

З.М. Кирєєва¹

ORCID: 0009-0003-9544-6944
kireeva5687@gmail.com

О.А. Скриник^{2,1}

ORCID: 0000-0003-0332-5073
skrynyk@nubip.edu.ua

Л.В. Паламарчук¹

ORCID: 0000-0001-9906-8870
palamarchuk.l@ukr.net

УДК 551.583.1+551.577

ГОМОГЕНІЗАЦІЯ ЧАСОВИХ РЯДІВ ОПАДІВ: ОСНОВНІ СУЧАСНІ ПІДХОДИ ТА ПРОГРАМНІ ЗАСОБИ

Проаналізовано сучасні програмні продукти, які можуть бути використані для проведення контролю якості та гомогенізації часових рядів метеорологічних величин, зокрема сум атмосферних опадів. Для виконання процедур контролю якості та гомогенізації часових рядів місячних сум опадів обрано програмний продукт HOMER. До переваг використання цієї програми можна віднести можливість для дослідника приймати рішення про підтвердження точок розриву на етапі об'єданого порівняння (joint detection), що дозволяє враховувати наявну інформацію про природну мінливість значень сум опадів у регіоні та не допускати надмірного згладжування їх величин. Проведено тестування процедури гомогенізації рядів місячних сум опадів метеорологічної мережі протягом 1946–2020 рр. з використанням HOMER. Визначено методичні підходи до виділення ділянок неоднорідностей в часових рядах (точок розриву). Встановлено, що величина амплітуди зсуву в точках розриву може змінюватися в межах 0,1–0,7 (у відносних одиницях стандартного відхилення). Проведено корекцію часових рядів місячних сум опадів за досліджуваний період та розраховано, що їх зміна залежить від величини і знаку амплітуди зміщення і складає 10–20% початкових сум опадів. Розроблені методичні прийоми та підходи до перевірки якості даних та їх гомогенізації з використанням програмного продукту HOMER можуть застосовуватися у подальших дослідженнях опадів.

Ключові слова: гомогенізація, програмний продукт HOMER, місячні суми опадів, кліматична однорідність даних, точки розриву.

ВСТУП

Головним джерелом даних для отримання різноманітних кліматичних показників, які використовуються у прогнозуванні змін клімату, оцінюванні впливу стану атмосфери на господарську діяльність та здоров'я людей, прогнозуванні погоди є дані стандартних метеорологічних спостережень, що проводяться на державних метеорологічних мережах. У [40] сформульовані сучасні вимоги Всесвітньої Метеорологічної Організації (ВМО) до якості та способу представлення результатів спостережень для подальшого коректного їх використання у розрахунках, верифікації моделювання та для інших потреб. Дані повинні бути отримані за стандартними методиками з допомогою вчасно каліброваних приладів, пройти перевірку якості. У часових рядах метеорологічних величин заповнюються усі пропуски й далі проводиться процедура гомогенізації. Поля метеорологічних величин у певному регіоні представляють у вигляді просторових сіток з різним розділенням (*gridding*), у вузли сіток інтерполюють дані спостережень станцій метеорологічної мережі. Одна з важливих процедур у наведеному

переліку — гомогенізація часових рядів, яка виконується після перевірки на якість отриманих даних.

Гомогенізація метеорологічних рядів — це процес виявлення та корекції аномальних змін у даних спостережень, які спричинені різними чинниками неприродного характеру. До таких можна віднести: перенесення станції в іншу точку спостережень, зміну методики, або приладів для вимірювання, зміну ландшафтного середовища навколо станції і тому подібне [18]. Отримані після гомогенізації ряди величин вважаються кліматично однорідними, тобто такими, що можуть показувати лише зміни і коливання регіонального та глобального клімату.

Процедура гомогенізації метеорологічних даних є важливим і обов'язковим процесом, оскільки дозволяє забезпечити у подальшому якісні та однорідні ряди для розрахунку кліматичних показників та кліматичних індексів.

Для проведення гомогенізації використовується ряд математичних методів та підходів, які змінюються відповідно до розвитку знань та технологій. Можна вважати, що історично процес гомогенізації

кліматичних даних зачинався ще з початком збору інформації про стан атмосфери у XIX сторіччі. У той час, коли були створені метеорологічні станції і почалося накопичення даних та формування їх часових рядів, перевірку на однорідність та відповідну корекцію проводили переважно для температурних показників, а рядам сум опадів, швидкості вітру та інших метеорологічних величин, у зв'язку з частою зміною умов вимірювання та складністю методик корекції, приділялась менша увага.

У вересні 1873 року на "Міжнародному метеорологічному конгресі" у Відні Карл Єлінек вперше звернув увагу на інформацію про національні багаторічні ряди даних [24], а через десятиліття у 1905 році Г. Хеллман [42] знову зауважив про відсутність однорідних кліматичних часових рядів через зміни в оточенні станцій та появу нових приладів і закликав працювати над тим, щоб дані станцій з багаторічними спостереженнями ("Säkularstationen") були максимально однорідними.

Методи гомогенізації змінювались з часом. Наприклад, раніше задокументовані точки розриву видаляли за допомогою паралельних вимірювань. Так спостереження в астрономічній обсерваторії Віденського університету були скориговані за допомогою багаторічних 24-годинних вимірювань астрономічної обсерваторії Празького університету. В свою чергу, вимірювання в Мілані з 1763 по 1834 роки були скориговані за допомогою вимірювань в Падуї [29].

Однак для більшості точок розривів у часових рядах була невідома величина зсуву, крім того, ймовірно, що ряди також містили незадокументовані неоднорідності. Таким чином, виникла потреба в статистичних методах виявлення та усунення розривів. На початку XX століття Конрад [17] для оцінювання однорідності рядів застосував критерій Хайдке [22], який використовував співвідношення двох рядів опадів та навіть оцінку отриманих результатів. Як наслідок, він рекомендував, крім названого критерію, використовувати ще додаткові критерії для перевірки однорідності рядів, такі як: критерій Гельмерта [23] і "ретельний" критерій Аббе [15]. Ці критерії продовжували використовувати і в 1940-х роках [14]. Через кілька років для виявлення та усунення неоднорідностей в рядах був запропонований принцип подвійної маси [28].

Юліус Ханн [21] вивчав мінливість абсолютних кількостей опадів і їх співвідношення між станціями та використовував це для контролю якості рядів. Цей підхід був використаний Брюкнером [10] для перевірки даних сум опадів на неоднорідність шляхом порівняння з сусідніми станціями. У своїй книзі "Методи в кліматології" Конрад і Поллак [16]

формалізували цей підхід відносної гомогенізації, який зараз є домінуючим методом виявлення та усунення наслідків штучних змін. Пізніше робоча група ВМО з кліматичних змін рекомендувала застосовувати такі підходи для створення еталонних рядів, використовуючи усереднення даних по багатьох станціях у відносно невеликій географічній області.

Подальший розвиток підходів до гомогенізації знаходимо у роботах Александерссона [7] та Моберга [9], де показано розроблений стандартний тест нормальної гомогенності (Standard Normal Homogeneity Test, SNHT), який знайшов широке впровадження в багатьох дослідженнях. SNHT — це метод, який можна використовувати для ряду з однією точкою розриву, але кліматичні ряди зазвичай містять більше одного розриву. Таким чином, важливим подальшим кроком стала розробка методів, спеціально створених для виявлення та виправлення кількох точок розриву в роботі з неоднорідними даними [30, 37]. Стаття Кауссінуса та Местре [12] містить перший опис такого методу, який одночасно коригує всі ряди виділеної мережі. В подальшому, подібні методи виявились більш точними, що знайшло своє підтвердження у порівняльному дослідженні EU COST ("European Cooperation in Science and Technology") Action ES0601 HOME [38].

Поточна робота зі створення відповідних наборів даних для досліджень змін клімату та його коливань сприяла постійному розвитку методів виявлення та корекції точок розривів у часових рядах кліматичних показників. Щоб відслідковувати удосконалення та адаптацію методів гомогенізації, Угорська метеорологічна служба розпочала серію "Семинарів з гомогенізації" у 1996 році [26].

Сучасні високі вимоги ВМО до якості даних метеорологічних спостережень, що сформульовані в [40], змушують науковців продовжувати працювати над удосконаленням методів гомогенізації даних та розробкою нових підходів до роботи з метеорологічною інформацією. На сьогодні створено значну кількість ефективних гомогенізаційних алгоритмів, процедур та програмних пакетів, які проводять спеціальний кліматичний контроль та мінімізують ті викиди значень показників, що мають некліматичну природу. У науковому та експертному середовищі нині активно обговорюються й оцінюються існуючі програмні продукти [20, 25, 37, 38], які можна використовувати для отримання однорідних часових рядів різних метеорологічних величин.

Наразі усі існуючі алгоритми та методи перевірки часових рядів на кліматологічну неоднорідність (або методи гомогенізації) можуть бути поділені на два основних типи: відносні та абсолютні.

При використанні відносних методів гомогенізації, завжди досліджують не один, а кілька рядів даних, отриманих на метеостанціях, що розташовані у типовому ландшафтному регіоні. Ці методи ґрунтуються на припущенні, що зміни в кількості опадів або температурні коливання на обраній метеостанції повинні бути подібними до тих, які спостерігаються на інших станціях цієї групи. Відносні методи можуть використовувати процедури стандартизації даних в рамках конкретної мережі станцій, порівняння середніх значень зі схожими станціями, аналіз сезонної варіації тощо. Як приклад використання відносного гомогенізаційного методу, можна навести емпіричне моделювання, яке включає створення моделі, що ґрунтується на взаємозв'язках між даними на різних станціях і використовується для визначення неоднорідності в результатах спостережень конкретної метеостанції. Таку модель можна розробити, користуючись кореляцією, регресійним аналізом, або аналізом головних компонентів, в залежності від характеру наявних даних і специфіки завдання.

У свою чергу, використання абсолютних методів передбачає опрацювання лише окремих рядів даних спостережень. Іншими словами, вибирається конкретний ряд, з яким проводяться необхідні статистичні процедури, такі як розрахунки критеріїв або проведення тестів, а на основі отриманих результатів робиться висновок щодо його гомогенності або неоднорідності. Головна ідея полягає в тому, що зміни в роботі приладів і інструментів, які використовуються для вимірювань конкретних метеорологічних величин, можуть суттєво впливати на отримані результати.

Згідно з висновками проведеного комплексного порівняльного аналізу гомогенізаційних методів і алгоритмів [38], в рамках Європейського проекту COST Action ES0601 HOME більш ефективними і коректними визнані відносні методи.

Протягом останніх десятиліть у багатьох країнах розроблена велика кількість програмних продуктів для гомогенізації кліматичних даних, які використовують різноманітні методи та алгоритми перевірок на наявність та усунення викидів і точок розриву часових рядів. Серед найбільш популярних та таких, що знаходять широке застосування у наукових дослідженнях є наступні відносні методи:

MASH (Multiple Analysis for Series Homogenization) представляє собою набір спеціалізованих програм, розроблених Угорською метеослужбою для оцінки якості кліматичних рядів і їх стандартизації. Методологія MASH передбачає використання критерію Шейрера-Рей-Хейра для перевірки однорідності даних, який ґрунтується на оцінюванні різниці/від-

ношення значень між спостереженнями та їх середніми показниками. У випадку, якщо результат тесту вказує на нерівномірність розподілу даних, для оцінки статистичної однорідності за нульовою гіпотезою використовується інструментарій рухомих блоків. Цей підхід дозволяє користувачам визначити, чи є виявлені зміни статистично значущими. Метод MASH належить до групи відносних і часто застосовується для виявлення поступових часових змін даних спостережень, оскільки він може фіксувати неоднорідності рядів, які формуються протягом кількох років. Цей програмний продукт використовують для гомогенізації різноманітних метеорологічних [4, 7] та гідрологічних показників, включаючи інформацію про рівень води в річках, опади та температуру.

Інше програмне забезпечення *RHtestsV4 (Regression-based Homogeneity Test Version 4.0)*, що використовує мову програмування R, застосовується для виявлення і корекції неоднорідностей рядів кліматичних даних. Тут пропонується підхід, що ґрунтується на методі відносних розбіжностей (*Relative Homogeneity Test, RHT*), порівнюються середні значення двох періодів та оцінюється ймовірність того, що вони різняться саме через неоднорідність даних. *RHtestsV4* розроблено науковцями Канадського метеорологічного центру (*Canadian Meteorological Centre*) [39].

ACMANT (Analysis of Climatological Mean and Anomalous Trends) є програмним забезпеченням, створеним Німецьким метеорологічним центром, що також призначене для гомогенізації кліматичних даних. Програма базується на статистичному методі, відомому як "модель змін", який використовується для визначення динаміки середніх значень та дисперсій кліматичних рядів. *ACMANT* включає різноманітні інструменти для аналізу та візуалізації даних, а також можливість експорту результатів для подальшого аналізу іншим програмним забезпеченням і також входить до групи відносних методів. Програмне забезпечення використовується для виявлення та коригування неточностей та розривів у кліматичних даних, а також для аналізу трендів у кліматичних рядах. В основі *ACMANT* лежать статистичні методи, такі як аналіз відношень середніх значень рядів, тест Фішера і різні методи регресійного аналізу. Результатом роботи цієї програми є гомогенізовані кліматичні дані, які можна використовувати для вивчення змін клімату за тривалі періоду часу. [19].

HOMER (HOMogenization softwarE in R) — це інтерактивне, напівавтоматичне сучасне програмне забезпечення, яке було розроблено для гомогенізації місячних та річних даних про атмосферні опади

або температуру повітря. HOMER був розроблений у 2007–2011 роках, під час виконання проєкту COST Action ES0601: (HOME) [25]. Основною метою проєкту було створення набору методів для виявлення та корекції кліматичних змін різних просторових та часових масштабів. Результатом виконання проєкту став синтез найбільш ефективних статистичних процедур, які існували на той час і були покладені в основу роботи програмного забезпечення HOMER.

Перелічені програмні продукти використовують різні методики для гомогенізації даних та надають різні можливості для користувачів. Вибір конкретного програмного забезпечення залежить від індивідуальних потреб користувача та специфіки дослідження, що проводиться.

Метою даного дослідження є аналіз програмних продуктів, що знаходяться у вільному доступі і використовуються для перевірки якості та гомогенізації часових рядів метеорологічних величин і, зокрема, включають прийоми та інструменти для гомогенізації рядів сум атмосферних опадів. Крім того, важливим завданням є підтвердження можливості використання обраного програмного продукту HOMER для проведення процедури гомогенізації місячних сум опадів, на прикладі тестування часових рядів отриманих на метеорологічних станціях України протягом 1946–2020 рр.

МАТЕРІАЛИ ТА МЕТОДИ ДОСЛІДЖЕННЯ

Для апробації програмного забезпечення HOMER було використано фактичні дані спостережень на 224 пунктах (метеорологічні станції та опадомірні пости) метеорологічної мережі України. Початкові дані отримані з таблиць ТМ-1 та ТМС архіву Центральної геофізичної обсерваторії імені Бориса Срезневського (ЦГО). Масив часових рядів щомісячних сум опадів за період 1946–2020 рр. сформований за результатами спостережень на 177 метеорологічних станціях та 47 опадомірних постах. Зауважимо, що до досліджуваного масиву включалися дані тільки окремих постів. Були задіяні лише ті, в часових рядах яких пропуски виявилися незначними і не перевищували 20% загальної кількості спостережень за обраний період.

Важливо відмітити ті зміни в методиках вимірювання та типах приладів, що використовувалися на метеорологічній мережі для вимірювання опадів, які могли вплинути на однорідність результатів вимірювань. До основних груп змін у проведенні спостережень за атмосферними опадами протягом досліджуваного періоду (1946–2020 рр.), можемо віднести наступні:

1. Так звані масові зміни, що відбувались на всіх метеостанціях майже одночасно. Сюди входить планова заміна приладів на усій мережі, а також зміни методики спостережень. На усіх метеостанціях в Україні, у 1950-х роках, було проведено заміну дощоміра Ніфера на опадомір Третьякова, що очевидно призвело до порушень гомогенності часових рядів. В 1966 р. відбувся перехід до чотириразових спостережень за опадами впродовж доби, що могло вплинути на точність визначення середніх добових та місячних значень. З 1967 р. до результатів вимірювань почали додавати поправку на змочування для всіх видів опадів. Проте, усі отримані у той час дані були перевірені на наявність викидів та розривів і для усунення виявлених неоднорідностей були введені спеціальні поправочні коефіцієнти для рядів місячних сум опадів. Значення коефіцієнтів подаються у кліматичних довідниках.

2. Локальні зміни, які фіксуються при вимірюваннях і викликані проблемами роботи приладів на окремих станціях у різні періоди. До цієї групи відноситься також перенесення метеостанцій на нове місце, зміни ландшафту біля метеомайданчика, зміна ступеню його відкритості.

Саме друга група змін умов вимірювання, а також наявність технічних помилок спостерігачів часто є причиною неправильних результатів спостережень, що порушують однорідність рядів даних та потребують їх перевірки.

Порівняльний аналіз існуючого спеціалізованого програмного забезпечення показав перевагу продукту HOMER для гомогенізації часових рядів опадів. Зазначене програмне забезпечення базується на *відносних методах* гомогенізації кліматичних даних, що, як зазначено вище, є найбільш ефективними у роботі з часовими рядами сум атмосферних опадів. Більше того, використання цього продукту вже показало хороші результати у проведенні аналізу на однорідність часових рядів температури повітря для території України, про що свідчать публікації вітчизняних науковців [4, 7]. HOMER також часто використовується міжнародними фахівцями для гомогенізації метеорологічних даних, отриманих на мережі метеостанцій в інших регіонах. Цей програмний продукт застосовується для оцінки однорідності не лише середньої температури повітря (Танзанія) [33], але і місячних максимальних та мінімальних температур (Піреней) [34]. Також опубліковані результати адаптації програмного продукту для аналізу даних спостережень за атмосферними опадами [13, 42].

Програмне забезпечення HOMER складається з ряду функцій повного процесу гомогенізації, серед яких наступні: контроль якості даних (*quality control*), що використовує покращений алгоритм

програми Climatol; виявлення та корекція помилок даних за допомогою адаптованих алгоритмів, взятих з PRODIGE (PRocedure Of Detection of Inhomogeneity) та ACMANT [19], таких як попарне порівняння (*pairwise-detection*), об'єднане порівняння (*joint detection*), а також використання моделі ANOVA (ANalysis Of VAriance) для корекції виявлених розривів однорідності ряду за допомогою дисперсійного аналізу. Крім того, HOMER надає широкий набір інструментів для аналізу та візуалізації набору даних у зручному форматі.

HOMER використовує мову програмування R. Зазначимо, що R — це мова програмування, яка використовується для статистичних обчислень і побудови графіки. Вона повністю підтримується основними операційними системами, такими як Windows, Linux, MacOS і має мінімальні відмінності, які не помітні кінцевому користувачеві. Для роботи з програмним забезпеченням HOMER потрібно мати встановлену версію R 2.15.3, або наступні варіанти версій, що зазначено у вимогах від розробників [31].

Програмне забезпечення HOMER використовує відносні гомогенізаційні методи [38], тому всі статистичні розрахунки в ньому виконують не для окремих рядів, а для певної обраної кількості або різницевого рядів для адитивних величин, або рядів відношень (для кумулятивних величин). Тобто, для гомогенізації довільного ряду за певною ознакою завжди має бути обрано деяку кількість інших

контрольних рядів, а потім розраховуються їх різниці або відношення, які будуть проаналізовані[3].

Проект COST Action ES0601 (HOME) започаткував стандартизацію формату великого об'єму місячних вхідних даних задля усунення проблем з сумісністю при використанні багатьох програм та проєктів, що у результаті дозволило пришвидшити взаємодію і зменшити час на підготовку і обробку вхідних даних.

HOMER також працює з рядами даних станцій, які згруповані у так звану "мережу" [27]. Мережа складається з декількох станцій, і на кожен окрему станцію формується свій файл з даними. Щоб задати мережу для HOMER, треба забезпечити файл зі списком станцій з необхідними метеорологічними показниками. Кожна мережа має містити лише один файл статусу (початковий або гомогенізований), і всі перераховані файли обов'язково повинні утримувати одну кліматичну змінну. Наприклад, мережа не може містити одночасно файли з даними про опади та тиск, для такого випадку мають бути створені окремі мережі.

Основна взаємодія між користувачем і HOMER здійснюється шляхом відповідей на запитання, що з'являються на екрані під час виконання послідовних опцій. Користувачу пропонується обрати потрібну процедуру шляхом натискання на відповідну кнопку на клавіатурі. Після завершення вибраної процедури, знову з'явиться головне меню HOMER, готове для наступного запиту. Меню поділене на дві секції: процедури контролю якості та гомогенізація (рис. 1).

Перший етап роботи — виявлення та видалення так званих викидів (*outliers*). В англомовній літературі цей процес називають контролем якості (*quality control* — QC). Викиди повинні бути вилучені до процедури гомогенізації, тому що вони можуть суттєво вплинути на її результати [7].

HOMER має дві процедури для контролю якості: набір статистичних процедур, які запозичені з програми Climatol, та власний швидкий контроль якості (*fast quality control*).

Основи контролю якості Fast Climatol. Fast Climatol передбачають статистичні розрахунки та відповідну їх візуалізацію. Результати накопичуються у вигляді pdf-файлів у відповідній папці, де розміщено набір графіків статистичного аналізу вхідних даних. Серед графічних зображень: гістограми наявної мережі; корелограма мережі, яка показує тісноту зв'язку між станціями в залежності від географічної

```

What do you wish, Master/Mistress?
FAST QUALITY CONTROL
-> Fast CLIMATOL checks           type i
-> Fast QC                         type f
-> Outlier file creation?         type o
-> Removal of outliers?           type r

HOMOGENISATION
-> Pairwise detection?            type d
-> Joint detection?                type j
-> Correction?                     type c
-> Visualization?                 type v
-> New neighbourhood              type n
-> Change hinteraction hoption :-) type h
-> Break file creation/modification? type b
-> Break file edition?            type e
-> Quit?                            type q

Your choice :|

```

Рис. 1. Вигляд основного меню HOMER

відстані. Прийнято вважати, що низька кореляція ($r < 0,2$) вказує на можливі помилки в результатах гомогенізації, якщо станції знаходяться у зоні низьких коефіцієнтів; дендрограма мережі, що демонструє зв'язки між рядами даних окремих станцій, які отримані після проведеного кластерного аналізу набору відповідних спостережень. Станції групуються в один кластер (верхня частина) і розділені на відповідну кількість груп за критерієм подібності статистичних параметрів рядів спостережень (нижня частина). Гілки дендрограми розділяють станції відповідно до міри статистичної відмінності між ними; розташування мережі, де зображена географічна карта з нанесеними станціями відповідно до їх географічних координат [3].

Далі проводиться етап швидкого контролю якості (Fast QC), що є стартовою фазою гомогенізації. На цьому етапі початкові дані опорної станції порівнюються з оточуючими станціями для виявлення механічних помилок на етапі вимірювання або їх фіксації. Для кожної опорної станції можливо обрати оточуючі станції двома шляхами — визначення коефіцієнту кореляції, або географічна близькість декількох станцій. Програмою розраховуються ряди аномалій для кожної станції і кожного місяця року окремо.

Безпосередньо до процедури гомогенізації у програмному продукті, відносять три основні етапи: попарне порівняння (*pairwise detection*), об'єднане порівняння (*joint detection*) та корекція (*correction*).

На етапі попарного порівняння для часових рядів сум опадів розглядаються відношення значень основної станції (*candidate station*) та її оточення (*referens station*) [7]. При цьому, програмним аналізом проводиться тільки ідентифікація можливих точок розриву. Вважається, що появу точок розриву в кожній парі основної станції та станції оточення можуть обумовлювати наступні причини: неоднорідності ряду вимірювань станції-кандидата, неоднорідність ряду опорної станції або помилкового чи справжнього кліматичного ефекту. Розрив буде зафіксований достовірним/обґрунтованим, якщо така позиція збігається у більшості рядів окремих станцій оточення (як мінімум у трьох рядах) [28].

Якщо користувач вибрав параметр “без взаємодії”, HOMER швидко обчислить, побудує та збереже в каталозі fig порівняння для кожної станції відповідно до попередньо вибраного методу вибору станцій оточення. Якщо користувач вибрав інтерактивну опцію, графіки не лише зберігатимуться на жорсткому диску користувача, але й для кожної станції-кандидата графік виявлення також з'являтиметься на екрані, потрібно буде клацнути мишею, щоб перейти до наступної станції. Для полегшення

аналізу використовують ряди відношень місячних сум опадів між основною станцією та станціями оточення та розміщують їх у порядку зростання середнього квадратичного відхилення (σ).

Наступним етапом є етап об'єданого порівняння. Тут важливо підкреслити, що попарне виявлення та об'єдане порівняння є окремими процедурами. HOMER використовує алгоритм *cgHseg*, який був створений біологом-науковцем для сегментації ДНК [36]. Виявлення попарного порівняння надає користувачеві графічний інструмент для візуального визначення точок розриву, тоді як об'єдане порівняння створить список точок розриву, а також, якщо потрібна взаємодія, інтерактивну графіку. Створені точки розриву у файлах можуть бути відредаговані користувачем, за умови, що їх формат збережено. Якщо вибрано інтерактивну опцію, то дослідник може використовувати систему на основі графіка, щоб додавати/видаляти контрольні точки. Для рядів сум опадів можливість редагування користувачем списку точок розриву має вагоме значення, бо надає можливість додати (або видалити) точку, виходячи з наявних знань користувача про природну мінливість опадів у досліджуваному регіоні.

Після визначення точок розриву у автоматичному, або інтерактивному режимі, та ручного виправлення точок у текстових файлах, ця інформація буде використана під час наступного етапу — корекції. Корекція використовує двофакторну модель ANOVA, адаптовану з PRODIGE. Детальний опис алгоритму описаний у роботі [32]. Наприкінці процедури виправлення над отриманими даними буде виконано процедуру парного порівняння. Це дозволяє користувачеві визначити потенційні залишкові неоднорідності. Якщо інтерактивний параметр увімкнено, графіки з'являтимуться на екрані так само, як і під час виконання процедур попарного порівняння.

Таким чином, HOMER може використовуватись як для простої перевірки на однорідність часових рядів, так і для проведення автоматизованого виявлення неточних даних та/або пропусків у спостереженнях так і їх корекції.

РЕЗУЛЬТАТИ ТА ЇХ ОБГОВОРЕННЯ

На основі даних спостережень за опадами протягом 1946–2020 рр. був сформований архів місячних сум за правилами HOMER. Після формування масиву даних та обрання робочої директорії проводиться початковий етап розрахунків — статистична оцінка даних (*Fast Climatol Check*), яка, серед іншого, дозволяє отримати загальну закономірність та окремі параметри розподілу досліджуваних часо-

вих рядів місячних сум атмосферних опадів. Формується масив діаграм, що представляє розподіл (повторюваність значень) сум опадів окремо для кожного місяця періоду досліджень, а також усереднений розподіл за весь період спостережень на усіх станціях. Уявлення про закономірність розподілу досліджуваної величини важливе і необхідне користувачу для вибору наступних опцій програмного забезпечення.

На рис. 2 наведено усереднений розподіл, який дозволяє констатувати, що максимум повторюваності місячних сум опадів у масиві даних знаходиться у діапазоні від 20 до 40 мм, а найнижчу ймовірність мають місячні суми опадів, значення яких перевищують 220–240 мм. На рис. 2 виділена градація значень з сумами опадів меншими 0,0 (мм), яка формується за рахунок пропусків даних, що неминуче зустрічаються у будь-якому великому масиві. При формуванні масиву для HOMER відсутні дані позначаються числом –999,9. Очевидно, що кількість пропусків є відносно невеликою у порівнянні з загальним обсягом даних.

Далі проводиться загальний контроль (Fast QC) якості початкових даних, для виявлення некліматичних похибок, які могли бути зроблені під час вимірювань і обробки даних. Основним методом пошуку аномалій (або викидів) на цьому етапі є порівняння значень, отриманих на опорному пункті (станції), зі значеннями на пунктах його оточення.

Зазначимо, що існує декілька способів вибору пунктів оточення, однак у даному дослідженні використовувався метод кореляції, який ґрунтується на розрахунку кореляційного зв'язку між часовими рядами даних опорного пункту та його оточення. Важливо відзначити, що граничне значення коефіці-

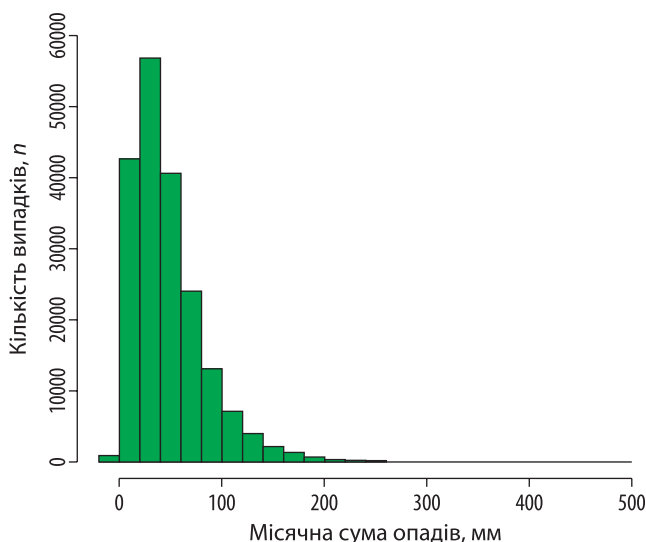


Рис. 2. Розподіл місячних сум опадів протягом 1946–2020 рр.

ента кореляції визначається дослідником. У нашому випадку було обрано значення 0,7, оскільки значення нижче цієї межі можуть призвести до включення у список оточення станцій, які мають принципово інші умови процесів утворення опадів.

У результаті проведення необхідних розрахунків HOMER створює відповідні текстові файли (рис. 3), де формується список станцій оточення, виявлених для опорної станції, та зазначається величина кореляційного зв'язку між ними.

Окрім цього, програмою подаються результати порівняння часових рядів станцій у вигляді графіків, на яких також зазначаються, так звані, викиди (відхилення). При цьому, величина такого викиду оцінюється за допомогою розрахунку відношень усереднених значень часового ряду опорної станції до її оточення. За допомогою графічних візуалізацій дослідник може проводити оцінку даних та контроль їх якості, розглядаючи кожний випадок окремо та самостійно встановлюючи граничну межу величини викиду. У нашому випадку, після аналізу отриманої інформації про встановлену мінливість опадів досліджуваного регіону [1, 6] було вирішено детально оцінювати і корегувати ті відхилення (викиди), величина яких лежить у проміжку від –1,5 і менше до 1,5 і більше. Виявлені відхилення у масиві даних перевірялися і корегувалися з використанням таблиць ТМ-1.

```

00033651 Kolomya
=====
CORRECTION NEIGHBORHOOD
00033526 0.877 Ivano-Frankivsk
00033658 0.825 Chernivtsi
00033524 0.821 Dolyna
00000078 0.812 Dora
00033645 0.810 Yaremche
00000056 0.799 Pidgajtsi
00033536 0.781 Chortkiv
00033657 0.774 Seliatyn
00000064 0.766 Galych
00033409 0.760 Berezhany
00000053 0.732 Zhvanets
00033297 0.726 Brody
00033557 0.716 Nova-Ushytsya
00033548 0.713 Kamyanets-Podilsky
00033421 0.703 Yampil
=====
Periode 1
00033651 1946 - 2020
UNCHANGED REFERENCE PERIOD
    
```

Рис. 3. Текстовий файл, що сформований на етапі Fast QC із переліком пунктів оточення для станції Коломия

Після оцінки якості рядів місячних сум опадів за весь досліджуваний період та виправлення виявлених помилок, наступним етапом стає можливість провести аналіз часових рядів станцій та перевірити їх однорідність. Перевірка однорідності ґрунтується на гіпотезі, яка стверджує, що у випадку однорідності часових рядів на двох сусідніх станціях, послідовність сформована з відношень значень сум опадів на цих станціях за тривалий період набуває узгодженого характеру. Однак, якщо величина відношення у якийсь момент часу показує різкі зміни, це свідчитиме про порушення однорідності в одному із порівнюваних рядів. Коли факт порушення однорідності у ряді встановлено, для його виправлення застосовуються статистичні методи [11].

Сама процедура гомогенізації має на меті виявлення впливу на однорідність часових рядів сум опадів тільки тих факторів які не пов'язані з коливаннями глобального чи регіонального клімату, а формуються завдяки технічним чи мікрокліматичним умовам вимірювання метеорологічних величин. До них можна віднести переміщення метеостанцій чи метеопостів, а також зміну природного ландшафту в районі метеорологічної станції. На основі існуючої інформації [32], можна стверджувати, що найчастіше за обраний період відзначаються переміщення станцій на відстань від десятків метрів до декількох кілометрів, особливо це було актуально у 1950–1960-х роках. Але подальший аналіз матеріалу показав, що неоднорідність у часових рядах може бути спричинена й іншими факторами. Наприклад, це може статися, коли порушується типовий річний розподіл опадів, характерний для даного регіону. Іншими словами, питання щодо причин, які призводять до порушення однорідності часових рядів опадів, є складним і вимагає подальших наукових досліджень [5].

Перевірка рядів на однорідність у HOMER починається з попарного порівняння (*pairwise detection*), під час якого здійснюється розрахунок та аналіз відношень між часовими рядами сум опадів основної станції та станцій оточення.

На цьому етапі програма лише виокремлює можливі точки розриву в рядах даних, а рішення про їх підтвердження приймається на наступному етапі — об'єднаного порівняння (*joint detection*).

Дослідник обирає, яким методом встановлення точок розриву буде користуватись. У варіанті інтерактивного методу точки розриву встановлюються безпосередньо дослідником, який приймає рішення на підставі власних міркувань, попереднього аналізу даних та порівнянні результатів з наявною інформацією. При автоматизованому варіанті програма виконує це самостійно. У нашому випадку, було обрано перший метод, тобто всі можливі точки розриву підтверджувались конкретно дослідником. При цьому, після проведеного попереднього аналізу, точками розриву було вирішено визначати такі, що повторювались на декількох (2–3) станціях в один момент. Результатом цього процесу є сформований програмою текстовий файл у робочій директорії, який містить список станцій, де були відмічені точки розриву однорідності у часових рядах опадів.

Під час попарного порівняння програма також генерує файли для кожного наявного пункту спостережень, у яких зберігається візуалізація результатів перевірки на однорідність рядів даних.

Після того, як точки розриву були зафіксовані та підтверджені, відбувається корегування неоднорідності часового ряду, за допомогою методів дисперсійного аналізу визначається величина амплітуди зміщення (зсуву) на ділянці неоднорідності, а корекція зсуву проводиться за допомогою двофакторної моделі ANOVA. При цьому, амплітуда зсуву є безрозмірною величиною, оскільки програма визначає її за співвідношенням сум опадів на двох станціях (мм/мм). Тобто, суть корекції полягає у зменшенні або збільшенні місячних сум опадів за певний період часу між точками розриву. Текстові файли зі скорегованими даними відображаються в робочій директорії та є доступними для перегляду і подальшої обробки.

Для прикладу розглянемо часовий ряд сум опадів станції Дніпро, що має зафіксовану і підтверджену точку розриву у 1973 році. За допомогою проведеного аналізу наявної в матеріалах інформації, причиною виникнення цієї точки було вирішено вважати перенесення станції, яке, згідно з довідником [2], мало місце в 1974 р. та могло вплинути на отримані результати вимірювань.

У *табл. 1* для порівняння представлені початкові та скореговані (гомогенізовані) місячні й річні суми

Таблиця 1. Результати гомогенізації місячних сум опадів (мм), метеорологічна станція Дніпро, 1973 рік

	I, мм	II, мм	III, мм	IV, мм	V, мм	VI, мм	VII, мм	VIII, мм	IX, мм	X, мм	XI, мм	XII, мм	Сума, мм
1973 р. (почат.)	31	60	62	19	128	25	47	69	5	48	38	41	573
1973 р. (гомог.)	34	67	69	21	142	28	52	77	6	53	42	46	635

Таблиця 2. **Результати гомогенізації місячних сум опадів (мм), метеорологічна станція Приколотне, 1994 рік**

	I, мм	II, мм	III, мм	IV, мм	V, мм	VI, мм	VII, мм	VIII, мм	IX, мм	X, мм	XI, мм	XII, мм	Сума, мм
1994 р. (почат.)	32	17	50	30	65	33	16	86	5	92	21	33	480
1994 р. (гомог.)	28	15	44	26	57	29	14	75	4	80	18	29	419

опадів за 1973 р. для станції Дніпро, коли було встановлено точку розриву у цей рік. Наведені значення гомогенізованих сум відрізняється від початкових, причому річна сума збільшилась на 10,8%. Очевидно, що процес вирівнювання ряду залежить від початкових значень величин і зі зменшенням суми початкових опадів, зменшується і величина її зміни при гомогенізації. Таким чином, величина скорегованої суми опадів січня збільшилась на 9,6%, в той час як травня — на 10,9%.

Іншим прикладом може бути точка розриву 1994 р. для станції Приколотне. У *табл. 2* наведені значення гомогенізованих місячних та річної сум,

які мають нижчі значення, порівняно з початковими даними, при чому річна сума зменшилась на 12,7%.

Із наведених у *табл. 1, 2* гомогенізованих часових рядів видно, що їх корекція відбувається з урахуванням величини і знаку амплітуди зміщення та початкових значень величин ряду до процедури гомогенізації. Зазначимо, що введення поправок, відбувається тільки для окремих ділянок часових рядів, а розрахунки їх величин проводяться окремо для кожного випадку.

На *рис. 4* наведені текстові файли, які подаються програмою на етапі Fast QC, де вказуються відпо-

```

00034504 Dnipro
=====
CORRECTION NEIGHBORHOOD
00034505 0.846 SyneInykove
00034407 0.822 Gubynykha
00034601 0.815 Zaporijhjhya
00033805 0.782 Nikopol
00034502 0.732 Pavlograd
00033705 0.724 Pomichna
00000273 0.723 Vasylkivka
00033848 0.720 Ochakiv
00033723 0.718 Komisarivka
00034401 0.717 Krasnograd
00033621 0.717 Kobelyaky
00034415 0.712 Izum
00033801 0.703 Loshkarivka
=====

Periode 1
00034504 1946 - 1973
13 1.109 522.522
Amplitude : 0.11

Periode 2
00034504 1974 - 2020
UNCHANGED REFERENCE PERIOD
    
```

a

```

00034301 Prykolotne
=====
CORRECTION NEIGHBORHOOD
00034319 0.789 Kupyansk
00034208 0.756 Zolochiv
00034409 0.755 Lozova
00034401 0.752 Krasnograd
00034415 0.747 Izum
00033621 0.747 Kobelyaky
00034300 0.747 Kharkiv
00034421 0.738 Svatove
00034317 0.733 Komsomolske
00033506 0.731 Poltava
00000265 0.709 Verkhnodniprovsk
00011111 0.701 Amvrosiivka
=====

Periode 1
00034301 1946 - 1994
13 0.873 580.174
Amplitude : -0.13

Periode 2
00034301 1995 - 2020
UNCHANGED REFERENCE PERIOD
    
```

б

Рис. 4. Програмні файли зі встановленими проміжками часу, для яких проводилась корекція, з зазначеною величиною амплітуди протягом 1946–2020 рр.: а — для станції Дніпро; б — для станції Приколотне

відні періоди часових рядів, для яких проводилась корекція даних. Також для них вказується величина та знак амплітуди. Бачимо, що амплітуда для Дніпра є додатньою, в той час як для Приколотного — від'ємною, що підтверджує попередні розрахунки.

Крім текстових файлів, програмне забезпечення HOMER також надає візуалізації розподілу річних сум атмосферних опадів для початкових та гомогенізованих даних окремо. Далі на *рис. 5* представлений такий часовий розподіл річних сум за період з 1946 по 2020 р. для станції Дніпро показані як початкові так і гомогенізовані дані. Оскільки в 1973 р. було зафіксовано точку розриву, то програмою були скореговані місячні та, відповідно, річні суми за попередній до точки період, тобто за 1946–1973 рр.

На *рис. 6* представлений аналогічний розподіл початкових та гомогенізованих рядів річних сум для станції Приколотне. У порівнянні з попередньою станцією тут корегувався довший відрізок часового ряду (48 років) і амплітуда зсуву була від'ємною, але мала трохи вище абсолютне значення. Візуально результат гомогенізації проявляється у зменшенні максимальних сум опадів впродовж 1946–1994 рр.

Проводяться розрахунки лінійного та поліноміального трендів для часових рядів річних сум опадів до і після процедури гомогенізації і візуалізуються як у вигляді кривих (*рис. 5, б*), так і значень коефіцієнтів відповідних рівнянь. Очевидно, що корекція рядів місячних і річних сум опадів спричиняє зміни значень коефіцієнтів лінійних та поліноміальних трендів розподілу опадів у часі, і змінює форми їх кривих. У цілому, для усього масиву даних отримуємо деяке згладжування екстремумів та зменшення мінливості показника, як у часовому так і у просторовому вимірі. Тут є певні неоднозначності, бо проведення процедури гомогенізації показало, що автоматичне спільне виявлення (*joint detection*) не є ідеальним і тому, можемо у результаті набуття однорідності рядом, втратити ті зміни і коливання, що уже безпосередньо пов'язані зі змінами регіонального клімату. Дослідник, маючи відповідну кваліфікацію й необхідну інформацію про закономірності розподілу опадів та міру їх мінливості у регіоні, повинен додатково провести оцінювання програмних змін ряду, у тому числі й враховуючи значення коефіцієнтів трендів, щоб по можливості не допустити надмірне "згладжування" часового ряду.

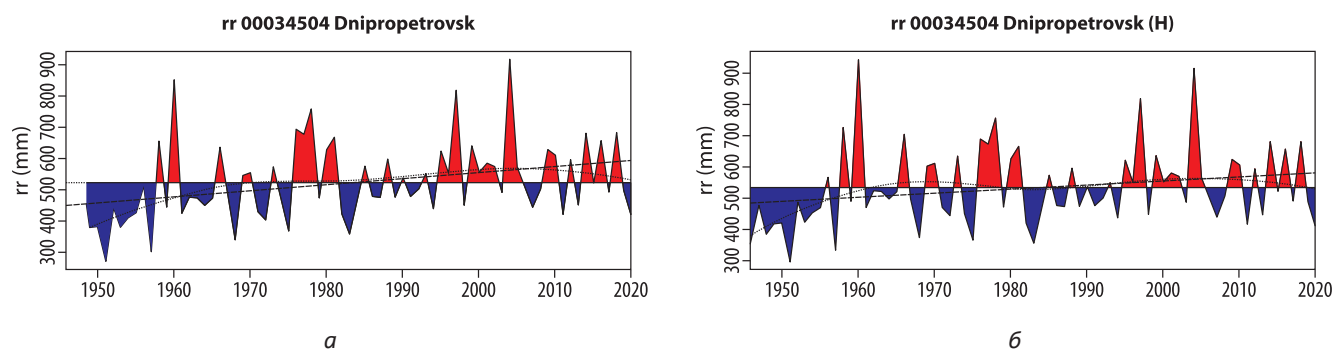


Рис. 5. Часовий розподіл річних сум опадів за період з 1946 р. по 2020 р., ст. Дніпро: *a* — до гомогенізації; *б* — після гомогенізації

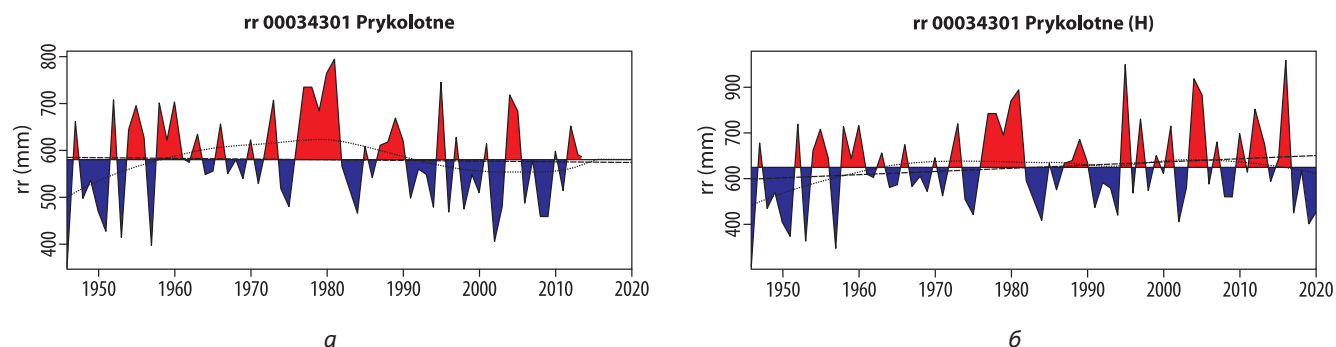


Рис. 6. Часовий розподіл річних сум опадів за період з 1946 р. по 2020 р., ст. Приколотне: *a* — до гомогенізації; *б* — після гомогенізації

ВИСНОВКИ

У зв'язку з необхідністю виконання сучасних вимог ВМО до якості та представлення даних спостережень, що використовуються при проведенні кліматичних досліджень (дані повинні отримуватися із застосуванням стандартних методик спостережень, не мати пропусків, пройти перевірку якості та однорідності), було виконано аналіз сучасних програмних продуктів та методів, які можуть забезпечити реалізацію необхідних процедур для часових рядів атмосферних опадів.

Аналізувалися найбільш поширені програмні продукти, що є у вільному доступі та найчастіше використовуються у роботах вітчизняних і зарубіжних дослідників: MASH, ACMANT, RHtestsV4, HOMER.

Для проведення гомогенізації часових рядів опадів у дослідженні використано програмне забезпечення HOMER, як продукт, що включає в себе необхідні функції для контролю якості та для повного процесу гомогенізації, використовує покращений алгоритм програми Climatol, проводить виявлення та корекцію помилок даних за допомогою адаптованих алгоритмів, взятих з PRODIGE та ACMANT, виявляє точки розриву (ділянки неоднорідності) у часових рядах та виконує корекцію часового ряду, використовуючи модель ANOVA.

Тестування процедури гомогенізації часових рядів місячних сум опадів проведено на основі даних спостережень метеорологічної мережі України за період 1946–2020 років. Гомогенізації рядів місячних сум опадів виконувалася шляхом створення для кожної станції окремої групи оточення, визначеної з урахуванням тісноти кореляційних зв'язків між рядами сум опадів у кожній парі станцій. До групи входило по 6–14 станцій.

Виконана статистична обробка відношень часових рядів для кожної пари станцій, установлювалися можливі ділянки неоднорідності рядів (точки розриву) та визначалася величина амплітуди зміщення (зсуву). У досліджуваному масиві спостережень величини амплітуда коливалися у межах 0,1–0,7.

Після фіксації точки розриву виконувалася корекція часового ряду в межах ділянки встановленої неоднорідності. Збільшення/зменшення місячних сум опадів залежало від величини та знаку амплітуди зміщення і становило 10–20% початкових сум опадів.

HOMER є простим у використанні та ефективним інструментом обробки і аналізу первинних масивів даних спостережень. Крім виконання статистичних розрахунків забезпечує якісну візуалізацію отриманих результатів, та надає можливість для уточнення результату проводити повторну процедуру гомогенізації. Для проведення статистичного аналізу програма може формувати як різниці ряди для адитивних величин (температура) так і ряди відношень для кумулятивних величин (опадів), що розширює можливості її застосування.

До переваг використання HOMER можна віднести можливість для дослідника приймати рішення про підтвердження (фіксацію) точок розриву на етапі об'єднаного порівняння (*joint detection*), що дозволяє враховувати наявну інформацію про природну мінливість значень сум опадів у регіоні та не допускати надмірного згладжування їх величин.

Отримані методичні прийоми та підходи до перевірки якості даних та їх гомогенізації з використанням програмного продукту HOMER можуть застосовуватися у подальших дослідженнях опадів.

ЛІТЕРАТУРА / REFERENCES

1. Балабух В.О. Мінливість дуже сильних дощів та сильних злив в Україні. *Наукові праці УкрНДГМІ*. 2008. **Вип. 257**. С. 61–72.
[Balabukh, V. (2004). Trajektorii tsykloniv, scho zumovliuiut' nebezpechnu i stykhijnu kil'kist' opadiv v Ukraini u teplotnyy period roku. *Naukovi pratsi UkrNDHMI*. **253**. 37–49].
2. Кліматичний довідник. Історія та фізико-географічний опис метеорологічних станцій України. К., 2011. 462 с.
[Klimatychnyi dovidnyk. Istorii ta fizyko-geohrafichnyi opys meteorolohichnykh stantsii Ukrainy. Kyiv, 2011. 462].
3. Киреєва З.М. Гомогенізація щомісячних рядів опадів за допомогою програмного забезпечення HOMER: кваліфікаційна робота магістра: 103 Науки про Землю / Киреєва Зорина Махмудсалімівна. Київ, 2023. 72 с. URL: <https://ir.library.knu.ua/items/cfae8b91-c8ef-4d8b-a55d-024adb2eb00d>.
[Kyreieva, Z.M. (2023). Homohenzatsiia shchomisiachnykh riadiv opadiv za dopomohoiu prohramnoho zabezpechennia HOMER: kvalifikatsiina robota mahistra: 103 Nauky pro Zemliu / Kyreieva Zoryna Makhmadsalimivna. Kyiv. 72].
4. Осадчий В.І. Скриник О.А., Сіденко В.П., Бойчук Д.О., Ошурок Д.О., Скриник О.Я. Гомогенізована база довгих рядів середньої місячної температури повітря. *Геоінформатика*. **1 (65)**. 2018. С. 54–68.
[Osadchyi, V.I., Skrynyk, O.A., Sidenko, V.P., Boichuk, D.O., Oshurok, D.O., Skrynyk, O.Y. (2018). Homohenzovana baza dovhykh riadiv serednoi misiachnoi temperatury povitria. *Heoinformatyka*. **1 (65)**. 54–68].
5. Паламарчук Л.В., Осадчий В.І., Скриник О.А., Киреєва З.М., Сіденко В.П., Ошурок Д.О., Скриник О.Я. Використання програмного забезпечення HOMER для контролю якості та гомогенізації рядів щомісячних сум опадів. *Гідрологія, гідрохімія і гідроекологія*. 2023. **1 (67)**. С. 58–77. DOI: <https://doi.org/10.17721/2306-5680.2023.1.7>.
[Palamarchuk, L.V., Osadchyi, V.I., Skrynyk, O.A., Kyreieva, Z.M., Sidenko, V.P., Oshurok, D.O., Skrynyk, O.Y. (2023). Vykorystannia prohramnoho zabezpechennia HOMER dlia kontroliu yakosti ta homohenzatsiia riadiv shchomisiachnykh sum opadiv. *Hidrolohiia, hidrokhiimia i hidroekolohiia*. **1 (67)**. 58–77].

6. Паламарчук Л.В., Сокур К.С., Заболоцька Т.М. Динаміка інтенсивності опадів та мезоструктурні особливості їх полів у теплий періоду року на рівнинній частині території України. *Гідрологія, гідрохімія і гідрогеологія*. 2019. **Вип. 4**. С. 95–111.
[Palamarchuk, L.V., Sokur, K.S., Zabolotska, T.M. (2019). Dynamika intensyvnosti opadiv ta mezostrukturni osoblyvosti yikh poliv u teplyi periodu roku na rivnynnii chastyni terytorii Ukrainy. *Hidrolohiia, hidrokhiimiia i hidroheolohiia*. **4** (55). 95–111].
7. Скриник О.А., Бойчук Д.О., Сіденко В.П. Виявлення та усунення кліматологічної неоднорідності у часових рядах кліматологічних показників. *Гідрологія, гідрохімія і гідрогеологія*. 2019. **Вип. 2** (53). С. 88–100.
[Skrynyk, O.A., Boichuk, D.O., Sidenko, V.P. (2019). Vyivlennia ta usunennia klimatolohichnoi neodnorodnosti u chasovykh riadakh klimatolohichnykh pokaznykiv. *Hidrolohiia, hidrokhiimiia i hidroheolohiia*. **2** (53). 88–100].
8. Alexandersson, A. (1986). A homogeneity test applied to precipitation data. *J. Climatol.* **6**. 661–675.
9. Alexandersson, H., Moberg, A. (1997). Homogenization of Swedish temperature data. 1. Homogeneity test for linear trends. *Int. J. Climatol.* **17**. 25–34.
10. Brückner, E. (1890). Klimaschwankungen seit 1700 nebst Bemerkungen über Klimaschwankungen der Diluvialzeit. E.D. Hölzel, Wien and Olnütz.
11. Cao, L.-J., Z.-W. (2012). Progress in research on homogenization of climate data. *Adv. Clim. Change Res.* **3**. Iss. 2. 59–67. DOI:10.3724/SP.J.1248.2012.00059.
12. Caussinus, H., Mestre O. (2004). Detection and correction of artificial shifts in climate series. *Appl. Statist.* **53**. Part 3. 405–425.
13. Coll, J., Curley, M., Walsh, S., Sweeney, J. (2014). Ireland with HOMER.
14. Conrad, V. (1944). *Methods in Climatology*. Harvard University Press, 228 p.
15. Conrad, V., Schreier O. (1927). Die Anwendung des Abbe'schen Kriteriums auf physikalische Beobachtungsreihen. *Gerland's Beiträge zur Geophysik*, XVII, 372.
16. Conrad, V., Pollak C. (1950). *Methods in Climatology*. Harvard University Press, Cambridge, MA, 459 p.
17. Conrad, V., (1925). Homogenitätsbestimmung meteorologischer Beobachtungsreihen. *Meteorologische Zeitschrift*, 482–485.
18. Cowtan, K.D. (2015). Homogenization of Temperature Data: An Assessment.
19. Domonkos, P. (2014). The ACMANT2 software package. Eighth Seminar for Homogenization and Quality Control in Climatological Databases and Third Conference on Spatial Interpolation Techniques In Climatology and Meteorology. WMO, WCDMP. **84**. 46–72.
20. Gorbachva, L., Khrystiuk, B., Shpyg, V., Pishniak, D. (2022). Estimation of tendencies homogeneity and stationarity of air temperature at the Ukrainian Antarctic "Akademik Vernadsky" station during 1951–2020. *Геофізичний журнал*. **T. 44**. 4. С. 183–194. DOI: <https://doi.org/10.24028/gjv44i4.264848>
21. Hann, J. (1880). Untersuchungen über die Regenverhältnisse von Österreich-Ungarn. II. Veränderlichkeit der Monats- und Jahresmengen. S.-B. Akad. Wiss. Wien.
22. Heidke, P. (1923). Quantitative Begriffsbestimmung homogener Temperatur- und Niederschlagsreihen. *Meteorologische Zeitschrift*, 114–115.
23. Helmert, F.R. (1907). Die Ausgleichrechnung nach der Methode der kleinsten Quadrate. 2. Auflage, Teubner Verlag.
24. Hofund Staatsdruckerei. (1873)/ Bericht über die Verhandlungen des internationalen Meteorologen-Congresses zu Wien, 2–10. September 1873, Protokolle und Beilagen.
25. HOME. Homepage of the COST Action ES06601 — Advance in Homogenization Methods of climate Series: An Integrated Approach (HOME), URL: <http://www.homogenization.org> (33); http://www.homogenization.org/HOMER_R/pdf (36).
26. Hungarian Meteorological Service (HMS). (1996). Proceedings of the First Seminar for Homogenization of Surface Climatological Data, Budapest, Hungary, 6–12 October 1996, 44 p.
27. Joelsson, M., Slättberg, N., Carnebring, A., Sturm, C., & Engström, E. (2020). Automation of the interactive mode of the homogenisation software HOMER for climatological applications.
28. Kohler, M.A. (1949). Double-mass analysis for testing the consistency of records and for making adjustments. *Bull. Amer. Meteorol. Soc.* **30**. 188–189.
29. Kreil, K. (1854a). Mehrjährige Beobachtungen in Wien vom Jahre 1775 bis 1850. Jahrbücher der k.k. Central-Anstalt für Meteorologie und Erdmagnetismus. I. Band — Jg 1848 und 1849. 35–74.
30. Mestre, O. (1999). Step-by-step procedures for choosing a model with change-points. In Proceedings of the second seminar for homogenisation of surface climatological data, Budapest, Hungary, WCDMP. No. 41, WMO-TD. No. 962, 15–26.
31. Mestre, O., Aguilar, E. (2011). HOME_R. Fast documentation. HOMER Training School. 36 p. http://www.homogenization.org/HOME_R.pdf
32. Mestre, O., Domonkos P., Pikard F., Auer I., Robin S., Lebarbier, E. (2013). a.o.Homer: a homogenization software — methods and applications. *Idojaras, Quart. J. Hungarion. Meteorol. Ser.* **117**. Iss. 1. 47–67.
33. Luhunga, P., Mutayoba, E., Ng'ongolo, H. (2014). Homogeneity of Monthly Mean Air Temperature of the United Republic of Tanzania with HOMER. *Atmospheric and Climate Sciences*. **4**. 1. 70–77. doi: 10.4236/acs.2014.41010.
34. Pérez-Zanón, N., Sigró, J., Domonkos, P., and Ashcroft, L. (2015). Comparison of HOMER and ACMANT homogenization methods using a central Pyrenees temperature dataset. *Adv. Sci. Res.* **12**. 111–119, <https://doi.org/10.5194/asr-12-111-2015>.
35. Peterson, T.C., Easterling D.R., Karl T.R., Groisman P., Nicholls N., Plummer N., Torok S., Auer I., Boehm R., Gullett D., Vincent L., Heino R., Tuomenvirta H., Mestre O., Szentimrey T., Salinger J., Forland E.J., Hanssen-Bauer I., Alexandersson H., Jones P., Parker D. (1998). Homogeneity adjustments of in situ atmospheric climate data: A review. *Int. J. Climatol.* **18**. 1493–1517.
36. Picard, F., Lebarbier, E., Hoebeke, M., Rigai, G., Thiam, B., Robin, S. (2011). Joint segmentation, calling, and normalization of multiple CGH profiles. *Biostatistics*. Jul. **12** (3). 413–28. doi: 10.1093/biostatistics/kxq076
37. Szentimrey, T. (1999). Multiple Analysis of Series for Homogenization (MASH). Proceedings of the second seminar for homogenization of surface climatological data, Budapest, Hungary; WMO, WCDMP. **41**. 27–46.
38. Venema, V., Mestre, O., Aguilar, E., Auer, I., Guijarro, J.A., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., Zahradnick, P., Viarre, J., Müller-Westermeier, G., Lakatos, M., Williams, C.N., Menne, M.J., Lindau, R., Rasol, D., Rustemeier, E., Kolokythas, K., Marinova, T., Andresen, L., Acquaotta, F., Fratianni, S., Cheval, S., Klancar, M., Brunetti, M., Gruber, Ch., Prohom Duran, M., Likso, T., Esteban, P., Brandsma, Th. (2012).

- Benchmarking homogenization algorithms for monthly data. *Climate of the Past*. **8**, 89–115, doi: 10.5194/cp-8-89-2012, See also the introductory blog post and a post on the weaknesses of the study.
39. Wang, Xiaolan, Feng, Y. (2013). RH tests V4 user manual. Environment Canada Science and Technology Branch Atmospheric Science and Technology Directorate Climate Research. Division Res. Rep.
40. WMO. Calculation of monthly and annual 30-years standard normals. World Meteorological Organization, World Climate Data and Monitoring Programme Series, 1989. WCDP. 10. WMO-TD.
41. Zentralanstalt für Meteorologie und Geodynamik. (1906). Bericht über die internationale meteorologische Direktorenkonferenz in Innsbruck, September 1905. Anhang zum Jahrbuch 1905. k.k. Hof-und Staatsdruckerei.
42. Zhang, L., Si, J., Jiapaer, G., Zhang, T., Mao, W., Dong, S. (2022). Spatial Homogenization Adjustment and Application of Weather Station Networks in Xinjiang, China. *Atmosphere*. **13 (11)**, 1840. <https://doi.org/10.3390/atmos13111840>.

Kyreieva Zoryna¹

ORCID: 0009-0003-9544-6944

kireeva5687@gmail.com

Skrynyk Olesya^{2,1}

ORCID: 0000-0003-0332-5073

skrynyk@nubip.edu.ua

Palamarchuk Lyudmyla¹

ORCID: 0000-0001-9906-8870

palamarchuk.l@ukr.net

¹ Ukrainian Hydrometeorological Institute of State Emergency Service of Ukraine and National Academy of Sciences of Ukraine

² National University of Life and Environmental Sciences of Ukraine

HOMOGENIZATION OF PRECIPITATION TIME SERIES: MAIN MODERN APPROACHES AND SOFTWARE TOOLS

In our work, we analyzed modern software products that are frequently used for quality control and homogenization of long climatological time series of different variables, in

particular, atmospheric precipitation. Based on our analysis, the HOMER software was selected to perform quality control and homogenization procedures for time series of monthly precipitation sums. The advantages of using HOMER include the possibility for the researcher to make a decision on confirming the breakpoints at the stage of joint comparison (joint detection), which allows considering the available information about the natural variability of atmospheric precipitation in the region and preventing oversmoothing of their values. The HOMER homogenization procedure was applied to monthly atmospheric precipitation time series collected in Ukraine for the period 1946–2020. Possible areas of inhomogeneity of the series (break points) were established and it was determined that the value of the shift amplitude is 0.1–0.7 (in relative units of standard deviation). The correction of the time series of the monthly precipitation sums for the studied period was carried out and it was calculated that their change depends on the magnitude and sign of the shift amplitude and is 10–20% of the initial amounts of precipitation. The considered methodological techniques and approaches to data quality control and their homogenization using the HOMER software can be applied in further precipitation studies.

Keywords: homogenization, software HOMER, monthly atmospheric precipitation sums, data quality, break points.

