pentru toate stațiile.

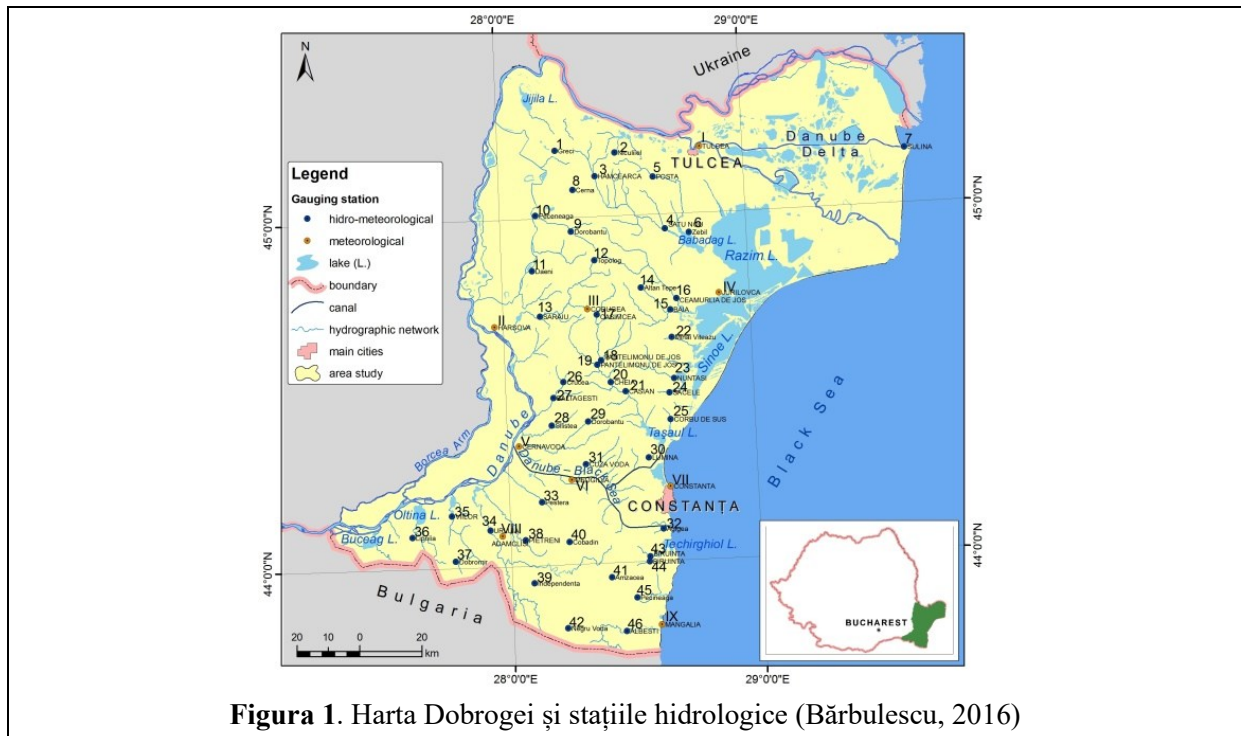


Figura 1. Harta Dobrogei și stațiile hidrologice (Bărbulescu, 2016)

2.3. Transformarea datelor

Atunci când seria originală nu urmează o distribuție normală, o transformare este necesară pentru o analiză parametrică precisă. Se poate aplica transformarea Yeo–Johnson (Yeo și Johnson, 2000), o variantă a transformării Box–Cox. Ecuația acestei transformări este:

$$\psi(\lambda, y) = \begin{cases} \frac{(y+1)^\lambda - 1}{\lambda}, & \text{dacă } \lambda \neq 0, y \geq 0 \\ \ln(y+1), & \text{dacă } \lambda = 0, y \geq 0 \\ -\frac{(-y+1)^{2-\lambda} - 1}{2-\lambda}, & \text{dacă } \lambda \neq 2, y < 0 \\ -\ln(-y+1), & \text{dacă } \lambda = 2, y < 0 \end{cases} \quad (1)$$

unde y reprezintă seria originală de precipitații, în timp ce $\psi(\lambda, y)$ este valoarea transformată.

Cel mai bun parametru al transformării, λ , este identificat folosind estimarea de verosimilitate maximă (maximum likelihood estimation). În schimb, normalitatea datelor transformate a fost evaluată utilizând testul Lilliefors (Dallal și Wilkinson, 1986). Statistica testului este:

$$D = \max |F_n(x) - F_0(x)| \quad (2)$$

unde $F_n(x)$ este funcția de repartiție empirică (ECDF) a datelor transformate, iar $F^*(y)$ este funcția de repartiție cumulativă (CDF) a unei distribuții normale estimate din eșantion. În ipoteza nulă a testului Lilliefors, se presupune că datele urmează o distribuție normală.

2.4. Date originale și transformate

Examinarea datelor brute fără nicio transformare păstrează caracteristicile originale ale seriilor de precipitații, reflectând variabilitatea observată în lumea reală. Proprietățile distribuționale naturale sunt deosebit de apreciate în studiile climatice și de mediu. Cercetările anterioare au demonstrat că seturile de date ecologice și de mediu beneficiază de pe urma menținerii datelor în formatul lor original. Acest lucru se datorează faptului că variabilitatea din aceste seturi de date conține informații intrinseci, utile atât pentru factorii de decizie, cât și pentru modelarea predictivă (Yeo și Johnson, 2000).

Fiabilitatea concluziilor statistice în condiții variabile ale datelor poate fi evaluată prin aplicarea simultană a testelor non-parametrice pe seturile de date originale și pe cele transformate. Dacă rezultatele obținute din datele originale și cele transformate sunt consecvente, acest lucru indică faptul că concluziile sunt precise, iar transformarea probabil nu a influențat interpretarea. Dimpotrivă, deoarece transformările îmbunătățesc frecvent validitatea analizelor parametrice prin reducerea unor probleme precum asimetria sau heteroscedasticitatea, discrepanțele semnificative dintre rezultate sugerează că transformarea a afectat puterea statistică sau ipotezele privind distribuția datelor (Helsel ș.a., 2020). Această strategie este utilă deoarece transformările ar putea să nu fie necesare dacă metodele non-parametrice oferă rezultate de încredere în ciuda naturii non-normale sau heteroscedastice a datelor (Wilks, 2020).

2.5 Teste parametrice și neparametrice

Atunci când condițiile de normalitate și homoscedasticitate sunt îndeplinite, testele parametrice au prioritate în această analiză. Atunci când se utilizează metode parametrice precum testele t și ANOVA, se presupune că datele urmează o distribuție normală cu varianțe omogene între grupuri. În aceste circumstanțe, testele parametrice sunt mai utile pentru analiză deoarece sunt sensibile la informațiile distribuționale, putând astfel detecta diferențe mai mici în varianță sau medie (Gatz și Smith, 1995). Ipoteza unei structuri suplimentare poate îmbunătăți precizia inferenței, iar această sensibilitate este deosebit de utilă în eșantioanele mai mari sau în seturile de date care sunt aproape normale.

Pe de altă parte, precizia poate fi compromisă atunci când ipotezele nu sunt îndeplinite, deoarece non-normalitatea și heteroscedasticitatea introduc erori sistematice (bias-uri) (Box, 1953). În astfel de cazuri, testele non-parametrice, care nu se bazează pe ipoteze distribuționale specifice, oferă o alternativă mai bună. Metode precum testul Mann–Whitney U pentru comparații pereche (Nachar, 2008) și testul Fligner–Killeen (Fligner, 1979) pentru evaluări globale sunt aplicate pentru a garanta că concluziile statistice rămân valabile într-o varietate mai largă de condiții ale datelor.

2.6 Analiza statistică a seriilor de precipitații medii lunare

Pe baza rezultatelor transformării, stațiile au fost clasificate ca fiind „normale” sau „non-normale”, iar ulterior analiza diferențelor dintre precipitațiile medii s-a efectuat separat pentru fiecare categorie, atât la scară locală (în perechi), cât și la scară globală. Pentru stațiile „normale”, am utilizat teste t (Student, 1908) pentru a testa ipoteza nulă conform căreia mediile stațiilor i și j sunt egale. O modalitate de a determina statistica t este prin:

$$t = \frac{\bar{X}_i - \bar{X}_j}{\sqrt{\frac{s_i^2}{n_i} - \frac{s_j^2}{n_j}}} \quad (3)$$

unde $\bar{X}_i$ și $\bar{X}_j$ sunt mediile eșantioanelor, s_i^2 și s_j^2 sunt dispersiile eșantioanelor, iar n_i și n_j sunt volumele eșantioanelor de la stațiile i și j , respectiv.

Pentru a face o analiză globală a diferențelor dintre medii, se utilizează analiza univariată a varianței (ANOVA univariată) (Fisher, 1992). ANOVA testează ipoteza nulă conform căreia toate mediile stațiilor sunt egale, fiind reprezentată prin statistica F :

$$F = MS_{between}/MS_{within} \quad (4)$$

$MS_{between}$ este pătratul mediu între grupuri și MS_{within} pătratul mediu în interiorul grupurilor.

Comparațiile post-hoc în perechi utilizând testul Student–Newman–Keuls (SNK) (Abdi și Williams, 2010) ajută la identificarea stațiilor specifice ale căror medii diferă, în cazul în care ANOVA indică diferențe semnificative. Acest lucru se realizează utilizând mediile ordonate ale grupurilor și valorile critice pentru diferențe calculate prin:

$$q = \frac{\bar{X}_{(r)} - \bar{X}_{(s)}}{\sqrt{\frac{MS_{within}}{n}}} \quad (5)$$

unde $\bar{X}_{(r)}$ și $\bar{X}_{(s)}$ sunt mediile grupurilor ierarhizate de la cea mai mare la cea mai mică, iar n este dimensiunea eșantionului pentru fiecare grup.

Pentru stațiile „non-normale”, au fost aplicate metode non-parametrice. Testul Mann–Whitney U (Mann și Whitney, 1947; Fay și Proschan, 2010) a fost utilizat pentru a evalua diferențele în perechi ale precipitațiilor medii dintre două stații. Ipoteza nulă este că distribuțiile precipitațiilor la cele două stații sunt identice. Statistica U este dată de

$$U = \min(U_i, U_j) \quad (6)$$

$$U_i = n_i n_j + \frac{n_i(n_i + 1)}{2} - R_i \quad (7)$$

$$U_j = n_i n_j + \frac{n_j(n_j + 1)}{2} - R_j \quad (8)$$

unde n_i și n_j sunt volumele eșantioanelor colectate la stațiile i și j , respectiv, și R_i , R_j sunt rangurile nivelurilor pentru cele două stații, combinate.

O analiză ANOVA (Welch, 1951) a fost efectuată pentru a face o comparație globală, cu varianțe inegale. Statistica Welch, F^* , pentru K grupuri, este definită prin:

$$F^* = \frac{\sum_{k=1}^K \frac{\omega_k (\bar{X}_k - \bar{X})^2}{(K-1)}}{1 + \frac{2(K-2)}{K^2-1} \Lambda} \quad (9)$$

unde $\omega_k = \frac{n_k}{s_k^2}$ este ponderea grupului k , bazată pe mărimea grupului n_k și varianțele s_k^2 , $\bar{X}_k$ este

media valorilor grupului k , $\bar{X} = \frac{\sum_{k=1}^K \omega_k \bar{X}_k}{\sum_{k=1}^K \omega_k}$ este media ponderată a tuturor grupurilor și $\Lambda =$

$\frac{(1 - \omega_k / \sum_{k=1}^K \omega_k)^2}{n_k - 1}$ reprezintă factorul de corecție pentru ajustarea varianțelor inegale.

Dacă sunt detectate diferențe semnificative, se aplică testul Dunn, cu corecția Bonferroni (Dunn, 1964) pentru comparații ale perechilor. Statistica testului este:

$$Z = \frac{R_i - R_j}{\sqrt{\left(\frac{N(N+1)}{12}\right) \left(\frac{1}{n_i} + \frac{1}{n_j}\right)}} \quad (10)$$

unde, R_i și R_j reprezintă sumele rangurilor pentru grupurile comparate, N este numărul total de observații din toate grupurile, iar n_i și n_j sunt dimensiunile eșantioanelor celor două grupuri. Corecția Bonferroni ajustează valorile p prin multiplicarea acestora cu numărul de comparații, k , pentru a controla creșterea probabilității de a comite o eroare de Tip I (fals pozitiv) în cazul testelor multiple.

$$p_{adjusted} = \min(p \times k, 1) \quad (11)$$

2.7 Analiza statistică a varianței

Pe lângă verificarea normalității, stabilită în secțiunea anterioară, este necesară evaluarea homoscedasticității, sau a omogenității varianțelor, pentru a analiza variabilitatea precipitațiilor între stațiile meteorologice. Ca o precondiție pentru utilizarea testelor parametrice, homoscedasticitatea presupune că seriile de date de la fiecare stație prezintă niveluri similare ale varianței. Pe baza testării normalității (Subsecțiunea 2.6), am clasificat stațiile în „normale” sau „non-normale”, aplicând ulterior testele de omogenitate a varianței pe aceste grupuri pentru a stabili abordarea adecvată pentru fiecare subset.

În cadrul analizei locale, sunt efectuate comparații în perechi ale varianțelor. Pentru stațiile „normale”, testul Levene (Levene, 1960) evaluează egalitatea varianțelor prin analizarea abaterilor absolute ale observațiilor față de media grupului. În schimb, pentru stațiile „non-normale”, s-a aplicat testul Brown–Forsythe (Brown și Forsythe, 1974), o variantă a testului Levene în care se utilizează mediana în locul mediei ca măsură a tendinței centrale. Ambele teste sunt reprezentate prin ecuația unificată

$$W = \frac{(N - k)}{(k - 1)} \cdot \frac{\sum_{j=1}^k n_j (M_j - M)^2}{\sum_{j=1}^k \sum_{i=1}^{n_j} (Z_{ij} - M_j)^2} \quad (12)$$

unde $Z_{ij} = |Y_{ij} - Y_j|$, cu Y_j media grupului (pentru testul Levene) sau mediana sa (pentru testul Brown-Forsythe). M_j reprezintă media valorilor Z_{ij} pentru grupul j . M este media totală a valorilor M_j ale tuturor grupurilor k . n_j este volumul grupului j și N este numărul total de observații ale tuturor grupurilor.

Pentru evaluarea globală, se aplică o analiză ANOVA unifactorială (Ecuația 4) în cazul stațiilor caracterizate prin „normalitate” și „homoscedasticitate”. Aceasta este urmată de testul post-hoc Student–Newman–Keuls (SNK) (Ecuația 5) dacă sunt detectate diferențe semnificative. În ceea ce privește stațiile care nu îndeplinesc oricare dintre aceste ipoteze, se utilizează testul non-parametric Fligner–Killeen (Fligner, 1979). Statistica testului, F , este calculată astfel:

$$F = \frac{\sum_{j=1}^k n_j (\text{rang}(Z_j) - \bar{R})^2}{\sum_{i=1}^N (\text{rang}(Z_i) - \bar{R})^2 / (N - 1)} \quad (13)$$

unde N reprezintă numărul total de observații ($N = \sum_{i=1}^k n_i$), n_j este dimensiunea eșantionului pentru grupul j , k este numărul de grupuri, $Z_i = |Y_i - \text{mediana}(Y)|$, $\bar{R}$ este media rangurilor.

Această metodă realizează ordonarea pe bază de ranguri pentru a atenua efectul valorilor aberante (outliers) și al non-normalității distribuționale (Conover ș.a., 1981). În cazul în care testul Fligner–Killeen detectează diferențe semnificative, se aplică testul Dunn cu corecția Bonferroni (Ecuația 10) pentru identificarea grupurilor între care există diferențe semnificative.

2.8 Analiza punctelor de schimbare

În domenii precum climatologia, unde modificările aparaturii de măsură, strămutările stațiilor sau factorii de mediu externi se corelează adesea cu schimbări structurale în seriile temporale, analiza punctelor de schimbare (change point) este esențială pentru interpretarea acestor fenomene. Detecția punctelor de schimbare este o provocare, în special în cazul seturilor de date care prezintă comportamente periodice și autocorelație, cum sunt înregistrările climatice. Ca răspuns la aceste dificultăți, un cadru bazat pe MDL (Minimum Description Length) dezvoltat pentru date climatice a fost prezentat de Lu ș.a. (2012). Această metodă poate detecta numeroase puncte de schimbare, luând în considerare ciclurile sezoniere și erorile autocorelate, deoarece combină penalizările pentru complexitatea modelului cu precizia inferenței bazate pe verosimilitate.

Metadatele, adică informațiile despre posibilele momente de schimbare derivate din istoricul stațiilor, au fost integrate ulterior în metodologia MDL de către Li ș.a. (2019). Pentru a îmbunătăți performanța de detecție atât în contexte multivariate, cât și univariate, acest cadru Bayesian MDL (BMDL) transpune cunoștințele de specialitate într-o structură probabilistică. În ceea ce privește omogenizarea datelor climatice, unde schimbările minore pot distorsiona evaluările tendințelor pe termen lung, astfel de inovații sunt esențiale. Pentru seriile zilnice de temperatură, granularitatea datelor necesită o eficiență computațională sporită, astfel încât paradigma MDL a fost adaptată în continuare de Hewaarachchi ș.a. (2017). În lucrarea lor, a fost dezvoltat și optimizat prin algoritmi genetici un model Bayesian de verosimilitate penalizată care ia în considerare autocorelația, variabilitatea sezonieră și complexitatea scării datelor.

Aceste completări extind cercetările fundamentale anterioare, cum ar fi cele realizate de Caussinus și Mestre (2004), care au fost primii ce au aplicat verosimilitatea penalizată la problema omogenizării seriilor climatice, și de Menne și Williams (2005), care au rezolvat problema punctelor de schimbare nedocumentate prin combinarea testării statistice cu serii de referință.

Metodele contemporane bazate pe MDL au depășit tehnicile convenționale de segmentare prin utilizarea modelelor autoregresive și a distribuțiilor apriori ierarhice (hierarchical priors). Aceste metode oferă un cadru eficient din punct de vedere computațional și robust din punct de vedere statistic pentru analiza întreruperilor structurale în seturi de date complexe. Colectiv, aceste studii ilustrează o schimbare de paradigmă în analiza punctelor de schimbare, îmbunătățindu-i eficacitatea în detectarea și reducerea impactului decalajelor artificiale, păstrând în același timp integritatea evaluărilor climatice pe termen lung.

2.9 MDL

Pornind de la lucrările lui ș.a. (2012), adoptăm cadrul lor metodologic pentru a formula funcția obiectiv MDL în vederea detecției punctelor de schimbare în seriile temporale climatice. Această abordare modelează datele din seria temporală Y_t sub forma unei regresii liniare pe bucăți cu efecte sezoniere, permițând decalaje de nivel în momentele de apariție a punctelor de schimbare și luând în considerare erorile autocorelate. Mai exact, presupunem că datele urmează modelul:

$$X_{nT+\nu} = \mu_\nu + \alpha(nT + \nu) + \delta_{nT+\nu} + \varepsilon_{nT+\nu} \quad (14)$$

unde μ_ν reprezintă valoarea medie sezonieră la perioada ν , β este parametrul tendinței liniare, Δ_j surprinde decalajul mediei la punctul de schimbare τ_j , iar $\varepsilon_{nT+\nu}$ desemnează erorile autoregresive periodice (PAR) de medie zero cu varianță periodică σ_ν^2 .

Păstrând o potrivire optimă pe datele observate, funcția obiectiv MDL este concepută pentru a penaliza complexitatea modelului. Lungimea totală de descriere este dată de:

$$\begin{aligned} MDL(\mathcal{M}) = & \frac{1}{2} \sum_{i=1}^{m+1} \ln(\tau_i - \tau_{i-1}) + \frac{pT \ln(2d)}{2} + \ln(m) + \ln(p) \\ & + \frac{1}{2} \sum_{t=1}^N \ln(v_t) + \frac{1}{2} \sum_{t=1}^N \frac{(X_t - \hat{X}_t)^2}{v_t} + \sum_{i=2}^m \ln(\tau_i) \end{aligned} \quad (15)$$

unde $\sigma_{v(t)}^2$ reprezintă varianța reziduurilor, $\mathbf{p} = (p_1, \dots, p_S)$ desemnează ordinul autoregresiei periodice, iar m este numărul de puncte de schimbare. Valorile $\boldsymbol{\tau} = (\tau_1, \tau_2, \dots, \tau_m)$ indică momentele de apariție a punctelor de schimbare, unde $\tau_0 = 1$ reprezintă începutul seriei, iar $\tau_{m+1} = N + 1$ reprezintă sfârșitul seriei temporale. Indicii $N_j = \tau_j - \tau_{j-1}$ exprimă lungimea fiecărui segment delimitat de puncte de schimbare consecutive. Termenul $\hat{X}_t$ se referă la valoarea estimată/predisă la pasul următor pentru X_t , calculată utilizând parametrii modelului estimați pentru segmentul respectiv.

Prin descurajarea modelelor excesiv de complexe, termenii de penalizare au ca scop prevenirea supra-ajustării (overfitting). Termenul $\ln(\tau_i - \tau_{i-1})$ penalizează punctele de

schimbare care creează segmente foarte scurte, menținând o segmentare semnificativă din punct de vedere fizic și statistic. Penalizările suplimentare, cum ar fi $\ln(m) + \ln(p) + \sum_{i=2}^m \ln(\tau_i)$, iau în considerare numărul total de puncte de schimbare, în timp ce termenii de $\frac{pT\ln(2d)}{2}$ penalizează complexitatea introdusă de modelul erorilor autoregresive periodice (PAR). Acești termeni de penalizare sunt echilibrați în raport cu termenii de fidelitate a datelor $\frac{1}{2} \sum_{t=1}^N \ln(v_t) + \frac{1}{2} \sum_{t=1}^N \frac{(X_t - \hat{X}_t)^2}{v_t}$.

Pentru optimizare, parametrii modelului sunt reformulați într-o reprezentare matriceală compactă

$$\mathbf{X} = \mathbf{D}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (16)$$

unde $\mathbf{X}$ este vectorul de dimensiune $N \times 1$ al datelor observate, $\mathbf{D}$ este matricea care captează indicatorii sezonal, componentele trendului și schimbarea de regim, $\boldsymbol{\beta}$ este vectorul parametrilor modelului (incluzând μ_v , α , și δ), și $\boldsymbol{\varepsilon}$ este vectorul reziduurilor. Cele $m + 1$ regimuri sunt reprezentate complet în matricea de design, ceea ce permite o estimare eficientă a parametrilor prin metoda celor mai mici pătrate generalizate (GLS). Acești parametri sunt ulterior optimizați utilizând algoritmul SA.

2.10 Simulated annealing

Tehnica de optimizare iterativă cunoscută sub numele de Simulated Annealing (SA) se bazează pe fenomenul fizic de recoacere din știința materialelor. Scopul acestui proces este obținerea unei configurații cristaline de energie joasă prin încălzirea și apoi răcirea lentă a unui material pentru a minimiza defectele. Pentru problemele de optimizare combinatorie cu funcții obiectiv complexe, cum ar fi detectarea punctelor de schimbare în datele din seriile temporale de precipitații, metoda propusă inițial de Kirkpatrick și colab. (1983) oferă performanțe excelente.

SA începe prin inițializarea soluției cu o mulțime vidă de puncte de schimbare ($\boldsymbol{\tau} = \emptyset$) și calcularea scorului MDL corespunzător. Temperatura inițială, T_0 , este setată la o valoare ridicată pentru a permite o explorare largă a spațiului soluțiilor. Temperatura scade treptat conform unei scheme de răcire definită ca:

$$T_{k+1} = \alpha T_k \quad (17)$$

unde α este rata de răcire (o valoare cuprinsă între 0 și 1), iar k denotă numărul iterației. Ecuația (17) asigură tranziția de la faza de explorare la cea de exploatare pe măsură ce algoritmul avansează.

La fiecare iterație, este propusă o nouă soluție candidată prin modificarea mulțimii curente de puncte de schimbare $\boldsymbol{\tau}$ prin intermediul uneia dintre cele trei acțiuni selectate aleatoriu:

1. Adăugare: Un nou punct de schimbare este eșantionat aleatoriu din intervalul $2, \dots, n - 1$ (pentru a evita punctele de frontieră) și adăugat la $\boldsymbol{\tau}$, cu condiția să nu existe deja.
2. Eliminare: Un punct de schimbare existent este eliminat din $\boldsymbol{\tau}$, cu condiția ca mulțimea să nu fie vidă.
3. Modificare: Un punct de schimbare existent este înlocuit aleatoriu cu o nouă poziție, asigurând evitarea duplicatelor.

Soluția propusă este sortată pentru a menține ordinea cronologică, iar duplicatele sunt eliminate.

Scorul MDL al soluției propuse este apoi calculat; dacă noua soluție îmbunătățește scorul ($\Delta MDL = MDL_{proposed} - MDL_{current} < 0$), aceasta este acceptată imediat. În caz contrar, este acceptată probabilistic pentru a permite explorarea soluțiilor suboptimale, cu o probabilitate dată de:

$$P_{accept} = \exp\left(-\frac{\Delta MDL}{T}\right) \quad (18)$$

ceea ce permite algoritmului să evadeze din minimele locale prin acceptarea ocazională a unor soluții mai slabe la începutul procesului, impunând treptat selecția soluțiilor mai bune pe măsură ce

temperatura scade. La fiecare iterație, algoritmul păstrează evidența celei mai bune soluții întâlnite până în prezent în ceea ce privește scorul MDL. Acest lucru asigură că rezultatul final este soluția cu cel mai mic scor MDL, indiferent dacă a fost acceptată sau nu în timpul procesului iterativ. Procesul continuă până când se atinge numărul maxim de iterații sau temperatura devine neglijabilă. Rezultatul final constă în cea mai bună mulțime de puncte de schimbare și scorul MDL corespunzător.

Pseudocodul algoritmului SA, așa cum a fost implementat în acest studiu, este sintetizat după cum urmează:

- Inițializare: Start cu $\tau = \emptyset$, calculează MDL_{curent} , setează $\tau_{optim} = \emptyset$ și inițializează $T = T_0$.
- Iterație:
 - Propune o nouă soluție candidată τ_{propus} prin adăugare, eliminare sau modificare de puncte de schimbare.
 - Calculează $\Delta MDL = MDL_{propus} - MDL_{curent}$.
 - Acceptă τ_{propus} dacă $\Delta MDL < 0$ sau cu probabilitatea dată de ecuație altfel.
 - Actualizează τ_{optim} dacă MDL_{propus} este cel mai mic scor înregistrat până în prezent.
 - Redu T conform (17).
- Rezultat: Returnează τ_{optim} și scorul MDL corespunzător.

3. Rezultate și discuții

3.1 Seriile de precipitații transformate

Distribuția inițială a datelor de precipitații a prezentat un tipar asimetric la dreapta, așa cum este ilustrat în Figura 2 (stânga), diagrama Q-Q corespunzătoare (panoul din dreapta) confirmând suplimentar abaterea de la normalitate. Datele originale au indicat o asimetrie substanțială și cozi grele (pachete mari de valori la extreme), ceea ce indică faptul că testele statistice parametrice ar putea produce rezultate distorsionate din cauza non-normalității.

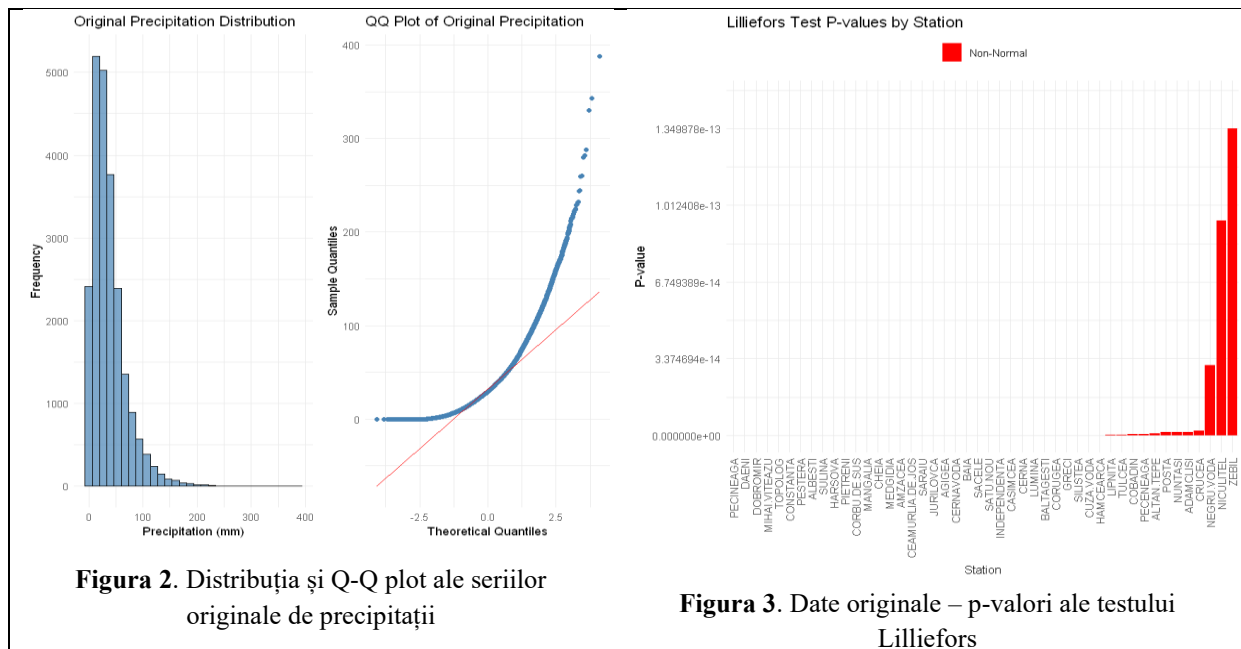


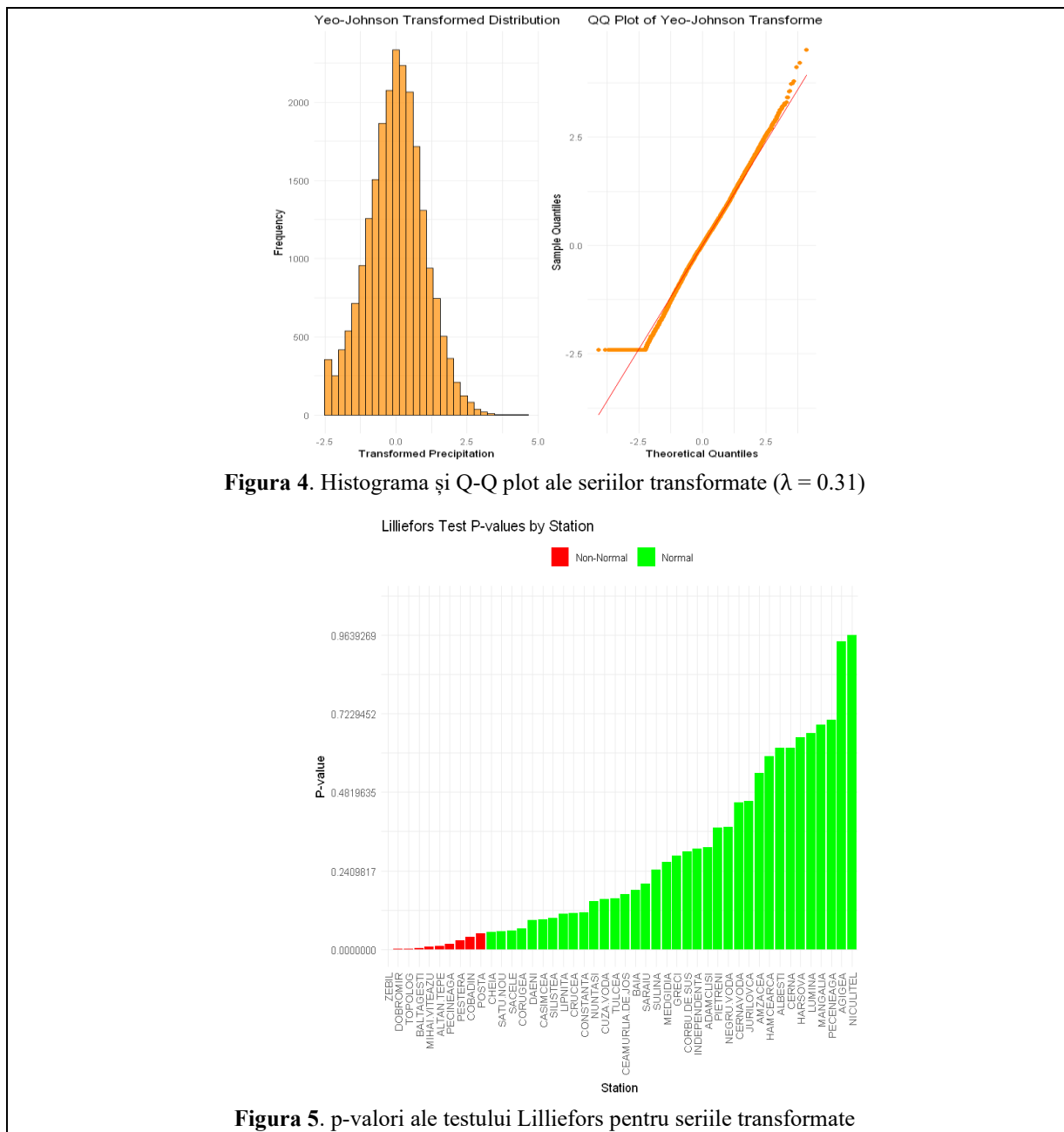
Figura 2. Distribuția și Q-Q plot ale seriilor originale de precipitații

Figura 3. Date originale – p-valori ale testului Lilliefors

Pe de altă parte, a fost aplicat testul Lilliefors pentru fiecare stație. Așa cum este ilustrat în Figura 3, valorile p pentru toate stațiile sunt aproape de zero, nedepășind pragul de semnificație de 0,05, ceea ce confirmă abaterea de la distribuția normală. Pentru a aborda această problemă, a fost aplicată transformarea Yeo–Johnson (Ecuația 1) pentru a stabili varianța și a reduce asimetria. După

transformare, distribuția precipitațiilor (Figura 4, stânga) a aproximat o formă mai simetrică, în clopot, sugerând o îmbunătățire a normalității. Această optimizare este coroborată de graficul Q-Q (Figura 4, dreapta), unde datele transformate se aliniază mai strâns cu cuantilele teoretice, indicând o aderență superioară la ipoteza de normalitate.

A fost realizat ulterior un test Lilliefors la nivel de stație pe datele transformate pentru a evalua normalitatea distribuției precipitațiilor pentru fiecare stație în parte. Așa cum este prezentat în Figura 5, valorile p pentru majoritatea stațiilor s-au situat peste pragul de 0,05, confirmând obținerea normalității la aceste stații în urma transformării. Cu toate acestea, un subgrup de stații (evidențiate cu roșu) a continuat să prezinte abateri semnificative de la normalitate chiar și după transformare, p – valorile asociate rămânând sub prag de semnificație.



3.2 Analiza valorilor medii

Comparațiile localizate care reprezintă rezultatele testului $t_{pereche}$, prezentate în Figura 6, indică semnificația la nivel de pereche a diferențelor dintre valorile medii pentru stațiile normale.

Fiecare celulă compară o pereche de stații, unde celulele verzi semnifică diferențe ne semnificative, iar celulele roșii indică diferențe semnificative ($p < 0,05$). Această vizualizare matricială relevă mai multe diferențe semnificative, reflectând variabilitatea nivelurilor de precipitații în întreaga regiune. În mod remarcabil, mai multe stații prezintă o grupare (clustering) în tiparele de semnificație, sugerând posibile similarități sau grupări regionale ale nivelurilor medii de precipitații.

Evaluarea globală ce reprezintă rezultatele testului post-hoc SNK pentru ANOVA univariată, prezentată în Figura 7, evidențiază semnificația de ansamblu a diferențelor dintre valorile medii pentru stațiile normale. Anumite grupuri de stații prezintă o ne semnificație constantă, indicând niveluri relativ omogene ale precipitațiilor în cadrul acestor clustere, în timp ce alte stații diferă considerabil. Aceste tipare pot reflecta factorii climatici sau geografici subiacenți care influențează dinamica regională a precipitațiilor.

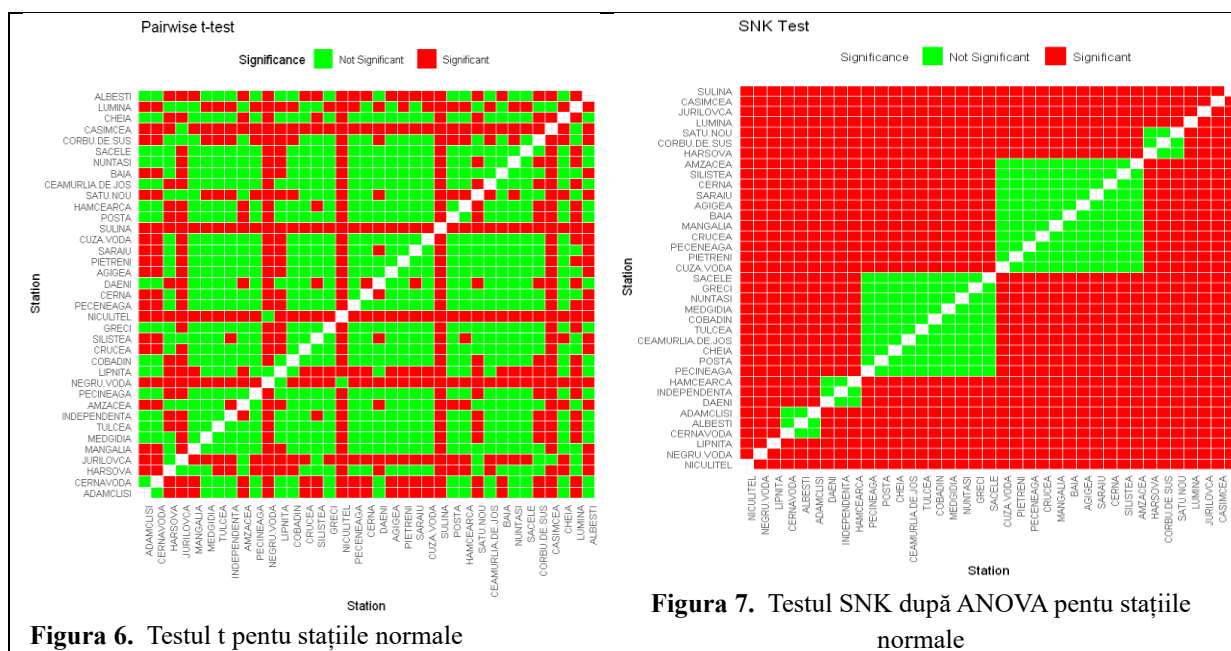


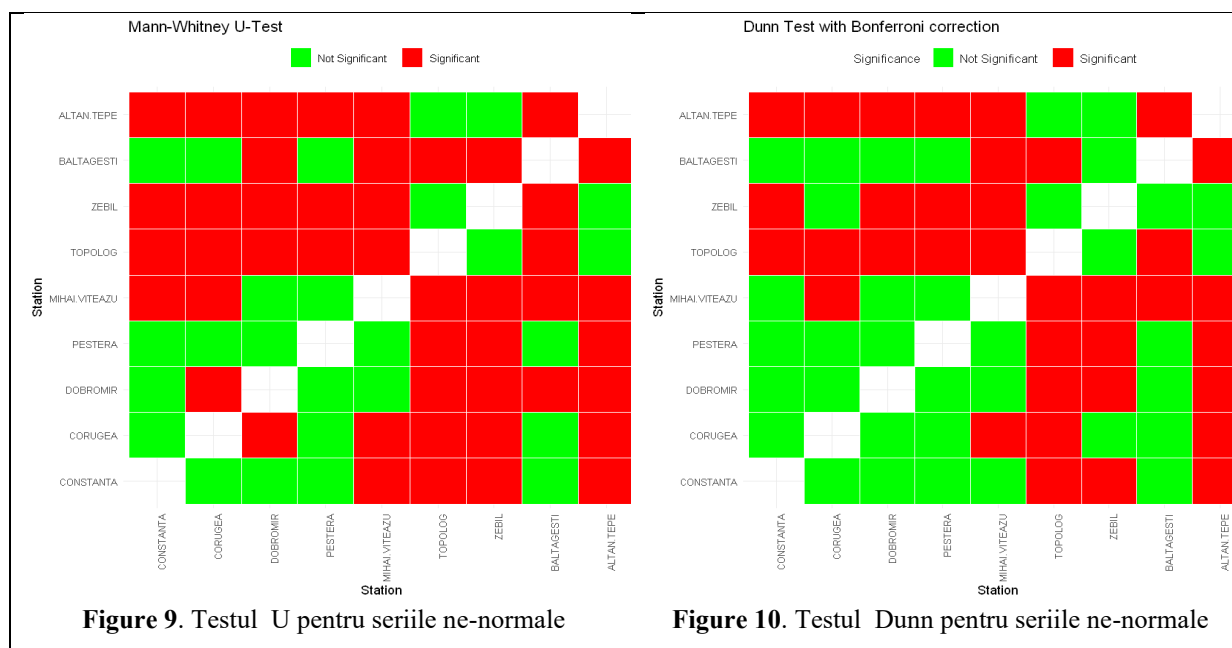
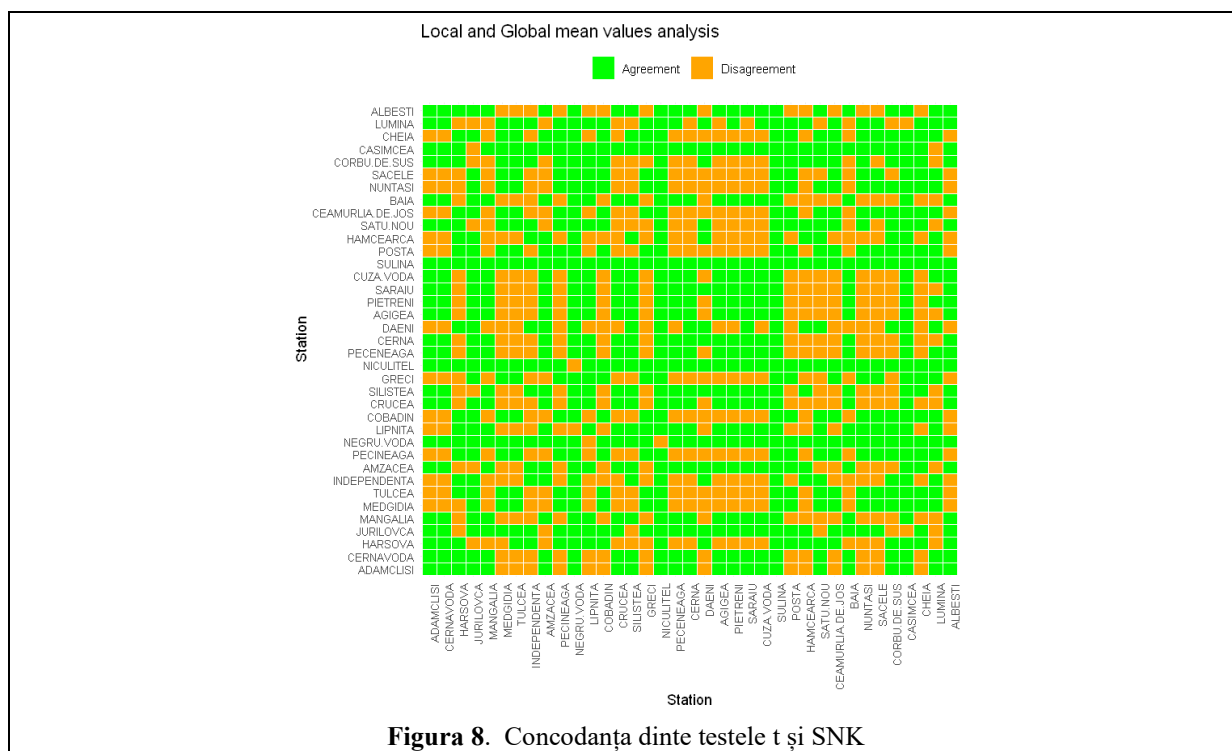
Figura 6. Testul t pentru stațiile normale

Figura 7. Testul SNK după ANOVA pentru stațiile normale

În continuarea analizei inițiale, Figura 8 prezintă o comparație între rezultatele obținute la nivel local (testul t) și cele de la nivel global (SNK), evidențiind gradul de concordanță sau discordanță dintre aceste două metode. Celulele verzi indică o concordanță între analiza locală și cea globală, în timp ce celulele portocalii denotă discordanță. Stațiile care prezintă o concordanță constantă atât în testele locale, cât și în cele globale reprezintă clustere cu caracteristici de precipitații stabile, neafectate de scara de analiză. Aceste stații aparțin, cel mai probabil, unor regiuni cu influențe climatice uniforme, constituind puncte de referință valoroase pentru înțelegerea tiparelor generale de precipitații din Dobrogea. În schimb, stațiile cu niveluri ridicate de discordanță pot fi mai sensibile la variații microclimatice sau topografice pe care testele locale le captează, dar care este posibil să nu se alinieze cu tendințele regionale mai largi identificate prin analiza globală.

Această analiză dublă oferă două concluzii principale.

În primul rând, stațiile cu rezultate consecvente de la un test la altul ar putea servi ca indicatori stabili pentru studierea tendințelor de precipitații pe termen lung și pentru strategiile regionale de gestionare a resurselor de apă. În al doilea rând, stațiile cu rezultate discordante justifică investigații suplimentare, deoarece pot evidenția efecte climatice localizate sau caracteristici geografice unice. Testul Mann–Whitney U, ale cărui rezultate sunt ilustrate în Figura 9, reprezintă comparația în perechi a valorilor medii pentru stațiile non-normale. Aceasta relevă o variabilitate a semnificației, anumite stații indicând clustere de semnificație ce pot puncta către similarități locale în nivelul precipitațiilor sau efecte microclimatice.



Pe lângă analiza locală, rezultatele evaluării globale prezentate în Figura 10 evidențiază semnificația la nivel de perechi pentru stațiile non-normale, clarificând suplimentar care perechi de stații prezintă diferențe semnificative în nivelurile mediane ale precipitațiilor și oferind o imagine de ansamblu asupra tiparelor regionale extinse.

Concordanța dintre analiza locală și cea globală este ilustrată în Figura 11, arătând că stațiile cu rezultate consecvente pot reflecta caracteristici de precipitații stabile la scări variabile, în timp ce stațiile care prezintă discrepante ar putea evidenția factori climatici sau geografici localizați, nesurprinși de evaluările mai largi.

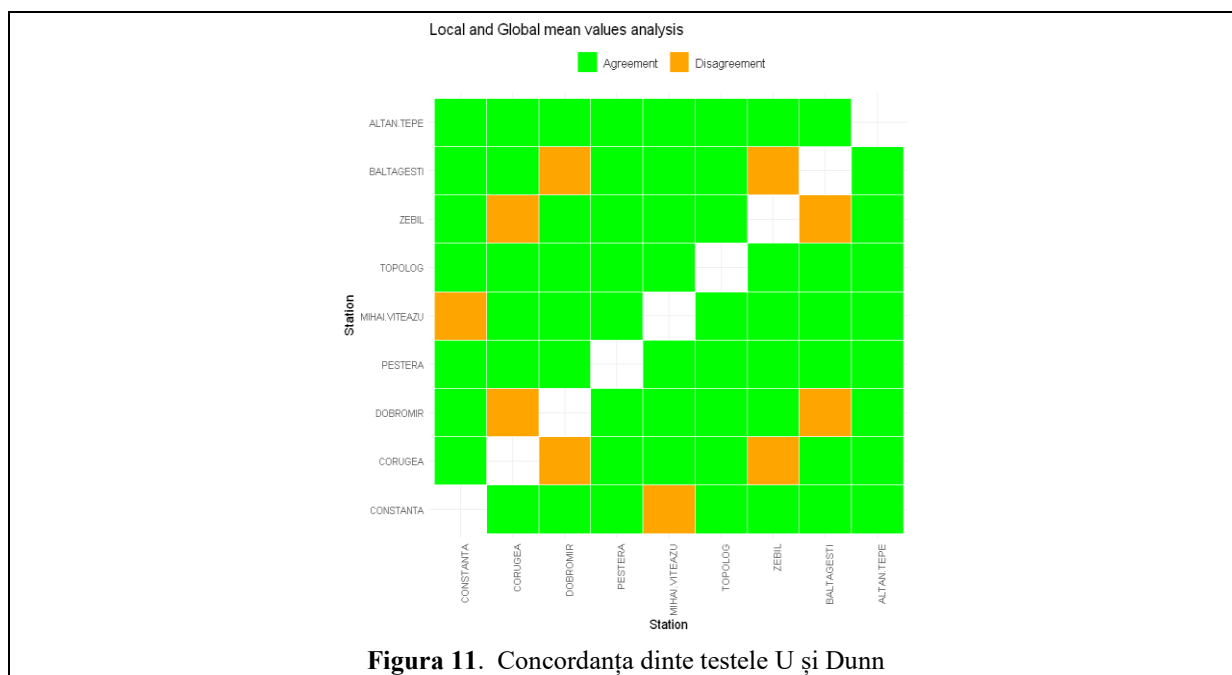


Figura 11. Concordanța dintre testele U și Dunn

3.3 Analiza de varianță

Analiza varianței pentru seriile transformate evidențiază o limitare semnificativă în îndeplinirea ipotezelor necesare pentru testele parametrice. Figura 12 arată că doar Agigea și Jurilovca îndeplinesc cerințele simultane de normalitate și homoscedasticitate, așa cum indică testul Levene ($p > 0.05$).

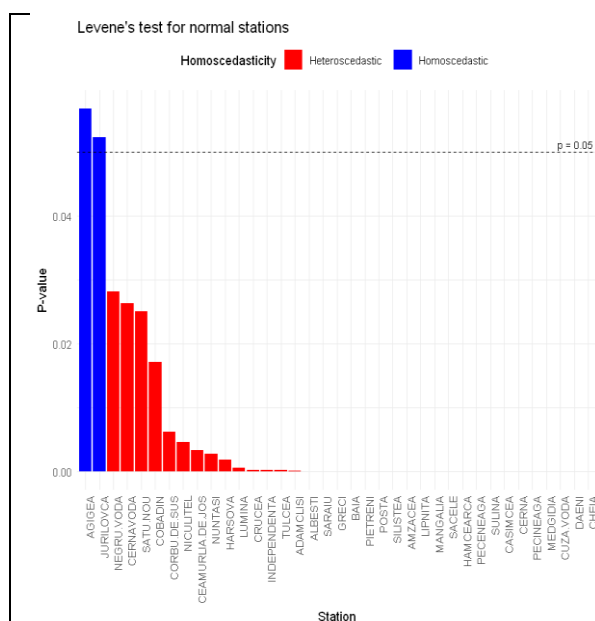


Figura 12. Homoscedasticitatea seriilor normale

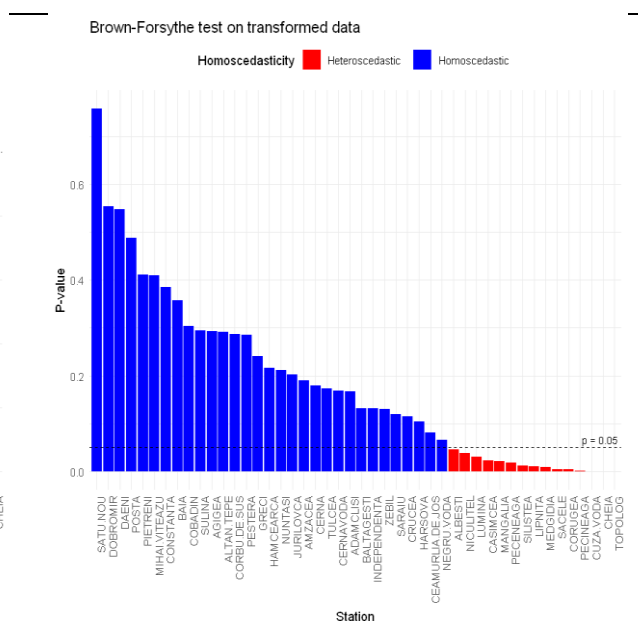


Figura 13. Homoscedasticitatea seriilor transformate

Având în vedere că majoritatea stațiilor nu îndeplinesc condiția de homoscedasticitate, se justifică aplicarea metodelor non-parametrice pentru analiza varianței la nivelul întregului set de date. O examinare ulterioară utilizând testul Brown–Forsythe (Figura 13) oferă o perspectivă suplimentară asupra homoscedasticității seriilor transformate individuale. Majoritatea stațiilor sunt reprezentate cu albastru, indicând diferențe ne semnificative ale varianțelor. În urma testelor individuale, s-a efectuat un test Brown–Forsythe în perechi (Figura 14). Rezultatele relevă diferențe uniforme și ne semnificative de la o pereche de stații la alta, sugerând că varianțele sunt constante la nivel local în

întregul set de date. Această consecvență se aliniază cu constatările nesemnificative din testele individuale, susținând suplimentar ipoteza de omogenitate

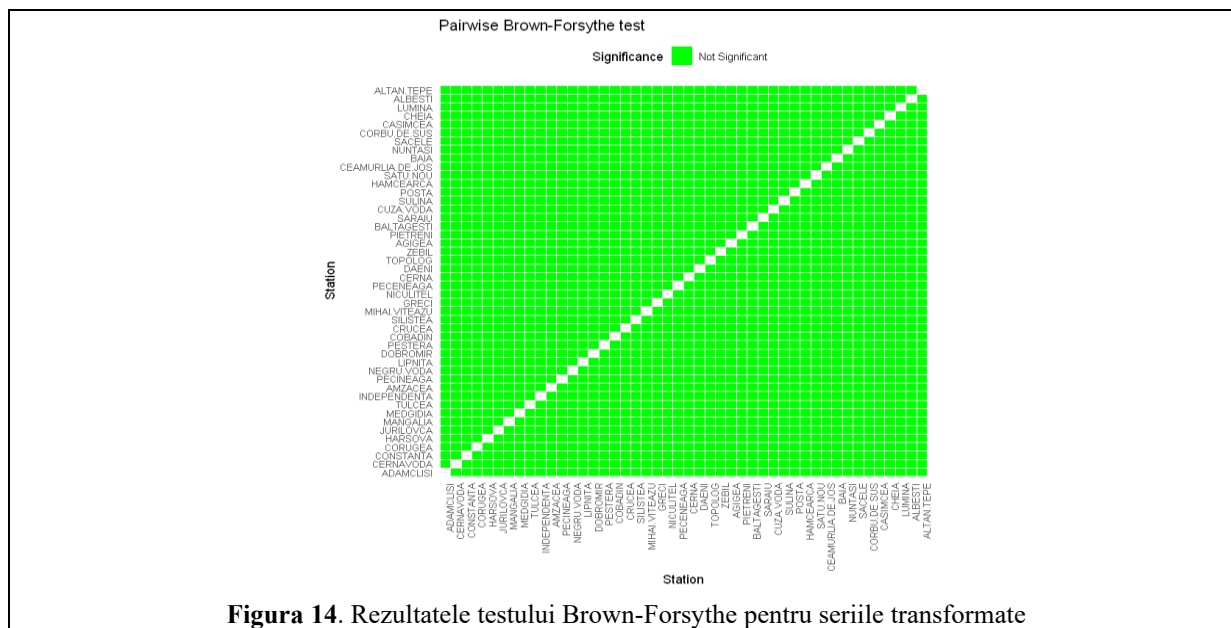


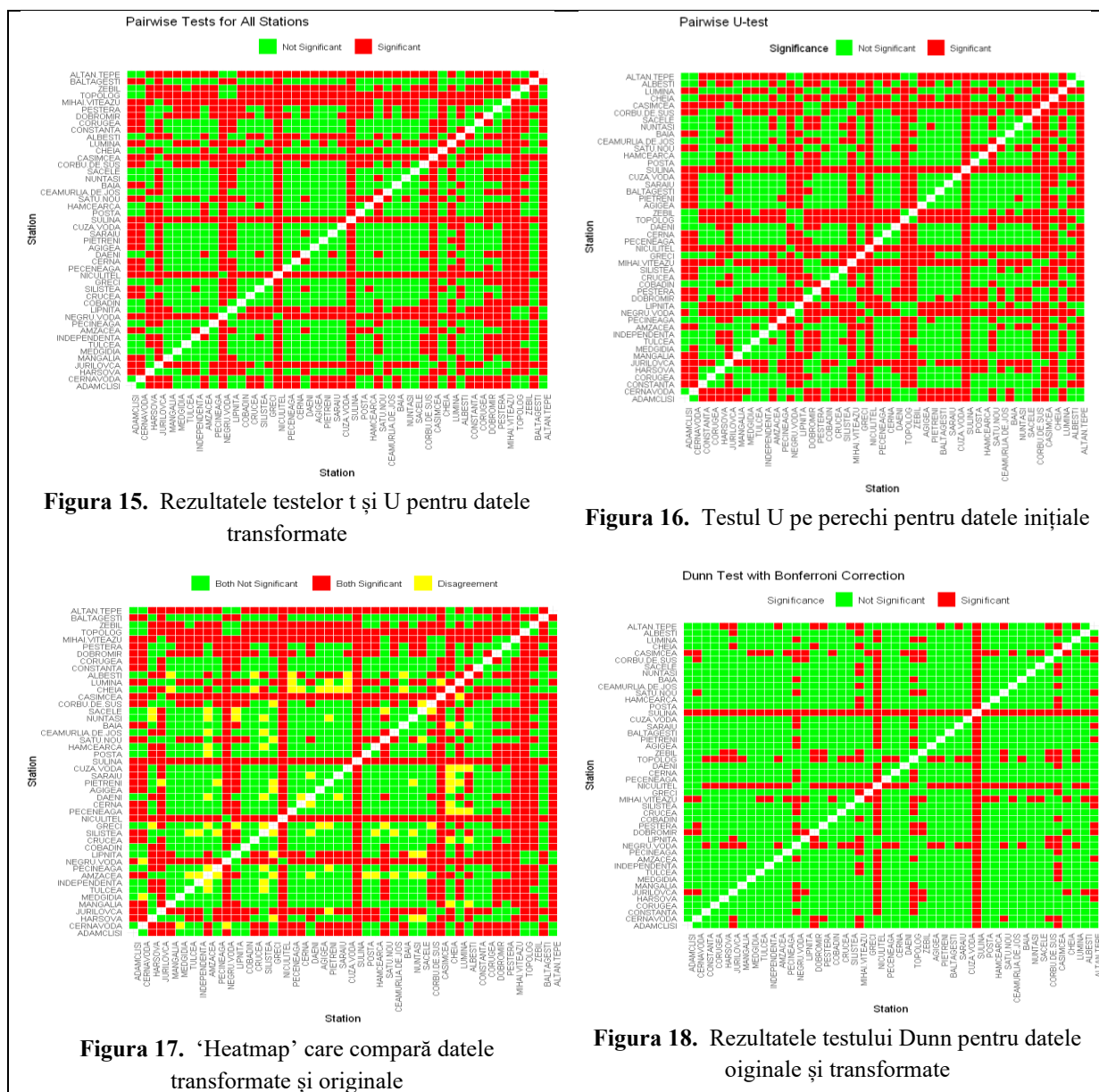
Figura 14. Rezultatele testului Brown-Forsythe pentru seriile transformate

Pentru a completa analiza în perechi, s-a aplicat testul Fligner–Killeen cu scopul de a evalua diferențele globale de varianță. O valoare $p = 0,8544$ indică absența unor diferențe semnificative din punct de vedere statistic în ceea ce privește varianțele. Acest rezultat coroborează suplimentar constatările obținute din testele locale în perechi, confirmând că diferențele de varianță nu sunt semnificative statistic la nivelul întregului set de date.

3.4 Analiza comparativă

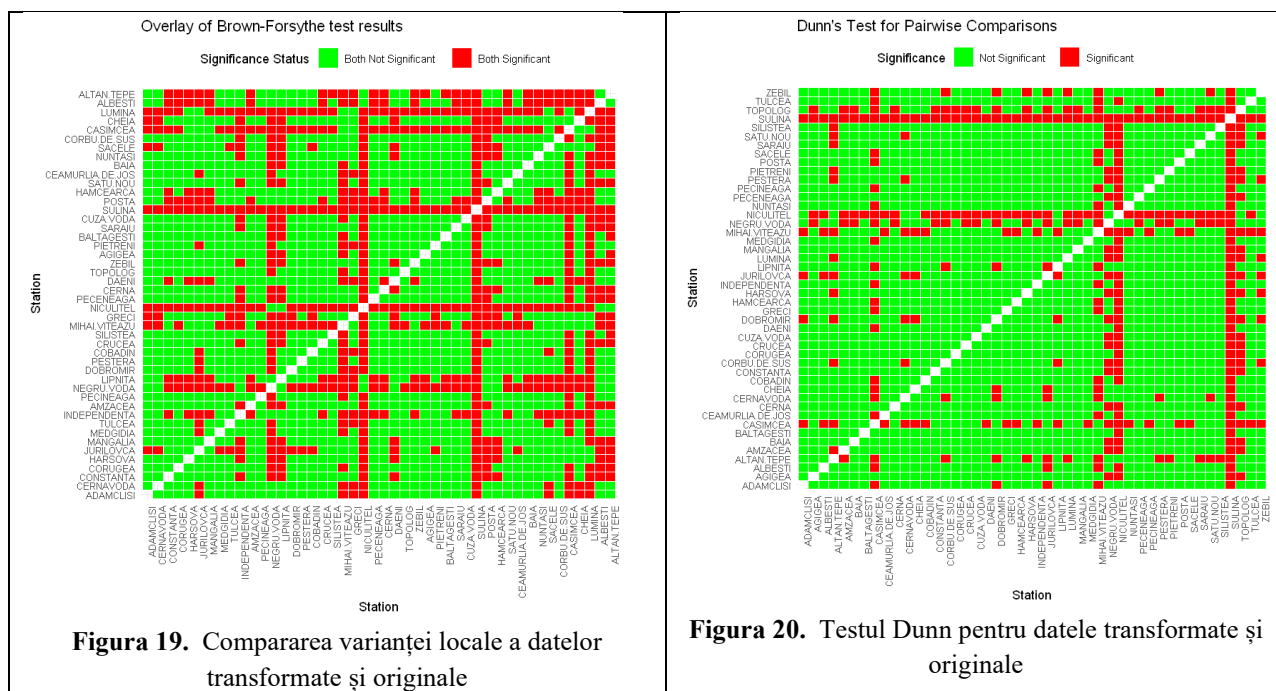
Figura 15 prezintă rezultatele combinate ale setului de date transformat pentru comparațiile în perechi, unde au fost aplicate atât analize parametrice (testul t), cât și non-parametrice (testul U), în funcție de proprietățile distribuționale specifice fiecărei stații. Figura 16 prezintă rezultatele testului U pentru datele originale, surprinzând comparația non-parametrică în perechi a mediilor precipitațiilor la nivelul tuturor stațiilor meteorologice. Concordanța dintre analizele efectuate pe seturile de date transformate și cele originale este reprezentată în Figura 17, reflectând gradul de acord dintre acestea. Celulele verzi indică un acord privind nesemnificația statistică, celulele roșii reprezintă un acord privind semnificația, iar celulele galbene indică o discrepanță între cele două analize.

Analiza vizuală din Figura 18 evidențiază rezultatul aplicării testului Dunn cu corecție Bonferroni pe seturile de date originale și cele transformate, relevând comparațiile globale non-parametrice în perechi. Ambele seturi de date mențin ordini ale rangurilor aproape identice la nivelul stațiilor, ca urmare a naturii transformării aplicate. În cadrul testelor non-parametrice, precum testul U și testul Dunn, rangurile constituie baza fundamentală a comparației statistice. Aceste teste evaluează dacă două sau mai multe grupuri diferă în ceea ce privește tendința centrală prin analiza pozițiilor relative ale observațiilor, și nu a valorilor lor absolute. Această abordare bazată pe ranguri permite comparațiilor să rămână eficiente în cazul distribuțiilor non-normale, reducând sensibilitatea la valorile aberante (outliers) și menținând accentul pe ordonarea generală a punctelor de date, demonstrând că transformarea datelor nu a modificat ordinea rangurilor. Prin urmare, relațiile, așa cum sunt reprezentate de ranguri, au rămas neschimbate, consecința fiind obținerea a două hărți termice (heatmaps) similare ce reflectă ambele abordări.



Rezultatele suprapuse reprezentate în Figura 19 relevă faptul că, pentru majoritatea perechilor de stații, transformarea fie a păstrat, fie nu a avut niciun impact asupra omogenității varianței. Mai exact, regiunile constante de semnificație sau nesemnificație sunt evidențiate cu roșu, respectiv verde. În mod remarcabil, absența oricăror regiuni galbene indică faptul că nu au existat discrepante între seturile de date originale și cele transformate. Acest rezultat este direct legat de conservarea ordinii rangurilor pe parcursul procesului de transformare. Prin menținerea structurii rangurilor, transformarea a asigurat că relațiile privind varianța au rămas consecvente în ambele seturi de date, conducând la rezultate similare, fără niciun conflict. Această sinteză vizuală susține astfel stabilitatea și eficacitatea transformării în conservarea relațiilor fundamentale de varianță.

În cele din urmă, am extins evaluarea la o comparație globală a varianței, aplicând testul Fligner-Killeen pe ambele seturi de date. Au fost detectate diferențe semnificative, iar testul Dunn a fost utilizat pentru a identifica grupurile specifice cu diferențe semnificative. Figura 20 prezintă rezultatele pentru setul de date original, menționând că setul de date transformat a generat aceeași hartă termică (heatmap). Similaritatea dintre rezultatele ambelor seturi de date se datorează, din nou, păstrării ordinii rangurilor pe tot parcursul procesului de transformare.



3.4 Detectarea punctelor de schimbare cu MDL

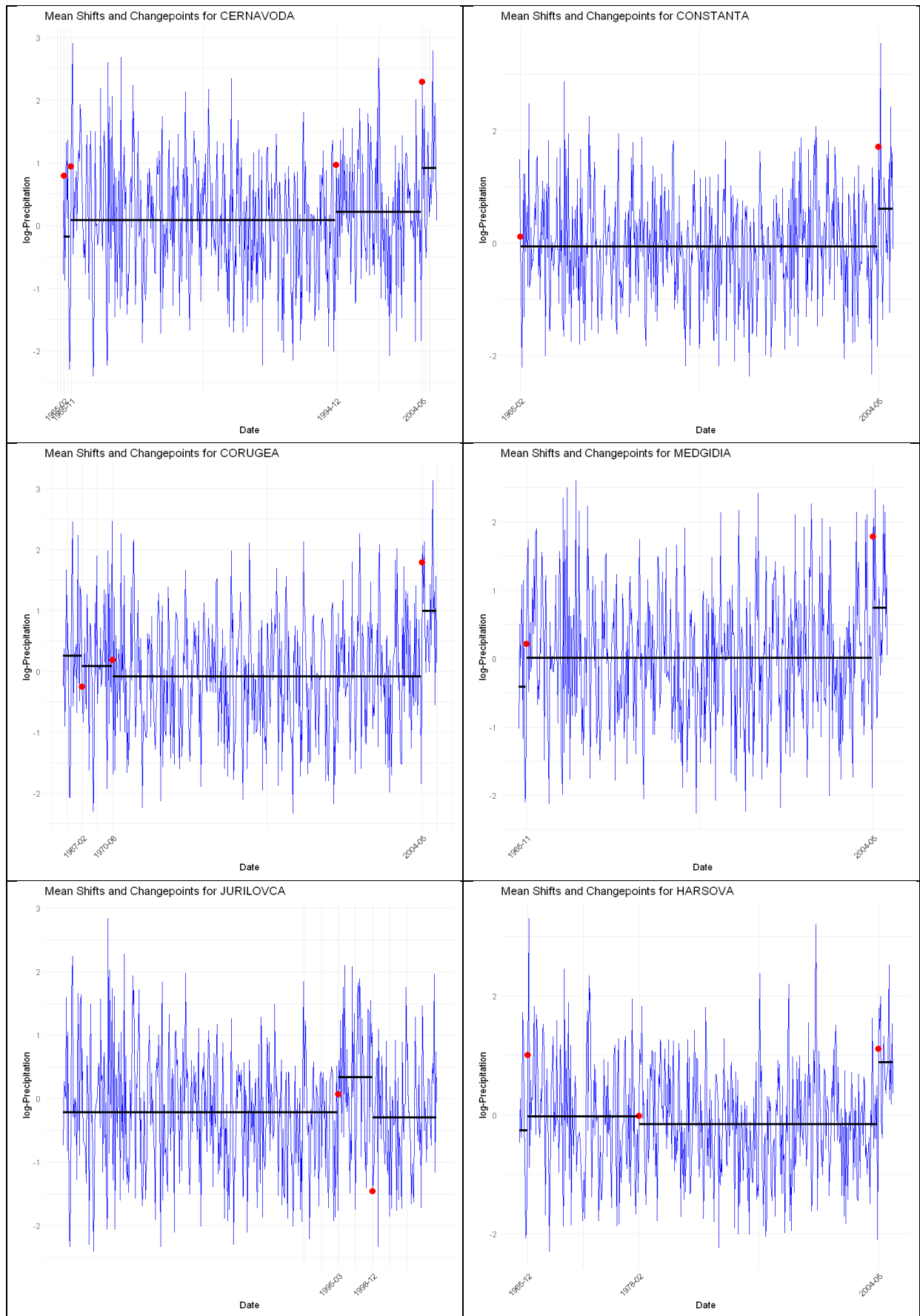
Pentru a valida cadrul MDL propus, am aplicat inițial metodologia pe serii temporale sintetice cu puncte de schimbare predefinite. Aceste seturi de date de test, generate sub ipoteze de normalitate, au permis evaluarea performanței algoritmului în identificarea pozițiilor cunoscute ale punctelor de schimbare. Pe parcursul a 1.000 de rulări de optimizare, am testat serii fără puncte de schimbare ($m = 0$), cu un singur punct de schimbare ($m = 1$) și cu două puncte de schimbare ($m = 2$). O sinteză a acestor rezultate este prezentată în Tabelul 1.

Tabel 1. Acuratețea detectării punctelor de schimbare

Număr de puncte de schimbare	Acuratețe
$m = 0$	99.99%
$m = 1$	99.7%
$m = 2$	98.8%

Aceste rezultate confirmă capacitatea cadrului metodologic propus de a detecta atât numărul, cât și poziționarea punctelor multiple de schimbare în condiții controlate.

În urma acestei validări, am aplicat cadrul metodologic pe datele reale de precipitații de la următoarele stații: Cernavodă, Constanța, Corugea, Hârșova, Jurilovca, Mangalia, Medgidia, Sulina, Tulcea și Adamclisi. Figura 21 ilustrează punctele de schimbare (punctele roșii) și modificările corespunzătoare în media segmentelor pentru stațiile analizate. Liniile orizontale reprezintă mediile precipitațiilor pe fiecare segment. Acestea confirmă prezența unor decalaje climatice detectabile, demonstrând că metodologia bazată pe MDL surprinde eficient punctele de schimbare în seturile de date climatice reale. Fiabilitatea acestor constatări ar putea fi consolidată prin intermediul mai multor abordări de validare. În primul rând, integrarea metadatelor, cum ar fi înregistrările istorice climatice, poate oferi dovezi contextuale care să susțină schimbările detectate. În plus, comparațiile cu alte metode consacrate de detecție a punctelor de schimbare, precum procedura de segmentare Hubert, CUSUM, testul Pettitt și testul Bayesian Lee–Heghinian, pot oferi un etalon pentru evaluarea cadrului MDL.



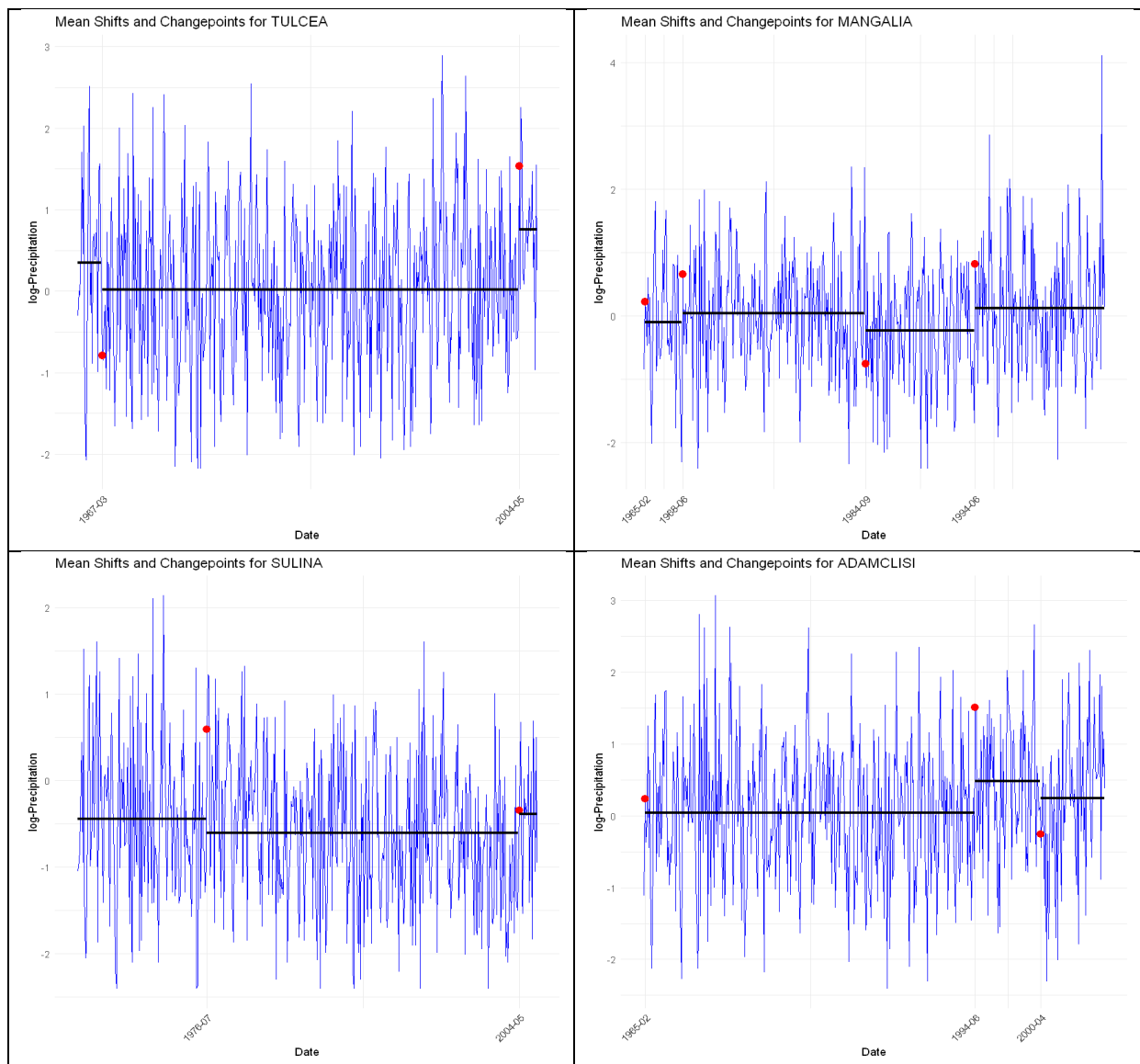


Figura 21. Puncte de schimbare ale celor 10 serii

Cercetările viitoare pot explora, de asemenea, abordări bayesiene pentru a optimiza cadrul MDL. Spre deosebire de metodele tradiționale, tehnicile bayesiene încorporează cunoștințe apriori și oferă distribuții posterioare atât pentru parametri, cât și pentru poziționarea punctelor de schimbare, oferind o perspectivă probabilistică asupra procesului de segmentare. Acest lucru ne poate permite nu doar să detectăm punctele de schimbare, ci și să cuantificăm incertitudinea asociată identificării acestora. Așa cum se menționează în lucrarea lui Lund, aceste modele pot include structuri ierarhice pentru a lua în considerare procesele climatice cunoscute, conducând la rezultate mai bine adaptate contextului și mai ușor de interpretat.

4. Concluzii

Acest studiu oferă o evaluare statistică a variabilității precipitațiilor din Dobrogea, utilizând date de la 46 de stații meteorologice pe o perioadă de 40 de ani (1965–2005). Pentru a aborda problema non-normalității din seriile de precipitații, s-a utilizat transformarea Yeo–Johnson. Această transformare a redus asimetria și a stabilizat varianțele, permițând aplicarea metodelor parametrice în majoritatea cazurilor. Validarea prin intermediul testului Lilliefors a confirmat obținerea normalității pentru majoritatea seriilor transformate, deși un subset de stații a prezentat abateri persistente,

necesitând aplicarea unor abordări non-parametrice. Omogenitatea varianțelor, o proprietate necesară pentru numeroase teste statistice, a fost evaluată prin testul Levene și echivalentele sale non-parametrice. Disparitățile de medie și varianță au indicat o heterogenitate spațială ridicată a precipitațiilor. Hărțile termice (heat maps) utilizate pentru a ilustra rezultatele testelor în perechi și globale au relevat grupări regionale de stații care ar putea fi influențate de factori climatici comuni.

Pe de altă parte, cadrul MDL, optimizat prin Simulated Annealing (SA), a identificat puncte de atât în seturile de date sintetice, cât și în cele reale. Validarea pe date sintetice a demonstrat precizia cadrului în detectarea punctelor de schimbare predefinite în condiții controlate. Mai mult, aplicarea pe seriile reale de precipitații a evidențiat decalaje structurale semnificative, cum ar fi cele din 1994 și 2000 pentru Adamclisi, respectiv din 1997 și 2004 pentru Cheia, reflectând modificări fundamentale în regimul precipitațiilor.

În cele din urmă, o direcție potențială pentru cercetări viitoare este indicată de tiparele de grupare din hărțile termice. Utilizând tehnici formale de analiză de cluster, cum ar fi k-means, ierarhizarea ierarhică sau DBSCAN, aceste grupări ar putea fi confirmate cantitativ. De asemenea, pentru validarea punctelor de schimbare detectate și diferențierea dintre variațiile climatice reale și artefactele de observație, este importantă integrarea metadatelor, cum ar fi istoricul strămutărilor de stații și modificările de aparatură. Metodele complementare de detecție a punctelor de schimbare ar putea spori suplimentar robustetea constatărilor. Extinderea analizei prin includerea modelării spațiale ar oferi un context mai larg acestor rezultate, consolidându-le relevanța pentru gestionarea regională a resurselor de apă și strategiile de adaptare la schimbările climatice. În ansamblu, aceste demersuri ar spori fiabilitatea studiilor de precipitații și ar sprijini luarea deciziilor fundamentate în abordarea provocărilor generate de creșterea variabilității climatice în medii semiaride precum Dobrogea.

Bibliografie

Abdi, H., Williams, L.J. Newman-Keuls Test and Tukey Test. Available online: <https://personal.utdallas.edu/~herve/abdi-NewmanKeuls2010-pretty.pdf> (accessed on 12 October 2024).

Bandoc, G., Prăvălie, R. Climatic water balance dynamics over the last five decades in Romania's most arid region, Dobrogea. *Journal of Geographical Sciences* **2015**, 25(11), 1307-1327. <https://doi.org/10.1007/s11442-015-1236-1>

Bărbulescu, A. A New Method for Estimation the Regional Precipitation. *Water Resources Management* **2016**, 30(1), 43-52 <https://doi.org/10.1007/s11269-015-1152-2>

Bărbulescu, A. On the Regional Temperature Series Evolution in the South-Eastern Part of Romania. *Applied Sciences* **2023**, 13(6), 3904. <https://doi.org/10.3390/app13063904>

Bărbulescu, A., Maftai, C. Evaluating the probable maximum precipitation. Case study from the Dobrogea region, Romania. *Romanian Reports in Physics* **2023**, 75(3), 704. <https://doi.org/10.59277/RomRepPhys.2023.75.704>.

Bărbulescu, A., Popescu-Bodorin, N. Assessing the history-based predictability of regional monthly precipitation data using statistical and fuzzy methods. *Stochastic Environmental Research and Risk Assessment* **2019**, 33(7), 1435-1451. <https://doi.org/10.1007/s00477-019-01702-1>

Bărbulescu, A., Postolache, F., Dumitriu, C. Ștefan. Estimating the precipitation amount at regional scale using a new tool, Climate Analyzer. *Hydrology* **2021**, 8(3), 125. <https://doi.org/10.3390/hydrology8030125>.

Box, G. E. P. Non-normality and tests on variances. *Biometrika* **1953**, 40(3-4). <https://doi.org/10.1093/biomet/40.3-4.318>.

Brown, M. B., Forsythe, A. B. Robust tests for the equality of variances. *Journal of the American Statistical Association* **1974**, 69(346). <https://doi.org/10.1080/01621459.1974.10482955>

Caussinus, H., Mestre, O. Detection and correction of artificial shifts in climate series. *Journal of the Royal Statistical Society. Series C: Applied Statistics* **2004**, 53(3), 405-435.

<https://doi.org/10.1111/j.1467-9876.2004.05155.x>

Chelcea, S., Ionita, M., Adler, M. Identification of Dry Periods in the Dobrogea Region. In I. Management Association (Ed.), *Civil and Environmental Engineering: Concepts, Methodologies, Tools, and Applications*, IGI Global Scientific Publishing, 2016, pp. 324-339. <https://doi.org/10.4018/978-1-4666-9619-8.ch012>

Conover, W.J., Johnson, M.E., Johnson, M.M. A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data. *Technometrics* **1981**, 23(4), 351-361. <https://doi.org/10.1080/00401706.1981.10487680>.

Dallal, G.E., Wilkinson, L. An analytic approximation to the distribution of lilliefors's test statistic for normality. *American Statistician* **1986**, 40(4), 294-296. <https://doi.org/10.1080/00031305.1986.10475419>.

Dunn, O.J. Multiple Comparisons Using Rank Sums. *Technometrics* **1964**, 6(3), 241-252. <https://doi.org/10.1080/00401706.1964.10490181>.

Fay, M.P.; Proschan, M.A. Wilcoxon–Mann–Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules. *Statistics Surveys* **2010**, 4, 1–39.

Fisher, R.A. Statistical Methods for Research Workers. In: *Breakthroughs in Statistics*. Kotz, S., Johnson, N.L., Eds.; Springer: New York, NY, 1992. https://doi.org/10.1007/978-1-4612-4380-9_6.

Fligner, M.A. A class of two-sample distribution-free tests for scale. *Journal of the American Statistical Association* **1979**, 74(368), 889-893. <https://doi.org/10.1080/01621459.1979.10481050>.

Gatz, D. F., Smith, L. The standard error of a weighted mean concentration-I. Bootstrapping vs other methods. *Atmospheric Environment* **1995**, 29(11), 1185-1193. [https://doi.org/10.1016/1352-2310\(94\)00210-C](https://doi.org/10.1016/1352-2310(94)00210-C).

Helsel, D.R., Hirsch, R.M., Ryberg, K.R., Archfield, S.A., Gilroy, E.J. *Statistical methods in water resources. Techniques and Methods* 4-A3, U.S. Geological Survey, 2020. <https://doi.org/10.3133/tm4a3>

Hewaarachchi, A.P., Li, Y., Lund, R., Rennie, J. Homogenization of daily temperature data. *Journal of Climate* **2017**, 30(3), 985-999. <https://doi.org/10.1175/JCLI-D-16-0139.1>

Kirkpatrick, S., Gelatt, C.D., Vecchi, M.P. Optimization by simulated annealing. *Science* **1983**, 220(4598), 671-680. <https://doi.org/10.1126/science.220.4598.671>

Levene, H. (1960). Robust tests for equality of variances. In: *Contributions to Probability and Statistics*, Olkin, I., Ed.; Stanford University Press: Palo Alto, USA, 278-292.

Li, Y., Lund, R., Hewaarachchi, A. Multiple change point detection with partial information on change point times. *Electronic Journal of Statistics* **2019**, 13(2), 2462-2520. <https://doi.org/10.1214/19-EJS1568>.

Lu, Q., Lund, R., Lee, T.C.M. An MDL approach to the climate segmentation problem. *Annals of Applied Statistics* **2012**, 6(1), 299-319. <https://doi.org/10.1214/09-AOAS289>.

Maftai, C., Bărbulescu, A. Statistical analysis of climate evolution in Dobrudja region. Proceedings of the World Congress on Engineering 2008, vol. II, WCE 2008, July 2 - 4, 2008, London, U.K., 1082-1087.

Mann, H.B., Whitney, D.R. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Annals of Mathematical Statistics* **1947**, 18(1), 50–60.

Menne, M.J., Williams, C.N. Detection of undocumented change points using multiple test statistics and composite reference series. *Journal of Climate* **2005**, 18(20), 4271–4286. <https://doi.org/10.1175/JCLI3524.1>

Nachar, N. The Mann-Whitney U: A Test for Assessing Whether Two Independent Samples Come from the Same Distribution. *Tutorials in Quantitative Methods for Psychology* **2008**, 4(1), 13-20. <https://doi.org/10.20982/tqmp.04.1.p013>

Student. The Probable Error of a Mean. *Biometrika* **1908**, 6(1), 1-25. <https://doi.org/10.2307/2331554>

Welch, B.L. On the Comparison of Several Mean Values: An Alternative Approach. *Biometrika* **1951**, 38(3/4), 330-336. <https://doi.org/10.2307/2332579>

Wilks, D.S.. *Statistical Methods in the Atmospheric Sciences*, 4th Edition, Elsevier, 2020.

Yeo, I.N.K., Johnson, R.A. A new family of power transformations to improve normality or symmetry. *Biometrika* **2000**, 87(4), 954-959. <https://doi.org/10.1093/biomet/87.4.954>.