

Підкатегорія. Аналіз кліматичних даних у RStudio

Дмитро Дядін, ХНУМГ ім. О. М. Бекетова

Про ресурс

Цей ресурс є частиною розділу «Кліматичні дані та інструменти» науково-освітньої платформи ClimEd. Він призначений для студентів, аспірантів, викладачів, дослідників і фахівців кліматозалежних секторів, які працюють із кліматичними даними, просторовими наборами, часовими рядами, реаналізом, кліматичними сценаріями та проєкціями.

Матеріал пояснює, як використовувати R і RStudio для автоматизованого завантаження, попереднього оброблення, просторового аналізу, агрегації, розрахунку кліматичних індикаторів, порівняння періодів і сценаріїв, підготовки таблиць, карт і графіків. Ресурс може застосовуватися під час реалізації курсів ClimEd, виконання практичних і лабораторних робіт, підготовки кваліфікаційних та наукових досліджень, оцінювання кліматичних ризиків і розроблення прикладних кліматичних сервісів.

Підкатегорії для підкатегорії «Аналіз кліматичних даних у RStudio»

1. Чому R і RStudio доцільно використовувати для кліматичних даних
2. Типова послідовність автоматизованого оброблення
3. Основні пакети R для кліматичних і просторових даних
4. Робота з часовими рядами та кліматичними індикаторами
5. Порівняння кліматичних періодів, моделей і сценаріїв
6. Приклад структури R-скрипта
7. Поеднання кліматичних даних із територіальними межами
8. Розрахунок інтегральних показників кліматичного ризику
9. Порівняння ручної роботи в ГІС та автоматизації в R
10. Практичні рекомендації та належна практика

1. Чому R і RStudio доцільно використовувати для кліматичних даних

R / RStudio є зручним середовищем для автоматизованої роботи з кліматичними даними, особливо тоді, коли потрібно обробляти великі масиви файлів NetCDF (.nc) без їх попереднього відкриття у QGIS або ArcGIS. На відміну від ручної роботи в ГІС, R дозволяє описати всю послідовність оброблення у вигляді скрипта: завантаження файлів, вибір змінної, обрізання за територією, агрегацію за часом, розрахунок індикаторів, збереження результатів і побудову графіків. Це суттєво прискорює аналіз, робить його відтворюваним і зменшує ризик помилок під час роботи з десятками або сотнями кліматичних шарів.

Термінологічна примітка

R - це мова програмування і середовище статистичних обчислень. RStudio - інтегроване середовище розроблення, яке полегшує написання, запуск і документування R-скриптів. У практичній роботі вони зазвичай використовуються разом.

2. Типова послідовність автоматизованого оброблення

Типовий робочий процес охоплює підготовку вихідних файлів, підключення необхідних пакетів, відкриття даних, вибір кліматичної змінної, визначення часової та просторової області, агрегацію, розрахунок показників, перевірку результатів і експорт.

1. Підключити пакети, потрібні для роботи з растровими, векторними та табличними даними.
2. Відкрити файл або серію файлів NetCDF, GeoTIFF, CSV чи іншого підтримуваного формату.
3. Обрати потрібну змінну, одиниці вимірювання, часовий діапазон і територію дослідження.
4. Перевірити систему координат, часовий вимір, пропущені значення та метадані.
5. Виконати просторове обрізання, маскування, перепроєктування або зміну роздільної здатності.
6. Агрегувати дані за місяцями, сезонами, роками чи іншими періодами.
7. Розрахувати кліматичні індикатори, статистики, аномалії або зміни.
8. Експортувати результати та задокументувати параметри аналізу.



Рисунок 1 – Процес оброблення кліматичних даних у R

3. Основні пакети R для кліматичних і просторових даних

Для роботи з кліматичними растрами в R найчастіше використовують пакети terra, stars, raster, ncdf4, sf, exactextractr, tidyverse та lubridate. Вони дозволяють напямку відкривати файли NetCDF, обирати потрібну кліматичну змінну, працювати з часовими вимірами, виконувати просторове обрізання за межами області, громади чи річкового басейну, перепроєктовувати дані, змінювати просторову роздільну здатність, маскувати значення за полігоном і експортувати результати у формати

GeoTIFF, NetCDF, CSV або Excel. Тобто R може виконувати ті самі базові операції, що й настільна ГІС, але у пакетному режимі та для великої кількості файлів одночасно.

Пакет	Основне призначення	Типові завдання
terra	Сучасна робота з растровими та векторними даними	Відкриття NetCDF/GeoTIFF, обрізання, маскування, агрегація, експорт
stars	Багатовимірні просторово-часові масиви	Робота з кубами даних і часовими вимірами
ncdf4	Низькорівнева робота з NetCDF	Читання змінних, вимірів, атрибутів і метаданих
sf	Векторні геопросторові дані	Межі громад, областей, басейнів; перепроєктування і просторові операції
exactextractr	Точна зональна статистика	Середні, суми та інші статистики растрів у межах полігонів
tidyverse	Таблична обробка і візуалізація	Очищення, трансформація, групування, підготовка таблиць і графіків
lubridate	Робота з датами і часом	Формування сезонів, років, місяців і часових інтервалів

Належна практика

Для нових проєктів доцільно надавати перевагу пакету terra замість застарілого пакета raster, якщо функціональність і сумісність конкретного робочого процесу це дозволяють. Версії R, RStudio і пакетів бажано фіксувати в описі методики або середовищі проєкту.

4. Робота з часовими рядами та кліматичними індикаторами

Особливо корисним R є для опрацювання часових рядів кліматичних даних. Наприклад, із щоденних значень температури можна автоматично розрахувати середньомісячні, сезонні або річні значення, кількість спекотних днів, тропічних ночей, морозних днів, тривалість хвиль спеки, суму активних температур або кількість днів із перевищенням заданого порогу. Для опадів можна обчислювати річні суми, максимальну добову кількість опадів, максимальну 5-денну суму, кількість днів із сильними опадами, тривалість бездошових періодів, індекси посухи або зміну опадів у відсотках відносно базового періоду.

Приклади температурних показників

- середні місячні, сезонні та річні значення температури;
- кількість спекотних днів і тропічних ночей;
- кількість морозних днів;
- тривалість і повторюваність хвиль спеки;
- суми активних або ефективних температур;
- кількість днів із перевищенням визначених температурних порогів.

Приклади показників опадів

- місячні, сезонні та річні суми опадів;
- максимальна добова кількість опадів;
- максимальна п'ятиденна сума опадів;
- кількість днів із сильними або дуже сильними опадами;
- тривалість бездошових періодів;

- індекси посухи та відносні зміни опадів.

Важливо

Перед розрахунком індикаторів необхідно перевірити календар, одиниці вимірювання, наявність пропусків, повноту років і правила агрегації. Для відтворюваності слід зберігати визначення кожного індикатора, порогові значення та базовий період.

5. Порівняння кліматичних періодів, моделей і сценаріїв

R також зручний для порівняння кліматичних періодів і сценаріїв. Наприклад, можна автоматично розрахувати середнє значення показника за базовий період 1991–2020 рр., окремо — за прогнозний період 2021–2040 рр., а потім отримати растр абсолютної або відносної зміни. Аналогічно можна порівнювати різні сценарії викидів парникових газів, кліматичні моделі або ансамблеві середні. Це особливо важливо для роботи з даними EURO-CORDEX, CMIP, ERA5-Land або E-OBS, де кількість файлів і часових зрізів може бути дуже великою.

- абсолютна зміна: майбутнє значення мінус значення базового періоду;
- відносна зміна: різниця, виражена у відсотках відносно базового періоду;
- ансамблеве середнє: середнє значення за кількома кліматичними моделями;
- діапазон або розкид ансамблю: характеристика міжмодельної невизначеності;
- порівняння сценаріїв: оцінювання залежності результатів від траєкторій майбутніх викидів.

6. Приклад структури R-скрипта

Простий R-скрипт для роботи з кліматичними даними може включати підключення пакетів, відкриття файлу, вибір змінної, визначення часового періоду, розрахунок показника та збереження результату. Схема на рисунку 2 демонструє логіку такого скрипта.

Приклад простого R-скрипта для обробки кліматичних даних (NetCDF)

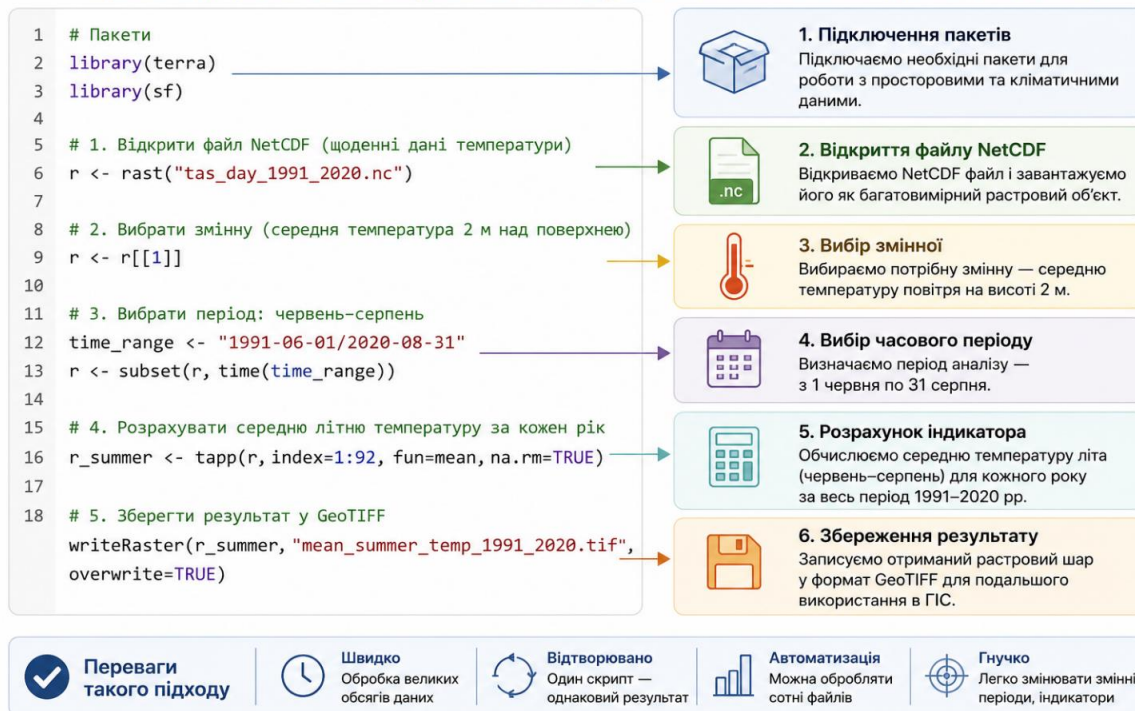


Рисунок 2 – Приклад R-скрипта з поясненням основних операцій

Рекомендація щодо коду

Для сайту або навчального курсу код краще подавати не лише у вигляді зображення, а також як текстовий блок,

який користувач може скопіювати. До коду слід додавати опис структури вхідного файлу, назви змінних, одиниці вимірювання та очікуваний результат.

7. Поєднання кліматичних даних із територіальними межами

Ще одна важлива перевага R — можливість поєднувати кліматичні дані з векторними межами територій. Наприклад, можна автоматично розрахувати середнє значення температури, опадів або будь-якого кліматичного індикатора для кожної громади, області, водозбірної басейну чи природоохоронної території. Результатом може бути таблиця, у якій для кожної територіальної одиниці наведено значення кліматичного показника, його зміну в майбутньому, клас небезпеки або інтегральний індекс. Такі таблиці потім легко використовувати в Excel, QGIS або статистичних моделях.

Перевірка просторової узгодженості

Перед зональною статистикою потрібно перевірити системи координат, геометрію полігонів, просторову роздільну здатність растру, правила оброблення частково перекритих комірок і значення NoData.

8. Розрахунок інтегральних показників кліматичного ризику

Крім простих розрахунків, R дозволяє створювати складніші індикатори кліматичних ризиків. Наприклад, можна нормалізувати кілька кліматичних показників до шкали 1–5, застосувати вагові коефіцієнти, об'єднати їх в інтегральний індекс кліматичної загрози, розрахувати чутливість, здатність до адаптації або вразливість. Усі ці операції можна виконати за єдиною методикою для багатьох секторів, сценаріїв і часових горизонтів, що значно підвищує узгодженість результатів.

- нормалізація показників до спільної шкали;
- перекласифікація за визначеними порогами;
- застосування вагових коефіцієнтів;
- об'єднання показників загрози, експозиції, чутливості та адаптивної спроможності;
- аналіз чутливості результатів до ваг і порогових значень;
- підготовка таблиць і карт інтегрального ризику або вразливості.

Інтерпретація

Інтегральний індекс не є безпосереднім вимірюванням ризику. Його значення залежить від вибраних показників, способів нормалізації, ваг і правил агрегації. Методика має бути прозорою, обґрунтованою та перевіреною за допомогою аналізу чутливості.

9. Порівняння ручної роботи в ГІС та автоматизації в R

Використання R дозволяє істотно прискорити процеси оброблення кліматичних даних завдяки автоматизації виконання повторюваних дій та відсутності необхідності відкриття кожного файлу в програмах ГІС. Якщо передбачається опрацювання великих NetCDF-файлів, які містять добові значення кліматичних показників за десятки років, R зазвичай є ефективнішим інструментом, оскільки дає змогу виконувати пакетне оброблення, керувати пам'яттю та автоматично повторювати однакові операції для багатьох файлів.

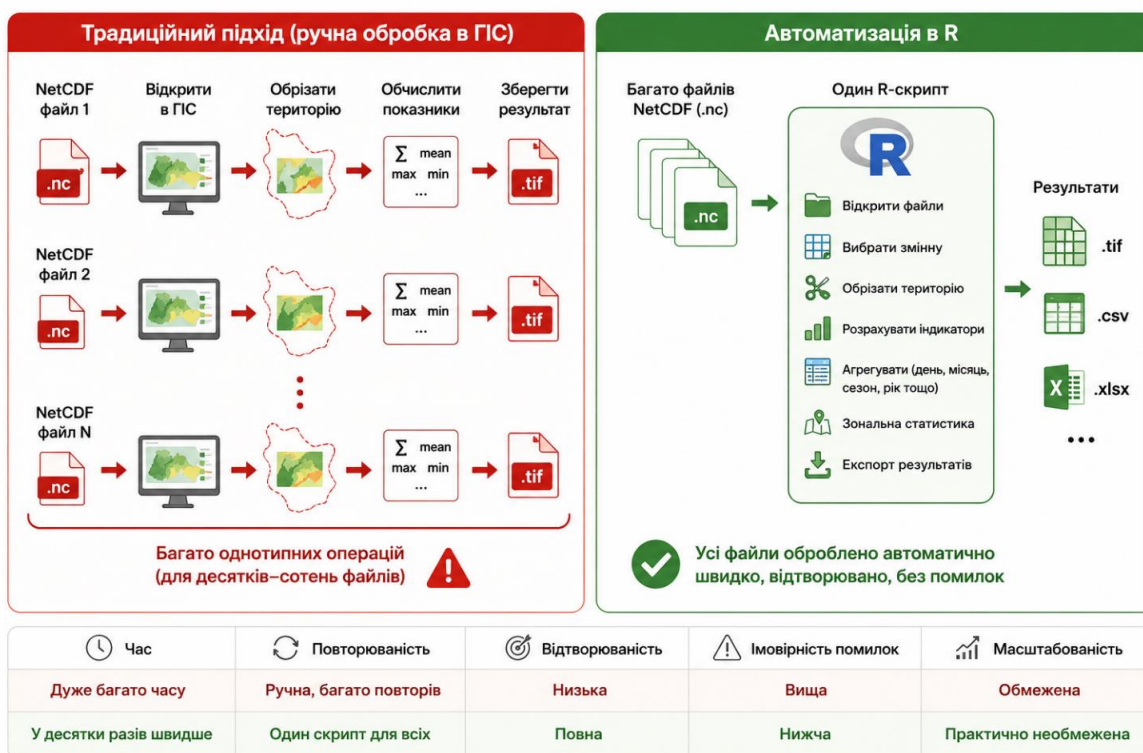


Рисунок 3 – Порівняння ручного та автоматизованого оброблення кліматичних даних у ГІС і R

Критерій	Ручна робота в ГІС	Автоматизація в R
Повторюваність	Операції повторюються вручну для кожного файлу	Один скрипт застосовується до всіх файлів
Відтворюваність	Залежить від документування ручних дій	Повна послідовність зафіксована в коді
Масштабованість	Обмежена для великої кількості файлів	Придатна для пакетного оброблення великих масивів
Візуальний контроль	Зручний інтерактивний перегляд	Потребує окремого формування графіків або карт
Типове застосування	Картографування, перевірка та редагування	Попередня обробка, статистика, індикатори й автоматизація

Таким чином, R / RStudio доцільно використовувати як інструмент попередньої обробки, автоматизованого аналізу та розрахунку кліматичних індикаторів. ГІС у такому підході використовується переважно на завершальному етапі — для перевірки, картографування та візуалізації результатів. Це дозволяє суттєво скоротити час роботи з кліматичними даними, уникнути повторюваних ручних операцій і забезпечити повну відтворюваність усіх етапів аналізу.

10. Практичні рекомендації та належна практика

- починати роботу з перевірки метаданих, назв змінних, одиниць вимірювання, календаря та системи координат;
- не завантажувати в оперативну пам'ять більше даних, ніж потрібно для конкретного розрахунку;
- обмежувати просторову й часову область на ранньому етапі оброблення;
- розділяти скрипт на логічні етапи та додавати зрозумілі коментарі;

- зберігати проміжні результати лише тоді, коли це необхідно для контролю або повторного використання;
- застосовувати однакові визначення індикаторів, базові періоди та правила оброблення для всіх моделей і сценаріїв;
- перевіряти результати за допомогою описової статистики, графіків і вибіркового перегляду карт у ГІС;
- фіксувати версії R, пакетів, вихідних наборів даних та дату їх завантаження;
- зберігати скрипти разом із README-файлом або методичним описом робочого процесу;
- не інтерпретувати модельні проєкції як точний прогноз для конкретної точки або дати.

Відтворюваність

До кожного результату бажано мати скрипт, перелік вхідних файлів, опис змінних, просторову область, часовий період, метод агрегації, визначення індикатора та параметри експорту.