

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Чорноморський національний університет імені Петра Могили
Факультет комп'ютерних наук
Кафедра інтелектуальних інформаційних систем

ДОПУЩЕНО ДО ЗАХИСТУ

В. о. завідувача кафедри інтелектуальних
інформаційних систем

_____ Євген СІДЕНКО
« ____ » _____ 2025 р.

КВАЛІФІКАЦІЙНА РОБОТА
НА ЗДОБУТТЯ ОСВІТНЬОГО СТУПЕНЯ МАГІСТРА
ІНФОРМАЦІЙНА СИСТЕМА АНАЛІЗУ ТЕНДЕНЦІЙ
ЗМІНИ КЛІМАТУ ЗА СУПУТНИКОВИМИ ДАНИМИ

Спеціальність 122 Комп'ютерні науки
Освітня програма «Інтелектуальні інформаційні системи»

Здобувач

_____ Антон РОМАНОВ
« ____ » _____ 2025 р.

Керівник канд. техн. наук, ст. викладач

_____ Віктор ГОЖИЙ
« ____ » _____ 2025 р.

Миколаїв – 2025

Чорноморський національний університет імені Петра Могили
(повне найменування закладу вищої освіти)

Факультет	Комп'ютерних наук
Кафедра	Інтелектуальних інформаційних систем
Рівень вищої освіти	Другий (магістерський)
Освітній ступень	Магістр
Спеціальність	122 Комп'ютерні науки
Освітня програма	Інтелектуальні інформаційні системи

ЗАТВЕРДЖУЮ

В. о. завідувача кафедри інтелектуальних
інформаційних систем

_____ Євген СІДЕНКО

«_____» _____ 2025 р.

ЗАВДАННЯ
на кваліфікаційну роботу здобувача

Романова Антона Валерійовича

(прізвище, ім'я, по батькові здобувача)

1. Тема кваліфікаційної роботи: «Інформаційна система аналізу тенденцій зміни клімату за супутниковими даними».

Керівник роботи: Гожий Віктор Олександрович, канд. техн. наук. старший викладач кафедри інтелектуальних інформаційних систем.

Затверджена наказом ЧНУ ім. Петра Могили від «07» липня 2025 р. № 184.

2. Строк представлення кваліфікаційної роботи «22» грудня 2025 р.

3. Очікуваний результат роботи та початкові дані, якщо такі потрібні система аналізу тенденцій зміни клімату за супутниковими та глобальними температурними даними; набір даних Global Surface Temperatures, що містить історичні глобальні температурні вимірювання (суші та океану), сформовані на основі довгострокових супутникових і наземних спостережень та доступні на платформі Kaggle.

4. Перелік питань, що підлягають розробці: проаналізувати поняття зміни клімату, його показники та основні чинники; вивчити теоретичні основи аналізу кліматичних часових рядів і сучасні методи їх обробки; поставити задачу аналізу глобальних температурних змін за супутниковими даними; проаналізувати структуру обраного набору даних, виконати його попередню обробку та розвідувальний аналіз; побудувати моделі для визначення довгострокових тенденцій, сезонних коливань і температурних аномалій; застосувати методи регресійного аналізу та інші алгоритми для оцінки трендів; оцінити результати аналізу та сформулювати висновки щодо динаміки глобальної температури.
5. Перелік графічних матеріалів: графіки часових рядів глобальної температури; графіки трендів і регресійних моделей; візуалізація температурних аномалій; презентація.

Керівник роботи

(Особистий підпис)

Віктор ГОЖИЙ
(Власне ім'я ПРІЗВИЩЕ)

Здобувач

(Особистий підпис)

Антон РОМАНОВ
(Власне ім'я ПРІЗВИЩЕ)

Дата видачі завдання «03» червня 2025 р.

КАЛЕНДАРНИЙ ПЛАН

кваліфікаційної роботи

Тема: «Інформаційна система аналізу тенденцій зміни клімату за супутниковими даними»

№	Найменування роботи	Початок	Закінчення	Примітки
1	Отримання завдання на виконання кваліфікаційної роботи	03.06.2025	07.06.2025	Виконано
2	Огляд літературних джерел за темою кваліфікаційної роботи	08.06.2025	24.06.2025	Виконано
3	Огляд існуючих методів аналізу і прогнозування часових рядів	25.06.2025	05.07.2025	Виконано
4	Аналіз та попередня обробка часового ряду	06.07.2025	15.09.2025	Виконано
5	Створення окремих прогнозних моделей, оцінка точності прогнозів	16.09.2025	24.11.2025	Виконано
6	Перший попередній захист КР на засіданні комісії кафедри	25.11.2025	25.11.2025	Виконано
7	Корегування роботи за результатами попереднього захисту	26.11.2025	08.12.2025	Виконано
8	Другий попередній захист КР на засіданні кафедри	09.12.2025	09.12.2025	Виконано
9	Доробка та остаточне оформлення КР	10.12.2025	17.12.2025	Виконано
10	Подання КР, її електронної копії та інших документів (відгуку, рецензії) до захисту	15.12.2025	16.12.2025	Виконано

Керівник роботи

(Особистий підпис)

Віктор ГОЖИЙ
(Власне ім'я ПРІЗВИЩЕ)

Здобувач

(Особистий підпис)

Антон РОМАНОВ
(Власне ім'я ПРІЗВИЩЕ)

Дата складання календарного плану
«03» червня 2025 р.

АНОТАЦІЯ

до кваліфікаційної роботи
здобувача групи 601м ЧНУ ім. Петра Могили

Романова Антона Валерійовича

на тему «**ІНФОРМАЦІЙНА СИСТЕМА АНАЛІЗУ ТЕНДЕНЦІЙ ЗМІНИ
КЛІМАТУ ЗА СУПУТНИКОВИМИ ДАНИМИ**»

Актуальність. Зміна клімату є однією з найгостріших глобальних проблем ХХІ століття. Довгострокові супутникові спостереження забезпечують об'єктивний, просторово повний та уніфікований набір кліматичних даних. Їх використання для аналізу тенденцій зміни температури та атмосферних характеристик дозволяє будувати обґрунтовані прогнози і приймати рішення.

Об'єкт дослідження – процес дослідження глобальних кліматичних параметрів.

Предмет дослідження – методи та алгоритми обробки кліматичних даних у вигляді часових рядів.

Метою кваліфікаційної роботи є підвищення ефективності аналізу зміни клімату.

Пояснювальна записка складається зі вступу, чотирьох розділів, висновків та додатків.

У першому розділі розглянуто теоретичні засади аналізу кліматичних змін, охарактеризовано сучасні підходи до оцінювання кліматичних тенденцій, джерела кліматичних даних та аналітичні методи дослідження кліматичних часових рядів.

У другому розділі проаналізовано інформаційні технології та програмні засоби обробки кліматичних даних, розглянуто інструментарій мови програмування R та методику аналізу температурних аномалій на основі супутникових спостережень.

У третьому розділі розроблено структуру інформаційної системи аналізу кліматичних тенденцій, виконано завантаження та попередню обробку кліматичних

даних, очищення і формування фінального набору часових рядів, а також реалізовано основні програмні модулі системи.

У четвертому розділі здійснено комплексний аналіз та візуалізацію кліматичних даних, виконано статистичний і трендовий аналіз температурних аномалій, STL-декомпозицію часових рядів, визначено найтепліші роки за весь період спостережень та проведено прогнозування глобальних температурних змін із використанням моделей ARIMA, ETS та нейронних мереж.

У результаті виконання роботи створено програмну інформаційну систему аналізу кліматичних тенденцій, реалізовану мовою програмування R, яка забезпечує автоматизацію процесів обробки, аналізу, візуалізації та прогнозування глобальних температурних змін. Отримані результати підтверджують наявність статистично значущого та прискореного тренду глобального потепління. Розроблена система може бути використана для подальших наукових досліджень, освітніх цілей та практичних кліматичних оцінок.

Кваліфікаційна робота містить 84 сторінки, 29 рисунків, 12 таблиць, 30 використаних джерел та 1 додаток.

Ключові слова: кліматичні зміни, супутникові дані, температурні аномалії, часові ряди, трендовий аналіз, прогнозування, ARIMA, STL-декомпозиція, нейронні мережі, інформаційна система, мова R., ARIMA, STL-декомпозиція, інформаційна система, мова R.

ABSTRACT

for master's qualification work
of the student of group 601m of Petro Mohyla Black Sea National University

Anton Romanov

on the topic «**INFORMATION SYSTEM FOR ANALYZING CLIMATE
CHANGE TRENDS BASED ON SATELLITE DATA**»

Relevance. Climate change is one of the most pressing global challenges of the 21st century. Long-term satellite observations provide an objective, spatially comprehensive, and unified set of climate data. Their use for analyzing temperature trends and atmospheric characteristics enables the development of scientifically grounded forecasts and supports informed decision-making.

Object of the study is the process of analyzing global climate parameters.

Subject of the study is the methods and algorithms for processing climate data in the form of time series.

The aim of the qualification work is to improve the efficiency of climate change analysis.

The explanatory note consists of an introduction, four chapters, conclusions, and appendices.

The first chapter examines the theoretical foundations of climate change analysis, describes modern approaches to assessing climate trends, sources of climate data, and analytical methods for studying climate time series.

The second chapter analyzes information technologies and software tools for climate data processing, considers the toolkit of the R programming language, and presents the methodology for analyzing temperature anomalies based on satellite observations.

The third chapter develops the structure of an information system for climate trend analysis, implements data loading and preprocessing, data cleaning and formation of the final time-series dataset, and realizes the core software modules of the system.

The fourth chapter performs a comprehensive analysis and visualization of climate data, including statistical and trend analysis of temperature anomalies, STL decomposition of time series, identification of the warmest years over the entire observation period, and forecasting of global temperature changes using ARIMA, ETS, and neural network models.

As a result of the study, a software-based information system for climate trend analysis was developed and implemented using the R programming language. The system automates data processing, analysis, visualization, and forecasting of global temperature changes. The obtained results confirm the presence of a statistically significant and accelerated trend of global warming. The developed system can be used for further scientific research, educational purposes, and practical climate assessments.

The qualification work contains 84 pages, 29 figures, 12 tables, 30 references, and 1 appendice.

Keywords: climate change, satellite data, temperature anomalies, time series, trend analysis, forecasting, ARIMA, STL decomposition, neural networks, information system, R programming language.

ЗМІСТ

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ	2
ВСТУП.....	4
1 ТЕОРЕТИЧНІ ЗАСАДИ АНАЛІЗУ КЛІМАТИЧНИХ ЗМІН.....	6
1.1 Кліматичні зміни як глобальна проблема сучасності.....	6
1.2 Підходи до оцінювання тенденцій кліматичних змін	9
1.3 Джерела кліматичних даних.....	12
1.4 Аналітичні методи дослідження кліматичних часових рядів.....	15
1.5 Аналіз існуючих досліджень	18
Висновки до розділу 1.....	19
2 ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА ЗАСОБИ АНАЛІЗУ КЛІМАТИЧНИХ ДАНИХ.....	21
2.1 Супутникові дані як джерело інформації для кліматичного аналізу.....	21
2.2 Засоби обробки кліматичних даних	23
2.3 Інструментарій мови R для кліматичного аналізу	25
2.4 Методика аналізу кліматичних тенденцій	28
Висновки до розділу 2.....	29
3 РОЗРОБКА ІНФОРМАЦІЙНОЇ СИСТЕМИ	31
3.1 Структура інформаційної системи	31
3.2 Підключення та конфігурація програмних бібліотек.....	33
3.3 Створення структури програмного проєкту	34
3.4 Первинний аналіз структури кліматичних даних	36
3.5 Попередня обробка кліматичних даних.....	38
3.6 Аналіз глобальних температурних даних	42
Висновки до розділу 3.....	50
4 ТРЕНДОВИЙ АНАЛІЗ ТА МОДЕЛЮВАННЯ ЗМІН КЛІМАТУ	53
4.1 Лінійна регресія тренду глобальних температурних аномалій	53
4.2 Непараметричні методи трендового аналізу глобальних температурних аномалій.....	57
4.3 Порівняльний аналіз параметричних і непараметричних методів оцінювання тренду.....	59
4.4 Прогнозування глобальних температурних аномалій.....	60
Висновки до розділу 4.....	69
ВИСНОВКИ.....	70
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ.....	72
ДОДАТОК А Код для побудови інформаційної системи.....	75

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАКИ

ARIMA	–	AutoRegressive Integrated Moving Average
CERES	–	Clouds and the Earth’s Radiant Energy System
ETS	–	Error Trend Seasonality
GSM	–	Global Surface Mean
IPCC	–	Intergovernmental Panel on Climate Change
MODIS	–	Moderate Resolution Imaging Spectroradiometer
NASA	–	National Aeronautics and Space Administration
NOAA	–	National Oceanic and Atmospheric Administration
STL	–	Seasonal Trend decomposition using Loess

ВСТУП

У контексті інтенсивного розвитку глобального потепління проблема дослідження кліматичних змін набуває виняткового наукового та практичного значення. Протягом останніх десятиліть фіксується стабільне зростання середньорічних температур, підвищення частоти екстремальних погодних явищ, танення льодовикових масивів та підняття рівня Світового океану. Сукупність цих процесів істотно впливає на функціонування природних екосистем, здоров'я населення, а також на енергетичну, економічну та продовольчу безпеку держав. У зв'язку з цим моніторинг кліматичних параметрів та кількісна оцінка їх довгострокових тенденцій є одним із пріоритетних завдань сучасної науки.

Супутникові технології відіграють ключову роль у дослідженні глобальних кліматичних процесів, оскільки забезпечують отримання просторово та часово узгоджених кліматичних даних у масштабах усієї планети. Такі дані дають змогу відстежувати зміни температури поверхні Землі, характеристик хмарності, радіаційного балансу та низки інших атмосферних параметрів у довготривалій перспективі. Водночас оброблення великих масивів супутникової інформації потребує застосування спеціалізованих методів статистичного, просторового та часово-рядного аналізу, а також відповідного програмного забезпечення.

Метою кваліфікаційної роботи є розроблення та обґрунтування методичних підходів до аналізу тенденцій зміни клімату на основі відкритих супутникових даних, а також дослідження інструментів, алгоритмів і програмних засобів, необхідних для створення системи виявлення та оцінювання трендів глобального потепління.

Отримані результати слугуватимуть теоретичним і практичним підґрунтям для побудови програмної системи аналізу кліматичних тенденцій, здатної підвищити ефективність екологічного моніторингу та забезпечити науково обґрунтоване прогнозування кліматичних змін.

Для досягнення поставленої мети необхідно виконати такі завдання:

- здійснити аналіз предметної області та вивчити теоретичні засади

дослідження кліматичних змін;

- проаналізувати джерела кліматичних даних та визначити формат, структуру і особливості їх використання;
- розробити вимоги до інформаційної системи аналізу кліматичних тенденцій;
- спроектувати архітектуру програмної системи, включаючи структуру каталогів, модулі обробки, зберігання та візуалізації даних;
- сформувати методику завантаження, попередньої обробки, очищення та підготовки кліматичних часових рядів;
- реалізувати засоби статистичного, трендового та графічного аналізу температурних аномалій;
- виконати декомпозицію часових рядів, визначення довгострокових тенденцій і ключових кліматичних показників;
- розробити програмну інформаційну систему мовою R, провести тестування її функціональних модулів та оцінити точність отриманих результатів.

Для реалізації методів обробки та аналізу кліматичних даних у роботі використано програмне середовище RStudio та пакети tidyverse, lubridate, ggplot2, forecast, tseries, які забезпечили зручне завантаження, трансформацію, статистичний аналіз, побудову графіків та моделювання часових рядів.

1 ТЕОРЕТИЧНІ ЗАСАДИ АНАЛІЗУ КЛІМАТИЧНИХ ЗМІН

1.1 Кліматичні зміни як глобальна проблема сучасності

Кліматичні зміни є одним із ключових глобальних викликів XXI століття – вони зачіпають всі компоненти природного середовища та соціально-економічну сферу, включно зі здоров'ям населення, продовольчою безпекою, інфраструктурою та екосистемами. Сучасні спостереження однозначно свідчать про те, що антропогенна діяльність, перш за все випуск парникових газів, є головною причиною спостережуваного потепління в атмосфері та океан. Це положення підкреслено у останньому оцінному звіті Міжурядової групи експертів з кліматичних змін (IPCC AR6) [1].

Основні прояви і наслідки кліматичних змін:

- підвищення глобальної середньої температури поверхні землі і океанів, що виражається в послідовних рекордних десятиліттях і зростанні частоти хвиль тепла;
- інтенсифікація екстремальних погодних явищ (зливи, повені, посухи, пожежі), які мають кумулятивні соціально-економічні наслідки [2];
- танення льодовиків та локальне скорочення криги в Арктиці, що сприяє підняттю рівня моря та зміні морських екосистем;
- зрушення кліматичних зон і зміна сезонних циклів, що впливає на сільське господарство, біорізноманіття та водні ресурси.

Динаміку глобальної середньої температури наведено в табл. 1.1.

Таблиця 1.1 – Динаміка глобальної середньої температури поверхні Землі за період 1901–2023 років [1, 3–4]

Період (роки)	Середнє відхилення температури від базового періоду (°C, відносно 1850–1900)	Основні характеристики періоду
1901–1930	+0,07	Початкове повільне зростання температури після «Малого льодовикового періоду».
1931–1960	+0,15	Локальне потепління, переважно в Північній півкулі.
1961–1990	+0,30	Стабілізація клімату, зростання концентрації CO ₂ до 350 ppm.
1991–2000	+0,45	Початок інтенсивного глобального потепління.
2001–2010	+0,65	Послідовне перевищення середньорічних температурних рекордів.
2011–2020	+1,02	Найтепліше десятиліття в історії спостережень.
2021–2023	+1,15	Рекордні температури 2023 р. – на $\approx 1,4$ °C вище допромислового рівня.

Як видно з таблиці 1.1, з початку XX століття середня температура поверхні Землі зросла більш ніж на 1°C, причому понад 70% цього підвищення припадає на період після 1970 року.

Згідно з оцінками NASA GISS і IPCC, потепління має нерівномірний просторовий розподіл – воно особливо виражене в північних широтах і над континентами, тоді як океани демонструють дещо менше підвищення через теплову інерційність.

Прискорення зростання температури після 1990-х років збігається з різким підвищенням антропогенних викидів CO₂, CH₄ та N₂O, що підтверджує висновки IPCC про переважно антропогенну природу сучасного потепління

Наукові підстави та доказова база змін клімату [5]:

- багатолінійні спостереження (наземні метеостанції, океанічні виміри, супутникові дані та палеокліматичні архіви) демонструють послідовне збільшення концентрацій парникових газів з індустріальної доби та відповідне потепління;
- аналіз глобальних часових рядів температур показує, що кожне з останніх чотирьох десятиліть було теплішим за попереднє, а сучасне середнє потепління щодо 1850–1900 становить близько $1,0\text{--}1,1^{\circ}\text{C}$;
- численні дослідження також вказують на зв'язок між антропогенним підвищенням концентрацій парникових газів і ростом частоти гарячих максимумів і екстремальних подій;

Роль супутникових спостережень у вивченні кліматичних змін [6]:

- супутники забезпечують просторово-повне і послідовне покриття Землі, що дозволяє відстежувати просторово-часові патерни температури, льоду, рослинності та інших кліматичних індикаторів;
- супутникові дані особливо цінні для віддалених і маломоніторених регіонів (полярні області, океани), де наземні вимірювання є обмеженими або відсутніми;
- разом із наземними серіями спостережень супутникові набори даних дозволяють формувати багатовимірні довгі часові ряди та перевіряти моделі кліматичної динаміки.

Соціально-економічний аспект проблеми:

- зміни клімату уже спричиняють прямі економічні збитки через пошкодження інфраструктури, зниження врожайності в окремих регіонах і збільшення витрат на охорону здоров'я через тепловий стрес;
- ризики кліматичної нерівності посилюються, оскільки найбільш уразливі країни та спільноти мають менше ресурсів для адаптації;
- наукові підходи й потреба в подальших дослідженнях:
- необхідні подальші дослідження з точнішої кількісної атрибуції екстремальних подій людському впливу, а також з вивчення нелінійних процесів і можливих порогових (tipping) ефектів у кліматичній системі;

– важливим є розвиток інтегрованих спостережних систем, що поєднують супутникові продукти, наземні виміри та моделі для підвищення надійності довгострокових оцінок кліматичних тенденцій.

Висновки до підрозділу 1.1

Кліматичні зміни є однією з найважливіших сучасних проблем, яка має як природні, так і соціально-економічні виміри. Консолідована доказова база свідчить про переважно антропогенний характер спостережуваного потепління, зростання частоти екстремальних явищ та широкими наслідками для біосфери та суспільства. Для прийняття ефективних управлінських рішень і адаптаційних заходів необхідні якісні, просторово-часові оцінки тенденцій – роль супутникових даних у цьому процесі є критичною.

1.2 Підходи до оцінювання тенденцій кліматичних змін

Зміна клімату – це складний багатовимірний процес, який охоплює атмосферу, гідросферу, літосферу та біосферу [7].

Для виявлення та кількісного опису кліматичних тенденцій застосовують систему показників, методів і аналітичних підходів, які дозволяють оцінити зміни у просторово-часовому вимірі.

1.2.1 Кліматичні індикатори

Кліматичні індикатори – це узагальнені показники, що відображають стан кліматичної системи та напрям її розвитку.

До основних індикаторів, які використовуються у світовій практиці, належать [8]:

- середня глобальна температура – найважливіший показник, який відображає енергетичний баланс системи Земля–атмосфера;
- кількість опадів – відображає зміни у водному циклі, частоту засух і паводків;

- альbedo поверхні – частка відбитої сонячної енергії; його зменшення внаслідок танення льодовиків підсилює нагрівання;
- вологість повітря та хмарність – визначають тепловий режим і процеси конвекції;
- рівень Світового океану – показник акумульованих наслідків глобального потепління;
- площа крижаного покриву та льодовиків – відображає стабільність кріосфери;
- вміст вуглекислого газу (CO_2) – основний антропогенний фактор кліматичних змін.

1.2.2 Показники глобального потепління

Кількісна оцінка потепління проводиться за допомогою спеціальних показників, що дозволяють порівнювати результати у часі та просторі.

Найуживаніші з них:

- середньорічне зростання температури ($^{\circ}\text{C}/\text{десятиліття}$);
- температурна аномалія – різниця між поточним і базовим періодом (найчастіше 1951–1980 pp.);
- індекс екстремальних температур (TX_x , TN_n) – максимальні і мінімальні температури за сезон;
- індекс теплових хвиль (Heat Wave Duration Index) – тривалість періодів із температурою вище порогових значень;
- індекс вологості (SPI , SPEI) – показник кліматичного балансу між опадами і випаровуванням.

1.2.3 Підходи до оцінювання кліматичних змін

Оцінювання тенденцій клімату базується на поєднанні спостережень, математичних моделей і статистичних методів.

Основні підходи поділяють на три великі групи [9]:

- емпіричні (описові) – базуються на аналізі спостережуваних даних: температур, опадів, рівня моря; передбачають обчислення середніх, стандартних відхилень, аномалій;
- моделювальні – використовують кліматичні моделі загальної циркуляції атмосфери (GCM), які імітують фізичні процеси в атмосфері та океані;
- індексні (інтегральні) – передбачають формування узагальнених показників зміни клімату, наприклад, індекс глобальної температурної аномалії, індекс ENSO або NDVI для рослинності.

Таким чином, ці підходи взаємодоповнюють одне одного: моделі забезпечують прогноз, емпіричний аналіз – підтвердження трендів, а індексні методи дозволяють стисло охарактеризувати стан системи.

Таблиця 1.2 – Основні підходи до оцінювання тенденцій кліматичних змін

Підхід	Джерела даних	Основна характеристика	Переваги	Обмеження
Емпіричний	Метеостанції, супутники	Аналіз спостережуваних даних	Простота, об'єктивність	Залежність від якості даних
Моделювальний	GCM, кліматичні симуляції	Прогноз майбутніх кліматичних станів	Враховує фізичні процеси	Висока обчислювальна складність
Індексний	Комбіновані джерела	Узагальнення багатьох параметрів	Порівнюваність у часі та просторі	Потребує калібрування

Як видно з таблиці, сучасна кліматологія поєднує емпіричні спостереження, моделювання та індексний аналіз для комплексної оцінки тенденцій.

Така багаторівнева структура дозволяє не лише фіксувати поточні зміни, але й прогнозувати подальший розвиток кліматичних процесів.

Висновки до підрозділу 1.2

Оцінювання тенденцій кліматичних змін базується на системному підході, що охоплює як спостережувані дані, так і моделювання фізичних процесів.

Ключову роль відіграють кліматичні індикатори – температура, опади, альbedo, рівень океану.

Різні методичні підходи забезпечують різні рівні деталізації: від опису поточних змін до прогнозування глобальних сценаріїв. Таке поєднання формує наукову основу для подальшого аналізу часових рядів кліматичних параметрів.

1.3 Джерела кліматичних даних

Для вивчення кліматичних тенденцій надзвичайно важливою є наявність достовірних, просторово повних і тривалих за часом рядів спостережень. Основними джерелами кліматичної інформації є глобальні метеорологічні архіви, супутникові місії та композитні кліматичні продукти, що інтегрують дані з різних сенсорів і станцій. Саме комбінація цих джерел забезпечує можливість комплексного аналізу змін температури, хмарності, випромінювання та інших атмосферних параметрів.

1.3.1 Глобальні метеорологічні та кліматичні бази даних

Метеорологічні архіви містять спостереження з наземних станцій, буїв, кораблів та радіозондів, які ведуться протягом десятиліть. Найбільш відомі з них:

- Global Historical Climatology Network (GHCN) – база Національного управління океанічних і атмосферних досліджень США (NOAA), що містить понад 100 тисяч метеостанцій у всьому світі; дані охоплюють період з початку XX ст. і регулярно оновлюються [10];
- ERA5 (ECMWF Reanalysis v5) – кліматичний реаналіз Європейського центру середньострокових прогнозів погоди (ECMWF), який інтегрує наземні, аерологічні та супутникові спостереження у єдину 3D-модель атмосфери [11];
- HadCRUT – глобальна температура поверхні Землі, яку розраховує Британський метеорологічний офіс (Hadley Centre) разом з Університетом Східної Англії; використовується у звітах IPCC [12];

– Berkeley Earth Dataset – незалежний набір кліматичних спостережень, який об'єднує дані з понад 40 тисяч станцій для виявлення довгострокових трендів температури [13].

Такі бази використовуються для валідації супутникових спостережень і перевірки результатів статистичних моделей у дослідженнях клімату.

1.3.2 Відкриті супутникові місії та програми спостережень

Дистанційне зондування Землі (ДЗЗ) стало основою сучасної кліматології, оскільки дозволяє проводити глобальні спостереження температури, хмарності, альбедо та випромінювання з високою частотою оновлення. Найважливіші відкриті супутникові місії включають:

– MODIS (Moderate Resolution Imaging Spectroradiometer) – сенсор, встановлений на супутниках Terra та Aqua (NASA), працює з 1999 року, надає щоденні глобальні спостереження температури поверхні, хмарності, рослинності та енергетичного балансу [14];

– NOAA AVHRR (Advanced Very High Resolution Radiometer) – серія метеорологічних супутників, які забезпечують найтриваліший безперервний ряд спостережень температури поверхні океану (SST) з 1981 року [15];

– Sentinel (програма Copernicus, ESA) – європейська програма моніторингу навколишнього середовища; Sentinel-3 забезпечує дані про температуру поверхні океану й суші, а Sentinel-5P – спостереження за атмосферними газами (CO₂, NO₂, O₃) [16];

– Landsat (NASA/USGS) – програма спостереження Землі, що діє з 1972 року; її знімки широко використовуються для вивчення змін покриття поверхні, індексів рослинності та локальних кліматичних ефектів [17];

– CERES (Clouds and the Earth's Radiant Energy System) – серія сенсорів для вимірювання відбитого сонячного та випроміненого теплового потоку, важливих для аналізу енергетичного балансу Землі [18].

Таблиця 1.3 – Основні відкриті супутникові місії для кліматичних досліджень

Супутникова місія	Агентство / програма	Основні параметри спостережень	Просторова роздільність	Період дії
MODIS (Terra/Aqua)	NASA	Температура поверхні, хмарність, альbedo, NDVI	250–1000 м	1999-дотепер
NOAA AVHRR	NOAA	Температура поверхні океану (SST), хмарність, відбиття	1 км	1981-дотепер
Sentinel-3	ESA (Copernicus)	Температура суші та океану, висота хвиль, кольори океану	300 м – 1 км	2016-дотепер
Sentinel-5P	ESA	Концентрації атмосферних газів (O ₃ , NO ₂ , CO ₂)	7×3,5 км	2017-дотепер
Landsat-8/9	NASA / USGS	Відбиття поверхні, рослинність, індекси LST	30–100 м	2013-дотепер
CERES	NASA	Радіаційний баланс Землі (довгохвильове/короткохвильове випромінювання)	~20 км	2000-дотепер

Як видно з таблиці 1.3, сучасні супутникові місії охоплюють усі основні компоненти кліматичної системи – атмосферу, океани, суходіл і радіаційний баланс.

MODIS та Sentinel-3 забезпечують багаторічні глобальні спостереження температури поверхні; NOAA AVHRR має унікальний архів морських температур; Sentinel-5P дає змогу відстежувати концентрації газів, відповідальних за парниковий ефект; програма Landsat доповнює ці дані високодетальними спостереженнями зміни земного покриву.

Поєднання цих джерел у межах єдиної системи аналізу дає можливість створювати повноцінні багатовимірні кліматичні моделі.

1.3.3 Принципи формування супутникових кліматичних наборів спостережень

Супутникові продукти формуються шляхом обробки радіометричних вимірювань з використанням алгоритмів корекції атмосфери, геоприв'язки та калібрування. Основні етапи:

- радіометрична обробка – перетворення сигналів сенсорів у фізичні величини (відбиття, яскравісна температура);

- геометрична корекція – усунення спотворень, пов'язаних з кутом спостереження, рельєфом і орбітою;
- атмосферна корекція – компенсація впливу парів води, аерозолів і озону;
- агрегація даних – усереднення вимірювань у просторових та часових межах для формування глобальних продуктів (наприклад, місячних або сезонних композитів);
- валідація – порівняння супутникових даних із наземними метеостанціями або буями для оцінки точності.

Таким чином, кожен кліматичний продукт проходить багатоступеневу перевірку перед публікацією, що гарантує високу достовірність спостережень.

Висновки до підрозділу 1.3

Джерела кліматичних даних формують основу для вивчення тенденцій зміни клімату.

Наземні метеорологічні архіви забезпечують довгострокові часові ряди, тоді як супутникові спостереження дають просторову повноту та глобальне покриття.

Використання відкритих місій, таких як MODIS, NOAA, Sentinel і Landsat, забезпечує дослідників доступом до високоякісних даних про температуру, хмарність, випромінювання та інші параметри атмосфери.

Завдяки стандартизації алгоритмів обробки і валідації ці дані стали надійною основою для побудови моделей, оцінки трендів і прогнозування кліматичних процесів.

У межах цієї роботи саме супутникові температурні спостереження обрано як основне джерело даних для подальшого аналізу тенденцій глобального потепління.

1.4 Аналітичні методи дослідження кліматичних часових рядів

Після визначення основних підходів і показників необхідним етапом є вибір конкретних аналітичних методів для роботи з часовими рядами кліматичних даних.

Такі ряди, як правило, мають довгу тривалість, наявність сезонних циклів, шумів і пропусків, що вимагає використання спеціальних статистичних і алгоритмічних засобів.

1.4.1 Класифікація методів

Аналітичні методи дослідження кліматичних часових рядів поділяються на кілька груп:

- описові статистичні методи – розрахунок середніх, стандартних відхилень, варіаційних коефіцієнтів;
- трендові методи – визначають напрям і швидкість зміни кліматичних показників (лінійна регресія, тест Манна–Кендалла, оцінка Сенса);
- сезонні методи – виявляють циклічність і повторюваність (декомпозиція STL, ковзні середні, спектральний аналіз);
- аномалійні методи – ідентифікують відхилення від середнього рівня (Z-score, IQR, виявлення екстремумів);
- інтелектуальні методи (машинне навчання) – класифікація, прогнозування та виявлення нелінійних закономірностей (нейронні мережі, Random Forest, ARIMA, LSTM).

1.4.2 Методи виявлення трендів

Виявлення тренду дозволяє визначити загальний напрям зміни кліматичних параметрів. Найчастіше застосовуються:

- тест Манна–Кендалла – непараметричний тест, який оцінює статистичну значущість тренду без вимоги нормальності розподілу [19];
- метод Сенса – оцінює нахил тренду, тобто середню зміну температури за рік [20];
- лінійна або поліноміальна регресія – використовується для моделювання тенденцій і прогнозу [21];

- ковзні середні – дозволяють згладити випадкові коливання, зберігаючи основну динаміку [22];
- вейвлет-аналіз – виявляє часові інтервали, де відбувається зміна характеру тренду [23].

Таблиця 1.4 – Основні аналітичні методи обробки кліматичних часових рядів

Метод	Тип	Мета застосування	Переваги	Обмеження
Манна–Кендалла	Непараметричний	Виявлення наявності тренду	Не залежить від розподілу даних	Не оцінює величину тренду
Сенса	Непараметричний	Оцінка нахилу тренду	Простота, стійкість	Чутливий до пропусків
STL-декомпозиція	Статистичний	Виділення сезонності	Гнучкість, наочність	Потребує рівномірних інтервалів
ARIMA	Стохастичний	Прогнозування часових рядів	Висока точність короткострокових прогнозів	Потребує стаціонарності
LSTM	Машинне навчання	Нелінійне прогнозування	Добре працює з великими даними	Висока складність реалізації

Як видно з таблиці 1.4, кожен метод має своє призначення залежно від типу даних і дослідницької мети. Тест Манна–Кендалла і Сенса підходять для статистичного виявлення трендів, STL-декомпозиція – для дослідження сезонності, а сучасні моделі ARIMA і LSTM – для прогнозування майбутніх кліматичних змін.

1.4.3 Аналіз сезонності та аномалій

Кліматичні дані мають виражену сезонність, тому її виокремлення дає змогу краще зрозуміти структуру температурних змін.

Основні методи:

- STL-декомпозиція – розділення ряду на тренд, сезонність і залишок [24];
- Fourier-аналіз – визначення періодичних компонентів (річні, піврічні цикли) [25];
- Z-score і IQR – кількісне виявлення температурних аномалій [26];

– кластеризація (k-means, DBSCAN) – просторове виявлення регіонів із схожими аномалійними характеристиками [27].

Висновки до підрозділу 1.4

Аналітичні методи дозволяють перетворити великі обсяги кліматичних даних у кількісні висновки щодо напрямів і закономірностей змін.

Поєднання класичних статистичних тестів і сучасних алгоритмів машинного навчання формує гнучкий інструментарій для аналізу трендів, сезонних коливань і прогнозування.

У межах цієї роботи ці методи будуть застосовані для аналізу супутникових температурних рядів, що забезпечить наочну оцінку просторово-часових тенденцій глобального потепління.

1.5 Аналіз існуючих досліджень

Аналіз аналогічних досліджень показав, що сучасні наукові підходи до аналізу кліматичних змін базуються на використанні довготривалих рядів супутникових та наземних спостережень у поєднанні з методами статистичного аналізу й прогнозування. Зокрема, у роботі J. Hansen [28] на основі набору даних GISTEMP встановлено, що середня глобальна температура поверхні Землі зросла приблизно на $1,1^{\circ}\text{C}$ з кінця XIX століття, при цьому темпи потепління після 1980 року істотно прискорилися. У Шостому звіті Міжурядової групи експертів зі зміни клімату (IPCC) на основі супутникових та наземних спостережень підтверджено, що кожне з останніх чотирьох десятиліть було теплішим за будь-які попередні з 1850 року, а прогнозоване потепління до 2100 року може сягнути від $1,5^{\circ}\text{C}$ до $4,4^{\circ}\text{C}$ залежно від сценарію викидів парникових газів. У дослідженні A. Benestad [29] методи регресійного аналізу та ARIMA використано для прогнозування глобальної температури, де встановлено середній темп зростання близько $0,18\text{--}0,22^{\circ}\text{C}$ за десятиліття. Крім того, у роботі S. Deo [30] із використанням експоненційного згладжування та нейронних мереж доведено, що моделі часових рядів

забезпечують високу точність прогнозу ($RMSE < 0,15^{\circ}C$) на інтервалі 10-20 років. Це підтверджує доцільність застосування методів аналізу часових рядів і прогнозування для дослідження тенденцій глобального потепління.

Висновки до розділу 1

У розділі розглянуто науково-методичні засади аналізу тенденцій зміни клімату на глобальному рівні.

Кліматичні зміни визначено як одну з ключових екологічних та соціально-економічних проблем XXI століття, що проявляється у зростанні середньорічних температур, збільшенні частоти екстремальних погодних явищ, таненні льодовиків і підвищенні рівня Світового океану.

Встановлено, що сучасні дослідження базуються на комплексному використанні метеорологічних і супутникових спостережень, які забезпечують просторово-часову повноту кліматичних даних.

Основними підходами до оцінювання тенденцій кліматичних змін є емпіричний аналіз спостережень, фізичне моделювання та формування індексів кліматичних аномалій.

Ключовими індикаторами виступають температура, кількість опадів, альbedo, вологість і рівень океану, що дозволяють відображати стан кліматичної системи у динаміці.

Показано, що джерела кліматичних даних представлені міжнародними супутниковими місіями (MODIS, NOAA, Sentinel, Landsat), глобальними архівами (NASA GISTEMP, ECMWF, Copernicus) і відкритими платформами (Kaggle, Google Earth Engine), які забезпечують відкритий доступ до високоточних вимірювань.

Принципи формування кліматичних наборів базуються на інтеграції дистанційних спостережень, калібруванні та перевірці достовірності даних.

Окрему увагу приділено аналітичним методам дослідження кліматичних часових рядів. До них належать класичні статистичні методи (тест Манна–

Кендалла, оцінювач Сенса, ковзні середні), сезонні методи (STL-декомпозиція, спектральний аналіз) та сучасні алгоритми машинного навчання (ARIMA, LSTM).

Вони дозволяють ідентифікувати тренди, сезонність і аномалії, а також здійснювати прогнозування кліматичних показників на основі великих супутникових даних.

Узагальнюючи викладене, можна зробити висновок, що ефективно дослідження кліматичних тенденцій потребує поєднання багаторівневих підходів – від збору та валідації даних до застосування сучасних аналітичних і візуалізаційних методів. Таке поєднання створює наукове підґрунтя для побудови комплексної системи аналізу кліматичних трендів, що буде реалізована у подальших розділах роботи.

2 ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА ЗАСОБИ АНАЛІЗУ КЛІМАТИЧНИХ ДАНИХ

2.1 Супутникові дані як джерело інформації для кліматичного аналізу

Супутникові спостереження є одним із найважливіших джерел інформації про стан кліматичної системи Землі.

Вони забезпечують безперервне, глобальне та об'єктивне вимірювання атмосферних і поверхневих параметрів, що робить їх незамінними для аналізу тенденцій зміни клімату.

На відміну від наземних метеостанцій, супутникові дані дозволяють охопити важкодоступні регіони – океани, полярні області, пустелі – і отримати єдину систему показників у просторово-часовому масштабі.

2.1.1 Особливості обробки супутникових температурних даних

Обробка супутникових температурних даних передбачає кілька ключових етапів:

- радіометричне калібрування – приведення сирих сигналів до фізичних величин температури випромінювання;
- геометрична корекція – усунення похибок, пов'язаних із положенням супутника, кутом спостереження та орбітальною динамікою;
- фільтрація шумів – усунення впливу хмарності, атмосферних аерозолів і технічних спотворень сигналу;
- перетворення до геопросторового формату – узгодження з координатною сіткою (широта–довгота) для подальшої візуалізації в ГІС або аналітичному середовищі;
- темпоральна агрегація – усереднення даних за місяць, рік або сезон для аналізу довгострокових тенденцій.

Такі етапи забезпечують уніфікацію даних і дозволяють проводити міжрічні порівняння без втрати точності.

2.1.2 Переваги використання супутникових вимірювань для глобального аналізу

Супутникові вимірювання мають низку переваг порівняно з традиційними наземними спостереженнями:

- забезпечують повне просторове покриття – від екватора до полюсів;
- дають високу часову роздільну здатність – від годинних до добових вимірювань;
- дозволяють моніторити важкодоступні регіони – океани, пустелі, гірські райони;
- створюють гомогенну базу даних для порівнянь між регіонами та періодами;
- можуть бути інтегровані з ГІС-системами для просторової візуалізації;
- підтримують автоматичну обробку великих масивів за допомогою мов програмування, зокрема R та Python.

Завдяки цим перевагам супутникові спостереження стали основним джерелом кліматичних індикаторів у звітах Міжурядової групи експертів зі зміни клімату (ІРСС, 2023) та програмах моніторингу Copernicus.

2.1.3 Структура і характеристика набору даних Global Surface Temperatures

У межах даного дослідження використовується відкритий набір даних Global Surface Temperatures, опублікований на платформі Kaggle.

Цей набір базується на комбінації супутникових та наземних вимірювань температури поверхні Землі і охоплює період з 1880 по 2023 рр.

Основні характеристики наведено у таблиці 2.1.

Таблиця 2.1 – Характеристика набору даних Global Surface Temperatures

Параметр	Опис
Джерело	Kaggle
Автор	Sujay Karadnis (2023)
Період спостережень	1880–2023 рр.
Частота вимірювань	Щомісячна
Одиниці вимірювання	°C (аномалія відносно базового періоду 1951–1980)
Географічне покриття	Глобальне (середні температури за всією поверхнею Землі)
Формат даних	CSV
Розмір файлу	≈ 120 тис. записів
Оновлення	Щорічно

Завдяки щомісячній роздільній здатності й широкому часовому охопленню, даний набір даних є репрезентативною основою для аналізу довгострокових кліматичних тенденцій, виявлення екстремальних років і моделювання трендів.

Супутникові дані забезпечують високоточну, систематичну і глобально узгоджену основу для аналізу кліматичних процесів. Вони дозволяють оцінювати зміни температури, хмарності, альбедо та інших параметрів із просторовою деталізацією, недоступною для традиційних метеостанцій.

Використання набору Global Surface Temperatures створює ефективну основу для виявлення довгострокових трендів і формування моделі глобального потепління.

2.2 Засоби обробки кліматичних даних

Обробка великих кліматичних масивів вимагає використання потужних програмних інструментів, здатних працювати з часовими рядами, геопросторовими сітками та статистичними моделями.

Сучасні дослідники користуються низкою аналітичних платформ – R, Python, MATLAB та QGIS, які забезпечують можливість аналізу, моделювання і візуалізації кліматичних даних.

Кожна з них має свої переваги, обмеження та сфери застосування (табл. 2.2).

Таблиця 2.2 – Порівняльна характеристика програмних платформ для аналізу кліматичних даних

Платформа	Основне призначення	Переваги	Недоліки
R	Статистичний аналіз, моделювання часових рядів, робота з просторовими даними	Відкритий код; велика кількість бібліотек для клімату (climate4R, raster, trend, ggplot2); висока якість графіків; активна наукова спільнота	Порівняно низька швидкодія при обробці надвеликих даних; потреба в програмних навичках
Python	Загальне програмування, машинне навчання, обробка даних	Потужні бібліотеки (NumPy, Pandas, SciPy, xarray, matplotlib); інтеграція з API супутникових сервісів; зручність автоматизації	Менш спеціалізований для статистики кліматичних часових рядів; складність відтворення наукових графіків у порівнянні з R
MATLAB	Наукові розрахунки, чисельне моделювання	Зручний у використанні інтерфейс; висока точність чисельних обчислень; наявність модулів для обробки кліматичних сіток	Комерційна ліцензія; обмежена кількість відкритих бібліотек; закрите середовище
QGIS	Геоінформаційна система для картографії	Інтерактивна робота з картами; інтеграція супутникових шарів (Sentinel, MODIS); візуалізація результатів	Обмежені можливості статистичного аналізу; низька продуктивність при великих часових рядах

Наведене порівняння демонструє, що кожна платформа має певну спеціалізацію. QGIS оптимальна для просторової візуалізації, MATLAB – для чисельного моделювання, Python – для гнучкої автоматизації процесів, а R – для статистичного та часово-просторового аналізу кліматичних даних.

Саме тому вибір R як основного інструменту є обґрунтованим для наукового дослідження тенденцій зміни клімату. Її переваги у контексті кліматичних досліджень полягають у наступному:

- широкий набір спеціалізованих пакетів для кліматичного аналізу – climate4R, trend, raster, sp, tidyverse, ggplot2;
- висока якість графічного відображення результатів – можливість створення багатошарових карт і динамічних графіків температурних аномалій;

- інтеграція з геопросторовими системами – підтримка форматів GeoTIFF, NetCDF, shapefile, що дозволяє працювати із супутниковими шарами MODIS, Sentinel або Landsat;
- активна міжнародна наукова спільнота, яка створює і підтримує відкриті бібліотеки для аналізу клімату;
- відкритість та відтворюваність результатів – код можна легко поширювати та перевіряти іншими дослідниками, що відповідає принципам FAIR-науки (Findable, Accessible, Interoperable, Reusable).

Таблиця 2.3 – Приклад бібліотек R, корисних для кліматичного аналізу

Пакет	Основне призначення
raster	Робота з багатовимірними супутниковими сітками та температурними шарами
trend	Оцінка кліматичних тенденцій за тестами Манна–Кендалла та Сенса
climate4R	Комплексний фреймворк для аналізу кліматичних даних, розроблений в рамках європейського проєкту <i>EUPORIAS</i>
ggplot2	Візуалізація часових рядів, карт і графіків аномалій
tidyverse	Уніфіковані інструменти для обробки даних (фільтрація, агрегування, очищення)

Отже, завдяки цим особливостям мова R дозволяє не лише аналізувати дані, а й створювати автоматизовані аналітичні системи для виявлення кліматичних трендів, оцінки сезонних коливань і прогнозування температурних змін.

2.3 Інструментарій мови R для кліматичного аналізу

Мова програмування R має розвинену екосистему бібліотек, що охоплюють практично всі аспекти аналізу кліматичних даних – від базового опрацювання числових рядів до просторового моделювання та інтерактивної візуалізації.

Бібліотеки R забезпечують зручну інтеграцію статистичних, геоінформаційних і візуалізаційних можливостей, що робить цю платформу однією з найпопулярніших у наукових дослідженнях клімату.

2.3.1 Основні напрями аналітичного застосування R

Інструментарій R у кліматологічних дослідженнях умовно поділяється на кілька функціональних напрямів:

- обробка та підготовка даних – імпорт, очищення, фільтрація, нормалізація кліматичних часових рядів;
- геопросторовий аналіз – робота з растровими та векторними шарами супутникових даних;
- статистичне моделювання – визначення трендів, сезонних коливань і виявлення аномалій;
- візуалізація результатів – створення графіків, карт, інтерактивних панелей і веб-додатків;
- автоматизація аналітичних процесів – побудова скриптів для пакетного опрацювання великих масивів даних.

Завдяки цьому R дозволяє реалізувати повний цикл дослідження – від завантаження сирих супутникових даних до публікації результатів у вигляді інтерактивних кліматичних карт.

2.3.2 Огляд основних бібліотек R для кліматичного аналізу

Найважливіші пакети мови R, що використовуються у даній кваліфікаційній роботі, наведено у таблиці 2.3.

Таблиця 2.3 – Основні бібліотеки R для аналізу кліматичних даних

Пакет	Основне призначення	Приклади застосування
raster	Робота з растровими кліматичними шарами та багатовимірними супутниковими даними	Зчитування супутникових зображень у форматах GeoTIFF, NetCDF; обчислення середніх температурних аномалій
terra	Сучасна альтернатива пакету <i>raster</i> з підвищеною швидкістю обробки	Обробка великих кліматичних матриць; обчислення статистичних характеристик температури та опадів

Кінець таблиці 2.3

sf	Робота з просторовими об'єктами (точки, лінії, полігони) за стандартом <i>Simple Features</i>	Побудова картографічних меж континентів, візуалізація розподілу кліматичних зон
tidyverse	Уніфікований набір інструментів для обробки даних (пакети <i>dplyr</i> , <i>readr</i> , <i>tidyr</i>)	Імпорт даних із CSV, фільтрація та групування за роками, розрахунок середніх температур
trend	Оцінка напрямку та сили кліматичних трендів за статистичними тестами	Застосування тестів Манна–Кендалла, Сенса, перевірка значущості тренду глобального потепління
ggplot2	Створення статичних та динамічних графіків високої якості [6];	Побудова графіків середньорічної температури, теплових карт і часових трендів
leaflet	Візуалізація кліматичних даних на інтерактивних картах	Створення онлайн-мап із температурними шарами та маркерами аномалій
shiny	Розробка інтерактивних веб-додатків для аналітики	Побудова панелей моніторингу кліматичних показників у режимі реального часу

Отже, бібліотеки R утворюють потужну інтегровану екосистему, у якій аналітичні, просторові та візуальні задачі взаємопов'язані.

Наприклад, дані можна імпортувати та обробити за допомогою tidyverse, провести статистичний аналіз трендів за допомогою trend, відобразити просторово за допомогою sf і leaflet, а результат інтегрувати у веб-додаток через shiny.

2.3.3 Можливості R для просторового, часово-рядного та візуального аналізу

Просторовий аналіз. R дозволяє працювати з геопросторовими даними різних форматів (GeoTIFF, shapefile, NetCDF), виконувати ресемплінг, обрізання шарів, зональну статистику та виявлення просторових аномалій. Пакети raster, terra та sf забезпечують поєднання супутникових зображень із кліматичними показниками, що дозволяє оцінювати вплив географічних чинників на зміну температури.

Часово-рядний аналіз. R має широкий набір інструментів для роботи з часовими рядами – від базових функцій до спеціалізованих пакетів (trend, zoo, xts). Це дозволяє виявляти довгострокові тенденції, циклічність, сезонність і аномальні значення у кліматичних даних. Зокрема, тест Манна–Кендалла та метод Сенса є найпоширенішими для визначення напрямку та швидкості кліматичних трендів.

Візуальний аналіз. Мова R пропонує широкий вибір інструментів для графічного відображення кліматичних процесів. Пакети ggplot2, leaflet та shiny дозволяють створювати як статичні графіки, так і інтерактивні карти й панелі моніторингу. Завдяки цьому можна представити складні просторово-часові процеси у зручному для сприйняття форматі.

2.4 Методика аналізу кліматичних тенденцій

Для дослідження тенденцій глобального потепління на основі супутникових температурних даних необхідно розробити чітку методику, що охоплює всі етапи – від завантаження вихідних спостережень до отримання аналітичних висновків.

Методика реалізована засобами мови R та передбачає автоматизовану обробку великих масивів кліматичної інформації з подальшою візуалізацією просторово-часових закономірностей.

Процес аналізу кліматичних тенденцій структуровано у п'ять основних етапів:

- імпорт даних – завантаження супутникових спостережень з відкритих джерел (Kaggle, NASA, NOAA) та приведення їх до уніфікованого формату CSV або NetCDF;
- попередня обробка – очищення даних від пропусків, дублювань, аномальних значень, формування часових рядів середньомісячних або середньорічних температур;
- трендовий аналіз – визначення напрямку й швидкості зміни температури за допомогою статистичних методів, зокрема тесту Манна–Кендалла, оцінки Сенса, ковзних середніх тощо;

- візуалізація результатів – побудова графіків, теплових карт та просторових діаграм з використанням пакетів ggplot2, leaflet та shiny;
- формування висновків – узагальнення результатів, оцінка достовірності трендів і просторових закономірностей зміни температури.

Таблиця 2.4 – Основні етапи методики аналізу кліматичних тенденцій

Етап	Опис	Основні інструменти
1. Імпорт даних	Отримання кліматичних даних із джерела <i>Global Surface Temperatures</i> (Kaggle), перевірка цілісності, конвертація у формат CSV.	readr, tidyverse, data.table
2. Попередня обробка	Видалення пропусків, згладжування шумів, групування за роками, нормалізація даних для подальшого аналізу.	dplyr, tidyr, lubridate
3. Трендовий аналіз	Визначення напрямку та сили кліматичних трендів за допомогою тестів Манна–Кендалла і Сенса; побудова ковзних середніх.	trend, zoo, stats
4. Просторовий аналіз	Географічна візуалізація зміни температури за координатами, поєднання з полігонами континентів.	sf, terra, raster, leaflet
5. Візуалізація результатів	Побудова графіків і теплових карт, створення інтерактивної панелі для аналізу динаміки кліматичних змін.	ggplot2, shiny, plotly
6. Формування висновків	Оцінка швидкості глобального потепління, виділення аномальних років, формування підсумкових показників.	dplyr, knitr, rmarkdown

Кожен етап логічно пов'язаний з попереднім – результати попередньої обробки використовуються для побудови трендів, які, у свою чергу, візуалізуються на інтерактивних картах.

Висновки до розділу 2

У цьому розділі було проведено комплексний аналіз сучасних інформаційних технологій, інструментів і засобів, що використовуються для опрацювання

супутникових кліматичних даних. Визначено, що супутникові спостереження є одним із найнадійніших джерел інформації для дослідження тенденцій зміни клімату, оскільки забезпечують глобальне покриття, тривалі часові ряди та високу точність вимірювань.

Проаналізовано структуру набору даних Global Surface Temperatures, який використовується як основне джерело температурних спостережень. Визначено, що його формат і обсяг дозволяють здійснювати детальний статистичний і просторово-часовий аналіз із можливістю побудови трендів і прогнозів.

Виконано порівняльний аналіз програмних платформ, що застосовуються для обробки кліматичних даних – R, Python, MATLAB, QGIS. Показано, що серед цих інструментів мова R є найбільш придатною для реалізації дослідження завдяки поєднанню засобів статистичного моделювання, геообробки та розвинених бібліотек візуалізації.

Розглянуто інструментарій мови R, зокрема пакети raster, terra, tidyverse, ggplot2, sf, leaflet, trend, shiny, які забезпечують повний цикл аналітичної обробки – від імпорту даних до побудови інтерактивних карт і графіків. Визначено, що використання цих бібліотек дозволяє створити модульну систему аналізу кліматичних процесів із високим рівнем автоматизації.

Розроблено методику аналізу кліматичних тенденцій, що включає етапи імпорту, очищення, статистичного аналізу, виявлення трендів, прогнозування та візуалізації результатів.

3 РОЗРОБКА ІНФОРМАЦІЙНОЇ СИСТЕМИ

3.1 Структура інформаційної системи

Інформаційна система аналізу тенденцій зміни клімату розроблена як модульний програмний комплекс, що забезпечує повний цикл опрацювання кліматичних даних – від їх завантаження та очищення до побудови статистичних оцінок, графічних візуалізацій, трендового аналізу та прогнозування. Такий підхід дозволяє гнучко розширювати функціонал системи, інтегрувати нові методи аналізу та забезпечувати відтворюваність отриманих результатів.

Архітектура системи побудована за принципом поділу на функціональні модулі, кожен з яких відповідає за окремий етап обробки даних (рис. 3.1). Основними компонентами інформаційної системи є:

- модуль завантаження даних забезпечує імпорт вхідних кліматичних наборів з локального середовища. Реалізований за допомогою бібліотек tidyverse, що дозволяє швидко зчитувати CSV-файли, перевіряти структуру таблиць та формувати базові описові характеристики;
- модуль попередньої обробки даних відповідає за очищення часових рядів, виявлення та інтерполяцію пропущених значень, перетворення форматів дат, формування узгодженого місячного ряду аномалій. На цьому етапі створюється головний робочий набір даних, що слугує основою для подальших аналітичних операцій;
- модуль аналітичної обробки включає методи розвідувального аналізу даних: обчислення статистичних показників, формування агрегованих рядів (річні, десятилітні), побудову гістограм, boxplot-діаграм, кривих ковзних середніх та сезонних профілів. У межах цього модуля реалізовані також засоби STL-декомпозиції з виділенням трендового та сезонного компонентів;
- модуль трендового аналізу містить інструменти оцінювання кліматичних тенденцій: лінійну регресію, непараметричний тест Манна–Кендалла, робастну оцінку Sen's slope та інші методи визначення напрямку і швидкості змін

температурних аномалій;

– модуль прогнозування призначений для побудови прогнозних моделей часу: ARIMA, ETS, нейронних мереж (NNAR), моделей згладжування та комбінованих прогнозів. Система дозволяє оцінити точність моделей та порівняти їх між собою;

– модуль візуалізації результатів використовує бібліотеку ggplot2 для створення наочних графічних представлень результатів дослідження: часові ряди, ковзні середні, діаграми розподілу, трендові графіки, сезонні профілі та прогнозні криві. Усі графіки можуть бути збережені у окремий каталог plots/ та використані у звітах;

– модуль експорту та збереження даних дозволяє зберігати проміжні та фінальні набори даних у форматах CSV або RDS, а також формувати таблиці для подальшого включення у текст кваліфікаційної роботи.



Рисунок 3.1 – Структура інформаційної системи

Подібна структура робить інформаційну систему універсальним інструментом для дослідження кліматичних тенденцій та побудови науково обґрунтованих прогнозів.

3.2 Підключення та конфігурація програмних бібліотек

Для реалізації інформаційної системи було підключено низку програмних бібліотек, що забезпечують завантаження даних, попередню обробку, статистичний аналіз, візуалізацію та прогнозування часових рядів. Завантаження бібліотек здійснювалося на початку роботи (рис. 3.2), що дало змогу сформувати узгоджене аналітичне середовище та забезпечити доступність усіх інструментів у межах одного програмного проєкту.

```
> install.packages(c(  
+   "tidyverse",    # базова обробка даних + ggplot2  
+   "lubridate",    # робота з датами  
+   "forecast",     # ARIMA, ETS, NNAR  
+   "zoo",           # інтерполяція пропусків  
+   "tseries",      # ADF-тест  
+   "trend",        # тест Манна-Кендалла  
+   "tsibble",      # сучасна робота з часовими рядами  
+   "fable",        # моделі прогнозування  
+   "fabletools"    # допоміжні інструменти для fable  
+ ))
```

Рисунок 3.2 – Підключення програмних бібліотек

У роботі використовуються засоби для імпорту та трансформації даних, такі як `tidyverse`, `readr`, `dplyr`, `tidyr` та `stringr`, які надають можливість швидкого читання CSV-файлів, очищення таблиць, фільтрації, агрегації, створення нових ознак і підготовки даних до подальшого аналізу. Для роботи з датами і часовими інтервалами застосовано `lubridate`, що забезпечує коректне перетворення форматів дат, виділення компонентів (рік, місяць) та зручні операції з часовими послідовностями.

Для статистичного аналізу та роботи з часовими рядами підключено бібліотеки `zoo`, `forecast`, `tseries`, `trend` та `fable`, які дозволяють виконувати декомпозицію рядів, обчислення ковзних середніх, перевірку стаціонарності, реалізацію лінійної регресії, проведення тесту Манна-Кендалла, обчислення

робастної оцінки Sen's slope, а також моделювання часових рядів методами ARIMA, ETS та нейронної авторегресії (NNAR).

Для візуалізації даних використано ggplot2, що входить до складу tidyverse та дозволяє будувати багаторівневі графіки, включаючи лінійні діаграми, boxplot-розподіли, сезонні профілі, комбінації тренду й фактичних значень, а також результати STL-декомпозиції. Побудовані графіки використовуються як інструмент для виявлення динамічних закономірностей у кліматичних даних.

Пакети forecast, fable та tseries також забезпечують засоби для прогнозування температурних аномалій, включаючи автоматичний підбір параметрів моделей, побудову прогнозів, формування довірчих інтервалів та оцінювання точності прогнозування за метриками MSE, RMSE та MAE.

Централізоване підключення бібліотек на початку проєкту забезпечує відтворюваність результатів дослідження, узгодженість програмного коду та можливість подальшого розширення системи за рахунок підключення нових інструментів. Така конфігурація створює стабільну та функціонально повну базу для виконання всіх етапів аналізу кліматичних даних.

3.3 Створення структури програмного проєкту

У межах розробки інформаційної системи важливим етапом є правильна організація структури проєкту та налаштування механізмів завантаження вхідних кліматичних даних. Для цього в межах програмного середовища RStudio було створено окрему файлову структуру (рис. 3.3), що включає каталоги data/ для зберігання наборів даних, scripts/ для програмних модулів, plots/ для збереження графічних результатів та output/ для експорту аналітичних матеріалів.

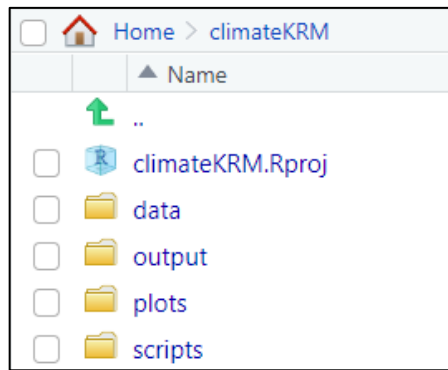


Рисунок 3.3 – Файлова структура створеного проєкту

Як вхідні дані використано відкритий набір Global Surface Temperatures, отриманий з платформи Kaggle, який містить чотири CSV-файли: `global_temps.csv`, `nh_temps.csv`, `sh_temps.csv` та `zonann_temps.csv`. Кожен із цих файлів містить дані температурних аномалій поверхні Землі за різні часові та просторові інтервали: глобальні показники, значення для Північної та Південної півкулі, а також зональні та річні температури.

Для завантаження CSV-файлів у систему використовувались засоби пакету `tidyverse`, зокрема функція `read_csv()`, яка забезпечує швидке зчитування структурованих даних (рис. 3.4).

```
> library(tidyverse)
> library(lubridate)
>
> global_temps <- read_csv(
+   file = "data/global_temps.csv"
+ )
Rows: 144 Columns: 19
— Column specification —————
Delimiter: ","
dbl (19): Year, Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct,...

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Рисунок 3.4 – Завантаження CSV-файлів у систему

Організована таким чином структура зберігання даних є основою для подальших етапів опрацювання, статистичного аналізу та побудови прогнозних моделей у межах інформаційної системи.

3.4 Первинний аналіз структури кліматичних даних

Після завантаження вхідних файлів було проведено первинний структурний аналіз усіх чотирьох таблиць, що входять до набору Global Surface Temperatures. На цьому етапі здійснено перевірку формату змінних, відповідності очікуваному обсягу даних, а також наявності потенційних пропусків. Огляд виконано за допомогою функцій `glimpse()` та `head()`, які забезпечують швидкий перегляд структури та перших рядків кожної таблиці.

Основна таблиця `global_temps` (рис. 3.5) містить 144 рядки (період 1880–2023 рр.) і 19 колонок, серед яких 12 змінних відповідають місячним температурним аномаліям, а додаткові поля включають сезонні та річні усереднення (J–D, D–N, DJF, MAM, JJA, SON). Усі значення представлені у числовому форматі, а колонка `Year` забезпечує можливість однозначної прив'язки кожного запису до календарного року.

```
> glimpse(global_temps)
Rows: 144
Columns: 19
$ Year    <dbl> 1880, 1881, 1882, 1883, 1884, 1885, 1886, 1887, 1888, 1889, 1890, 1891, 1892, 1893, 1894, 1895, 1896, 1897, 1...
$ Jan     <dbl> -0.19, -0.20, 0.16, -0.30, -0.13, -0.59, -0.44, -0.72, -0.34, -0.09, -0.42, -0.33, -0.29, -0.81, -0.53, -0.43,...
$ Feb     <dbl> -0.25, -0.15, 0.13, -0.37, -0.09, -0.34, -0.51, -0.57, -0.36, 0.16, -0.45, -0.47, -0.11, -0.57, -0.31, -0.43,...
$ Mar     <dbl> -0.09, 0.03, 0.04, -0.13, -0.37, -0.27, -0.43, -0.36, -0.41, 0.06, -0.40, -0.18, -0.40, -0.22, -0.23, -0.34, ...
$ Apr     <dbl> -0.17, 0.05, -0.16, -0.19, -0.40, -0.42, -0.28, -0.35, -0.20, 0.10, -0.30, -0.27, -0.33, -0.27, -0.45, -0.22,...
$ May     <dbl> -0.10, 0.05, -0.14, -0.18, -0.34, -0.45, -0.24, -0.31, -0.22, -0.01, -0.39, -0.16, -0.23, -0.33, -0.32, -0.28,...
$ Jun     <dbl> -0.21, -0.19, -0.22, -0.08, -0.35, -0.44, -0.35, -0.25, -0.17, -0.10, -0.25, -0.20, -0.22, -0.25, -0.41, -0.2...
$ Jul     <dbl> -0.18, 0.00, -0.17, -0.08, -0.31, -0.34, -0.18, -0.26, -0.11, -0.08, -0.28, -0.17, -0.31, -0.14, -0.25, -0.18...
$ Aug     <dbl> -0.11, -0.04, -0.08, -0.14, -0.28, -0.32, -0.31, -0.36, -0.16, -0.20, -0.39, -0.17, -0.27, -0.25, -0.25, -0.1...
$ Sep     <dbl> -0.15, -0.16, -0.15, -0.23, -0.28, -0.29, -0.24, -0.26, -0.12, -0.24, -0.37, -0.15, -0.17, -0.22, -0.29, -0.1...
$ Oct     <dbl> -0.24, -0.22, -0.24, -0.12, -0.25, -0.24, -0.28, -0.36, 0.01, -0.25, -0.25, -0.22, -0.14, -0.19, -0.24, -0.10...
$ Nov     <dbl> -0.22, -0.19, -0.17, -0.24, -0.34, -0.24, -0.28, -0.27, 0.03, -0.33, -0.43, -0.31, -0.42, -0.19, -0.27, -0.17...
$ Dec     <dbl> -0.18, -0.08, -0.36, -0.11, -0.31, -0.11, -0.26, -0.33, -0.04, -0.29, -0.31, -0.04, -0.38, -0.34, -0.22, -0.1...
$ `J-D`   <dbl> -0.17, -0.09, -0.11, -0.18, -0.29, -0.34, -0.32, -0.37, -0.17, -0.11, -0.35, -0.22, -0.27, -0.32, -0.31, -0.2...
$ `D-N`   <dbl> NA, -0.10, -0.09, -0.20, -0.27, -0.35, -0.30, -0.36, -0.20, -0.09, -0.35, -0.25, -0.25, -0.32, -0.32, -0.24, ...
$ DJF     <dbl> NA, -0.18, 0.07, -0.34, -0.11, -0.41, -0.35, -0.52, -0.34, 0.01, -0.39, -0.37, -0.15, -0.59, -0.39, -0.36, -0...
$ MAM     <dbl> -0.12, 0.04, -0.09, -0.17, -0.37, -0.38, -0.32, -0.34, -0.28, 0.05, -0.36, -0.20, -0.32, -0.27, -0.34, -0.28,...
$ JJA     <dbl> -0.17, -0.08, -0.16, -0.10, -0.32, -0.36, -0.28, -0.29, -0.14, -0.13, -0.30, -0.18, -0.27, -0.21, -0.30, -0.1...
$ SON     <dbl> -0.20, -0.19, -0.19, -0.20, -0.29, -0.25, -0.27, -0.29, -0.03, -0.28, -0.35, -0.23, -0.24, -0.20, -0.26, -0.1...
```

```
> head(global_temps)
# A tibble: 6 × 19
  Year   Jan   Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov   Dec `J-D` `D-N` DJF   MAM   JJA   SON
  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 1880 -0.19 -0.25 -0.09 -0.17 -0.1 -0.21 -0.18 -0.11 -0.15 -0.24 -0.22 -0.18 -0.17 NA    NA    -0.12 -0.17 -0.2
2 1881 -0.2 -0.15 0.03 0.05 0.05 -0.19 0    -0.04 -0.16 -0.22 -0.19 -0.08 -0.09 -0.1 -0.18 0.04 -0.08 -0.19
3 1882 0.16 0.13 0.04 -0.16 -0.14 -0.22 -0.17 -0.08 -0.15 -0.24 -0.17 -0.36 -0.11 -0.09 0.07 -0.09 -0.16 -0.19
4 1883 -0.3 -0.37 -0.13 -0.19 -0.18 -0.08 -0.08 -0.14 -0.23 -0.12 -0.24 -0.11 -0.18 -0.2 -0.34 -0.17 -0.1 -0.2
5 1884 -0.13 -0.09 -0.37 -0.4 -0.34 -0.35 -0.31 -0.28 -0.28 -0.25 -0.34 -0.31 -0.29 -0.27 -0.11 -0.37 -0.32 -0.29
6 1885 -0.59 -0.34 -0.27 -0.42 -0.45 -0.44 -0.34 -0.32 -0.29 -0.24 -0.24 -0.11 -0.34 -0.35 -0.41 -0.38 -0.36 -0.25
```

Рисунок 3.5 – Структура та перші рядки таблиці `global_temps`

Таблиця `nh_temps` (рис. 3.6) має аналогічну структуру, однак містить дані лише для Північної півкулі. У ній також представлено повний набір місячних

аномалій та сезонні середні значення, що дозволяє проводити порівняльний аналіз між півкулями. Структура таблиці sh_temps (рис. 3.7) повністю відповідає nh_temps, але відображає температурні аномалії для Південної півкулі. Первинний огляд цих двох таблиць підтвердив їх повну структурну відповідність глобальному набору, а відмінність полягає лише у географічній області охоплення.

```
> glimpse(nh_temps)
Rows: 144
Columns: 19
$ Year <dbl> 1880, 1881, 1882, 1883, 1884, 1885, 1886, 1887, 1888, 1889, 1890, 1891, 1892, 1893, 1894, 1895, 1896, 1897, 1898, 1899, ...
$ Jan <dbl> -0.37, -0.32, 0.24, -0.59, -0.18, -1.02, -0.76, -1.09, -0.50, -0.29, -0.64, -0.46, -0.50, -1.52, -0.92, -0.71, -0.36, -0. ...
$ Feb <dbl> -0.52, -0.24, 0.20, -0.67, -0.12, -0.47, -0.85, -0.71, -0.62, 0.28, -0.59, -0.69, -0.17, -0.98, -0.43, -0.81, -0.18, -0. ...
$ Mar <dbl> -0.24, -0.05, 0.01, -0.16, -0.65, -0.25, -0.72, -0.45, -0.64, -0.02, -0.48, -0.16, -0.67, -0.32, -0.21, -0.50, -0.49, -0. ...
$ Apr <dbl> -0.31, -0.01, -0.31, -0.31, -0.60, -0.50, -0.38, -0.39, -0.22, 0.16, -0.25, -0.35, -0.45, -0.30, -0.54, -0.19, -0.60, -0. ...
$ May <dbl> -0.07, 0.02, -0.24, -0.26, -0.37, -0.59, -0.34, -0.26, -0.16, -0.04, -0.36, -0.15, -0.23, -0.43, -0.26, -0.30, -0.16, 0. ...
$ Jun <dbl> -0.17, -0.34, -0.29, -0.12, -0.42, -0.45, -0.38, -0.21, -0.03, -0.07, -0.16, -0.12, -0.05, -0.19, -0.41, -0.21, 0.08, -0. ...
$ Jul <dbl> -0.19, 0.07, -0.29, -0.06, -0.42, -0.35, -0.14, -0.25, -0.01, -0.08, -0.23, -0.12, -0.38, -0.14, -0.21, -0.19, -0.01, 0. ...
$ Aug <dbl> -0.27, -0.05, -0.17, -0.23, -0.52, -0.42, -0.43, -0.55, -0.22, -0.21, -0.39, -0.13, -0.33, -0.34, -0.27, -0.28, -0.11, ...
$ Sep <dbl> -0.24, -0.27, -0.27, -0.35, -0.46, -0.41, -0.34, -0.21, -0.20, -0.31, -0.36, -0.08, -0.21, -0.29, -0.33, -0.06, -0.13, ...
$ Oct <dbl> -0.33, -0.45, -0.53, -0.17, -0.45, -0.37, -0.32, -0.49, -0.04, -0.42, -0.10, -0.28, -0.28, -0.29, -0.28, -0.09, 0.04, -0. ...
$ Nov <dbl> -0.44, -0.38, -0.35, -0.45, -0.58, -0.39, -0.40, -0.28, -0.02, -0.62, -0.63, -0.53, -0.83, -0.34, -0.49, -0.12, -0.20, ...
$ Dec <dbl> -0.41, -0.25, -0.70, -0.16, -0.48, -0.12, -0.22, -0.44, -0.24, -0.55, -0.47, -0.03, -0.65, -0.62, -0.36, -0.22, -0.21, ...
$ `J-D` <dbl> -0.30, -0.19, -0.22, -0.30, -0.44, -0.44, -0.44, -0.45, -0.24, -0.18, -0.39, -0.26, -0.40, -0.48, -0.39, -0.31, -0.19, ...
$ `D-N` <dbl> NA, -0.20, -0.19, -0.34, -0.41, -0.47, -0.43, -0.43, -0.26, -0.15, -0.40, -0.30, -0.35, -0.48, -0.41, -0.32, -0.20, -0. ...
$ DJF <dbl> NA, -0.32, 0.06, -0.65, -0.15, -0.66, -0.58, -0.67, -0.52, -0.08, -0.59, -0.54, -0.24, -1.05, -0.65, -0.63, -0.26, -0.3 ...
$ MAM <dbl> -0.20, -0.01, -0.18, -0.24, -0.54, -0.44, -0.48, -0.37, -0.34, 0.03, -0.36, -0.22, -0.45, -0.35, -0.34, -0.33, -0.42, -0. ...
$ JJA <dbl> -0.21, -0.11, -0.25, -0.14, -0.45, -0.40, -0.32, -0.34, -0.08, -0.12, -0.26, -0.12, -0.25, -0.22, -0.29, -0.23, -0.01, ...
$ SON <dbl> -0.34, -0.37, -0.38, -0.33, -0.50, -0.39, -0.35, -0.33, -0.09, -0.45, -0.36, -0.30, -0.44, -0.31, -0.36, -0.09, -0.10, ...
```

```
> head(nh_temps)
# A tibble: 6 × 19
  Year Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec `J-D` `D-N` DJF MAM JJA SON
<dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 1880 -0.37 -0.52 -0.24 -0.31 -0.07 -0.17 -0.19 -0.27 -0.24 -0.33 -0.44 -0.41 -0.3 NA NA -0.2 -0.21 -0.34
2 1881 -0.32 -0.24 -0.05 -0.01 0.02 0.34 0.07 -0.05 -0.27 -0.45 -0.38 -0.25 -0.19 -0.2 -0.32 -0.01 -0.11 -0.37
3 1882 0.24 0.2 0.01 -0.31 -0.24 -0.29 -0.29 -0.17 -0.27 -0.53 -0.35 -0.7 -0.22 -0.19 0.06 -0.18 -0.25 -0.38
4 1883 -0.59 -0.67 -0.16 -0.31 -0.26 -0.12 -0.06 -0.23 -0.35 -0.17 -0.45 -0.16 -0.3 -0.34 -0.65 -0.24 -0.14 -0.33
5 1884 -0.18 -0.12 -0.65 -0.6 -0.37 -0.42 -0.42 -0.52 -0.46 -0.45 -0.58 -0.48 -0.44 -0.41 -0.15 -0.54 -0.45 -0.5
6 1885 -1.02 -0.47 -0.25 -0.5 -0.59 -0.45 -0.35 -0.42 -0.41 -0.37 -0.39 -0.12 -0.44 -0.47 -0.66 -0.44 -0.4 -0.39
```

Рисунок 3.6 – Структура та перші рядки таблиці nh_temps

```
> glimpse(sh_temps)
Rows: 144
Columns: 19
$ Year <dbl> 1880, 1881, 1882, 1883, 1884, 1885, 1886, 1887, 1888, 1889, 1890, 1891, 1892, 1893, 1894, 1895, 1896, 1897, 1898, 1899, ...
$ Jan <dbl> 0.00, -0.09, 0.07, -0.02, -0.09, -0.20, -0.14, -0.37, -0.20, 0.08, -0.20, -0.21, -0.09, -0.13, -0.15, -0.15, -0.08, 0.1...
$ Feb <dbl> 0.03, -0.06, 0.07, -0.08, -0.06, -0.23, -0.20, -0.44, -0.12, 0.03, -0.32, -0.25, -0.05, -0.17, -0.18, -0.06, -0.07, 0.0...
$ Mar <dbl> 0.06, 0.09, 0.07, -0.09, -0.10, -0.30, -0.17, -0.27, -0.21, 0.13, -0.33, -0.20, -0.14, -0.12, -0.24, -0.18, -0.03, 0.12...
$ Apr <dbl> -0.02, 0.10, -0.02, -0.07, -0.22, -0.36, -0.20, -0.31, -0.19, 0.03, -0.34, -0.19, -0.21, -0.24, -0.36, -0.26, -0.01, 0.0...
$ May <dbl> -0.13, 0.08, -0.04, -0.10, -0.32, -0.33, -0.15, -0.35, -0.29, 0.02, -0.43, -0.17, -0.23, -0.23, -0.37, -0.26, -0.15, -0.0...
$ Jun <dbl> -0.25, -0.06, -0.15, -0.02, -0.30, -0.43, -0.32, -0.28, -0.31, -0.13, -0.33, -0.28, -0.38, -0.32, -0.41, -0.22, -0.29, ...
$ Jul <dbl> -0.17, -0.08, -0.05, -0.09, -0.22, -0.34, -0.22, -0.27, -0.20, -0.07, -0.32, -0.23, -0.25, -0.14, -0.28, -0.17, -0.04, ...
$ Aug <dbl> 0.06, -0.03, 0.01, -0.05, -0.07, -0.23, -0.19, -0.17, -0.10, -0.20, -0.39, -0.22, -0.21, -0.17, -0.23, -0.09, 0.01, -0.0...
$ Sep <dbl> -0.05, -0.05, -0.05, -0.11, -0.11, -0.19, -0.15, -0.30, -0.04, -0.17, -0.38, -0.23, -0.13, -0.16, -0.25, -0.18, -0.01, ...
$ Oct <dbl> -0.15, -0.01, 0.03, -0.06, -0.07, -0.13, -0.24, -0.22, 0.06, -0.10, -0.39, -0.16, 0.00, -0.09, -0.19, -0.11, 0.09, -0.0...
$ Nov <dbl> 0.00, 0.00, -0.01, -0.04, -0.11, -0.11, -0.15, -0.25, 0.07, -0.06, -0.25, -0.10, -0.03, -0.04, -0.06, -0.20, 0.12, -0.0...
$ Dec <dbl> 0.05, 0.09, -0.07, -0.06, -0.15, -0.09, -0.29, -0.23, 0.15, -0.04, -0.16, -0.05, -0.11, -0.06, -0.09, -0.05, 0.13, -0.1...
$ `J-D` <dbl> -0.05, 0.00, -0.01, -0.07, -0.15, -0.24, -0.20, -0.29, -0.12, -0.04, -0.32, -0.19, -0.15, -0.15, -0.23, -0.16, -0.03, ...
$ `D-N` <dbl> NA, 0.00, 0.00, -0.07, -0.14, -0.25, -0.19, -0.29, -0.15, -0.02, -0.31, -0.20, -0.15, -0.16, -0.23, -0.16, -0.04, -0.03...
$ DJF <dbl> NA, -0.03, 0.08, -0.06, -0.07, -0.19, -0.15, -0.36, -0.19, 0.09, -0.19, -0.21, -0.06, -0.14, -0.13, -0.10, -0.07, 0.10...
$ MAM <dbl> -0.03, 0.09, 0.00, -0.09, -0.21, -0.33, -0.17, -0.31, -0.23, 0.06, -0.37, -0.19, -0.19, -0.33, -0.23, -0.06, -0.0...
$ JJA <dbl> -0.12, -0.05, -0.06, -0.05, -0.19, -0.33, -0.24, -0.24, -0.20, -0.13, -0.35, -0.24, -0.28, -0.21, -0.31, -0.16, -0.11, ...
$ SON <dbl> -0.07, -0.02, -0.01, -0.07, -0.10, -0.14, -0.18, -0.26, 0.03, -0.11, -0.34, -0.16, -0.05, -0.10, -0.16, -0.16, 0.07, -0.0...
```

```
> head(sh_temps)
# A tibble: 6 × 19
  Year Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec `J-D` `D-N` DJF MAM JJA SON
<dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 1880 0 0.03 0.06 -0.02 -0.13 -0.25 -0.17 0.06 -0.05 -0.15 0 0.05 -0.05 NA NA -0.03 -0.12 -0.07
2 1881 -0.09 -0.06 0.09 0.1 0.08 -0.06 -0.08 -0.03 -0.05 -0.01 0 0.09 0 0 -0.03 0.09 -0.05 -0.02
3 1882 0.07 0.07 0.07 -0.02 -0.04 -0.15 -0.05 0.01 -0.05 0.03 -0.01 -0.07 -0.01 0 0.08 0 -0.06 -0.01
4 1883 -0.02 -0.08 -0.09 -0.07 -0.1 -0.02 -0.09 -0.05 -0.11 -0.06 -0.04 -0.06 -0.07 -0.07 -0.06 -0.09 -0.05 -0.07
5 1884 -0.09 -0.06 -0.1 -0.22 -0.32 -0.3 -0.22 -0.07 -0.11 -0.07 -0.11 -0.15 -0.15 -0.14 -0.07 -0.21 -0.19 -0.1
6 1885 -0.2 -0.23 -0.3 -0.36 -0.33 -0.43 -0.34 -0.23 -0.19 -0.13 -0.11 -0.09 -0.24 -0.25 -0.19 -0.33 -0.33 -0.14
```

Рисунок 3.7 – Структура та перші рядки таблиці sh_temps

Четверта таблиця `zonann_temps` (рис. 3.8) містить річні, а не місячні значення та має компактнішу структуру (15 колонок). Окрім глобального та півкульних рядів, вона включає зональні температурні аномалії для широтних поясів: від 90°N до 90°S з розбивкою на 24° сегменти. Така деталізація дозволяє проводити широтний аналіз кліматичних змін, що є важливою складовою досліджень регіональних кліматичних процесів. На відміну від трьох інших таблиць, цей набір не містить помісячних значень, але є фундаментальним для перевірки загальновісних трендів та просторового аналізу.

```
> glimpse(zonann_temps)
Rows: 143
Columns: 15
$ Year      <dbl> 1880, 1881, 1882, 1883, 1884, 1885, 1886, 1887, 1888, 1889, 1890, 1891, 1892, 1893, 1894, 1895, 1896, 1897, 1898, 1...
$ Glob      <dbl> -0.17, -0.09, -0.11, -0.18, -0.29, -0.34, -0.32, -0.37, -0.17, -0.11, -0.35, -0.22, -0.27, -0.32, -0.31, -0.23, -0.0...
$ NHem      <dbl> -0.30, -0.19, -0.22, -0.30, -0.44, -0.44, -0.44, -0.45, -0.24, -0.18, -0.39, -0.26, -0.40, -0.48, -0.39, -0.31, -0.0...
$ SHem      <dbl> -0.05, 0.00, -0.01, -0.07, -0.15, -0.24, -0.20, -0.29, -0.12, -0.04, -0.32, -0.19, -0.15, -0.15, -0.23, -0.16, -0.0...
$ `24N-90N` <dbl> -0.39, -0.37, -0.33, -0.37, -0.62, -0.67, -0.58, -0.54, -0.43, -0.30, -0.47, -0.43, -0.44, -0.52, -0.33, -0.41, -0.0...
$ `24S-24N` <dbl> -0.13, 0.10, -0.05, -0.17, -0.15, -0.13, -0.20, -0.30, 0.10, 0.08, -0.36, -0.07, -0.29, -0.37, -0.36, -0.09, 0.10, ...
$ `90S-24S` <dbl> -0.01, -0.07, 0.01, -0.01, -0.14, -0.27, -0.21, -0.28, -0.29, -0.16, -0.22, -0.21, -0.08, -0.05, -0.24, -0.26, -0.1...
$ `64N-90N` <dbl> -0.82, -0.92, -1.41, -0.19, -1.31, -1.51, -1.59, -1.77, -1.31, -0.69, -1.15, -1.13, -1.16, -0.70, -0.96, -0.70, -0.0...
$ `44N-64N` <dbl> -0.50, -0.47, -0.28, -0.57, -0.65, -0.69, -0.51, -0.51, -0.43, -0.23, -0.57, -0.33, -0.65, -0.54, -0.32, -0.53, -0.0...
$ `24N-44N` <dbl> -0.30, -0.23, -0.18, -0.29, -0.49, -0.48, -0.43, -0.32, -0.27, -0.29, -0.28, -0.38, -0.18, -0.48, -0.20, -0.28, -0.0...
$ `EQU-24N` <dbl> -0.15, 0.10, -0.05, -0.18, -0.12, -0.06, -0.21, -0.29, 0.07, 0.02, -0.26, 0.01, -0.32, -0.42, -0.49, -0.14, 0.06, 0.0...
$ `24S-EQU` <dbl> -0.11, 0.10, -0.05, -0.16, -0.17, -0.20, -0.18, -0.31, 0.13, 0.14, -0.47, -0.16, -0.26, -0.32, -0.23, -0.03, 0.14, ...
$ `44S-24S` <dbl> -0.04, -0.05, 0.01, -0.04, -0.19, -0.33, -0.20, -0.27, -0.31, -0.15, -0.17, -0.18, 0.03, 0.07, -0.14, -0.19, -0.15, ...
$ `64S-44S` <dbl> 0.05, -0.07, 0.04, 0.07, -0.02, -0.15, -0.20, -0.26, -0.22, -0.15, -0.27, -0.23, -0.21, -0.20, -0.35, -0.32, -0.13, ...
$ `90S-64S` <dbl> 0.66, 0.59, 0.62, 0.49, 0.64, 0.81, 0.58, 0.31, 0.15, 0.44, 0.14, -0.16, 0.08, -0.05, -0.38, -0.59, 0.25, 0.40, 0.3...
#> head(zonann_temps)
# A tibble: 6 × 15
  Year  Glob  NHem  SHem `24N-90N` `24S-24N` `90S-24S` `64N-90N` `44N-64N` `24N-44N` `EQU-24N` `24S-EQU` `44S-24S` `64S-44S` `90S-64S`
  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 1880 -0.17 -0.3 -0.05 -0.39 -0.13 -0.01 -0.82 -0.5 -0.3 -0.15 -0.11 -0.04 0.05 0.66
2 1881 -0.09 -0.19 0 -0.37 0.1 -0.07 -0.92 -0.47 -0.23 0.1 0.1 -0.05 -0.07 0.59
3 1882 -0.11 -0.22 -0.01 -0.33 -0.05 0.01 -1.41 -0.28 -0.18 -0.05 -0.05 0.01 0.04 0.62
4 1883 -0.18 -0.3 -0.07 -0.37 -0.17 -0.01 -0.19 -0.57 -0.29 -0.18 -0.16 -0.04 0.07 0.49
5 1884 -0.29 -0.44 -0.15 -0.62 -0.15 -0.14 -1.31 -0.65 -0.49 -0.12 -0.17 -0.19 -0.02 0.64
6 1885 -0.34 -0.44 -0.24 -0.67 -0.13 -0.27 -1.51 -0.69 -0.48 -0.06 -0.2 -0.33 -0.15 0.81
```

Рисунок 3.8 – Структура та перші рядки таблиці `zonann_temps`

Перегляд перших рядків кожної таблиці підтвердив коректність імпорту та відсутність структурних помилок, зокрема зсувів колонок або некоректних форматів значень. У всіх наборах даних виявлено незначні пропуски у сезонних та широтних колонках, що є очікуваною особливістю вихідного джерела й буде усунено на етапі попередньої обробки.

3.5 Попередня обробка кліматичних даних

Попередня обробка вихідних даних є ключовим етапом у формуванні якісного часового ряду, придатного для подальшого статистичного аналізу та побудови прогнозних моделей. На першому етапі було виконано перетворення

таблиці з місячними температурними аномаліями із «широкого» формату (рік + 12 місяців) у «довгий», що забезпечило створення окремого запису для кожної комбінації «рік–місяць». Для цього застосовувалась функція `pivot_longer()` пакета `tidyverse`, а формування календарної дати відбувалось із використанням функції `make_date()` із пакета `lubridate` (рис. 3.9). Сформована змінна `Date` дозволила впорядкувати дані хронологічно та забезпечила можливість застосування методів аналізу часових рядів.

```
> # Порядок місяців у вихідній таблиці
> month_levels <- c("Jan", "Feb", "Mar", "Apr", "May", "Jun",
+ "Jul", "Aug", "Sep", "Oct", "Nov", "Dec")
> # Перетворення широкого формату (рік + 12 місяців) у довгий формат "рік-місяць-аномалія"
> global_series <- global_temps %>%
+   select(Year, all_of(month_levels)) %>%
+   pivot_longer(
+     cols = -Year,
+     names_to = "Month",
+     values_to = "Temp_Anomaly"
+   ) %>%
+   mutate(
+     Month = factor(Month, levels = month_levels),
+     Month_Num = as.integer(Month),
+     Date = make_date(year = Year, month = Month_Num, day = 1)
+   ) %>%
+   arrange(Date)
> glimpse(global_series)
Rows: 1,728
Columns: 5
$ Year      <dbl> 1880, 1880, 1880, 1880, 1880, 1880, 1880, 1880, 1880, 1880, 1880, 1880, 1881, 1881, 1881, 1881, 1881, 1881, 1881...
$ Month     <fct> Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec, Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, No...
$ Temp_Anomaly <dbl> -0.19, -0.25, -0.09, -0.17, -0.10, -0.21, -0.18, -0.11, -0.15, -0.24, -0.22, -0.18, -0.20, -0.15, 0.03, 0.05, 0...
$ Month_Num <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11,...
$ Date      <date> 1880-01-01, 1880-02-01, 1880-03-01, 1880-04-01, 1880-05-01, 1880-06-01, 1880-07-01, 1880-08-01, 1880-09-01, 188...
```

Рисунок 3.9 – Перетворення широкого формату в довгий формат

На наступному етапі оцінено повноту даних за допомогою функцій `sum(is.na())`, `nrow()` та `is.na()`, що дозволило визначити кількість і частку пропущених значень (рис. 3.10). У глобальному ряді виявлено 7 пропусків (0,405% від вибірки), які потребували усунення для можливості подальшого аналізу.

```
> total_na <- sum(is.na(global_series$Temp_Anomaly))
> n_obs <- nrow(global_series)
> na_percent <- round(100 * total_na / n_obs, 3)
>
> total_na
[1] 7
> na_percent
[1] 0.405
```

Рисунок 3.10 – Перевірка кількості пропущених значень

Відновлення пропущених значень здійснювалося методом лінійної інтерполяції, що передбачає обчислення невідомих величин на основі двох сусідніх відомих точок. Математично інтерполяція визначається формулою:

$$x(t) = x(t_1) + \frac{x(t_2) - x(t_1)}{t_2 - t_1} (t - t_1), \quad (3.1)$$

де $x(t)$ – інтерпольоване значення показника у момент часу,

$x(t_1)$ – значення показника у попередній відомій точці,

$x(t_2)$ – значення показника у наступній відомій точці,

t_1 – момент часу попереднього відомого спостереження,

t_2 – момент часу наступного відомого спостереження,

t – момент часу, для якого потрібно відновити значення.

Для потенційних пропусків на межах ряду додатково застосовано метод перенесення останнього наявного значення (`na.locf()`), який гарантує коректне заповнення крайових точок.

Крім інтерполяції, під час попередньої обробки було сформовано допоміжні характеристики ряду, зокрема ковзні середні, що дозволяють згладити короткострокові коливання та виокремити довгострокову тенденцію. Ковзне середнє з вікном у k місяців визначається як:

$$MA_t = \frac{1}{k} \sum_{i=0}^{k-1} x_{t-i}, \quad (3.2)$$

де MA_t – ковзне середнє у момент часу,

x_{t-i} – значення показника у поточному та попередніх місяцях,

k – кількість спостережень у вікні згладжування,

t – номер поточного моменту часу.

Використання цього показника дозволяє зменшити стохастичні коливання та забезпечує наочну інтерпретацію тренду.

Після заповнення пропущених значень було проведено повторну перевірку і виявлено, що кількість пропущених значень у наборі – 0 (рис. 3.11).

```
> global_series_interpolated <- global_series %>%
+   mutate(
+     # Лінійна інтерполяція за часом
+     Temp_Anomaly_Interp = zoo::na.approx(
+       Temp_Anomaly,
+       x      = Date,
+       na.rm = FALSE
+     ),
+     Temp_Anomaly_Interp = zoo::na.locf(Temp_Anomaly_Interp, na.rm = FALSE),
+     Temp_Anomaly_Interp = zoo::na.locf(Temp_Anomaly_Interp, fromLast = TRUE)
+   )
> total_na_after <- sum(is.na(global_series_interpolated$Temp_Anomaly_Interp))
> total_na_after
[1] 0
```

Рисунок 3.11 – Перевірка кількості пропущених значень після здійснення інтерполяції

Після завершення всіх етапів обробки сформовано фінальний очищений часовий ряд `global_clean` (рис. 3.12), який містить дату, рік, назву місяця, порядковий номер місяця та інтерпольоване значення температурної аномалії.

```
> global_clean <- global_series_interpolated %>%
+   select(
+     Date,
+     Year,
+     Month,
+     Month_Num,
+     Temp_Anomaly = Temp_Anomaly_Interp
+   )
> glimpse(global_clean)
Rows: 1,728
Columns: 5
$ Date      <date> 1880-01-01, 1880-02-01, 1880-03-01, 1880-04-01, 1880-05-01, 1880-06-01, 1880-07-01, 1880-08-01, 1880-09-01, 188...
$ Year      <dbl> 1880, 1880, 1880, 1880, 1880, 1880, 1880, 1880, 1880, 1880, 1880, 1880, 1881, 1881, 1881, 1881, 1881, 1881, 1881...
$ Month     <fct> Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec, Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, No...
$ Month_Num <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, ...
$ Temp_Anomaly <dbl> -0.19, -0.25, -0.09, -0.17, -0.10, -0.21, -0.18, -0.11, -0.15, -0.24, -0.22, -0.18, -0.20, -0.15, 0.03, 0.05, 0.0...
```

Рисунок 3.12 – Структура фінального очищеного часового ряду `global_clean`

Цей набір даних використовується в подальшому для статистичного аналізу, сезонної декомпозиції, оцінювання трендів та побудови прогнозних моделей.

3.6 Аналіз глобальних температурних даних

3.6.1 Динаміка глобальних температурних аномалій

Первинний огляд часової динаміки температурних аномалій дозволяє визначити загальну структуру ряду, характер його коливань та наявність довгострокових змін. На рисунку 3.13 наведено повний помісячний ряд температурних аномалій за весь період спостережень, а також дві згладжені криві – ковзні середні з вікном 12 та 60 місяців.

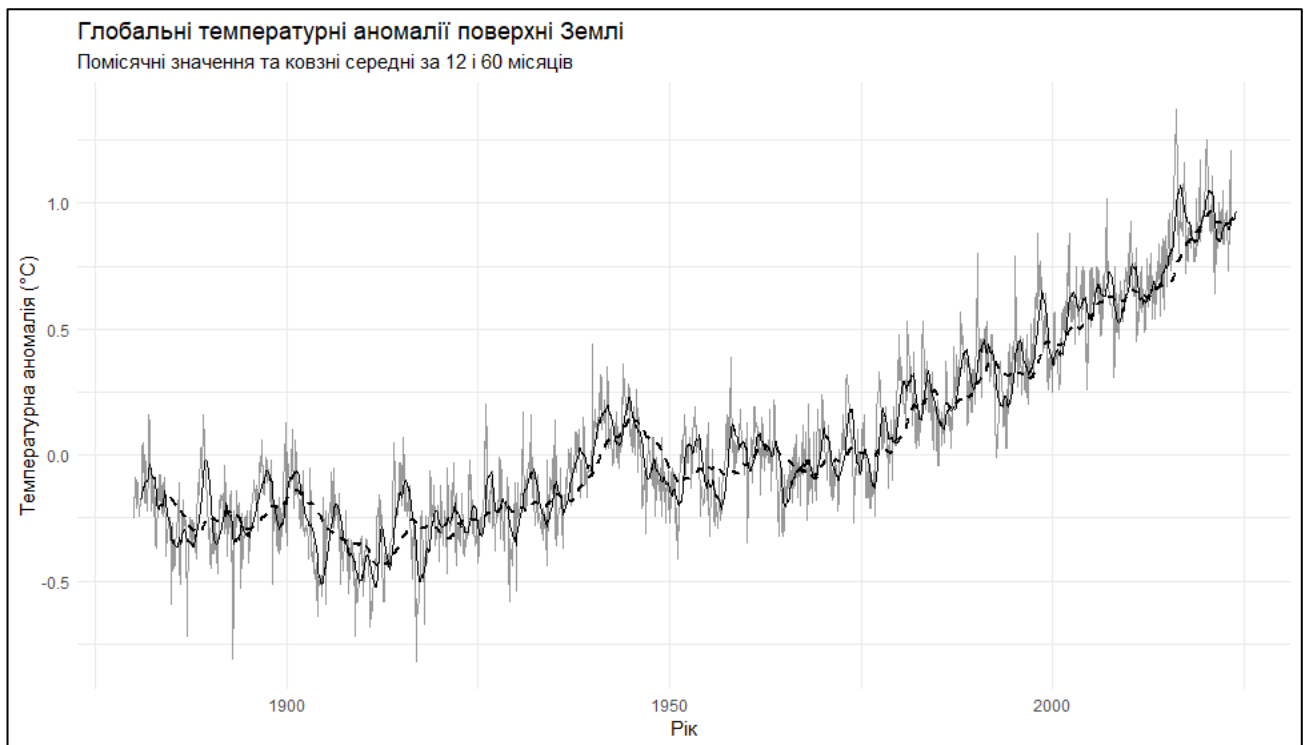


Рисунок 3.13 – Динаміка глобальних температурних аномалій із ковзними середніми за 12 та 60 місяців

На графіку зображено повний часовий ряд глобальних температурних аномалій за період 1880–2023 років. Тонка світла лінія відображає помісячні коливання температури, які характеризуються значною варіабельністю в короткостроковій перспективі. Дві згладжені криві – річне (12 місяців) та

п'ятирічне (60 місяців) ковзні середні – демонструють загальну тенденцію зміни клімату.

Графік чітко показує поступове зростання температурних аномалій упродовж XX століття з різким прискоренням після 1980 року. Починаючи з кінця XX – початку XXI століття, аномалії стабільно перебувають у позитивній зоні, що свідчить про систематичне й довготривале глобальне потепління. П'ятирічне ковзне середнє згладжує короткотермінові коливання і виокремлює стійкий висхідний тренд, який є характерною ознакою сучасних кліматичних змін.

Для детальнішого вивчення сучасного періоду було побудовано окремий фрагмент ряду для 1950–2023 років (рис. 3.14).

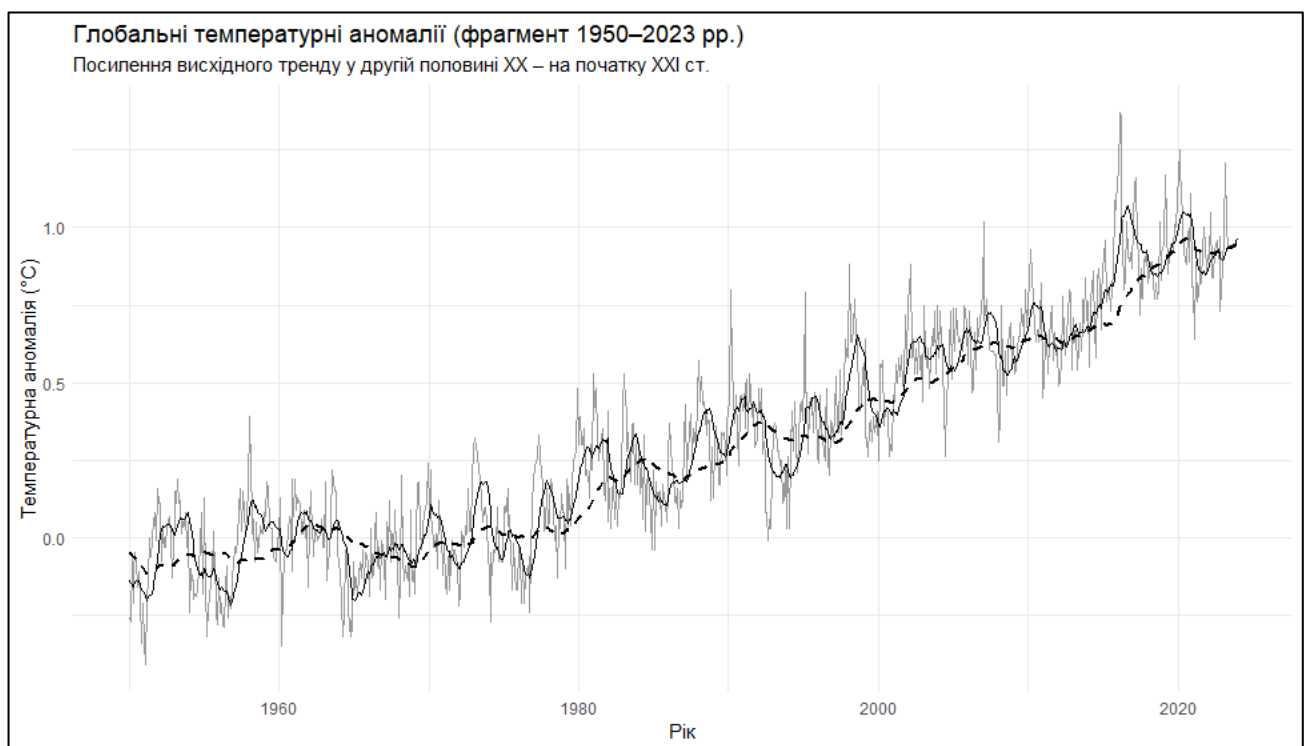


Рисунок 3.14 – Фрагмент глобальних температурних аномалій за 1950–2023 роки

Із графіка видно, що в 1950–1970-х роках температурні аномалії коливалися поблизу нульового рівня, інколи переходячи у від'ємні значення. Починаючи з 1980-х років, рівень аномалій систематично зростає, а після 2000 року майже всі значення стають стабільно позитивними. У 2010–2020-х роках спостерігаються

найвищі температурні відхилення за весь період спостережень, що є характерною ознакою прискореного глобального потепління.

Графік наочно підтверджує, що інтенсивність підвищення глобальної температури зростає саме в останні десятиліття, а сучасний кліматичний режим суттєво відрізняється від умов середини XX століття.

3.6.2 Розподіл глобальних температурних аномалій

Дослідження розподілу значень температурних аномалій дозволяє оцінити, як змінювався характер кліматичних відхилень протягом усього періоду спостережень. На рисунку 3.15 наведено гістограму та криву щільності розподілу.

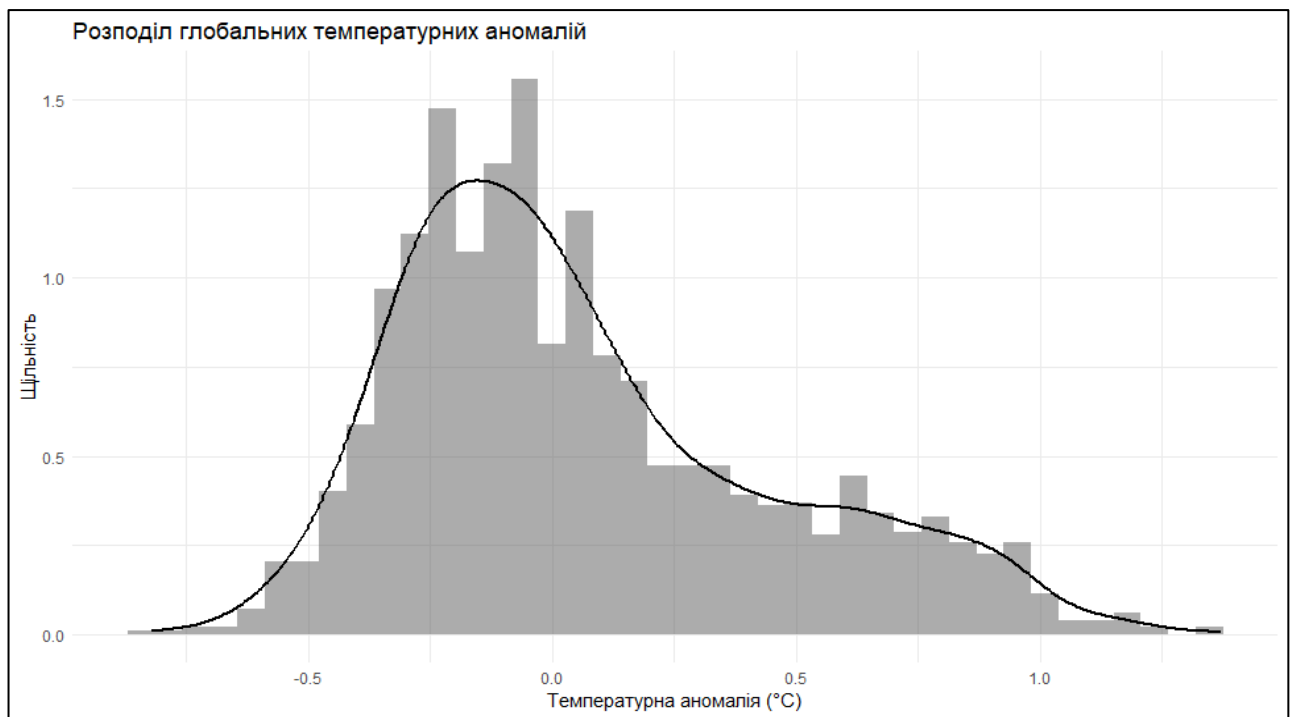


Рисунок 3.15 – Розподіл глобальних температурних аномалій

Графік демонструє, що більшість значень зосереджена поблизу нульової позначки або в діапазоні від -0.3 до $+0.2$ °C, що відображає типові кліматичні умови кінця XIX – першої половини XX століття.

Розподіл є асиметричним: спостерігається зміщення вправо та наявність «довгого хвоста» в зоні позитивних аномалій. Така форма розподілу вказує на те,

що у пізніші десятиліття частота та інтенсивність тепліших періодів зростала, що узгоджується з сучасними тенденціями глобального потепління. Наявність значної кількості значень у діапазоні $+0.5...+1.0$ °C свідчить про те, що підвищені температурні аномалії стали характерною рисою клімату XXI століття.

Таким чином, отриманий розподіл підтверджує систематичне зростання температурного фону та зміщення кліматичної системи у тепліший режим.

3.6.3 Температурні аномалії у розрізі десятиліть

Окреме уявлення про масштаб потепління дає аналіз розподілу температурних аномалій окремо для кожного десятиліття. Для цього було побудовано boxplot-діаграми, що відображають медіану, кватилі та екстремальні значення (рис. 3.16).

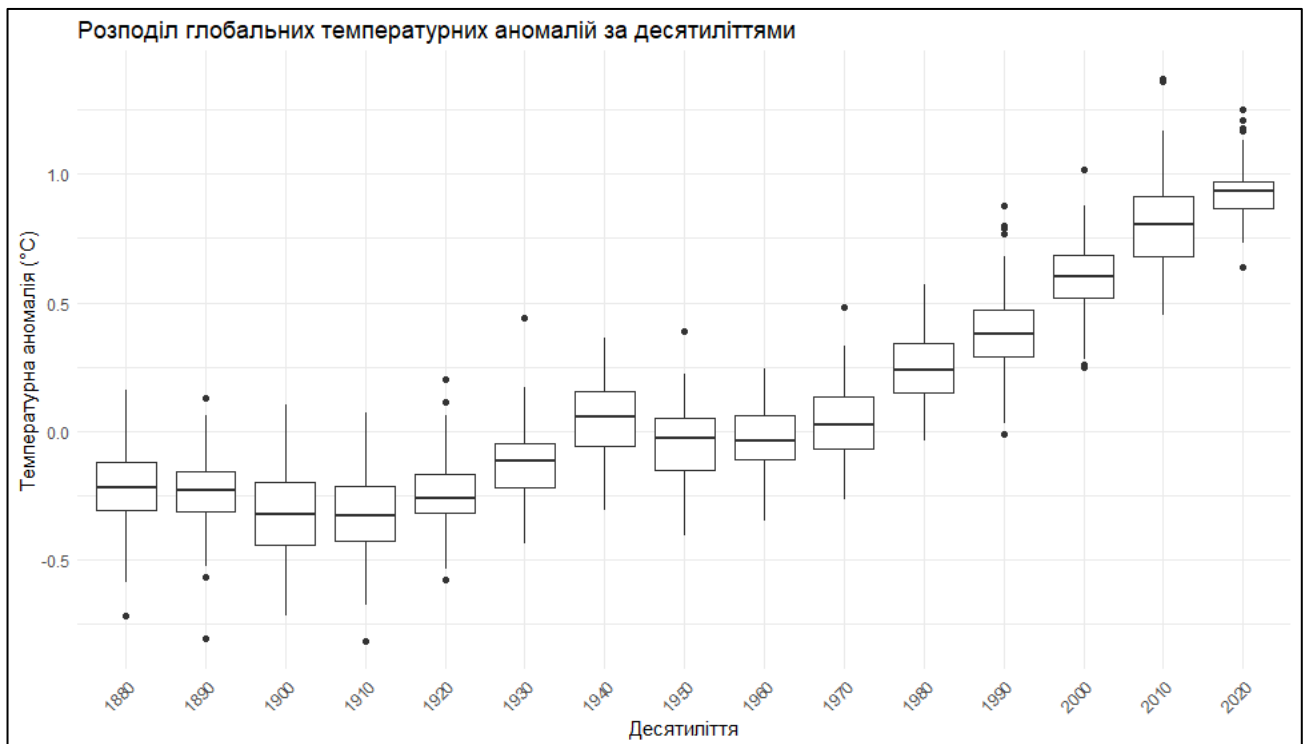


Рисунок 3.16 – Розподіл глобальних температурних аномалій за десятиліттями

Графік демонструє чітку та послідовну еволюцію кліматичної системи від відносно стабільних умов кінця XIX–початку XX століття до різкого підвищення температури в останні десятиліття.

У десятиліттях 1880–1940-х років медіана аномалій перебуває поблизу нульового рівня або має незначні від’ємні значення, а розмах коливань відносно симетричний. Починаючи з 1950-х років, медіана поступово зростає, а в 1980-х роках переходить у стійко позитивну зону. У 1990–2020-х роках зростання прискорюється: медіана, верхні квартилі та максимальні значення суттєво підвищуються, а розподіл значно зміщується вправо.

Особливо виразні зміни спостерігаються у 2010-х і 2020-х роках, де майже всі значення знаходяться вище $+0,5\text{ }^{\circ}\text{C}$, а окремі аномалії перевищують $+1,0\text{ }^{\circ}\text{C}$. Наявність численних позитивних «викидів» у новітніх десятиліттях також свідчить про збільшення частоти екстремально теплих місяців.

Таким чином, графік підтверджує суттєве прискорення глобального потепління в останні 40–50 років та демонструє перехід кліматичної системи у якісно новий, тепліший режим.

3.6.4 Сезонні особливості глобальних температурних аномалій

Дослідження сезонності дозволяє визначити, чи є підвищення температури рівномірним протягом року, чи воно проявляється лише у певні сезони. Для цього обчислено середні аномалії для кожного місяця за весь період спостережень (рис. 3.17).



Рисунок 3.17 – Середні глобальні температурні аномалії за місяцями

Графік показує, що сезонні коливання температурних аномалій є відносно невеликими: більшість значень знаходиться в діапазоні 0.04-0.09 °C. Найвищі аномалії спостерігаються у березні та жовтні–листопаді, тоді як найменші – у червні та травні. Це свідчить про те, що глобальне потепління проявляється не через різке збільшення сезонної амплітуди, а як загальне підвищення температурного фону протягом року.

Сезонний профіль є досить стабільним і має лише помірні коливання, що підтверджує: основним фактором сучасних кліматичних змін є довгостроковий тренд потепління, а сезонність відіграє другорядну роль.

3.6.5 Аналіз середніх річних глобальних температурних аномалій та визначення найтепліших років

Здійснимо агрегування помісячних значень до річних середніх температурних аномалій. Такий підхід дозволяє зменшити випадкові короткострокові коливання та отримати більш стабільний індикатор

довгострокових змін кліматичної системи. На основі річних значень було визначено топ-10 найтепліших років за увесь період спостережень (рис. 3.18).

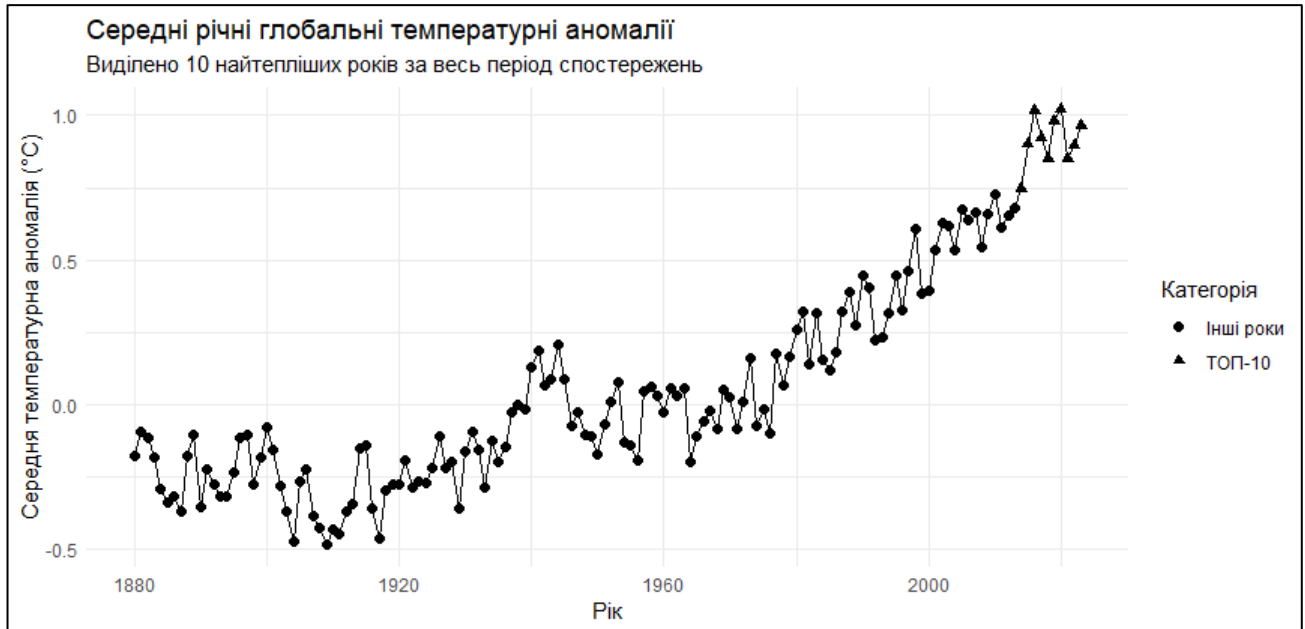


Рисунок 3.18 – Середні річні глобальні температурні аномалії з виділенням топ-10 найтепліших років

Усі роки, що входять до топ-10 найтепліших, зосереджені у часовому проміжку 2010–2023 рр., що свідчить про безпрецедентне прискорення глобального потепління у XXI столітті. Найвищі аномалії зафіксовано у 2016, 2020 та 2023 роках, де середня температура перевищує $+1.0\text{ }^{\circ}\text{C}$ відносно кліматичної норми.

Графік демонструє не лише висхідний тренд, але й поступове збільшення варіабельності у останні десятиліття, що може бути пов'язано з посиленням впливу природних кліматичних циклів (ENSO) на фоні загального антропогенного потепління.

3.6.6 STL-декомпозиція глобальних температурних аномалій

Для більш глибокого аналізу структури часового ряду було виконано STL-декомпозицію (Seasonal–Trend–Loess), яка дозволяє розділити помісячний

температурний ряд на три складові: $T(t)$ – довгостроковий тренд, $S(t)$ – сезонність, $R(t)$ – залишкова (випадкова) компонента.

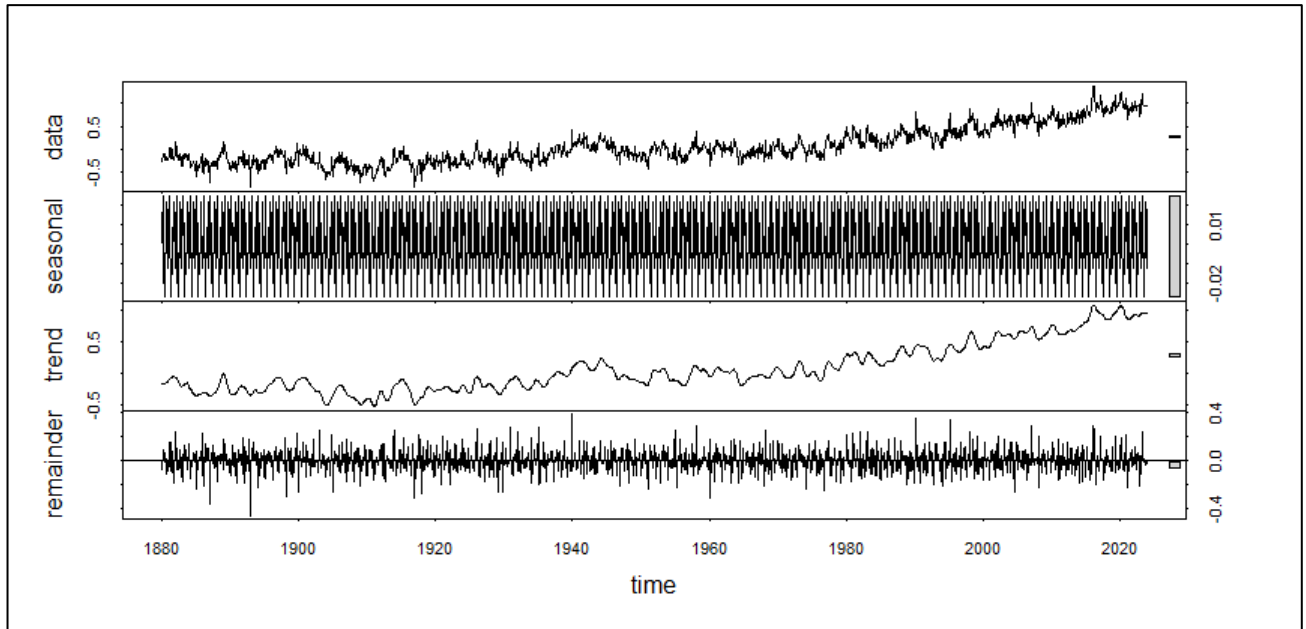


Рисунок 3.16 – STL-декомпозиція глобальних температурних аномалій

Метод дозволяє розкласти часовий ряд на три основні компоненти — вихідні дані (data), сезонну складову (seasonal), трендову складову (trend) та залишок (remainder).

1. Вихідні дані (data): у верхній панелі зображено вихідний часовий ряд помісячних аномалій. Видно значну короткострокову мінливість, особливо в ранні роки спостережень, та стійке підвищення середнього рівня у XXI столітті. Графік підтверджує результат попереднього аналізу: найбільші значення аномалій спостерігаються у останні два десятиліття.

2. Сезонна складова (seasonal): сезонний компонент демонструє регулярні річні коливання, які мають стабільну форму та амплітуду протягом усього періоду. Це свідчить про те, що глобальні зміни клімату не впливають суттєво на сезонність температурних аномалій — вона залишається практично незмінною. Амплітуда сезонного циклу є малою, що узгоджується з висновком про домінування довгострокового тренду над сезонними ефектами.

3. Трендова складова (trend): тренд є найважливішою частиною декомпозиції, оскільки відображає довгострокові зміни кліматичної системи. На графіку видно плавне зростання від кінця XIX століття з різким прискоренням у другій половині XX століття та особливо після 1980 року. У XXI столітті трендова крива досягає максимальних значень, що підтверджує системне й прискорене глобальне потепління.

4. Залишкова складова (remainder): залишок містить випадкові коливання, які не можуть бути пояснені трендом або сезонністю. У цій компоненті проявляються впливи короточасних кліматичних явищ, таких як вулканічні виверження, океанічні цикли ENSO (Ель-Ніньйо / Ла-Нінья), а також природний шум кліматичної системи. Характер коливань залишається приблизно однаковим упродовж усього ряду, що свідчить про відсутність системного зростання “шуму” у кліматичних даних.

STL-декомпозиція підтверджує, що головним фактором довгострокових змін глобальних температурних аномалій є стійкий висхідний тренд, тоді як сезонність залишається стабільною, а залишкова варіабельність не демонструє суттєвих змін у часі.

Висновки до розділу 3

У третьому розділі виконано розробку інформаційної системи аналізу тенденцій зміни клімату як модульного програмного комплексу в середовищі R. Запропонована архітектура орієнтована на повний цикл роботи з кліматичними даними: від імпорту вихідних таблиць до отримання аналітичних результатів, наочних візуалізацій, трендових оцінок і підготовки даних для подальшого моделювання та прогнозування.

Сформовано узгоджене програмне середовище шляхом підключення та конфігурації необхідних бібліотек і організації структури проєкту в RStudio. Створені каталоги для даних, скриптів, графіків і вихідних результатів забезпечили відтворюваність виконання обчислень, стандартизований підхід до збереження

проміжних та фінальних артефактів, а також спростили масштабування системи та підтримку коду.

На етапі первинного аналізу структури даних проведено огляд і валідацію всіх таблиць набору Global Surface Temperatures. Встановлено, що набори `global_temps`, `nh_temps` та `sh_temps` мають узгоджену структуру з місячними й сезонними показниками, тоді як `zonann_temps` містить річні значення та широтні пояси, що створює передумови для подальших порівнянь між півкулями та просторового аналізу. Перевірка структури підтвердила коректність імпорту та відсутність критичних структурних помилок, а виявлені незначні пропуски визначено як очікувану особливість вихідного джерела даних.

Реалізовано ключовий етап попередньої обробки: трансформацію даних у «довгий» формат із формуванням календарної змінної `Date`, перевірку повноти вибірки та усунення пропусків. Виявлено 7 пропущених значень (0,405% від вибірки), які відновлено методом лінійної інтерполяції з додатковим коректним заповненням крайових значень. У результаті сформовано фінальний очищений часовий ряд `global_clean`, придатний для застосування методів аналізу часових рядів, сезонної декомпозиції, трендового аналізу та прогнозування.

Виконано базовий аналіз глобальних температурних аномалій і побудовано комплекс графічних представлень, що описують динаміку, розподіл та структурні властивості даних. Отримані візуалізації з ковзними середніми підтвердили наявність стійкого висхідного тренду з прискоренням після 1980 року, аналіз у розрізі десятиліть продемонстрував зміщення розподілу в бік позитивних аномалій у кінці XX – на початку XXI століття, а сезонний профіль засвідчив відносно стабільну сезонність на тлі зростання середнього температурного фону. Агрегування до річних значень і визначення топ-10 найтепліших років показало їх концентрацію у періоді 2010-2023 рр., що є індикатором інтенсивного потепління в новітній період. Додатково STL-декомпозиція підтвердила домінування довгострокового тренду над сезонною складовою та відносну стабільність випадкової компоненти.

Таким чином, у розділі 3 створено працездатну основу інформаційної системи: налаштовано програмне середовище, організовано структуру проєкту, виконано імпорт, очищення та формування фінального набору даних і реалізовано ключові процедури первинного аналітичного дослідження. Отримані результати забезпечили підготовку даних і аналітичної бази для подальшого трендового аналізу та моделювання змін клімату в наступному розділі роботи.

4 ТРЕНДОВИЙ АНАЛІЗ ТА МОДЕЛЮВАННЯ ЗМІН КЛІМАТУ

4.1 Лінійна регресія тренду глобальних температурних аномалій

Для кількісної оцінки довгострокових змін глобальних температурних аномалій у роботі застосовано параметричний метод лінійної регресії. Даний підхід дозволяє визначити напрямок, інтенсивність та статистичну значущість тренду температурних змін у часі, а також оцінити, яка частка загальної варіації даних пояснюється лінійною тенденцією.

На першому етапі аналізу було сформовано річний часовий ряд середніх глобальних температурних аномалій (рис. 4.1). Для цього помісячні значення очищеного набору даних було агреговано за роками шляхом обчислення середнього значення температурної аномалії для кожного календарного року. Така агрегація дозволяє зменшити вплив короткострокових сезонних коливань і випадкових флуктуацій, зосередивши увагу на довгострокових кліматичних змінах.

```
> yearly_stats <- global_clean %>%  
+   group_by(Year) %>%  
+   summarise(  
+     Mean_Anomaly = mean(Temp_Anomaly, na.rm = TRUE),  
+     .groups = "drop"  
+   )  
> head(yearly_stats)  
# A tibble: 6 × 2  
  Year Mean_Anomaly  
  <dbl>      <dbl>  
1 1880    -0.174  
2 1881    -0.0917  
3 1882    -0.113  
4 1883    -0.181  
5 1884    -0.288  
6 1885    -0.338
```

Рисунок 4.1 – Формування річних середніх значень

На другому етапі до отриманого річного ряду було застосовано лінійну регресійну модель (рис. 4.2), у якій середня річна температурна аномалія

розглядається як залежна змінна, а календарний рік — як незалежна змінна. Таким чином, модель описує зміну температурних аномалій у часі та дозволяє оцінити середню швидкість зростання температури за одиницю часу.

Математично лінійний тренд описується рівнянням:

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t, \quad (4.1)$$

де y_t – середня річна глобальна температурна аномалія,

t – рік спостереження,

β_0 – вільний член,

β_1 – коефіцієнт тренду,

ε_e – випадкова похибка.

```
> trend_lm <- lm(Mean_Anomaly ~ Year, data = yearly_stats)
> summary(trend_lm)

Call:
lm(formula = Mean_Anomaly ~ Year, data = yearly_stats)

Residuals:
    Min       1Q   Median       3Q      Max
-0.36454 -0.13090 -0.02962  0.13248  0.43997

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.537e+01  7.084e-01  -21.7   <2e-16 ***
Year          7.910e-03  3.629e-04   21.8   <2e-16 ***
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.181 on 142 degrees of freedom
Multiple R-squared:  0.7699,    Adjusted R-squared:  0.7682
F-statistic:  475 on 1 and 142 DF,  p-value: < 2.2e-16
```

Рисунок 4.2 – Побудова лінійної регресійної моделі Mean_Anomaly ~ Year

У результаті оцінювання параметрів моделі отримано значення коефіцієнта тренду $\beta_1 = 0,00791$ °C на рік, що відповідає зростанню температури приблизно на 0,0791 °C за десятиліття. Це означає, що в середньому глобальна температура

поверхні Землі підвищується майже на одну десяту градуса кожні десять років. Статистична значущість коефіцієнта підтверджується надзвичайно малим значенням p-value ($p < 2,2 \cdot 10^{-16}$), що дозволяє відхилити нульову гіпотезу про відсутність тренду.

Якість побудованої моделі оцінювалася за допомогою коефіцієнта детермінації R^2 , який становить 0,7699 (рис. 4.3). Це свідчить про те, що близько 77 % варіації середніх річних температурних аномалій пояснюється лінійною залежністю від часу, тоді як решта варіації обумовлена міжрічними кліматичними коливаннями та випадковими факторами.

```
> trend_coef_year <- coef(trend_lm)["Year"]           # °C за 1 рік
> trend_coef_decade <- 10 * trend_coef_year          # °C за 10 років
>
> trend_coef_year
      Year
0.007910096
> trend_coef_decade
      Year
0.07910096
> model_summary <- summary(trend_lm)
>
> r2_value <- model_summary$r.squared
> p_value_year <- model_summary$coefficients["Year", "Pr(>|t|)"]
>
> r2_value
[1] 0.7698571
> p_value_year
[1] 3.826155e-47
```

Рисунок 4.3 – Обрахунок коефіцієнта детермінації R^2

Для наочної інтерпретації результатів побудовано графік, на якому відображено фактичні середні річні значення глобальних температурних аномалій разом із лінією регресійного тренду (рис. 4.4). Графік демонструє, що, незважаючи на помітну міжрічну варіабельність, загальний напрям зміни температури є чітко висхідним. Найбільші відхилення від лінії тренду спостерігаються у роки з аномальними кліматичними подіями, однак вони не змінюють загального характеру довгострокового потепління.

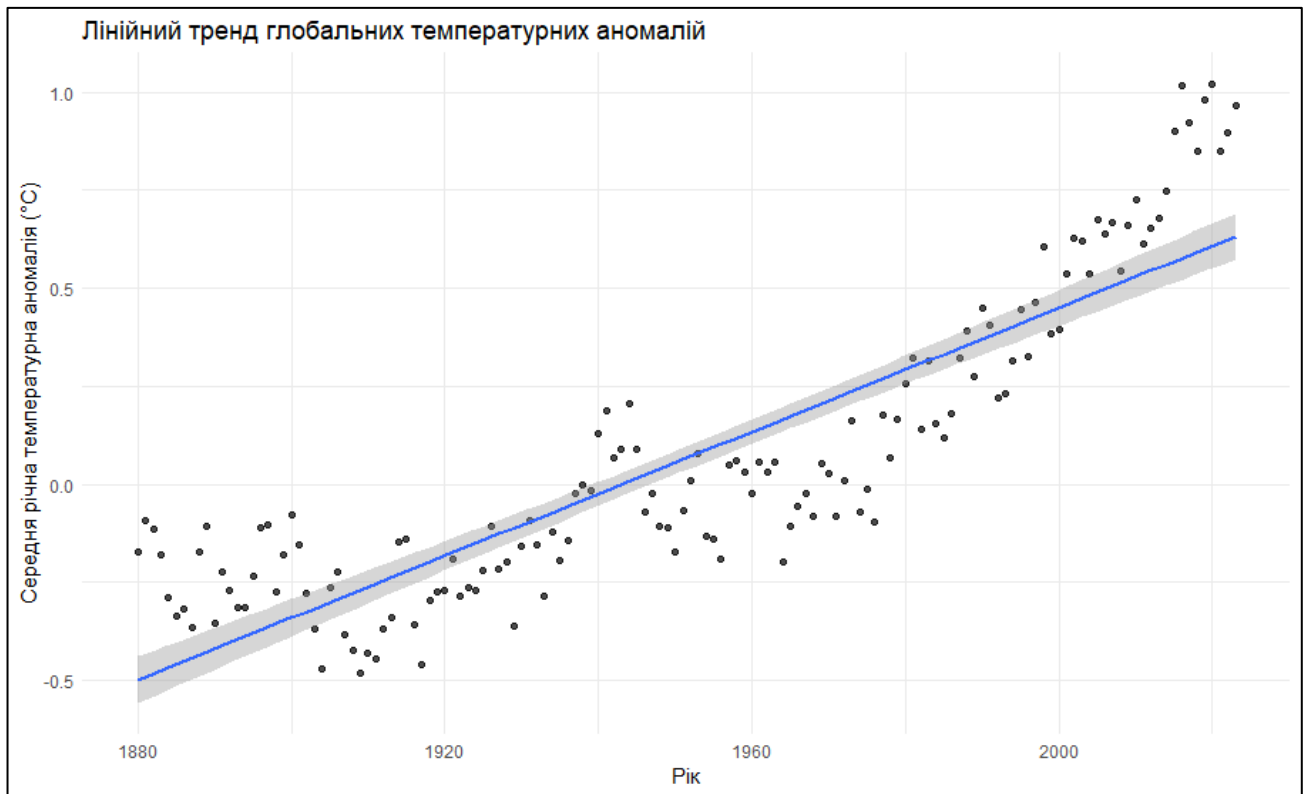


Рисунок 4.4 – Лінійний тренд глобальних температурних аномалій

Графік наочно демонструє стійкий висхідний характер температурних змін упродовж усього періоду спостережень. У кінці XIX – на початку XX століття значення середніх річних аномалій переважно перебували в зоні від’ємних значень, що відповідає холоднішому кліматичному режиму того часу. Починаючи з середини XX століття спостерігається поступове підвищення температури, яке суттєво прискорюється після 1980-х років.

Розсіювання точок навколо лінії тренду свідчить про наявність міжрічної варіабельності, зумовленої природними кліматичними коливаннями та короткостроковими аномальними подіями. Водночас вузький довірчий інтервал навколо лінії регресії підтверджує надійність оцінки тренду та його статистичну значущість.

Отримані результати лінійної регресії підтверджують наявність стійкого зростаючого тренду глобальних температурних аномалій і узгоджуються з висновками міжнародних кліматичних досліджень. Разом з тим, з огляду на

можливу наявність автокореляції та нелінійних ефектів у часовому ряді, доцільним є доповнення параметричного аналізу непараметричними методами, які не залежать від припущень щодо розподілу даних.

4.2 Непараметричні методи трендового аналізу глобальних температурних аномалій

Для підтвердження результатів параметричного трендового аналізу та зменшення впливу можливих порушень припущень лінійної регресії у роботі застосовано непараметричні методи аналізу часових рядів. Основною перевагою таких методів є відсутність вимог до нормальності розподілу даних, лінійності залежності та стійкість до викидів. У межах дослідження використано тест Манна-Кендалла та робастну оцінку швидкості тренду Sen's slope.

На першому етапі проведено тест Манна-Кендалла (рис. 4.5), який дозволяє перевірити наявність монотонного тренду у часовому ряді без припущень щодо його функціональної форми. Нульова гіпотеза тесту полягає у відсутності тренду, альтернативна – у наявності зростаючого або спадного тренду.

За результатами тесту отримано значення статистики $z = 12,809$ при кількості спостережень $n = 144$ та $p\text{-value} < 2,2 \cdot 10^{-16}$. Отримане значення $p\text{-value}$ є значно меншим за рівень значущості $0,05$, що дозволяє відхилити нульову гіпотезу та підтвердити наявність статистично значущого зростаючого тренду глобальних температурних аномалій. Додатково коефіцієнт рангової кореляції Кендалла становить $\tau = 0,7208$, що свідчить про сильний позитивний зв'язок між температурною аномалією та часом.

```
> mk_test <- mk.test(yearly_stats$Mean_Anomaly)
> mk_test

      Mann-Kendall trend test

data:  yearly_stats$Mean_Anomaly
z = 12.809, n = 144, p-value < 2.2e-16
alternative hypothesis: true S is not equal to 0
sample estimates:
             S             varS             tau
7.41700e+03  3.35180e+05  7.20797e-01
```

Рисунок 4.5 – Тест Манна-Кендалла

На другому етапі застосовано робастну оцінку швидкості тренду Sen's slope, яка визначає медіану всіх можливих кутових коефіцієнтів між парами точок часового ряду. Такий підхід є стійким до викидів і забезпечує надійну оцінку інтенсивності тренду.

Оцінка Sen's slope (рис. 4.6) становить 0,00787 °C/рік, а довірчий інтервал має вигляд [0,00700; 0,00869] °C/рік. Отримане значення практично збігається з оцінкою швидкості тренду, отриманою за допомогою лінійної регресії, що підтверджує стабільність та надійність результатів.

```
> mk_test <- mk.test(yearly_stats$Mean_Anomaly)
> mk_test

      Mann-Kendall trend test

data:  yearly_stats$Mean_Anomaly
z = 12.809, n = 144, p-value < 2.2e-16
alternative hypothesis: true S is not equal to 0
sample estimates:
             S             varS             tau
7.41700e+03  3.35180e+05  7.20797e-01
```

Рисунок 4.6 – Тест Sen's slope

Таким чином, застосування непараметричних методів трендового аналізу підтвердило наявність сильного, статистично значущого та монотонно

зростаючого тренду глобальних температурних аномалій. Узгодженість результатів параметричного та непараметричного підходів свідчить про те, що виявлена тенденція глобального потепління не є артефактом вибраної моделі, а відображає реальні довгострокові кліматичні зміни.

4.3 Порівняльний аналіз параметричних і непараметричних методів оцінювання тренду

Для підвищення достовірності висновків щодо наявності та інтенсивності кліматичних змін у роботі було застосовано як параметричні, так і непараметричні методи трендового аналізу. Такий підхід дозволяє порівняти результати, отримані за різних статистичних припущень, і підтвердити стабільність виявленої тенденції глобального потепління.

Параметричний аналіз ґрунтувався на використанні лінійної регресійної моделі, яка описує залежність середньої річної температурної аномалії від часу. За результатами регресії встановлено статистично значущий зростаючий тренд зі швидкістю $0,00791\text{ }^{\circ}\text{C}$ на рік ($0,0791\text{ }^{\circ}\text{C}$ за десятиліття) при $p\text{-value} < 2,2 \cdot 10^{-16}$ та коефіцієнті детермінації $R^2 = 0,7699$. Отримане значення R^2 свідчить про те, що майже 77 % варіації середніх річних температурних аномалій пояснюється лінійною часовою тенденцією.

Для перевірки стійкості отриманих результатів було застосовано непараметричні методи трендового аналізу, які не потребують припущень щодо розподілу даних і є менш чутливими до викидів. Тест Манна–Кендалла підтвердив наявність статистично значущого монотонного зростаючого тренду, що підтверджується надзвичайно малим значенням $p\text{-value}$ ($p = 1,45 \cdot 10^{-37}$). Це дозволяє впевнено відхилити нульову гіпотезу про відсутність тренду у часовому ряді.

Оцінка швидкості тренду за допомогою робастного методу Sen's slope становить $0,00787\text{ }^{\circ}\text{C}$ на рік із 95 % довірчим інтервалом $[0,00700; 0,00869]\text{ }^{\circ}\text{C/рік}$. Отримане значення практично збігається з результатом лінійної регресії, що

свідчить про узгодженість параметричних і непараметричних підходів та підтверджує реальний характер виявленого тренду.

Таблиця 4.1 – Порівняння результатів трендового аналізу глобальних температурних аномалій

Метод аналізу	Швидкість тренду	Статистична значущість	Особливості методу
Лінійна регресія	0,00791 °C/рік (0,0791 °C/десятиліття)	$p\text{-value} < 2,2 \cdot 10^{-16}$, $R^2 = 0,7699$	Параметричний метод, чутливий до розподілу даних
Тест Манна–Кендалла	Наявність зростаючого тренду	$p = 1,45 \cdot 10^{-37}$	Непараметричний тест, стійкий до викидів
Sen's slope	0,00787 °C/рік	95 % ДІ [0,00700; 0,00869]	Робастна оцінка швидкості тренду

Таким чином, використання різних методів трендового аналізу дозволило не лише виявити наявність довгострокового потепління, а й кількісно оцінити його інтенсивність. Узгоджені результати, отримані за допомогою незалежних статистичних підходів, значно підвищують надійність висновків щодо сучасних тенденцій зміни клімату.

4.4 Прогнозування глобальних температурних аномалій

Прогнозування кліматичних показників є важливим етапом аналізу часових рядів, оскільки дозволяє не лише описати історичні тенденції, але й оцінити можливий напрямок та інтенсивність майбутніх змін клімату. У межах даної кваліфікаційної роботи прогнозування глобальних температурних аномалій здійснюється на основі середніх річних значень, що забезпечує фокусування на довгострокових кліматичних процесах і зменшує вплив сезонних та короткострокових коливань.

Для побудови прогнозів використано кілька різних підходів до моделювання часових рядів, реалізованих у середовищі R. Застосування декількох моделей дозволяє порівняти результати, оцінити стабільність прогнозів та підвищити надійність висновків.

4.4.1 Прогнозування на основі моделі ARIMA

Модель ARIMA (AutoRegressive Integrated Moving Average) є одним із базових статистичних інструментів прогнозування часових рядів і широко застосовується у кліматичних дослідженнях для аналізу довгострокових тенденцій. Основною перевагою даної моделі є її здатність враховувати автокореляційну структуру ряду та ефективно описувати трендову складову після приведення ряду до стаціонарного вигляду.

У межах цієї роботи для прогнозування глобальних температурних аномалій використано середні річні значення, що дозволяє усунути сезонні коливання та зосередитися на повільних кліматичних процесах. Перед побудовою моделі часовий ряд було автоматично проаналізовано на стаціонарність, після чого виконано необхідне диференціювання. Вибір оптимальної структури моделі ARIMA (рис. 4.7) здійснено за допомогою функції `auto.arima()` мови програмування R, яка підбирає параметри моделі шляхом мінімізації інформаційного критерію Акаїке.

```
> # Перетворення річних даних у часовий ряд
> temp_ts <- ts(
+   yearly_stats$Mean_Anomaly,
+   start = min(yearly_stats$Year),
+   frequency = 1
+ )
> model_arima <- auto.arima(temp_ts)
> h <- 20
> forecast_arima <- forecast(model_arima, h = h)
> autoplot(forecast_arima) +
+   labs(
+     title = "Прогноз глобальних температурних аномалій (ARIMA)",
+     x = "Рік",
+     y = "Температурна аномалія (°C)"
+   )
```

Рисунок 4.7 – Побудова моделі ARIMA

Побудована модель адекватно відтворює історичну динаміку глобальних температурних аномалій і коректно враховує наявну автокореляцію між

послідовними спостереженнями. Це підтверджується відсутністю виражених систематичних закономірностей у залишках моделі та їхньою близькістю до білого шуму.

На основі отриманої ARIMA-моделі виконано прогнозування температурних аномалій на горизонт 20 років уперед (рис. 4.8). Прогноз демонструє подальше зростання середніх річних температурних аномалій, що узгоджується з результатами параметричного та непараметричного трендового аналізу, наведеними у попередніх підрозділах. Довірчі інтервали прогнозу поступово розширюються зі збільшенням горизонту прогнозування, що відображає зростання невизначеності, характерне для кліматичних часових рядів.



Рисунок 4.8 – Прогноз глобальних температурних аномалій (ARIMA)

Графік демонструє чітке збереження висхідної тенденції температурних аномалій у прогнозному періоді. Згідно з моделлю ARIMA, очікується подальше зростання середньої глобальної температурної аномалії, що узгоджується з результатами трендового аналізу, отриманими на основі історичних даних.

Відсутність різких зламів або стабілізації у прогнозній траєкторії свідчить про інерційний характер кліматичної системи та домінування довгострокового потепління.

Розширення довірчих інтервалів зі збільшенням горизонту прогнозування відображає зростання невизначеності оцінок, що є типовим для довгострокового прогнозування кліматичних часових рядів. Водночас навіть нижня межа 95 % довірчого інтервалу залишається у зоні додатних температурних аномалій, що вказує на високу ймовірність подальшого потепління.

4.4.2 Прогнозування методом експоненційного згладжування (ETS)

Методи експоненційного згладжування є класом адаптивних моделей прогнозування часових рядів, які ґрунтуються на принципі надання більшої ваги останнім спостереженням. Такий підхід є доцільним для кліматичних часових рядів, оскільки дозволяє оперативно враховувати поступові структурні зміни та довгострокові тенденції, характерні для процесів глобального потепління.

У межах даного дослідження для прогнозування середніх річних глобальних температурних аномалій використано модель ETS (Error, Trend, Seasonality), реалізовану в пакеті `forecast` мови програмування R (рис. 4.9 – рис. 4.10). Оскільки аналіз проводиться на річних даних, сезонна складова у моделі відсутня, а основну увагу зосереджено на описі рівня та тренду часового ряду. Вибір оптимальної структури моделі здійснюється автоматично шляхом мінімізації інформаційних критеріїв, що забезпечує об'єктивність побудови прогнозу.

```
> model_ets <- ets(temp_ts)
> h <- 20
> forecast_ets <- forecast(model_ets, h = h)
> autoplot(forecast_ets) +
+   labs(
+     title = "Прогноз глобальних температурних аномалій (ETS)",
+     x = "Рік",
+     y = "Температурна аномалія (°C)"
+   )
```

Рисунок 4.9 – Побудова моделі ETS

На відміну від моделей класу ARIMA, які базуються на автокореляційній структурі ряду, метод експоненційного згладжування безпосередньо моделює зміну рівня та тренду в часі. Це дозволяє отримати більш плавну прогнозу траєкторію, що добре відповідає інерційному характеру кліматичної системи. Згладжування короткострокових флуктуацій підсилює видимість довгострокового потепління та зменшує вплив міжрічної варіабельності.

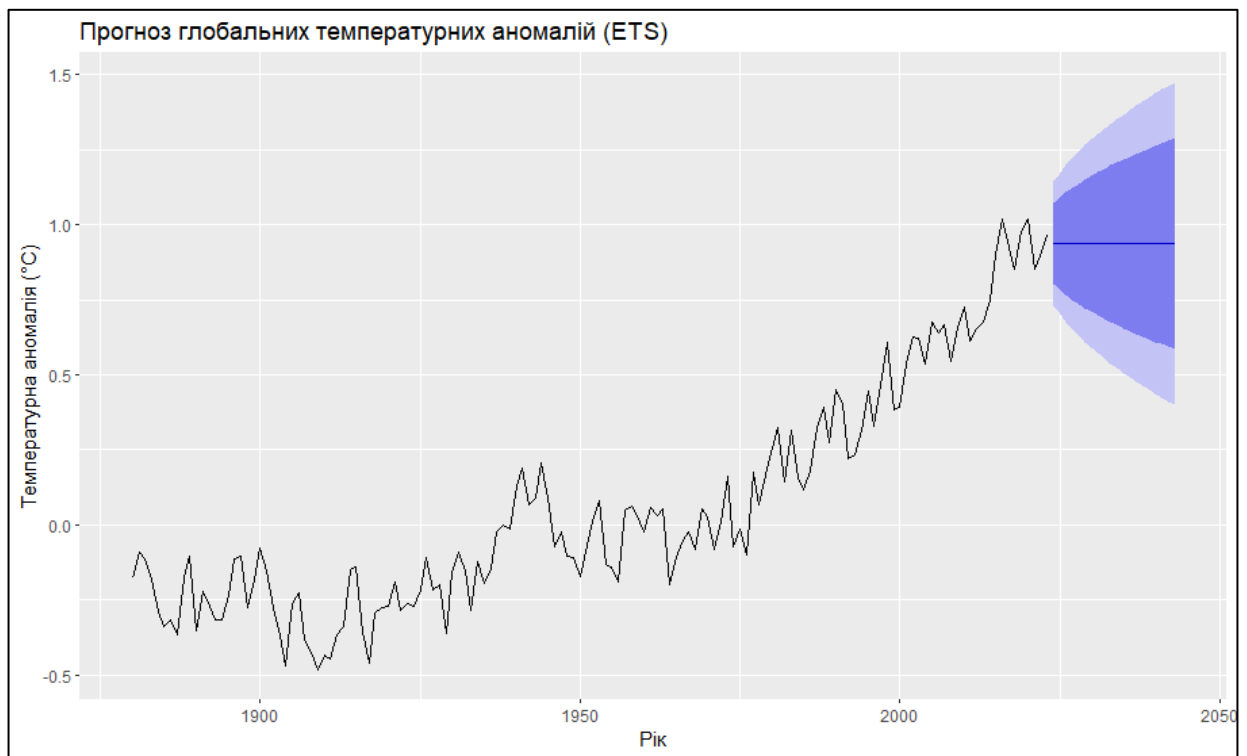


Рисунок 4.10 – Прогноз глобальних температурних аномалій (ETS)

Прогноз, отриманий за моделлю ETS, демонструє плавну та монотонно зростаючу траєкторію температурних аномалій, що свідчить про стабільний характер довгострокового потепління. На відміну від моделі ARIMA, експоненційне згладжування менш чутливе до короткострокових флуктуацій, унаслідок чого прогнозна крива має більш згладжений вигляд.

Зі збільшенням горизонту прогнозування спостерігається поступове розширення довірчих інтервалів, що відображає зростання невизначеності прогнозних оцінок. Водночас навіть нижня межа 95 % довірчого інтервалу

залишається у зоні додатних температурних аномалій, що підтверджує високу ймовірність подальшого зростання глобальної температури.

4.4.3 Прогнозування з використанням нейронних мереж (NNAR)

Нейронні мережі для аналізу часових рядів є сучасним інструментом прогнозування, який дозволяє враховувати складні та нелінійні залежності між спостереженнями. На відміну від класичних статистичних моделей, таких як ARIMA та ETS, нейромережеві підходи не потребують попереднього задання аналітичної форми тренду та можуть автоматично адаптуватися до структури даних.

У межах даної роботи для прогнозування середніх річних глобальних температурних аномалій застосовано модель NNAR (Neural Network AutoRegressive), реалізовану у пакеті forecast мови програмування R (рис. 4.11 – рис. 4.12). Дана модель використовує лагові значення часового ряду як вхідні змінні багатошарової нейронної мережі, що дозволяє відтворювати як лінійні, так і нелінійні закономірності кліматичної динаміки.

```
> model_nnar <- nnetar(temp_ts)
> h <- 20
> forecast_nnar <- forecast(model_nnar, h = h)
> autoplot(forecast_nnar) +
+   labs(
+     title = "Прогноз глобальних температурних аномалій (NNAR)",
+     x = "Рік",
+     y = "Температурна аномалія (°C)"
+   )
```

Рисунок 4.11 – Побудова моделі NNAR

Перед побудовою моделі часовий ряд було нормалізовано автоматичними засобами бібліотеки, після чого виконано навчання нейронної мережі на історичних даних. Процес навчання полягає в мінімізації помилки між фактичними та змодельованими значеннями температурних аномалій, що забезпечує оптимальне налаштування ваг нейронів.

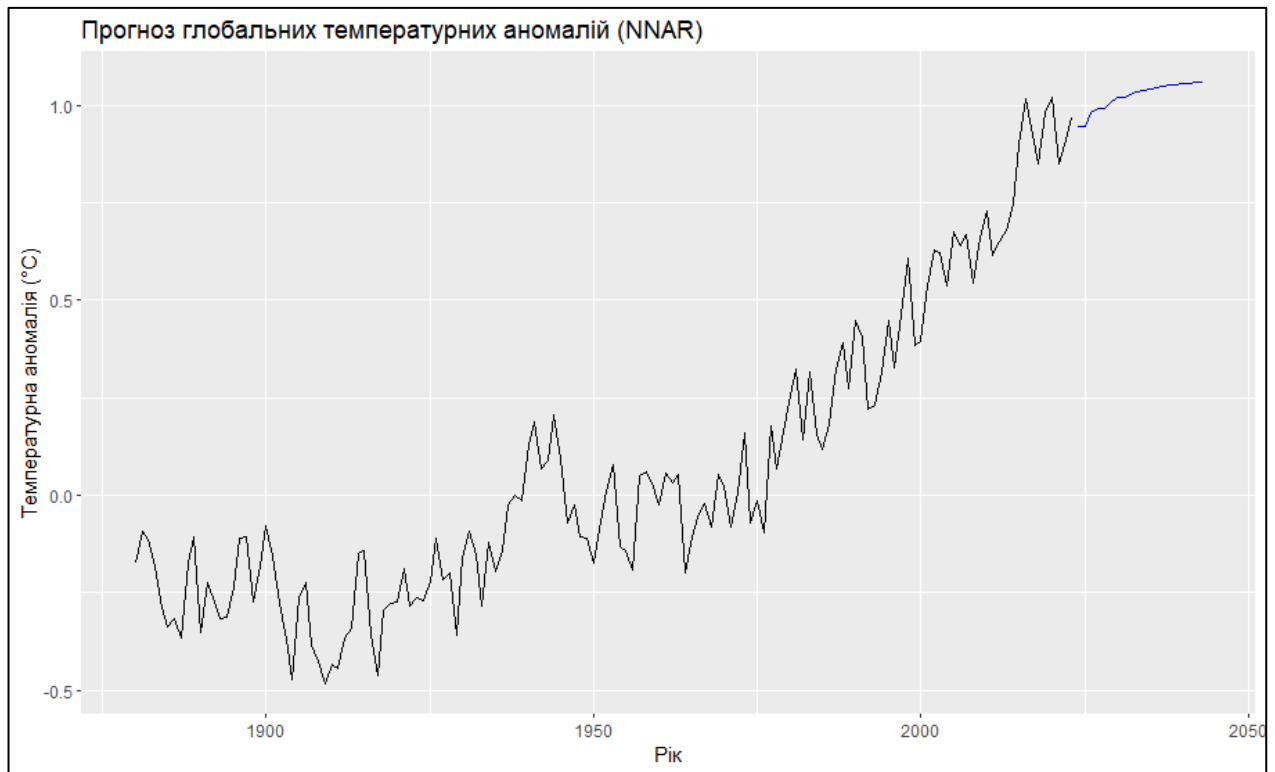


Рисунок 4.12 – Прогноз глобальних температурних аномалій (NNAR)

Прогноз, сформований нейронною мережею, демонструє подальше зростання температурних аномалій, що узгоджується з результатами статистичного та трендового аналізу, виконаного у попередніх розділах. На відміну від моделей ARIMA та ETS, нейромережева модель формує більш гладку прогнозну траєкторію з поступовим уповільненням темпів зростання, що може відображати ефект усереднення навченої моделі або обмеження екстраполяції за межі навчального інтервалу.

Згладжений характер прогнозу NNAR свідчить про орієнтацію моделі на відтворення загальної довгострокової тенденції без деталізації короткострокових коливань. Це є типовою особливістю нейромережевих моделей, які добре узагальнюють дані, але можуть втрачати локальні варіації при прогнозуванні на далекий горизонт.

4.4.4 Порівняльний аналіз результатів прогнозування та оцінка точності моделей

Для кількісної оцінки якості прогнозних моделей ARIMA, ETS та NNAR використано стандартні показники точності прогнозування, зокрема середню помилку (ME), середньоквадратичну помилку (RMSE), середню абсолютну помилку (MAE), середню відносну абсолютну помилку (MAPE) та масштабовану абсолютну помилку (MASE). Оцінювання виконано на навчальній вибірці, що дозволяє порівняти здатність моделей відтворювати історичну динаміку глобальних температурних аномалій (рис. 4.13).

```
> accuracy_arima <- accuracy(model_arima)
> accuracy_ets <- accuracy(model_ets)
> accuracy_nnar <- accuracy(model_nnar)
>
> accuracy_arima
              ME      RMSE      MAE      MPE      MAPE
Training set -0.0005938131 0.09836333 0.08208807 -3.751765 83.88674
              MASE      ACF1
Training set 0.8917081 -0.02052807
> accuracy_ets
              ME      RMSE      MAE      MPE      MAPE
Training set 0.01376526 0.1048406 0.08815656 -50.63405 139.6872
              MASE      ACF1
Training set 0.9576291 0.08771486
> accuracy_nnar
              ME      RMSE      MAE      MPE      MAPE
Training set -0.0002732169 0.09740696 0.08209226 -38.11392 120.1857
              MASE      ACF1
Training set 0.8917536 -0.06473426
```

Рисунок 4.13 – Оцінка точності моделей

Якісні характеристики прогнозних моделей наведено в таблиці 4.2.

Таблиця 4.2 – Порівняння прогностичних моделей глобальних температурних аномалій

Модель	Тип підходу	Характер прогнозу	Переваги	Обмеження
ARIMA	Статистична параметрична	Чіткий висхідний тренд, помірна варіативність	Добре враховує автокореляцію, інтерпретованість	Чутлива до структурних змін
ETS	Адаптивне експоненційне згладжування	Плавний, згладжений тренд	Стійка до шуму, добре виділяє довгострокові зміни	Ширші довірчі інтервали
NNAR	Нейромережева	Згладжений нелінійний прогноз	Враховує нелінійності, сучасний підхід	Менша інтерпретованість

Результати оцінювання наведено в таблиці 4.3. Аналіз показує, що всі розглянуті моделі демонструють близькі значення абсолютних і квадратичних помилок, що свідчить про їхню адекватність для моделювання довгострокових кліматичних процесів. Найменше значення RMSE отримано для нейромережевої моделі NNAR (0,0974 °C), що вказує на її високу здатність до апроксимації навчальних даних. Водночас модель ARIMA демонструє порівнянні результати з $RMSE = 0,0984$ °C та $MAE = 0,0821$ °C, що підтверджує ефективність класичного статистичного підходу.

Таблиця 4.3 – Показники точності прогностичних моделей

Модель	ME	RMSE	MAE	MAPE	MASE
ARIMA	-0,00059	0,09836	0,08209	83,89	0,8917
ETS	0,01377	0,10484	0,08816	139,69	0,9576
NNAR	-0,00027	0,09741	0,08209	120,19	0,8918

Модель ETS характеризується дещо більшими значеннями похибок ($RMSE = 0,1048$ °C, $MAE = 0,0882$ °C), що може бути пов'язано з більш агресивним згладжуванням часових коливань. Разом із тим такий характер прогнозу є прийнятним для стратегічних оцінок довгострокових кліматичних тенденцій.

Показник MASE для моделей ARIMA та NNAR є меншим за 1, що свідчить про їхню перевагу над наївною моделлю прогнозування. Відносно високі значення MAPE для всіх моделей зумовлені наявністю від'ємних та близьких до нуля значень температурних аномалій у початковій частині ряду, що є типовою проблемою при застосуванні відносних показників до кліматичних даних.

Модель ARIMA забезпечує добрий баланс між точністю та інтерпретованістю, ETS формує згладжений і стабільний прогноз, а модель NNAR демонструє найменші помилки апроксимації та здатність враховувати нелінійні залежності.

Висновки до розділу 4

У розділі 4 виконано трендовий аналіз і моделювання змін клімату на основі середніх річних глобальних температурних аномалій, що дозволило кількісно оцінити довгострокову тенденцію потепління та підтвердити її статистичну значущість.

Параметричний підхід на основі лінійної регресії дав змогу оцінити середню швидкість зростання температурних аномалій та пояснювальну здатність моделі. Непараметричні методи (тест Манна–Кендалла та оцінка Sen's slope) використано для перевірки стійкості висновків без жорстких припущень щодо розподілу даних і чутливості до викидів.

Окрему увагу приділено прогнозуванню глобальних температурних аномалій із застосуванням кількох підходів моделювання часових рядів. Побудовано прогнозні моделі ARIMA, ETS та нейромережеву модель NNAR, сформовано прогнозні траєкторії з довірчими інтервалами та виконано порівняльну оцінку їх точності за основними метриками (ME, RMSE, MAE, MAPE, MASE). Отримані результати засвідчили близькі значення похибок для всіх моделей, що підтверджує їх придатність до задач довгострокового прогнозування та дозволяє використовувати комплексний підхід для підвищення надійності прогнозних оцінок.

ВИСНОВКИ

У результаті виконання кваліфікаційної роботи магістра було досягнуто поставленої мети та вирішено основні завдання, визначені на початковому етапі дослідження. Розроблено інформаційну систему аналізу тенденцій зміни клімату за супутниковими даними, яка забезпечує автоматизований цикл роботи з кліматичними часовими рядами: завантаження, попередню обробку, аналіз, візуалізацію, оцінювання трендів і прогнозування глобальних температурних змін.

Проведений аналіз предметної області підтвердив, що кліматичні зміни характеризуються стійким зростанням глобальної температури та потребують застосування надійних методів обробки й моделювання великих масивів спостережень. На основі опрацювання наукових джерел визначено ключові підходи до оцінювання кліматичних тенденцій, охарактеризовано супутникові та комбіновані кліматичні дані, а також обґрунтовано використання методів аналізу часових рядів і прогнозування для задач моніторингу глобального потепління.

Практичну частину реалізовано мовою програмування R із використанням відкритого набору даних Global Surface Temperatures (Kaggle), який містить багаторічні глобальні й регіональні температурні аномалії. У межах розробки інформаційної системи сформовано структуру програмного проєкту, налаштовано бібліотеки та реалізовано модулі імпорту й обробки даних. Виконано перетворення даних у формат, придатний для аналізу часових рядів, здійснено перевірку повноти та усунення пропусків (зокрема лінійну інтерполяцію), сформовано фінальний очищений часовий ряд і побудовано комплекс графічних матеріалів для розвідувального та статистичного аналізу, включаючи сезонні профілі, аналіз за десятиліттями, визначення найтепліших років і STL-декомпозицію.

У розділі трендового аналізу отримано кількісні оцінки інтенсивності потепління та підтверджено їх статистичну значущість. Параметричний підхід (лінійна регресія) та непараметричні методи (тест Манна–Кендалла і робастна оцінка Sen's slope) дали узгоджені результати, що підтверджує стійкий монотонно зростаючий тренд глобальних температурних аномалій. Для моделювання

майбутньої динаміки виконано прогнозування за моделями ARIMA, ETS та нейромережевою моделлю NNAR, а також проведено оцінювання точності прогнозів за стандартними метриками (ME, RMSE, MAE, MAPE, MASE). Порівняльний аналіз засвідчив, що моделі демонструють близькі значення похибок, що дозволяє застосовувати комплексний підхід до прогнозування для підвищення надійності оцінок.

Таким чином, у межах кваліфікаційної роботи створено програмну інформаційну систему, яка забезпечує відтворюваний і статистично обґрунтований аналіз кліматичних тенденцій, формує наочні візуалізації та підтримує прогнозування глобальних температурних змін. Розроблені програмні модулі та отримані результати можуть бути використані для подальших наукових досліджень, навчальних цілей, а також як основа для розширення системи у напрямку просторового аналізу за широтними зонами та інтеграції додаткових кліматичних показників.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. IPCC, 2021. Climate Change 2021: The Physical Science Basis. Summary for Policymakers. Intergovernmental Panel on Climate Change (AR6 WGI). URL: <https://www.ipcc.ch/report/ar6/wg1/> (дата звернення: 08.10.2025).
2. Robinson, W.A., 2021. Climate change and extreme weather: A review. Journal article (2021). URL: https://www.tandfonline.com/doi/full/10.1080/10962247.2021.1942319?utm_source=chatgpt.com (дата звернення: 08.10.2025).
3. NASA GISS. GISTEMP v4 – GISS Surface Temperature Analysis. Available at: <https://data.giss.nasa.gov/gistemp/> (дата звернення: 09.10.2025).
4. NOAA National Centers for Environmental Information (NCEI). State of the Climate – Global Climate Report 2024. Available at: <https://www.ncei.noaa.gov/> (дата звернення: 09.10.2025).
5. Abbass, K., et al., 2022. A review of the global climate change impacts, adaptation and mitigation. Environmental Science and Policy (open access). URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8978769/> (дата звернення: 09.10.2025).
6. Thompson, D.W.J., et al. (2021). Trends and Variability in Global and Regional Surface Temperatures. Reviews of Geophysics, 59(3), e2020RG000704.
7. IPCC, 2013. Climate Change 2013: The Physical Science Basis. Intergovernmental Panel on Climate Change (AR5 WG1). URL: <https://www.ipcc.ch/report/ar5/wg1/> (дата звернення: 10.10.2025).
8. World Meteorological Organization (WMO), 2020. State of the Global Climate. URL: <https://public.wmo.int/> (дата звернення: 10.10.2025).
9. Wilks, D.S., 2019. Statistical Methods in the Atmospheric Sciences. Academic Press.
10. Global Historical Climatology Network (GHCN). URL: <https://www.ncei.noaa.gov/products/land-based-station/global-historical-climatology-network-daily> (дата звернення: 10.10.2025).

11. ERA5 (ECMWF Reanalysis v5) URL: <https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5> (дата звернення: 10.10.2025).
12. HadCRUT (Hadley Centre / University of East Anglia). URL: <https://www.metoffice.gov.uk/hadobs/hadcruz5/> (дата звернення: 10.10.2025).
13. Berkeley Earth Dataset. URL: <https://berkeleyearth.org/data/> (дата звернення: 10.10.2025).
14. MODIS (Moderate Resolution Imaging Spectroradiometer), NASA. URL: <https://modis.gsfc.nasa.gov/> (дата звернення: 10.10.2025).
15. NOAA AVHRR (Advanced Very High Resolution Radiometer). URL: <https://www.ncei.noaa.gov/products/avhrr> (дата звернення: 10.10.2025).
16. Copernicus Sentinel Missions, European Space Agency (ESA). URL: <https://sentinel.esa.int/web/sentinel/home> (дата звернення: 10.10.2025).
17. Landsat Missions, NASA / U.S. Geological Survey (USGS). URL: <https://landsat.gsfc.nasa.gov/> (дата звернення: 10.10.2025).
18. CERES (Clouds and the Earth's Radiant Energy System), NASA. URL: <https://ceres.larc.nasa.gov/> (дата звернення: 10.10.2025).
19. Mann, H.B. (1945); Kendall, M.G. (1975). Mann–Kendall Trend Test. URL: <https://cran.r-project.org/web/packages/trend/trend.pdf> (дата звернення: 15.10.2025) Sen, P.K. (1968).
20. Estimates of the Regression Coefficient Based on Kendall's Tau. URL: <https://www.jstor.org/stable/2238201> (дата звернення: 15.10.2025).
21. Draper, N.R., Smith, H. (1998). Applied Regression Analysis. URL: <https://onlinelibrary.wiley.com/doi/book/10.1002/9781118625590> (дата звернення: 15.10.2025).
22. Chatfield, C. (2003). The Analysis of Time Series: An Introduction. URL: <https://www.crcpress.com/The-Analysis-of-Time-Series-An-Introduction/Chatfield/p/book/9781584883173> (дата звернення: 15.10.2025).

23. Torrence, C., Compo, G.P. (1998). A Practical Guide to Wavelet Analysis. URL: <https://paos.colorado.edu/research/wavelets/> (дата звернення: 15.10.2025).
24. Cleveland, R.B., Cleveland, W.S., McRae, J.E., Terpenning, I. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess. URL: <https://www.jstor.org/stable/2237638> (дата звернення: 20.10.2025).
25. Bloomfield, P. (2000). Fourier Analysis of Time Series: An Introduction. URL: <https://onlinelibrary.wiley.com/doi/book/10.1002/9780470521400> (дата звернення: 20.10.2025).
26. Tukey, J.W. (1977). Exploratory Data Analysis. URL: <https://www.jstor.org/stable/2237845> (дата звернення: 10.10.2025).
27. Jain, A.K. (2010). Data Clustering: 50 Years Beyond K-means. URL: <https://ieeexplore.ieee.org/document/5432202> (дата звернення: 20.10.2025).
28. Hansen, J., Ruedy, R., Sato, M., Lo, K. (2010). Global Surface Temperature Change. Reviews of Geophysics.
29. Benestad, A. E. (2013). Association between trends in daily rainfall percentiles and the global mean temperature. Journal of Geophysical Research: Atmospheres.
30. Deo, S., Şahin, M. (2015). Application of artificial neural networks for climate change prediction. Atmospheric Research.

ДОДАТОК А

Код для побудови інформаційної системи

```
dir.create("data")
dir.create("scripts")
dir.create("plots")
dir.create("output")

list.files()

install.packages(c(
  "tidyverse",    # базова обробка даних + ggplot2
  "lubridate",    # робота з датами
  "forecast",     # ARIMA, ETS, NNAR
  "zoo",          # інтерполяція пропусків
  "tseries",      # ADF-тест
  "trend",        # тест Манна-Кендалла
  "tsibble",      # сучасна робота з часовими рядами
  "fable",        # моделі прогнозування (як альтернатива forecast)
  "fabletools"    # допоміжні інструменти для fable
))

library(tidyverse)
library(lubridate)
library(forecast)
library(zoo)
library(tseries)
library(trend)
library(tsibble)
library(fable)
library(fabletools)

# 1) Глобальні температури
global_temps <- read_csv("data/global_temps.csv")

# 2) Температури Північної півкулі
nh_temps <- read_csv("data/nh_temps.csv")

# 3) Температури Південної півкулі
sh_temps <- read_csv("data/sh_temps.csv")

# 4) Зональні та річні температури
zonann_temps <- read_csv("data/zonann_temps.csv")

glimpse(global_temps)
glimpse(nh_temps)
glimpse(sh_temps)
glimpse(zonann_temps)

# Перегляд перших рядків (за потреби)
head(global_temps)
head(nh_temps)
head(sh_temps)
head(zonann_temps)
```

```
# Порядок місяців у вихідній таблиці
month_levels <- c("Jan", "Feb", "Mar", "Apr", "May", "Jun",
                 "Jul", "Aug", "Sep", "Oct", "Nov", "Dec")

# Перетворення широкого формату (рік + 12 місяців) у довгий формат "рік-місяць-аномалія"
global_series <- global_temps %>%
  select(Year, all_of(month_levels)) %>%
  pivot_longer(
    cols      = -Year,
    names_to  = "Month",
    values_to = "Temp_Anomaly"
  ) %>%
  mutate(
    Month      = factor(Month, levels = month_levels),
    Month_Num  = as.integer(Month),
    Date       = make_date(year = Year, month = Month_Num, day = 1)
  ) %>%
  arrange(Date)

# Перевіримо структуру отриманого часового ряду
glimpse(global_series)

# Оцінка кількості пропусків
total_na <- sum(is.na(global_series$Temp_Anomaly))
n_obs    <- nrow(global_series)
na_percent <- round(100 * total_na / n_obs, 3)

total_na
na_percent

# Інтерполяція пропущених значень
global_series_interpolated <- global_series %>%
  mutate(
    # Лінійна інтерполяція за часом
    Temp_Anomaly_Interp = zoo::na.approx(
      Temp_Anomaly,
      x      = Date,
      na.rm  = FALSE
    ),
    Temp_Anomaly_Interp = zoo::na.locf(Temp_Anomaly_Interp, na.rm = FALSE),
    Temp_Anomaly_Interp = zoo::na.locf(Temp_Anomaly_Interp, fromLast = TRUE)
  )

# Перевірка відсутності пропусків після інтерполяції
total_na_after <- sum(is.na(global_series_interpolated$Temp_Anomaly_Interp))
total_na_after

# Формування фінального очищеного набору
global_clean <- global_series_interpolated %>%
  select(
    Date,
    Year,
    Month,
    Month_Num,
    Temp_Anomaly = Temp_Anomaly_Interp
  )
```

```
glimpse(global_clean)

# Глобальні температурні аномалії + ковзні середні
global_trend <- global_clean %>%
  arrange(Date) %>%
  mutate(
    MA_12 = rollmean(Temp_Anomaly, k = 12, fill = NA, align = "right"),
    MA_60 = rollmean(Temp_Anomaly, k = 60, fill = NA, align = "right")
  )

ggplot(global_trend, aes(x = Date)) +
  geom_line(aes(y = Temp_Anomaly), alpha = 0.4) +
  geom_line(aes(y = MA_12), linewidth = 0.6) +
  geom_line(aes(y = MA_60), linewidth = 0.9, linetype = "dashed") +
  labs(
    title = "Глобальні температурні аномалії поверхні Землі",
    subtitle = "Помісячні значення та ковзні середні за 12 і 60 місяців",
    x = "Рік",
    y = "Температурна аномалія (°C)"
  ) +
  theme_minimal()

# Збільшення фрагмента: 1950–2023 рр.

global_trend_zoom <- global_trend %>%
  filter(Year >= 1950)

ggplot(global_trend_zoom, aes(x = Date)) +
  geom_line(aes(y = Temp_Anomaly), alpha = 0.4) +
  geom_line(aes(y = MA_12), linewidth = 0.7) +
  geom_line(aes(y = MA_60), linewidth = 1, linetype = "dashed") +
  labs(
    title = "Глобальні температурні аномалії (фрагмент 1950–2023 рр.)",
    subtitle = "Посилення висхідного тренду у другій половині XX – на початку XXI
ст.",
    x = "Рік",
    y = "Температурна аномалія (°C)"
  ) +
  theme_minimal()

# Розподіл глобальних температурних аномалій

ggplot(global_clean, aes(x = Temp_Anomaly)) +
  geom_histogram(aes(y = ..density..), bins = 40, alpha = 0.5) +
  geom_density(linewidth = 1) +
  labs(
    title = "Розподіл глобальних температурних аномалій",
    x = "Температурна аномалія (°C)",
    y = "Щільність"
  ) +
  theme_minimal()

# Boxplot глобальних аномалій за десятиліттями

global_decades <- global_clean %>%
  mutate(
    Decade = floor(Year / 10) * 10
```

```
)

ggplot(global_decades, aes(x = factor(Decade), y = Temp_Anomaly)) +
  geom_boxplot() +
  labs(
    title = "Розподіл глобальних температурних аномалій за десятиліттями",
    x      = "Десятиліття",
    y      = "Температурна аномалія (°C)"
  ) +
  theme_minimal() +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1)
  )

# Середні річні глобальні аномалії + ТОП-10 найтепліших років

yearly_stats <- global_clean %>%
  group_by(Year) %>%
  summarise(Mean_Anomaly = mean(Temp_Anomaly), .groups = "drop")

top10_hot <- yearly_stats %>%
  arrange(desc(Mean_Anomaly)) %>%
  slice(1:10)

yearly_stats <- yearly_stats %>%
  mutate(Category = ifelse(Year %in% top10_hot$Year, "ТОП-10", "Інші роки"))

ggplot(yearly_stats, aes(x = Year, y = Mean_Anomaly, color = Category)) +
  geom_line(linewidth = 0.7) +
  geom_point(size = 2) +
  scale_color_manual(values = c("black", "red")) +
  labs(
    title = "Середні річні глобальні температурні аномалії",
    subtitle = "Виділено 10 найтепліших років за весь період спостережень",
    x = "Рік",
    y = "Середня річна температурна аномалія (°C)",
    color = "Категорія"
  ) +
  theme_minimal()

# STL-декомпозиція глобальних температурних аномалій
ts_global <- ts(global_clean$Temp_Anomaly,
               start = c(min(global_clean$Year), global_clean$Month_Num[1]),
               frequency = 12)

stl_result <- stl(ts_global, s.window = "periodic", robust = TRUE)

plot(stl_result,
     main = "STL-декомпозиція глобальних температурних аномалій")

yearly_stats <- global_clean %>%
  group_by(Year) %>%
  summarise(
    Mean_Anomaly = mean(Temp_Anomaly, na.rm = TRUE),
    .groups = "drop"
  )

head(yearly_stats)

# Лінійна регресійна модель: Mean_Anomaly ~ Year
```

```
trend_lm <- lm(Mean_Anomaly ~ Year, data = yearly_stats)

# Вивід результатів моделі
summary(trend_lm)

trend_coef_year <- coef(trend_lm)["Year"]          # °C за 1 рік
trend_coef_decade <- 10 * trend_coef_year          # °C за 10 років

trend_coef_year
trend_coef_decade

model_summary <- summary(trend_lm)

r2_value <- model_summary$r.squared
p_value_year <- model_summary$coefficients["Year", "Pr(>|t|)"]

r2_value
p_value_year

# Середні річні значення + лінія тренду
ggplot(yearly_stats, aes(x = Year, y = Mean_Anomaly)) +
  geom_point(alpha = 0.7) +
  geom_smooth(method = "lm", se = TRUE, linewidth = 0.9) +
  labs(
    title = "Лінійний тренд глобальних температурних аномалій",
    x = "Рік",
    y = "Середня річна температурна аномалія (°C)"
  ) +
  theme_minimal()

# Тест Манна-Кендалла
mk_test <- mk.test(yearly_stats$Mean_Anomaly)

mk_test

# Напрямок тренду
trend_direction <- ifelse(mk_test$z > 0, "зростаючий", "спадний")
trend_direction

# Sen's slope
sen_slope <- sens.slope(yearly_stats$Mean_Anomaly)

sen_slope

mk_z <- mk_test$z
mk_p <- mk_test$p.value
mk_tau <- Kendall(yearly_stats$Mean_Anomaly)$tau

sen_value <- sen_slope$estimates
sen_ci <- sen_slope$conf.int

mk_z
mk_p
mk_tau
```

```
sen_value
sen_ci

# Підготовка часового ряду

# Перетворення річних даних у часовий ряд
temp_ts <- ts(
  yearly_stats$Mean_Anomaly,
  start = min(yearly_stats$Year),
  frequency = 1
)

# Побудова базових прогнозних моделей

model_arima <- auto.arima(temp_ts)

model_ets <- ets(temp_ts)

model_nnar <- nnetar(temp_ts)

# Прогноз на 20 років уперед

h <- 20

forecast_arima <- forecast(model_arima, h = h)
forecast_ets <- forecast(model_ets, h = h)
forecast_nnar <- forecast(model_nnar, h = h)

# Візуалізація прогнозів

autoplot(forecast_arima) +
  labs(
    title = "Прогноз глобальних температурних аномалій (ARIMA)",
    x = "Рік",
    y = "Температурна аномалія (°C)"
  )

autoplot(forecast_ets) +
  labs(
    title = "Прогноз глобальних температурних аномалій (ETS)",
    x = "Рік",
    y = "Температурна аномалія (°C)"
  )

autoplot(forecast_nnar) +
  labs(
    title = "Прогноз глобальних температурних аномалій (NNAR)",
    x = "Рік",
    y = "Температурна аномалія (°C)"
  )

# Оцінка точності моделей

accuracy_arima <- accuracy(model_arima)
accuracy_ets <- accuracy(model_ets)
accuracy_nnar <- accuracy(model_nnar)

accuracy_arima
accuracy_ets
accuracy_nnar
```