

Галіцин В.К., д.е.н., професор,
професор кафедри комп'ютерної математики та інформаційної безпеки,
Київський національний економічний університет
імені Вадима Гетьмана

Жук Д.В., студент 3 курсу спеціальності «Системний аналіз»,
кафедра комп'ютерної математики та інформаційної безпеки,
Київський національний економічний університет
імені Вадима Гетьмана

Петриченко А.В., студентка 3 курсу спеціальності «Системний аналіз»,
кафедра комп'ютерної математики та інформаційної безпеки,
Київський національний економічний університет
імені Вадима Гетьмана.

Galicin V.K., Doctor of Economic Sciences,
Professor, Professor Department of Computer Mathematics
and Information Security,
KNEU named after Vadym Hetman

Zhuk D.V., Student of the 3rd year, specialization «Systems analysis»,
Department of Computer Mathematics and Information Security,
KNEU named after Vadym Hetman

Petrychenko A.V., Student of the 3rd year, specialization «Systems
analysis»,
Department of Computer Mathematics and Information Security,
KNEU named after Vadym Hetman

ВИПАДКОВІ ПРОЦЕСИ В МЕТЕОРОЛОГІЇ

RANDOM PROCESSES IN METEOROLOGY

Анотація. Зміна клімату — це спостережувані та прогнозовані довгострокові зміни середнього клімату, а також мінливість клімату, викликана діяльністю людини, включаючи такі аномалії, як посухи, сильні шторми та повені. Клімат нашої планети постійно змінювався протягом усієї геологічної історії Землі, і ці зміни супроводжувалися значними коливаннями середніх глобальних температур. Одним із головних чинників зміни глобальної середньої температури є спалювання вугільного палива, вугілля, газу та нафти, що збільшило концентрацію парникових газів, таких як вуглекислий газ, у нашій атмосфері. Ця робота в першу чергу спрямована на вивчення проблем прогнозування та моніторингу такого погодного явища, як локальна зміна температури — на основі моделей часових рядів, представлених у цій роботі, можна зробити необхідні прогнози також для вимірювання рівня снігового покриву, рівень опадів. Використання моделей прогнозування температури (таких як метод Холта-Вінтера або бібліотека Пророка) має велике практичне значення, починаючи від прогнозування температури на наступний рік

для сівби, врожайності культур тощо. У даній роботі було проведено довгостроковий аналіз середньодобової температури у Київській області мовою програмування Python та її бібліотеками для аналізу даних Pandas, Numpy, Scikit-Learn тощо за допомогою моделей експоненційного згладжування та бібліотеки Prophet. За допомогою цих моделей можна обробляти та прогнозувати різні типи даних, такі як: температура, рівень опадів, глибина снігового покриву. Крім того, прогнози можуть бути використані для різноманітних цілей людської діяльності. Матеріали статті мають науково-методичний характер.

Ключові слова: випадковий процес; часові ряди; метеорологія; зміна клімату; модель прогнозування; навчання під наглядом; експоненціальне згладжування; метод Холта-Вінтера; бібліотека Пророк.

Abstract. Climate change is the observed and projected long-term changes in average climate, as well as climate variability, caused by human activity, including such anomalies as droughts, severe storms and floods. The climate of our planet has been constantly changing throughout the entire geological history of the Earth, and these changes were accompanied by significant fluctuations in average global temperatures. One of the main drivers of global average temperature change is the burning of fossil fuels, coal, gas and oil, which has increased the concentration of greenhouse gases such as carbon dioxide in our atmosphere. This work is primarily aimed at studying the problems of predicting such a weather phenomenon as a local change in temperature — based on the time series models presented in this work, it is possible to make the necessary predictions also for measuring the level of snow cover, precipitation level. The use of temperature prediction models (such as Holt-Winter's method or Prophet library) is of great practical importance, starting from predicting the temperature for the next year for sowing, crop yields, etc. In this work, a long-term analysis of the average daily temperature in the Kyiv region was carried out using the Python programming language and its data analysis libraries Pandas, Numpy, Scikit-Learn, etc. using exponential smoothing models and the Prophet library. These models can be used to process and forecast different types of data, such as: temperature, precipitation, snow depth. In addition, forecasts can be used for various purposes of human activity. The materials of the article have a scientific and methodological nature.

Keywords: Random process; time series; meteorology; climate change; prediction model; supervised learning; exponential smoothing; Holt-Winter's method; Prophet.

Постановка проблеми. Зміна клімату та наслідки цього процесу для планети — один з важливих напрямів сучасних метеорологічних досліджень. Метеорологія наука про будову і властивості земної атмосфери та фізичних процесів, що в ній відбуваються. Перша офіційна згадка про цю науку належить Аристотелю, який приблизно в 340 р. до н. е. написав книгу під назвою «Метеорологія», що містила все, що було відомо на той час про погоду та клімат. Назва походила від грецького слова «Μετέωρο», що означало «щось високо» і стосувалося усього, що спостерігалось в атмосфері [2].

Математичні методи та моделі використовують для розв'язання побудови прогнозу погоди і загальної циркуляції атмосферних процесів, чисельних алгоритмів для їх вирішення. Рівняння гідротермодинаміки атмосферних процесів є досить скла-

дними, тому існує необхідність розробки ефективніших алгоритмів, здатних з високою точністю описати широкий спектр завдань динамічної метеорології та прогнозу погоди. Побудова алгоритмів розв'язання таких завдань тісно пов'язана з проблемою апроксимації рівнянь, генералізації даних і стійкості отриманих розв'язків, які є основними проблемами під час конструкції нових чисельних алгоритмів загалом. З наукової точки зору передбачення погоди — одне з найскладніших завдань фізики атмосфери. Удосконалення відповідного комп'ютерного обладнання дозволяє реалізовувати математичні підходи і методи дослідження атмосфери. Кожен із них дає змогу в тій чи тій мірі прогнозувати метеорологічні явища.

Аналіз останніх досліджень і публікацій. Для отримання достовірного прогнозу погоди, у міжнародних кліматичних центрах збирається інформація про поточну погоду на Землі. Вона отримується з тисяч метеостанцій, метеопостів, зондів, радіобуїв: наземних, літальних, плаваючих, що покривають необхідну територію густою сіткою. Для того щоб модель провела розрахунок прогнозу погоди, їй необхідно зібрати якомога більше даних про поточну погоду. Чим густіше на території розміщені метеостанції, тим точніше буде розрахований прогноз [4].

Статистичні моделі спостережуваних щоденних змін («Генератори погоди»), можуть заповнювати відсутні дані або створювати нескінченно довгі синтетичні — ці моделі погодних часових рядів імітують ключові властивості спостережуваних метеорологічних записів (тобто щоденні середні значення, дисперсії та коваріації, частоти, екстремуми тощо). Вони використовуються в моделюванні змін сільського господарства, екосистем чи клімату, оскільки спостережувані наземні метеорологічні дані часто є неадекватними у контексті їх тривалості, повноти чи просторового охоплення.

Щоденні симулятори погоди є, безумовно, найпоширенішими як через широку доступність даних про погоду в часовому масштабі, так і з великою кількістю моделей впливів, які керуються щоденними даними про погоду [5]. Демонстрацію роботи випадкового генератора погоди подано на рис. 1.

Існує два взаємодоповнюючі способи перегляду погодних моделей. По-перше, коли ці моделі використовуються для моделювання за методом Монте-Карло, їх можна розглядати як складні генератори випадкових чисел, результати яких статистично нагадують щоденні дані про погоду в певному місці [1].

Random Weather Generator

Type: **Climate** **Season** **Supernatural**

Temperate Summer Rare **Select**

Weather:



Description: Windy

Temperature: Moderate

High: 75°F (23°C)

Low: 59°F (15°C)

Relative: Warmer than normal

Wind Force: Moderate

Wind Speed: 11 mph (17 kph)

Moderate Wind: A steady wind with a 50% chance of extinguishing candles, torches, and similar unprotected flames.

Рис. 1. Випадковий генератор погоди

Джерело: [9].

Важливо зазначити, що генератори погоди не є алгоритмами прогнозування погоди, і тому сильно відрізняються від детермінованих погодних моделей, які працюють шляхом чисельного інтегрування диференціальних рівнянь у частинних похідних, що описують потоки рідини. Приклад моделювання за вказаним методом продемонстровано на рис. 2. По-друге, це стохастичні моделі щоденних (і відповідно довготривалих) коливань погоди. З цієї позиції параметри стохастичної моделі погоди містять стислу дистиляцію певних аспектів місцевого клімату на локальній території.

Опади не лише є найважливішою метеорологічною змінною для багатьох застосувань, а й наявність або відсутність опадів також зазвичай впливає на статистику багатьох змінних без опадів, які потрібно моделювати. Дані про атмосферні опади демонструють відмінні та складні характеристики, які ускладнюють статистичні моделі, необхідні для їх опису. На додаток до кореляції між значеннями в послідовні періоди часу, яка є типовою для всіх погодних змінних, кількість опадів є унікальною за змішаним характером, оскільки це є як дискретна, так і неперервна змінна [7].

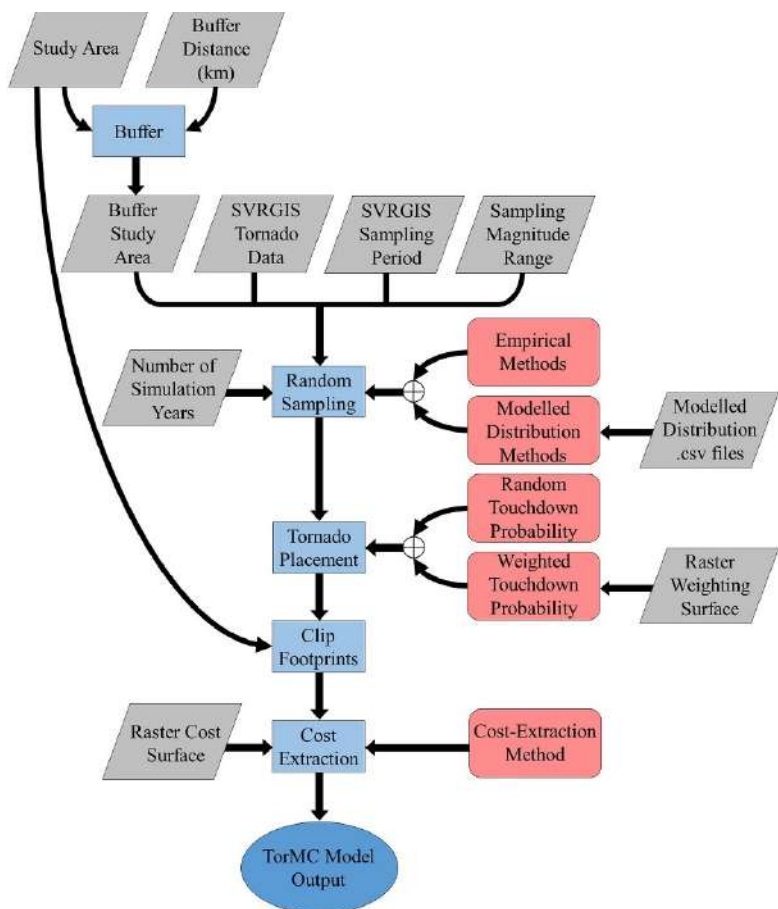


Рис. 2. Приклад моделі для розрахунку впливу торнадо за методом Монте-Карло

Джерело: [10].

Іншими словами, опади дуже часто точно дорівнюють нулю, і отже, існує розрив у розподілі ймовірностей даних про опади між нульовими та ненульовими спостереженнями. Процес випадання опадів проявляється у двох станах погоди — вологому або сухому. Ключовим аспектом стохастичних моделей погоди є представлення тенденції вологих і сухих днів виявляти стійкість або позитивну послідовну кореляцію, так що вологий і сухий цикли мають тенденцію повторюватися у часі сильніше, ніж очікується.

Процес визначення інтенсивності опадів належить до моделювання ненульової кількості опадів.

Через розвиток науково-технічного прогрес зростає не лише завчасність прогнозів, а й їхня точність — це дає можливість отримати прогностичну інформацію уже не лише для кожного пункту, а й для окремих районів міст, навіть вулиць. «Проте для того, щоб зробити такі розрахунки для України, необхідно суттєво реорганізувати існуючу мережу спостережень, адже вона уже давно не задовольняє існуючі потреби в інформації. Саме це завдання і є одним з пріоритетних для ДСНС України. Технічному переоснащенню мережі спостережень в Україні сприятиме співпраця з компанією «Baron — Critical Weather Intelligence», з якою нещодавно підписано Протокол про співробітництво у сфері гідрометеорологічної діяльності», — повідомляють у ДСНС України.

Основним методом, який використовують синоптики в усьому світі для прогнозу погоди, є синоптичний. Цей метод передбачає аналіз і прогноз атмосферних процесів і умов погоди за допомогою синоптичних карт, аерологічних діаграм, вертикальних розрізів атмосфери та інших засобів. Дані численних моделей погоди все ще є допоміжними засобами. Вони дозволяють отримати інформацію лише про основні характеристики атмосфери — температуру, тиск, вологість, швидкість вітру та їхню зміну. Використовуючи цю інформацію, синоптики можуть спрогнозувати явища погоди, їхню інтенсивність та локалізацію і попередити про них населення та відповідні органи влади.

Зміна клімату — це спостережувані та прогнозовані довгострокові зміни клімату, а також мінливість клімату, спричинена діяльністю людини, включаючи такі аномалії, як посухи, сильні шторми та повені. Удосконалення якості прогнозів погоди відбувається за двома напрямками: поліпшення справджуваності прогнозу і збільшення його завчасності. Короткостроковий прогноз визначається переважно початковим станом атмосфери. Для середньострокових прогнозів потрібне як детальне знання початкових полів метеорологічних величин, так і уміння описати впливи зовнішніх чинників, які призводять до нового стану рівноваги системи, що прогнозується. Довгостроковий прогноз меншою мірою залежить від початкового стану атмосфери і включає елементи моделювання клімату.

Мета статті — аналіз проблем прогнозування такого погодніх явища, як локальна зміна температури, на основі моделей часових рядів та їх застосування також для вимірювання рівня снігового покриву, рівню опадів.

Основний матеріал дослідження. З метою довгострокового аналізу та прогнозу середнього рівня температури у Київській обл. було використано відкриті дані з сайту NOAA (National Oceanic and Atmospheric Administration). Київська обл., на якій проводиться довгостроковий прогноз та аналіз показників заморів, розташована на півночі України. Клімат помірно континентальний, помірний, з достатньою вологістю.

NOAA — це федеральне відомство в структурі Міністерства торгівлі, яке займається різними видами метеорологічних, геодезичних досліджень і прогнозів для США та інших країн, вивченням світового океану і атмосфери. На сайті NOAA розміщені відкриті дані, де за певний проміжок часу можна знайти місцеві кліматологічні дані, морські дані, річні та сезонні норми тощо. Наявні відкриті дані з вимірювань станцій 3 у Київській обл.: KIEV, UP; BORYSPIL, UP; NEMESHAIEVO, UP. Сторінку з сайту з прикладом запиту на пошук денних загальних показників погоди у Київській обл. подано на рис. 3. На рис. 4 показано деякі деталі про локацію, що досліджується. Інформація про наявні станції в Київській обл. представлено на рис. 5.

■ Climate Data Online Search

Start searching here to find past weather and climate data. Search within a date range and select specific type of search. All fields are required.

Select Weather Observation Type/Dataset

Daily Summaries

Select Date Range

2022-01-01 to 2022-08-26

Search For

Countries

Enter a Search Term

UKRAINE

SEARCH

Рис. 3. Пошук денних загальних показників погоди у Київській обл.

Джерело: [11].

Daily Summaries Location Details

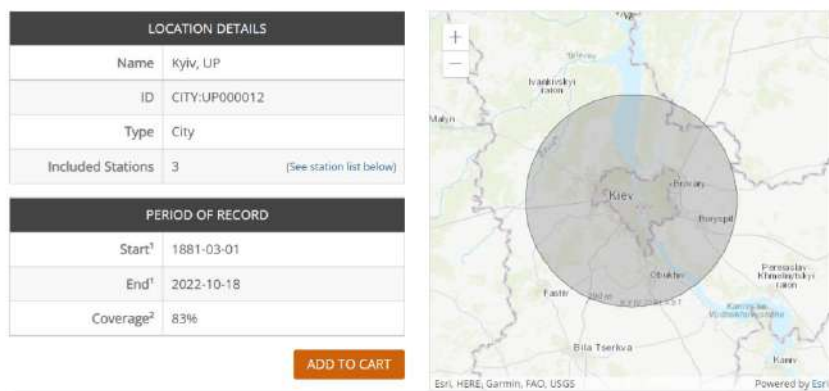


Рис. 4. Деталі шуканої локації, % покриття станцій

Джерело: [12].

Location Station List & Summarized Data Inventory

Available Data Types	STATION NAME & ID *	START ¹ ‡	END ¹ ‡	COVERAGE ² ‡	
Air Temperature	BORYSPIL, UP GHCND:UPM00033347	1959-07-01	2021-12-31	44%	ADD
Precipitation	KIEV, UP GHCND:UPM00033345	1881-03-01	2022-10-18	83%	ADD
Included Stations	NEMESHAIVO, UP GHCND:UPM00033342	2000-03-26	2009-06-16	55%	ADD
Station List					
Additional Information					
Documents					

1-3 of 3

Рис. 5. Наявні станції у Київській обл., періоди замірів спостережень

Джерело: [13].

Структура отриманих даних:

- STATION — номер певної станції;
- NAME — назва певної станції;
- LATITUDE, LONGITUDE — широта, довгота;
- DATE — дата;
- TAVG — середня температура;
- TMAX — максимальна температура;

- TMIN — мінімальна температура;
- PRCP — рівень опадів;
- SNWD — глибина сніжного покриву.

Дослідження даних проводилося мовою програмування Python, включаючи бібліотеки для роботи з масивами великих даних Pandas, Numpy, засобів візуалізації Plotly, Matplotlib, статистичних моделей Statsmodels, класичного машинного навчання з вчителем Scikit-Learn. Виконавши розвідковий аналіз даних (Exploratory Data Analysis), на перший погляд можна помітити, що:

- 1) дозмір дата сету — 24844 записи та 11 колонок;
- 2) максимальні, максимальні, середні, середньоквадратичне відхилення рівнів температур, опадів, глибини сніжного покриву на 3 станціях в Київській області;
- 3) існує багато пропущених даних, особливо рівня глибини снігу (пояснюється сезонністю), температур максимальних та мінімальних;
- 4) унікальних записів на кожній станції — різна кількість, так як заміри не проводилися кожен день;
- 5) станція NEMESHAEVO, UP почала працювати набагато пізніше (у 2000 р.), ніж станції KIEV, UP і BORYSPIL, UP.

На рис. 6 продемонстровано статистику даних та аналіз кількості пропущених значень. На рис. 7 подано приклад перших замірів у 2000 р.

	LATITUDE	LONGITUDE	ELEVATION	PRCP	SNWD	TAVG	TMAX	TMIN
count	24884.000000	24884.000000	24884.000000	8825.000000	2468.000000	22698.000000	13534.000000	12332.000000
mean	50.391401	30.678729	129.332744	2.399127	120.184765	8.748498	13.378077	4.697389
std	0.077932	0.284881	48.803056	6.259729	127.966285	9.986619	11.164683	8.962894
min	50.333000	30.100000	0.000000	0.000000	10.000000	-26.500000	-21.700000	-43.000000
25%	50.333000	30.533100	122.000000	0.000000	30.000000	1.000000	3.800000	-1.400000
50%	50.400000	30.533100	122.000000	0.200000	79.000000	8.900000	14.000000	4.700000
75%	50.400000	30.957000	166.000000	2.200000	180.000000	17.200000	22.800000	12.200000
max	50.600000	30.967000	186.000000	255.000000	2230.000000	36.200000	94.600000	26.800000
STATION	0							
NAME	0							
LATITUDE	0							
LONGITUDE	0							
ELEVATION	0							
DATE	0							
PRCP	16059							
SNWD	22416							
TAVG	2186							
TMAX	11350							
TMIN	12552							
dtype: int64								

Рис. 6. Статистики отриманих даних, кількість пропущених значень (NaN values)

Джерело: розроблено авторами.

	STATION	NAME	LATITUDE	LONGITUDE	ELEVATION	DATE	PRCP	SNWD	TAVG	TMAX	TMIN
0	UPM00033342	NEMESHAЕVO, UP	50.6	30.1	0.0	2000-03-26	NaN	NaN	NaN	NaN	12.6
1	UPM00033342	NEMESHAЕVO, UP	50.6	30.1	0.0	2000-04-25	NaN	NaN	NaN	21.0	NaN
2	UPM00033342	NEMESHAЕVO, UP	50.6	30.1	0.0	2000-04-26	NaN	NaN	NaN	23.1	12.6
3	UPM00033342	NEMESHAЕVO, UP	50.6	30.1	0.0	2000-04-27	NaN	NaN	23.1	23.0	11.5
4	UPM00033342	NEMESHAЕVO, UP	50.6	30.1	0.0	2000-04-28	NaN	NaN	23.0	24.4	15.3

Рис. 7. Приклад перших замірів на станції NEMESHAЕVO, UP у Київській обл., початок замірів 2000 р.

За методом Найменших квадратів візуалізуємо лінію тренду випадкового процесу, а також сам випадковий процес, де на осі OX — час (дискретна величина), на осі OY — середньодобова температура (неперервна величина).

Випадковий процес середньодобової температури є стаціонарним у випадку замірів на станціях KIEV, UP і BORYSPIL, UP, має виражену сезонність по роках та порах року. Діаграми зображено на рис. 8–10.

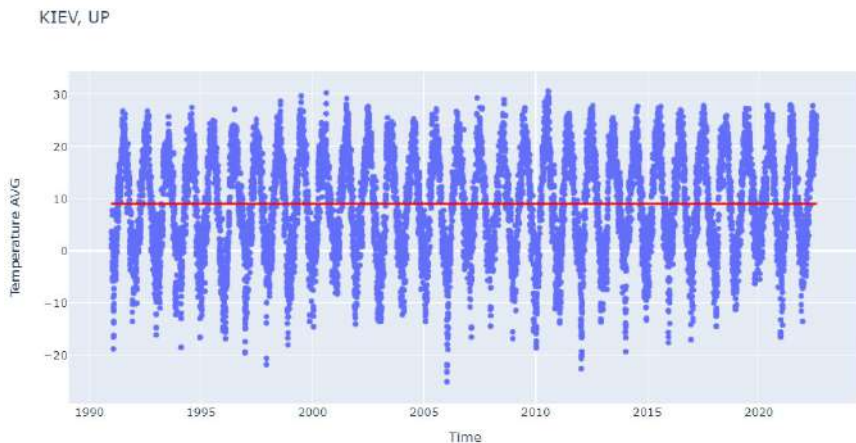


Рис. 8. Показники замірів середньодобової температури на станції KIEV, UP

Джерело: розроблено авторами.

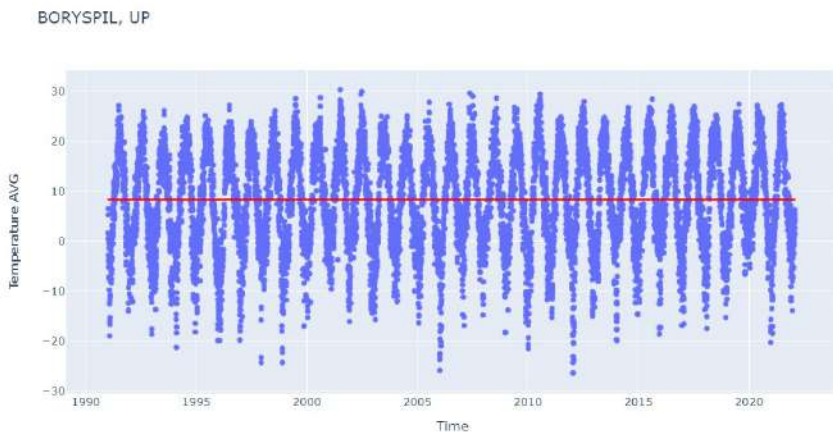


Рис. 9. Показники замірів середньодобової температури на станції BORYSPIL, UP

Джерело: розроблено авторами.

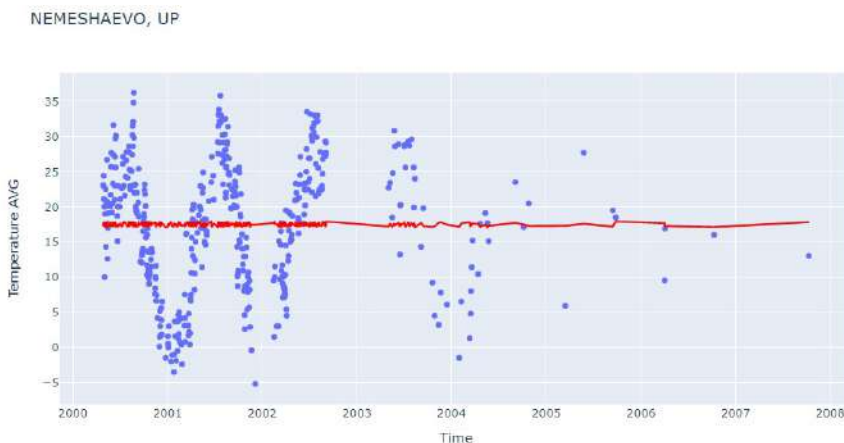


Рис. 10. Показники замірів середньодобової температури на станції NEMESHAЕVO, UP

Джерело: розроблено авторами.

У випадку недостатніх і непостійних замірів на станції NEMESHAЕVO, UP можна сказати, що цей процес не є стаціонарним — заміри припинилися у 2008 р. Для перевірки стаціонар-

ністі випадкового процесу можна використати критерій Дікі-Фуллера та p -value. Якщо p -value менше 0,05, то нульову гіпотезу (відкинути нестационарність) буде відхилено, і цей випадковий процес буде стаціонарним. В іншому випадку він не буде стаціонарним [6].

Результати проведення тесту Дікі Фуллера подано на рис. 11–13.

```
adf, pval, usedlag, nobs, crit_vals, icbest = adfuller(temp1.values) # KIEV, UP station
print('ADF test statistic:', adf)
print('ADF p-values:', pval)
print('ADF number of lags used:', usedlag)
print('ADF number of observations:', nobs)
print('ADF critical values:', crit_vals)
print('ADF best information criterion:', icbest)

ADF test statistic: -7.814145099346631
ADF p-values: 6.942206731164266e-12
ADF number of lags used: 40
ADF number of observations: 11275
ADF critical values: {'1%': -3.4309301143596738, '5%': -2.861796379231577, '10%': -2.5669064655552334}
ADF best information criterion: 52077.163820151254
```

Рис. 11. Результати тесту Дікі-Фуллера для випадкового процесу замірів станції KIEV, UP

Джерело: розроблено авторами.

```
adf, pval, usedlag, nobs, crit_vals, icbest = adfuller(temp2.values) # BORYSPIL, UP station
print('ADF test statistic:', adf)
print('ADF p-values:', pval)
print('ADF number of lags used:', usedlag)
print('ADF number of observations:', nobs)
print('ADF critical values:', crit_vals)
print('ADF best information criterion:', icbest)

ADF test statistic: -7.752913154242219
ADF p-values: 9.91031944752788e-12
ADF number of lags used: 39
ADF number of observations: 10902
ADF critical values: {'1%': -3.4309499670139245, '5%': -2.8618051521469448, '10%': -2.5669111353567367}
ADF best information criterion: 51982.25663392912
```

Рис. 12. Результати тесту Дікі-Фуллера для випадкового процесу замірів станції BORYSPIL, UP

Джерело: розроблено авторами.

Отже, є три випадкові процеси з дискретним часом спостереження. Згідно результату статистичних тестів Дікі — Фуллера, перші два процеси мають чітку стаціонарність по роках. Останній має занадто багато відсутніх значень і є нестационарним, його не можна точно описати з позиції сезонності.

Декомпозиція часових рядів — це статистична задача, яка розбиває часовий ряд на кілька компонентів, кожен з яких представляє одну з базових категорій закономірностей, а саме: тренд, сезонність і залишки [8].

```

adf, pval, usedlag, nobs, crit_vals, icbest = adfuller(temp3.values) # NEMESHAEO, UP station
print('ADF test statistic:', adf)
print('ADF p-values:', pval)
print('ADF number of lags used:', usedlag)
print('ADF number of observations:', nobs)
print('ADF critical values:', crit_vals)
print('ADF best information criterion:', icbest)

ADF test statistic: -2.8267475129688497
ADF p-values: 0.054553482945667635
ADF number of lags used: 3
ADF number of observations: 436
ADF critical values: {'1%': -3.445437655635993, '5%': -2.8681918844944785, '10%': -2.5703132171113543}
ADF best information criterion: 2378.713320634922

```

Рис. 13. Результати тесту Дікі-Фуллера для випадкового процесу замірів станції NEMESHAEO, UP

Джерело: розроблено авторами.

Припустимо, що у нас є дві адитивні моделі, тобто такі, що складаються з лінійного тренду та сезонного циклу з однаковою частотою та амплітудою. Сезонні та залишкові компоненти представлені на рис. 14.

Seasonal aspect of observations

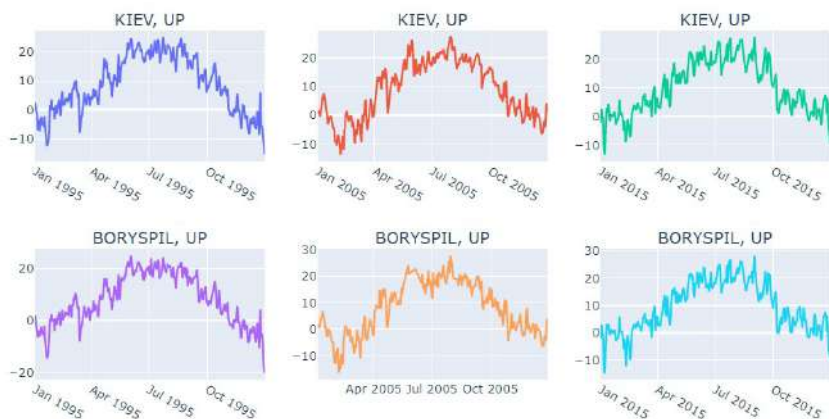


Рис. 14. Сезонний компонент випадкових процесів

Джерело: розроблено авторами.

Усі спостереження мають однакову сезонну структуру: починаючи з низької середньої температури на початку року взимку, потім навесні середня температура все більше зростає, влітку —

під середньодобової температури року, середньодобова осіння температура знижується, в кінці циклу закінчується низькою середньо добовою температурою. На рис. 15 продемонстровано залишковий компонент випадкових процесів.

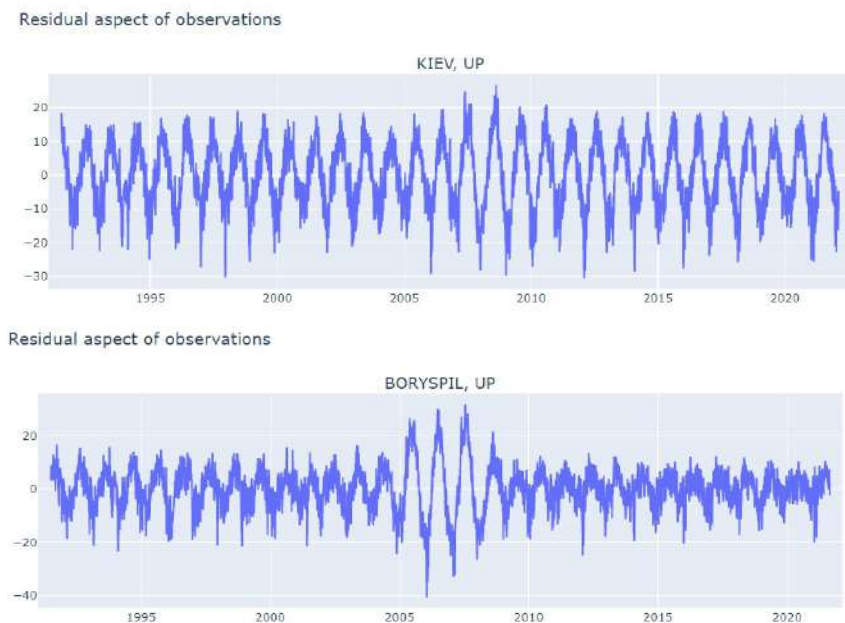


Рис. 15. Залишковий компонент випадкових процесів

Джерело: розроблено авторами.

Типи моделей прогнозування часових рядів [3]:

- *Класичні моделі часових рядів* — це сімейство моделей, які традиційно часто використовуються в багатьох областях прогнозування. Вони здебільшого базуються на часових змінах у часовому ряді та адекватно відображають залежності з однофакторними часовими рядами. Ці моделі зазвичай застосовуються лише до часових рядів і не є корисними для інших типів машинного навчання.

- *Контрольовані моделі* — це сімейство моделей, які використовуються для багатьох завдань машинного навчання. Модель машинного навчання використовує чітко визначені входні змінні та одну або кілька вихідних (цільових) змінних. Контрольовані моделі можна використовувати для часових рядів, за умови, що є спосіб виокремити сезонність і помістити її в змінну.

Експоненційне згладжування — це метод математичного перетворення, який застосовується при прогнозуванні часових рядів. При кожній наступній ітерації враховуються всі попередні значення ряду, але ступінь врахування зменшується за експонентою (s_t — згладжений ряд; c_t — первинний ряд; $\alpha \in (0, 1)$ — коефіцієнт згладжування, який обирається апіорі). Рекурентна формула має вигляд:

$$s_t = \begin{cases} c_1: t = 1 \\ s_{t-1} + \alpha * (c_t - s_{t-1}): t > 1 \end{cases}$$

KIEV, UP, HWES1, HWES2



KIEV, UP, HWES3



Рис. 16. Одиничне, подвійне, потрійне експоненційне згладжування
Джерело: розроблено авторами.

На рис. 16 побудовано графіки відповідно одиничного, подвійного та потрійного експоненційного згладжування. Розроблений авторами на основі потрійного експоненційного згладжування прогноз на рік вперед подано на рис. 17. Використано модель Хольта — Уінтерса. Основні характеристики отриманої моделі за методом Хольта-Уінтерса (коефіцієнта R^2 , похибок MSE, MAE) подано на рис. 18.

KIEV, UP, HWES3_ADD, Prediction



Рис. 17. Прогноз на рік вперед за допомогою потрійного експоненційного згладжування (модель Хольта — Уінтерса)

Джерело: розроблено авторами.

```
In [51]: mean_squared_error(temp1.values[365:730], data) # MSE
```

```
Out[51]: 26.316476334574727
```

```
In [52]: mean_absolute_error(temp1.values[365:730], data) # MAE
```

```
Out[52]: 4.036518487071952
```

```
In [53]: r2_score(temp1.values[365:730], data) # R^2
```

```
Out[53]: 0.7024258165900131
```

Рис. 18. Результати методу Хольта—Уінтерса (коефіцієнта R^2 , похибок MSE, MAE)

Джерело: розроблено авторами.

Бібліотека Prophet — це збірка процедур для прогнозування даних часових рядів на основі адитивної моделі, де лінійні або нелінійні тренди узгоджуються з річною, тижневою та денною сезонністю.

На рис. 19 подано отриманий авторами прогноз на рік складених з використанням бібліотеки Prophet. На рис. 20 показано відповідні результативні характеристики останніх прогнозів.

KIEV, UP, Prophet

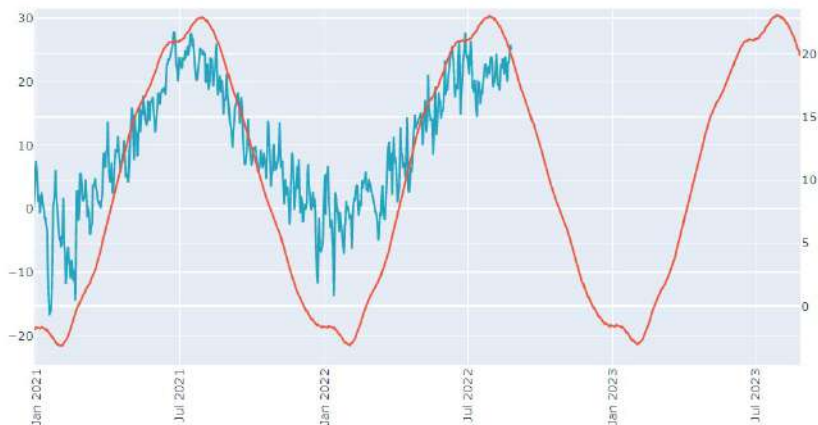


Рис. 19. Прогноз на рік вперед за допомогою бібліотеки Prophet

Джерело: розроблено авторами.

```
In [46]: mean_squared_error(temp1.values, forecast.iloc[0:11316]["yhat"].values) # MSE
Out[46]: 17.296278436536348

In [47]: mean_absolute_error(temp1.values, forecast.iloc[0:11316]["yhat"].values) # MAE
Out[47]: 3.296445347214469

In [48]: r2_score(temp1.values, forecast.iloc[0:11316]["yhat"].values) # R^2
Out[48]: 0.8233631153055813
```

Рис. 20. Результати бібліотеки Хольта-Уінтерса (коефіцієнта R^2 , похибок MSE, MAE)

Джерело: розроблено авторами.

Висновки та пропозиції. Було проведено довгостроковий аналіз середньодобової температури у Київській обл. мовою програмування Python та її бібліотеками для аналізу даних Pandas, Numpy, Scikit-Learn тощо за допомогою моделей експоненційного згладжування, бібліотеки Prophet. За результатами даної роботи можна зробити такі висновки:

- Бібліотека Prophet має вищий показник R^2 і нижчий MSE, MAE відповідно. Незважаючи на те, що Prophet є моделлю half-black box, вона досить добре прогнозує за допомогою керованих методів машинного навчання — модель генералізує дані, а також розпізнає сезонність і стаціонарність.

- Використовуючи моделі класичного прогнозування часових рядів (наприклад, експоненціальне згладжування, метод Хольта — Уінтерса), керованого машинне навчання (Prophet), стає можливим їх масштабування, прогнозування на тривалі періоди часу.

- За допомогою цих моделей можна обробляти та прогнозувати різні типи даних, такі як: температура, рівень опадів, глибина снігового покриву. Крім того, прогнози можуть бути використані для різноманітних цілей людської діяльності.

Підвищення завчасності й справджуваності прогнозів метеорологічних величин, небезпечних і стихійних явищ відкривають виняткові можливості і перспективи математичного моделювання і ще більшого його практичного застосування в задачах попередження і зменшення негативних наслідків від надзвичайних ситуацій природного і техногенного характеру. Математичне моделювання на сучасному етапі є однією з важливих і ефективних систем інформаційного забезпечення, експертного аналізу, обґрунтування організаційних рішень і підвищення рівня керування. Розробка і впровадження безперервних прогнозувальних систем наразі стають все більш актуальними у зв'язку зі впровадженням автоматизованих систем оцінки ризику, прийняття рішень і керування. Однією з таких технік уточнення прогнозу може бути використання машинного навчання.

Особливо важливим, на наш погляд, аспектом даної методології є то, що вона може використовуватись для подальшого дослідження.

Бібліографічні посилання

1. Юрченко М.Є. Прогнозування та аналіз часових рядів. Методичні вказівки до практичних занять та самостійної роботи студентів спеціа-

льності 051 «Економіка» освітня програма «Економічна кібернетика», «Економічна аналітика». Чернівці: ЧНТУ, 2018. 88 с.

2. Дорошенко А.Ю., Бекетов О.Г., Прусов В.А., Тирчак Ю.М., Яценко О.А. Формалізоване проектування та генерація паралельної програми чисельного прогнозування погоди. *Проблеми програмування*. 2014. № 2–3. С. 72–81.

3. Валєєв К.Г., Джалладова І.А. Теорія ймовірностей та теорія випадкових процесів: навч. посібник. Київ:КНЕУ, 2009. 378 с.

4. Прусов В.А., Дорошенко А.Ю. Моделювання природних і техногенних процесів в атмосфері. Київ: Наукова думка, 2006. 542 с

5. Прусов В.А., Сніжко С.І. Математичне моделювання атмосферних процесів. Київ: Ніка-Центр. 2005. 496 с.

6. Shumway, Robert H. a David S. STOFFER. Time Series Analysis and Its Applications: With R Examples. Third Edition. New York: Springer-Verlag, 2011. 606 p.

7. Grimmett, Geoffrey R. a David STIRZAKER. Probability and random processes. 3rd ed. — Oxford: Oxford University Press, 2001. xii. 596 с.

8. Hamilton, James Douglas. Time series analysis. Princeton, N.J.: Princeton University Press, 1994. xiv. 799 с.

9. Random Weather Generator. URL: <https://donjon.bin.sh/d20/weather/>

10. A Monte Carlo model for estimating tornado impacts. URL: <https://rmets.onlinelibrary.wiley.com/doi/full/10.1002/met.1552>

11. NOAA Climate Data online. URL: <https://www.ncei.noaa.gov/cdo-web/>

12. NOAA Climate Data search. URL: <https://www.ncei.noaa.gov/cdo-web/search>

13. NOAA Climate Data station list. URL: <https://www.ncei.noaa.gov/cdo-web/datasets/GHCND/locations/CITY:UP000012/detail>

Статтю подано до редакції 25.11.2022