

В.П. Сіденко¹

ORCID: 0000-0002-4143-2913
vladyslavsidenko@gmail.com

І.І. Кравченко^{1,2}

ORCID: 0009-0006-4653-1853
igorkravchenko578@gmail.com

З.М. Кирєєва¹

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

Д.Б. Пінчук^{1,3}

ORCID: 0000-0002-2054-3761
leodig44@gmail.com

УДК 551.583

ПОРЯТУНОК ДАНИХ (DATA RESCUE) ТА КОНТРОЛЬ ЯКОСТІ ДОБОВИХ ЧАСОВИХ РЯДІВ ТЕМПЕРАТУРИ ПОВІТРЯ (СЕРЕДНЬОЇ, МАКСИМАЛЬНОЇ ТА МІНІМАЛЬНОЇ) ТА АТМОСФЕРНИХ ОПАДІВ УКРАЇНИ

У роботі представлено результати роботи щодо порятунку даних (Data Rescue) паперових носіїв (метеорологічних таблиць), які містять записи добових значень середньої (ТМ), максимальної (ТХ) та мінімальної (ТН) приземних температур повітря та сум атмосферних опадів (RR). Виконано оцифрування добових значень ТМ, ТХ та ТН отриманих на 176 метеорологічних станціях національної гідрометеорологічної моніторингової мережі. Найбільшу кількість станцій (178) опрацьовано при оцифруванні даних атмосферних опадів. Загальна кількість оцифрованих значень становить 3 571 778. Оцифровані значення заповнюють пропуски, які були у цифровій базі даних добових значень основних кліматичних параметрів (ТМ, ТХ, ТН та RR), сформованій в Українському гідрометеорологічному інституті. За допомогою сучасного, добре апробованого спеціалізованого програмного забезпечення INQC та Climatol, проведено контроль якості цифрової бази даних. Кількість виявлених грубих помилок становить 3 933 (від 9 до 2015 залежно від метеорологічного параметра, проте — не більше 0,04%). Більше величин належать до категорії можливих помилок, викидів, підозрілих значень та груп підозрілих значень. Їх частка від усіх значень окремих величин дорівнює 0,1–0,14%. За результатами проведеної процедури контролю якості було здійснено перевірку та корекцію всіх виявлених грубих помилок відповідно до даних в оригінальних паперових носіях.

Ключові слова: добові часові ряди, температура повітря, екстремальні температури, атмосферні опади, data rescue, контроль якості, INQC, Climatol.

ВСТУП

Висока актуальність досліджень змін клімату зумовлює необхідність створення високоякісних баз даних первинної метеорологічної інформації. Такі бази даних можуть бути покладені в основу великої кількості робіт, так чи інакше пов'язаних із кліматом та його змінами. Часто, однією з перепон до використання в кліматологічних та метеорологічних дослідженнях баз даних первинної метеорологічної інформації є кліматологічна неоднорідність цих рядів [27], яку усувають за допомогою процедури гомогенізації. Таку процедуру для довгих рядів метеорологічних/кліматологічних даних виконують як на рівні середніх за місяць значень [15, 31], так і для даних добового часового масштабу [13, 14, 16, 20–22]. Своєю чергою, головним лімітуючим/обмежуючим фактором при виконанні гомогенізації може бути велика кількість пропущених даних та загальна низька якість рядів досліджуваних станцій (наявність викидів чи інших помилкових значень).

Результати процедур контролю якості та гомогенізації рядів метеорологічних величин представлено в численних дослідженнях, як на локальному чи регіональному, так і на національному, міжнаціо-

нальному або ж глобальному рівнях [15, 21, 22, 25, 30 та ін.]. В Україні низку досліджень кліматологічної неоднорідності первинної метеорологічної інформації та її подальшого аналізу на місячному часовому масштабі було проведено в Українському гідрометеорологічному інституті [1–3, 23, 24, 27, 28]. Далі слід виконати дослідження неоднорідностей у кліматологічних рядах з більш високою часовою роздільністю, зокрема на добовому масштабі. Такий підхід дозволить провести комплексний різномасштабний порівняльний аналіз неоднорідностей в рядах спостережень. Також гомогенізовані довгі ряди добових значень метеорологічних спостережень будуть використані як основа для проведення спеціалізованої геостатистичної інтерполяції (геопросторового моделювання) та створення сіткових баз даних для дослідження зміни клімату, включаючи кліматологічні аспекти екстремальних погодних подій та явищ. Варто зазначити, що часова роздільність вихідних даних для кліматологічного дослідження екстремальних погодних явищ повинна бути не менша, ніж добова. Це зумовлено тим, що вони мають коротку тривалість (від кількох годин до кількох днів) [4]. Для того, щоб заповнити про-

печення — якщо в автоматизованому. Якість оцифрування також сильно залежить від стану, в якому знаходяться паперові носії, та від того, чи це оригінальні записи, чи переписані заново, оскільки при переписуванні також можуть припуститися помилки. Процедури створення цифрового фотоархіву та оцифрування даних було проведено відповідно до рекомендацій ВМО з порятунку даних [17].

Контроль якості. Часові ряди даних метеорологічних спостережень та вимірювань практично завжди містять в собі помилки різної величини та походження. Саме тому до них необхідно застосувати процедуру контролю якості. Метою такої процедури є виявлення та вилучення помилок, які в подальшому аналізі можуть призвести до отримання хибних чи необґрунтованих висновків щодо стану кліматичної системи, змін клімату та його мінливості. Контроль якості було виконано у декілька етапів.

Перший етап — “ручний”, був проведений за допомогою вбудованих засобів табличного процесора Excel. Він є важливим, оскільки дозволяє очистити ряди від різного роду друкарських помилок, які могли виникнути під час диджиталізації даних. Крім того, він дозволяє очистити часові ряди від нечислових значень, що є важливим для коректної роботи програмного забезпечення, яке буде застосоване на подальших стадіях процедури контролю якості. Слід зазначити, що на цьому етапі перевірка рядів здійснювалась для періоду, в якому було виконано процес ручного оцифрування даних з паперових носіїв. Натомість головні етапи контролю якості для всього періоду покриття бази даних (1946–2020 рр.) було виконано за допомогою спеціалізованих кліматологічних програмних засобів INQC [5] та Climatol [18]. Ці програмні продукти часом використовують і вони є добре апробованими.

Наступний (основний) етап контролю якості — аналіз часових рядів за допомогою INQC (Indices Quality Control of Climatological Daily Time Series) [5]. INQC — це програмне забезпечення (ПЗ), що виконується в середовищі R. INQC можна використовувати для контролю якості кліматологічних даних з добовою роздільною здатністю. Воно являє собою набір функцій/тестів для виявлення помилкових чи підозрілих значень у часових рядах основних метеорологічних величин (середньої, максимальної та мінімальної приземних температур повітря, атмосферних опадів та тиску, відносної вологості, швидкості вітру та ін.). INQC працює, застосовуючи серію тестів до даних, виявляючи помилкові значення (наприклад, від’ємні опади або температури вище/нижче певних встановлених меж), підозрілі значення (наприклад, екстремальні значення, які

“випадають” із розподілу ймовірностей значень метеорологічної величини, що аналізується) та сукупно підозрілі значення (які повторюються надто багато разів за окремі короткі періоди). Короткий опис застосованих INQC-тестів представлено в [29]. Під час проведення контролю якості за допомогою ПЗ INQC кожному значенню в часових рядах даних присвоюється додаткове значення на основі системи цілочисельних міток контролю якості, які надають інформацію користувачу про статус перевічених даних для подальшого аналізу і прийняття рішень щодо можливого виправлення чи вилучення помилкового значення. INQC має 6 міток контролю якості для позначення категорій, до яких потрапляє значення ряду, а саме: значення, що пройшли контроль якості; грубі помилки; ймовірні помилки; викиди та підозрілі значення; група підозрілих значень (колективно підозрілі значення) та пропущені значення ряду. На цьому етапі контроль якості здійснювався для усього часового періоду дослідження, а саме для 1946–2020 років для 178 станцій.

Як додатковий етап контролю якості було використано Climatol (Climate Tools: Series Homogenization and Derived Products) [18]. Climatol — це програмне забезпечення без графічного інтерфейсу, яке теж працює в середовищі R і яке створене для виконання процедур контролю якості, заповнення пропущених даних та гомогенізації рядів широкого набору метеорологічних величин з місячною та з добовою роздільною здатністю. За допомогою Climatol були проаналізовані дані про викиди як на рівні осереднених за місяць значень, так і на рівні добових значень.

Важливо зазначити, що різниця між INQC та Climatol при проведенні контролю якості полягає в тому, що перше ПЗ реалізує метод “абсолютного” контролю якості й використовує для виявлення помилкових та підозрілих значень лише статистичні характеристики досліджуваного ряду. У випадку Climatol реалізується процедура “відносного” контролю якості, оскільки для виявлення викидів використовуються дані рядів сусідніх станцій.

Для того, щоб перевірити наявності інших, не врахованих в INQC та Climatol, типових помилок, було вирішено провести додатковий етап контролю якості для виявлення випадків, коли для певної станції порушується нерівність $TN \leq TM \leq TX$. Перевірку виконано на основі створених власних скриптів у середовищі R.

Враховуючи велику кількість даних та необхідність проведення декількох ітерацій/циклів контролю якості часових рядів, процес проведення розрахунків необхідно максимально автоматизувати. З цієї метою було створено скрипти та програмні

коди, які дозволяють значно мінімізувати участь користувача у всьому ланцюжку основних етапів аналізу емпіричних даних.

РЕЗУЛЬТАТИ ТА ЇХ ОБГОВОРЕННЯ

Порятунок даних. В результаті проведення детальної інвентаризації наявної цифрової метеорологічної/кліматологічної інформації добового часового масштабу було визначено повноту рядів окремо для кожного року для 178 станцій мережі метеорологічних спостережень України у наявних цифрових наборах рядів добових значень середньої, максимальної та мінімальної приземних температур повітря та добових сум опадів. Для розуміння суті такої оцінки потрібно визначити поняття повноти ряду. В цьому випадку, повнотою ряду будемо вважати відношення фактичної кількості членів ряду (без пропущених значень) до їх загальної (максимально можливої) кількості у досліджуваному періоді, виражене у відсотках. Було встановлено, що в межах часового діапазону 1946–2020 рр. найбільша кількість пропущених даних зосереджена в 1946–1960 рр. (15 років), де повнота становить 20–40%. Часовий період 1976–2020 років в середньому має задовільне та високе значення повноти рядів (понад 80–90%).

На основі створеного архіву відсканованих паперових носіїв було проведено оцифрування по 176 рядах добових значень середньої, максимальної та мінімальної температури, а також 178 рядах добових значень сум опадів для різних часових періодів.

Процедура порятунку даних (оцифрування) дозволила доповнити наявну цифрову базу даних 3571 778 добовими значеннями кліматологічних величин. Серед них, загальна кількість оцифрованих значень середньої добової температури повітря становить 640 702; кількість оцифрованих значень максимальної та мінімальної добових температур повітря складає по 652 963 та 644 843 відповідно; для рядів добових сум атмосферних опадів оциф-

ровано 1 633 270 значення. Найбільше роботи з оцифрування рядів середньої температури було виконано для періоду 1946–1960 років, тоді як для років періоду з 1961 по 1991 було виконано оцифрування невеликих фрагментів пропущених даних з метою підвищення повноти рядів. Для рядів максимальної та мінімальної температур повітря більшість даних оцифровані для періоду 1946–1959 років, оцифровано окремі пропущені періоди даних з 1960–1991 роки. Для рядів значень добових сум опадів більшість даних була оцифрована для періоду 1946–1975 років. Як і для рядів температури, для рядів опадів було оцифровано певну кількість пропущених даних для періоду 1976–1993 років. Теплові карти повноти добових часових рядів до та після оцифрування показано на *рис. 2*. Як видно з теплових карт, не всі пропущені дані були оцифровані. Однак, кількість пропущених даних суттєво зменшилась у порівнянні з початковим станом бази даних.

Контроль якості. Первинний аналіз оцифрованих даних показав, що типовими помилками під час ручного оцифрування є пропуск знаку мінус перед числом, пропуск роздільника цілих та десяткових значень (символу крапки «.» або коми «,»), що призводить до внесення значень у десять разів вищих ніж насправді. При оцифруванні рукописних таблиць рядів низької якості можна внести хибне значення, переплутавши подібні один до одного варіанти написання різних цифр та літер. Іноді трапляється так, що ряди різних станцій зі схожими назвами можуть бути помилково об'єднані в один. До того ж дані однієї станції можуть бути помилково розділені у різні ряди. Не варто забувати й про те, що і в метаданих про станції, їх перенесення та будь-які інші зміни теж можуть міститися помилки. Перевірка за допомогою вбудованих засобів Excel виявила 56 грубих помилок. Їх найбільшу кількість було знайдено в рядах максимальної та середньої температур повітря, дещо меншу кількість в рядах мінімальної температури повітря та добової суми опадів (*табл. 1*).

Таблиця 1. Розподіл виявлених помилок за метрологічними величинами під час проведення попереднього етапу контролю якості часових рядів

Величина	Кількість помилок	Відсоток, %
Середня добова температура повітря, ТМ	18	32
Максимальна добова температура повітря, ТХ	26	46
Мінімальна добова температура повітря, ТН	7	13
Добова сума опадів, RR	5	9
Загалом	56	100

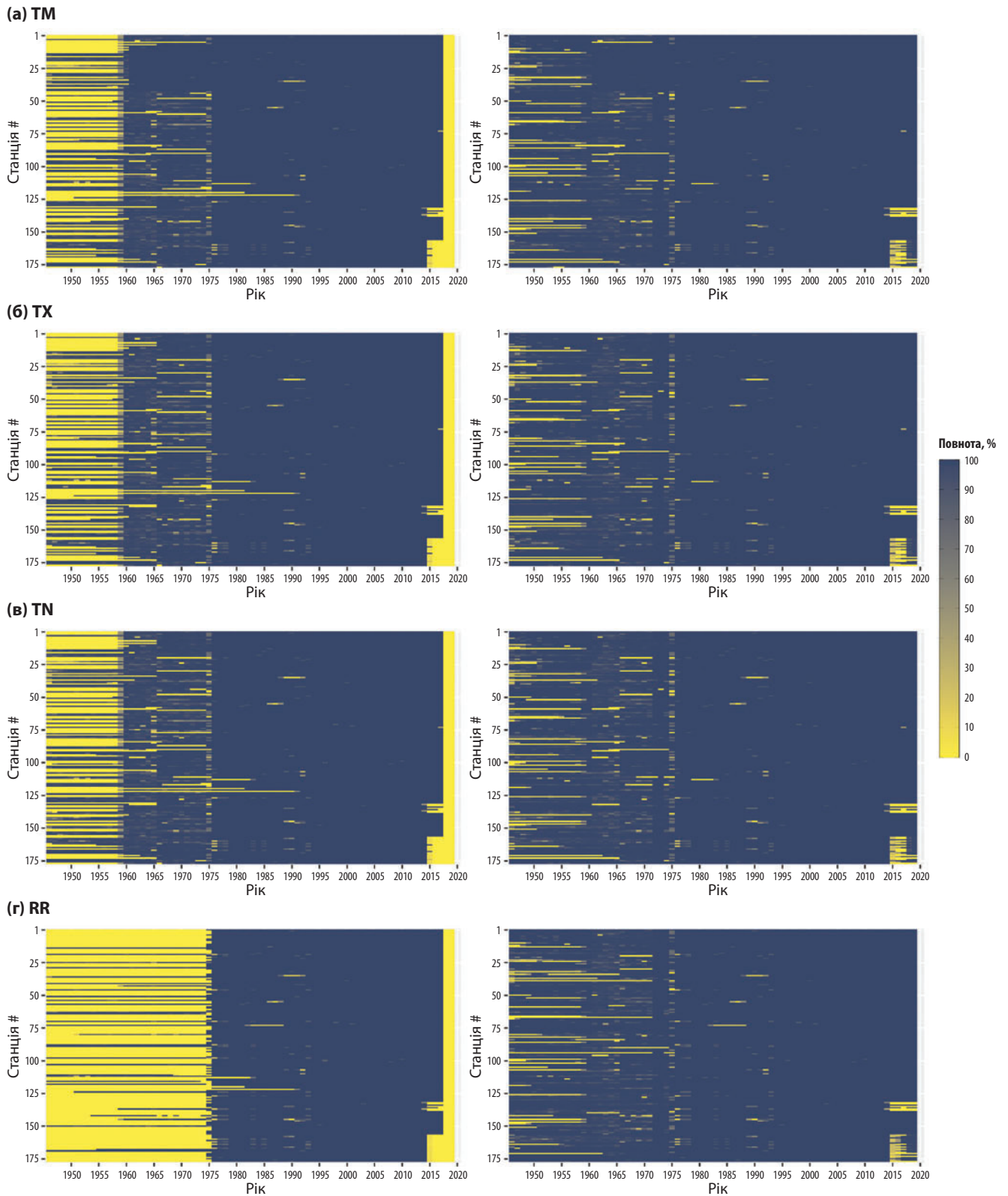


Рис. 2. Теплові карти повноти добових часових рядів до (зліва) та після (справа) оцифрування даних: а — ТМ; б — ТХ; в — ТН; г — РР

Серед виявлених на попередньому етапі помилок найчастіше зустрічається пропуск символу коми (розділового десяткового знаку) (55,4%) або

ж навпаки, подвійний ввід символу (25%). Решту 19,6% склали помилки, що стосувалися набору при оцифруванні зайвих цифр, символів або літер за-

мість цифр, неправильного порядку цифр тощо. виправлення усіх виявлених випадків було здійснено відповідно до даних у фотокопіях оригінальних паперових носіїв.

За результатами головного етапу контролю якості (за допомогою INQC), в усіх 4 наборах рядів досліджуваних метеорологічних величин (ТМ, ТХ, ТН, RR) кількість значень, які пройшли контроль якості становить понад 94% від усієї сукупності значень для досліджуваного часового періоду відповідних величин. Кількість виявлених помилок в абсолютних значеннях становить 3 933 (від 9 в рядах RR до 2015 в рядах TN), проте — не більше 0,04% від усієї сукупності значень кожної з величин. Така кількість помилок спричинена великою кількістю помилок категорії txtn — по 1543. Більше величин належать до категорії можливих помилок, викидів, підозрілих значень та груп підозрілих значень. Їх частка від всіх значень окремих величин дорівнює 0,1–0,14%. Однак в абсолютних значеннях ця кількість, а саме до 6758 значень, є досить значною. Відсоток про-

пусків сягає приблизно 4,6–5,1% значень у кожному з наборів часових рядів досліджуваних величин. Оскільки під час перевірки даних за допомогою INQC було виявлено досить значну кількість помилок категорії txtn, який перевіряє чи виконується умова $TN \leq TX$, здійснено перевірку за допомогою власних тестів: tntm та tmtx. Ці тести були розраховані для кожної пари відповідних рядів однієї метеорологічної станції за допомогою власних скриптів/програмних кодів. Така перевірка показала, що в рядах є 1331 випадок, коли $TN > TM$ та 1 422 випадки, коли $TM > TX$.

Результатом проведеної процедури контролю якості було виявлення помилок та підозрілих значень температур повітря та атмосферних опадів, які подано в табл. 2.

Під час перевірки викидів (виявлених за допомогою ПЗ Climatol) на рівні осереднених за місяць значень у багатьох випадках були знайдені можливі причини помилок. Наприклад, це примітки про несправність мінімального термометра, закресле-

Таблиця 2. Розподіл випадків виявлених помилок та підозрілих значень після проведення контролю якості за величинами для кожного тесту/функції INQC

Тест	ТМ	ТХ	ТН	RR
drywetlong	×	×	×	1653
duplas	0	0	0	0
flat	426	558	533	2439
friki	1438	1605	1603	3676
frikilight	8137	8143	8230	10909
IQRoutliers	126	291	323	1180
jumpABS	226	694	747	0
jumpQUANT	1739	274	1054	0
paretogadget	×	×	×	6
physics (large)	52	46	72	11
physics (small)	11	11	7	13
repeatedvalue	×	×	×	501
rounding	0	193	162	1079
rounding (roundmax)	0	192	167	384
roundprecip	×	×	×	0
suspectacumprec	×	×	×	63
toomany (annual)	0	0	0	2209
toomany (toomanymonth)	0	81	50	2025
txtn	×	1543	1543	×
weirddate	0	0	0	0
Загалом	12155	13631	14491	26148

Примітки. 1) × — тест не застосовується; 2) короткий опис застосованих тестів наведений у [29].

Таблиця 3. Кількість перевірених, виправлених, підтверджених та вилучених значень після усіх етапів процедури контролю якості

Контроль якості	Величина				Загалом
	ТМ	ТХ	ТН	RR	
Перевірені	6171	6252	6587	6099	25109
в т. ч.:					
виправлені	1447	2246	2620	261	6574
підтверджені	3080	2358	2270	4202	11910
вилучені	1644	1648	1697	1636	6625

ні дані (ймовірно, записані помилково), а також пропуски у вимірах. Ці примітки, переважно, були розміщені на полях оригінальних паперових носіїв даних. Під час перевірки викидів місячних рядів дані були перевірені на рівні добових значень для кожного дня місяця з викидів. У рядах середньої температури було виправлено 175 значень, а підтверджено — 45. В рядах максимальної та мінімальної температури виправлено 91 та 112 значень, а підтверджено 61 та 68 значень відповідно.

Всі знайдені помилки на кожному з етапів контролю якості, а саме грубі помилки, знайдені завдяки INQC, помилки, знайдені за допомогою власних скриптів (тести tntm та tmtx) та викиди, виявлені внаслідок тестових запусків Climatol, були скомпановані разом та перевірені вручну. Слід зазначити, що для дати кожної помилки було перевірено кожну з досліджуваних метеорологічних величин. Такий підхід було застосовано для підвищення точності контролю якості, а також тому, що на паперових носіях величини часто розташовуються на одних листах у сусідніх клітинках.

За результатами такої комплексної процедури контролю якості було виконано перевірку всіх виявлених грубих помилок та викидів. Перевірку було здійснено за даними у фотокопіях паперових носіїв (у випадку, якщо вони були наявні). Всі знайдені помилки було виправлено у відповідних наборах рядів даних. Якщо не було можливості перевірити певні помилки, то такі грубі помилки були вилучені та позначені як пропущені. Результати перевірки представлені у таблиці 3. Окрім цього, якщо при перевірці була підозра, що всі значення певного місяця неправильні, то вони перевірялись додатково. Таким чином було знайдено та виправлено 72 помилкових значення середньої, 63 значення максимальної та 65 значень мінімальної температури.

Окрім цього, завдяки процедурі контролю якості в оригінальних рядах добових сум атмосферних опадів на станції Умань для періоду 1946–1965 років

було виявлено та виправлено велику кількість помилок, не включених до *таблиці 3*. Після перевірки з паперовими носіями, було виявлено пропуск символу десяткового розділювача. Кількість таких виправлених значень склала 7305.

ВИСНОВКИ

У роботі представлено результати інвентаризації даних наявної цифрової метеорологічної/кліматологічної інформації добового часового масштабу, а саме проаналізовано повноту рядів у наявних цифрових наборах рядів добових значень середньої, максимальної та мінімальної температур повітря та добових сум атмосферних опадів. Встановлено, що за межами періоду 1946–2020 років середня для кожного з чотирьох наборів рядів повнота рядів не є достатньою для проведення подальших досліджень, таких, як процедура гомогенізації добових значень.

Виконано оцифрування 176 рядів добових значень для ТМ, ТХ та ТН, а також 178 рядів добових значень RR для різних часових періодів (переважно 1946–1960 та 1946–1975). Загальна кількість оцифрованих значень становить 3 571 778 (з них: 640 702 оцифрованих значень середньої добової температури повітря, 652 963 значень максимальної добової температури повітря, 644 843 значень мінімальної добової температури повітря та 1 633 270 оцифрованих значень для рядів добових сум опадів).

Проведено контроль якості даних у декілька етапів з використанням ПЗ INQC, Climatol та власних скриптів/тестів. На першому етапі виявлено та виправлено 56 грубих помилок. На другому, головному етапі контролю якості, кількість виявлених помилок становить 3 933 (від 9 до 2015 в залежності від набору даних), що є не більше 0,04% від усієї можливої для досліджуваного періоду сукупності значень кожної з величин. Кількість величин, що належать до ймовірних помилок, викидів, підозрілих значень та колективно підозрілих значень є дещо більшою. Відсоток таких значень від максимально

можливої сукупності значень окремих величин сягає 0,1–0,14%. За допомогою власних скриптів виявлено 1331 помилкових значень, коли $TN > TM$ та 1422 випадки, коли $TM > TX$. За результатами проведеної процедури контролю якості було здійснено перевірку всіх виявлених грубих помилок та викидів за даними у фотокопіях паперових носіїв.

Застосування вищезазначених процедур порятунку даних та контролю якості до оригінальних кліматологічних рядів дозволило значно підвищити повноту рядів, а також покращити якість довгих рядів метеорологічних величин та є важливим для проведення подальших досліджень на основі опрацьованих даних.

ЛІТЕРАТУРА / REFERENCES

1. Осадчий В.І., Агуілар Е., Скриник О.А., Бойчук Д.О., Сіденко В.П., Скриник О.Я. Добова асиметрія кліматичних змін температури повітря в Україні. *Український географічний журнал*. 2018. **3**. С. 21–30. DOI: <https://doi.org/10.15407/ugz2018.03.021> [Osadchyi, V.I., Aguilar, E., Skrynyk, O.A., Boichuk, D.O., Sidenko, V.P., Skrynyk, O.Y. (2018). Daily asymmetry of temperature changes in Ukraine. *Ukrainian Geographical Journal*. **3**. 21–30. DOI: <https://doi.org/10.15407/ugz2018.03.021>].
2. Осадчий В.І., Скриник О.А., Сіденко В.П., Бойчук Д.О., Ошурок Д.О., Скриник О.Я. Гомогенізована база даних довгих часових рядів середньої місячної температури. *Геоінформатика*. 2018. **1** (65). С. 46–58. [Osadchyi, V.I., Skrynyk, O.A., Sidenko, V.P., Boichuk, D.O., Oshurok, D.O., Skrynyk, O.Y. (2018). Homogenized database of long monthly mean air temperature time series. *Geoinformatika*. **1** (65). 45–68].
3. Паламарчук Л.В., Осадчий В.І., Скриник О.А., Киреева 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). Application of the HOMER software to quality control and homogenize time series of monthly precipitation sums. *Hydrology, hydrochemistry and hydroecology*. **1** (67). 58–77. DOI: <https://doi.org/10.17721/2306-5680.2023.1.7>].
4. Сіденко В. П. Кліматологічні дослідження екстремальних погодних умов, подій та явищ в Україні та світі. *Гідрологія, гідрохімія та гідроекологія*. 2022. **2** (64). С. 53–71. DOI: <https://doi.org/10.17721/2306-5680.2022.2.5>. [Sidenko, V.P. (2022). Climatic studies of extreme weather conditions, events and phenomena in Ukraine and the world. *Hydrology, hydrochemistry and hydroecology*. **2** (64). DOI: <https://doi.org/10.17721/2306-5680.2022.2.5>].
5. Aguilar, E., Skrynyk, O. (2021). INQC: Quality control of climatological daily time series. R-Packages. URL: <https://CRAN.R-project.org/package=INQC> (accessed 12/01/2024).
6. Ashcroft, L., Coll, J. R., Gilabert, A., Domonkos, P., Brunet, M., Aguilar, E., Castella, M., Sigro, J., Harris, I., Uden, P., Jones, P. (2018). A rescued dataset of sub-daily meteorological observations for Europe and the southern Mediterranean region, 1877–2012. *Earth System Science Data*. **10** (3). 1613–1635. DOI: <https://doi.org/10.5194/ESSD-10-1613-2018>.
7. Ashcroft, L., Gergis, J., Karoly, D.J. (2014). A historical climate dataset for southeastern Australia, 1788–1859. *Geoscience Data Journal*. **1** (2). 158–178. DOI: <https://doi.org/10.1002/GDJ3.19>.
8. Brönnimann, S., Annis, J., Dann, W., Ewen, T., Grant, A.N., Griesser, T., Krähenmann, S., Mohr, C., Scherer, M., Vogler, C. (2006). A guide for digitising manuscript climate data. *Climate of the Past*. **2** (2). 137–144. DOI: <https://doi.org/10.5194/CP-2-137-2006>.
9. Brönnimann, S., Allan, R., Ashcroft, L., Baer, S., Barriendos, M., Brázdil, R., Brugnara, Y., Brunet, M., Brunetti, M., Chimani, B., Cornes, R., Domínguez-Castro, F., Filipiak, J., Founda, D., Herrera, R.G., Gergis, J., Grab, S., Hannak, L., Huhtamaa, H., Jacobsen, K.S., Jones, P., Jourdain, S., Kiss, A., Lin, K.E., Lorrey, A., Lundstad, E., Luterbacher, J., Mauelshagen, F., Maugeri, M., Maughan, N., Moberg, A., Neukom, R., Nicholson, S., Noone, S., Nordli, Ø., Ólafsdóttir, K.B., Pearce, P.R., Pfister, L., Pribyl, K., Przybylak, R., Pudmenzky, C., Rasol, D., Reichenbach, D., Řezníčková, L., Rodrigo, F.S., Rohr, C., Skrynyk, O., Slonosky, V., Thorne, P., Valente, M.A., Vaquero, J.M., Westcott, N.E., Williamson, F. (2019). Wyszyński Przemys. Unlocking pre-1850 instrumental meteorological records a global inventory. *Bulletin of the American Meteorological Society*. **100** (12). E5389–E5413. DOI: <https://doi.org/10.1175/BAMS-D-19-0040.1>.
10. Brönnimann, S., Brugnara, Y., Allan, R.J., Brunet, M., Compo, G.P., Crouthamel, R.I., Jones, P.D., Jourdain, S., Luterbacher, J., Siegmund, P., Valente, M.A., Wilkinson, C.W. (2018). A roadmap to climate data rescue services. *Geoscience Data Journal*. **5** (1). 28–39. DOI: <https://doi.org/10.1002/GDJ3.56>.
11. Brunet, M., Gilabert, A., Jones, P., Efthymiadis, D. (2014). A historical surface climate dataset from station observations in Mediterranean North Africa and Middle East areas. *Geoscience Data Journal*. **1** (2). 121–128. DOI: <https://doi.org/10.1002/GDJ3.12>.
12. Brunet, M., Jones, P. (2011). Data rescue initiatives: bringing historical climate data into the 21st century. *Climate Research*. **47** (1–2). 29–40. DOI: <https://doi.org/10.3354/CR00960>.
13. Camuffo, D., Valle, A. della, Becherini, F., Zanini, V. (2020). Three centuries of daily precipitation in Padua, Italy, 1713–2018: history, relocations, gaps, homogeneity and raw data. *Climatic Change*. **162** (2). 923–942. DOI: <https://doi.org/10.1007/S10584-020-02717-2>.
14. Costa, A.C., Soares, A. (2008). Homogenization of climate data: review and new perspectives using geostatistics. *Mathematical Geosciences*. **41** (3). 291–305. DOI: <https://doi.org/10.1007/s11004-008-9203-3>.
15. Delvaux, C., Ingels, R., Vrábel, V., Journée, M., Bertrand, C. (2019). Quality control and homogenization of the Belgian historical temperature data. *International Journal of Climatology*. **39** (1). 157–171. DOI: <https://doi.org/10.1002/JOC.5792>.
16. Frick, C., Steiner, H., Mazurkiewicz, A., Riediger, U., Rauthe, M., Reich, T., Gratzki, A. (2014). Central European high-resolution gridded daily data sets (hyras): mean temperature and relative humidity. *Meteorologische Zeitschrift*. **23** (1). 15–32. DOI: <https://doi.org/10.1127/0941-2948/2014/0560>.
17. Guidelines on best practices for climate data rescue. 2016 edition. Geneva, Switzerland: World Meteorological Organization, 2016. ISBN 978-92-63-11182-1.
18. Guijarro, J.A. (2019). Climatol: Climate tools (series homogenization and derived products). R-Packages. URL: <https://CRAN.R-project.org/package=climatol> (accessed 12/01/2024).

19. Hawkins, E., Burt, S., Brohan, P., Lockwood, M., Richardson, H., Roy, M., Thomas, S. (2019). Hourly weather observations from the Scottish Highlands (1883–1904) rescued by volunteer citizen scientists. *Geoscience Data Journal*. **6** (2). 160–173. DOI: <https://doi.org/10.1002/GDJ3.79>.
20. Li, Z., Cao, L., Zhu, Y., Yan, Z. (2016). Comparison of two homogenized datasets of daily maximum/mean/minimum temperature in China during 1960–2013. *Journal of Meteorological Research*. **30** (1). 53–66. DOI: <https://doi.org/10.1007/s13351-016-5054-x>.
21. Mahmood, R., Jia, S. (2016). Quality control and homogenization of daily meteorological data in the trans-boundary region of the Jhelum River basin. *Journal of Geographical Sciences*. **26** (12). 1661–1674. DOI: <https://doi.org/10.1007/s11442-016-1351-7>.
22. Mateus, C., Potito, A. (2021). Development of a quality-controlled and homogenised long-term daily maximum and minimum air temperature network dataset for Ireland. *Climate*. **9** (11). 158. DOI: <https://doi.org/10.3390/cli9110158>.
23. Osadchyi, V., Skrynyk, O., Palamarchuk, L., Skrynyk, O., Osypov, V., Oshurok, D., Sidenko, V. (2022). Dataset of gridded time series of monthly air temperature (min, max, mean) and atmospheric precipitation for Ukraine covering the period of 1946–2020. Data in Brief. **44**. 108553. DOI: <https://doi.org/10.1016/j.dib.2022.108553>.
24. Osadchyi, V., Skrynyk, O., Radchenko, R., Skrynyk, O. (2017). Homogenization of Ukrainian air temperature data. *International Journal of Climatology*. **38** (1). 497–505. DOI: <https://doi.org/10.1002/joc.5191>.
25. Randriamarolaza, L.Y.A., Aguilar, E., Skrynyk, O., Vicente-Serrano, S.M., Domínguez-Castro, F. (2021). Indices for daily temperature and precipitation in Madagascar, based on quality-controlled and homogenized data, 1950–2018. *International Journal of Climatology*. **42** (1). DOI: <https://doi.org/10.1002/joc.7243>.
26. Rodrigo, F.S. (2020). Recovering climate data from documentary sources: A study on the climate in the south of Spain from 1792 to 1808. *Atmosphere*. **11** (3). DOI: <https://doi.org/10.3390/ATMOS11030296>.
27. Skrynyk, O., Aguilar, E., Skrynyk, O., Sidenko, V., Boichuk, D., Osadchyi, V. (2019). Quality control and homogenization of monthly extreme air temperature of Ukraine. *International Journal of Climatology*. **39** (4). 2071–2079. DOI: <https://doi.org/10.1002/JOC.5934>.
28. Skrynyk, O., Luterbacher, J., Allan, R., Boichuk, D., Sidenko, V., Skrynyk, O., Palarz, A., Oshurok, D., Xoplaki, E., Osadchyi, V. (2020). Ukrainian early (pre-1850) historical weather observations. *Geoscience Data Journal*. **8** (1). P. 55–73. DOI: <https://doi.org/10.1002/gdj3.108>.
29. Skrynyk, O., Sidenko, V., Aguilar, E., Guijarro, J., Skrynyk, O., Palamarchuk, L., Oshurok, D., Osypov, V., Osadchyi, V. (2023). Data quality control and homogenization of daily precipitation and air temperature (mean, max and min) time series of Ukraine. *International Journal of Climatology*. C. joc.8080. DOI: <https://doi.org/10.1002/joc.8080>.
30. Štěpánek, P., Zahradníček, P., Skalák, P. (2009). Data quality control and homogenization of air temperature and precipitation series in the area of the Czech Republic in the period 1961–2007. *Advances in Science and Research*. **3** (1). 23–26. DOI: <https://doi.org/10.5194/asr-3-23-2009>.
31. Venema, V.K.C., Mestre, O., Aguilar, E., Auer, I., Guijarro, J.A., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., Zahradnicek, 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., Acquafina, F., Fratiani, S., Cheval, S., Klancar, M., Brunetti, M., Gruber, C., Prohom Duran, M., Likso, T., Esteban, P., Brandsma, T. (2012). Benchmarking homogenization algorithms for monthly data. *Climate of the Past*. **8** (1). 89–115. DOI: <https://doi.org/10.5194/cp-8-89-2012>.

Sidenko Vladyslav¹; ORCID: 0000–0002–4143–2913;
*vladyslavsidenko@gmail.com

Kravchenko Ihor^{1,2}; ORCID: : 0009–0006–4653–1853
igorkravchenko578@gmail.com

Kyreieva Zoryna¹; ORCID: 0009–0003–9544–6944;
kireeva5687@gmail.com

Pinchuk Dmytro^{1,3}; ORCID: 0000–0002–2054–3761
leodig44@gmail.com

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

² National University of Life and Environmental Sciences
of Ukraine, Kyiv

³ Taras Shevchenko National University of Kyiv, Kyiv

DATA RESCUE AND QUALITY CONTROL OF DAILY TIME SERIES OF AIR TEMPERATURE (MEAN, MAXIMUM AND MINIMUM) AND ATMOSPHERIC PRECIPITATION IN UKRAINE

This paper presents the results of the digitization of hard copies (meteorological tables) containing records of daily values

of mean (TM), maximum (TX) and minimum (TN) surface air temperatures and atmospheric precipitation sums (RR). The daily values of TM, TX and TN obtained at 176 meteorological stations of the national hydrometeorological monitoring network were digitized. The largest number of stations (178) were processed for digitizing atmospheric precipitation data. The total number of digitized values is 3,571,778. The digitized values fill in the gaps in the digital database of daily values of the essential climatic variables (TM, TX, TN and RR), which was created at the Ukrainian Hydrometeorological Institute. The quality control of the digital database was carried out using state-of-the-art, well-tested dedicated software INQC and Climatol. The number of detected gross errors is 3,933 and ranges from 9 to 2015, depending on the meteorological parameter (however, not more than 0.04% of the total set of values of each variable). A slightly larger number of values were recorded that fell into the category of probable errors, outliers, suspicious values, and collectively suspicious values. The percentage of such values from the total amount of values for each dataset reaches up to 0.14%. Based on the results of the quality control procedure, all identified errors were checked and corrected in correspondence with the data in the original hard copies.

Keywords: daily time series, air temperature, extreme temperatures, atmospheric precipitation, data rescue, quality control, INQC, Climatol.