

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені ІГОРЯ  
СІКОРСЬКОГО»**

**Навчально-науковий фізико-технічний інститут  
Кафедра математичного моделювання та аналізу  
даних**

До захисту  
допущено:  
В.о. завідувача  
кафедри  
\_\_\_\_\_ Г.О. Яйлимова

«\_\_\_» \_\_\_\_\_ 2024р.

**Дипломна робота**  
на здобуття ступеня бакалавра  
зі спеціальності 113 «Прикладна Математика» на тему:  
«Екстраполяція якості повітря в міському середовищі з обмеженою  
кількістю станцій»

Виконала: студентка 4 курсу, групи ФІ-01

Городецька Дар'я Костянтинівна \_\_\_\_\_

Науковий керівник: ст. викладачка кафедри ММАД ННФТІ,

д-р філософії Яйлимова Ганна Олексіївна \_\_\_\_\_

Рецензент: к.т.н., ст. дослідник, зав. відділу ІКД НАНУ та ДКАУ

Яйлимов Богдан Ялкапович \_\_\_\_\_

Засвідчую, що у цій  
дипломній роботі немає  
запозичень з праць  
інших авторів без  
відповідних посилань

Студент \_\_\_\_\_

Київ – 2024 року

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ**  
**«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ**  
**імені Ігоря СІКОРСЬКОГО»**  
**НАВЧАЛЬНО-НАУКОВИЙ ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ**  
**Кафедра математичного моделювання та аналізу даних**

Рівень вищої освіти — перший (бакалаврський)

Спеціальність: 113 Прикладна математика,

Освітня програма: Математичні методи моделювання, розпізнавання образів та Комп'ютерного зору

ЗАТВЕРДЖУЮ

В. о. завідувача кафедри  
\_\_\_\_\_ Г.О. Яйлимова

«\_\_\_» \_\_\_\_\_ 2024р.

**ЗАВДАННЯ**

на дипломну роботу

Студентка: Городецька Дар'я Костянтинівна

1. Тема роботи: «Екстраполяція якості повітря в міському середовищі з обмеженою кількістю станцій»,  
науковий керівник роботи ст. викладачка кафедри ММАД ННФТІ, д-р філософії Яйлимова Г.О.,  
затверджені наказом по університету № \_\_\_\_\_ від «\_\_\_» \_\_\_\_\_ 2024 р
2. Термін подання студентом роботи: «\_\_\_» \_\_\_\_\_ 2024 р
3. Вихідні дані до роботи: дані з платформи SaveEcoBot за 2022 р. щодо показників якості повітря міста Київ.
4. Зміст роботи: аналіз світового досвіду розв'язання питання обмеженої кількості станцій моніторингу показників якості повітря, реалізація та оцінка результатів наявних методів екстраполяції даних, створення суцільної карти забруднень міста Київ.
5. Перелік ілюстративного матеріалу: презентація доповіді.
6. Дата видачі завдання: 1 лютого 2024 р.

### Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання	Примітка
1	Узгодження теми роботи із науковим керівником	Лютий 2024 р.	Виконано
2	Огляд проблеми та світової практики вимірювання якості повітря	Лютий – Березень 2024 р.	Виконано
3	Аналіз недоліків та прогалин в знайдених дослідженнях	Березень 2024 р.	Виконано
4	Збір та обробка даних щодо показників якості повітря за 2022 р.	Квітень 2024 р.	Виконано
5	Розробка та вдосконалення наявних моделей екстраполяції даних. Оцінка точності та порівняння реалізованих моделей	Травень 2024 р.	Виконано
6	Написання пояснювальної записки до дипломної роботи, створення презентації та захист дипломної роботи.	Травень – Червень 2024 р.	Виконано

Студент  
Науковий керівник роботи

\_\_\_\_\_  
Городецька Д.К.  
\_\_\_\_\_  
Яйлимова Г. О.

## РЕФЕРАТ

Дипломна робота за обсягом становить 45 сторінок, містить 1 таблицю та 11 рисунків, 4 додатки. Для дослідження було використано 11 бібліографічних найменувань.

З початком пандемії COVID-19 та повномасштабного вторгнення росії в Україну 24 лютого 2022 року, країна зіткнулася з багатьма проблемами, одна з яких досі не отримує належної уваги. Це проблема погіршення якості повітря. Пандемія спричинила численні зміни в екології, зокрема в якості повітря через зміну промислового виробництва та транспортних звичок. Воєнні дії призвели до додаткового забруднення навколишнього середовища, що також вплинуло на якість повітря. Внаслідок цих подій, місто Київ, як і багато інших регіонів України, зіткнулося з суттєвими викликами щодо моніторингу та оцінки стану якості повітря. Даних моніторингу недостатньо через малу кількість станцій заміру показників, тому для розв'язання поставленої задачі в дипломній роботі використовуються IDW модель екстраполяції даних та Griddata модель, яка використовується для екстраполяції нерегулярно розподілених даних на регулярну сітку, порівняльний аналіз моделей для визначення найкращої моделі.

В роботі досліджено дані показників якості повітря міста Київ за 2022 р. для навчання двох моделей екстраполяції даних.

Ключові слова: математичне моделювання, екологія, якість повітря, обмежена кількість станцій моніторингу, екстраполяція даних.

## ABSTRACT

The thesis is 45 pages long, and contains 1 table and 11 figures, 4 applications. The research was based on 11 bibliographic references.

With the outbreak of the COVID-19 pandemic and Russia's full-scale invasion of Ukraine on 24 February 2022, the country has faced many challenges, one of which still does not receive adequate attention. This is the problem of deteriorating air quality. The pandemic has caused numerous changes in the environment, including air quality due to changes in industrial production and transport habits. Military operations have led to additional environmental pollution, which has also affected air quality. As a result of these events, the city of Kyiv, like many other regions of Ukraine, has faced significant challenges in monitoring and assessing air quality. Monitoring data is insufficient due to the small number of measurement stations, so to solve this problem, the theses uses the IDW data extrapolation model and the Griddata model, which is used to extrapolate irregularly distributed data to a regular grid, and a comparative analyzes of the models to determine the best model.

In this paper, the data of air quality indicators for the city of Kyiv for 2022 are studied to train two data extrapolation models.

Keywords: mathematical modeling, ecology, air quality, limited number of monitoring statistics, data extrapolation.

## ЗМІСТ

<b>ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ І ТЕРМІНІВ.....</b>	<b>8</b>
<b>ВСТУП.....</b>	<b>9</b>
<b>РОЗДІЛ 1.....</b>	<b>11</b>
1.1 Постановка задачі .....	11
1.2 Значення якості повітря для міського середовища .....	11
1.3 Методи розв’язання питання обмеженої кількості станцій у світовій практиці .....	15
<b>Висновки до розділу 1.....</b>	<b>20</b>
<b>РОЗДІЛ 2.....</b>	<b>21</b>
2.1 Загальні відомості про екстраполяцію даних .....	21
2.2 IDW .....	23
2.3 Griddata .....	25
<b>Висновки до розділу 2.....</b>	<b>27</b>
<b>РОЗДІЛ 3.....</b>	<b>29</b>
3.1 Джерела даних для екстраполяції .....	29
3.2 Реалізація моделі IDW .....	32
3.3 Реалізація моделі Griddata.....	33
3.4 Аналіз та оцінка точності створених моделей .....	35
<b>Висновки до розділу 3.....</b>	<b>42</b>
<b>Висновки.....</b>	<b>44</b>
<b>Перелік використаних джерел.....</b>	<b>45</b>
<b>ДОДАТКИ .....</b>	<b>47</b>

<b>Додаток А. Перші рядки створеного датасету data.csv .....</b>	<b>47</b>
<b>Додаток Б. Програмний код обробки даних .....</b>	<b>49</b>
<b>Додаток В. Програмний код реалізації моделі IDW .....</b>	<b>50</b>
<b>Додаток Д. Програмний код реалізації моделі Griddata .....</b>	<b>51</b>

## **ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ І ТЕРМІНІВ**

IDW — Inverse Distance Weighting

TNGAPMS — Концепція Системи Моніторингу Забруднення Повітря Наступного Покоління

SSN — Статична Сенсорна Мережа

CSN — Спільнотна Сенсорна Мережа

VSN — Транспортна Сенсорна Мережа

GPS — Global Positioning System

ГІС — Географічні інформаційні системи

MSE — Середньоквадратична помилка

MAE — Середня абсолютна помилка

CO — Монооксид Вуглецю

SO<sub>2</sub> — Сірчистий Ангідрид

NO<sub>x</sub> — Оксиди Азоту

PM — Дрібні тверді частинки

PM<sub>2.5</sub> — Дрібні тверді частинки розміром менше 2.5 мк

PM<sub>10</sub> — Дрібні тверді частинки розміром менше 10 мк

МГУА — метод групового урахування аргументів

## ВСТУП

**Актуальність дослідження:** У наш час відбувається неперервний розвиток нових технологій, міст і збільшення частки міського населення. Однак цей процес супроводжується багатьма проблемами, такими як погіршення якості повітря, забруднення водойм, і зменшення зелених насаджень. Це становить одну з головних загроз для здоров'я міських мешканців, спричиняючи мільйони передчасних смертей у всьому світі. Тому моніторинг якості повітря стає надзвичайно важливим інструментом для забезпечення здоров'я населення і планування дій для покращення якості повітря. В Україні, як і в багатьох інших країнах, для моніторингу якості повітря традиційно використовуються наземні станції. Однак, з огляду на великі території міст і обмежену кількість таких станцій, існує важлива потреба в розробці та впровадженні більш комплексних методів оцінки якості повітря. Це стосується не лише великих міст, а й менших міських поселень, де моніторинг якості повітря часто є обмеженим або недоступним.

**Мета дослідження:** Дослідження розв'язання питання обмеженої кількості станцій. Реалізація та аналіз наявних методів для екстраполяції якості повітря в міському середовищі, де дані моніторингу обмежені.

**Завдання дослідження:**

1. Аналіз розв'язання питання обмеженої кількості станцій моніторингу показників якості повітря, спираючись на досвід закордонних колег.
2. Створення набору даних показників якості повітря міста Київ за 2022 р. та обробка їх для використання в подальшій роботі.
3. Розробка двох моделей екстраполяції даних, налаштування їх параметрів для отримання оптимальних результатів.
4. Оцінка результатів реалізованих моделей та оцінка їх ефективності, точності на реальних даних. Створення суцільної карти забруднень міста Київ за різні часові проміжки 2022 р.

**Об'єкт дослідження:** Показники якості повітря зі станцій моніторингу в

міському середовищі.

**Предмет дослідження:** : Методи математичного моделювання щодо екстраполяції даних про якість повітря, отриманих з обмеженої кількості станцій моніторингу міста Києва, включаючи методи IDW, Griddata.

**Наукова новизна** полягає у дослідженні досвіду розв'язання питання обмеженої кількості станцій моніторингу, створення набору даних показників якості повітря з усіх станцій міста Київ за 2022 р. та розробка двох моделей для екстраполяції даних про якість повітря, які поєднують в собі методи математичного моделювання. Вивчення ефективності різних методів екстраполяції та їх комбінацій для точного прогнозування якості повітря в міському середовищі. За допомогою реалізованих моделей створення суцільної карти забруднень міста Київ, яка надає інформацію щодо забруднення повітря в різні часові проміжки 2022 р.

**Практичне значення** полягає в розробці системи моніторингу якості повітря, яка дозволяє отримувати точні дані про стан повітря навіть в умовах обмеженої кількості станцій, що може бути корисним для муніципальних органів влади та екологічних організацій. Забезпечення можливості швидкого реагування на зміни якості повітря, що сприяє захисту здоров'я населення та покращенню екологічної безпеки в містах, особливо в умовах воєнного часу.

# РОЗДІЛ 1

## Огляд проблеми та світової практики вимірювання якості повітря

### 1.1 Постановка задачі

Метою даної дипломної роботи є реалізація та впровадження методів екстраполяції якості повітря в міському середовищі на основі обмеженої кількості станцій моніторингу. Задача екстраполяції якості повітря має важливі застосування в різних сферах, включаючи:

- забезпечення міських адміністрацій актуальними даними про якість повітря для прийняття рішень щодо заходів з покращення екологічної ситуації,
- надання даних для оцінки впливу забруднення повітря на здоров'я населення й зменшення негативних наслідків,
- створення систем раннього попередження про погіршення якості повітря та вжиття заходів для запобігання надзвичайним ситуаціям,
- оптимізація розміщення нових станцій моніторингу для підвищення ефективності та точності оцінки якості повітря.

### 1.2 Значення якості повітря для міського середовища

Атмосфера Землі — основа планети Земля, вона відіграє важливу роль у забезпеченні життєдіяльності рослин, бактерій, тварин, мікроорганізмів, людей. Ця тонка оболонка дозволяє зберігати тепло і вологу, створюючи умови для обміну кисню та дихання. Активності природних водних циклів напряду пов'язана зі станом атмосфери, до того ж вона захищає всі організми від шкідливого впливу ультрафіолетового випромінювання.

Згідно з дослідженням [1] 1950 року населення світу зросло більш ніж удвічі, а кількість автомобілів збільшилася в 10 разів. За цей же період частка міського населення зростає в 4 рази, і на початок 2000-х років майже половина населення світу проживала в містах. Близько 20 міських агломерацій мають

населення понад 10 мільйонів осіб.

Протягом тривалих періодів забруднення повітря у великих містах зазвичай зростає на етапі розростання міст, досягає піка, а потім зменшується завдяки впровадженню заходів з його зменшення. У розвинених країнах Західного світу забруднення повітря в містах зменшилося, зокрема рівні діоксиду сірки та сажі суттєво знизилися. Однак, в останні десятиліття зростання трафіку привернуло увагу до проблем з оксидами азоту, органічними сполуками та дрібними частинками. У деяких містах фотохімічне забруднення стало важливою проблемою, але у Північній Європі це явище здебільшого спостерігається у сільських районах, де рівень озону іноді вищий, ніж у міських центрах. Міста Східної Європи залишаються сильно забрудненими, хоча останні політичні зміни та впровадження нових технологій покращили ситуацію.

У країнах, що розвиваються, швидка урбанізація призвела до неконтрольованого зростання й погіршення стану навколишнього середовища, де рівень забруднення повітря все ще зростає. Окрім локального забруднення, міська діяльність є значним джерелом транскордонного забруднення та зростання глобальної концентрації парникових газів. Впровадження чистих, енергоефективних технологій сприятиме розв'язанню цих проблем, тоді як планування міста з розповсюдженням діяльності може призвести до збільшення трафіку і погіршення екологічної ситуації.

Небезпечні хімічні речовини потрапляють у навколишнє середовище внаслідок ряду природних та антропогенних дій і можуть спричинити негативний вплив на здоров'я людини та навколишнє середовище [2]. Збільшене спалювання викопного палива відповідає за прогресивну зміну складу атмосфери. Забруднювачі повітря, такі як CO, SO<sub>2</sub>, NO<sub>x</sub>, важкі метали та PM<sub>2.5</sub> і PM<sub>10</sub>, відрізняються за хімічним складом, реакційними властивостями, виділенням, часом розпаду та здатністю розповсюджуватись на великі чи короткі відстані. PM частинки можуть випромінюватися безпосередньо з різних джерел, таких як фабрики, електростанції, транспорт і

лісові пожежі, або вони можуть утворюватися в атмосфері в результаті хімічних реакцій в промисловості. Концентрації РМ можна розглядати як індикатор збільшення або зменшення економічної діяльності міста. Дослідження [3] має на меті показати результати аналізу розподілу часових рядів РМ<sub>2.5</sub> по містах України, а також перевірити кореляцію цих змін з даними ООН про біженців. Зміни в річних середніх концентраціях РМ<sub>2.5</sub> та РМ<sub>10</sub> у повітрі над міськими територіями можуть відображати зміни в кількості населення, яке проживає в цих зонах. Наприклад, якщо концентрації РМ знижуються, це може вказувати на зменшення кількості людей в цьому районі, можливо, через міграцію або виїзд мешканців. Навпаки, збільшення концентрацій РМ може свідчити про збільшення кількості людей, що може бути результатом міграції до цього району.

Забруднення повітря має гострий і хронічний вплив на здоров'я людини, впливаючи на ряд різних систем і органів, все починається від незначного подразнення верхніх дихальних шляхів до хронічних респіраторних і серцевих захворювань, раку легень, гострих респіраторних інфекцій у дітей і хронічного бронхіту у дорослих, загострення вже присутніх захворювань серця або астматичних нападів. Короткочасне та довгострокове опромінення також пов'язане з передчасною смертю та скороченням тривалості життя.

Не дивлячись, що ряд фізичних дій (вулкани, пожежі тощо) можуть вивільняти різні забруднювальні речовини в навколишнє середовище, антропогенна діяльність є основною причиною забруднення повітря. Небезпечні хімічні речовини можуть випадково потрапляти в навколишнє середовище, але деякі забруднювачі повітря викидаються з промислових об'єктів та інших видів діяльності та можуть негативно впливати на здоров'я людини та навколишнє середовище. Забруднювачем повітря є будь-яка речовина, яка може завдати шкоди людям, тваринам, рослинності чи матеріалам. Що стосується людей, забруднювач повітря може спричинити чи сприяти збільшенню смертності, чи серйозних захворювань або може становити поточну чи потенційну небезпеку для здоров'я людини. Визначення

того, чи становить речовина ризик для здоров'я людини, базується на клінічних, епідеміологічних дослідженнях та/або дослідженнях на тваринах, які демонструють, що вплив речовини пов'язаний із впливом на здоров'я.

За допомогою супутникових даних у моніторингу якості повітря, можливо виявити екологічні загрози та знайти підтримку політичних рішень щодо сталого розвитку міських територій. Для цього потрібно використання супутникових даних разом з іншими джерелами інформації для забезпечення більш комплексного моніторингу якості повітря, створення інструментів для підтримки міських планувальників та ухвалення рішень, які допомагають у вирішенні екологічних проблем та підвищенні якості життя у містах та взаємодія з громадськістю. Дослідження [4] вказує на те, що супутникові дані можуть бути використані для покращення розуміння міських екосистем та створення стабільних міст, що важливо в умовах зростання міського населення та посилення екологічних викликів, пов'язаних із забрудненням повітря.

Проте з 2022 року забруднення повітря від заводів, виробництв відійшло на другий план, бо справжньою проблемою наразі є воєнні дії на території України. Забруднення підземних вод, водойм, морів, підтоплення територій і просідання ґрунту, забруднення повітря — це все невичерпний список наслідків російського вторгнення в Україну. Саме тому зараз постало питання детального моніторингу повітря в міських умовах. Для цього можна використовувати станції моніторингу, які збирають різні показники повітря, ці системи оснащені високоточним обладнанням для вимірювання рівнів забруднювальних речовин у повітрі, таких як діоксид азоту, озон, тверді частки тощо. Наявних станцій на території України не вистачає для проведення повноцінного аналізу, до того ж подивившись за межі міст можна побачити, що станцій там взагалі немає. Проблема обмеженості виходить з того, що традиційні системи складаються з невеликої кількості станцій, розташованих у ключових локаціях міста або регіону. Однак, через високу вартість та складність обслуговування такого обладнання, їхня кількість залишається обмеженою, що призводить до недостатнього географічного покриття та втрати

інформації про якість повітря у багатьох районах. Для розв'язання цієї проблеми потрібно розглянути наявні рішення для моніторингу якості повітря на прикладі світового досвіду, а саме розробку та впровадження мобільних та низьковартісних сенсорних мереж, що дозволяють значно розширити мережу моніторингу шляхом збільшення кількості точок збору даних. Також використання методів математичного моделювання для прогнозування стану повітря на різних відстанях.

### **1.3 Методи розв'язання питання обмеженої кількості станцій у світовій практиці**

Проблема обмеженої кількості моніторингових станцій полягає в тому, що традиційні моніторингові системи складаються з невеликої кількості станцій, розташованих у ключових локаціях міста або регіону. Ці системи оснащені високоточним обладнанням для вимірювання рівнів забруднювальних речовин у повітрі, таких як діоксид азоту, озон, тверді частки тощо. Проте, через високу вартість та складність обслуговування такого обладнання, кількість таких станцій залишається обмеженою, що призводить до недостатнього географічного покриття та втрати інформації про якість повітря в багатьох районах.

Таким чином, ми можемо мати доступ до даних про стан повітря в деяких районах Києва, але за межами міста перегляд стану повітря буде проблематичним. Ситуація стає ще гіршою в інших містах та селах, де датчики відсутні взагалі. Для розв'язання цієї проблеми необхідно розглянути наявні рішення для моніторингу якості повітря, зокрема, світовий досвід у розробці та впровадженні мобільних та низьковартісних сенсорних мереж, які дозволяють значно розширити мережу моніторингу шляхом збільшення кількості точок збору даних.

Крім того, використання методів математичного моделювання для прогнозування стану повітря на різних відстанях може суттєво підвищити ефективність моніторингу. Це дозволить забезпечити більш точне та повне

розуміння стану повітря, що є надзвичайно важливим для прийняття обґрунтованих рішень щодо покращення екологічної ситуації та захисту здоров'я населення.

Оптимізація розташування станцій моніторингу якості атмосферного повітря має велике значення для забезпечення високоякісних даних регіонального моніторингу забруднення атмосферного повітря. Регулярний моніторинг якості повітря в міських районах має вирішальне значення для забезпечення здоров'я населення та підтримання екологічної безпеки. Для всебічного визначення стану якості повітря в міському середовищі було б ідеально мати фактично нескінченну кількість станцій моніторингу, які надають просторово безперервні дані про якість повітря. Однак завжди існують практичні обмеження, такі як бюджетні обмеження, для підтримки великої кількості станцій моніторингу. Таким чином, оцінка загального стану якості повітря міського середовища за допомогою мінімальної кількості станцій є важливою інженерною проблемою.

В Гонконзі моніторинг якості повітря відбувається за допомогою інноваційних технологій. Проблема нестачі датчиків була розглянута у дослідженні [5], вона розв'язується за допомогою TNGAPMS. Існують три основні типи сенсорних мереж на основі цієї моделі: SSN, CSN, VSN.

- SSN: статичні сенсори, які розміщуються на вуличних ліхтарях або стовпах. Вони характеризуються високою надійністю даних і можливістю використання різноманітних сенсорів на одному вузлі, але потребують планування розміщення та затрат на обслуговування. На Рис. 1.1 зображено приклад цієї моделі.

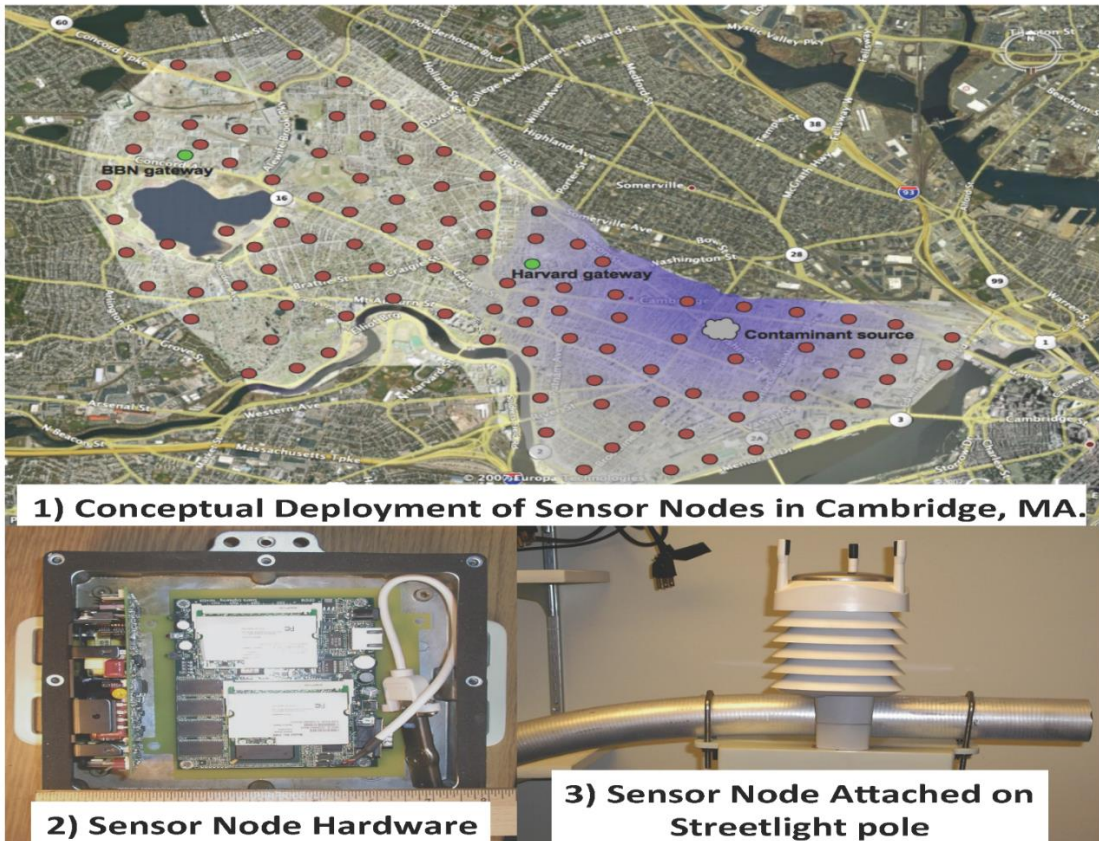


Рис. 1.1. Приклад архітектури та прототипу системи SSN

- CSN: мобільні сенсори, що носять з собою люди. Відсутня проблема частоти вимірів, але можлива низька точність даних та питань приватності. На Рис. 1.2 зображено приклад цієї моделі.

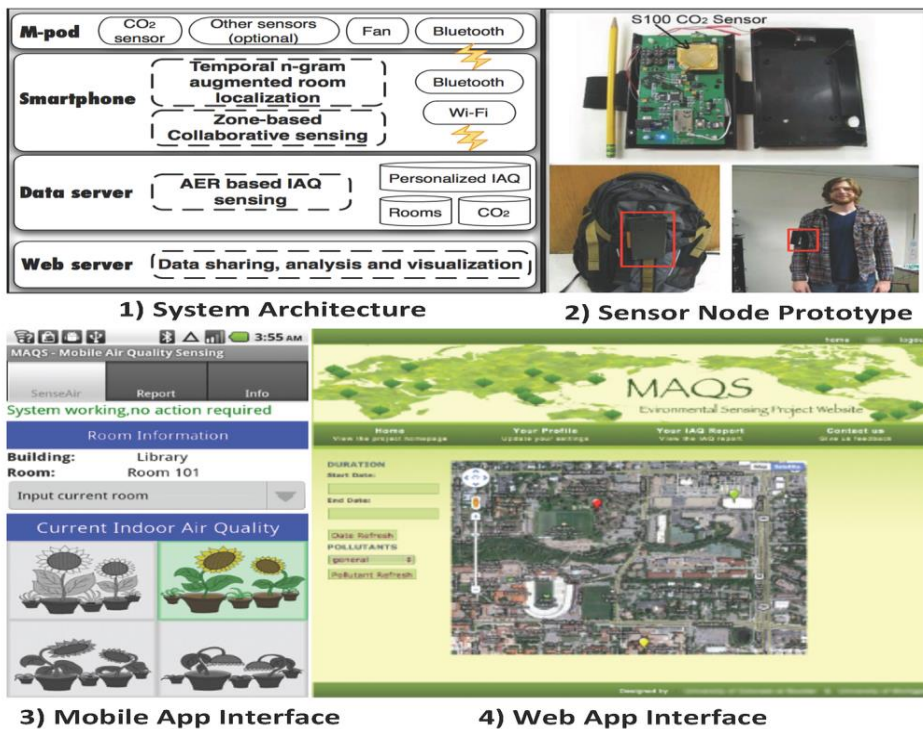


Рис. 1.2. Приклад архітектури та прототипу системи CSN

- VSN: сенсори на транспортних засобах (автобуси або таксі) для збору даних по широкій географічній території. На Рис. 1.3 зображено приклад цієї моделі.

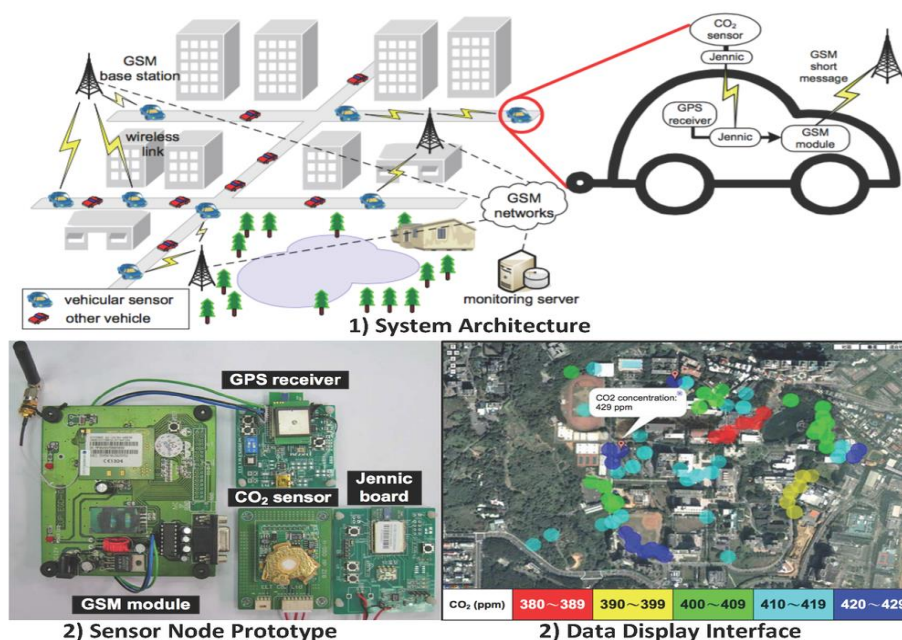


Рис. 1.3. Приклад архітектури та прототипу системи VSN

За допомогою цих мереж проблема розв'язується на державному рівні, бо існує велика кількість датчиків, що дає змогу створити повноцінне покриття, але введення даної системи в Україні неможливе. По-перше, досить висока вартість закупки та обслуговування обладнання. По-друге, наразі використання систем GPS буде небезпечним через можливість відстежувати якісь параметри у режимі реального часу.

Окрім великої кількості датчиків варто розглянути використання сучасних технологій для екстраполяції даних, не збільшуючи кількості датчиків. Нейромережеві технології здобули значну популярність у багатьох галузях, включаючи комп'ютерну лінгвістику та автоматичну обробку текстів природною мовою. Застосування нейронних мереж значно підвищує якість розв'язання стандартних задач, знижує обсяг роботи з текстами й дозволяє розв'язувати нові завдання, які були недоступні для стандартних методів машинного навчання. Метою статті [6] є представлення алгоритму вибору вхідних змінних і застосування глибоких нейронних мереж для короткострокового прогнозування показників вітру. Дрібні та глибокі нейронні

мережі в поєднанні з алгоритмом вибору вхідних змінних порівнюються в задачі короткострокового прогнозування вітру для набору різних місць. Результати показують, що ретельно відібрані глибокі нейронні мережі перевершують дрібні. Використання вибору вхідних змінних зменшує складність нейронної мережі та спрощує глибоке навчання нейронної мережі.

Для запобігання різких змін у показниках якості повітря необхідні ефективні та точні моделі прогнозування. Технологічний розвиток зробив значний внесок у розробку успішних моделей прогнозування на основі глибокого навчання. З високим відсотком успіху дослідження [7] запропонувало гібридну модель прогнозування на основі моделей згорткових нейронних мереж і рекурентних нейронних мереж. Запропонована гібридна модель була застосована до набору даних, який містить 14-параметричні великомасштабні метеорологічні дані, які ніколи не використовувалися для прогнозування якості повітря. Експериментальні результати порівнювалися з популярними глибоким навчанням, машинним навчанням і статистичними методами.

В питанні обмеженої кількості станцій важливу роль відіграє правильний підбір даних для математичних моделей. Одним з варіантів вибору даних постають супутникові знімки. Супутникові дані забезпечують глобальне покриття, що дозволяє отримувати інформацію про якість повітря навіть у важкодоступних та віддалених районах, де немає наземних станцій моніторингу, супутники забезпечують регулярні спостереження для відстежування зміни якості повітря у режимі реального часу або з невеликими інтервалами. Дослідження [8] підкреслює переваги використання супутникових даних для моніторингу якості повітря в умовах обмеженої кількості наземних станцій. Це забезпечує більш повну просторово-часову картину розподілу забруднення, покращує можливості моніторингу навколишнього середовища. Розглянуто математичну модель, яка дозволяє оцінювати якість повітря щодо PM<sub>2.5</sub> та PM<sub>10</sub> на основі супутникових спостережень, сфокусованих на місті Київ. Основним методом виступає МГУА, що дозволяє відновлювати нелінійні

функціональні залежності між часовими рядами супутникових та наземних змінних.

Знаючи точне місце розташування джерел забруднення, можна розробити та впровадити більш ефективні стратегії зниження забруднення, що сприятиме покращенню якості повітря та зменшенню негативного впливу на здоров'я населення. Нова математична модель [9] для точного визначення місця знаходження забруднювачів навколишнього середовища на основі даних сенсорів інтегрує дані з різних сенсорів, що значно підвищує точність і надійність виявлення забруднювачів. Використовуючи геометричні принципи, модель визначає місце розташування забруднювачів. Для вирішення екологічних проблем, особливо в міських умовах, де рівень забруднення може значно відрізнятись на невеликих відстанях, важливим є поєднання сенсорних даних з передовими математичними методами.

## **Висновки до розділу 1**

В даному розділі показано, що атмосфера Землі є критично важливою для підтримки життя на планеті, забезпечуючи оптимальні умови для існування різноманітних форм життя та природних процесів. Тому актуальним є впровадження сучасних технологій моніторингу та адаптація міжнародного досвіду для боротьби з атмосферним забрудненням.

Використання супутникових даних та інноваційних математичних моделей для моніторингу якості повітря є важливим елементом у процесі аналізу стану повітря. Супутникові спостереження надають можливість отримувати точні та детальні дані про забруднення атмосфери на великій території, що дозволяє оперативно реагувати на зміни та вживати необхідних заходів. Дані супутникових спостережень інтегруються з іншими джерелами інформації, що дозволяє отримувати більш комплексні дані про стан атмосфери.

## РОЗДІЛ 2

### Екстраполяція даних

#### 2.1 Загальні відомості про екстраполяцію даних

Сучасна практика розв'язання складних завдань вимагає повного аналізу всіх даних та встановлення функціональних залежностей між ними. Цей аналіз використовується для розробки математичних моделей, які дозволяють прогнозувати поведінку системи в різних ситуаціях. Проте дуже часто безпосередні вимірювання є неможливими або обмеженими. Тоді застосовується екстраполяція даних, яка ґрунтується на використанні вже наявних даних для узагальнення та передбачення значень у невідомих або майбутніх умовах. Цей підхід дозволяє здійснювати обґрунтовані прогнози та приймати важливі рішення в умовах невизначеності.

Екстраполяція даних — це метод, який використовується для прогнозування даних за межами відомої інформації. Цей процес залучає моделі, які були адаптовані до наявних даних, і використовується для передбачення значень в невідомому майбутньому. Екстраполяція даних грає важливу роль в різних наукових дослідженнях та проєктах у різних галузях, бо вона дозволяє розширювати наше розуміння процесів та явищ за межами вже наявних даних. Екстраполяція даних відіграє особливу роль, коли мова йде про моніторинг якості повітря в міському середовищі, бо традиційні системи моніторингу з обмеженою кількістю станцій не можуть забезпечити повне географічне покриття. Моделі машинного та глибокого навчання можна використовувати для прогнозування якості повітря в районах без станцій моніторингу. Це не тільки покращує загальне розуміння екологічної ситуації, але й дозволяє швидше реагувати на можливі проблеми.

Використання даних із різних джерел дозволяє точніше екстраполювати моделі, що сприятиме більш точному моніторингу якості повітря. Наукові дослідження для прогнозування результатів за умов, коли фактичне

спостереження неможливе, для прогнозування економічних показників, цін на акції, інфляції тощо технічних систем і пристроїв для прогнозування поведінки систем за нових умов.

Один з основних методів екстраполяції — лінійна екстраполяція, яка використовує лінійні моделі для прогнозування даних, що демонструють прямі тенденції. Однак, для більших наборів даних можуть застосовуватися більш розширені методи, такі як поліноміальна або нелінійна екстраполяція. Поліноміальна екстраполяція використовує моделі вищих ступенів для розв'язання задач, де взаємозв'язки між даними є складнішими. Нелінійна екстраполяція застосовується до даних, які включають складні, непередбачувані зміни.

Найбільша проблема екстраполяції даних полягає у можливості отримання ненадійних прогнозів, коли дані мають велику варіативність або коли прогнозування проводиться на значно далеку відстань за межі доступних даних. Прогнози можуть бути чутливі до припущень, які використовуються в моделях, маленькі зміни у вихідних даних або у параметрах моделі можуть вплинути на результати екстраполяції. Тому перед застосуванням екстраполяції необхідно ретельно обробити вхідні дані, а також урахувати можливість впливу непередбачуваних факторів на результати прогнозування.

Розробка математичної моделі та її моделювання є ключовими етапами екстраполяції даних. Математичні моделі потрібні для аналізу та передбачення поведінки досліджуваних об'єктів поза межами спостережуваних даних. Точність прогнозів залежить від того, наскільки добре модель відтворює реальні процеси та взаємодії між змінними. Для цього потрібно проаналізувати поведінку моделі в межах відомих даних і згодом випробувати її поза спостережуваним діапазоном.

## 2.2 IDW

Основна ідея методу полягає в тому, що вплив відомих точок на прогнозовану точку зменшується зі збільшенням відстані. Значення у невідомій точці визначаються з використанням значень відомих точок, але кожне з цих значень зважується відповідно до відстані до невідомої точки. Ближчі точки мають більший вплив на прогнозоване значення, ніж ті, що знаходяться далі. Це робиться шляхом використання вагових коефіцієнтів, які є обернено пропорційними до відстані між відомими та невідомими точками. IDW широко використовується в ГІС, екології, метеорології та інших областях, де потрібна ефективна екстраполяція даних між відомими точками. Метод IDW використовується тоді, коли важлива просторова точність передбачення. Екологічні дослідження можуть використовувати його для створення детальних карт забруднень на основі обмеженої кількості вимірювальних станцій. У метеорології IDW застосовується для прогнозування погоди в областях, де дані спостережень обмежені, таким чином забезпечуючи більше покриття.

Застосовуючи його в ГІС можна створювати карти, які відображають просторовий розподіл різних показників залежно від отриманої задачі. Це корисно в управлінні природними ресурсами, міському плануванні та інших сферах, де важливе розуміння просторових варіацій даних.

Проте варто враховувати, що вплив відомої точки плавно зменшується зі збільшенням відстані, що не завжди точно в реальних ситуаціях. Природні бар'єри (гори та річки) можуть впливати на розподіл забруднювальних речовин і погодні умови, але не враховуються в методі IDW. В таких випадках варто використовувати інші методи екстраполяції, наприклад, кригінг, який враховує просторову кореляцію між точками даних.

Формула 2.1 для прогнозування значення в точці  $p$  за допомогою IDW виглядає наступним чином:

$$V(p) = \frac{\sum_{i=1}^n \left( \frac{V_i}{d(p, p_i)^p} \right)}{\sum_{i=1}^n \left( \frac{1}{d(p, p_i)^p} \right)}, \quad (2.1)$$

де  $V(p)$  — інтерпольоване значення в точці  $p$ ,

$V_i$  — значення в  $i$ -тій відомій точці,

$d(p, p_i)$  — відстань між точкою  $p$  та  $i$ -тою відомою точкою,

$p$  — параметр ступеня, який регулює вплив відстані на вагові коефіцієнти,

$n$  — кількість відомих точок, які використовуються для інтерполяції.

Параметр  $p$  у формулі 2.1 визначає, наскільки сильно вплив точок зменшується зі збільшенням відстані. Частіше за все,  $p$  обирають у діапазоні від 1 до 3. Вибір значення  $p$  залежить від специфіки даних та мети аналізу:

$p = 1$  дає лінійне зменшення ваги з відстанню,

$p = 2$  дає квадратичне зменшення, що є звичним для багатьох прикладів,

Вищі значення  $p$  збільшують вплив найближчих точок і значно зменшують вплив точок, що знаходяться далі. IDW має певні вимоги, щоб забезпечити адекватність та точність отриманих результатів. Нижче наведено шість основних вимог, які потрібно врахувати під час реалізації цього методу:

- 1) Дані повинні мати однакову варіантність по всьому діапазону значень. Вплив точок на прогнозовану точку не повинен змінюватися залежно від величини або рівня змінної.
- 2) Точки даних повинні бути незалежними одна від одної.
- 3) IDW є нелінійним методом, проте припущення про лінійну залежність між зваженими значеннями і прогнозованими результатами допомагає спростити модель і забезпечити її стабільність.
- 4) Дані повинні поширюватись нормально або мають розподіл, близький до нормального. Це спрощує аналіз та інтерпретацію результатів.
- 5) Для ефективного використання IDW потрібно, щоб дані були рівномірно розподілені по досліджуваній області. Рідкісні або нерівномірно розподілені точки можуть вести до неправильної інтерпретації результатів.

- б) Дані повинні бути вільними від аномалій або викидів, оскільки екстремальні значення можуть сильно вплинути на результати через велику вагу, яку вони отримують у залежності від відстані.

## 2.3 Griddata

Цей метод застосовується для визначення значень у точках, де дані відсутні, на основі значень у відомих точках. Наприклад, якщо ми маємо інформацію про значення в кількох точках на площині, ми можемо використовувати цей метод для прогнозування значень у проміжних точках. Функція Griddata є частиною бібліотеки SciPy і пропонує кілька методів екстраполяції: лінійна, кубічна та метод найближчого сусіда. Ідея даного методу враховує розташування відомих точок і їхні значення, щоб передбачити, яке значення має бути у невідомих точках, розташованих між ними. Це дозволяє створювати більш точну картину розподілу значень у просторі, навіть якщо деякі дані відсутні. Він використовується в ГІС, екології, метеорології та інших галузях, де потрібна екстраполяція даних між відомими точками. Його варто використовувати, коли необхідно забезпечити точність просторових прогнозів. Лінійна інтерполяція є одним з найпростіших і найпоширеніших методів, вона заснована на тому, що значення між відомими точками змінюються лінійно, тобто прогнозоване значення у невідомій точці є лінійною комбінацією значень у найближчих відомих точках. Кубічна інтерполяція використовує кубічні поліноми для розрахунку значень у невідомих точках, значення будуть змінюватись не лінійно, а згідно з кубічним поліномом, що дозволяє отримати більш плавний та природний перехід між точками. Інтерполяція найближчого сусіда передбачає, що значення у невідомій точці дорівнює значенню найближчої відомої точки. Цей метод не враховує інші точки, крім найближчої. Загалом кожен з цих методів має свої переваги та недоліки, і вибір конкретного методу залежить від специфіки даних та вимог до точності і швидкості обчислень. У таблиці 2.1 наведений

порівняльний аналіз трьох видів інтерполяції:

<b>Вид</b>	<b>Лінійна інтерполяція</b>	<b>Кубічна інтерполяція</b>	<b>Інтерполяція найближчого сусіда</b>
<b>Простота реалізації</b>	Помірна	Складна	Проста
<b>Точність</b>	Помірна для лінійних змін	Висока для нелінійних змін	Низька для плавних змін
<b>Обчислювальна складність</b>	Середня	Висока	Низька

Таблиця 2.1. Порівняльний аналіз трьох видів інтерполяції

Формула для прогнозування значення в точці  $p$  за допомогою Griddata залежить від обраного методу інтерполяції. У формулі 2.2 наведено розрахунок значення в точці  $p$  для лінійної інтерполяції:

$$V(p) = \frac{\sum_{i=1}^n \left( \frac{v_i}{d(p, p_i)} \right)}{\sum_{i=1}^n \left( \frac{1}{d(p, p_i)} \right)},$$

(2.2)

де  $V(p)$  — інтерпольоване значення в точці  $p$ ,

$v_i$  — значення в  $i$ -тій відомій точці,

$d(p, p_i)$  — відстань між точкою  $p$  та  $i$ -тою відомою точкою,

$n$  — кількість відомих точок, які використовуються для інтерполяції.

У формулі 2.3 наведено розрахунок значення в точці  $p$  для кубічної інтерполяції:

$$V(p) = \frac{\sum_{i=1}^n \left( \frac{V_i}{d(p, p_i)^3} \right)}{\sum_{i=1}^n \left( \frac{1}{d(p, p_i)^3} \right)}, \quad (2.3)$$

де  $V(p)$  — інтерпольоване значення в точці  $p$ ,

$V_i$  — значення в  $i$ -тій відомій точці,

$d(p, p_i)$  — відстань між точкою  $p$  та  $i$ -тою відомою точкою,

$n$  — кількість відомих точок, які використовуються для інтерполяції.

У формулі 2.4 наведено розрахунок значення в точці  $p$  для інтерполяції найближчого сусіда:

$$V(p) = V \text{ nearest } (p), \quad (2.4)$$

де  $V(p)$  — інтерпольоване значення в точці  $p$ ,

$V \text{ nearest } (p)$  — значення найближчої відомої точки до точки  $p$ .

Щоб забезпечити точність отриманих результатів, під час реалізації методу Griddata потрібно враховувати такі вимоги:

- 1) Дані повинні мати однакову варіантність по всьому діапазону значень, і вплив точок на прогнозовану точку не повинен змінюватися залежно від величини або рівня змінної.
- 2) Точки даних повинні бути незалежними одна від одної.
- 3) Дані повинні поширюватись нормально або мати розподіл, близький до нормального.
- 4) Дані мають бути рівномірно розподілені по досліджуваній області. Нерівномірно розподілені точки можуть призвести до неправильної інтерпретації результатів.
- 5) Дані повинні бути очищені від аномалій або викидів, бо екстремальні значення можуть сильно вплинути на результати інтерполяції.

## Висновки до розділу 2

У розділі описані теоретичні основи для екстраполяції даних. Було розглянуто методи IDW та Griddata, зроблений порівняльний аналіз трьох видів

інтерполяції даних. Особливу увагу було приділено методам математичного моделювання, які забезпечують засоби для розуміння та прогнозування поведінки складних систем в різних умовах. Метод IDW використовується для прогнозування значень у точках, де дані відсутні, з урахуванням зменшення впливу відомих точок зі збільшенням відстані до них. Було розглянуто його основні характеристики, переваги та обмеження. Метод Griddata дозволяє передбачити дані з використанням різних підходів, таких як лінійна, кубічна інтерполяція та метод найближчого сусіда. Кожен з цих підходів має свої переваги та недоліки, які були детально розглянуті в роботі. Завдяки перегляду різних методів екстраполяції, включаючи лінійну та нелінійну екстраполяцію, було встановлено основу для наступної практичної частини дослідження.

Наступний етап дослідження передбачає застосування розглянутих методів до фактичних даних показників якості повітря в місті Київ. Це дозволяє протестувати ефективність і точність кожного методу в різних умовах. Практичне застосування методів екстраполяції включає аналіз точності прогнозів, порівняння результатів та визначення найкращого підходу для конкретного типу даних.

Під час реалізації потрібно буде використовувати новітні технології і програмне забезпечення для реалізації цих методів. Програмні пакети та бібліотеки Python, такі як SciPy та NumPy, надають потужні інструменти для виконання інтерполяції та екстраполяції даних. За допомогою цих інструментів процес аналізу автоматизується і стає більш ефективним.

В майбутньому результати цього дослідження можуть бути застосовані для вирішення практичних завдань в різних галузях, таких як екологія, питання якості повітря та вплив подій на це, планування подій.

## РОЗДІЛ 3

### Побудова та оцінка моделей

#### 3.1 Джерела даних для екстраполяції

В нашому світі велику роль у розробці сучасних рішень для глобальних викликів екології відіграють наукові дослідження, які засновані на використанні передових технологій. У контексті екологічних досліджень, зокрема екстраполяції якості повітря у міських умовах, важливою задачею стає використання точних даних і надійних методів аналізу. Наявність обмеженої кількості моніторингових станцій ще більше підкреслює важливість застосування розвинутих математичних моделей та алгоритмів для забезпечення якісних прогнозів. Сьогодні існує велика кількість підходів, які дозволяють розв'язувати специфічні задачі аналізу якості повітря, забезпечуючи високу точність результатів і можливість їх практичного застосування.

Ефективний моніторинг якості повітря вимагає не лише точності та надійності, а й різноманіття використовуваних джерел даних. Для даного дослідження було обрано інтегративний підхід до збору даних для того, щоб створити повноцінну картину стану атмосферного повітря міста Київ. Проте викликом стала обмежена кількість станцій моніторингу, тому виникла задача створення моделей екстраполяції даних. Для аналізу були обрані дані по якості повітря за 2022 рік. Це дало змогу оцінити зміни, які відбулись на фоні завершення пандемії Covid-19 та початку повномасштабного вторгнення росії на територію України.

Використовуючи платформу SaveEcoBot [10], яка агрегує інформацію з різних моніторингових станцій, включно з державними системами, місцевими ініціативами та комерційними пристроями, було зібрано дані з різних точок. Дані були піддані ретельній перевірці та очищенню для забезпечення їхньої наповненості та надійності. Процес обробки включав корекцію форматів даних,

видалення нечислових значень і викидів, а також інтеграцію декількох джерел в єдиний датасет, оскільки було важливо зробити найбільше покриття Києва. Наступним етапом було додавання географічних координат кожної станції та об'єднання інформації в єдиний формат для подальшого аналізу. Перші рядки набору даних надано в додатку А. Отже, датасет містить наступну інформацію:

1. `device_id`: Ідентифікатор пристрою, що збирає дані. Цей стовпець містить унікальний код або номер, який дозволяє ідентифікувати кожен датчик у мережі моніторингу, використовується для відстеження вихідних даних кожного пристрою, виявлення потенційних проблем з обладнанням або розміщенням пристроїв. Розділення даних на навчальну та тестову частину відбувалось саме по індивідуальному номеру щоб виключити однакові дані в різних частинах.

2. `phenomenon`: Цей стовпець описує тип феномену, який вимірюється. Це можуть бути такі показники як рівень діоксиду азоту, озону, твердих часток PM2.5 або PM10, вологість, температура та різні значення. Знання феномену дозволяє розуміти, які забруднювачі або параметри повітря аналізуються.

3. `value`: Конкретне значення, зареєстроване датчиком. Цей стовпець містить кількісні дані, що відображають інтенсивність або концентрацію забруднювача, зазначену у стовпці `phenomenon`. Дані використовуються для оцінки якості повітря та розрахунку індексу якості повітря в певному місці.

4. `logged_at`: Дата та час реєстрації кожного виміру. Цей стовпець дозволяє аналізувати дані у часовій послідовності та спостерігати за динамікою змін якості повітря, за допомогою нього можна виявити циклічні зміни або тренди пов'язані з певними годинами доби, днями тижня, сезонами року.

5. `latitude` та `longitude`: Географічні координати, де знаходиться датчик. Ці стовпці містять важливі просторові дані, що дозволяють аналізувати географічний розподіл якості повітря, ідентифікувати особливо забруднені зони або регіони, де якість повітря є кращою. Вони також допомагають інтегрувати дані моніторингу в геопросторові системи аналізу та картографічні сервіси

Після інтеграції, фінальний датасет піддавався структуруванню та аналізу для ідентифікації випадків, які могли б вплинути на моніторинг і прогнозування якості повітря. Програмний код обробки даних надано в додатку Б. Розглянемо опис функцій обробки даних:

1. `clean_data(data)`: Функція перетворює дані у колонках `latitude`, `longitude`, і `value` до числового формату, замінюючи нечислові значення на `NaN` для легшого виявлення та видалення. В результаті, усі дані в цих колонках стають числовими і повними.

2. `filter_data(data)`: Ця функція застосовує фільтри за географічними координатами та категоріями, вибираючи дані в заданому діапазоні широти і довготи, а також фільтруючи по феномену `'pm10'` для фокусування на певній забруднювальній речовині.

3. `remove_outliers(data)`: Функція видаляє викиди з датасету за допомогою IQR, видаляючи значення за межами  $[Q1 - 1.5IQR, Q3 + 1.5IQR]$ , що допомагає уникнути аномалій у даних.

4. `scale_data(data)`: Функція застосовує масштабування Min-Max до значень `'value'`, нормалізуючи їх у діапазон від 0 до 1 для підготовки до статистичного аналізу та машинного навчання.

Важливим зауваженням є те, що розділення даних на навчальну (70 %) та тестову вибірку (30 %) відбувалось за унікальним ідентифікатором. Це дало змогу уникнути однакових станцій в різних вибірках, що приводило б до некоректної роботи моделі. Включення даних за конкретний часовий проміжок дозволило виявити як короткотермінові, так і довготривалі зміни в якості повітря, залежно від різних факторів. На Рис. 3.1 зображено розподіл станцій на мапі Києва.

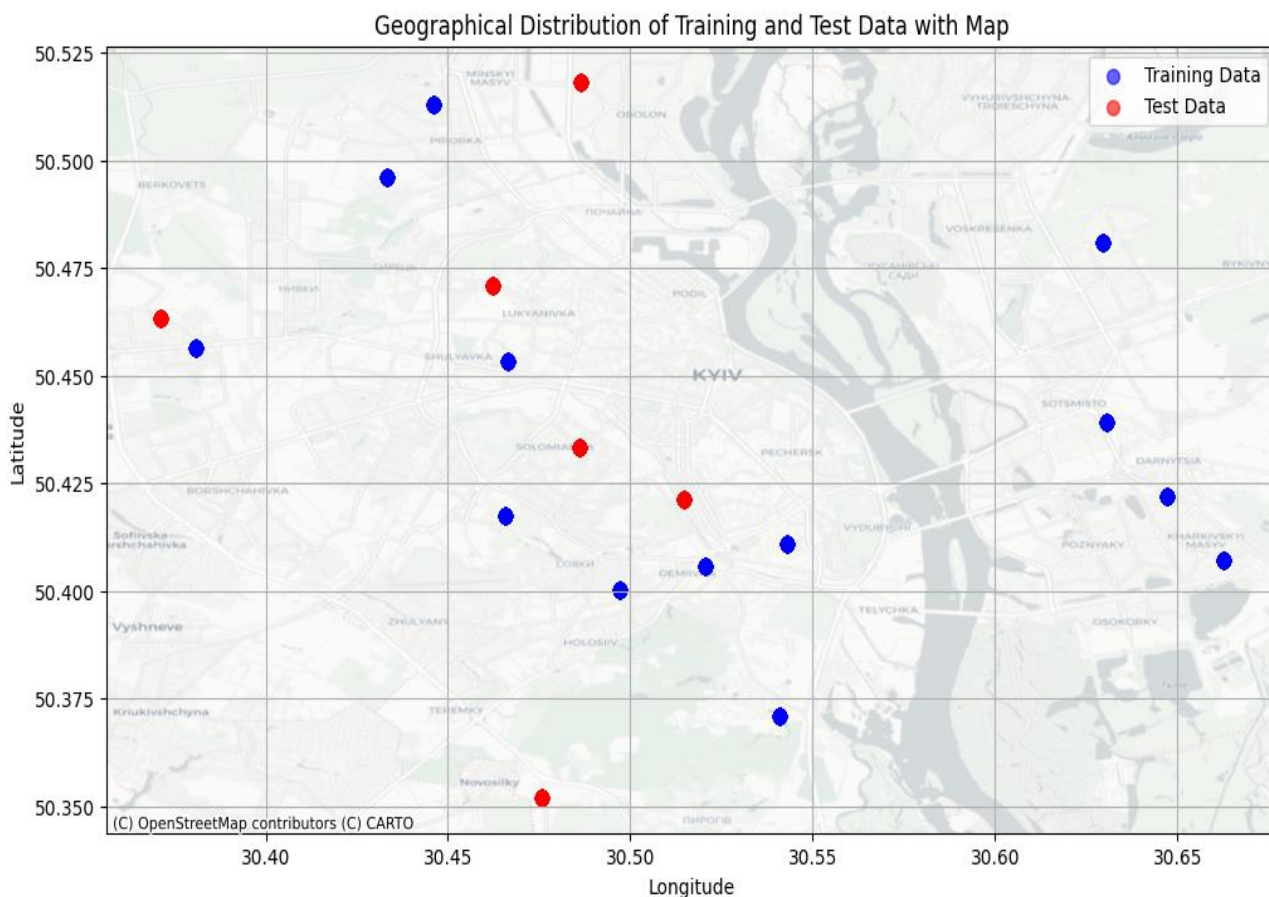


Рис. 3.1. Географічний розподіл навчальних і тестових даних зі значеннями показників якості повітря

Загалом можна побачити, як інтегративний підхід до збору та аналізу даних може значно підсилити якість та охоплення моніторингу якості повітря, забезпечуючи більш комплексну оцінку для рішень, які впливають на громадське здоров'я та політику.

### 3.2 Реалізація моделі IDW

Процес розробки моделі екстраполяції якості повітря на основі методу IDW та стратифікованої крос-валідації включає кілька ключових етапів, що допомагають забезпечити високу точність та надійність прогнозів. Підготовка до моделювання починається зі збору та обробки вихідних даних про якість повітря, таких як рівні забруднювачів, метеодані та географічні координати. Після цього до оброблених даних застосовується відповідна модель екстраполяції якості повітря, вона реалізована через IDW, він використовується

для передбачення значень в невідомих точках, базуючись на відомих значеннях з навколишніх місць. Цей метод заснований на припущенні, що вплив забруднювального фактора зменшується зі збільшенням відстані від його джерела.

Використовуючи початкові координати точок (широта та довгота), параметри інтерполяції (ступінь  $p$  та радіус згладжування) ініціалізуються. Ці параметри визначають вплив відстані на ваги та границі, в межах яких точки будуть впливати на результат. Для кожної точки, яку потрібно передбачити, обчислюються відстані до відомих точок. Ваги для кожної відомої точки визначаються як обернена величина відстані до степеня  $p$ , обмежена радіусом згладжування. Якщо відстань перевищує радіус, вага ставиться рівною нулю, щоб уникнути впливу дуже далеких точок. Ваги нормалізуються шляхом поділу суми вагованих значень на суму ваг, що забезпечує отримання зваженого середнього значення. Це зважене середнє використовується як інтерпольоване значення для кожної цільової точки.

Далі проводиться навчання моделі, під час якого параметри методу IDW оптимізуються за допомогою тренувальних даних. Оптимізація параметрів дозволяє моделі максимально точно відтворити вихідні дані. Використовуючи методи, такі як крос-валідація, можна визначити оптимальні значення параметрів  $p$  та радіуса згладжування для конкретних даних. Метрики MSE та MAE допомагають оцінити якість моделі. Програмний код реалізації моделі надано в додатку В. В кінці побудовано графіки географічного розподілу тренувальних даних зі значеннями показників якості повітря.

### **3.3 Реалізація моделі Griddata**

Метод Griddata є числовим методом екстраполяції просторових даних, який використовується для прогнозування значень у точках, де дані відсутні, на основі значень у відомих точках. В процесі роботи було реалізувати три види моделі Griddata: лінійна, кубічна та метод найближчого сусіда. Це було зроблено, щоб зрозуміти який підхід буде найкращим для даного набору даних,

оскільки залежність між показниками зі станцій моніторингу якості повітря та географічними координатами не має певну природу. Використовувалась вбудована функція `Griddata` з використанням бібліотеки `SciPy`. Програмний код надано в додатку Д. Реалізації прогнозування даних за допомогою методу `Griddata` включала в себе:

- 1) Метод `Griddata` є числовим методом екстраполяції просторових даних, який використовується для прогнозування значень у точках, де дані відсутні, на основі значень у відомих точках. В процесі роботи було реалізувати три види моделі `Griddata`: лінійна, кубічна та метод найближчого сусіда. Це було зроблено, щоб зрозуміти який підхід буде найкращим для даного набору даних, оскільки залежність між показниками зі станцій моніторингу якості повітря та географічними координатами не має певну природу. Використовувалась вбудована функція `Griddata` з використанням бібліотеки `SciPy`. Програмний код надано в додатку Д. Реалізації прогнозування даних за допомогою методу `Griddata` включала в себе:
- 2) Створюється регулярна сітка координат у межах заданого діапазону широти (від 50.30 до 50.60) і довготи (від 30.20 до 30.70) з роздільною здатністю 200 точок по кожній осі. Вона буде використовуватися для пошуку значень якості повітря у всіх точках на сітці.
- 3) Задання списку методів інтерполяцій: метод найближчого сусіда (`nearest`), лінійна інтерполяція (`linear`) та кубічна інтерполяція (`cubic`).
- 4) Для кожного методу інтерполяції виконується інтерполяція значень якості повітря в точках сітки. Використовується функція `Griddata` з бібліотеки `SciPy`, яка приймає координати (широту та довготу) і відповідні значення якості повітря з навчальної вибірки. Метод інтерполяції обирається відповідно до поточного методу зі списку. Функція `Griddata` виконує інтерполяцію значень у точках сітки на основі обраного методу.
- 5) Прогнозування значень якості повітря у точках тестової вибірки. Цей крок дозволяє оцінити точність моделі, порівнюючи прогнозовані значення з фактичними значеннями тестової вибірки.

- 6) Обчислення MSE та MAE для оцінки точності моделі інтерполяції. MSE визначає середнє квадратичне відхилення між прогнозованими та фактичними значеннями, тоді як MAE вимірює середню абсолютну різницю між прогнозованими та фактичними значеннями. Це дозволяє оцінити, наскільки добре модель інтерполяції прогнозує значення якості повітря.
- 7) На екран виводяться значення MSE та MAE для поточного методу інтерполяції, що дозволяє порівняти точність різних методів.
- 8) Візуалізація результатів інтерполяції. Кожен метод інтерполяції візуалізується окремо. Створюється контурний графік, що показує розподіл значень якості повітря у точках сітки. На графіку також позначаються точки розташування станцій моніторингу, що дозволяє візуально оцінити відповідність інтерпольованих значень фактичним даним.

### **3.4 Аналіз та оцінка точності створених моделей**

Для оцінки результатів моделі IDW було побудовано графіки географічного розподілу тренувальних даних зі значеннями показників якості повітря. Побудова графіків включає кілька кроків, що дозволяють візуалізувати результати прогнозування якості повітря за допомогою моделі IDW:

- Підготовка даних для використання у моделі. Координати та значення забруднювачів отримуються з тренувальних даних.
- Створення сітки координат для прогнозування, що включає широкий діапазон значень широти та довготи. Ця сітка буде використовуватись для створення графічного представлення прогнозованих значень на всій області.
- Використання створеної моделі для прогнозування значень забруднювачів на цій сітці координат. Прогнозовані значення отримуються для кожної точки сітки.
- Отримані прогнозовані значення перетворюються назад у форматі сітки, що відповідає розміру сітки координат.

- Візуалізація результатів. Візуалізація ілюструє просторове розподілення концентрацій PM10, де червоні точки вказують на розташування моніторингових станцій. Контурні лінії і кольорова шкала представляють екстрапольовані рівні забруднення повітря, дозволяючи візуально оцінити області з вищою та нижчою концентрацією забруднювачів.

За часовий проміжок було обрано 3 періоди:

- січень 2022 року. Час завершення пандемії Covid-19 та ще відсутність повномасштабного вторгнення. MSE (0.08503895425404585), MAE (0.24033852786960433). На Рис. 3.2 зображено карту забруднень.

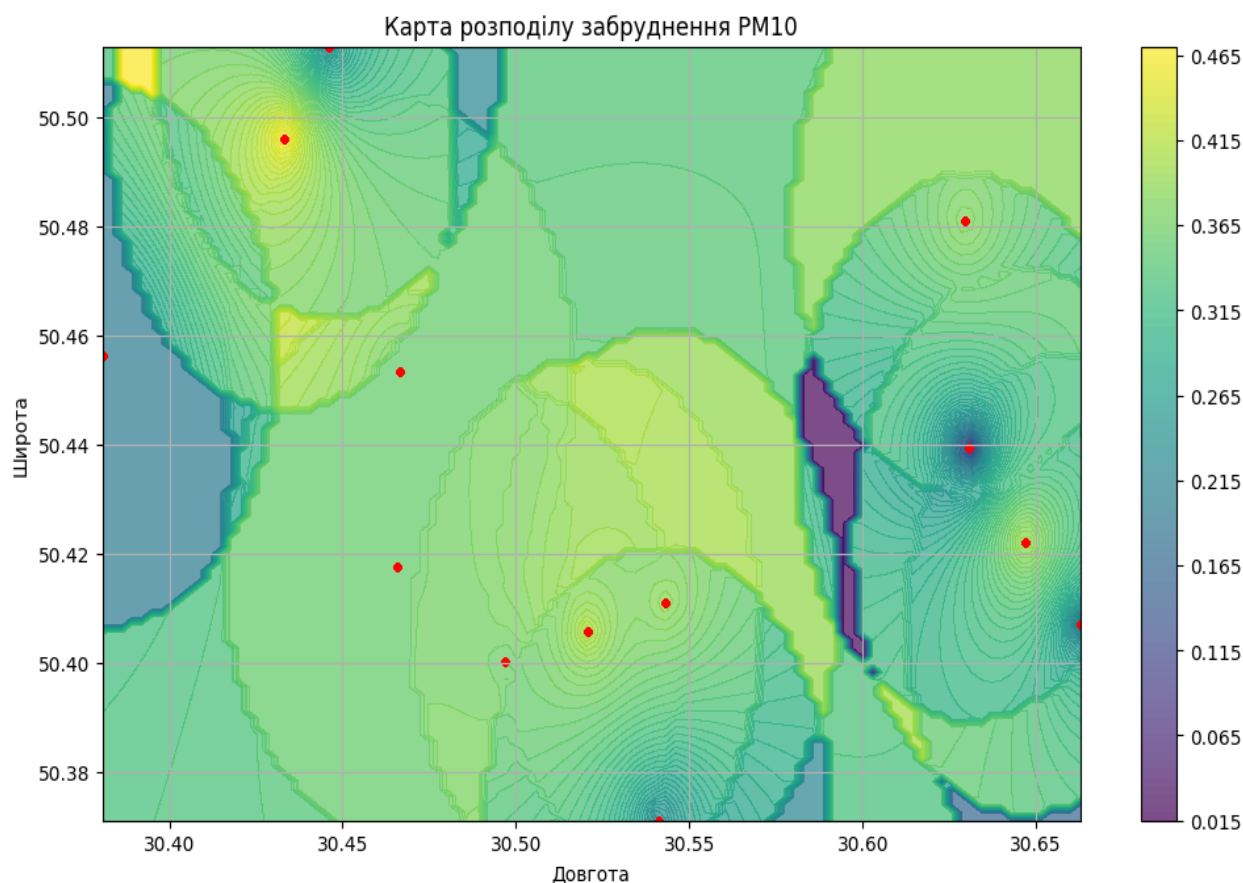


Рис. 3.2. Географічний розподіл тренувальних даних за січень 2022 року зі значеннями показників якості повітря

- березень 2022 року. Початок повномасштабного вторгнення росії на території України. MSE (0.11103703438436767), MAE (0.31219274785821566). На Рис. 3.3 зображено карту забруднень.

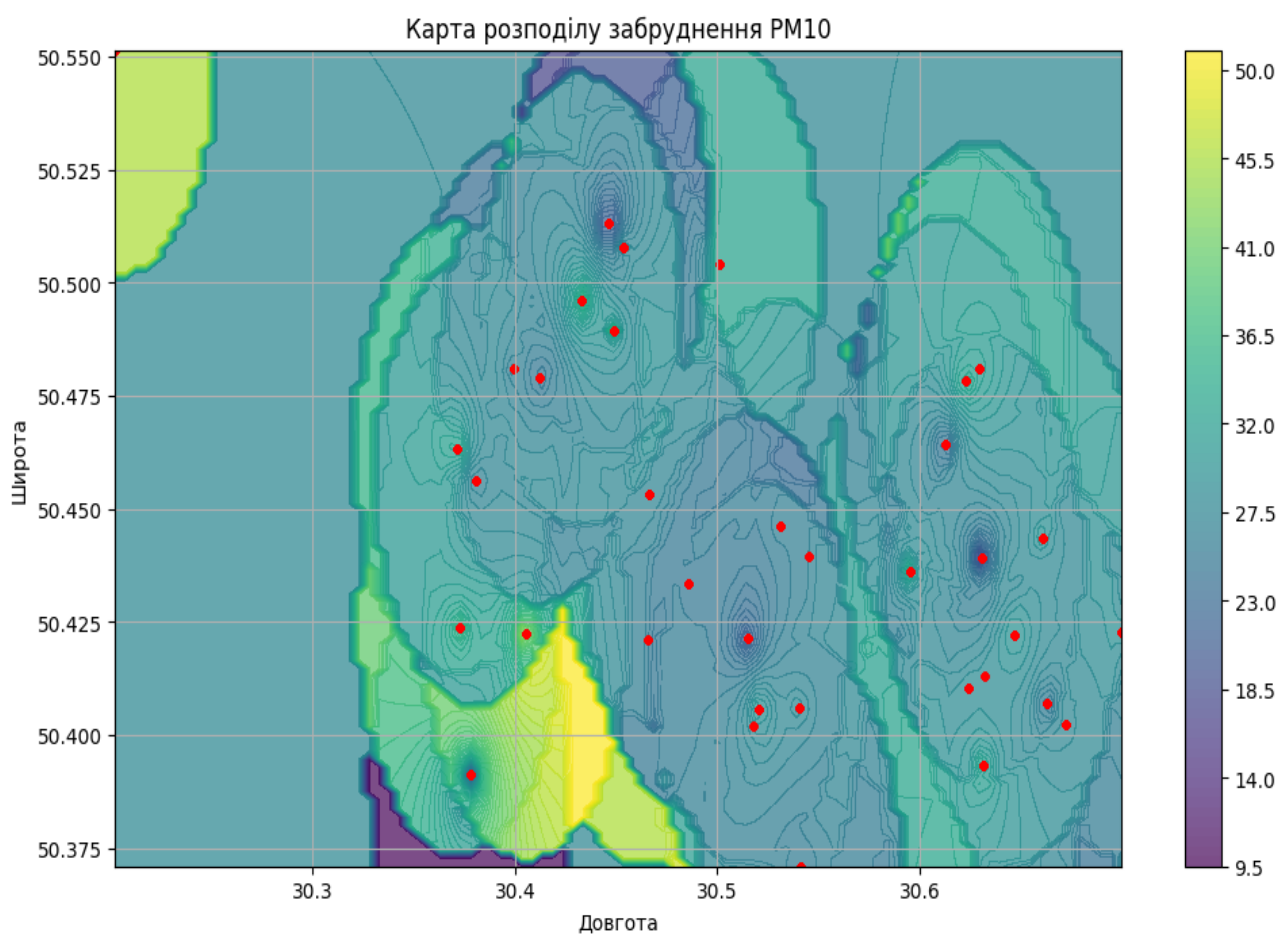


Рис. 3.3. Географічний розподіл тренувальних даних за березень 2022 року зі значеннями показників якості повітря

-жовтень 2022 роки. Згідно з інформацією [11] 10 жовтня 2022 року російські війська завдали першого масованого ракетного удару по енергетичній інфраструктурі України, а також наймасовішого від початку повномасштабного вторгнення ракетного удару по всій території України. Росіяни застосували крилаті ракети повітряного, морського та наземного базування, балістичні ракети, зенітні керовані ракети, розвідувальні та ударні БПЛА типу «Shahed-136». Внаслідок масованого ракетного удару постраждала інфраструктура понад 40 населених пунктів у Київській, Львівській, Вінницькій, Дніпропетровській, Запорізькій та Донецькій областях, зокрема, це Київ, Львів,

Попівка, Чугунівка, Ладизин, Кривий Ріг, Запоріжжя, Павлоград та Нікополь. MSE (0.1521819327442096), MAE (0.35252272071026713). На Рис. 3.4 зображено карту забруднень.

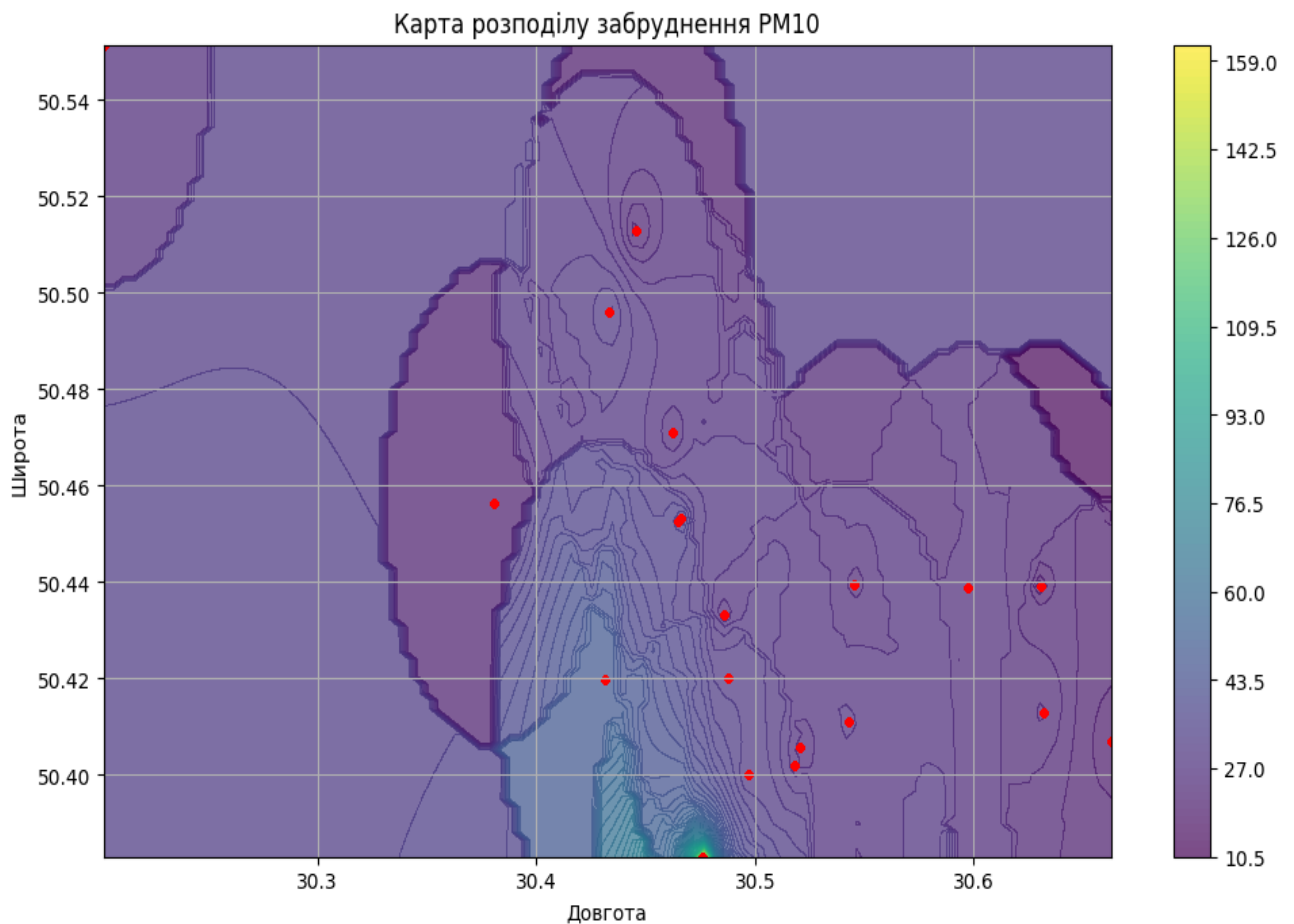


Рис. 3.4. Географічний розподіл тренувальних даних за жовтень 2022 року зі значеннями показників якості повітря

На основі аналізу трьох карт розподілу забруднення PM10 можна зробити кілька висновків про те, як відрізняються умови забруднення в різні місяці. По-перше, відрізняється загальний розподіл забруднення. Карта за січень показує порівняно високі рівні забруднення у північній частині карти, зі зниженням концентрацій у напрямку до південної частини. Карта за березень демонструє дещо більш рівномірний розподіл з високими концентраціями у центральних та південних районах. Карта за жовтень відрізняється найвищими концентраціями у південно-західній частині, з меншими концентраціями у північних районах. Зокрема, на третій карті спостерігаються значно вищі пікові концентрації (до 159.0), порівняно з першою та другою картами (максимуми до 46.5 та 50.0

відповідно). Що показує вплив воєнних дій на екологію, а саме на стан повітря. З обстрілами повітря стало гіршим в 3 рази.

Для оцінки точності та ефективності прогнозування даних про якість повітря за допомогою методу Griddata було використано три методи інтерполяції: метод найближчого сусіда (nearest), лінійна інтерполяція (linear) та кубічна інтерполяція (cubic).

На Рис 3.5 зображено фактичні та прогнозовані значення рівня забруднення PM10 для тестових даних.

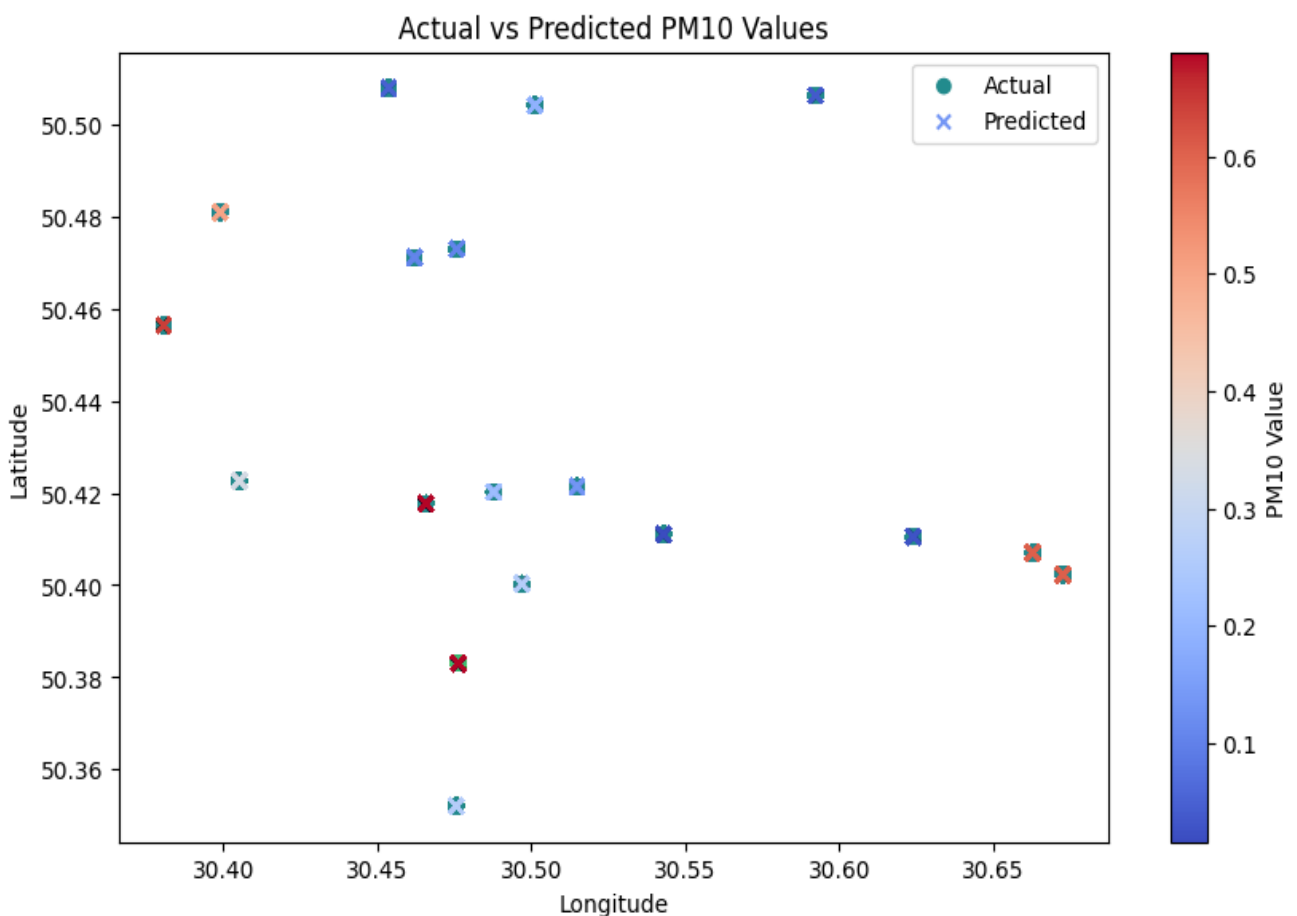


Рис 3.5. Порівняння отриманих та передбачуваних результатів методом Griddata

Фактичні значення позначені зеленими точками, а прогнозовані значення - синіми хрестиками. Колірна шкала праворуч відображає значення PM10. Цей графік допомагає візуально оцінити, наскільки добре метод інтерполяції прогнозує значення PM10 у точках тестової вибірки. Збіг або близькість хрестиків та точок вказує на високу точність моделі.

Були отримані такі показники точності:

- Метод інтерполяції: nearest. MSE моделі інтерполяції: 0.3236411970394231.  
MAE моделі інтерполяції: 0.251579828841705.

Візуалізація методу найближчого сусіда зображена на Рис 3.6.

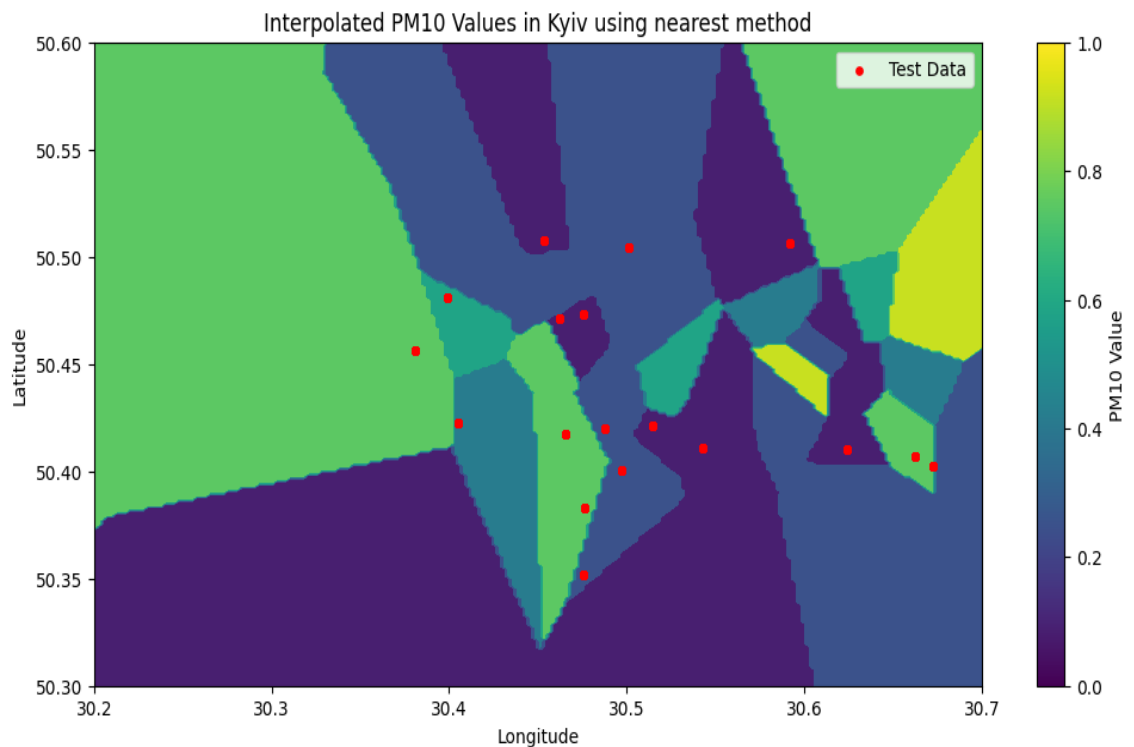


Рис 3.6. Візуалізація значень показників якості повітря за допомогою методу nearest

Метод демонструє чіткі області, де кожна зона представляє ділянку, найближчу до певної точки з відомими даними. Це може призводити до різких змін значень між сусідніми зонами, що не завжди відображає реальний розподіл забруднення повітря. Метрики свідчать про відносно добру точність прогнозів, хоча і з певними похибками.

- Метод інтерполяції: linear. MSE моделі інтерполяції: 0.3509634398584189.  
MAE моделі інтерполяції: 0.2929835346395786.

Візуалізація методу найближчого сусіда зображена на Рис 3.7.

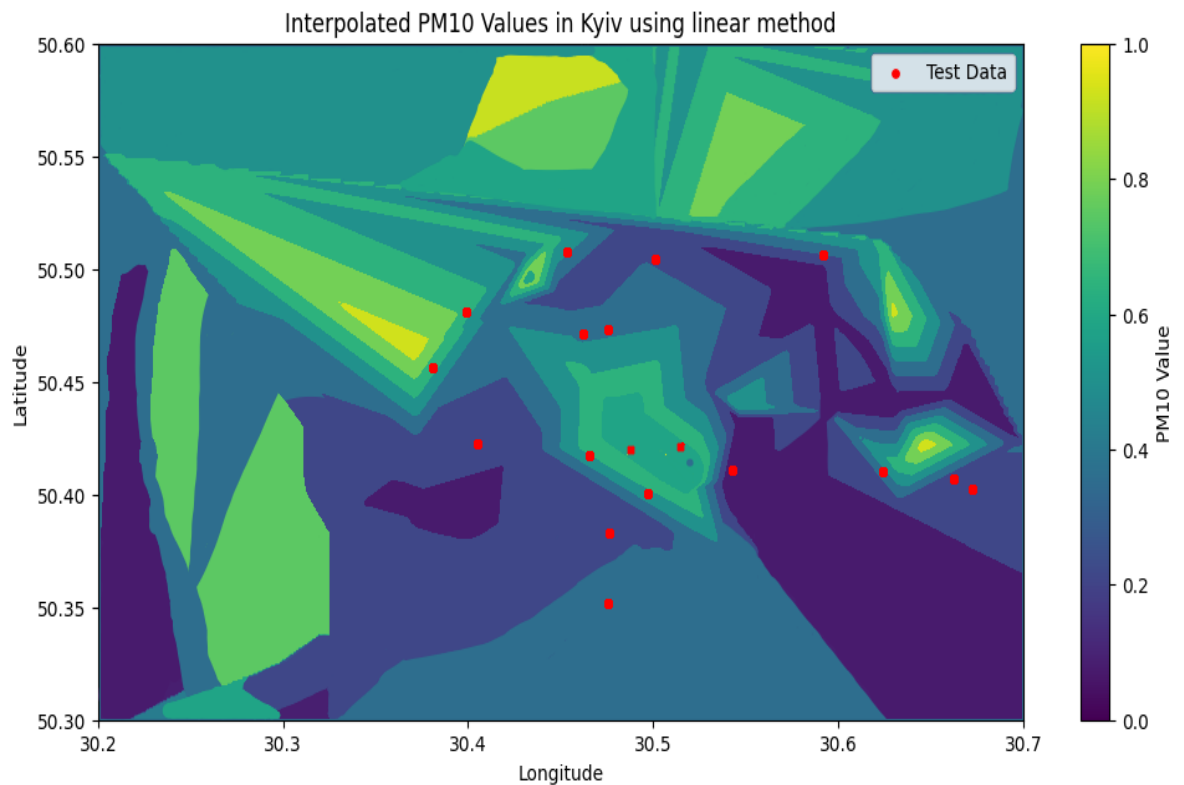


Рис 3.7. Візуалізація значень показників якості повітря за допомогою методу linear

Лінійна інтерполяція надає більш плавні переходи між точками, що дозволяє краще відобразити зміни рівня забруднення повітря на території. Проте метрики точності свідчать про дещо гіршу точність порівняно з методом найближчого сусіда.

- Метод інтерполяції: cubic. MSE моделі інтерполяції: 0.4009419846592176.  
MAE моделі інтерполяції: 0.37853437029829127

Візуалізація методу найближчого сусіда зображена на Рис 3.8.

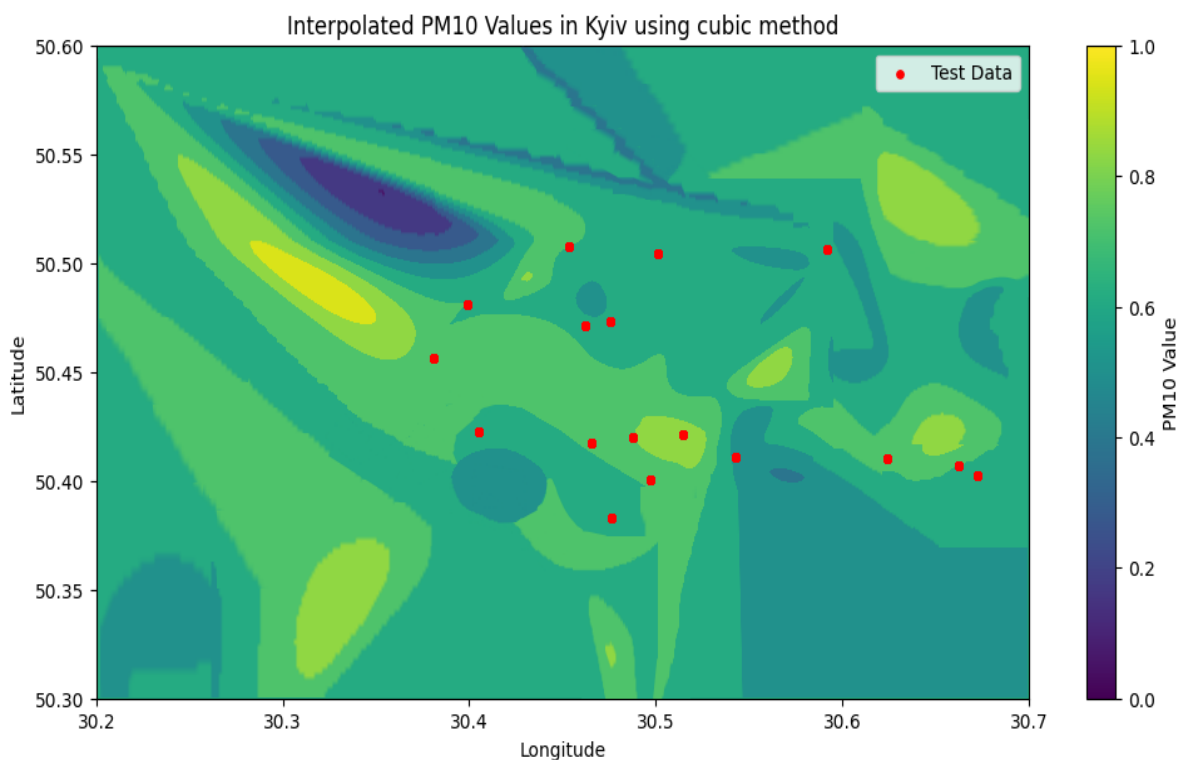


Рис 3.8. Візуалізація значень показників якості повітря за допомогою методу cubic

Вибір методу інтерполяції залежить від характеру даних та специфіки досліджуваного регіону. Подальше вдосконалення моделі може включати комбінування різних методів або використання більш складних алгоритмів машинного навчання для покращення точності прогнозування.

Кубічна інтерполяція забезпечує найгладший перехід між точками, що може бути корисним для регіонів з плавними змінами рівня забруднення. Метод найближчого сусіда підходить для регіонів з чітко вираженими зонами забруднення, де значення змінюються різко між зонами. Лінійна інтерполяція це кращий вибір для областей з поступовими змінами рівня забруднення. Плавніші переходи між точками, хоча точність дещо гірша.

### Висновки до розділу 3

У розділі були розглянуті методи екстраполяції даних для оцінки якості повітря в міському середовищі з обмеженою кількістю моніторингових станцій, зокрема методи IDW та Griddata. Проведено аналіз точності кожного з

методів, спираючись на переваги та обмеження. Метод IDW для екстраполяції рівнів забруднення PM10, враховуючи результати оцінки точності, продемонстрував достатньо точні результати. Висновок про ефективність моделі базується на вивченні показників MSE та MAE, які отримані після стратифікованої крос-валідації. Також було отримано суцільну карту забруднень, яка демонструє більш наочно результати. Можливими проблемами методу є виникнення неточностей у випадках наявності природних бар'єрів, які впливають на розподіл забруднюючих речовин і погодних умов.

Під час реалізації методу Griddata реалізовано три види інтерполяції (лінійна, кубічна та метод найближчого сусіда), що дозволило обрати найоптимальніший підхід для конкретного набору даних. Інтерполяція методом найближчого сусіда продемонструвала найбільшу точність для нелінійних змін даних, що підтверджено низькими значеннями MSE та MAE. Лінійна інтерполяція виявилась найпростішою у реалізації та забезпечила достатню точність для лінійних змін. Кубічна інтерполяція мала найнижчу точність для плавних змін, але виявився ефективним для швидких обчислень. Отже, метод Griddata показав себе як ефективний інструмент для просторового прогнозування, забезпечуючи точні результати навіть у випадках з нерегулярним розподілом даних, проте модель IDW дала кращі результати показників точності.

Аналіз результатів моделі IDW на прикладі трьох часових періодів (січень, березень, жовтень 2022 року) показав значний вплив воєнних дій на забруднення повітря. Карти забруднення показали, що в жовтні концентрація забруднюючих речовин зросла в три рази порівняно з початком року, що вказує на погіршення екологічної ситуації внаслідок обстрілів.

## Висновки

У даній дипломній роботі було досліджено методи екстраполяції якості повітря в умовах обмеженої кількості моніторингових станцій. Було проведено аналіз світового досвіду, розглянуто методи розв'язання проблеми обмеженої кількості станцій моніторингу, застосовані в різних країнах, і визначено можливість їх адаптації до умов України.

Після теоретичних досліджень була проведена розробка та обробка датасету. Зібрано та підготовлено дані про показники якості повітря міста Київ за 2022 рік з платформи SaveEcoBot. Проведено очищення даних та видалення аномалій, що дозволило забезпечити їхню надійність для подальшого аналізу.

На отриманому набору даних була виконана реалізація методів екстраполяції. Розроблено та реалізовано моделі екстраполяції даних, включаючи методи IDW та Griddata. Проведено порівняльний аналіз їхньої точності та визначено оптимальні параметри для кожної моделі. Зроблено оцінку точності створених моделей за допомогою метрик MSE та MAE на реальних даних. В результаті отримано інформацію, що методи IDW та метод найближчого сусіда Griddata показали найкращі результати.

Досліджено вплив воєнних дій на екологію та якість повітря. Аналіз результатів за три різні часові періоди показав значний вплив воєнних дій на повітря в Києві. Виявлено, що концентрація забруднюючих речовин значно зросла після початку обстрілів. Результати роботи можуть бути використані для покращення системи моніторингу якості повітря в Україні, забезпечуючи точні дані про стан повітря навіть за умов обмеженої кількості станцій. Це дозволить органам влади та екологічним організаціям швидше реагувати на зміни якості повітря та приймати обґрунтовані рішення для покращення екологічної ситуації.

## Перелік використаних джерел

1. Urban air quality. / Jes Fenger [та ін.] // Atmospheric Environment. – 1999. – Volume 33:29, pp. 4877-4900, ISSN 1352-2310, [https://doi.org/10.1016/S1352-2310\(99\)00290-3](https://doi.org/10.1016/S1352-2310(99)00290-3).
2. Human health effects of air pollution. / Kampa M., Castanas E.. // Environmental Pollution. – 2008. - pp. 362-367.
3. Air Quality as Proxy for Assesment of Economic Activity. / Hanna Yailymova, Andrii Kolotii, Nataliia Kussul, Andrii Shelestov. // Proceedings of IEEE EUROCON 2023, – July 6-8, 2023. - pp. 89-92. Doi: 10.1109/EUROCON56442.2023.10198882.
4. Earth observation: An integral part of a smart and sustainable city. / Evangelos Gerasopoulos, Jennifer Bailey, Eleni Athanasopoulou, Orestis Speyer, David Kocman, Astrid Raudner, Alexia Tsouni, Haris Kontoes, Christer Johansson, Charalampos Georgiadis, Volker Matthias, Nataliia Kussul, Mariella Aquilino, Pauli Paasonen. // Environmental Science & Policy, 132, pp. 296-307. DOI: 10.1016/j.envsci.2022.02.033.
5. A Survey of Wireless Sensor Network Based Air Pollution Monitoring Systems.  
W. Y. Yi, K. M. Lo, T. Mak, K. S. Leung, Y. Leung, M. L. Meng // Sensors. – 2015. - pp. 280-284.
6. Deep neural networks for ultra-short-term wind forecasting. / Dalto M., Matusko J., Vasak M.. // Proceedings of the IEEE International Conference on Industrial Technology. – 2015. pp. 91-106, DOI: 10.1109/ICIT.2015.7125335.
7. An efficient hybrid weather prediction model based on deep learning. / Utku A., Can U. // International Journal of Environmental Science and Technology. 2023; 20(3), DOI: 10.1007/s13762-023-05092-4.
8. Модель оцінки якості повітря за супутниковими даними на основі методу групового урахування аргументів. / В.В. Хайдуров, Б.Я. Яйлимов, А.Ю. Шелестов. // Проблеми керування та інформатики. – 2023. - 68(5), с. 93-106. doi: 10.34229/1028-0979-2023-5-8.

9. Mathematical Model for Determining the Geometric Location of the Environmental Pollutant Based on Sensor Data. / Vladyslav Khaidurov, Bohdan Yailymov, Andrii Shelestov. // Proceedings of the 12th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS2023). – 7-9 September, 2023. - pp. 703-707. Doi: 10.1109/IDAACS58523.2023.10348700.
10. SaveEcoBot. — URL: <https://www.saveecobot.com/>.
11. Хронологія російського вторгнення в Україну (жовтень 2022) — (2023, вересня 8) / Вікіпедія. // URL: [https://uk.wikipedia.org/w/index.php?title=%D0%A5%D1%80%D0%BE%D0%BD%D0%BE%D0%BB%D0%BE%D0%B3%D1%96%D1%8F\\_%D1%80%D0%BE%D1%81%D1%96%D0%B9%D1%81%D1%8C%D0%BA%D0%BE%D0%B3%D0%BE\\_%D0%B2%D1%82%D0%BE%D1%80%D0%B3%D0%BD%D0%B5%D0%BD%D0%BD%D1%8F\\_%D0%B2\\_%D0%A3%D0%BA%D1%80%D0%B0%D1%97%D0%BD%D1%83\\_\(%D0%B6%D0%BE%D0%B2%D1%82%D0%B5%D0%BD%D1%8C\\_2022\)&oldid=40360](https://uk.wikipedia.org/w/index.php?title=%D0%A5%D1%80%D0%BE%D0%BD%D0%BE%D0%BB%D0%BE%D0%B3%D1%96%D1%8F_%D1%80%D0%BE%D1%81%D1%96%D0%B9%D1%81%D1%8C%D0%BA%D0%BE%D0%B3%D0%BE_%D0%B2%D1%82%D0%BE%D1%80%D0%B3%D0%BD%D0%B5%D0%BD%D0%BD%D1%8F_%D0%B2_%D0%A3%D0%BA%D1%80%D0%B0%D1%97%D0%BD%D1%83_(%D0%B6%D0%BE%D0%B2%D1%82%D0%B5%D0%BD%D1%8C_2022)&oldid=40360)

## ДОДАТКИ

### Додаток А. Перші рядки створеного датасету data.csv

```
device_id,phenomenon,value,logged_at,latitude,longitude
1376,pm10,44.53,2022-01-01 00:01:20,50.42138399294448,30.514839882005163
1376,pm25,21.23,2022-01-01 00:01:20,50.42138399294448,30.514839882005163
1376,temperature,6.73,2022-01-01
00:01:20,50.42138399294448,30.514839882005163
1376,humidity,87.07,2022-01-01 00:01:20,50.42138399294448,30.514839882005163
1376,pressure_pa,996950.0,2022-01-01
00:01:20,50.42138399294448,30.514839882005163
1376,heca_temperature,14.12,2022-01-01
00:01:20,50.42138399294448,30.514839882005163
1376,heca_humidity,59.56,2022-01-01
00:01:20,50.42138399294448,30.514839882005163
1376,min_micro,85.0,2022-01-01 00:01:20,50.42138399294448,30.514839882005163
1376,max_micro,21794.0,2022-01-01
00:01:20,50.42138399294448,30.514839882005163
1376,signal,-68.0,2022-01-01 00:01:20,50.42138399294448,30.514839882005163
1376,pm10,47.33,2022-01-01 00:03:49,50.42138399294448,30.514839882005163
1376,pm25,21.4,2022-01-01 00:03:49,50.42138399294448,30.514839882005163
1376,temperature,6.71,2022-01-01
00:03:49,50.42138399294448,30.514839882005163
1376,humidity,87.06,2022-01-01 00:03:49,50.42138399294448,30.514839882005163
1376,pressure_pa,996875.4,2022-01-01
00:03:49,50.42138399294448,30.514839882005163
1376,heca_temperature,13.6,2022-01-01
00:03:49,50.42138399294448,30.514839882005163
1376,heca_humidity,59.21,2022-01-01
00:03:49,50.42138399294448,30.514839882005163
1376,min_micro,85.0,2022-01-01 00:03:49,50.42138399294448,30.514839882005163
1376,max_micro,21086.0,2022-01-01
00:03:49,50.42138399294448,30.514839882005163
1376,signal,-67.0,2022-01-01 00:03:49,50.42138399294448,30.514839882005163
1376,pm10,46.92,2022-01-01 00:06:16,50.42138399294448,30.514839882005163
1376,pm25,22.5,2022-01-01 00:06:16,50.42138399294448,30.514839882005163
1376,temperature,6.71,2022-01-01
00:06:16,50.42138399294448,30.514839882005163
1376,humidity,87.08,2022-01-01 00:06:16,50.42138399294448,30.514839882005163
1376,pressure_pa,996848.1,2022-01-01
00:06:16,50.42138399294448,30.514839882005163
1376,heca_temperature,12.22,2022-01-01
```

00:06:16,50.42138399294448,30.514839882005163  
1376,heca\_humidity,62.85,2022-01-01  
00:06:16,50.42138399294448,30.514839882005163  
1376,min\_micro,84.0,2022-01-01 00:06:16,50.42138399294448,30.514839882005163  
1376,max\_micro,21098.0,2022-01-01  
00:06:16,50.42138399294448,30.514839882005163  
1376,signal,-67.0,2022-01-01 00:06:16,50.42138399294448,30.514839882005163

## Додаток Б. Програмний код обробки даних

```
# Функції для передпроцесингу даних
def clean_data(data):
    data['latitude'] = pd.to_numeric(data['latitude'], errors='coerce')
    data['longitude'] = pd.to_numeric(data['longitude'], errors='coerce')
    data['value'] = pd.to_numeric(data['value'], errors='coerce')
    data.dropna(subset=['latitude', 'longitude', 'value'], inplace=True)
    return data

def filter_data(data):
    data = data[(data['latitude'] >= 50.30) & (data['latitude'] <= 50.60)]
    data = data[(data['longitude'] >= 30.20) & (data['longitude'] <=
30.70)]
    data = data[data['phenomenon'] == 'pm10']
    return data

def remove_outliers(data):
    Q1 = data['value'].quantile(0.25)
    Q3 = data['value'].quantile(0.75)
    IQR = Q3 - Q1
    return data[~((data['value'] < (Q1 - 1.5 * IQR)) | (data['value'] > (Q3
+ 1.5 * IQR)))]

def scale_data(data):
    scaler = MinMaxScaler()
    data.loc[:, 'value'] = scaler.fit_transform(data[['value']])
    return data

# Завантаження та підготовка даних
data_path = 'data.csv'
data = pd.read_csv(data_path, low_memory=False)
data['logged_at'] = pd.to_datetime(data['logged_at'], errors='coerce')
data = data[(data['logged_at'] >= '2022-10-01') & (data['logged_at'] <=
'2022-10-31')]
data = clean_data(data)
data = filter_data(data)
data['value_bin'] = pd.qcut(data['value'], 5, labels=False,
duplicates='drop')

# Розділення даних за унікальними ідентифікаторами пристроїв
unique_devices = data['device_id'].unique()
train_devices, test_devices = train_test_split(unique_devices,
test_size=0.3, random_state=42)
train_data = data[data['device_id'].isin(train_devices)]
test_data = data[data['device_id'].isin(test_devices)]
```

## Додаток В. Програмний код реалізації моделі IDW

```
# Інтерполяція за методом IDW
def spatial_smoothing_IDW(train_coords, train_values, predict_coords,
power, smoothing_radius):
    interpolated_values = []
    for p_coord in predict_coords:
        distances = np.sqrt(np.sum((train_coords - p_coord) ** 2, axis=1))
        weights = np.where(distances < smoothing_radius, 1 /
np.maximum(distances, 1e-10)**power, 0)
        if np.sum(weights) == 0:
            weights = 1 / distances**power
        weighted_values = weights * train_values
        interpolated_value = weighted_values.sum() / weights.sum()
        interpolated_values.append(interpolated_value)
    return np.array(interpolated_values)

# Проведення стратифікованої крос-валідації
def perform_cross_validation(train_data, n_splits=5, power=1,
smoothing_radius=0.05):
    skf = StratifiedKFold(n_splits=n_splits)
    mse_scores = []
    mae_scores = []
    r2_scores = []
    for train_index, test_index in skf.split(train_data,
train_data['value_bin']):
        fold_train_data = train_data.iloc[train_index]
        fold_test_data = train_data.iloc[test_index]
        fold_train_data = remove_outliers(fold_train_data)
        fold_train_data = scale_data(fold_train_data)
        fold_test_data = scale_data(fold_test_data)
        train_coords = fold_train_data[['latitude', 'longitude']].values
        train_values = fold_train_data['value'].values
        test_coords = fold_test_data[['latitude', 'longitude']].values
        test_values = fold_test_data['value'].values

        predicted_values = spatial_smoothing_IDW(train_coords,
train_values, test_coords, power, smoothing_radius)
        mse = mean_squared_error(test_values, predicted_values)
        mae = mean_absolute_error(test_values, predicted_values)
        mse_scores.append(mse)
        mae_scores.append(mae)
    return np.mean(mse_scores), np.mean(mae_scores)

# Проведення стратифікованої крос-валідації на тренувальних даних
average_mse, average_mae= perform_cross_validation(train_data)
print("Середньоквадратична помилка (MSE):", average_mse)
print("Середня абсолютна помилка (MAE):", average_mae)
```

## Додаток Д. Програмний код реалізації моделі Griddata

```
# Створюємо сітку координат
grid_x, grid_y = np.mgrid[50.30:50.60:200j, 30.20:30.70:200j]

# Список методів інтерполяції
methods = ['nearest', 'linear', 'cubic']

# Основний код
for method in methods:
    # Інтерполяція за допомогою методу
    predicted_values = Griddata(
        train_data[['latitude', 'longitude']].values,
        train_data['value'].values,
        (grid_x, grid_y),
        method=method,
        fill_value=np.nanmean(train_data['value'].values) # Додаємо
екстраполяцію
    )

    # Прогнозування значень для тестової вибірки
    predicted_test_values = Griddata(
        train_data[['latitude', 'longitude']].values,
        train_data['value'].values,
        test_data[['latitude', 'longitude']].values,
        method=method,
        fill_value=np.nanmean(train_data['value'].values) # Додаємо
екстраполяцію
    )

    # Обчислення MSE та MAE
    mse = mean_squared_error(test_data['value'].values,
predicted_test_values)
    mae = mean_absolute_error(test_data['value'].values,
predicted_test_values)

    # Виведення результатів
    print(f"Метод інтерполяції: {method}")
    print(f"Середньоквадратична помилка (MSE) моделі інтерполяції: {mse}")
    print(f"Середня абсолютна помилка (MAE) моделі інтерполяції: {mae}")

    # Візуалізація результатів інтерполяції
    plt.figure(figsize=(12, 6))
    plt.contourf(grid_y, grid_x, predicted_values, cmap='viridis') #
Створюємо суцільну карту
    plt.scatter(test_data['longitude'], test_data['latitude'], c='red',
s=15, label='Test Data')
    plt.colorbar(label='PM10 Value')
    plt.xlabel('Longitude')
```

```
plt.ylabel('Latitude')
plt.title(f'Interpolated PM10 Values in Kyiv using {method} method')
plt.legend()
plt.show()
```