

Міністерство освіти і науки України
Харківський національний університет імені В. Н. Каразіна
Навчально-науковий інститут
«Українська інженерно-педагогічна академія»
Кафедра (автоматизації, метрології та енергоефективних технологій)

КВАЛІФІКАЦІЙНА РОБОТА

магістра

на тему

«Інформаційна система моніторингу довкілля м. Харкова»
(тема кваліфікаційної роботи)

Виконав: студент 2 курсу, групи ДІВс24мг
спеціальності: 175 Інформаційно-вимірю-
(шифр і назва напрямку підготовки, спеціальності)
вальні технології

_____/Едуард РОМАНОВ
(підпис) (ім'я та прізвище)

Керівник _____/Наталія АНТОНЕНКО
(підпис) (ім'я та прізвище)

Рецензент _____/Тетяна ГОНТАР
(підпис) (ім'я та прізвище)

«До захисту допущено»

Завідувач кафедри _____/Геннадій КАНЮК
(підпис) (ім'я та прізвище)

Нормоконтроль _____/Євген КЛЮЧКА
(підпис) (ім'я та прізвище)

Секретар ЕК _____/Євген КЛЮЧКА
(підпис) (ім'я та прізвище)

Харків – 2025 рік

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМ. В.Н. КАРАЗІНА**

Навчально-науковий інститут «Українська інженерно-педагогічна
академія»

Кафедра автоматизації, метрології та енергоефективних технологій

Спеціальність 175 Інформаційно-вимірювальні технології

Освітньо-професійна програма «Якість, стандартизація та сертифікація»

ЗАТВЕРДЖУЮ

Завідувач кафедри

(підпис)

д.т.н., проф. Геннадій КАНЮК

« ____ » _____ 2025 р.

ЗАВДАННЯ

на кваліфікаційну роботу (дипломну роботу/дипломний проєкт)

другого (магістерського) рівня вищої освіти

здобувачу вищої освіти Едуарду РОМАНОВУ

(ім'я, прізвище)

1. Тема «Інформаційна система моніторингу довкілля м. Харкова»
затверджена наказом по університету № 4801-5/3664 від 06.10.2025 р.
2. Термін здачі закінченої роботи «10» грудня 2025 р.
3. Вихідні дані до роботи/проєкту: Закони України, Постанови Верховної Ради, Постанови Кабінету Міністрів, теоретичні та практичні розробки вітчизняних та зарубіжних авторів за темою роботи, періодичні видання, статистичні дані, галузева нормативна документація, технологічна документація.
4. Зміст роботи/проєкту (перелік питань, що їх належить розробити): Вступ. Моніторинг якості довкілля та атмосферного повітря. Об'єкти та методи дослідження. Розрахунки та аналітика. Висновки.
5. Перелік графічного матеріалу (презентаційний матеріал): Презентація, виконана в програмі Microsoft PowerPoint

6. Консультант:

Розділ	Консультант	Підпис, дата		Оцінка (бали)
		Завдання видав	Завдання прийняв	
1-3	К.т.н., доц. Наталія АНТОНЕНКО			

7. Дата видачі завдання «06» жовтня 2025 р.

Керівник роботи

Наталія АНТОНЕНКО
(підпис) (ім'я, прізвище)

Завдання прийняв до виконання

Едуард РОМАНОВ
(підпис) (ім'я, прізвище)

КАЛЕНДАРНИЙ ПЛАН-ГРАФІК
виконання кваліфікаційної роботи
(дипломної роботи/дипломного проєкту)

№ з/п	Назва етапів роботи та питань, які мають бути розроблені відповідно до завдання	Термін виконання	Позначки керівника про виконання завдань
1	Аналітичний огляд публікацій за темою роботи	06.10.2025 – 15.10.2025	вик.
2	Моніторинг якості довкілля та атмосферного повітря	16.10.2025 – 05.11.2025	вик.
3	Об'єкти та методи дослідження	06.11.2025 – 16.11.2025	вик.
4	Розрахунки та аналітика	16.11.2025 – 25.11.2025	вик.
5	Формулювання висновків по роботі. Оформлення пояснювальної записки та презентації	до 10.12.2025	вик.

Здобувач вищої освіти

Едуард РОМАНОВ
(підпис) (ім'я, прізвище)

Нормоконтроль

Євген КЛЮЧКА
(підпис) (ім'я, прізвище)

РЕФЕРАТ

Пояснювальна записка дипломної роботи складає: 62 сторінок, 8 рисунків, 6 таблиць, 22 переліки посилань.

Ключові слова: ІНФОРМАЦІЙНА СИСТЕМА, ЕКОЛОГІЧНИЙ МОНІТОРИНГ, СТАН ПОВІТРЯ, ПРОГНОЗУВАННЯ ЯКОСТІ ПОВІТРЯ, МОДЕЛЬ ARIMA, СТАТИСТИЧНІ МЕТОДИ ОБРОБКИ ДАНИХ, ЗАБРУДНЮЮЧА РЕЧОВИНА

Об'єкт дослідження – інформаційна система екологічного моніторингу довкілля.

Предмет дослідження – модель ARIMA для прогнозування якості повітря на основі статистичних методів обробки даних.

Мета дослідження – розробка інформаційної системи екологічного моніторингу для прогнозування якості повітря.

В роботі зібрано та підготовлено дані про концентрації домішок в атмосфері м. Харків, Україна; побудовано модель ARIMA для кожної із забруднюючих речовин з використанням отриманих даних; оцінено якість моделі ARIMA за допомогою середньоквадратичної помилки (RMSE); візуалізовано фактичні та прогнозовані значення для кожної забруднюючої речовини; виконано прогнозування майбутніх значень концентрацій домішок на основі моделі ARIMA.

Розроблено інформаційну систему екологічного моніторингу, яка поєднує дані про якість повітря у минулому та передбачає майбутні значення якості на основі отриманої моделі. Це дозволяє оперативно визначати стан повітря у м. Харкова та вживати заходів щодо зниження ризиків для здоров'я людей.

ABSTRACT

An explaining message consists of: 62 pages, 8 pictures, 6 tables, 22 lists of references.

Keywords: INFORMATION SYSTEM, ENVIRONMENTAL MONITORING, AIR CONDITION, AIR QUALITY FORECASTING, ARIMA MODEL, STATISTICAL DATA PROCESSING METHODS, POLLUTANT

The object of the study is an information system for environmental monitoring.

The subject of the study is the ARIMA model for air quality forecasting based on statistical data processing methods.

The purpose of the study is to develop an information system for environmental monitoring for air quality forecasting.

The work collected and prepared data on the concentrations of impurities in the atmosphere of Kharkiv, Ukraine; built an ARIMA model for each pollutant using the obtained data; assessed the quality of the ARIMA model using the root mean square error (RMSE); visualized the actual and predicted values for each pollutant; performed a forecast of future values of impurity concentrations based on the ARIMA model.

An information system for environmental monitoring was developed that combines data on air quality in the past and predicts future quality values based on the obtained model. This allows you to quickly determine the state of the air in Kharkiv and take measures to reduce risks to human health.

ЗМІСТ

ВСТУП	8
РОЗДІЛ 1 МОНІТОРИНГ ЯКОСТІ ДОВКІЛЛЯ ТА АТМОСФЕРНОГО ПОВІТРЯ.....	11
1.1 Аналіз загального стану повітряного басейну	12
1.2 Цифрові платформи екологічного моніторингу в Україні	14
1.2.1 SaveEcoBot — громадський інструмент екологічної прозорості .	14
1.2.2 Державна екологічна інспекція України — офіційний регулятор	15
1.2.3 Порівняльна характеристика	16
1.3 Моніторинг якості атмосферного повітря в Україні: виклики, інструменти та локальний контекст Харкова.....	18
1.3.1 Значення моніторингу повітря в умовах урбанізації	18
1.3.2 Інструменти моніторингу: від державного контролю до громадських ініціатив.....	18
1.3.3 Харків: локальний контекст і перспективи	19
Висновки по розділу	20
РОЗДІЛ 2 ОБ'ЄКТИ ТА МЕТОДИ ДОСЛІДЖЕННЯ	22
2.1 Об'єкт дослідження	22
2.2 Метод збирання даних.....	22
2.3 Попередній аналіз даних (EDA)	24
2.3 Кореляційний аналіз	25
2.4 Аналіз часових рядів.....	28
2.5 Моделі для прогнозування.....	28

2.6 Показники для оцінки моделі та продуктивності.....	29
Висновки по розділу	30
РОЗДІЛ 3 РОЗРАХУНКИ ТА АНАЛІТИКА	31
3.1 Попередня обробка даних	31
3.2 Перевірка на стаціонарність	34
3.3 Наслідки та інтерпретація	39
3.4 Побудови моделей	40
3.5 Прогнозування.....	41
3.6 Оцінка ефективності роботи	46
Висновки по розділу	54
ВИСНОВКИ.....	57
ПЕРЕЛІК ПОСИЛАНЬ.....	60

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ І ПОЗНАЧЕНЬ

AQI — індекс якості повітря

ДЕІ — Державна екологічна інспекція України

RMSE — середньоквадратична помилка

EDA — попередній аналіз даних

ARIMA-модель — AutoRegressive Integrated Moving Average model

ADF — розширений тест Дікі-Фуллера

MAE — середня абсолютна помилка

MPE — середня відсоткова помилка

MAPE — абсолютна процентна помилка

AIC — інформаційний критерій Акаїки

BIC — Байєсовський інформаційний критерій

AQI — індекс якості повітря

STL — процедура декомпозиції сезонного тренду на основі Loess

ВСТУП

Прогнозування якості повітря є активною областю досліджень в останні роки гостро постає проблема погіршення якості повітря у великих містах. Це пов'язано зі стрімким збільшенням автотранспорту в межах міста, незадовільною роботою фільтрів паливно-енергетичного комплексу, збільшенням у опалювальний сезон викидів приватного сектора. Одним із показників забрудненості атмосферного повітря є індекс якості повітря (AQI). Аналіз часових рядів, зокрема, широко використовується завдяки своїй ефективності у відображенні часових залежностей екологічних даних.

В даний час різними організаціями, що займаються контролем та оцінкою якістю повітря, активно накопичуються історичні дані метеорологічних показників, включаючи інформацію про забруднення атмосфери. Це призводить до значного збільшення обсягу даних, доступних прогнозування. Разом з тим розвиток програмних і апаратних засобів дає можливість реалізації складних алгоритмів прогнозування. Це дозволить покращити точність прогнозів та задовольнити все більш вимоги до прогнозування погоди та якості повітря.

У низці досліджень використовувалася ARIMA. Наприклад, Кумар і Джайн використовували модель ARIMA для прогнозування рівнів PM_{2,5} у Делі, Індія, [1]. Так само Муніка і Шастри успішно застосували ARIMA для прогнозування забруднення повітря Вишакхапатнаме, Індія, [2]. Інші дослідники розширили застосування традиційних моделей ARIMA. Пірес та інші представили процедуру декомпозиції сезонного тренду на основі Loess (STL) та ARIMA для прогнозування PM₁₀ у Португалії[3]. Астіта М. та інші об'єднали ARIMA з іншими методами машинного навчання для підвищення точності прогнозування якості повітря в Хартфорд, США.[4] Наведені вище дослідження демонструють потенціал аналізу часових рядів та, зокрема, ARIMA для прогнозування якості повітря. Цей проект спрямований на застосування цих методів для прогнозування параметрів якості повітря, вносячи свій внесок у глобальні зусилля з моніторингу та управління довкіллям. У зв'язку з цим

виникає необхідність розробки ефективних методів та інструментів для моніторингу та прогнозування якості повітря.

Наведені у кваліфікаційній магістерській роботі дослідження демонструють потенціал аналізу часових рядів та, зокрема, ARIMA для прогнозування якості повітря. Проект спрямовано на застосування статистичних методів обробки інформації для прогнозування параметрів якості повітря у м. Харків, Україна, вносячи свій внесок у глобальні зусилля з моніторингу та управління довкіллям.

Метою цієї кваліфікаційної магістерської роботи є розробка інформаційної системи екологічного моніторингу, яка використовує модель ARIMA для прогнозування якості повітря на основі історичних даних. Аналіз результатів, отриманих при роботі системи монітерингу дозволить передбачати майбутні параметри якості повітря та вживати заходів щодо зниження ризиків забруднення повітря.

Для досягнення цієї мети було поставлено такі завдання:

- зібрати та підготувати дані про концентрації домішок в атмосфері м. Харків, Україна;
- побудувати модель ARIMA для кожної із забруднюючих речовин з використанням отриманих даних;
- оцінити якість моделі ARIMA за допомогою середньоквадратичної помилки (RMSE);
- візуалізувати фактичні та прогнозовані значення для кожної забруднюючої речовини;
- прогнозувати майбутні значення концентрацій домішок на основі моделі ARIMA.

Наукова та практична новизна та значимість роботи:

У роботі розглядається побудова моделі ARIMA для прогнозування якості повітря. Розроблена інформаційна система екологічного моніторингу поєднує історичні дані про якість повітря та передбачає майбутні значення на основі

моделі. Це дозволяє оперативно визначати стан повітря та вживати заходів щодо зниження ризиків для здоров'я людей.

РОЗДІЛ 1 МОНІТОРИНГ ЯКОСТІ ДОВКІЛЛЯ ТА АТМОСФЕРНОГО ПОВІТРЯ

Проблема забруднення атмосферного повітря є актуальною та потребує безперервної уваги та контролю. Наростаюча кількість автотранспортних засобів, використання систем опалення у приватному секторі та промислові викиди значно впливають на якість повітря у містах та урбанізованих територіях. Щоб досягти поліпшення атмосферного повітря та забезпечити більш здорове довкілля для всіх, необхідно активно здійснювати контроль та проводити дослідження. Застосування сучасних засобів вимірювання та зв'язку, а також нових комп'ютерних технологій відіграє ключову роль у досягненні ефективніших результатів.

У цій роботі розглянуто моніторинг якості повітря, який дозволить відстежувати рівні забруднення та вживати відповідних заходів у разі перевищення нормативів. Система включає комплекс вимірювальної апаратури, програмне забезпечення для збору і обробки даних, а також базу даних для зберігання інформації. Аналіз та порівняння існуючих математичних моделей та чисельних методів у задачах екологічного моніторингу атмосфери, а також збирання відомостей про показники концентрації домішок в атмосфері місті Харків є важливим етапом дослідження. Отримані дані можуть бути широко використані в галузі екології та охорони навколишнього середовища, а також є основою для подальших досліджень та розробок, спрямованих на покращення якості атмосферного повітря в населених пунктах.

Актуальністю магістерської кваліфікаційної роботи «Інформаційна система моніторингу довкілля м. Харкова» обумовлена тим, що інформаційна система дозволяє розраховувати концентрацію забруднюючих речовин, розрахувати комплексний індекс забруднення

промислового об'єкта, зберігати дані, аналізувати стан приземного повітряного шару, прогнозувати можливу концентрацію забруднюючих речовин на забруднених речовинах.

1.1 Аналіз загального стану повітряного басейну

Забруднення атмосферного повітря – привнесення у нього чи виникнення у ньому нових (зазвичай не характерних йому) шкідливих хімічних, фізичних, біологічних агентів. Воно може бути природним (природним) та антропогенним (техногенним). «Природне забруднення повітря викликано природними процесами (вулканічна діяльність, вітрова ерозія, масове цвітіння рослин, дим від лісових і степових пожеж та інших.)» Антропогенне забруднення пов'язані з викидом забруднюючих речовин, у результаті діяльності. За агрегатним станом викиди шкідливих речовин в атмосферу класифікуються наступним чином:

- 1) газоподібні (діоксид сірки, оксиди азоту, оксид вуглецю, вуглеводні та ін.);
- 2) рідкі (кислоти, луги, розчини солей та ін.);
- 3) тверді (важкі метали, канцерогенні речовини, органічний та неорганічний пил, сажа, смолисті речовини та ін.).

Якщо ж розглянути екологічну ситуацію міста Харкова та області, то рівень забруднення атмосфери, незважаючи на скорочення виробництва, залишається досить високим. Найбільший рівень забруднення повітря спостерігається у промисловому районі міста – Індустріальному.

В даний час у плані аналізу та оцінки екологічних та техногенних небезпек виняткова роль відводиться системі екологічного моніторингу. У цій галузі для прогнозування розвитку екологічно небезпечних ситуацій недостатньо старої практики, що ґрунтується на спостереженні, накопиченні даних та складанні бюлетенів забруднення навколишнього

середовища. Необхідний оперативний екологічний контроль, а це, отже, потрібна нова стратегія та нові методи, які дозволять концентрувати увагу на найближчих та майбутніх тенденціях та першорядних завданнях. В роботі А.К. Муртазов [6] дано визначення, екологічний моніторинг – інформаційна система спостережень, оцінки та прогнозу змін у стані навколишнього середовища, створена з метою виділення антропогенної складової цих змін на тлі природних процесів. Говорячи про систему екологічного моніторингу, ми маємо на увазі, що вона має накопичувати, систематизувати та аналізувати інформацію: про стан навколишнього середовища; про причини спостережуваних та ймовірних змін стану (тобто про джерела та фактори впливу); про допустимість змін та навантажень на середовище в цілому. Відповідно до наведених визначень і покладених на систему функцій, можна виділити три основні напрямки діяльності, які включає

- екологічний моніторинг:
- спостереження за факторами впливу і станом середовища; □ оцінку фактичного стану середовища;
- прогноз стану навколишнього природного середовища та оцінку прогнозованого стану.

Необхідно також зазначити, що сама система моніторингу не включає діяльність з управління якістю середовища, але є джерелом інформації необхідної для ухвалення екологічно значимих рішень. Системний метод екологічного моніторингу ґрунтується на експертизі екологічного впливу шкідливих викидів на навколишнє середовище та забезпечує комплексний облік вимірювань та зіставлення їх зі стандартними показниками, вираженими через якісні та кількісні характеристики екологічної безпеки. Метод оцінки екологічних впливів розробки системи містить комплекс заходів, які включають ідентифікацію, аналіз, стеження і моніторинг

екологічних ризиків від своїх запланованих значень. При цьому інформаційна модель екологічних впливів шкідливих речовин включає ідентифікацію: явних і неявних небезпек для людини, природи і суспільства в цілому. Використання сучасних інформаційних технологій у цій галузі дозволяє розробити автоматизовану систему екологічного моніторингу довкілля задля забезпечення екологічної безпеки різноманітних об'єктів і людей. Екологічний моніторинг, що мається на увазі в даній роботі, є системою стаціонарних пунктів спостереження за станом забруднення атмосфери на території м. Харкова, зокрема за поширенням забруднюючих речовин від стаціонарних джерел забруднення, тобто від промислових підприємств. Як відомо, частка забруднення атмосфери промисловими підприємствами досить велика, вони є одними з основних забруднювачів, а відповідно безпосередньо впливають на загальний стан приземного шару атмосферного повітря, на стан здоров'я жителів того чи іншого населеного пункту.

1.2 Цифрові платформи екологічного моніторингу в Україні

SaveEcoBot та сайт Держекоінспекції України — це два ключові інструменти екологічної прозорості в Україні. Перший орієнтований на громадянську обізнаність та онлайн-моніторинг, другий — на державний контроль та звітність. У цьому підрозділі надано докладний огляд їхніх можливостей та особливостей.

1.2.1 SaveEcoBot — громадський інструмент екологічної прозорості

У сучасних умовах зростаючої урбанізації та промислового навантаження на довкілля, питання доступу до екологічної інформації набуває особливої актуальності. Одним із найуспішніших прикладів

цифрової екологічної ініціативи в Україні є платформа SaveEcoBot, створена ГО «SaveDnipro».

Цей сервіс забезпечує відкритий доступ до даних про якість повітря, радіаційний фон, екологічні дозволи та перевірки підприємств. Його головна перевага — інтерактивність і доступність для широкого загалу. Користувачі можуть переглядати карту забруднення в реальному часі, підписуватися на сповіщення у Telegram або Viber, а також подавати скарги на екологічні порушення через чат-бот.

Особливої уваги заслуговує інтеграція з державними реєстрами, що дозволяє відстежувати діяльність підприємств, які мають дозвіл на викиди, або були об'єктами екологічних перевірок. Таким чином, SaveEcoBot виконує функцію громадського контролю, сприяючи прозорості та підзвітності.

1.2.2 Державна екологічна інспекція України — офіційний регулятор

На противагу громадському інструменту, Державна екологічна інспекція України (ДЕІ) є центральним органом виконавчої влади, що здійснює державний контроль у сфері охорони навколишнього природного середовища. Її діяльність регламентується законодавством України, а сайт www.dei.gov.ua слугує офіційним джерелом екологічної інформації.

На сайті ДЕІ публікуються:

- результати перевірок підприємств;
- статистика екологічних порушень;
- нормативні документи та методики контролю;
- інтерактивні карти стану довкілля;
- електронні сервіси для подання звернень та запитів.

Хоча оновлення даних відбувається не в реальному часі, як у SaveEcoBot, офіційний статус інформації забезпечує її достовірність та

юридичну силу. Сайт орієнтований переважно на фахівців, проте доступний і для громадян, які бажають подати скаргу або ознайомитися з екологічною ситуацією у своєму регіоні.

1.2.3 Порівняльна характеристика

Порівняльну характеристику SaveEcoBot та сайту Держекоінспекції України — двох ключових інструментів екологічної прозорості в Україні наведено у табл. 1.1.

Таблиця 1.1

Порівняльна характеристика SaveEcoBot та сайту Держекоінспекції
України

Критерій	SaveEcoBot	ДЕІ
Джерело даних	Сенсори, відкриті реєстри	Офіційні перевірки
Оновлення	Щогодини	Після завершення перевірок
Орієнтація	Громадяни, активісти	Фахівці, підприємства
Інтерактивність	Висока (карта, бот, сповіщення)	Середня (електронний кабінет)
Юридичний статус	Громадська ініціатива	Державний орган

Обидві платформи — SaveEcoBot і сайт ДЕІ — відіграють важливу роль у формуванні екологічної свідомості та забезпеченні доступу до інформації. Їхнє поєднання дозволяє громадянам не лише реагувати на екологічні загрози, а й впливати на прийняття рішень, що стосуються довкілля. У контексті цифрової трансформації державного управління, такі інструменти є прикладом ефективної взаємодії між громадянським суспільством і владою.

1.3 Моніторинг якості атмосферного повітря в Україні: виклики, інструменти та локальний контекст Харкова

1.3.1 Значення моніторингу повітря в умовах урбанізації

Атмосферне повітря — це не просто середовище, в якому ми живемо. Це основа нашого здоров'я, добробуту та екологічної стабільності. У великих містах, таких як Харків, які мають потужну промислову базу, розвинену транспортну інфраструктуру та щільну забудову, моніторинг якості повітря стає не розкішшю, а необхідністю.

Забруднення повітря дрібнодисперсними частинками (PM2.5, PM10), оксидами азоту, озоном та іншими речовинами має прямий вплив на захворюваність населення, особливо серед дітей, літніх людей та осіб з хронічними хворобами. Тому своєчасне виявлення перевищень норм — це не просто технічна задача, а питання соціальної відповідальності.

1.3.2 Інструменти моніторингу: від державного контролю до громадських ініціатив

В Україні існує дві основні моделі екологічного моніторингу: державна та громадська.

Державна модель — через ДЕІ

Державна екологічна інспекція України (ДЕІ) здійснює перевірки підприємств, контролює дотримання екологічних норм, публікує звіти про викиди та порушення. Її сайт www.dei.gov.ua містить розділи з публічною інформацією, нормативними документами та результатами інспекцій. Однак, оновлення даних відбувається після завершення перевірок, що не дозволяє оперативно реагувати на зміни екологічної ситуації.

Громадська модель — через SaveEcoBot.

SaveEcoBot — це цифрова платформа, яка збирає дані з понад 500 сенсорів по всій Україні, включаючи Харків. Вона дозволяє:

- переглядати карту забруднення в реальному часі;
- отримувати сповіщення про перевищення норм;
- подавати скарги на екологічні порушення;
- перевіряти дозвільні документи підприємств.

Це приклад громадської екологічної демократії, коли кожен мешканець може не лише спостерігати, а й діяти.

1.3.3 Харків: локальний контекст і перспективи

Харків — друге за населенням місто України, з потужним промисловим спадком. Тут розташовані підприємства машинобудування, хімічної промисловості, енергетики. У поєднанні з інтенсивним автомобільним рухом це створює високий ризик забруднення атмосфери.

У 2024 році Харківська міська рада затвердила Програму державного моніторингу атмосферного повітря агломерації «Харків» на 2024–2028 роки. Вона передбачає:

- створення нових пунктів спостереження;
- запуск пересувної лабораторії;
- впровадження аналітичної системи обробки даних;
- оцінку ефективності заходів щодо зменшення викидів.

Це свідчить про перехід від реактивної до проактивної екологічної політики, що є позитивним сигналом для міста.

Попри наявність інструментів, більшість громадян не усвідомлює реального стану повітря, яким вони дихають щодня. Чи відомо, що рівень PM_{2.5} у центрі Харкова може перевищувати норму вдвічі у години пік? Чи відомо, що навіть «чисте» на вигляд повітря може містити озон або формальдегід у небезпечних концентраціях?

Це питання не лише до влади, а й до мешканців міста. Моніторинг — це не лише технічна процедура, а форма екологічної свідомості. Коли інформація є відомою, можна діяти: обирати маршрути, змінювати поведінку, вимагати змін.

Моніторинг якості повітря — це ключовий інструмент екологічної безпеки. У Харкові, як і в інших містах України, він має розвиватися на основі поєднання державного контролю, громадської участі та цифрових технологій. Лише тоді мешканці зможуть не просто реагувати на екологічні загрози, а передбачати їх, запобігати та формувати здорове міське середовище.

Висновки по розділу

У першому розділі магістерської кваліфікаційної роботи обгрунтовано актуальність дослідження та сформульовані задачі. Ретельно розглянуто поняття оперативного екологічного контролю та системи екологічного моніторингу. Обгрунтовано нову стратегію та нові методи, які дозволять концентрувати увагу на найближчих та майбутніх тенденціях та першорядних завданнях.

Визначено основні напрямки діяльності у галузі екологічного моніторингу.

Обґрунтовано, що системний метод екологічного моніторингу ґрунтується на експертизі екологічного впливу шкідливих викидів на навколишнє середовище та забезпечує комплексний облік вимірювань та зіставлення їх зі стандартними показниками, вираженими через якісні та кількісні характеристики екологічної безпеки. Проаналізовано комплекс заходів щодо оцінки екологічних впливів.

Проведено порівняльний аналіз двох ключових інструментів екологічної прозорості в Україні. Обґрунтовано, що моніторинг якості повітря є ключовим інструментом екологічної безпеки.

РОЗДІЛ 2 ОБ'ЄКТИ ТА МЕТОДИ ДОСЛІДЖЕННЯ

2.1 Об'єкт дослідження

Основна увага в цьому дослідженні приділяється проаналізованам даним, отриманим з платформи, яка надає детальну інформацію про атмосферні умови у місті Харків. Набір даних містить середньодобові вимірювання забруднюючих речовин PM_{2.5}, PM₁₀, NO₂ та CO. Зазначені забруднюючі речовини є основними маркерами забруднення атмосфери і потенційно можуть призвести до значних наслідків для здоров'я населення. Метою даного дослідження є аналіз структури та тенденцій забруднення з метою отримання всебічного уявлення про умови якості повітря у місті Харків. Кінцева мета полягає в тому, щоб допомогти у розробці ефективних стратегій боротьби із забрудненням повітря та охорони здоров'я населення. У роботі було використано поєднання дослідницького аналізу даних, аналізу часових рядів, підходів до прогнозування та оцінки ефективності з використанням RMSE (середньоквадратична помилка) як основний показник для досягнення цілей дослідження.

2.2 Метод збирання даних

Дані про якість повітря, використані у цьому дослідженні, були отримані із сайту aqicn.org. Сайт надає актуальну інформацію про якість повітря в режимі реального часу для багатьох місць по всьому світу, включаючи Харків, Україна, як показано на рис. 2.1. Дані включають вимірювання різних забруднюючих речовин, які є найважливішими показниками загальної якості повітря.

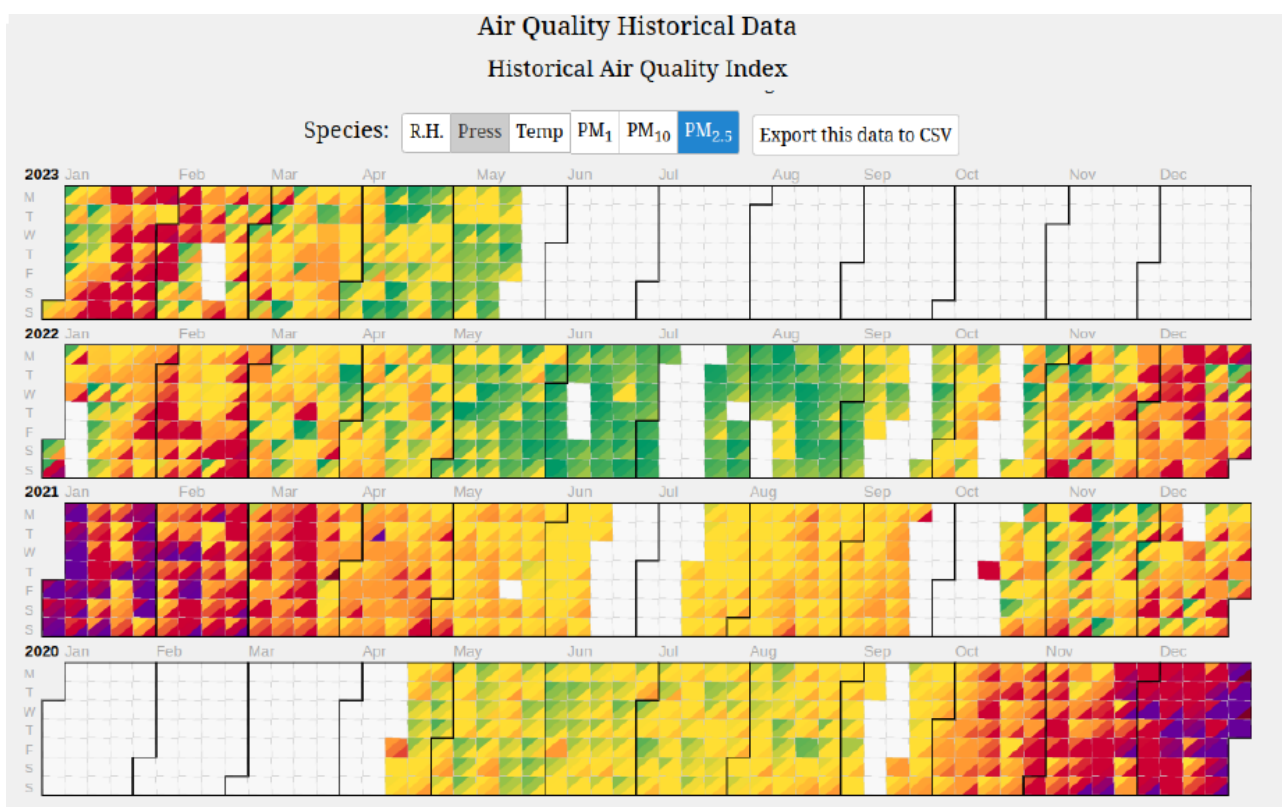


Рис. 2.1 Дані з одного з датчиків, розташованих у м. Харків

Для автоматизації процесу отримання даних був використаний Selenium, потужний інструмент управління веб-браузерами за допомогою програм і автоматизації завдань браузера. Selenium використовувався для навігації по веб-сайту, вибору відповідних параметрів та часових рамок та, нарешті, для завантаження даних про якість повітря. Використання Selenium для веб-скрейпінгу забезпечує ефективний та систематичний збір даних. Після завантаження дані були імпортовані в pandas DataFrame для подальшої обробки та аналізу. Отриманий DataFrame включає колонки для дати спостереження та показників чотирьох забруднюючих речовин (PM2.5, PM10, NO2 та CO).

	date	pm25	pm10	no2	co
0	2023-05-01	39.0	26.0	17.0	4.0
1	2023-05-02	49.0	23.0	17.0	2.0
2	2023-05-03	46.0	12.0	19.0	3.0
3	2023-05-04	35.0	10.0	23.0	3.0
4	2023-05-05	41.0	15.0	22.0	3.0

Рис. 2.2 Таблиця із даними зібраними у травні 2023 року

2.3 Попередній аналіз даних (EDA)

На першій стадії цього дослідження було проведено попередній аналіз даних, щоб мати повне уявлення про набір даних про якість повітря. Було розглянуто розподіл даних, зведену статистику та візуалізацію забруднюючих речовин PM2.5, PM10, NO2 та CO. Завдяки цьому дослідженню ми змогли виявити будь-які відсутні цифри, викиди чи розбіжності у даних, які вимагали попередньої обробки чи додаткових

досліджень. Щоб узагальнити основні тенденції та мінливість даних про забруднювачі, використовувалися статистичні інструменти, включаючи середнє значення, медіану, стандартне відхилення та прямокутні графіки. Крім того, гістограми, лінійні графіки та точкові діаграми допомогли виявити тенденції, закономірності та можливі зв'язки між забруднювачами.

2.3 Кореляційний аналіз

Лінійний кореляційний аналіз дозволяє встановити прямі зв'язки між змінними величинами за абсолютними значеннями. Формула розрахунку коефіцієнта кореляції побудована в такий спосіб, що й зв'язок між ознаками має лінійний характер, коефіцієнт Пірсона точно встановлює тісноту зв'язку. Тому він називається також коефіцієнтом лінійної кореляції Пірсона.

У загальному вигляді формула для підрахунку коефіцієнта кореляції така:

$$r_{xy} = \frac{\sum(x_i - M_x)(y_i - M_y)}{\sqrt{\sum(x_i - M_x)^2(y_i - M_y)^2}} \quad (2.1)$$

де x_i - значення, що приймаються змінною X ;

y_i - значення, що приймаються змінною Y ;

M_x - середня по X ;

M_y - середня по Y .

Перш ніж перейти до моделі ARIMA, було проведено аналіз кореляцій між змінними. Наведена нижче кореляційна матриця відображає коефіцієнти кореляції Пірсона [7], які варіюються від -1 до 1. Коефіцієнт, близький до 1, вказує на сильну позитивну кореляцію, тоді як коефіцієнт

близький до -1 вказує на сильну негативну кореляцію. Ця таблиця, яку представлено на рис. 2.3, дає цінну інформацію про взаємозв'язку між чотирма забруднювачами.

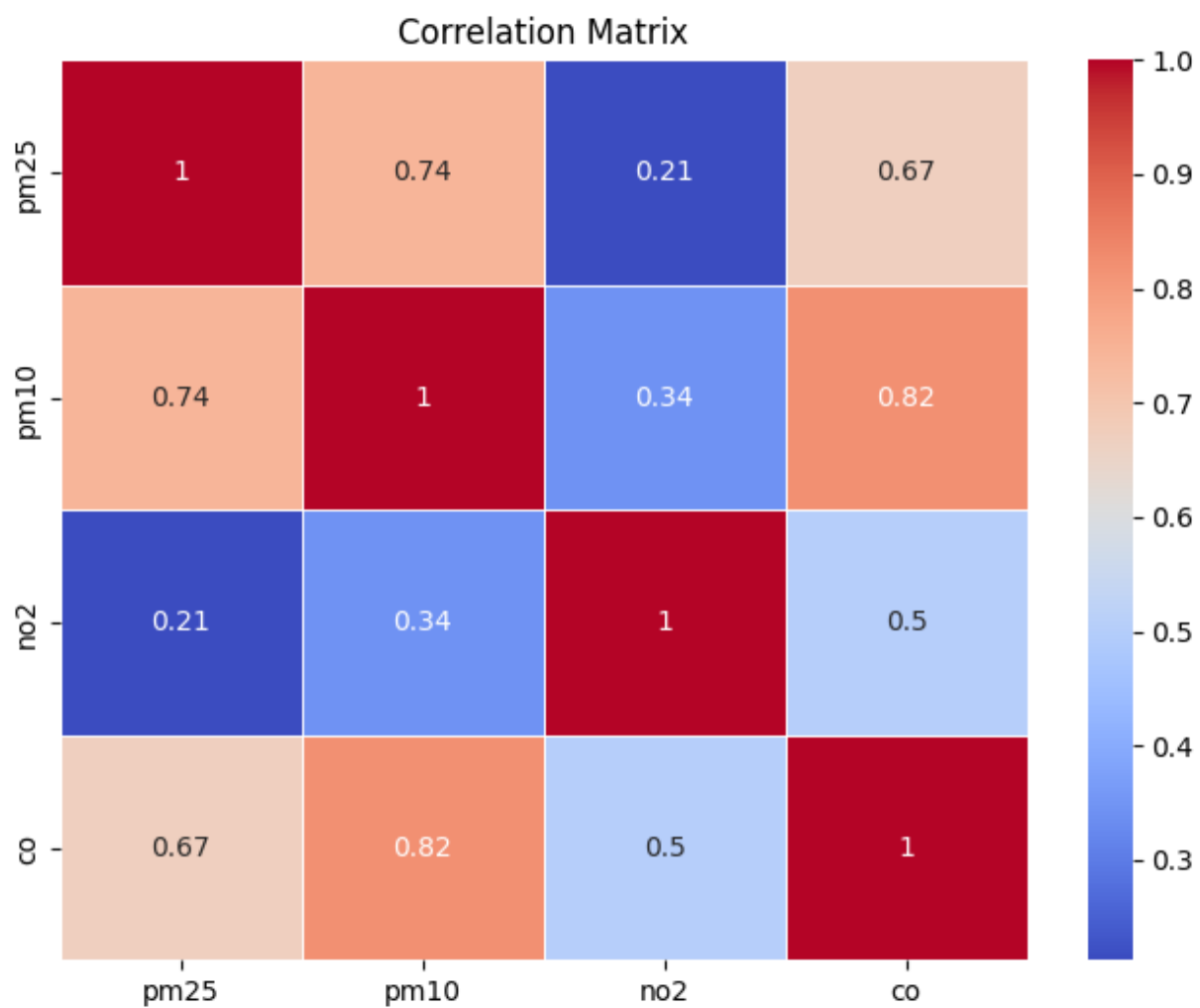


Рис. 2.3 Теплова карта між цільовими значеннями

2.4 Аналіз часових рядів

Використані підходи до аналізу часових рядів для вивчення основних закономірностей, тенденцій та залежностей забруднюючих речовин PM_{2.5}, PM₁₀, NO₂ та CO через тимчасовий характер даних про якість повітря. Щоб зрозуміти взаємозалежність та ефекти запізнення всередині часових рядів, ми досліджували функції автокореляції та часткової автокореляції. Крім того, ми використовували статистичний аналіз та візуальний огляд графіків часових рядів, щоб оцінити стаціонарність даних. Щоб встановити стаціонарність, ми використовували відповідні перетворення чи розмежування за необхідності. Методи декомпозиції були використані при аналізі сезонності для визначення сезонних елементів та закономірностей у даних. Цей етап полегшив вибір моделей, які належним чином відбивали тимчасову динаміку забруднюючих речовин.

2.5 Моделі для прогнозування

ARIMA-моделі (AutoRegressive Integrated Moving Average model), або інтегровані моделі авторегресії-ковзного середнього, вперше запропоновані Боксом і Дженкінсом [8], найбільш популярні та ефективні статистичні моделі для прогнозу часових рядів. В їх основі лежить фундаментальний принцип, що майбутні значення тимчасового ряду генеруються деякою лінійною функцією минулих спостережень і випадковою помилкою (білим шумом). В силу того, що інерційність у часових рядах забруднень велика і, незважаючи на появу штучних нейронних мереж, ARIMA-моделі широко використовують для прогнозу забруднень. Так, наприклад, [9] ARIMA-техніка застосована для короткострокового (максимум на 24 години) прогнозу концентрації CO в атмосферному повітрі, в [10] запропоновані

сезонні ARIMA-моделі для прогнозу індексу якості повітря, в [11] ARIMA-техніка застосована для прогнозу забруднення повітря NO₂, PM_{2.5}.

У цій роботі ми використовували складні методи прогнозування, особливо моделі ARIMA (Авторегресійне інтегроване ковзне середнє), для прогнозування майбутніх рівнів забруднення в місті Харків. Модель ARIMA включала компоненти авторегресії та ковзного середнього. Прогнози були складені з використанням моделей після навчання з використанням попередніх даних про забруднення. Ми змогли передбачити рівні забруднюючих речовин у бажані майбутні проміжки часу завдяки методу прогнозування, який включав повторні обчислення на основі значень і параметрів моделі, що спостерігалися раніше.

Загалом ARIMA(p, d, q) модель має форму

$$\phi(B)(1-B)^d y_t = \theta(B)\varepsilon_t,$$

де y_t – значення модельного ряду на момент часу t ;

ε_t – випадкова помилка з нульовим середнім та постійною дисперсією;

$\phi(B)$, $\theta(B)$ – поліноми ступеня p та q ;

B – лаговий оператор.

2.6 Показники для оцінки моделі та продуктивності

Ефективність наших моделей прогнозування оцінювалась із використанням статистики RMSE (середньоквадратична помилка). Середній розмір розбіжностей між очікуваним та фактичним рівнями забруднення оцінили RMSE. На кращу продуктивність моделі за умови правильного обліку мінливості даних вказували нижчі значення RMSE. Зіставивши прогнозовані значення з фактичними спостереженнями з набору

тестових даних, який був відокремлений від навчальних даних, ми змогли перевірити моделі. Завдяки методу оцінки ми змогли оцінити точність та надійність моделей прогнозування, а також їхню застосовність для прогнозування майбутніх рівнів забруднення.

Висновки по розділу

У другому розділі магістерської кваліфікаційної роботи розглянуто об'єкти та методи дослідження, ретельно розглянуті його основні етапи.

Обґрунтовано ефективність використання статистичних інструментів, включаючи середнє значення, медіану, стандартне відхилення прямокутні графіки, гістограми, лінійні графіки та точкові діаграми для виявлення тенденцій, закономірностей та можливих зв'язків між забруднювачами повітря.

РОЗДІЛ 3 РОЗРАХУНКИ ТА АНАЛІТИКА

У цьому розділі наведено розрахунки та аналізи, проведені на основі даних про якість повітря, отриманих з aqicn.org для міста Харків. Аналітичний процес складається з кількох етапів, а саме попередньої обробки даних, оцінки параметрів моделі, прогнозування та оцінки продуктивності.

3.1 Попередня обробка даних

Перш ніж розпочати розрахунки та аналітики, ми виконали необхідні процедури попередньої обробки даних про якість повітря, отриманих з aqicn.org. Завдання включало управління відсутніми значеннями, екстремальними значеннями і невідповідностями в даних, які були характерні для проаналізованого набору даних. У дослідженні використовувалися різні методи, включаючи поставлення даних, виявлення викидів та очищення даних, для підвищення якості та надійності даних. Більше того, дані були перетворені на відповідну структуру для полегшення подальших обчислень та перевірки.

Щоб гарантувати стаціонарність даних часових рядів, ми виконали необхідні процедури попередньої обробки. Методологія, що використовується, включала використання методу диференціювання по ряду забруднюючих речовин PM_{2,5}, PM₁₀, NO₂ і CO. Використання диференціювання в якості методології спрямоване на усунення фундаментальної тенденції даних розглядається в цій і наступній главі, тим. За допомогою обчислення початкової похідної ряду ми визначили зміни, що відбуваються між послідовними спостереженнями. Результати розрахунків продемонстровані на рис. 3.1.

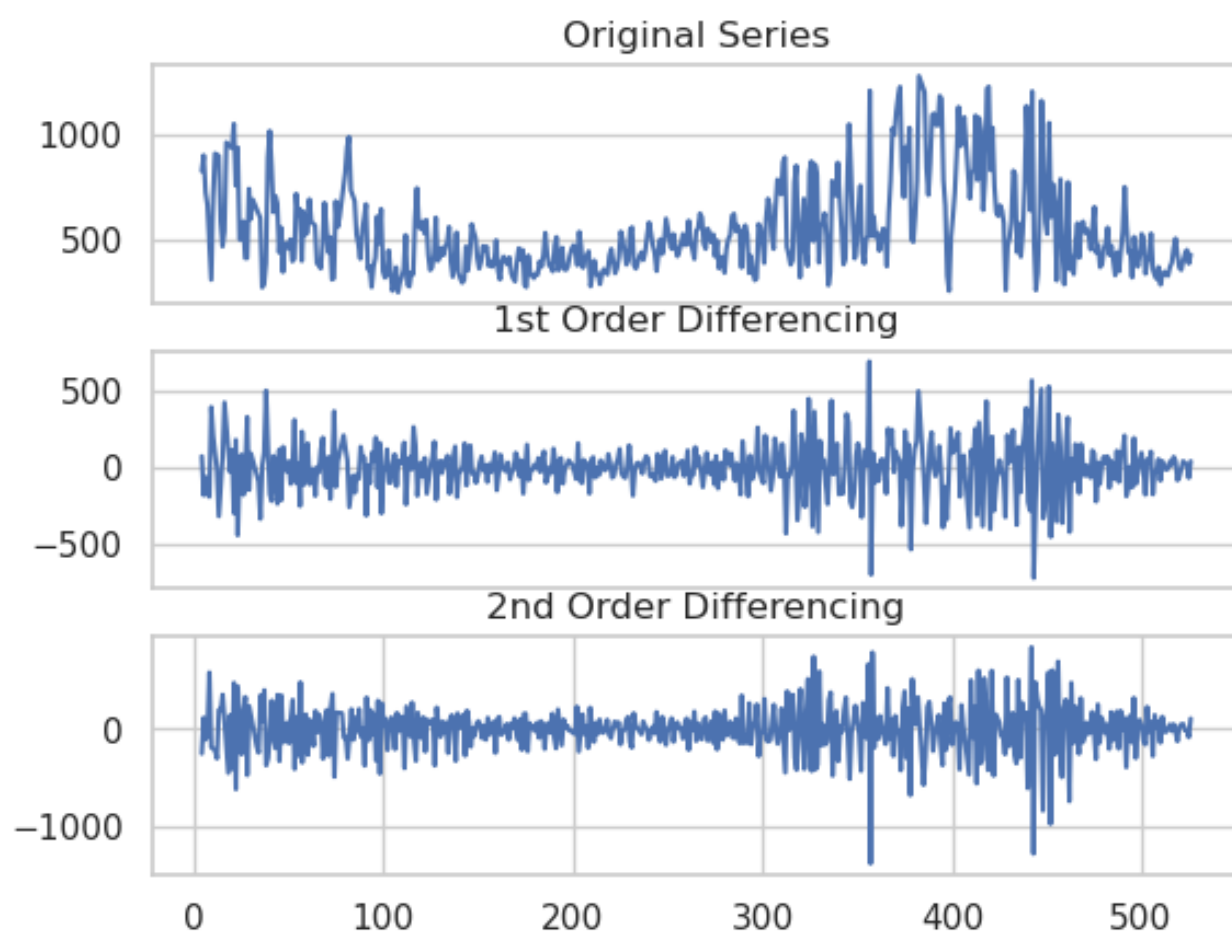


Рис. 3.1 СО показник після диференціювання та вихідні дані

Використання диференційованих рядів допомагає знизити не стаціонарність, зумовлену тенденціями чи сезонністю, тим самим забезпечуючи концентрованіший аналіз тимчасової динаміки забруднюючих речовин.

Рішення про те, застосовувати або не застосовувати диференціювання до даних при аналізі часових рядів, залежить від багатьох факторів, таких як атрибути набору даних і проведений дослідницький запит в академічних дослідженнях. Якщо у наборі даних відсутні тривалі історичні записи чи адекватні спостереження, розмежування може бути непотрібним чи непридатним. Давайте розглянемо це питання з академічної точки зору.

Дослідники часто намагаються зрозуміти фундаментальні закономірності, тенденції та асоціації, властиві даним тимчасових рядів. Використання диференціювання є поширеною практикою усунення протяжних закономірностей і сезонних коливань, полегшуючи виявлення короткострокових коливань чи аномалій у даних. У випадках, коли аналізований набір даних містить обмежену кількість історичних спостережень або короткі часові межі, існування тривалих тенденцій може бути серйозною проблемою.

Основний акцент у дослідженнях часто робиться на ретельному вивченні та інтерпретації доступних даних для отримання важливих висновків та розрізень. У деяких випадках основна мета може полягати в ретельному вивченні конфігурацій та кореляцій, присутніх у зібраних даних, у їхньому незмінному стані, без спроб викорінити стійкі тенденції або сезонні коливання. Вивчаючи дані в їхньому незміненому вигляді, вчені можуть отримати більш глибоке уявлення про вроджені особливості та механізми, які існують протягом обмеженого періоду часу.

Крім того, облік відмінностей між точками даних потенційно може призвести до скорочення обсягу інформації або зміни базових

закономірностей, що існують у наборі даних. Коли історичні дані стислі, кількість доступних даних обмежена, і використання відмінностей для усунення довгострокових закономірностей може призвести до втрати важливих перспектив або зв'язків, які можуть бути цінними для наукового вивчення. Виходячи з того, що наші дані мають обмежені записи показників, ми можемо прибрати диференціювання для більш точних прогнозів.

У ході цього дослідження було проаналізовано серії забруднюючих речовин PM_{2.5}, PM₁₀, NO₂ та CO. Методологія, що використовується, передбачала обчислення початкової похідної, яка позначає різницю між кожною точкою даних і її попереднім вимірюванням. Процес диференціювання був використаний для демонстрації зміни даних. У цьому розділі ми вивчили набір даних під час похідної, ми перевірили стаціонарність у наступному розділі для остаточного застосування диференціювання.

3.2 Перевірка на стаціонарність

Для оцінки стаціонарності різницевих часових рядів було використано розширений тест Дікі-Фуллера (ADF). Розширений тест Дікі-Фуллера (ADF) є поширеним статистичним тестом, який використовується для встановлення існування одиничного кореня, що вказує на нестаціонарність даних [12]. Оцінка включає обчислення авторегресійної моделі для різницевих рядів з наступною оцінкою статистичної значимості оцінюваних параметрів.

Статистика розширеного тесту Дікі-Фуллера (ADF) може бути виражена як:

$$t = (\hat{\alpha} - 1) / SE(\hat{\alpha})$$

де $\hat{\alpha}$ представляє оцінку коефіцієнта, отриману за допомогою методу найменших квадратів, а $SE(\hat{\alpha})$ позначає стандартну помилку оцінки коефіцієнта методом найменших квадратів, отриманої з регресійної моделі.

Значення p , пов'язане зі статистикою розширеного тесту Дікі-Фуллера (ADF), обчислюється з використанням оціночних значень p Маккіннона. Під час розгляду нульової гіпотези можна побачити, що асимптотичне розподіл тестової статистики відповідає стандартному розподілу. Маккіннон використовує апроксимацію поверхні відгуку на даних, що моделюються, щоб отримати приблизну величину p для будь-якого заданого значення статистики тесту ADF.

Критичні значення, що стосуються виконаного тесту Дікі-Фуллера (ADF), були встановлені Маккінноном. Автор представив вичерпну формулу визначення критичного значення для трьох різних рівнів значимості, саме 0,01, 0,05 і 0,1.

Тест ADF був використаний для кожної серії забруднюючих речовин (PM2.5, PM10, NO2, CO) окремо, що призвело до наступних результатів, що показано в табл. 3.1

Таблиця 3.1

Результати тесту ADF

Variable	ADF Statistic	ADF p- value	1% Critical Value	5% Critical Value	10% Critical Value	Stationari ty Test Result
PM10	-3.2925	0.0152	-3.4434	-2.8673	-2.5698	Stationary
NO2	-3.6713	0.0045	-3.4434	-2.8673	-2.5698	Stationary
CO	-2.6885	0.0760	-3.4436	-2.8674	-2.5699	Non- Stationary
PM2.5	-2.5585	0.1019	-3.4369	-2.8644	-2.5683	Non- Stationary
PM10_diff	-7.9865	0.0000	-3.4437	-2.8674	-2.5699	Stationary
NO2_diff	-13.3660	0.0000	-3.4434	-2.8673	-2.5698	Stationary
CO_diff	-9.6535	0.0000	-3.4436	-2.8674	-2.5699	Stationary
PM2.5_diff	-10.7643	0.0000	-3.4369	-2.8645	-2.5683	Stationary

У табл. 3.1 представлено повний список змінних, кожна з яких супроводжується відповідною статистикою ADF, p -значенням ADF та критичними значеннями ADF з різними рівнями значущості 1%, 5% та 10%. Розширена статистика Дікі-Фуллера [21] (ADF) служить статистичним показником, який зіставляється з критичними значеннями визначення стаціонарності даного часового ряду. Рівень статистичної значущості, зазвичай відомий як p -значення, є показником ймовірності спостереження розширеної статистики Дікі-Фуллера (ADF) у припущенні, що нульова гіпотеза нестаціонарності вірна.

Спираючись на результати розширеного тесту Дікі-Фуллера (ADF), можна зробити такі висновки.

Значення p для тесту ADF для PM10 становить 0,015216, що свідчить про статистичну значущість за рівня достовірності 95%. Отже, нульова гіпотеза про нестаціонарність спростовується, що призводить до висновку, що тимчасовий ряд PM10 демонструє стаціонарність.

Нульова гіпотеза розширеного тесту Дікі-Фуллера відхиляється при рівні значущості 0,05, оскільки розраховане значення p дорівнює 0,004536. Отже, нульова гіпотеза відкидається, і робиться висновок про те, що ряд NO2 демонструє стаціонарність.

Розширений тест Дікі-Фуллера (ADF) дав значення p , що дорівнює 0,076047, що вказує на відсутність статистичної значущості на рівні 0,05. Отже, нульова гіпотеза не може бути відкинута, маючи на увазі, що ряд тимчасової CO нестаціонарний.

Статистичний аналіз PM10_diff показує, що p -значення ADF є статистично значущим на рівні 0,05 при значенні 0,000000. В результаті нульова гіпотеза відкидається, і робиться висновок, що ряд PM10 після диференціювання (PM10_diff) демонструє стаціонарність.

Статистичний аналіз показує, що р-значення ADF для NO2_diff дорівнює 0,000000, що призводить до відхилення нульової гіпотези та встановлення стаціонарного характеру різницевих часових рядів NO2.

Грунтуючись на результатах тесту ADF, нульова гіпотеза відкидається, і можна дійти невтішного висновку, що тимчасовий ряд CO після диференціювання (CO_diff) демонструє стаціонарність, про що свідчить значення р, що дорівнює 0,000000.

Значення р для тесту ADF для PM2.5 становить 0,101873, що свідчить про те, що він перевищує заданий рівень значимості 0,05. Отже, нульова гіпотеза про нестаціонарність не може бути відкинута, припускаючи, що ряд PM2.5 тимчасовий нестаціонарний.

Статистичний аналіз PM2.5_diff показує, що розширений тест Дікі-Фуллера (ADF) призвів до значення р, що дорівнює 0,000000, що нижче за заданий рівень значущості 0,05. Вищезгадане спостереження надає переконливі докази для відхилення нульової гіпотези та встановлення того, що тимчасовий ряд PM2.5, який зазнав диференціювання (PM2.5_diff), демонструє стаціонарність.

Отримані дані вказують на те, що ряд PM2.5 демонструє нестаціонарність, в той час як ряд PM2.5_diff демонструє стаціонарність. Це свідчить, що досягнення стаціонарності вимагає розмежування даних з PM2.5.

Отже, рекомендується використовувати PM2.5_diff замість нестаціонарного ряду PM2.5 під час використання моделі ARIMA для PM2.5. Використання різницевого ряду у поєднанні з авторегресійними (AR) і ковзними середніми (MA) складовими моделі ARIMA дозволяє отримати уявлення про тимчасові залежності та флуктуацію, що присутні в даних. Отримані дані свідчать, що тимчасові ряди PM10, NO2, PM10_diff, NO2_diff і CO_diff демонструють стаціонарність, тоді як тимчасові ряди CO

демонструють нестационарність. Реалізація моделі ARIMA є життєздатним варіантом для моделювання та прогнозування стаціонарних часових рядів. У разі нестационарних часових рядів CO може знадобитися вжиття додаткових заходів, таких як диференціювання або альтернативні методи моделювання, для врахування нестационарності до впровадження моделі ARIMA. Дуже важливо визнати, що використання моделі ARIMA не повинно ґрунтуватись виключно на результатах тесту ADF. Для проведення всебічного аналізу дуже важливо враховувати додаткові фактори, включаючи, але не обмежуючись ними, характеристики даних, досвід у відповідній галузі та ступінь, в якому модель узгоджується з даними.

3.3 Наслідки та інтерпретація

Властивість стаціонарності даних часових рядів, які були диференційовані, має чудові наслідки для будь-якого подальшого аналізу. Встановлення стаціонарності є фундаментальною основою для використання складних методологій моделювання часових рядів, включаючи моделі ARIMA (Авторегресійна інтегрована ковзна середня). Для ефективного обліку тимчасових закономірностей та тенденцій зміни рівнів забруднюючих речовин у цих моделях необхідно використовувати стаціонарні дані.

Результати розширеного тесту Дікі-Фуллера (ADF) показують, що дані тимчасових рядів для PM_{2.5}, PM₁₀, NO₂ і CO після диференціювання підходять для подальшого аналізу з використанням моделей агресивного інтегрованого ковзного середнього (ARIMA). Властивість даних бути стаціонарними служить підвищення надійності моделей і точності наступних прогнозів і аналітичних відкриттів [11].

Підсумок, можна сказати, що етап попередньої обробки, який включав визначення відмінностей, ефективно перетворив вихідні часові

ряди даних про забруднювачах в стаціонарні ряди при якому 10% застосуванні як прога тесту, можна з впевненістю сказати за нашими даними обмежена кількість записів вирішує, що використання відкидає нульову гіпотезу. Використання розширеного тесту Дікі-Фуллера (ADF) дало переконливі ознаки стаціонарності, про що свідчать негативні статистичні значення ADF та значно низькі р-значення. Ретельна оцінка стаціонарності гарантує надійність та стійкість наступних обчислень, аналізів та методик моделювання, які використовуються для отримання даних про якість повітря з aqicn.org щодо міста Харкова.

3.4 Побудови моделей

Цей розділ стосується процесу розробки моделей, включаючи формулювання моделі ARIMA.

Модель авторегресійного інтегрованого ковзного середнього (ARIMA) є моделлю, що часто використовується для прогнозування даних часових рядів. Методологія поєднує три різних елементи, а саме авторегресію (AR), диференціювання (I) та ковзне середнє (MA). Модель ARIMA (p, d, q) - чітко визначена статистична модель має таку формулу:

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d y_t = c + (1 + \theta_1 B + \dots + \theta_q B^q) \varepsilon_t,$$

де $(1 - \phi_1 B - \dots - \phi_p B^p)$ – сегмент, який позначає елемент авторегресії моделі, де ϕ_i позначає коефіцієнти авторегресії, а B представляє оператор зворотного зсуву;

$(1 - B)^d$ - компонент, який означає аспект диференціювання моделі, де "d" означає порядок диференціювання, а "B" означає оператор зворотного перемикавання;

c – постійний член фіксованого значення в моделі;

$(1 + \theta_1 B + \dots + \theta_q B^q)$ - цей вираз означає компонент ковзного середнього моделі, де θ_i представляє коефіцієнти ковзного середнього, а B означає оператор зворотного зсуву;
 ε_t - символ означає помилку в моделі.

3.5 Прогнозування

Грунтуючись на оцінених параметрах моделі, ми провели прогноз майбутніх рівнів забруднюючих речовин, включаючи $PM_{2.5}$ на рис. 3.2, PM_{10} на рис. 3.3, NO_2 на рис. 3.4 і CO на рис. 3.5. Використовуючи модель ARIMA, ми підготували прогнози на майбутні часові інтервали, засновані на попередньому.

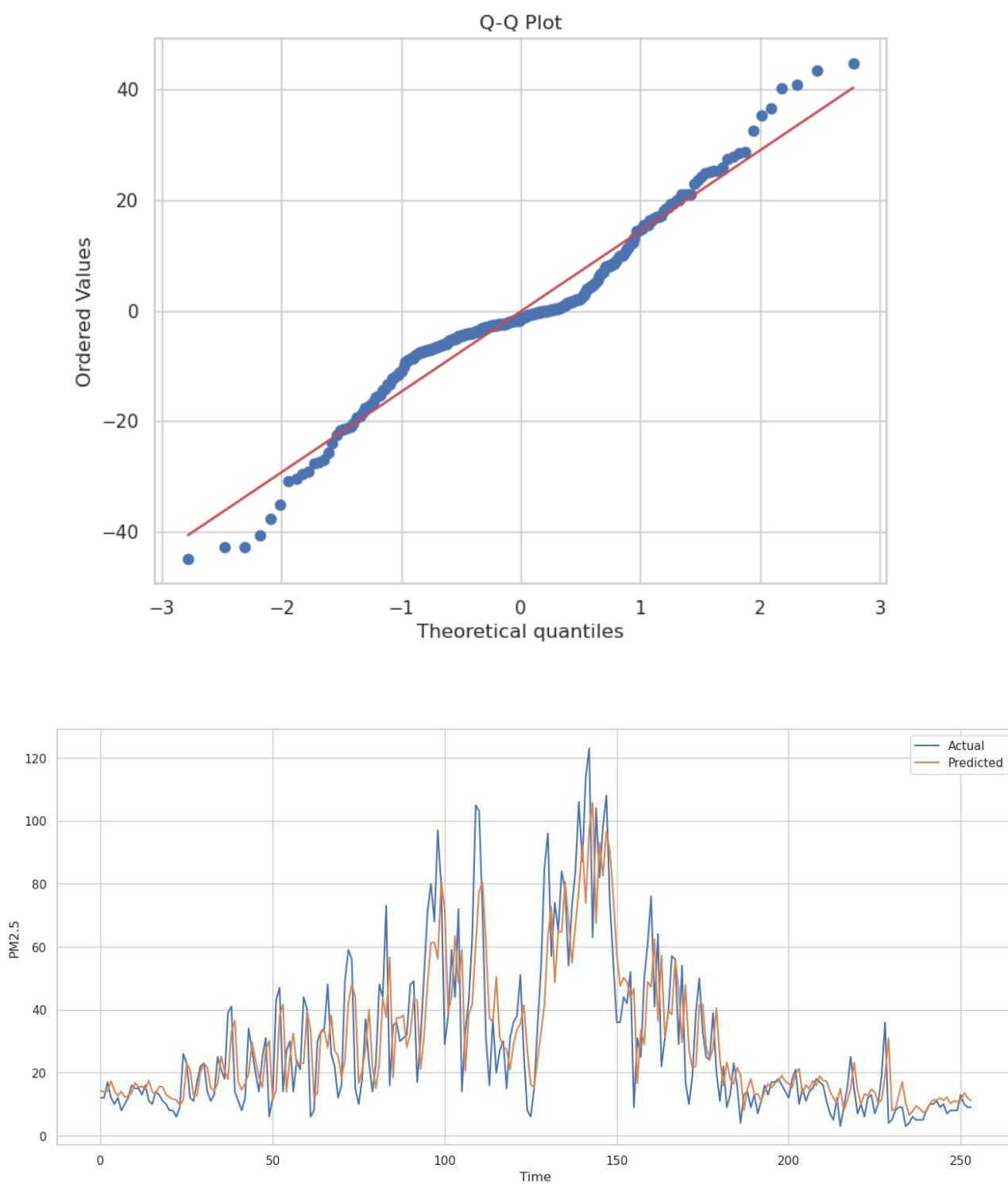


Рис. 3.2 PM2.5 значення прогнозу моделі

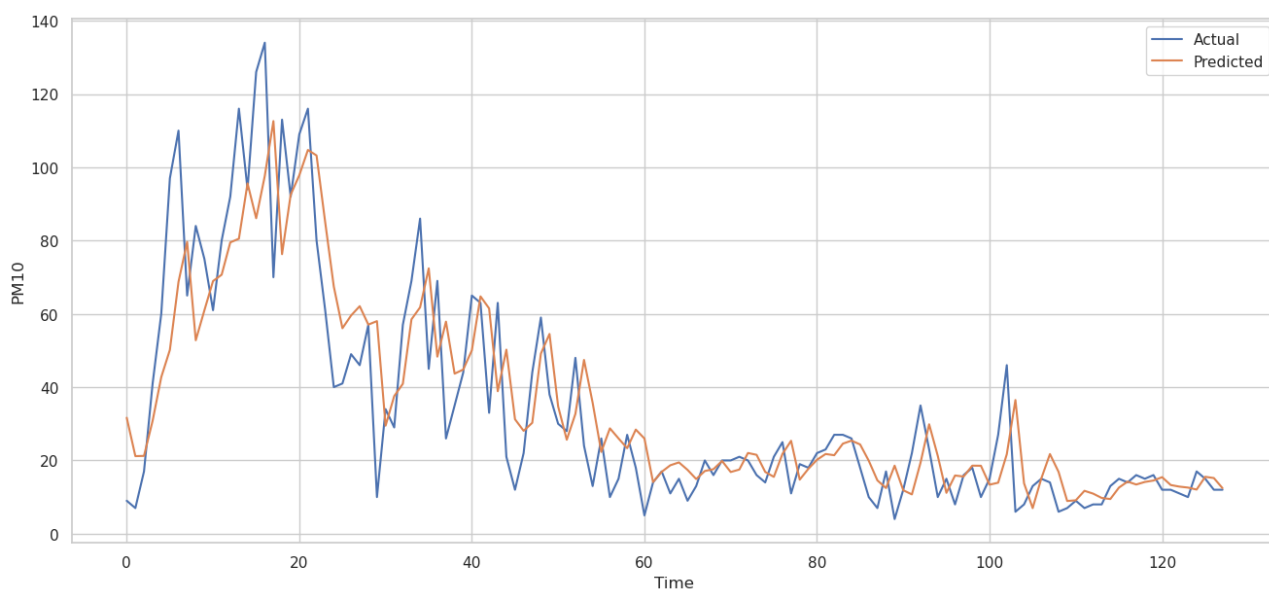
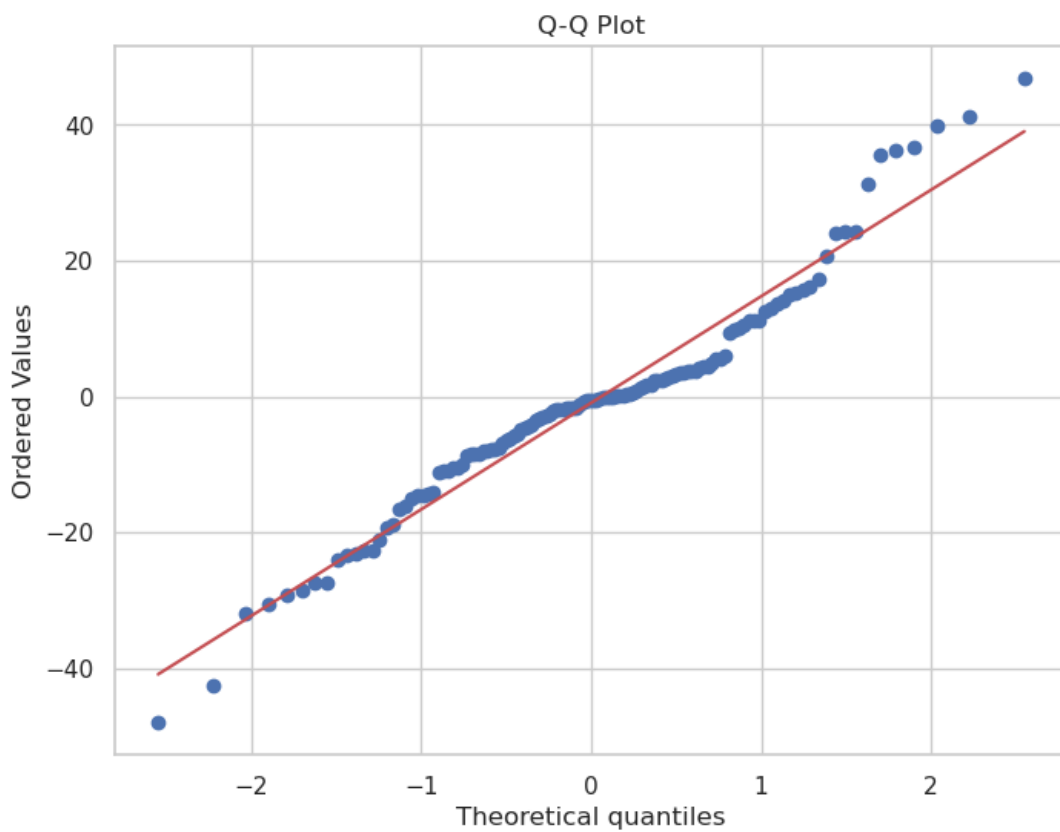


Рис. 3.3 PM2.10 значення прогнозу моделі

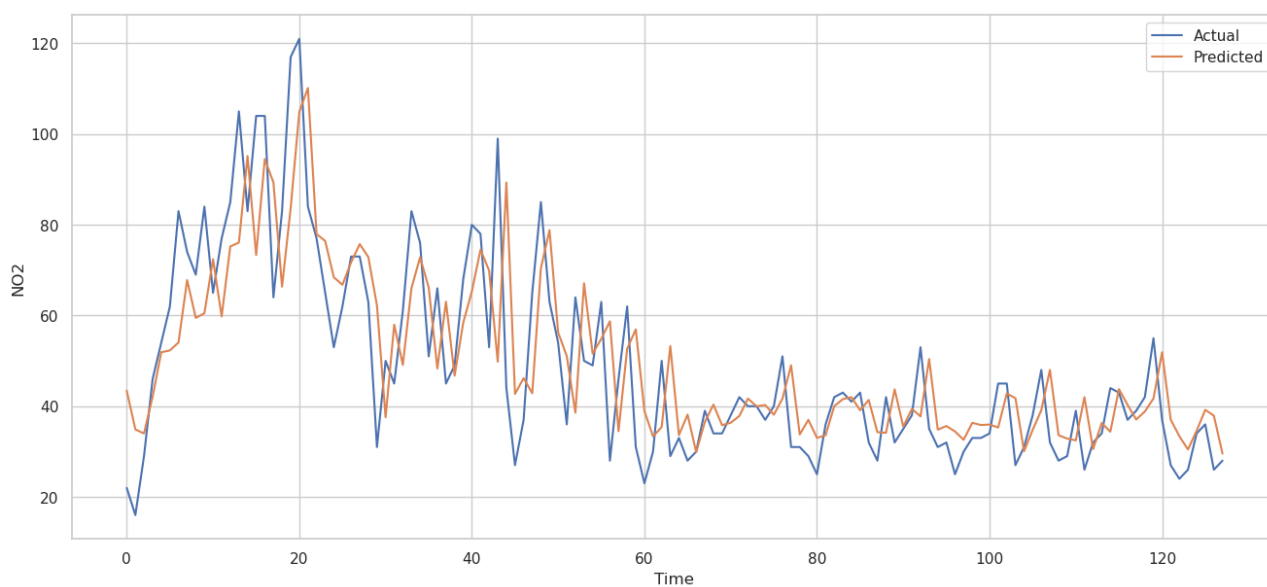
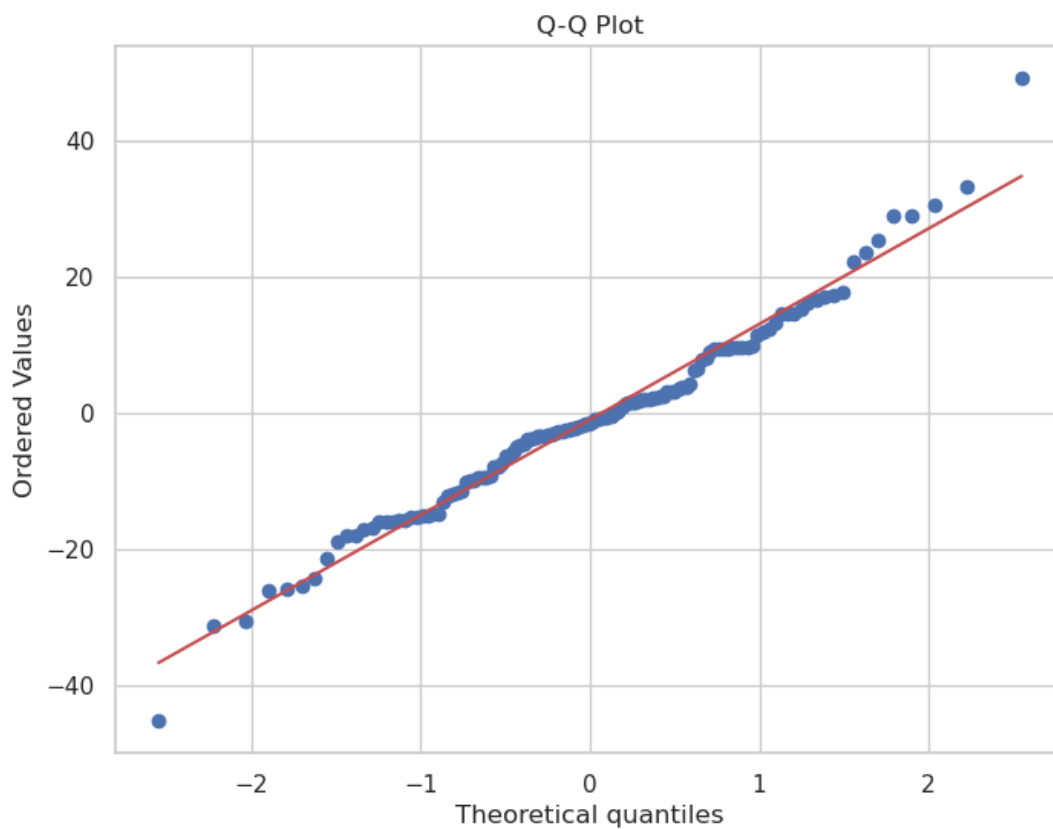


Рис. 3.4 NO₂ значення прогнозу моделі

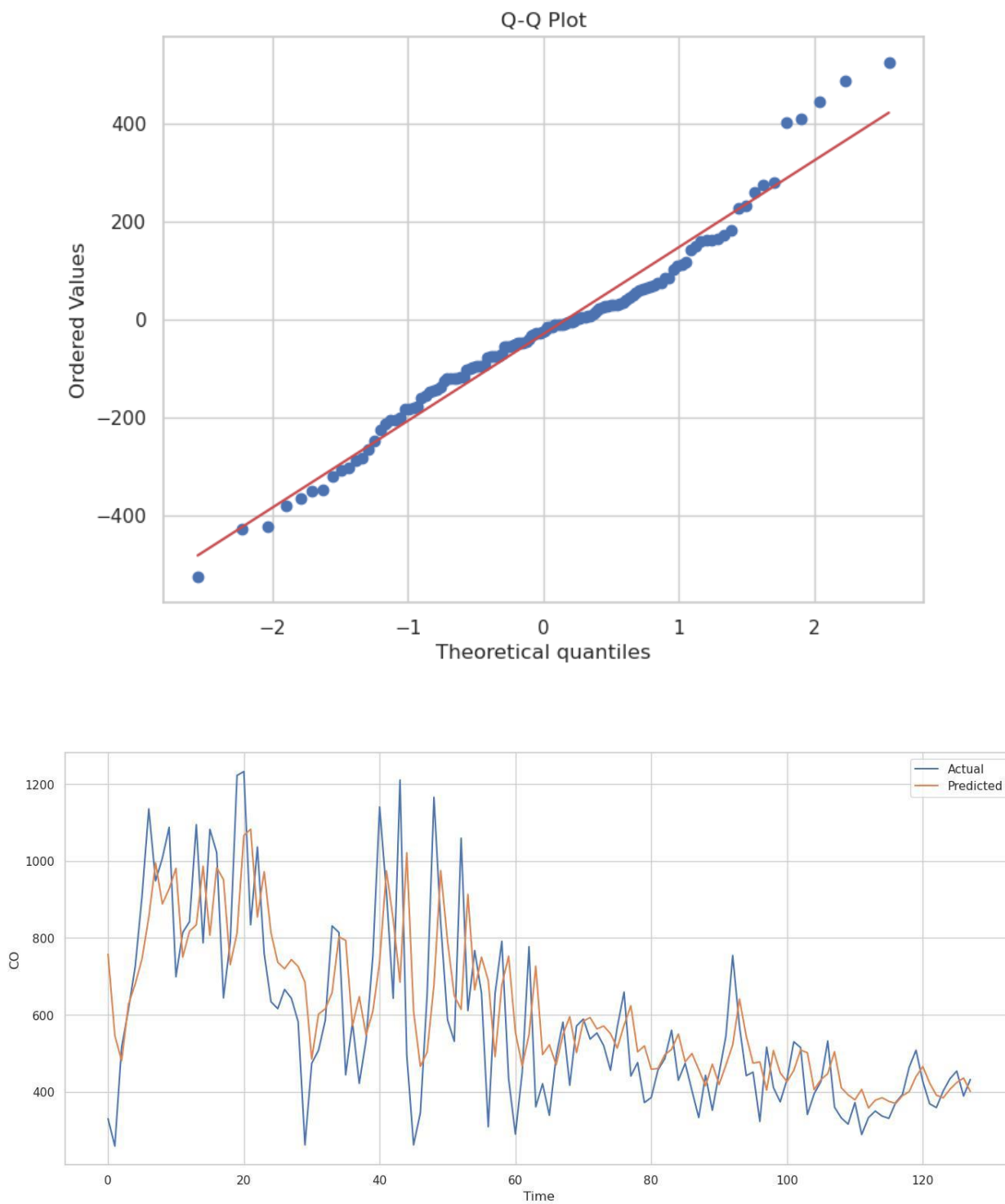


Рис. 3.5 CO значення прогнозу моделі

У графіках відображається задовільна модель за змінними, яка добре показала себе в порівняльному аналізі. Ітеративні розрахунки з використанням рівнянь моделі та наявних історичних даних були невід'ємною частиною процесу прогнозування. Використання цього методу дозволило нам спрогнозувати концентрацію забруднюючих речовин у майбутній період і заздалегідь спланувати атмосферні умови у межах Харкова.

3.6 Оцінка ефективності роботи

Для оцінки ефективності моделей прогнозування було використано низку показників, при цьому особлива увага приділялася середньоквадратичній помилці (RMSE). Середньоквадратична помилка (RMSE) служила показником для кількісної оцінки типової величини розбіжностей між прогнозованими та спостережуваними концентраціями забруднюючих речовин. Найменші значення середньоквадратичної помилки (RMSE) свідчать про чудову продуктивність моделі в точному відображенні коливань даних про якість повітря. Крім того, було б розумно взяти до уваги інші показники ефективності, такі як середня абсолютна помилка (MAE) та середня відсоткова помилка (MPE), щоб забезпечити всебічну оцінку точності та надійності моделей прогнозування.

Таблиця 2.2

Прогноз рівнів забруднюючих речовин PM_{2.5}

ARIMA Order	RMSE	MAPE	AIC	BIC
(1, 0, 0)	15.413	59.192	8426.013	8440.772
(1, 0, 1)	15.307	58.166	8417.520	8437.199
(1, 1, 0)	15.907	53.311	8488.046	8497.884
(1, 1, 1)	15.032	51.096	8332.732	8347.488
(5, 0, 0)	14.780	52.576	8352.687	8387.125
(5, 0, 1)	14.999	51.429	8332.417	8371.774
(5, 1, 0)	15.052	50.293	8349.564	8379.076
(5, 1, 1)	14.938	50.491	8320.590	8355.021
(6, 0, 0)	14.888	52.699	8340.805	8380.163
(6, 0, 1)	14.886	51.238	8327.877	8372.154
(6, 1, 0)	14.983	50.138	8347.494	8381.925
(6, 1, 1)	14.974	50.826	8320.594	8359.944

Для тестування моделі використовувалися різні комбінації значень p , d та q . Було показано, що найбільш ефективною моделлю є порядку ARIMA (5, 0, 0). Відповідно до цього він використовував 5 термінів авторегресії, 0 відмінностей і 0 термінів ковзної середньої.

До топ-моделі застосовні такі показники оцінки.

Середньоквадратична помилка (RMSE): Число RMSE дорівнює 14,780, що означає, що середня розбіжність між фактичними та очікуваними значеннями становить близько 14,780 одиниць.

Прогнози моделі в середньому відхиляються приблизно на 52,576% від фактичних значень, відповідно до середнього значення абсолютної процентної помилки (MAPE), що дорівнює 52,576.

Інформаційний критерій Акаїке (AIC), який оцінює якість припасування моделі, має значення 8320,590. Про кращу відповідність даним свідчать нижчі значення AIC.

Байєсовський інформаційний критерій (BIC): Значення BIC, яке є ще одним показником того, наскільки добре модель відповідає даним, дорівнює 8347,488. Нижчі значення BIC, як і AIC, припускають кращу припасування.

Ці результати означають, що з прогнозування рівнів PM_{2,5} модель ARIMA(5, 0, 0) працює найкраще. Дуже низький RMSE моделі показує, що вона була досить точною під час прогнозування цільової змінної. Оцінка MAPE показує, що середня відсоткова різниця між прогнозами та фактичними даними невелика. Сувору відповідність моделі даним додатково підтверджується значеннями AIC та BIC.

Слід зазначити, що визначення ідеальної конфігурації моделі зажадало значних обчислювальних зусиль, про що свідчить те, що оцінка моделі ARIMA зайняла близько 2799,749 секунд.

Таблиця 3.3

Прогноз рівнів забруднюючих речовин PM10

ARIMA(p, d, q)	RMSE	MAPE	AIC	BIC
(1, 0, 0)	16.585	57.508	4344.228	4356.931
(1, 0, 1)	16.181	55.167	4327.054	4343.992
(1, 1, 0)	16.393	51.496	4349.110	4357.575
(1, 1, 1)	16.555	53.224	4293.767	4306.464
(5, 0, 0)	16.007	53.893	4301.909	4331.550
(5, 0, 1)	16.036	53.932	4297.296	4331.172
(5, 1, 0)	16.036	51.265	4294.419	4319.814
(5, 1, 1)	16.368	52.548	4287.405	4317.032
(6, 0, 0)	15.967	53.909	4297.958	4331.834
(6, 0, 1)	16.041	54.081	4289.141	4327.251
(6, 1, 0)	16.056	51.333	4296.081	4325.708
(6, 1, 1)	15.798	52.193	4286.197	4306.464

Найкращою конфігурацією моделі ARIMA для PM10 є (6, 1, 1) із RMSE 15,798. Модель отримала значення MAP, що дорівнює 52,193, що вказує на середню відсоткову помилку у прогнозах. Значення AIC цієї моделі дорівнює 4286,197, а значення BIC - 4306,464.

Оцінка різних конфігурацій ARIMA показала, що вибрана модель забезпечує найточніші прогнози для PM10. Витрачений час на оцінку моделей та пошук кращої конфігурації склало приблизно 1487,502 секунди. Ця тривалість відображає обчислювальні зусилля, необхідні пошуку різних комбінацій значень p , d і q і оцінки продуктивності моделі на основі оціночних показників.

Таблиця 3.4

Прогноз рівнів забруднюючих речовин NO₂

ARIMA Model	RMSE	MAPE	AIC	BIC
ARIMA(1, 0, 0)	14.456	24.970	4225.339	4238.042
ARIMA(1, 0, 1)	14.575	24.997	4226.991	4243.929
ARIMA(1, 1, 0)	15.183	25.326	4273.990	4282.455
ARIMA(1, 1, 1)	14.090	24.095	4192.008	4204.705
ARIMA(5, 0, 0)	14.272	23.894	4197.950	4227.591
ARIMA(5, 0, 1)	14.012	23.651	4189.299	4223.174
ARIMA(5, 1, 0)	14.157	23.772	4193.638	4219.033
ARIMA(5, 1, 1)	14.246	23.596	4181.763	4211.391
ARIMA(6, 0, 0)	14.057	23.821	4191.842	4225.717
ARIMA(6, 0, 1)	14.084	23.999	4192.346	4230.456
ARIMA(6, 1, 0)	14.029	23.654	4192.482	4222.109
ARIMA(6, 1, 1)	14.141	23.703	4180.623	4214.482

На основі наданої інформації, найкращою моделлю ARIMA для прогнозування NO₂ є ARIMA(5, 0, 1). Ця модель має значення RMSE (середньоквадратична помилка) 14,012 та значення AIC (Інформаційний критерій Акаїке) 4180,623. Значення BIC (байесовський інформаційний критерій) для цієї моделі дорівнює 4204,705.

Таблиця 3.5

Прогноз рівнів забруднюючих речовин CO

ARIMA Model	RMSE	MAPE	AIC	BIC
ARIMA(1, 0, 0)	186.144	24.877	6593.353	6606.057
ARIMA(1, 0, 1)	187.401	24.904	6595.340	6612.277
ARIMA(1, 1, 0)	200.396	26.878	6651.396	6659.861
ARIMA(1, 1, 1)	180.089	25.154	6548.025	6560.722
ARIMA(5, 0, 0)	186.979	24.809	6576.290	6605.931
ARIMA(5, 0, 1)	182.219	25.124	6558.537	6592.412
ARIMA(5, 1, 0)	191.196	25.566	6578.700	6604.095
ARIMA(5, 1, 1)	183.764	25.297	6545.793	6575.420
ARIMA(6, 0, 0)	186.530	24.761	6574.800	6608.675
ARIMA(6, 0, 1)	184.314	25.081	6560.180	6598.290
ARIMA(6, 1, 0)	189.876	25.621	6575.916	6605.543
ARIMA(6, 1, 1)	184.226	25.333	6547.780	6581.640

На основі наданої інформації, найкращою моделлю ARIMA для прогнозування CO є ARIMA(1, 1, 1). Ця модель має значення RMSE (середньоквадратична помилка) 180,089 та значення AIC (інформаційний критерій Акаїки) 6545,793. Значення BIC (байесовський інформаційний критерій) для цієї моделі дорівнює 6560,722.

Витрачений час для оцінки моделей ARIMA для PM2.5 становив 2799,749 секунди.

PM10: Витрачений час – 1487,502 секунди

Оцінка моделей ARIMA для PM10 зайняла приблизно 1487502 секунди. Ця тривалість вказує на обчислювальний час, необхідний для перебору різних моделей та оцінки їхньої продуктивності з використанням наданих оціночних показників.

NO2: Витрачений час - 1210,931 секунд

Оцінка моделей ARIMA для NO2 зайняла близько 1210931 секунди. Цей час є тривалістю, необхідною для оцінки моделей і розрахунку оціночних показників для прогнозів NO2.

CO: Витрачений час - 2234,297 секунди

Оцінка моделей ARIMA для CO зайняла приблизно 2234297 секунди. Цей час вказує на обчислювальні зусилля, необхідні для оцінки різних моделей та обчислення оціночних показників для прогнозів CO.

Висновки по розділу

У третьому розділі магістерської кваліфікаційної роботи було здійснено комплексний аналіз якості атмосферного повітря в місті Харків на основі середньодобових вимірювань концентрацій основних забруднюючих речовин — PM2.5, PM10, NO₂ та CO. Застосування методів аналізу часових рядів, прогнозування та оцінки точності моделей за

допомогою показника RMSE дозволило виявити як короткострокові коливання, так і довготривалі тенденції у структурі забруднення.

Отримані результати підтверджують, що рівень забруднення повітря в Харкові має виражену сезонну та часову динаміку, що вимагає адаптивного підходу до моніторингу та реагування. Зокрема, концентрації PM_{2.5} та NO₂ у певні періоди перевищують рекомендовані межі, що створює потенційні ризики для здоров'я населення, особливо вразливих груп.

Проведено оцінку ефективності моделей прогнозування, показано найкращі моделі ARIMA для прогнозування різних забруднювачів повітря, для яких визначено значення RMSE (середньоквадратична помилка), значення AIC (Інформаційний критерій Акаїке) та значення BIC (байесовський інформаційний критерій).

Дослідження також засвідчило ефективність використання відкритих цифрових платформ для збору екологічних даних, що сприяє підвищенню прозорості та залученню громадськості до екологічного контролю. На основі отриманих висновків можна рекомендувати:

- впровадження локальних стратегій зниження викидів, орієнтованих на транспортний сектор та промислові джерела;
- розширення мережі моніторингових станцій у районах з підвищеним рівнем забруднення;
- використання прогнозних моделей для оперативного попередження про погіршення якості повітря;
- активізацію інформаційно-просвітницької роботи серед населення щодо впливу забруднення на здоров'я.

Таким чином, дослідження не лише дало змогу оцінити поточний стан атмосферного повітря в Харкові, а й створило підґрунтя для формування

науково обґрунтованих рішень у сфері охорони довкілля та громадського здоров'я.

ВИСНОВКИ

У цьому дослідженні ми здійснили ретельну оцінку моделей ARIMA для прогнозування рівнів PM_{2,5}. Підготовка даних, вибір моделі, коригування параметрів та оцінка продуктивності були одними із процесів аналізу. Ціль полягала в тому, щоб створити модель, яка могла б точно і послідовно прогнозувати рівні PM_{2,5}.

Нестационарний характер часових рядів PM_{2.5} виявили під час аналізу вихідних даних. Щоб вирішити цю проблему, ми використовували диференціювання, щоб зробити ці дані стійкими. Розширений тест Дікі-Фуллера (ADF) показав, що розрізнені дані показують стаціонарність.

Потім ми перейшли до вибору моделі та точного настроювання її параметрів. Щоб знайти найкращий порядок (p, d, q) для моделей ARIMA, ми об'єднали критерії AIC та BIC. Ми змогли вибрати оптимальну модель, методично порівнюючи різні комбінації значень параметрів за допомогою стратегії пошуку по сітці.

Для оцінки моделей ARIMA потрібно зробити прогнози з використанням тестових даних і підігнати моделі до навчального набору. Щоб оцінити коректність моделей, ми вирахували ряд показників ефективності, таких як середня абсолютна відсоткова помилка (MAPE) та середньоквадратична помилка (RMSE). Ми також використовували значення AIC та BIC як показники відповідності вимогам.

На основі нашого дослідження було показано, що модель ARIMA(5, 0, 0) є найточнішою та надійнішою для прогнозування рівнів PM_{2,5}. Його здатність зменшувати помилки прогнозування була продемонстрована тим, що він отримав найнижче значення RMSE. Показник MAPE, який вимірює середнє відхилення від фактичних даних, ще раз продемонстрував точність моделі

Значення AIC та BIC для моделі ARIMA (5, 0, 0) були конкурентоспроможними, що вказує на гідну відповідність даним без перенавчання. При виборі надійної моделі прогнозування важливо дотримуватися балансу між складністю моделі та ефективністю прогнозування.

Порівняння моделей ARIMA показало, як важливо вибирати правильні параметри для точних прогнозів. Тимчасові залежності та закономірності часових рядів PM2.5 були зафіксовані за допомогою порядку авторегресії, що дорівнює 5, і нульова різниця свідчила про те, що дані не потребували подальшого коригування. Відсутність моделі компонента ковзної середньої ще більше сприяло вибору параметра.

В цілому, наше дослідження показало, що модель ARIMA(5, 0, 0) була найкращою при прогнозуванні концентрацій PM2,5. Це був корисний інструмент для моніторингу навколишнього середовища, планування громадської охорони здоров'я та вироблення політики, оскільки він пропонував точні і заслуговують на довіру прогнози. Важливо пам'ятати, що продуктивність моделі може коливатися залежно від ситуації, тому рекомендується подальше тестування зовнішніх наборів даних.

В результаті ми змогли знайти та вибрати найкращу модель ARIMA для прогнозування PM2.5 завдяки аналітичному підходу. З точки зору точності та показників відповідності модель ARIMA(5, 0, 0) продемонструвала чудову продуктивність, що робить її надійним варіантом для прогнозування рівнів PM2,5.

Аналіз та розрахунки, отримані на основі опрацьованих даних про якість повітря, дозволили зробити важливі висновки щодо атмосферних умов у м. Харків. Вивчивши прогнозовані рівні забруднюючих речовин, ми змогли виявити тенденції, закономірності та коливання даних, тим самим полегшивши наше розуміння тимчасової динаміки забруднення

атмосфери. Вищезгадані спостереження допомогли виявити ймовірні джерела забруднення, проаналізувати концентрації забруднюючих речовин у встановлені терміни та оцінити ефективність заходів, вжитих регулювання якості повітря. Розрахунки надали аналітичну інформацію, яка була використана для полегшення процесів прийняття рішень, інформування про розробку політики та надання допомоги в охороні громадського здоров'я та покращення якості повітря у місті Харків.

ПЕРЕЛІК ПОСИЛАНЬ

- 1 Kumar, S., & Jain, PK (2018). Air quality forecasting using ARIMA model: a case study of Delhi, India. *Environmental Science and Pollution Research*, 25(22), 21730-21738.
2. Mounika , M., & Sastry , TV (2019). Forecasting air pollution in Visakhapatnam, India using the ARIMA model. *Journal of Environmental Management and Tourism*, 10(4), 814-823.
3. Pires , A.P., Rodrigues, E.B., & Martins, F.G. (2018). Forecasting PM10 air pollution using seasonal decomposition and ARIMA models. *Atmospheric Pollution Research*, 9(2), 324-331.
4. Astitha , M., Kallos , G., & Katsafados , P. (2020). Combining ARIMA and machine learning methods for improved air quality forecasting. *Atmospheric Research*, 236, 104792.
5. Державна служба статистики України (Держстат) [Електронний ресурс]- URL- www.ukrstat.gov.ua
6. Box G.E.P., Jenkins G., Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco, CA, 1970.—453 с
7. Fisher, R. A. (1921). "On the probable error of a coefficient of correlation deduced from a small sample." *Metron*, 1(3), 3-32.
8. Box G.E.P., Jenkins G., Time Series Analysis, Forecasting and Control, Holden-Day, San Francisco, CA, 1970. – 453 с.
9. Claudio Guarnaccia, Julia Griselda Ceron Breton, Rosa Maria Ceron Breton, Carmine Tepedino, Joseph Quartieri, and Nikos E. Mastorakis. ARIMA models application to air pollution data in Monterrey, Mexico// AIP Conference Proceedings 1982, 020041, 2018; <https://doi.org/10.1063/1.5045447>.
10. Muhammad Hisyam Lee, Nur Haizum Abd. Rahman, Suhartono, Mohd Talib Latif, Ma-ria Elena Nor and Nur Arina Bazilah Kamisan. Seasonal

ARIMA for Forecasting Air Pollution Index: A Case Study // American Journal of Applied Sciences 9 (4): 570-578, 2012 ISSN 1546-9239, 2012.

11. Abhilash M.S.K., Thakur A., Gupta D., Sreevidya B. Time Series Analysis of Air Pollution in Bengaluru Using ARIMA Model. In: Perez G., Tiwari S., Trivedi M., Mishra K. (eds) Ambient Communications and Computer Systems. Advances in Intelligent Systems and Computing, vol 696. Springer, Singapore, 2018.

12. MacKinnon, J.G. Critical Values for Cointegration Tests // Queen's University, Dept of Economics, Working Papers, 2010.
<http://ideas.repec.org/p/qed/wpaper/1227.html>.

13. Ковальчук І.П., Ковальчук О.І. Моніторинг довкілля: навч. посібник. – Львів: Видавництво Львівської політехніки, 2013. – 312 с. ISBN 978-617-607-456-7.

14. Бондаренко В.І., Коваль С.М. Екологічний моніторинг атмосферного повітря: методичні основи та практичні аспекти. – Київ: Наукова думка, 2018. – 198 с. ISBN 978-966-00-1578-2.

15. Горова А.І., Козловський В.В. Моніторинг забруднення атмосферного повітря в Україні: стан, проблеми, перспективи // Екологічна безпека та природокористування. – 2021. – №1. – С. 45–52. ISSN 2413-7558.

16. Савицький О.І. Системи екологічного моніторингу: навч. посібник. – Івано-Франківськ: ІФНТУНГ, 2016. – 224 с. ISBN 978-966-641-634-3.

17. SaveEcoBot – онлайн-карта моніторингу якості повітря в Україні.
<https://www.saveecobot.com/maps>

18. Варламов Є.Н., Палагута О.А. Система екологічних індикаторних показників для оцінки стану навколишнього природного середовища в Україні // Регіональні геосистеми. – 2013. – №1. – С. 47–53. 19.

19. Савицький О.І. Системи екологічного моніторингу: навч. посібник. – Івано-Франківськ: ІФНТУНГ, 2016. – 224 с. ISBN 978-966-641-634-3
20. Ковальчук І.П., Ковальчук О.І. Моніторинг довкілля: навч. посібник. – Львів: Видавництво Львівської політехніки, 2013. – 312 с. ISBN 978-617-607-456-7
21. Бондаренко В.І., Коваль С.М. Екологічний моніторинг атмосферного повітря: методичні основи та практичні аспекти. – Київ: Наукова думка, 2018. – 198 с. ISBN 978-966-00-1578-2
22. Горова А.І., Козловський В.В. Моніторинг забруднення атмосферного повітря в Україні: стан, проблеми, перспективи // Екологічна безпека та природокористування. – 2021. – №1. – С. 45–52. ISSN 2413-7558