

УДК 004.9:504.3.054

**В. Є. Копняк; В. Б. Мокін, д-р техн. наук, проф.; С. О. Жуков, канд. техн. наук, доц.;  
І. В. Варчук, канд. техн. наук, доц.; Т. В. Скринник**

## **МЕТОД БУСТИНГУ ГЕТЕРОСКЕДАСТИЧНИХ МОДЕЛЕЙ ДЛЯ ПРОГНОЗУВАННЯ КОНЦЕНТРАЦІЙ ПИЛУ САХАРИ В АТМОСФЕРНОМУ ПОВІТРІ УКРАЇНИ**

*У статті представлено новий метод бустингу гетероскедастичних моделей та його прикладне застосування на прикладі прогнозування концентрації пилу Сахари в атмосферному повітрі України. Останнім часом спостерігається підвищена частота переносу дрібнодисперсного пилу з пустелі Сахара через Середземне море на територію Європи, зокрема – в Україну. Це явище ускладнює прогнозування якості атмосферного повітря через руйнування усталених закономірностей забруднення, оскільки додаються нові чинники, які потребують спеціальних моделей для адекватного опису. Особливий метеорологічний режим поширення пилу Сахари дозволяє припустити, що дисперсія залишків моделі ARIMA може бути випадковим процесом і для її опису доцільно використовувати гетероскедастичні моделі, такі як GARCH. Проте, традиційні GARCH-моделі є ефективними лише за наявності одного домінуючого випадкового процесу. У разі додавання декількох таких процесів, традиційні моделі втрачають ефективність.*

*У цій роботі запропоновано застосування бустингового підходу для побудови ансамблевих моделей, що, на відміну від наявного, включають кооперацію дерев рішень та гетероскедастичних моделей, для моделювання складних гетероскедастичних процесів. Запропонований метод, як це прийнято для бустингових моделей, ґрунтується на ітеративному процесі підбору моделей, де кожна наступна модель враховує похибки попередньої. Для перевірки ефективності методу використано дані громадського моніторингу атмосферного повітря EcoCity, зокрема дані по Вінницькій області за показником PM<sub>1</sub>, які вказують на періоди, коли концентрація дрібнодисперсного пилу Сахари в атмосферному повітрі області досягала особливо аномальних значень.*

*Доведено, що процес поширення пилу Сахари у Вінницькій області є гетероскедастичним. Побудовано моделі SARIMAX та типові GARCH-моделі з використанням Python-бібліотек statsmodels та arch. Виявлено, що модель ARIMA демонструє, на диво, кращі результати у порівнянні з класичними GARCH-моделями з різними параметрами, що свідчить про недостатню ефективність цих GARCH-моделей. Запропонований метод бустингових гетероскедастичних моделей дозволяє досягти значно більшої точності за усі ці моделі на усьому діапазоні значень, окрім значення найбільшої аномалії, яке передбачити неможливо. А отже, розроблений у статті метод прогнозування є ефективним підходом для розв'язання складних задач прогнозування, прикладом яких є прогнозування якості атмосферного повітря під час поширення пилу Сахари в Україні.*

**Ключові слова:** штучний інтелект, часові ряди, гетероскедастичні моделі, GARCH, ARIMA, SARIMAX, бустинг моделей машинного навчання, пил Сахари, якість атмосферного повітря, екологічний моніторинг, прогнозування забруднення, дрібнодисперсний пил, PM<sub>1</sub>, EcoCity, ансамблеві моделі, машинне навчання.

### **Вступ**

Останнім часом, особливості метеорологічних умов призводять до того, що дрібнодисперсний пил з пустелі Сахара переноситься через Середземне море та поширюється на території Європи, зокрема – в Україні. Це фіксується системами як державного, так і громадського моніторингу якості атмосферного повітря.

За даними спостережень 2024 року таке явище спостерігалось у різні періоди з 17 березня по 28 квітня (рис. 1, 2) [1 – 3]. Це призводить до значних наслідків для екосистем, включаючи підвищення концентрацій шкідливих речовин у повітрі, що негативно впливає на здоров'я людей, рослин і тварин. Висока концентрація дрібнодисперсного пилу може пригнічувати ріст рослин, зменшувати видимість і сприяти виникненню респіраторних захворювань у людей. Крім того, пил з Сахари може переносити патогени та інші забруднюючі речовини, що ще більше погіршує стан екосистем [4 – 6].

Ще одним наслідком цього є ускладнення прогнозування якості атмосферного повітря. Домінуючий забруднюючий вплив руйнує усталені закономірності. Для його прогнозування потрібні інші моделі та підходи. Для опису часових рядів традиційно використовуються моделі ARIMA та їх варіації [7]. Але особливий метеорологічний режим поширення пилу Сахари дозволяє припускати, що дисперсія залишків моделі ARIMA може теж бути випадковим процесом, який описується певною моделлю. Для опису таких процесів гарно підходять гетероскедастичні моделі [8, 9]. Однак, процес моделювання ускладнюється тим, що традиційні чинники забруднення не зникають, а – додаються, а отже, потрібні більш складні ансамблеві моделі та методи їх ідентифікації.

**Метою статті** є підвищення точності прогнозування часових рядів з домінуючим впливом з використанням гетероскедастичних моделей на прикладі моделювання концентрацій пилу в атмосферному повітрі України під час активного поширення пилу Сахари. Це дозволить точніше оцінити діапазон значень короткострокового прогнозування якості атмосферного повітря під час цього особливого метеорологічного явища.

### Ідея методу

У роботі [10] зазначено, що особливим класом часового ряду є гетероскедастичний процес з дисперсією, яка змінюється у часі за певним стохастичним законом. Для його прогнозування використовуються спеціальні гетероскедастичні моделі (GARCH-моделі). У тій же роботі [10] наведено огляд таких моделей, ідентифікація яких автоматизована в Python-бібліотеці arch.

Однак, слід розуміти, що GARCH-модель є ефективною, коли має місце один основний випадковий процес зі змінною у часі дисперсією за певною моделлю. Якщо ж має місце сума двох таких процесів, тоді ця модель не буде ефективною. Один закон буде накладатись на інший. Саме такий приклад досліджено у роботі [10], де доведено, що процес зміни показника «PM10» на станції № 650 («Турбів») мережі EcoCity за даними 2020 – 2023 рр. є гетероскедастичним, але показано, що класичні моделі ARCH, GARCH, EGARCH, APARCH, HARCH є менш адекватними, аніж модель ARIMA, яка теж дає чималу похибку. Однією з причин цього може бути те, що станція фіксує не один гетероскедастичний процес, а – декілька одночасно. Для таких складних випадків пропонується будувати ансамблеву модель.

У роботі [11] наведено огляд основних видів ансамблевих моделей машинного навчання: бегінг, бустинг, стекінг, голосування. Якщо розглядати складний процес як сукупність простіших процесів, тоді найкращим варіантом є бустинг. Це – ітеративний процес підбору моделі. На кожній ітерації аналізується нев'язка у передбаченнях за попередньою моделлю. Нова модель намагається врахувати саме цю нев'язку. Результатом є комплекс моделей, які застосовуються послідовно, кожна з яких враховує певні особливості даних та зменшує результуючу похибку. Зазвичай, цей метод застосовується для поєднання простих моделей (наприклад, дерев рішень DecisionTree). У роботі [12] була зроблена спроба побудови ансамблю гетероскедастичних моделей на основі бустингу (метод WildBoostGARCH), але цей підхід не отримав поширення. Автори вважають, що причиною цього є те, що процес бустингу є ефективним, у разі використання простих моделей, а не тільки складних гетероскедастичних.

Основною ідеєю методу, який пропонується, є побудова бустингу на основі кооперації дерев рішень та гетероскедастичних моделей. Цей метод варто застосовувати тільки для гетероскедастичних рядів даних.

Розглянемо найбільш узагальнений випадок, коли задача є регресійною, а не класифікаційною, тобто, коли значеннями ряду є, наприклад, дробові числа. А тоді для бустингу слід використовувати регресійний варіант дерева рішень «DecisionTreeRegressor» (DTR) бібліотеки Sklearn на Python [11]. А як гетероскедастичну модель пропонуємо використовувати один з найбільш узагальнених її видів, який дозволяє враховувати асиметричні зміни дисперсії і враховує знак залишків, що дозволяє по-різному описувати

вплив позитивних і негативних відхилень на волатильність ряду: «ExponentialGARCH» (EGARCH) [10]. Для запуску алгоритму бустингу на нульовій ітерації ( $t = 0$ , де  $t$  – номер ітерації) пропонуємо використовувати одну з найбільш ефективних ансамблевих бустингових моделей – «RandomForestRegressor» (RFR) [11], а далі вже будуть працювати пари моделей «DTR- EGARCH».

### Алгоритм застосування методу

У загальному випадку, вхідні для моделювання дані  $X$  можуть бути таблицею з багатьма ознаками (і бути як часовим рядом, так і – часовим перетином), а цільова ознака  $y$  – це окремий часовий ряд тієї ж довжини. У першому ж наближенні розглянемо найбільш поширений на практиці варіант, коли вхідний датасет  $X$  містить тільки один ряд значень у певні моменти часу, які одночасно є і вхідними даними, і цільовою ознакою  $y$ . Для того, щоб відібрати оптимальний варіант моделі, використовуємо класичний прийом поділу вхідних даних  $X$  на навчальні  $X_T$  і валідаційні  $X_V$ , наприклад, випадковим чином, у співвідношенні 80 % на 20 % [11]. А як тестові, використаємо дані  $y = X$ , щоб проаналізувати як добре оптимальна модель опише весь ряд спостережень. Це – скоріше, задача моделювання, аніж – прогнозування, коли досліджується адекватність моделі усім вхідним даним. Якщо вона – адекватна, тоді її потім можна використовувати і для прогнозування.

Пропонується такий метод прогнозування гетероскедастичних рядів даних  $X$  та цільової ознаки  $y$  на основі класичного алгоритму бустингу, але з урахуванням кооперації дерев рішень DTR іоднієї з гетероскедастичних моделей EGARCH:

1. Ініціалізація залишків  $r$  на нульовій ітерації ( $t = 0$ ):

$$r[0] = y. \quad (1)$$

2. Початкова ітерація:

- Навчання моделі  $f_0(X_T)$  RandomForestRegressor:

$$f_0(X_T) = RFR(X_T, r[0]). \quad (2)$$

- Обчислення прогнозів на нульовій ітерації ( $t = 0$ )  $y_{pred}[0]$  шляхом підставлення в модель (2) валідаційних даних  $X_V$ :

$$y_{pred}[0] = f_0(X_V). \quad (3)$$

- Обчислення залишків на наступній ітерації, де, як це прийнято в моделях бустингу [12], від залишків  $r[0]$  віднімаються прогнози (3) з урахуванням коефіцієнта  $L$ , який часто називають «learning\_rate» («швидкість навчання»):

$$r[1] = r[0] - L y_{pred}[0] = \theta(r[0], y_{pred}). \quad (4)$$

- Навчання моделі EGARCH (EG), в яку підставляються залишки  $r[1]$ :

$$\sigma^2[1] = EG(r[1]). \quad (5)$$

- Коригування залишків з урахуванням дуже малої константи  $\xi$  (наприклад  $10^{-6}$ ), щоб уникнути ділення на нуль у знаменнику дробу – це типовий прийом із [12]:

$$r^*[1] = \frac{r[1]}{\sqrt{\sigma^2[1] + \xi}} = \Omega_1(r[1], \sigma^2[1], \xi). \quad (6)$$

3. Наступні ітерації  $t = 1 \dots N - 1$ :

- Навчання моделі DecisionTreeRegressor:

$$f_t(X_T) = DTR(X_T, r^*[t]). \quad (7)$$

- Обчислення прогнозів:

$$y_{pred}[t] = f_t(X_V). \quad (8)$$

- Обчислення залишків:

$$r[t+1] = r^*[t] - Ly_{pred}[t]. \quad (9)$$

- Навчання моделі EGARCH:

$$\sigma^2[t+1] = EG(r[t+1]). \quad (10)$$

- Коригування залишків:

$$r^*[t+1] = \frac{r[t+1]}{\sqrt{\sigma^2[t+1] + \xi}} = \Omega_{t+1}(\cdot). \quad (11)$$

4. Кінцевий прогноз – зважена сума передбачень на усіх ітераціях:

$$y_\Sigma = \sum_{t=0}^{N-1} Ly_{pred}[t]. \quad (12)$$

5. Оцінювання точності рішення за метрикою MSE («MeanSquareError» – «середньоквадратична помилка»):

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - y_{\Sigma,i})^2. \quad (13)$$

На рис. 1 наведена блок-схема запропонованого алгоритму.

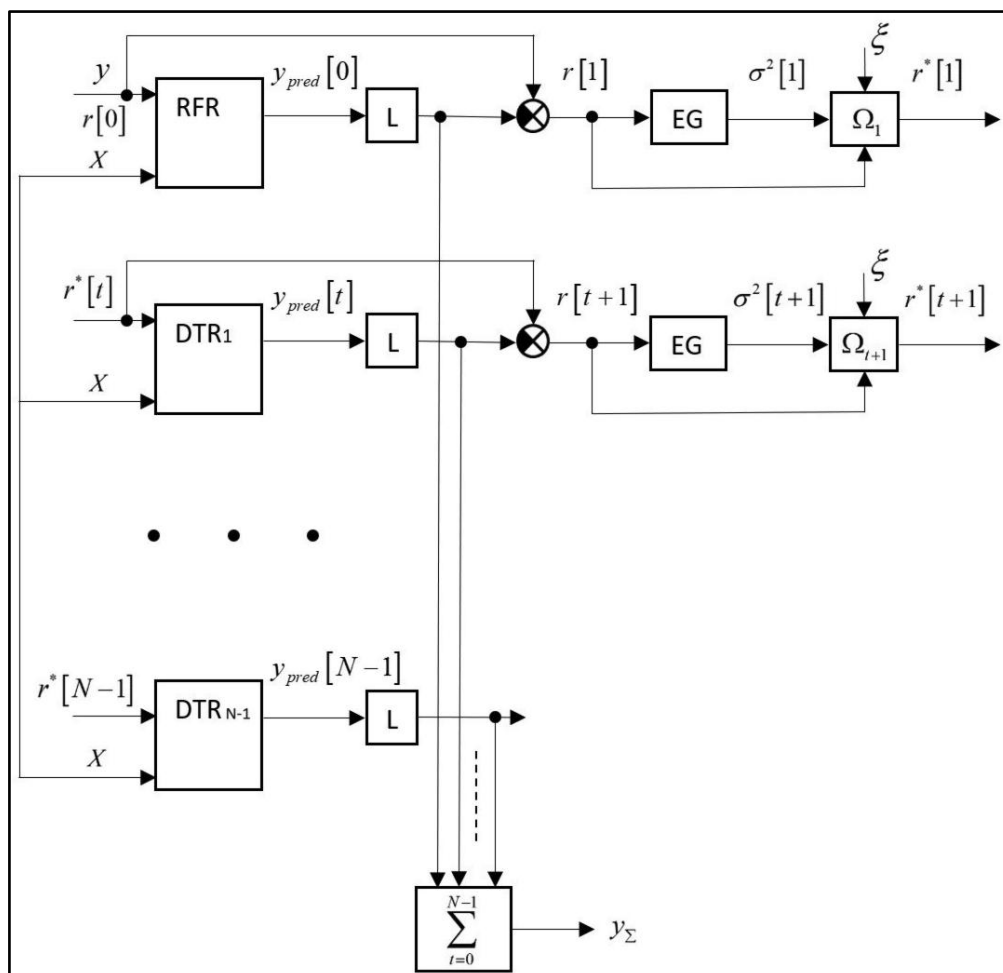


Рис. 1. Блок-схема запропонованого алгоритму методу прогнозування гетероскедастичних рядів даних на основі бустингу та кооперації дерев рішень і гетероскедастичних моделей

Для застосування запропонованого алгоритму на практиці пропонується таке доповнення операцій на початку:

1. За даними спостережень, з урахуванням джерел інформації про специфіку предметної області та об'єкта досліджень, виявити період(и) часу, коли потенційно мав місце домінуючий процес. Зробити вибірку даних за цей час.
2. Перевірити чи є вибірка гетероскедастичним процесом, наприклад з використанням критерію Шапіро-Вілкса [9]. Якщо умова підтвердиться, тоді перейти на п. 3, інакше – в кінець алгоритму.
3. За необхідності, застосувати типові операції передоброблення даних (імпутинг, стандартизація). Поділити дані на навчальні та валідаційні, наприклад, 80 % на 20 %, відповідно.
4. Припасувати бустингову гетероскедастичну модель до цієї вибірки (рис. 1).
5. Здійснити передбачення даних та порахувати метрики та похибки.

Цей алгоритм можна застосовувати до різних показників якості повітря, пов'язаних з поширенням забруднень у повітрі. Найбільш яскраво його можна продемонструвати для прогнозування найбільших концентрацій, пов'язаних з впливом пилу пустелі Сахари на якість атмосферного повітря в Україні. Доведемо це.

### Приклад застосування запропонованого методу для задачі поширення пилу Сахари в атмосферному повітрі в Україні

Аналіз джерел показав, що пил Сахари досяг України в кінці березня 2024 року, поширюючись із західних регіонів до решти країни. Хмара пилу з Сахари накрила Німеччину, Францію та Швейцарію, поширившись далі на Україну. Це явище було помітним, зокрема, у Вінницькій області, де на небі з'явилась біла завіса. Супутникові

зображення також підтверджують поширення пилу Сахари над Україною (рис. 2, 3) [13, 14]. Пилові бурі з Сахари зазвичай містять мікроскопічні частинки, які можуть подорожувати тисячі кілометрів і впливати на якість повітря та видимість.



Рис. 2. Супутниковий знімок станом на 1 квітня 2024 року

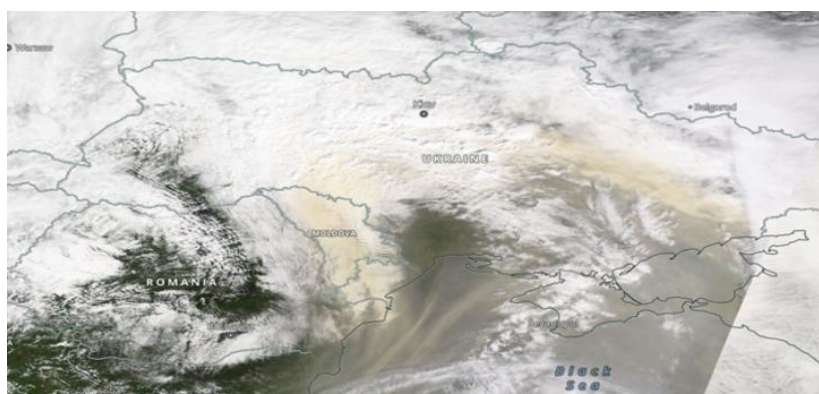


Рис. 3. Супутниковий знімок станом на 24 квітня 2024 року

Було використано дані «Кабінету дослідника якості повітря України» (<https://archive.eco-city.org.ua>) мережі громадського моніторингу атмосферного повітря EcoCity, до якого Вінницький національний технічний університет (ВНТУ) має авторизований доступ. З цього Кабінету дослідника аспірантом ВНТУ Копняком В. Є. було здійснено імпорту даних спостережень по Вінницькій області за показником «PM1» (концентрація найбільш дрібнодисперсного пилу, який переноситься на великі відстані) за період з 25.03.2024 р. по 10.05.2024 р. та завантажено у Kaggle-датасет «AirQualityMonitoringfromEcoCity» [15]. Наприклад, за даними станції № 1315, яка встановлена в науково-дослідній лабораторії екологічних досліджень та екологічного моніторингу кафедри системного аналізу та інформаційних технологій (САІТ) ВНТУ, чітко видно в кінці березня і на початку квітня період часу з аномально великими значеннями (рис. 4), який відповідає поширенню пилу Сахари у Вінницькій області (рис. 2).

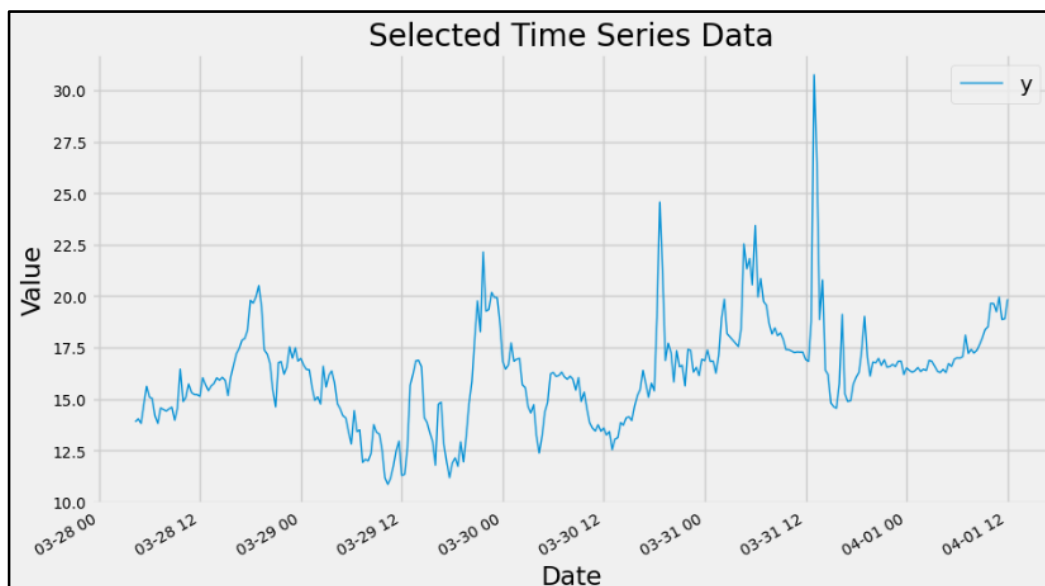


Рис. 4. Графік даних спостережень показника «PM1» на станції № 1315 (м. Вінниця, ВНТУ) мережі EсоCity під час проходження хмари пилу з Сахари в атмосферному повітрі Вінницької області

Аналіз рядів даних з рис. 4 на гетероскедастичність наведено на рис. 5 [16].

```
# Test for heteroskedasticity
arch_test = het_arch(resid, maxlag=lags_max)
shapiro_test = shapiro(st_resid)

print(f'Lagrange multplier p-value: {arch_test[1]}')
print(f'Shapiro-Wilks p-value: {shapiro_test[1]}')

if arch_test[1]>0.05:
    print("Time series is heteroskedastic")
else: print("Time series is homoscedastic")

Lagrange multplier p-value: 0.9262081876457032
Shapiro-Wilks p-value: 1.0658690596798692e-10
Time series is heteroskedastic
```

Рис. 5. Результат перевірки нульової гіпотези щодо гетероскедастичності даних спостережень показника «PM1» на станціях мережі EсоCity під час проходження хмари піску з Сахари в атмосферному повітрі Вінницької області(з авторського Kaggle-ноутбуку [16])

Здійснено стандартизацію даних та побудовано модель SARIMAX з бібліотеки statsmodels.tsa.arima\_model і типові моделі з бібліотеки arch з використанням авторської програми-ноутбуку на базі платформи Kaggle [16] (рис. 6).

model_name	AIC	BIC	params
ARIMA	1060.904542	1079.571248	[1, 0, 2]
EGARCH	1961.595627	1980.246126	[1, 2]
EGARCH	1963.386028	1985.766627	[2, 2]
EGARCH	1970.596960	1989.247459	[2, 1]
EGARCH	1978.098144	1993.018543	[1, 1]
APARCH	1999.483736	2021.864335	[1, 2]
APARCH	2001.205673	2027.316371	[2, 2]
APARCH	2002.252056	2020.902555	[1, 1]
GARCH	2002.445240	2021.095739	[1, 2]
APARCH	2004.252056	2026.632654	[2, 1]
GARCH	2004.445241	2026.825839	[2, 2]
GARCH	2005.004308	2019.924707	[1, 1]
ARCH	2005.652019	2020.572418	[2, 1]
ARCH	2005.652019	2020.572418	[2, 2]
HARCH	2005.652019	2020.572418	[2, 1]
HARCH	2005.652019	2020.572418	[2, 2]
GARCH	2007.004308	2025.654807	[2, 1]
ARCH	2012.279722	2023.470021	[1, 2]
ARCH	2012.279722	2023.470021	[1, 1]
HARCH	2012.279722	2023.470021	[1, 1]
HARCH	2012.279722	2023.470021	[1, 2]

Рис. 6. Точність типових моделей гетероскедастичних рядів даних

Як видно з рис. 6, хоча ряд і є гетероскедастичним, але модель ARIMA демонструє кращі результати, хоча й не дуже гарні. Серед GARCH-моделей кращою є EGARCH-модель. Подібний висновок був обґрунтований й у роботі [10].

На рис. 7 наведено результат ідентифікації різних параметрів моделей з рис. 1, який дає доволі високу точність за метриками  $r2\_score$  (враховує коректність напрямку зростання значень), і за метрикою «MAPE» («Mean Absolute Percentage Error» — середня абсолютна відносна похибка, яка є аналогом відносної похибки у машинному навчанні моделей).

n_estimators	learning_rate	max_depth	rmse	r2_score	mape	n_estimators	learning_rate	max_depth	rmse	r2_score	mape
30	0.038	6	0.525978	0.969829	0.015924	25	0.04	5	0.619013	0.958211	0.0064
30	0.038	5	0.530082	0.969356	0.015963	50	0.02	6	0.638552	0.955532	0.007497
30	0.04	5	0.564941	0.965193	0.021762	25	0.04	7	0.632898	0.956316	0.008254
25	0.04	5	0.619013	0.958211	0.0064	25	0.038	5	0.697322	0.94697	0.008905
30	0.04	7	0.619418	0.958157	0.02373	25	0.038	6	0.693737	0.947513	0.009052
25	0.04	7	0.632898	0.956316	0.008254	25	0.038	7	0.711547	0.944784	0.01001
50	0.02	6	0.638552	0.955532	0.007497	25	0.04	6	0.722393	0.943088	0.010866
30	0.038	7	0.640419	0.955271	0.019301	30	0.038	6	0.525978	0.969829	0.015924
25	0.038	6	0.693737	0.947513	0.009052	30	0.038	5	0.530082	0.969356	0.015963
25	0.038	5	0.697322	0.94697	0.008905	30	0.038	7	0.640419	0.955271	0.019301
25	0.038	7	0.711547	0.944784	0.01001	50	0.02	7	0.925277	0.906631	0.01973
25	0.04	6	0.722393	0.943088	0.010866	30	0.04	5	0.564941	0.965193	0.021762
30	0.04	6	0.742197	0.939925	0.026406	50	0.02	5	1.060899	0.877254	0.022432
50	0.02	7	0.925277	0.906631	0.01973	30	0.04	7	0.619418	0.958157	0.02373
50	0.02	5	1.060899	0.877254	0.022432	30	0.04	6	0.742197	0.939925	0.026406
30	0.02	6	1.492676	0.75701	0.046486	30	0.02	5	1.576136	0.729077	0.044712
30	0.02	7	1.49818	0.755214	0.046474	30	0.02	7	1.49818	0.755214	0.046474
30	0.02	5	1.576136	0.729077	0.044712	30	0.02	6	1.492676	0.75701	0.046486
25	0.02	6	1.74112	0.66939	0.057347	25	0.02	7	1.751656	0.665377	0.057316
25	0.02	5	1.743243	0.668584	0.057425	25	0.02	6	1.74112	0.66939	0.057347
25	0.02	7	1.751656	0.665377	0.057316	25	0.02	5	1.743243	0.668584	0.057425
50	0.038	6	1.867448	0.619675	0.091782	50	0.038	6	1.867448	0.619675	0.091782
50	0.04	7	2.085526	0.525661	0.107327	50	0.038	5	2.147649	0.496981	0.094501
50	0.038	5	2.147649	0.496981	0.094501	50	0.04	5	2.408582	0.367325	0.105215
50	0.04	6	2.355412	0.39495	0.128006	50	0.04	7	2.085526	0.525661	0.107327
50	0.04	5	2.408582	0.367325	0.105215	50	0.04	6	2.355412	0.39495	0.128006
50	0.038	7	40.412197	-177.107771	1.096049	50	0.038	7	40.412197	-177.107771	1.096049

а)

б)

Рис. 7. Точність типових моделей гетероскедастичних рядів даних: а) відсортовані за метрикою  $r2\_score$ ; б) відсортовані за метрикою MAPE[16]

На рис. 8 а, б наведено графіки, які ілюструють точність прогнозування усіх даних вибірки з рис. 1 і які є оптимальними за 2-ма основними метриками з параметрами, які вказані у першому рядку таблиць на рис. 7 а, б, відповідно.



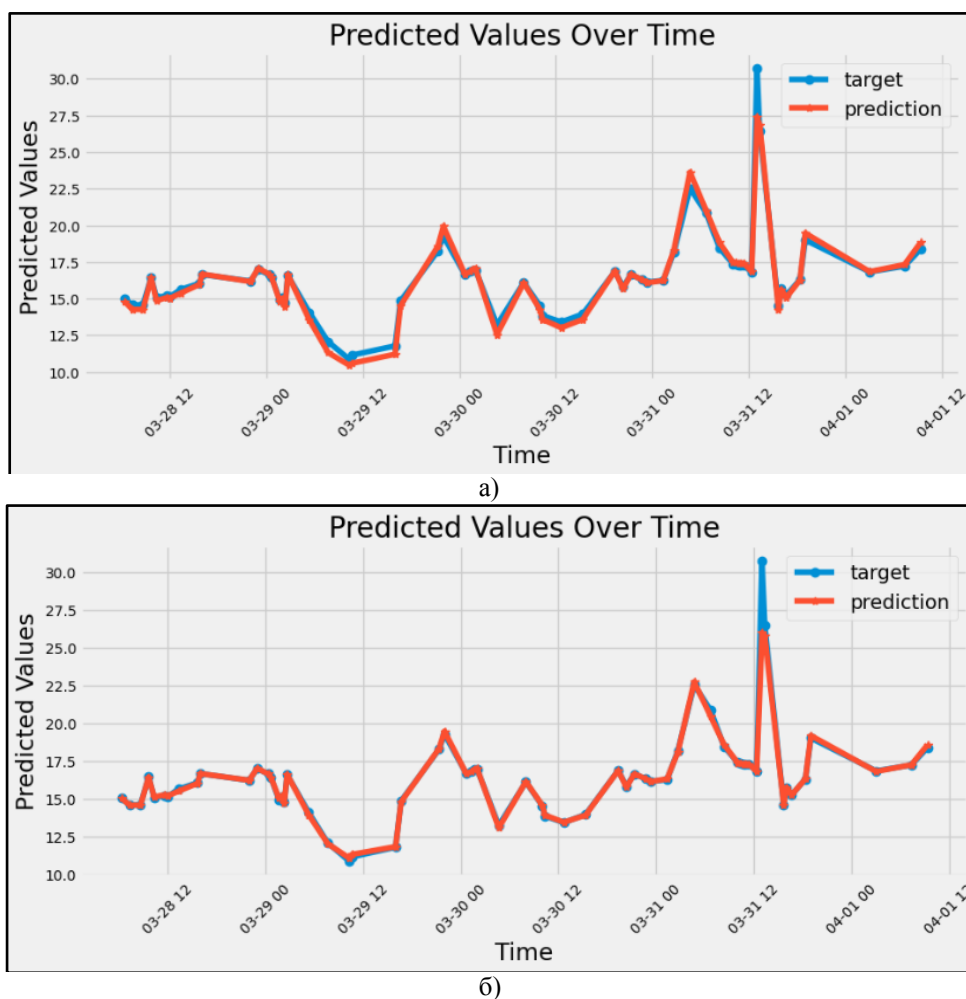


Рис. 8. Результат передбачення даних за заданими та прогнозними значенням: а) модель, оптимальна за метрикою  $r^2\_score$ ; б) модель, оптимальна за метрикою MAPE

Як видно, на рис. 8, вдалось досягти доволі високої точності. Основною причиною відхилення є особливо високе аномальне значення основного піку забруднення атмосферного повітря пилом Сахари 31 березня 2024 р. о 12:00, яке неможливо передбачити цією моделлю.

Отже, запропонований метод довів свою ефективність під час розв'язання складної прикладної задачі.

### Висновки

У статті було представлено метод моделювання та прогнозування концентрацій пилу Сахари в атмосферному повітрі України за допомогою бустингових гетероскедастичних моделей. Проведене дослідження показало, що домінуючий вплив пилу з пустелі Сахара на якість атмосферного повітря значно ускладнює прогнозування традиційними методами через руйнування усталених закономірностей забруднення. Враховуючи це, було запропоновано застосовувати більш складні ансамблеві моделі, що враховують більш складну динаміку процесу під впливом різних факторів забруднення одночасно.

Основною ідеєю запропонованого методу є побудова бустингової моделі на основі кооперації дерев рішень і гетероскедастичних моделей. Бустинговий підхід дозволяє поступово зменшувати похибки прогнозування за рахунок ітеративного підбору моделей, що враховують похибки моделей, побудованих на попередніх ітераціях. Проведений аналіз даних громадського моніторингу EcoCity за показником концентрації дрібнодисперсного пилу «PM1» у м. Вінниці підтвердив ефективність запропонованого методу.

Запропонований метод включає кілька важливих кроків: виявлення періодів, коли має місце вплив домінуючого забруднювального процесу, перевірка вибірки на гетероскедастичність та, у разі позитивного результату перевірки: передоброблення даних, побудова бустингової моделі на основі кооперації дерев рішень і гетероскедастичних моделей та оцінювання метрик точності.

Ефективність методу підтверджується високою адекватністю моделі на усій вибірці даних, де мав місце вплив пилу Сахари в атмосферному повітрі м. Вінниці, що є критично важливим для надійного прогнозування даних та, на його основі – прийняття рішень у сфері екологічного моніторингу і захисту здоров'я населення. Запропонований підхід дозволяє адекватно враховувати вплив складних метеорологічних умов та додаткових факторів забруднення, що є ключовим для розробки ефективних стратегій управління якістю атмосферного повітря.

Таким чином, бустинговий підхід з кооперацією дерев рішень та гетероскедастичних моделей відкриває нові можливості для моделювання та прогнозування складних гетероскедастичних процесів. Цей метод може бути застосований для різних показників якості повітря, що дозволить значно покращити точність прогнозів і, відповідно, ефективність управлінських заходів з реагування на це явище, спрямованих на захист здоров'я населення, в умовах активного поширення пилу Сахари та інших подібних явищ.

#### СПИСОК ЛІТЕРАТУРИ

1. The Copernicus Atmosphere Monitoring Service [Electronic resource]. Access mode : <https://atmosphere.copernicus.eu/climate-atmosphere-podcast-understanding-impact-saharan-dust-storms>.
2. Портал системи громадського моніторингу якості повітря EcoCity [Електронний ресурс]. Режим доступу : <https://eco-city.org.ua/>.
3. WorldView [Electronic resource]. Access mode : <https://worldview.earthdata.nasa.gov>.
4. The Saharan dust plume : Current knowledge on the impact on health, human activities, and the ecosystem, with comments on research gaps [Electronic resource] / Olaoluwa Omoniyi Olarewaju, Olufunke Omowumi Fajinmi, Theophilus Clavell Davies // Authorea Preprints. – 2023. – Access mode : [https://d197for5662m48.cloudfront.net/documents/publicationstatus/131361/preprint\\_pdf/058d51d7471809a08d290dcf8a5e0950.pdf](https://d197for5662m48.cloudfront.net/documents/publicationstatus/131361/preprint_pdf/058d51d7471809a08d290dcf8a5e0950.pdf). DOI:10.22541/essoar.167979167.72582095/v1.
5. Wang Q. The impact of Sahara dust on air quality and public health in European countries / Q. Wang, J. Gu, X. Wang // Atmospheric Environment. – November 2020. – Volume 241. – P. 117771. DOI : <https://doi.org/10.1016/j.atmosenv.2020.117771>.
6. Knight J. Dust from the Central Sahara : Environmental and Cultural Impacts / J. Knight, S. Merlo, A. Zerboni // Landscapes and Landforms of the Central Sahara. World Geomorphological Landscapes. Springer, Cham, 2024. – P. 131 – 141. DOI: [https://doi.org/10.1007/978-3-031-47160-5\\_14](https://doi.org/10.1007/978-3-031-47160-5_14).
7. Методологія та організація наукових досліджень : підручник – вид. 3-є, змін. та доп. / [Б. І. Мокін, О. Б. Мокін, В. Б. Мокін]. – Вінниця : ВНТУ, 2023. – 230 с.
8. How to Model Volatility with ARCH and GARCH for Time Series Forecasting in Python [Electronic resource] / Jason Brownlee // Machine Learning Mastery. Time Series. – 24 August 2018. – Access mode : <https://machinelearningmastery.com/develop-arch-and-garch-models-for-time-series-forecasting-in-python>.
9. Volatility Clustering and GARCH / Nicholas Holloway // Kaggle notebook. – 18 March 2019. – Access mode : <https://www.kaggle.com/code/nholloway/volatility-clustering-and-garch>.
10. Дослідження проблем із гетероскедастичністю даних моніторингу якості атмосферного повітря [Електронний ресурс] / В. Є. Копняк, В. Б. Мокін // Матеріали ЛІ Науково-технічної конференції факультету інтелектуальних інформаційних технологій та автоматизації Вінницького національного технічного університету, Вінниця, 21 – 23 червня 2023 р. – Режим доступу : <https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2023/paper/view/18962/15728>.
11. Наука про дані: машинне навчання та інтелектуальний аналіз даних : електронний навчальний посібник комбінованого (локального та мережевого) використання [Електронний ресурс] / В. Б. Мокін, М. В. Дратований – Вінниця : ВНТУ, 2024. – 258 с. – Режим доступу : <https://iq.vntu.edu.ua/repository/getfile.php/8163.pdf>.
12. Boosting GARCH and neural networks for the prediction of heteroskedastic time series / J. M. Matías, M. Febrero-Bande, W. González-Manteiga [et al.] // Mathematical and Computer Modelling. – 2010. – Volume 51, Issues 3 – 4. – P. 256 – 271. – DOI : <https://doi.org/10.1016/j.mcm.2009.08.013>.
13. Satellite images: Dust from the Sahara reached Ukraine [Electronic resource] / Access mode : <https://universemagazine.com/en/satellite-images-dust-from-the-sahara-reached-ukraine/>.
14. Why is Sahara dust blowing into Europe? [Electronic resource] / Access mode : <https://www.dw.com/en/why-is-sahara-dust-blowing-into-europe/a-68927274>.

15. Air Quality Monitoring from EcoCity [Electronic resource] / V. Mokin, D. Shmundiak, V. Kopniak // Kaggle Dataset. – May 2024. – Access mode : <https://www.kaggle.com/datasets/vbmokin/air-quality-monitoring-from-ecocity>.
16. Sahara's Dust in UA: Boosting GARCH models [Electronic resource] / V. Mokin, V. Kopniak // Kaggle Notebook. – June 2024. – Access mode : <https://www.kaggle.com/code/vbmokin/sahara-s-dust-in-ua-boosting-garch-models/settings>.

Стаття надійшла до редакції 20.06.2024.

Стаття пройшла рецензування 24.06.2024.

**Копняк Володимир Євгенович** – аспірант кафедри системного аналізу та інформаційних технологій, e-mail: [vkopnyak@gmail.com](mailto:vkopnyak@gmail.com).

**Мокін Віталій Борисович** – д-р техн. наук, професор, завідувач кафедри системного аналізу та інформаційних технологій, e-mail: [vbmokin@gmail.com](mailto:vbmokin@gmail.com).

**Жуков Сергій Олександрович** – канд. техн. наук, доцент, доцент кафедри системного аналізу та інформаційних технологій, e-mail: [sazhukov@gmail.com](mailto:sazhukov@gmail.com).

**Варчук Ілона В'ячеславівна** – канд. техн. наук, доцент, доцент кафедри системного аналізу та інформаційних технологій, e-mail: [ilona.varchuk@gmail.com](mailto:ilona.varchuk@gmail.com).

**Скринник Тарас Васильович** – студент гр. 2ICT-20б кафедри системного аналізу та інформаційних технологій, e-mail: [tarasskrynyk1@gmail.com](mailto:tarasskrynyk1@gmail.com).

Вінницький національний технічний університет.