

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ
Кафедра інформаційної безпеки

«До захисту допущено»
В.о. завідувача кафедри

_____ М.В.Грайворонський
(підпис)

“ ” _____ 2019 р.

Дипломна робота
на здобуття ступеня бакалавра

з напрямку підготовки 6.040301 «Прикладна математика»

на тему: Регресійна модель концентрацій забруднювачів повітря для України на основі
точкових наземних та супутникових даних

Виконав: студент 4 курсу, групи ФІ-51
(шифр групи)

Козловський Євгеній Петрович
(прізвище, ім'я, по батькові) (підпис)

Керівник доцент, к.т.н. Лавренюк А. М.
(посада, науковий ступінь, вчене звання, прізвище та ініціали) (підпис)

Консультант _____
(назва розділу) (посада, вчене звання, науковий ступінь, прізвище, ініціали) (підпис)

Рецензент _____
(посада, науковий ступінь, вчене звання, науковий ступінь, прізвище та ініціали) (підпис)

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

Студент _____
(підпис)

Київ - 2019 року

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ
Кафедра інформаційної безпеки

Рівень вищої освіти – перший (бакалаврський)

Напрямок підготовки 6.040301 «Прикладна математика»

ЗАТВЕРДЖУЮ

В.о. завідувача кафедри

_____ М.В.Грайворонський
(підпис)

«__» _____ 2019 р.

ЗАВДАННЯ
на дипломну роботу студенту

Козловському Євгенію Петровичу

(прізвище, ім'я, по батькові)

1. Тема роботи: Регресійна модель концентрацій забруднювачів повітря для України на основі точкових наземних та супутникових даних, науковий керівник роботи: к.т.н., доцент Лавренюк А. М.,
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом по університету від «27 травня» 2019 р. № 1414-с

2. Термін подання студентом роботи 10 червня 2019 р.

3. Вихідні дані до роботи: Дані з відкритих джерел та наземні виміри концентрацій забруднювачів повітря, регресійний аналіз, метод найменших квадратів

4. Зміст роботи: Ознайомитись з навчальною літературою для програмного забезпечення, ознайомитись з математичною літературою, виміряти концентрації забруднювачів повітря, знайти дані про концентрації забруднювачів повітря у відкритому доступі, побудувати регресійні моделі

5. Перелік ілюстративного матеріалу (із зазначенням плакатів, презентацій тощо) _____

6. Дата видачі завдання: 5 жовтня 2018 р.

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів дипломної роботи	Примітка
1	Ознайомлення з навчальною літературою для програмного забезпечення	05.10.18-19.05.19	
2	Ознайомлення з математичною літературою	05.10.18-19.05.19	
3	Пошук геоданих	12.02.19-05.05.19	
4	Аналіз отриманих наборів даних	12.02.19-12.05.19	
5	Розробка регресійних моделей	17.04.19-19.05.19	

Студент

(підпис)

Є. П. Козловський
(ініціали, прізвище)

Керівник роботи

(підпис)

А. М. Лавренюк
(ініціали, прізвище)

РЕФЕРАТ

Дипломна робота містить: 52 сторінки, 17 рисунків, 2 таблиці, 14 джерел літератури.

Мета роботи: побудова регресійних моделей концентрацій забруднювачів повітря для України на основі точкових наземних вимірів та даних з CAMS.

Об'єктом дослідження є процеси забруднення повітря.

Предметом дослідження є моделі для отримання концентрацій забруднювачів повітря, наближених до реально виміряних.

Було побудовано та досліджено різні регресійні моделі: поліноміальні, лінійні, ізотонічні, для різних забруднювачів повітря: PM_{10} , $PM_{2.5}$, CO, O₃, NO₂. Якість цих моделей було порівняно за коефіцієнтом детермінації R^2 . Найкращі результати показали саме поліноміальні моделі, а лінійні - доволі низькі, це свідчить про нелінійний характер зв'язку між даними. Ізотонічні моделі показали результати, кращі ніж лінійні, але гірші ніж поліноміальні.

Ключові слова: прогнозування, регресійний аналіз, забруднювачі повітря, супутникові дані, наземні вимірювання, лінійна регресія, поліноміальна регресія.

ABSTRACT

Thesis contains: 52 pages, 17 figures, 2 tables, 14 sources of literature.

Purpose: construction of regression models of concentrations of air pollutants for Ukraine on the basis of point ground measurements and data from CAMS.

The object of the study is the processes of air pollution.

The subject of the study is a model for obtaining concentrations of air pollutants close to those actually measured.

Different regression models were constructed and investigated: polynomial, linear, isotonic, for various air pollutants: PM_{10} , $PM_{2.5}$, CO, O₃, NO₂. The quality of these models was compared with the determination coefficient R^2 . The best results have shown polynomial models, and the linear ones are rather low, indicating the nonlinear nature of the connection between the data. Isotonic models showed results better than linear, but worse than polynomial ones.

Keywords: forecasting, regression analysis, air pollutants, satellite data, ground measurements, linear regression, polynomial regression.

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів.....	7
Вступ.....	8
1 Сучасний стан оцінювання та прогнозування забруднення повітря в Україні та світі.....	10
1.1 Огляд літературних джерел з теми дослідження.....	10
1.2 Повітря – його якісні властивості та вплив на здоров'я людини.....	16
1.3 Міжнародні джерела даних.....	20
Висновки до розділу 1	24
2 Прогнозування рівня забруднення повітря за допомогою регресійного моделювання.....	25
2.1 Регресійний аналіз.....	25
2.2 Програмні інструменти та формати даних.....	31
Висновки до розділу 2	37
3 Побудова регресійної моделі на основі точкових наземних та супутникових вимірів.....	38
Висновки до розділу 3	49
Висновки.....	50
Перелік джерел посилань.....	51

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

ECMWF – Європейський центр прогнозу погоди

CAMS – The Copernicus Atmosphere Monitoring Service

ООП – об’єктно-орієнтоване програмування

ІКД – Інститут космічних досліджень

USEPA – Агенство охорони навколишнього середовища США

MLR – множинна лінійна регресія

ANN – штучна нейронна мережа

PM₁₀ – Зважені частинки розміром менше 10 мкм

PM_{2.5} – Зважені частинки розміром менше 2.5 мкм

CO – Монооксид вуглецю

O₃ – Озон

NO₂ – Оксид азоту

ГІС – геоінформаційна система

QGIS – quantum geographic information system

OSGeo – The Open Source Geospatial Foundation

µg/m³ – мікрограм поділити на метр в кубі

ppm – parts-per-million

ДЗЗ – дистанційне зондування землі

ЕЕА – Європейське агентство з охорони навколишнього середовища

R² – коефіцієнт детермінації

ВСТУП

Повітря, з яким ми постійно контактуємо, суттєво впливає на організм людини та на все навколишнє середовище. Воно є складною сумішшю газів, серед яких є ті, що спричиняють більше шкоди ніж користі, тому за якістю повітря потрібно слідкувати.

На сьогоднішній день існує два методи визначення концентрацій забруднювачів повітря: вимірювання наземними приладами та обробка даних з супутників, кожен із методів має свої переваги і недоліки. Тому використання наземних і супутникових даних має відбуватись одночасно, проте для оцінки та прогнозування якості нижніх шарів атмосфери актуальним є застосування регресійного моделювання для отримання концентрацій забруднювачів повітря, наближених до реально вимірних. Застосування математичного моделювання та заснованих на ньому програмних пакетів, дозволяє вирішувати завдання розрахунку забруднення атмосферного повітря населених пунктів.

Хоча вже були деякі дослідження на основі регресійного аналізу концентрацій забруднювачів повітря з різними предикторами, в цій роботі використано в якості предиктору дані з CAMS для території України, щоб побудувати модель яка буде надавати дані про концентрації забруднювачів повітря наближені до реально вимірних на рівні землі.

Метою даної роботи є побудова регресійних моделей концентрацій забруднювачів повітря для України на основі точкових наземних вимірів та даних з CAMS.

Досягнення головної мети вимагає виконання наступних завдань:

- вивчити опубліковані джерела з питань моделювання концентрацій забруднювачів повітря на основі точкових наземних та супутникових даних;
- проаналізувати існуючі методи дослідження якості повітря;
- побудувати регресійні моделі концентрацій забруднювачів повітря для України на основі точкових наземних вимірів та даних з CAMS;

Об'єктом дослідження є процеси забруднення повітря.

Предметом дослідження є моделі для отримання концентрацій забруднювачів повітря, наближених до реально вимірних.

Теоретичною та методологічною основою дослідження є праці вітчизняних і зарубіжних вчених в галузях математичного моделювання, забруднення навколишнього середовища.

При розв'язанні поставлених у роботі завдань було використано регресійний аналіз. Математичні та програмні розрахунки виконані на персональному комп'ютері з використанням мови програмування Python та програми Qgis, для роботи з метеорологічними та геоданими.

Вперше було побудовано різні регресійні моделі, а саме: поліноміальні, лінійні, ізотонічні, концентрацій PM_{10} , $PM_{2.5}$, CO, O₃, NO₂ для України на основі точкових наземних вимірів, отриманих за допомогою апаратів: Air quality monitor sdl607 та розробки Noosphere Engineering School, та даних з CAMS. Якість цих моделей було порівняно за коефіцієнтом детермінації R^2 . Виявилось, що найкращі результати показують саме поліноміальні моделі, а лінійні - доволі низькі, це свідчить про нелінійний характер зв'язку між даними. Ізотонічні моделі показали результати, кращі ніж лінійні, але гірші ніж поліноміальні.

Це дозволить отримувати концентрації забруднювачів повітря, наближені до реально вимірних, що можна використовувати у багатьох сферах, наприклад, для моніторингу якості повітря, інших дослідів, заснованих на характеристиках навколишнього середовища на певній місцевості.

1 СУЧАСНИЙ СТАН ОЦІНЮВАННЯ ТА ПРОГНОЗУВАННЯ ЗАБРУДНЕННЯ ПОВІТРЯ В УКРАЇНІ ТА СВІТІ

1.1 Огляд літературних джерел з теми дослідження

На сьогоднішній день все більшої популярності набуває моніторинг та оцінювання стану атмосферного повітря, зокрема, за допомогою обробки даних із метеорологічних супутників. Це надає певні переваги, такі як здійснення моніторингу стану повітря у будь-якому місці землі, низька вартість виконання, широкий територіальний охоплення.

Аналіз літературних джерел з теми показав велику кількість вже проведених досліджень стану повітря та прогнозування забруднення з використанням регресійних моделей, як в Україні так і за її межами.

Так, на прикладі території Столиці Еквадору - Кіто, був проведений регресійний аналіз для $PM_{2.5}$. Це дослідження [1] демонструє, що продуктивність прогнозування $PM_{2.5}$ залежить від доступної вхідної інформації. Перший висновок показує, що можна отримати розумний прогноз концентрацій $PM_{2.5}$ лише за допомогою публічних даних про трафік руху машин. Прогнозування забруднення навколишнього середовища $PM_{2.5}$ на основі трафіку можна значно покращити, використовуючи три моделі на день замість однієї, особливо для ранкових годин. Протягом ранкової години пік фіксують кумулятивне зростання концентрацій $PM_{2.5}$. Використання доступної метеорологічної станції даних додатково покращує точність прогнозування. У цьому випадку дерево регресійної моделі дає кращий прогноз, ніж модель лінійної регресії. Як і очікувалося, найкраща модель отримана шляхом включення гібридних джерел даних таких як час, трафік, та концентрації атмосферних забруднювачів.

$$PM_{2.5} = 0.2032 * RH + 0.6507 * temperature - 0.0021 * SR + 0.454 * \\ * Xwind + 0.225 * NO_2 + 0.2159 * O_3 + 1.0707 * SO_2 + 8.8163 * CO - \\ -23.9476 \quad (1.1)$$

Ці вісім вибраних ознак складаються з критеріїв забруднюючих речовин (CO, NO₂, O₃, SO₂) і метеорологічних факторів (вологість, сонячна радіація, температура, швидкість вітру і напрямок).

Запропонований підхід легко узагальнюється для інших міст світу. Зберігання та регресійний аналіз двомісячних даних був достатнім для побудови моделей, які здатні прогнозувати дрібнодисперсний матеріал з високою точністю. Подальша робота буде зосереджена на поліпшенні продуктивності моделі у вечірні години пік. Очікується, що більш досконалі моделі будуть отримані шляхом включення додаткових спостережень і функцій у набір даних. Наприклад, деякі додаткові дослідження передбачають вивчення впливу глибини планетного прикордонного шару на розрідження забруднення PM_{2.5}.

Крім того, є мотивація дослідити поточну продуктивність моделі з даними, отриманими обладнанням гіршої якості. У цьому дослідженні забруднення повітря та метеорологія були виміряні за допомогою обладнання, затвердженого USEPA, недоступного для великої частини міст у країнах, що розвиваються, що обмежує дослідження забруднення повітря у багатьох містах. Було показано, що малі міста часто є більш забрудненими, ніж великі агломерації.

Також вже був проведений регресійний аналіз для O₃, у місті Афіни, Греція [2-3]. Так у роботі [2], лінійна модель була використана для регресії та прогнозування щоденних рівнів забруднення повітря для побудови оперативного модуля прогнозування концентрації озону, розробленого з нуля, для басейну Афін, Греція. Найбільш важливими факторами, що впливають на концентрацію озону, виявилися максимальна температура, максимальний рівень озону попереднього дня та швидкість вітру.

Загальний огляд цього дослідження підтверджує труднощі, що виникають у статистичному моделюванні атмосфери. Низька спроможність передбачення

моделі не дивує, оскільки стохастична, випадкова складова в регресії завжди існує і залишається не з'ясована. Враховуючи різноманітність міського ландшафту, особливості міського мікроклімату та складність визначення якості повітря, кількість даних фактичних вимірювань необхідних для створення далекосяжних і надійних передбачень призводить до проблеми, якою важко керувати. Модель статистичного лінійного регресійного аналізу стає в цьому сенсі привабливою, оскільки дає змогу швидко оцінити ситуацію, характерну для місця розташування.

Навіть у цьому випадку, враховуючи межі регресійного аналізу, використання площ та часу, що охоплюють дані моніторингу, залишається обов'язковою умовою. Тому можна лише підкреслити, що статистична методологія повинна керуватися розумінням основних фізичних механізмів.

Розроблена модель регресії першого порядку заснована на всіх предикторах, та була створена з набору даних 1990-1998 років. Коефіцієнт визначення моделі дорівнює 0,593, що свідчить про те, що всі включені змінні становлять 59,30% від початкової варіації змінної відповіді, тоді як інші 40,70% представляють варіацію залишків.

$$[O_3]=2.169 + 0.170 * \max \text{ temp} + 0.487 * \text{prev } [O_3] - 0.00618 * \text{ws} \quad (1.2)$$

де $[O_3]$ - концентрація озону, $\max \text{ temp}$ - максимальна добова температура, $\text{prev } [O_3]$ - максимальна концентрація озону попереднього дня, ws - середня швидкість вітру [2].

А у проведеному дослідженні [3] зроблено спробу прогнозувати щоденну максимальну концентрацію озону на рівні землі протягом наступних 24 годин у межах Афін. Для цього застосували моделі з множинною лінійною регресією (MLR) проти моделі прогнозування, заснованої на підході до штучної нейронної мережі (ANN). Наявність основних метеорологічних параметрів має велике значення для прогнозування концентрації озону. Моделювання базувалося на записаних метеорологічних даних та даних про забруднення повітря з

тринадцяти моніторингових пунктів у межах Афін (мережа Міністерства охорони навколишнього середовища, енергетики та клімату Греції) протягом п'яти років з 2001 по 2005 роки.

Побудована модель множинної лінійної регресії (MLR) має вигляд:

$$y = 1.4271 + 0.6562 * x_1 + 0.0101 * x_2 + 0.0076 * x_3 \quad (1.3)$$

де x_1 - натуральний логарифм максимальної добової концентрації озону попереднього дня, x_2 - максимальна добова температура повітря попереднього дня, x_3 - середньодобова швидкість вітру попереднього дня.

Проведений аналіз показав, що коефіцієнт детермінації (MLR: 0,653, ANN: 0,666) між спостережуваними та прогнозованими концентраціями озону за 2005 рік [3].

Уряд Китаю висунув ряд агресивних заходів контролю для вирішення екологічних проблем, таких як погана видимість, з першого року свого 11-го п'ятирічного плану (2006–2010). Нещодавно зареєстровані видимість, якість повітря та метеорологічні дані в чотирьох великих мегаполісах (Пекін, Шанхай, Гуанчжоу та Ченду) в різних регіонах (та кліматичних зонах) Китаю були проаналізовані з метою оцінки ступеня впливу цих дій на цю проблему. Концентрації навколишнього середовища трьох основних забруднювачів повітря (SO_2 , NO_2 та PM_{10}) у цих містах зменшилися в 2005–2009 роках. Однак поліпшення видимості спостерігалось лише в Пекіні та Гуанчжоу; вона залишалася стабільною в Шанхаї і мала тенденцію до зменшення в Ченду.

Результати підкреслюють той факт, що співвідношення між якістю повітря і видимістю є складним. Розроблено оптимальні емпіричні регресійні моделі, що базуються на вимірних даних про якість повітря та на метеорологічних параметрах, щоб краще ізолювати можливі причинні кореляції між видимістю та якістю повітря, а також метеорологічними умовами. Наші результати показують, що поліпшення видимості як в Пекіні, так і в Гуанчжоу було зумовлене зниженням концентрації PM_{10} . У Гуанчжоу поліпшення атмосферної видимості

також сприяло зниженню концентрації SO₂ взимку. Навпаки, менша швидкість вітру, разом з можливими змінами концентрації і складу дрібних частинок, могла б пояснити, чому в Шанхаї чи Ченду не було знайдено жодного поліпшення тенденції видимості.

Було отримано наступні моделі:

Для Пекіну:

If [RH] < 51%,

$$[\text{Vis}] = -4.75 \times \ln[\text{PM}_{10}] - 0.07 \times [\text{NO}_2] - 0.15 \times [\text{RH}] + 51.52 \quad (1.4)$$

Otherwise,

$$[\text{Vis}] = -7.51 \times \ln[\text{PM}_{10}] - 0.27 \times [\text{RH}] + 67.31 \quad (1.5)$$

Для Шанхаю:

If [WS] < 3 m/s,

$$[\text{Vis}] = -7.05 \times \ln[\text{PM}_{10}] - 0.18 \times [\text{RH}] + 0.05 \times [\text{Temp}] + 59.27 \quad (1.6)$$

Otherwise,

$$[\text{Vis}] = -7.12 \times \ln[\text{PM}_{10}] - 0.23 \times [\text{RH}] + 0.13 \times [\text{Temp}] + 62.08 \quad (1.7)$$

Для Гуанчжоу:

If [Temp] < 23°C,

$$[\text{Vis}] = -0.18 \times [\text{RH}] - 2.63 \times \ln[\text{PM}_{10}] - 0.02 \times [\text{SO}_2] + 35.86 \quad (1.8)$$

Otherwise,

$$[\text{Vis}] = -5.17 \times \ln[\text{PM}_{10}] + 0.52 \times [\text{Temp}] - 0.15 \times [\text{RH}] + 30.44 \quad (1.9)$$

Для Ченду:

If [RH] > 75% and [PM₁₀] > 100 µg/m³,

$$[\text{Vis}] = -0.30 \times [\text{RH}] + 0.08 \times [\text{Temp}] + 2.01 \times [\text{WS}] - 0.01 \times [\text{PM}_{10}] + 30.01 \quad (1.10)$$

Otherwise,

$$[\text{Vis}] = -0.12 \times [\text{RH}] - 0.01 \times [\text{PM}_{10}] + 0.05 \times [\text{Temp}] + 0.56 \times [\text{WS}] + 18.36 \quad (1.11)$$

де [Vis], [Temp], [RH], [WS], [SO₂], [NO₂] і [PM₁₀] - означають видимість (одиниця: км), температуру (одиниця: ° C), відносну вологість (одиниця: %),

швидкість вітру (одиниця: м / с), концентрації SO_2 , NO_2 і PM_{10} (одиниця: мкг / м³) відповідно [4].

Дослідження питань придатності сучасних супутникових систем та їх продуктів для аналізу стану повітря, встановлення математичних залежностей між супутниковими даними та результатами наземних спостережень, виявлення особливостей розподілу забруднення атмосферного повітря проводились і вітчизняними науковцями.

В.В. Путренко, С.Ю. Назаренко у своїй роботі [5], запропонували наукові та практичні засади моніторингу стану атмосферного повітря на основі використання даних супутникової зйомки Землі з метою дослідження вмісту часток $\text{PM}_{2.5}$ та NO_2 . На основі регресійного та географічно зваженого регресійного аналізу встановлено зв'язок між даними космічного моніторингу та наземних спостережень.

Садихов А.Б., Гасанов Р.Е., Ісмаїлова А.М. на основі лінійної регресії розробили методи прогнозування середньодобової концентрації забруднення атмосфери [6].

Шелестов А.Ю., Куссуль Н.М., Скакун С.В., Міронов А.І., Остапенко В.А., Яйлимов Б.Я. [7] в межах проекту JESAM-Ukraine та FP7 ImagineS весною та влітку 2013 року на території тестового полігону (с. Пшеничне, НУБіП України) проводили збір наземної інформації [7,8], для валідації глобальних супутникових продуктів: LAI (індекс листової поверхні), FAPAR (частка поглиненої фотосинтетичної активної радіації), FCover (частка рослинного покриву). Наземні дослідження проводились за протоколом Validation of LAnd European Remote sensing Instruments (VALERI) у відповідності до Європейських методик, стандартів CEOS Cal/Val та стандартів комітету GEO. Проаналізовано два типи регресійних моделей: лінійну та експоненціальну. Моделі побудовано для різних типів супутникових даних: Landsat-8 (просторове розрізнення 30 м), RapidEye (5 м) та SPOT-4 (20 м). Дані RapidEye та SPOT-4 представлено Європейським космічним агентством (ЄКА) в рамках ініціативи Take-5. На основі експериментальних досліджень показано, що експоненціальні регресійні моделі

найкраще описують залежність між NDVI та біофізичними параметрами у випадках значного насичення індексу NDVI. Середнє значення коефіцієнту детермінації експоненціальних моделей дорівнює 0,6, в той час як для лінійних моделей відповідний параметр дорівнює 0,5.

1.2 Повітря – його якісні властивості та вплив на здоров'я людини

У багатьох містах якість атмосферного повітря вже давно стоїть гострою та болючою проблемою. На це більше не можна закривати очі, потрібно активно аналізувати та вирішувати цю проблему. Піклування про здоров'я це дуже важливо. Величезний вплив на організм людини має повітря, яким всі дихають. Для життя людині, як і всім живим істотам потрібен кисень, але це лише одна хімічна сполука серед екзотичного різноманіття їх у повітрі. Окрім корисного та необхідного нам кисню у повітрі, ми також вдихаємо шкідливі і руйнівні Ozone (O_3), Nitrogen dioxide (NO_2), PM_{10} , $PM_{2.5}$, Carbon monoxide (CO).

PM — particulate matter. Зважені частинки це широкий і різноманітний клас шкідливих забруднювачів нашого повітря. Більшість людського населення не знає про цю загрозу для свого здоров'я.

Найчастіше з цього сімейства згадують саме про PM_{10} та $PM_{2.5}$. PM_{10} це позначення для часток речовин діаметр яких менше 10 мкм. $PM_{2.5}$ це позначення для часток речовин діаметр яких менше 2.5 мкм. Ці розміри можна легко уявити згадавши, що товщина людського волосся приблизно дорівнює 100 мкм. Вони настільки малі, що можуть потрапляти у легені, а не затримуватись у носовій порожнині, верхніх дихальних шляхах, бронхах. Частки $PM_{2.5}$ можуть потрапляти в альвеоли разом з повітрям, це бульбашки, що слугують для газообміну між кровоносними судами та легкими. З ними пов'язані захворювання не тільки дихальної системи а також і серцево-судинної. Ці частки можуть бути антропогенного або природного походження.

Частки $PM_{2.5}$ мають розміри менше 10 мкм. Виділяють два основних типи часток: вони можуть бути такі, що містять вуглець і метали, які у свою чергу можуть бути розділені на підрозділи на підставі своїх магнітних властивостей. У $PM_{2.5}$, які часто називають дрібнодисперсними зваженими частинками, також входять ультра дрібнодисперсні частинки діаметром менше 0,1 мкм. На більшості території Європи $PM_{2.5}$ складають 50-70% PM_{10} . У повітрі вміст часток може бути вимірний за допомогою конденсаційного лічильника часток, в якому частинки змішуються з парами спирту і потім остигають, при цьому пар конденсується на частинках і потім їх можна поррахувати за допомогою світлового сканера. Частки можуть бути антропогенного або природного походження. Через велику кількість і можливість проникати глибоко в легені $PM_{2.5}$ мають великий вплив на здоров'я дихальної системи. В Україні в діючій системі спостереження за станом атмосферного повітря контроль за вмістом $PM_{2.5}$ не здійснюється, що дає можливість для використання даних ДЗЗ для моніторингу та прогнозування впливу на здоров'я населення [5,9].

Carbon monoxide (CO) - монооксид вуглецю, це газ без кольору, смаку та запаху. Він заважає переносити кисень гемоглобіну в нашому організмі. Він утворюється внаслідок спалювання органічних речовин, наприклад, деревини чи вугілля при недостатньому доступі кисню, основне антропогенне джерело це вихлопні гази двигунів внутрішнього згорання.

У таблиці 1.1 наведено основні симптоми та ознаки які можуть виникнути у людини при зміні проценту концентрації.

Таблиця 1.1 – Ознаки та симптоми

Концентрація (%)	Загальні ознаки та симптоми
1	2
0,3-0,7	Відсутність ознак і симптомів. Нормальний ендогенний рівень.

Кінець таблиці 1.1

1	2
2,5-5	Відсутність симптомів. Компенсаційне збільшення кровотоку до деяких життєво важливих органів. У пацієнтів з серйозною серцево-судинною недостатністю може бути відсутнім компенсаційний резерв. Біль в грудях у пацієнтів страждають стенокардією викликається меншою кількістю фізичних зусиль.
5-10	Візуальний світловий поріг дещо збільшується.
10-20	Легкий головний біль. Аномалії візуально викликаної відповіді. Можлива невелика задишка при фізичних зусиллях. Може бути летальний для плода. Може бути летальний для пацієнтів з серйозною серцевою недостатністю.
20-30	Слабкий або помірний головний біль і пульсації в скронях. Прилив крові до шкіри. Нудота. Втрата спритності рук.
30-40	Серйозний головний біль, запаморочення, нудота і блювота. Слабкість. Дратівливість і порушення суджень. Непритомність при фізичних зусиллях.
40-50	Те ж, що і вище, але в більш серйозній формі і з більшою ймовірністю колапсу і непритомності.
50-60	Можлива кома з переривчастими конвульсіями і диханням Чейн-Стокса.
60-70	Кома з переривчастими конвульсіями. Пригнічення респіраторної системи і серцевої діяльності. Можливий летальний результат.
70-80	Слабкий пульс і уповільнення дихання. Пригнічення дихального центру, яке може викликати смерть.

Ozone (O_3) - дуже токсичний газ, що має блідо-голубий колір та гострий запах, що нагадує хлор. Молекула озону складається з трьох атомів кисню. Вона нестабільна і постійно намагається стати двохатомною молекулою кисню, так би мовити, здихатись одного зі своїх атомів, при комфортних для людей умовах, температурі та тиску. Якщо їй це вдається, на виході з'явиться звичайна молекула O_2 та один вільний атом кисню який буде дуже активно намагатися окислити щось інше. Тільки фтор є більш потужним окиснювачем ніж озон. Саме тому озон входить у категорію речовин першого класу загрози і для людини є виключно шкідливим. Незважаючи на це, він все ж приносить певну користь.

Високо в атмосфері він захищає землю від потужного ультрафіолетового випромінювання. Озон має руйнівний вплив на дихальну систему та підвищує ймовірність розвитку астматичного нападу, підвищує активність тромбоцитів та збільшує артеріальний тиск тому він є загрозою і для серцево-судинної системи також.

Діоксид азоту (NO_2) - це отруйний газ без запаху, жовто-бурого кольору, але при низьких температурах стає прозорим. Він має високу хімічну активність, виступає окиснювачем в реакціях з речовинами неметалами. Перетворюється в азотну кислоту при контакті з водою. Утворює нітрити та нітрати з лужним середовищем. Має другий клас загрози. Вливає переважно на органи дихальної системи. Викликає наслідки різного рівня від слабкого подразнення слизової очей до набряку легень. Сприяє зменшенню концентрації гемоглобіну в крові.

Двоокис азоту (NO_2) утворюється в навколишньому повітрі в результаті окислення оксиду азоту (NO), який у свою чергу з'являється в атмосфері природним шляхом і в результаті людської діяльності. Оксиди азоту з'являються в більшій мірі через згоряння викопного палива. Процес проходить внаслідок хімічної реакції між атмосферним N_2 і O_2 при теплі з утворенням NO , який потім входить в реакцію з O_2 з утворенням NO_2 . Швидкість реакції залежить від температури горіння. Таким чином, NO_2 не є складовою викопного палива, на відміну від SO_2 , а результатом каталітичної реакції тепла з атмосферним N_2 і O_2 в процесі горіння [5,10]. Особливо небезпечні оксиди азоту в містах, де вони взаємодіють з вуглецем вихлопних газів, де утворюють фотохімічний туман – смог [5,11]. Діоксид азоту (NO_2) є важливим при формуванні озону в тропосфері і несе загрозу для здоров'я людини і екосистеми. Тож NO_2 виділяється в тропосферу в результаті антропогенних явищ, наприклад: використання викопного палива, спалювання біопалива, спалювання біомаси і природних, наприклад: лісові пожежі і блискавки [5].

1.3 Міжнародні джерела даних

Для проведення даного дослідження використовуються дані з геопорталу Служби моніторингу атмосфери Copernicus (CAMS). CAMS є однією з шести служб, які формують Copernicus, програму спостереження Землі Європейського Союзу, яка розглядає нашу планету і її навколишнє середовище для кінцевої користі всіх громадян Європи. Основна концепція роботи Copernicus – надання інформації про якість повітря і склад атмосфери в межах і за межами Європи на основі супутникових і наземних вимірювань і спостережень в комбінації з моделями прогнозування (рисунок 1.1).

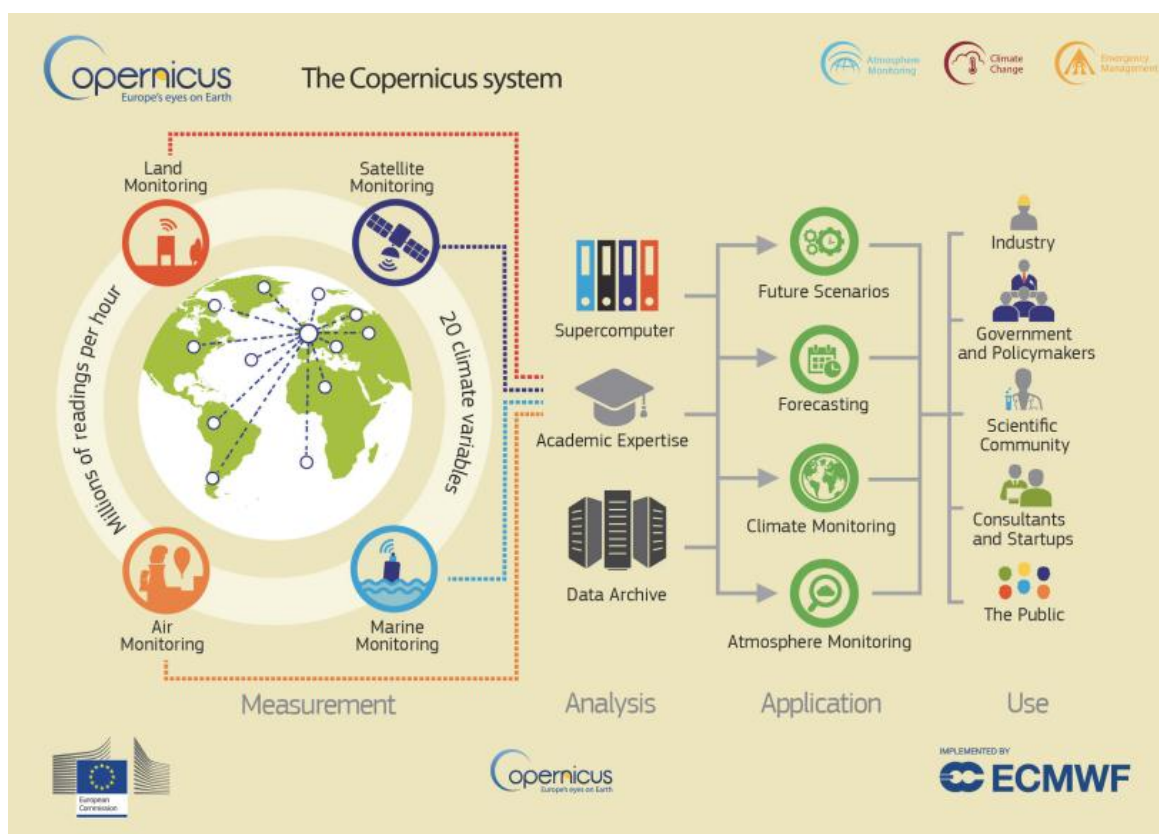


Рисунок 1.1 – Організація роботи системи Copernicus [12]

Copernicus надає користувачам дані через різні сервіси. Вони включають Службу зміни клімату (C3S) та Служба моніторингу атмосфер (CAMS), обидва керуються Європейським центром середньострокових прогнозів погоди (ECMWF). The Copernicus Atmosphere Monitoring Service (CAMS) забезпечує

безперервні дані та інформацію про склад атмосфери. Головними завданнями моніторингу атмосфери Copernicus є: опис поточної ситуації; прогноз ситуації на кілька днів попереду; аналіз послідовних ретроспективних даних за останні роки. Служба підтримує багато застосувань у різних областях, включаючи охорону здоров'я, моніторинг навколишнього середовища, поновлювані джерела енергії, метеорологію та кліматологію.

Надання послуг CAMS зосереджено на п'яти основних напрямках:

1. Якість повітря і склад атмосфери;
2. Озоновий шар і ультрафіолетове випромінювання;
3. Викиди та поверхневі потоки;
4. Сонячна радіація;
5. Примус до клімату.

Сервіс забезпечує щоденну інформацію про загальний склад атмосфери за допомогою компонентів моніторингу і прогнозування, таких як парникові гази (вуглекислий газ і метан), реактивні гази (наприклад, оксид вуглецю, окислені сполуки азоту, діоксид сірки), озон і аерозолі; забезпечує практичний аналіз в реальному часі та 4-денні прогнози, а також повторний аналіз якості європейського повітря, що дозволяє постійно оцінювати повітря, яким ми дихаємо; надає громадським та приватним організаціям, що займаються використанням сонячної енергії, відповідну і точну інформацію про ресурси сонячної радіації на поверхні Землі, що має важливе значення в областях, таких як охорона здоров'я, сільське господарство та відновлювані джерела енергії.

Всі інформаційні дані надані безкоштовно і без обмежень з метою підвищити рівень обізнаності про стан навколишнього атмосферного середовища серед політиків, бізнесменів і громадян.

Сервіс CAMS базується на семи сучасних моделях якості повітря, розроблених у Європі:

- 1) CHIMERE від INERIS (Франція),
- 2) EMEP від MET Norway (Норвегія),
- 3) EURAD-IM з Кельнського університету (Німеччина),

- 4) LOTOS- EUROS з KNMI і TNO (Нідерланди),
- 5) MATCH з SMHI (Швеція),
- 6) MOCAGE з METEO-FRANCE (Франція),
- 7) SILAM з FMI (Фінляндія).

Окрім них є ще ENSEMBLE модель, що є поєднанням усіх вищезгаданих моделей. Дані з цієї моделі доступні для території всієї Європи (25°W - 45°E , 30°N - 70°N).

CAMS впроваджується Європейським центром середньострокових прогнозів погоди (ECMWF) від імені Європейської Комісії. ECMWF є міжнародною незалежною організацією, що підтримується 34 країнами. Це як науково-дослідний інститут, так і цілодобова оперативний сервіс, що виробляє та поширює чисельні погодні прогнози для своїх країн-членів.

Для забезпечення та подальшого розвитку CAMS, ECMWF працює з багатьма постачальниками послуг по всій Європі. Роблячи це, CAMS поєднує досвід та інфраструктуру, які існують в Європі, щоб забезпечити цілий ряд послуг, які не мають собі рівних серед інших організацій у світі.

CAMS реалізується за допомогою Європейського центру середньорічних прогнозів погоди (ECMWF) і за підтримки понад 30 організацій, розташованих по всій Європі. Інформація про якість повітря оцінюється за допомогою Європейського індексу якості повітря, що відповідає визначенню ЕЕА. Індекс розраховується для п'яти найважливіших забруднювачів повітря, які регулюються Європейським законодавством: O_3 (озон), NO_2 (діоксид азоту), SO_2 (діоксид сірки), $\text{PM}_{2.5}$ і PM_{10} (зважені частинки діаметром менше 2,5 мікрометрів і 10 мікрометрів відповідно). Для кожної забруднюючої речовини, значення індексу лежить у межах від 1 (добре) до 5 (дуже погано). Європейський індекс якості повітря розраховується за різними забруднювачами повітря окремо за до концентраціями (в залежності від забруднювача повітря, в середньому за день чи моментально): чим більша концентрація, тим вище індекс. Європейський індекс якості повітря представлений цілим числом, відповідним п'яти діапазнам концентрацій, характерним для кожної забруднюючої речовини.

Загальний погодинний Європейський індекс якості повітря визначається як найвище значення з п'яти індивідуальних індексів забруднюючих речовин, розрахованих за один і той же час. Для прикладу, якщо індекси O_3 , NO_2 , SO_2 , $PM_{2.5}$ і PM_{10} рівні 1, 3, 1, 2, 2 відповідно, середній індекс буде дорівнює трьом. Загальний щоденний Європейський індекс якості повітря - це найвище значення загального погодинного Європейського індексу якості повітря за відповідний день. Загальний щоденний Європейський індекс якості повітря використовується у новинах на каналі Euronews.

CAMS може надавати прогнозовані дані. Точно так же, як і прогноз погоди, ця інформація заснована на передових математичних моделях і відображає закони фізики в комбінації з минулими спостереженнями. Внаслідок цього не виключена ймовірність неточного прогнозу. Однак оцінка за останні шість місяців 2017 року показала, що прогнози були коректними в межах одного значення індексу (були точними, більше або менше на одне значення) в більш ніж 98% випадків.

Дані які він надає, отримані методом розрахункової сітки з осередками розміром приблизно 10 км на 10 км по горизонталі, це означає, що вони не можуть відображати локальні ефекти (такі, як дорога з інтенсивним рухом в межах декількох сотень метрів і т. д.). У великих містах значення, надані CAMS також відображають концентрації так званого «міського фону», відповідно до районів міста, які не перебувають під впливом місцевих джерел, таких, як автомобільний рух безпосередньо (наприклад, в середині великого парку або в спальному районі). Очікується, що райони, що знаходяться під впливом місцевих джерел, можуть мати більш високу концентрацію NO_2 , $PM_{2.5}$, PM_{10} і SO_2 і більш низьку концентрацію O_3 .

Висновки до розділу 1

Повітря є сумішшю різноманітних хімічних елементів та сполук. Одні можуть бути корисними для людей та навколишнього середовища, а інші навпаки. Хімічні сполуки діляться на різні класи за рівнем шкоди, що вони можуть спричинити. В організмі живих істот вони можуть негативно впливати на різні важливі органи та їх системи, наприклад, серцево-судинну та дихальну. Багато окремих вітчизняних та зарубіжних науковців вже давно ведуть дослідження в цій області. За допомогою регресійного аналізу можна прогнозувати якість повітря за багатьма параметрами. Існують міжнародні компанії, що утримують великі відкриті бази даних з метеорологічною інформацією, які можна використовувати у різних дослідженнях.

2 ПРОГНОЗУВАННЯ РІВНЯ ЗАБРУДНЕННЯ ПОВІТРЯ ЗА ДОПОМОГОЮ РЕГРЕСІЙНОГО МОДЕЛЮВАННЯ

2.1 Регресійний аналіз

У сучасних компаній немає недостачі в інформації. Вони можуть збирати різні особисті дані про користувачів своїх сервісів в інтернеті, роботу своїх програм, показники приладів на промисловості, внутрішні процеси самої компанії. Деякі з цих даних виходять через рівні інтервали, наприклад, кожні п'ять хвилин або кожні півгодини, інші вимірюються безперервно. При невеликих додаткових витратах часу і зусиль завжди можна мати неперервний потік даних. Їх аналіз можна проводити періодично. В результаті такого аналізу виходять кількісні дані про властивості, що мають значення для компанії або для споживача. У багатьох компаній знаходяться гігантські об'єми подібних даних. І часто цифри просто колекціонуються без будь-якого розуміння мети або сенсу або в ім'я цілей, що ставилися в колишні роки. Незважаючи на те, що цих цілей більше вже не існує, дані все ж збираються.

Загалом, дані такого типу представляються у вигляді таблиць чисел. У цих числах можуть бути завуальовані деякі співвідношення або ж ці співвідношення можуть безпосередньо витікати з даних. Для будь-яких завдань з мінливою кількістю змінних, представляє інтерес дослідження впливу деяких змінних на інші. Таким впливом, звичайно, може бути простий функціональний зв'язок між змінними; однак, у багатьох фізичних процесах це скоріше виключення, ніж правило. Часто існує функціональний зв'язок, занадто складний для розуміння або для опису в простих термінах. У такому випадку слід прагнути підібрати апроксимацію цього функціонального зв'язку за допомогою якої-небудь простої математичної функції (скажімо, такої, як поліном), яка включає відповідні змінні, і згладжувати або апроксимувати «справжню» функцію в певній обмеженій області зміни цих змінних. При дослідженні такої згладженої функції

можна більше дізнатися про розглянуту «істинну» залежність та оцінити окремі або спільні ефекти зміни деяких важливих змінних.

Навіть тоді, коли за змістом не існує фізичного зв'язку між змінними, можна прагнути до того, щоб відобразити її за допомогою математичного рівняння даного виду. Якщо рівняння фізично безглузде, то воно тим не менш може виявитись доволі цінним для передбачення значень ряду змінних через відомі значення інших змінних, можливо, при визначених певних обмеженнях.

Найпростішим прикладом цієї процедури служить підбір прямої по парам спостережень $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$. Якщо задача включає більше змінних, то основним стає матричний підхід. Можна розрізнити два основних типи змінних. Назвемо їх предикторами або незалежними змінними (факторами) і залежними змінними або змінними відгуками. Під предикторами або факторами розуміють такі змінні, для яких зазвичай можна встановлювати бажані значення, або ті, які можна тільки спостерігати, але не керувати ними. В результаті навмисних змін або змін, що сталися з незалежними змінними випадково, з'являється ефект, який передається на інші змінні, на відгуки. Загалом, важливим є те, як зміни предикторів впливають на значення відгуків. Якщо можемо виявити просте співвідношення або залежність відгуку від одного або декількох факторів, то це, звичайно успішно. Поділ на предиктори і відгуки не завжди цілком чіткий і іноді залежить від наших цілей. Можна розглядати відгук проміжної стадії процесу як предиктор для, скажімо, наступної стадії. Практично, однак, ролі змінних зазвичай легко помітні. У конкретній масі даних дві або кілька змінних можуть змінюватися одночасно деяким певним чином, можливо, пов'язаним з методом, що лежить в основі експерименту. Це зазвичай не бажано - перш за все тому, що обмежується інформація про індивідуальну роль чинників, але часто неминуче.

Розглянемо метод аналізу, названий методом найменших квадратів. Його можна застосовувати для обробки даних експерименту і для отримання розумних висновків про властивості обраного рівняння. Цей метод часто називають регресійним аналізом. У багатьох експериментальних роботах необхідно

дослідити, як зміна однієї змінної впливає на іншу. Іноді дві змінні пов'язані точним рівнянням прямої лінії. Можна переконатися, що знати вид залежності дуже корисно і коли залежність строго лінійна, і коли вона лінійна лише для середніх значень. Прямолінійна залежність може бути корисна, навіть якщо відомо, що таке співвідношення взагалі не може бути вірним. Розглянемо залежність відгуку від фактору, показану на рис. 2.1.

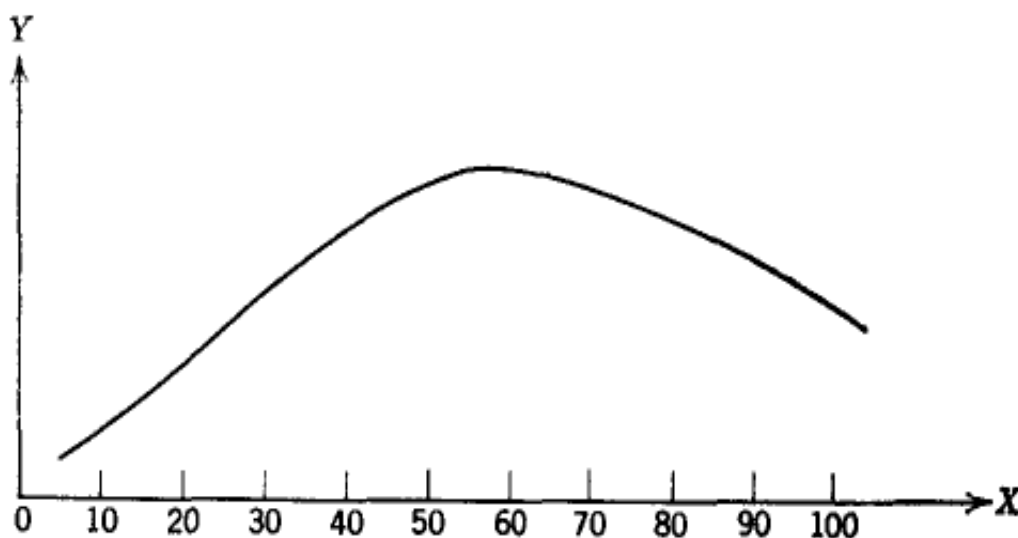


Рисунок 2.1 - Залежність відгуку від фактору

Вона, очевидно, нелінійна в діапазоні $0 < x < 100$, але якщо розглядати головним чином інтервал $0 < x < 45$, то лінійне рівняння для спостережень в цих межах могло б забезпечити цілком адекватне уявлення. Звичайно, побудоване рівняння не можна застосувати для значень X , що виходять за ці межі, так як воно не може забезпечити розумного передбачення.

Нехай потрібно дослідити, яким чином відгук Y залежить від факторів X_1, X_2, \dots, X_k . Отримаємо рівняння регресії для даних, які «покривають» деяку область в «просторі X ». Нехай точка $X_0 = (X_{10}, X_{20}, \dots, X_{k0})$ лежить поза області, що покривається початковими даними. Хоча математично можна отримати прогноз значення $Y(X_0)$ для відгуку в точці X_0 , потрібно розуміти, що довіряти такому прогнозу вкрай небезпечно і небезпека зростає при видаленні X_0 від вихідної області, якщо, звичайно, не додається деяка наявна додаткова інформація, що робить рівняння регресії придатним в широкій області простору

Х. Зауважимо, що іноді складно зрозуміти відразу, що цікава нам точка лежить за межами даної області багатовимірного простору [13].

Розглянемо МНК для рівняння поліному виду 2.1:

$$y = \sum_{j=0}^k b_j x^j \quad (2.1)$$

Потрібно мінімізувати суму квадратів:

$$S = \sum_{i=1}^n (\sum_{j=0}^k b_j x_i^j - y_i)^2 \rightarrow \min \quad (2.2)$$

Прирівняємо до 0 часткові похідні по всім $k+1$ змінним $S(b_0, b_1, \dots, b_k)$

$$S'_{b_p} = 2 \sum_{i=1}^n x_i^p (\sum_{j=0}^k b_j x_i^j - y_i) = 0, \quad p = \overline{0, k} \quad (2.3)$$

$$\sum_{i=1}^n x_i^p (b_0 + b_1 x_i + \dots + b_k x_i^k) - \sum_{i=1}^n x_i^p y_i = 0, \quad p = \overline{0, k} \quad (2.4)$$

$$\begin{cases} b_0 + b_1 \bar{x} + \dots + b_k \bar{x}^k = \bar{y} \\ \dots \\ b_0 \bar{x}^k + b_1 \bar{x}^{k+1} + \dots + b_k \bar{x}^{2k} = \bar{x}^k \bar{y} \end{cases} \quad (2.5)$$

Розглянемо МНК для рівняння прямої виду $y = ax + b$. Потрібно мінімізувати суму квадратів:

$$S = \sum_{i=1}^n (y_i - (ax_i + b))^2 \rightarrow \min \quad (2.6)$$

Візьмемо часткові похідні від S по a та b .

$$\begin{aligned} \frac{dS}{da} = \sum_{i=1}^n \left(2(y_i - (ax_i + b)) * (y_i - (ax_i + b))'_a \right) &= 2 \sum_{i=1}^n \left((y_i - ax_i - \right. \\ \left. -b) * (0 - (x_i + 0)) \right) &= 2 \sum_{i=1}^n ((y_i - ax_i - b) * (-x_i)) = 2 \sum_{i=1}^n (ax_i^2 + bx_i - \\ &-x_i y_i) \end{aligned} \quad (2.7)$$

$$\begin{aligned} \frac{dS}{db} = \sum_{i=1}^n \left(2(y_i - (ax_i + b)) * (y_i - (ax_i + b))'_b \right) &= 2 \sum_{i=1}^n \left((y_i - ax_i - \right. \\ \left. b) * (0 - (0 + 1)) \right) &= 2 \sum_{i=1}^n (ax_i + b - y_i) \end{aligned} \quad (2.8)$$

Прирівняємо похідні до 0 та розв'яжемо отриману систему.

$$\begin{cases} 2 \sum_{i=1}^n (ax_i^2 + bx_i - x_i y_i) = 0 \\ 2 \sum_{i=1}^n (ax_i + b - y_i) = 0 \end{cases} \quad (2.9)$$

$$a = \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i - N \sum_{i=1}^n x_i y_i}{(\sum_{i=1}^n x_i)^2 - N \sum_{i=1}^n x_i^2} \quad (2.10)$$

$$b = \frac{1}{N} (\sum_{i=1}^n y_i - a \sum_{i=1}^n x_i) \quad (2.11)$$

Ізотонічна регресійна модель працює за технікою підгонки лінії вільної форми до послідовності даних з наступними обмеженнями: лінія вільної форми повинна не зменшуватися (або не збільшуватися) скрізь, і повинна лежати якомога ближче до спостережень. Для цього потрібно мінімізувати $\sum_{i=1}^n \omega_i (y_i - \hat{y}_i)^2$, де $\omega > 0$ та $\hat{y}_{min} = \hat{y}_1 \leq \hat{y}_2 \leq \dots \leq \hat{y}_n = \hat{y}_{max}$.

Коефіцієнт детермінації показує долю дисперсії залежної змінної, що пояснюється розглянутою моделлю залежності. Вона показує, наскільки умовна дисперсія моделі відрізняється від дисперсії реальних значень. Для моделі залежності випадкової величини y від предикторів x , він визначається наступним чином

$$R^2 = 1 - \frac{V(y|x)}{V(y)} \quad (2.12)$$

$$R^2 = 1 - \frac{\sum(\hat{y}_i - y_i)^2}{\sum(y_i - \bar{y})^2} \quad (2.13)$$

де y_i - реальні значення y в кожному спостереженні, \hat{y}_i - значення, передбачені моделлю, \bar{y} - середнє по всіх реальних значеннях y_i . Якщо цей коефіцієнт близький до 1, то умовна дисперсія моделі досить мала і дуже ймовірно, що модель непогано описує дані. Якщо ж коефіцієнт R^2 сильно менше, наприклад, менше 0.5, то, з великою часткою впевненості модель не відображає реальний стан речей.

Однак, у статистики R^2 є один серйозний недолік: при збільшенні числа предикторів ця статистика може тільки зростати. Тому може здатися, що модель з більшою кількістю предикторів краща, ніж модель з меншою, навіть якщо всі нові предиктори ніяк не впливають на залежну змінну. Отже порівняння моделей з різним числом змінних за допомогою коефіцієнта детермінації, взагалі кажучи, некоректне. Для того, щоб, все ж таки, можна було порівнювати моделі з різною кількістю ознак так, щоб кількість ознак не впливала на статистику R^2 , була придумана статистика скоригований коефіцієнт детермінації, що використовує незміщені оцінки дисперсій.

$$R_{adj}^2 = 1 - (1 - R^2) \frac{n-1}{n-k} \leq R^2 \quad (2.14)$$

де n - кількість спостережень, а k - кількість параметрів. Він дає штраф за велику кількість предикторів, якщо нові незалежні змінні дають великий внесок в якість моделі, значення цієї статистики зростає, якщо ні - то навпаки зменшується.

2.2 Програмні інструменти та формати даних

Під час виконання дипломної роботи було використано різноманітне програмне забезпечення, основним з яких була високорівнева мова програмування Python. Вона має ряд переваг, через які її обирають багато розробників як інструмент для автоматизації, розробки програмного забезпечення та аналізу даних.

Одні з основних переваг мови Python полягають в зручності читання та використання інструментів у світі мов сценаріїв. Програмний код на мові Python зручно читати, а отже, його багаторазове використання і обслуговування проходить значно легше, аніж використання програмного коду на інших мовах програмування. Стиль оформлення програмного коду на мові Python, дозволяє легко його читати, навіть тим, хто не брав участі в його створенні. Крім того, Python використовує найсучасніші способи багаторазового використання програмного коду, таке як об'єктно-орієнтоване програмування (ООП).

У порівнянні з мовами, що вимагають компілювати їх код або строго типізованими мовами, наприклад, C, C ++ і Java, Python суттєво збільшує продуктивність праці розробника. Обсяг програмного коду на мові Python зазвичай складає третю або навіть п'яту частину аналогічного програмного коду на мові C ++ або Java. Це призводить до зменшення обсягів набору з клавіатури, зменшення часу, необхідного для налагодження і зменшення зусиль на супровід. Програми на мові Python запускаються одразу ж, оминаючи тривалі етапи компіляції і зв'язування, що є необхідними для деяких інших мов програмування, це значно підвищує продуктивність праці програміста.

Багато програм на мові Python працює однаково на всіх основних платформах. Експорт програмного коду з операційної системи Linux в Windows і навпаки, часто полягає в простому перенесенні файлів програм з одного комп'ютера до іншого. Більш того, Python надає багато можливостей для створення програм для роботи з базами даних, графічних інтерфейсів для користувачів, веб-додатків та інших видів програм. Інтерфейси операційних

систем, що можуть запускати програми і обробляти каталоги, в мові Python реалізовані переносним способом.

В середині Python міститься багато зібраних переносимих і функціональних можливостей, з яких складається стандартна бібліотека. Вона надає багато можливостей, що необхідні в прикладних програмах, такі як пошук тексту по шаблону і робота з мережевими функціями. Окрім того, Python дозволяє використовувати і ваші власні бібліотеки і бібліотеки створені сторонніми розробниками. З числа сторонніх розробок можна назвати інструменти для проведення математичних розрахунків, розробки веб-сайтів, розробку ігор, доступ до послідовного порту та інше. Наприклад, бібліотека NumPy є вільним і значно потужнішим аналогом системи для математичних обчислень Matlab.

Код мови Python може легко працювати з іншими частинами програми завдяки різним механізмам інтеграції. Це дозволяє використовувати Python для розширення і налаштування функціональних можливостей програм. У сучасності програмний код на мові Python здатний сам викликатися з програм, написаних на мові C / C ++, викликати функції з розширень на мові C / C ++, інтегруватися зі складовими компонентами мови Java, проводити обмін даними через послідовний порт або через мережу за допомогою таких протоколів, як XML-RPC, SOAP і CORBA та взаємодіяти з такими платформами, як .NET і COM. Python - не відокремлений інструмент.

Завдяки невимушеності мови Python і наявності вбудованих інструментів процес програмування може приносити задоволення. На перший погляд це важко назвати перевагою, тим не менше, задоволення, отримане від роботи, безпосередньо впливає продуктивність праці.

Python є інструментом, що дозволяє програмістам отримувати більшу віддачу при менших зусиллях. Він оптимізований для досягнення високої швидкості розробки - простий синтаксис, динамічна типізація, відсутність етапу компіляції і вбудовані інструментальні засоби дозволяють програмістам створювати програми за менший час, ніж при використанні деяких інших

інструментів. В результаті Python збільшує продуктивність праці розробника в багато разів у порівнянні з традиційними мовами програмування.

За своєю природою Python має простий, легкий для читання синтаксис і ясну модель програмування. Python «Кожному по плечу», тобто характеристики мови взаємодіють обмеженим числом не суперечливих способів і природно впливають з невеликого кола базових концепцій. Це робить мову простою в освоєнні, розумінні і запам'ятовуванні. На практиці програмістам, які використовують мову Python, майже не доводиться вдаватися до довідкових посібників - це несуперечлива система, на виході якої, на подив багатьох, виходить професіональний програмний код [14].

Географічні інформаційні системи - основний інструмент для всіх, чия робота пов'язана зі створенням, управлінням або аналізом геопросторових даних в різних сферах охорони природи, архітектури і містобудування, лісового, сільського та водного господарства, земельного кадастру, метеорології, транспорту і логістики. Таким чином, практичні навички роботи з ГІС трансформувалася з області вузькоспеціалізованої компетенції до загальноприйнятого стандарту.

QGIS - вільна кросплатформена геоінформаційна система, що складається з настільної і серверної частин. Вона працює в Windows і в більшості платформ Unix (включаючи Mac OS), підтримує безліч векторних і растрових форматів і баз даних, а також має багатий набір вбудованих інструментів.

Робота над QGIS була розпочата в травні 2002 р., а в 2007 р. вона стала проектом Open Source Geospatial Foundation (OSGeo) - міжнародної некомерційної організації, створеної для підтримки спільної розробки і використання геоінформаційного ПЗ з відкритим кодом.

На сьогоднішній день це одна з найбільш динамічно ростучих і функціональних настільних ГІС. Виходячи з умов ліцензії GNU General Public License, використання, копіювання та поширення QGIS для будь-яких цілей, в т. ч. комерційних, не вимагає фінансових відрахувань. Завдяки відкритості програмного коду, користувачі не тільки можуть вивчати особливості пристрою

QGIS, але і модифікувати її відповідно до власних потреб. Розробка QGIS ведеться міжнародною групою розробників, яка з 2014 р. перейшла на чотирьох місячний цикл релізів. Таким чином, нова версія виходить три рази в рік. Вона гнучко взаємодіє з різними апаратними базами, операційними системами та програмним забезпеченням, способами уявлення геоданих і їх просторовими характеристиками. Вона може підтримувати різні формати і моделі даних, а саме: більше 60 форматів растрових даних (бібліотека Geospatial Data Abstraction Library - GDAL), більш 20 - векторних (OGR Simple Features Library), взаємодіяти з базами геоданих, OGC-сервісами, взаємодіяти з даними в різних проекціях і системах координат (в т. ч. і користувацьких) через бібліотеку проекцій Proj.4. QGIS спочатку замислювалася як переглядач просторових таблиць PostGIS, але з часом перетворилася на повну ГІС, здатну вирішувати широкий спектр завдань, що наведений у таблиці 2.1.

Таблиця 2.1 – Можливості програми QGIS

1	2
Створення геоданих	<ul style="list-style-type: none"> • геокодування; • створення і редагування векторних (shape) файлів, в т.ч. з підтримкою топологій; • створення і редагування атрибутивних даних; • інструменти для імпорту і експорту даних GPS; • створення і редагування таблиць просторових баз даних; • вивантаження і редагування даних OpenStreetMap.
Управління геоданими	<ul style="list-style-type: none"> • підтримка стандартних проекцій (більше 2 700), а також параметрів переходу між різними системами координат; • створення призначених для користувача проекцій; • перепроеціювання «на льоту»; • перепроеціювання векторних і растрових шарів; • перевірка топології; • перегляд / пошук атрибутів (SQL-запити); • визначення / вибірка об'єктів (SQL-запити).

Кінець таблиці 2.1

1	2
Аналіз геоданих	<ul style="list-style-type: none"> • функції геообробки: буферні зони, відсікання, об'єднання та ін.; • просторові запити; • калькулятор полів атрибутів; • калькулятор растрів; • морфометричний аналіз; • найбільш широкі можливості аналізу геоданих надає фреймворк геообробки QGIS, а саме - доступ до більш ніж 500 алгоритмів інших Відкритих ГІС (зокрема, GRASS, SAGA, Orfeo Toolbox), користувацьким скриптам. Крім того, він дозволяє автоматизувати процеси завдяки можливостям пакетної геообробки і створення моделей аналізу.
Подання геоданих	<ul style="list-style-type: none"> • зміна символіки векторних і растрових шарів; • підпис об'єктів; • компоновка карт для створення карт і атласів; • публікація карт в Інтернет.

CAMS надає дані у форматі netCDF. На відміну від таблиць формату csv, netCDF має складну структуру. NetCDF (network Common Data Form) - це двійковий формат файлів, призначений для створення, доступу і публікації наукових даних. Він широко використовується фахівцями метеорологами і океанографами для зберігання змінних, як, наприклад, даних про температуру, тиск, швидкості вітру і висоти хвиль. Дані NetCDF є:

- такі, що описують самі себе, що означає, що в файл netCDF входить інформація про його дані, в тому числі де були отримані елементи цих даних і які одиниці вимірювання використовувалися.

- портативними, або крос-платформеними, тобто файл netCDF, створений в операційному середовищі одного типу, часто може читатися програмними засобами, що працюють в іншій операційній системі.

- масштабовними, тобто є можливість доступу до невеликого піднабору великого файлу netCDF, без зчитування всього файлу netCDF.

Файл netCDF містить вимірювання, змінні і атрибути. Спільно ці компоненти використовуються для збереження значення даних, а також

відносин між полями даних в наборі даних, що складається з масивів. Осі координат netCDF мають і ім'я, і розмір. Розміром осі координат є будь-яке позитивне ціле число. У файлі netCDF тільки одна вісь координат може мати розмір UNLIMITED. Така вісь координат є необмеженою або віссю запису. За допомогою осі координат можна уявляти реальний фізичний вимір, наприклад, час, широту, довготу або висоту. За допомогою осі координат також можна індексувати інші величини, наприклад, номер прогону станції чи моделі.

Змінна представляє масив значення одного типу. У файлі netCDF змінні використовуються для зберігання даних. У змінної є ім'я, тип даних і форма, описані за допомогою списку її осей, які задаються при створенні змінної. Число осей є рангом (або розмірністю). Скалярна змінна має розмірність 0, вектор має розмірність 1, а матриця має розмірність 2. Одновимірною змінною, що має таке ж ім'я, як і вимір, є координатною змінною. Вона пов'язана з виміром однієї або декількох змінних даних і зазвичай визначає фізичну координату, яка відповідає цьому виміру. Координатні змінні не мають особливого значення для бібліотеки netCDF. Однак, програмне забезпечення, що використовує цю бібліотеку, має обробляти координатні змінні особливим чином.

Атрибути NetCDF використовуються для зберігання допоміжних даних або метаданих. Більшість атрибутів надають відомості про певну змінну. Ці атрибути визначаються за іменем змінної, що використовується разом з ім'ям самого атрибута. Атрибути, що надають відомості про весь файл netCDF, є глобальними атрибутами.

У правилах найменування визначаються метадані, що містять точний опис даних, що зберігаються в кожній змінній, а також їх просторові і тимчасові властивості. Завдяки правилам найменування користувачі даних з різних джерел можуть вирішувати, які кількості можна порівняти. Ім'я правила найменування представлено в файлі netCDF у вигляді глобального атрибута.

Висновки до розділу 2

Регресійний аналіз є потужним математичним апаратом для дослідження даних, що дозволяє передбачувати значення залежної змінної за допомогою незалежних, визначати вклад окремих незалежних змінних у варіацію залежної. Одним із важливих методів регресійного аналізу є метод найменших квадратів, що дозволяє побудувати лінійну функцію, яка найкращим чином наближає шукану криву.

Методи регресійного аналізу легко алгоритмізуються, що дозволяє легко застосовувати їх на обчислювальних системах за допомогою сучасного програмного забезпечення, яке значно полегшує роботу науковцям і дослідникам. Існує багато програмного забезпечення для математичних розрахунків. Однак для реалізації розробленої регресійної моделі концентрацій забруднювачів повітря на основі наземних вимірів та даних з CAMS, було обрано мову програмування Python, яка виявилась зручною для швидкого моделювання. Для роботи з даними з вузької предметної області існують вузькоспеціалізовані програми, такі як Qgis для геоданих.

3 ПОБУДОВА РЕГРЕСІЙНОЇ МОДЕЛІ НА ОСНОВІ ТОЧКОВИХ НАЗЕМНИХ ТА СУПУТНИКОВИХ ВИМІРІВ

Для виконання дослідження та побудови рівнянь регресії концентрацій забруднювачів повітря було використано дані, надані сервісом CAMS, та дані, які були виміряні на рівні земної поверхні.

CAMS надає дані по пікселям, що охоплюють велику територію площі землі. За допомогою апаратів, що вимірюють концентрації забруднювачів повітря: Air quality monitor sdl607 та розробки Noosphere Engineering School було отримано дані на рівні землі. Так як зазначені апарати надають дані майже за кожну секунду, а CAMS надає дані погодинно, було усереднено дані з наземних апаратів для години. Співставивши дані за часом вимірювання та координатами місця, в якому було виміряно наземні дані, із координатами пікселя, який накривав цю територію в час проведення наземних вимірювань, отримано дані для побудови регресійної моделі.

В результаті проведення розрахунків було отримано рівняння, для яких незалежною змінною є дані, отримані з міжнародного сервісу CAMS, вони подаються у $\mu\text{g}/\text{m}^3$, а залежна змінна - це дані виміряні на рівні землі, вони подаються у $\mu\text{g}/\text{m}^3$ для PM_{10} , $\text{PM}_{2.5}$ та у ppm для CO , O_3 , NO_2 .

Для кожного із забруднювачів повітря було побудовано поліноміальні моделі регресії.

Функція поліноміальної регресії для PM_{10} матиме вигляд 3.1.

$$\text{PM}_{10} = 369 - 94.4 * \text{PM}_{10c} + 8.43 * \text{PM}_{10c}^2 - 0.295 * \text{PM}_{10c}^3 + 3.58\text{E-}03 * \text{PM}_{10c}^4 \quad (3.1)$$

де PM_{10} - концентрація PM_{10} виміряна, PM_{10c} - концентрація PM_{10} з CAMS.
 $R^2 = 0,788$.

Поліноміальний зв'язок між даними з CAMS та наземними вимірами концентрації PM_{10} наведено на рисунку 3.1.

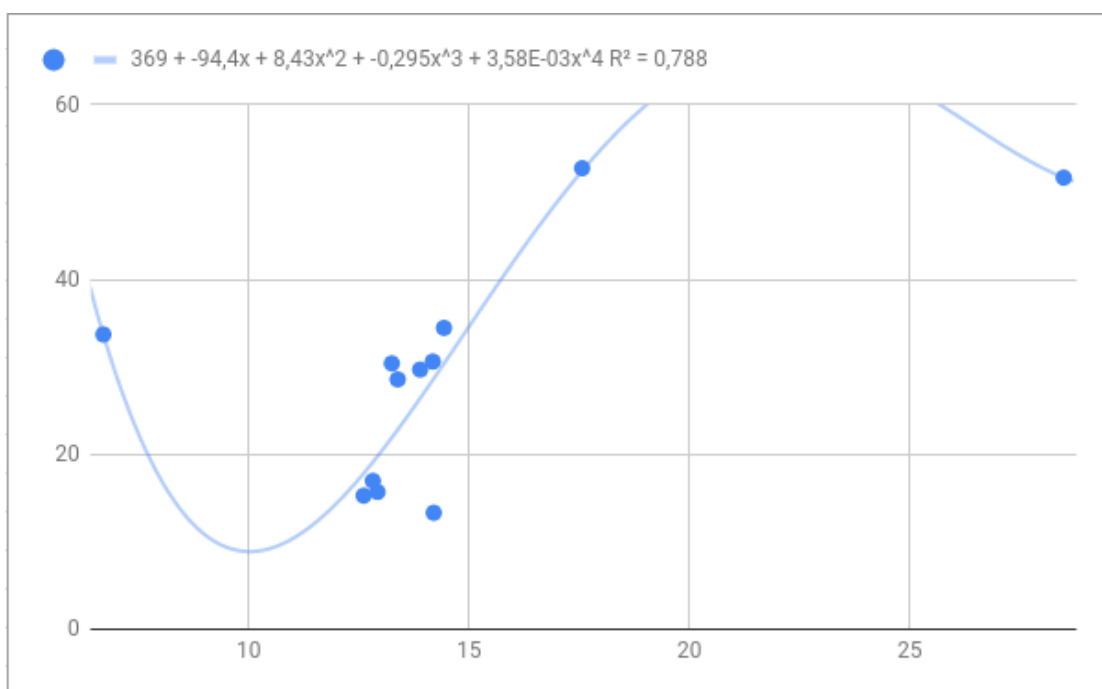


Рисунок 3.1 - Поліноміальний зв'язок між даними з SAMS та наземними вимірами концентрації PM_{10}

Поліноміальний зв'язок між даними з SAMS та наземними вимірами концентрації $PM_{2.5}$ наведено на рисунку 3.2.

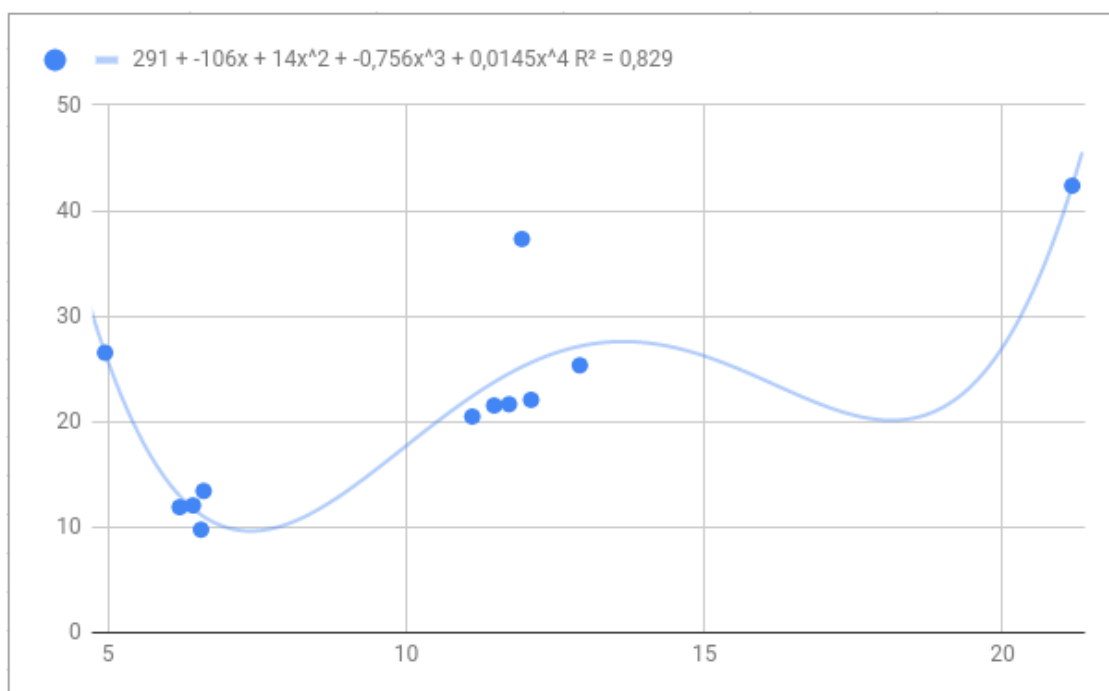


Рисунок 3.2 - Поліноміальний зв'язок між даними з SAMS та наземними вимірами концентрації $PM_{2.5}$

Функція поліноміальної регресії для $PM_{2.5}$ матиме вигляд 3.2.

$$PM_{2.5} = 291 - 106 * PM_{2.5C} + 14 * PM_{2.5C}^2 - 0.756 * PM_{2.5C}^3 + 0.0145 * PM_{2.5C}^4 \quad (3.2)$$

де $PM_{2.5}$ - концентрація $PM_{2.5}$ виміряна, $PM_{2.5C}$ - концентрація $PM_{2.5}$ з CAMS.

$$R^2 = 0.829.$$

Поліноміальний зв'язок між даними з CAMS та наземними вимірами концентрації CO наведено на рисунку 3.3.

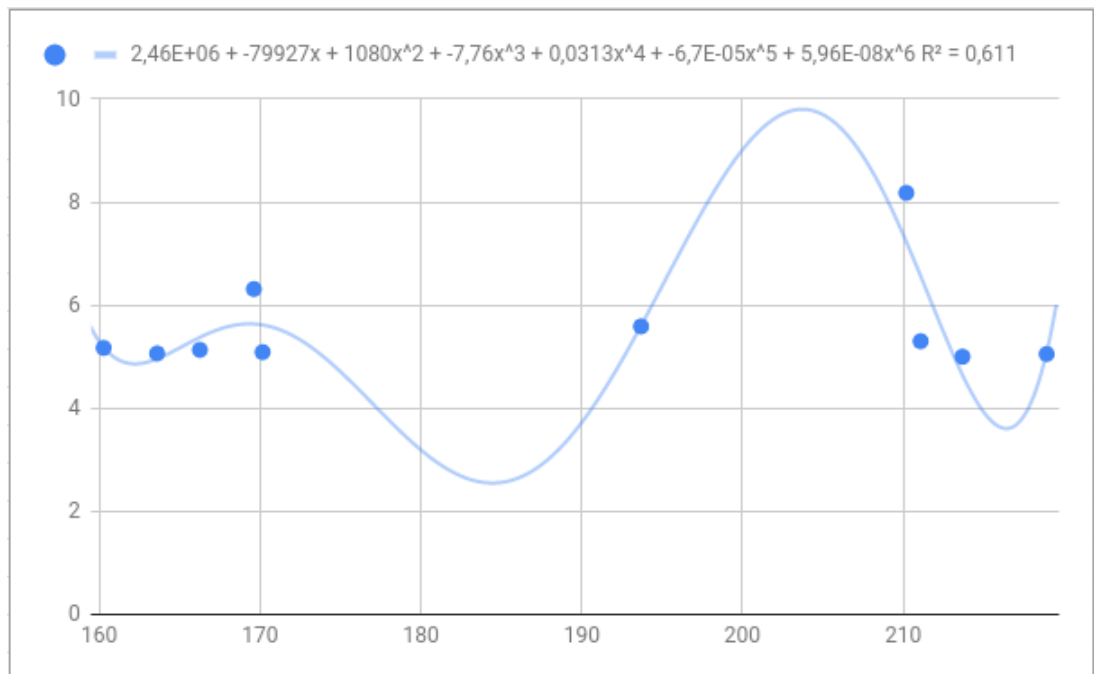


Рисунок 3.3 - Поліноміальний зв'язок між даними з CAMS та наземними вимірами концентрації CO

Функція поліноміальної регресії для CO матиме вигляд 3.3.

$$CO = 2.46E+06 - 79927 * CO_c + 1080 * CO_c^2 - 7.76 * CO_c^3 + 0.0313 * CO_c^4 - 6.7E-05 * CO_c^5 + 5.96E-08 * CO_c^6 \quad (3.3)$$

де CO - концентрація CO виміряна, CO_c - концентрація CO з CAMS.

$$R^2 = 0.611.$$

Функція поліноміальної регресії для O_3 матиме вигляд 3.4.

$$O_3 = 2.36E+07 - 1.11E+06 * O_{3c} + 20689 * O_{3c}^2 - 193 * O_{3c}^3 + 0.903 * O_{3c}^4 - 1.69E-03 * O_{3c}^5 \quad (3.4)$$

де O_3 - концентрація O_3 виміряна, O_{3c} - концентрація O_3 з CAMS.

$$R^2 = 0.58.$$

Поліноміальний зв'язок між даними з CAMS та наземними вимірами концентрації O_3 наведено на рисунку 3.4.

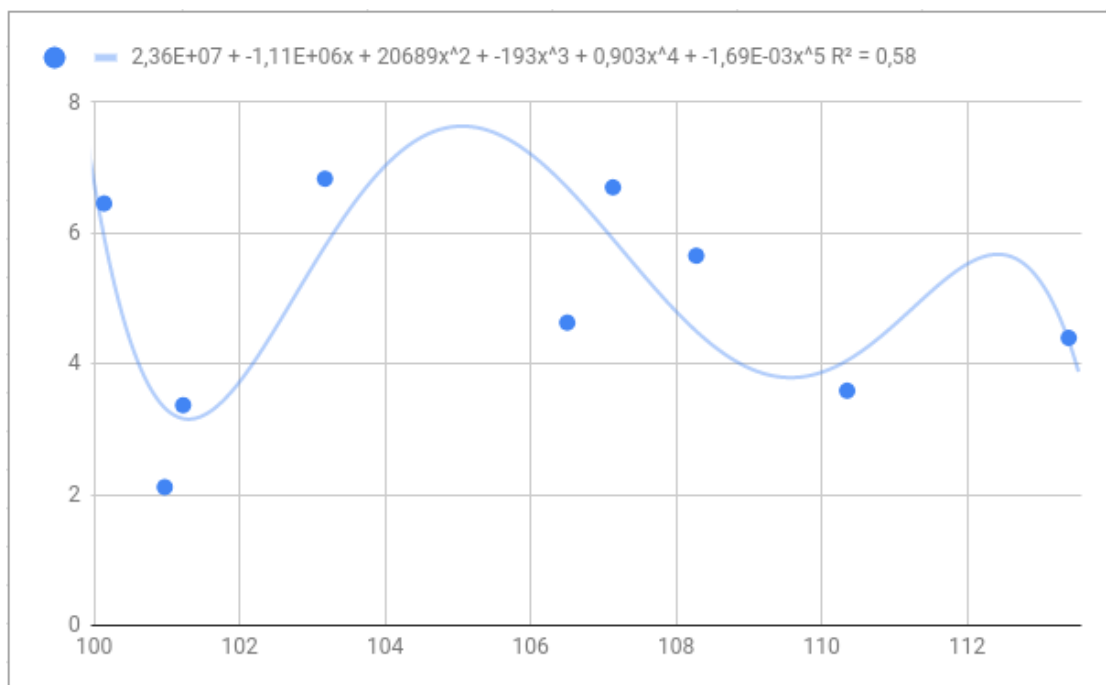


Рисунок 3.4 - Поліноміальний зв'язок між даними з CAMS та наземними вимірами концентрації O_3

Функція поліноміальної регресії для NO_2 матиме вигляд 3.5.

$$NO_2 = -356 + 920 * NO_{2c} - 950 * NO_{2c}^2 + 522 * NO_{2c}^3 - 168 * NO_{2c}^4 + 32.6 * NO_{2c}^5 - 3.75 * NO_{2c}^6 + 0.235 * NO_{2c}^7 - 6.16E-03 * NO_{2c}^8 \quad (3.5)$$

де NO_2 - концентрація NO_2 виміряна, NO_{2c} - концентрація NO_2 з CAMS.

$$R^2 = 0.929.$$

Поліноміальний зв'язок між даними з CAMS та наземними вимірами концентрації NO₂ наведено на рисунку 3.5.

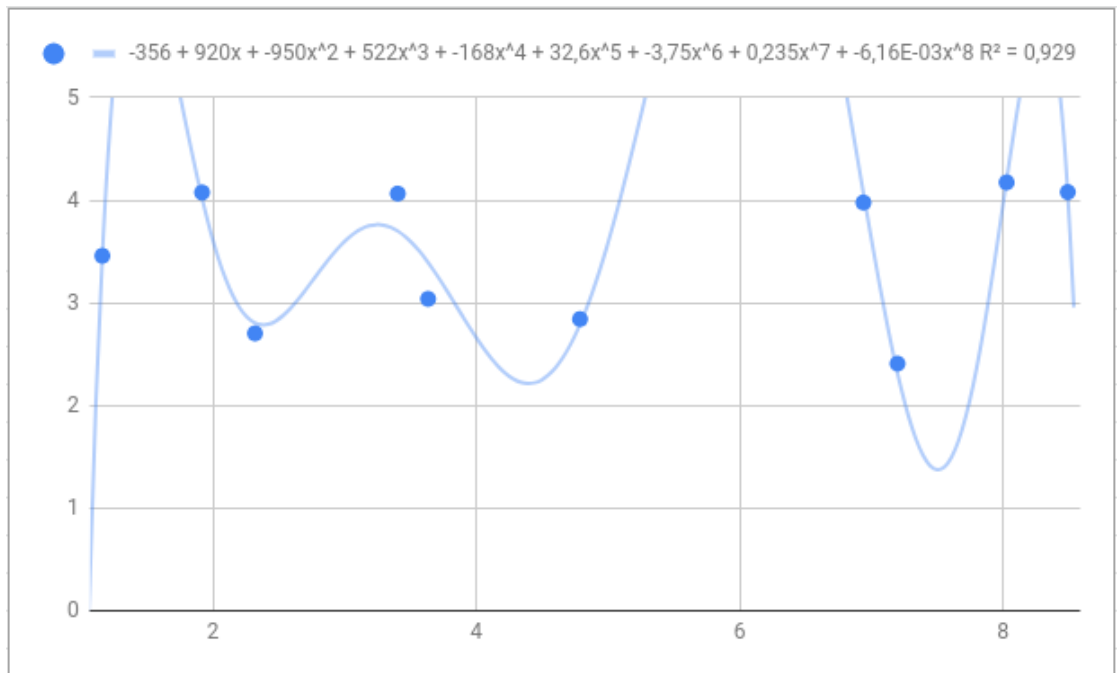


Рисунок 3.5 - Поліноміальний зв'язок між даними з CAMS та наземними вимірами концентрації NO₂

Для порівняння також побудовано лінійні моделі регресії для кожного із забруднювачів повітря.

Функція регресії для PM₁₀ матиме вигляд 3.6.

$$PM_{10} = 7.37006318 + 1.5173128 * PM_{10C} \quad (3.6)$$

де PM₁₀ - концентрація PM₁₀ виміряна, PM_{10C} - концентрація PM₁₀ з CAMS.

$$R^2 = 0.33920542770590123.$$

Лінійний зв'язок між даними з CAMS та наземними вимірами концентрації PM₁₀ наведено на рисунку 3.6.

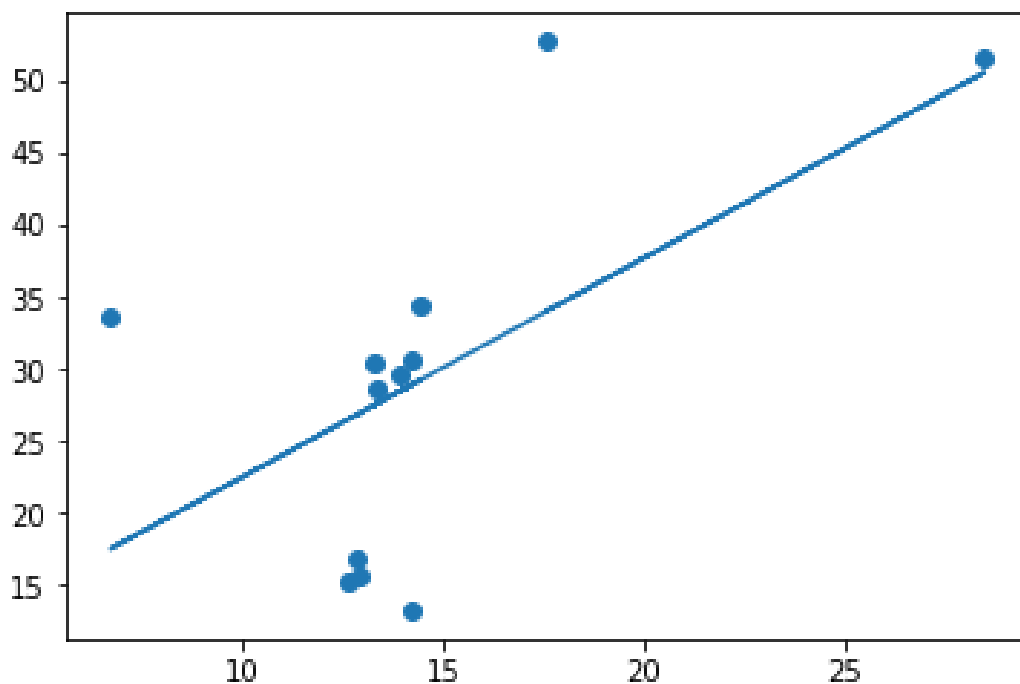


Рисунок 3.6 - Лінійний зв'язок між даними з SAM5 та наземними вимірами концентрації PM₁₀

На рисунку 3.7 наведено лінійний зв'язок між даними з SAM5 та наземними вимірами концентрації PM_{2.5}:

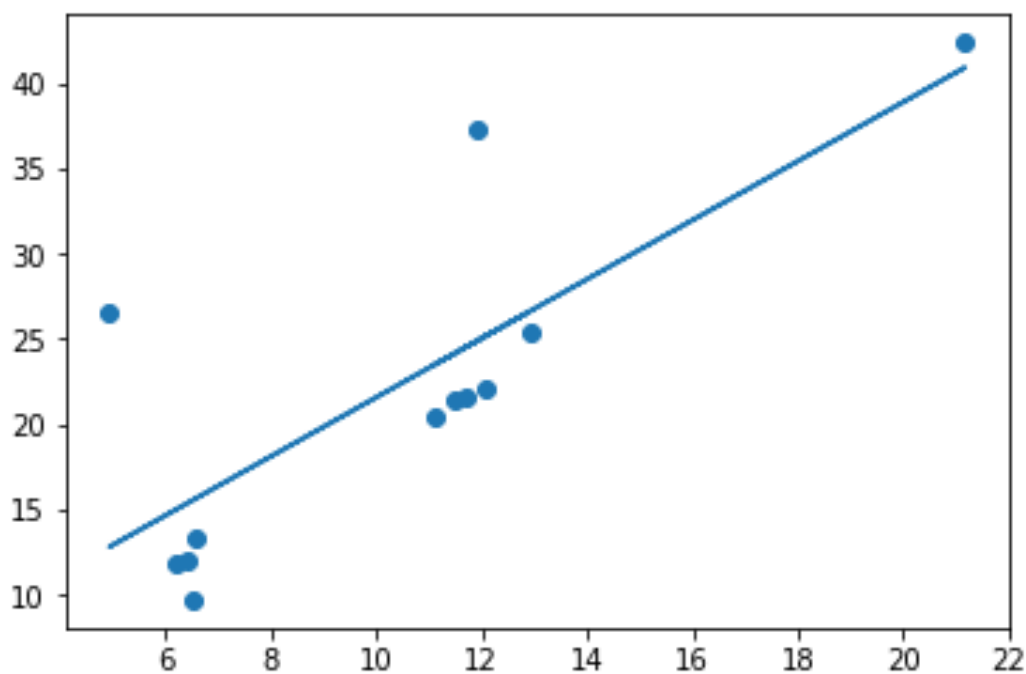


Рисунок 3.7 - Лінійний зв'язок між даними з SAM5 та наземними вимірами концентрації PM_{2.5}

Функція регресії для $PM_{2.5}$ матиме вигляд 3.7.

$$PM_{2.5} = 4.25851712 + 1.73174602 * PM_{2.5C} \quad (3.7)$$

де $PM_{2.5}$ - концентрація $PM_{2.5}$ виміряна, $PM_{2.5C}$ - концентрація $PM_{2.5}$ з SAMS.

$$R^2 = 0.6038027537234036.$$

Далі розглянемо лінійний зв'язок між даними з SAMS та наземними вимірами концентрації CO за допомогою рисунку 3.8.

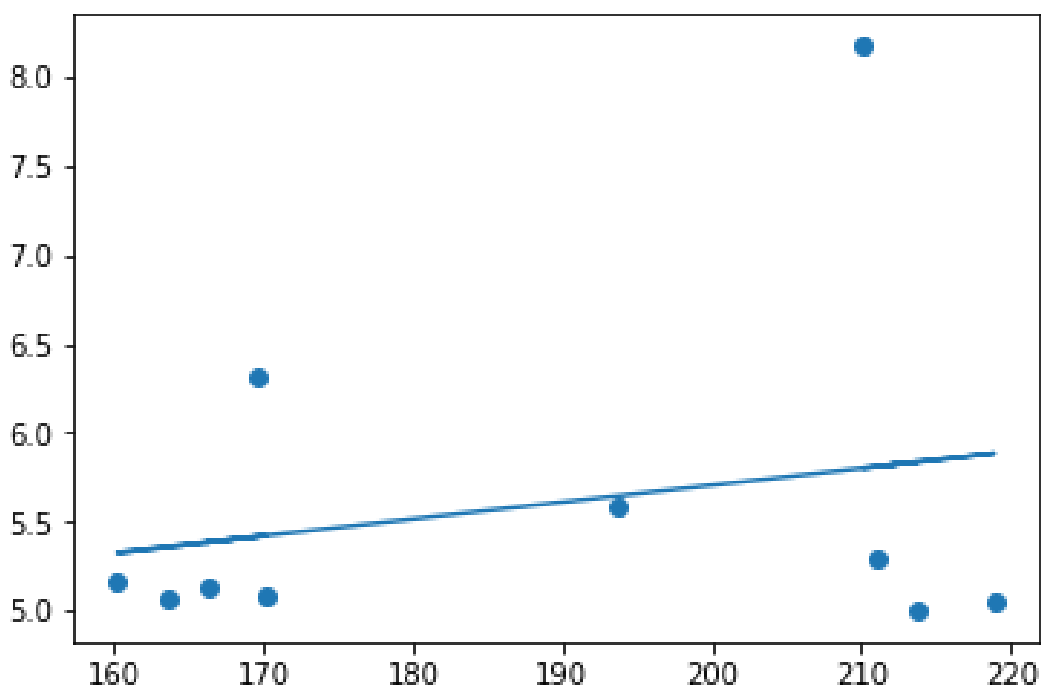


Рисунок 3.8 - Лінійний зв'язок між даними з SAMS та наземними вимірами концентрації CO

Функція регресії для CO матиме вигляд 3.8.

$$CO = 3.80423911 + 0.00949962 * CO_C \quad (3.8)$$

де CO - концентрація CO виміряна, CO_C - концентрація CO з SAMS.

$$R^2 = 0.05265413785334838.$$

Лінійний зв'язок між даними з CAMS та наземними вимірами концентрації O_3 наведено на рисунку 3.9.

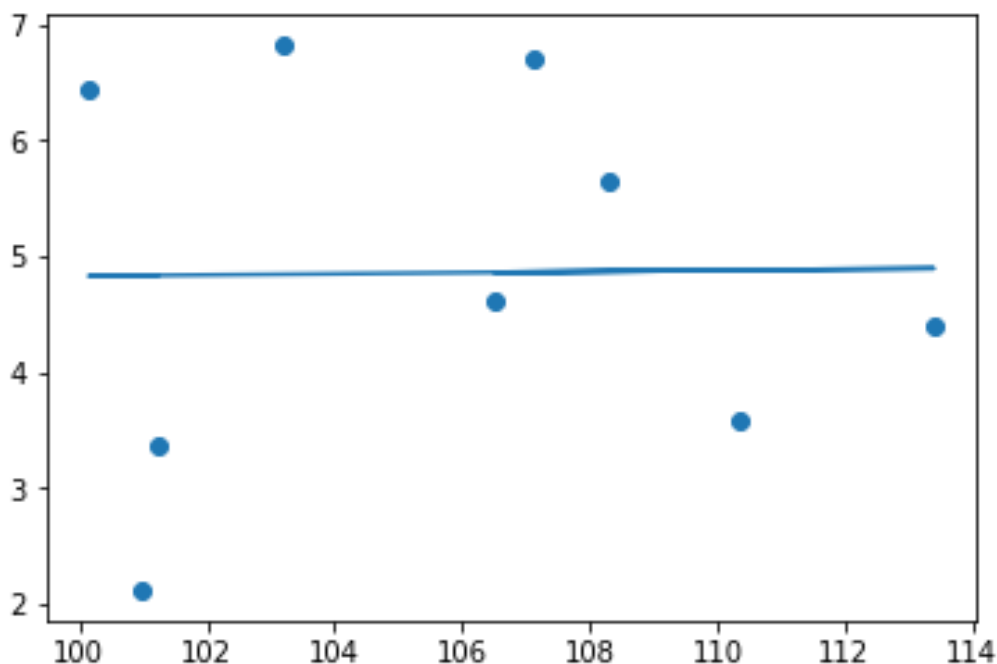


Рисунок 3.9 - Лінійний зв'язок між даними з CAMS та наземними вимірами концентрації O_3

Функція регресії для O_3 матиме вигляд 3.9.

$$O_3 = 4.33217159 + 0.00498511 * O_{3C} \quad (3.9)$$

де O_3 - концентрація O_3 виміряна, O_{3C} - концентрація O_3 з CAMS.

$$R^2 = 0.0001912419156294698.$$

Функція регресії для NO_2 матиме вигляд 3.10.

$$NO_2 = 3.26586271 + 0.04498713 * NO_{2C} \quad (3.10)$$

де: NO_2 - концентрація NO_2 виміряна, NO_{2C} - концентрація NO_2 з CAMS.

$$R^2 = 0.0319749269316717.$$

Для NO_2 лінійний зв'язок між даними з CAMS та наземними вимірами концентрації матиме вигляд наведений на рисунку 3.10.

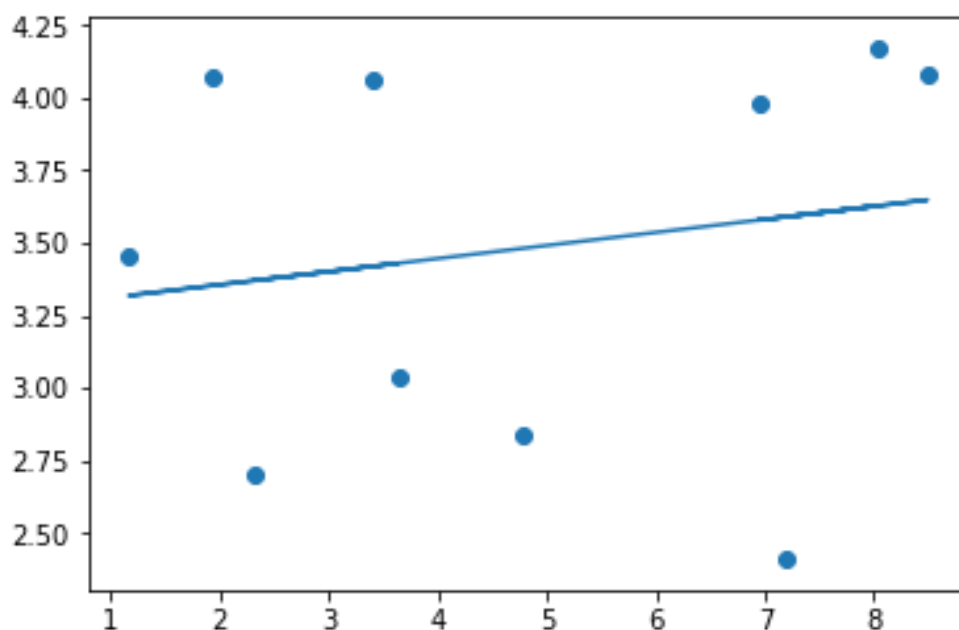


Рисунок 3.10 – Лінійний зв'язок між даними з SAMS та наземними вимірами концентрації NO_2

Також для більшого порівняння якості різних моделей, які оцінюють дані SAMS та наземні виміри концентрації забруднювачів повітря, були побудовані ізотонічні регресійні моделі.

Для PM_{10} ізотонічна регресійна модель матиме вигляд зображений на рисунку 3.11.

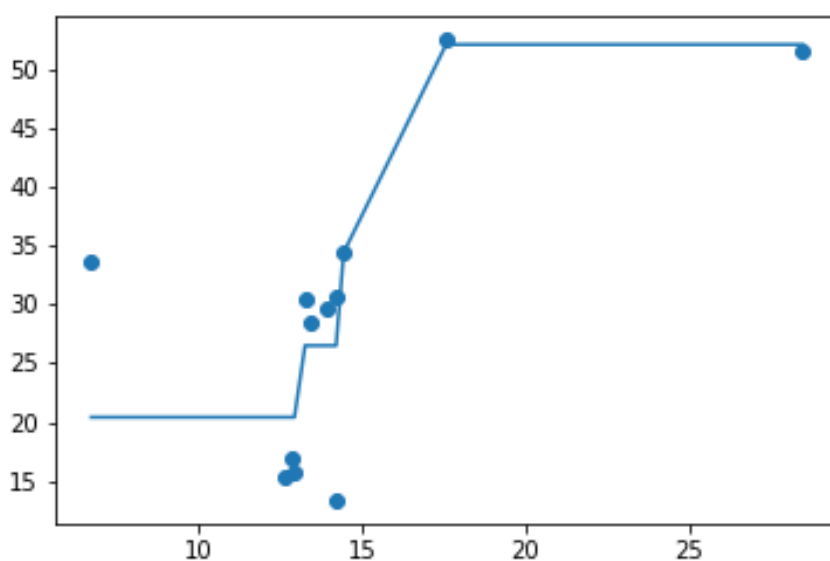


Рисунок 3.11 - Ізотонічна регресія для PM_{10}

Оцінка $R^2 = 0.7571511318329747$.

Для $PM_{2.5}$ ізотонічна регресійна модель матиме вигляд зображений на рисунку 3.12.

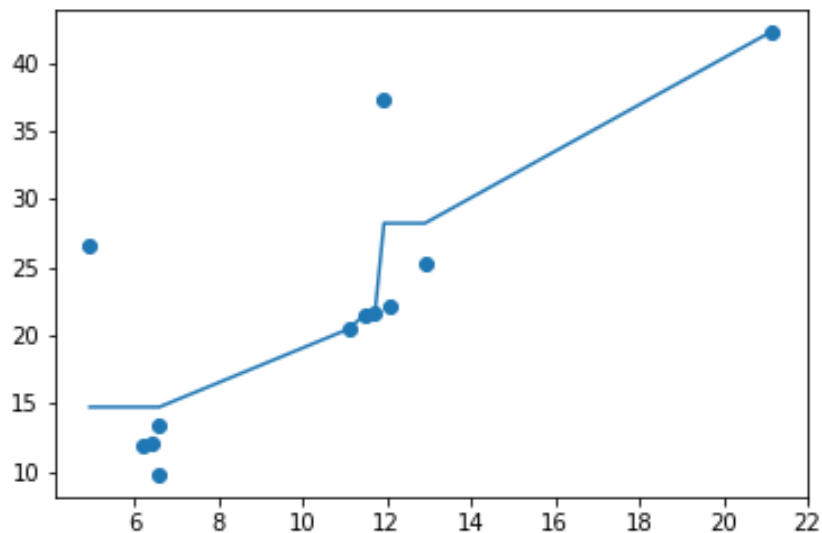


Рисунок 3.12 - Ізотонічна регресія для $PM_{2.5}$

Оцінка $R^2 = 0.7204647612842787$.

Для CO ізотонічна регресійна модель матиме вигляд зображений на рисунку 3.13.

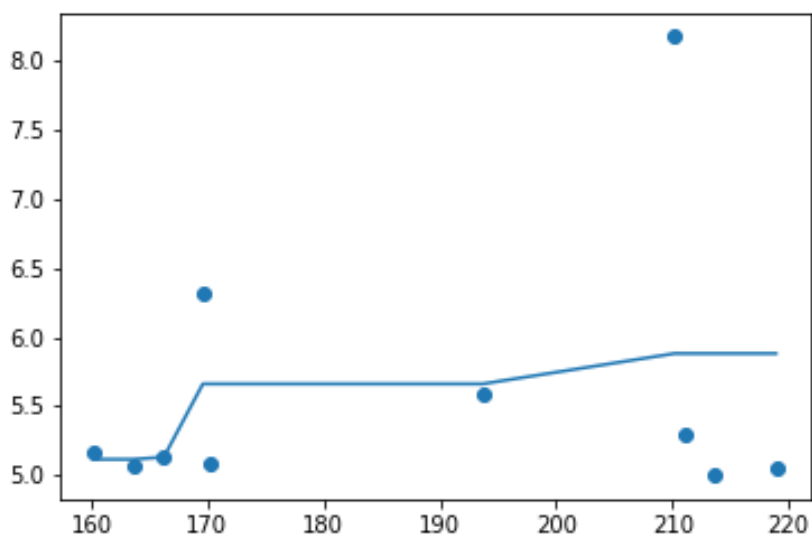


Рисунок 3.13 - Ізотонічна регресія для CO

Оцінка $R^2 = 0.11472749286959838$.

Для O_3 ізотонічна регресійна модель матиме вигляд зображений на рисунку 3.14.

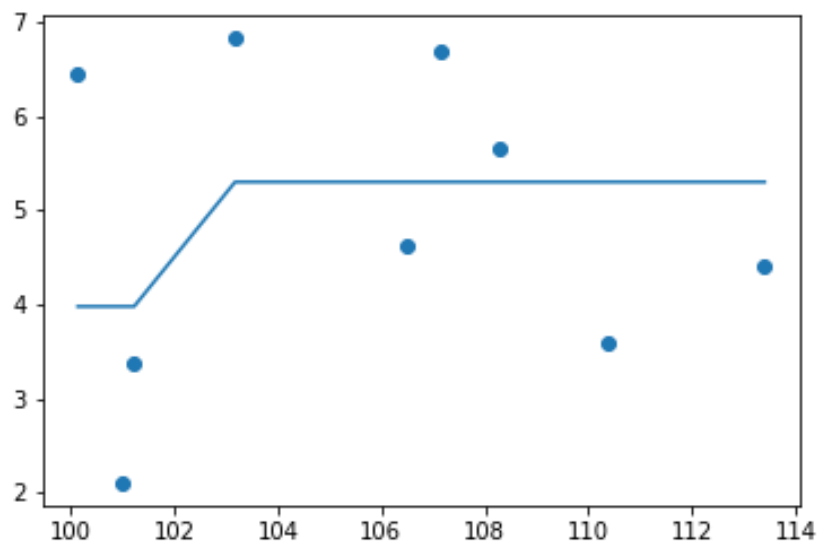


Рисунок 3.14 - Ізотонічна регресія для O_3

Оцінка $R^2 = 0.15827165702218537$.

Для NO_2 ізотонічна регресійна модель матиме вигляд зображений на рисунку 3.15.

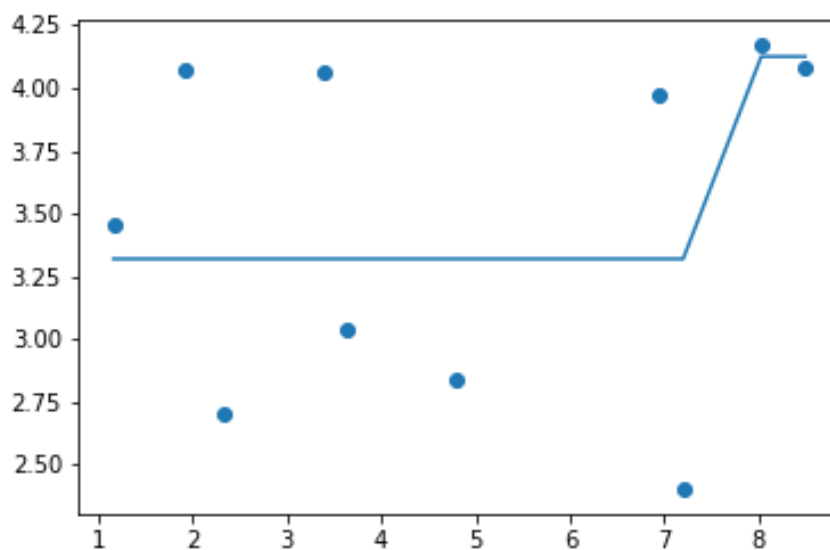


Рисунок 3.15 - Ізотонічна регресія для NO_2

Оцінка $R^2 = 0.2509686069248477$.

Таким чином, побудувавши поліноміальні, лінійні та ізотонічні моделі для прогнозування концентрацій забруднювачів повітря: PM_{10} , $PM_{2.5}$, CO, O_3 , NO_2 та порівнявши їх оцінки R^2 , можна зробити висновок, що найбільш точний прогноз даних наближених до реально виміряних дають саме поліноміальні моделі. Виявилось, що лінійні моделі погано описують вхідні дані, а ізотонічні роблять це набагато краще за лінійні, але гірше, ніж поліноміальні.

Висновки до розділу 3

Для роботи та аналізу даних про навколишнє середовище застосовувались методи регресійного аналізу. Побудувавши регресійні моделі трьох типів, а саме поліноміальні, лінійні та ізотонічні їх було порівняно за тим як добре вони описують вхідні дані. Найкращими з них виявились саме поліноміальні, вони найточніше описують вхідні дані на відміну від інших. Це свідчить про те, що залежність даних з CAMS від наземних вимірювань концентрацій забруднювачів повітря є нелінійною.

ВИСНОВКИ

Якість повітря має велике значення для людей та навколишнього середовища, за цим потрібно слідкувати. Сучасні методики аналізу допоможуть у визначенні рівня якості повітря, для інформування населення та державних органів. Багато вітчизняних і зарубіжних науковців ведуть дослідження в цій галузі за допомогою регресійного аналізу, використовуючи, в якості незалежної змінної, різноманітні показники та дані з навколишнього середовища, вимірювальних апаратів, онлайн сервісів в якості незалежних змінних, проте для цього дослідження обрано дані з CAMS. Велика міжнародна організація Copernicus, яка підтримує CAMS, вже веде спостереження за якістю повітря у багатьох місцях землі. Проте вона надає дані про концентрації забруднювачів повітря на рівні землі, розраховані за допомогою власних математичних моделей, а не виміряні.

У регресійному аналізі проводиться багато однотипних розрахунків, які зручно алгоритмізувати. Сучасне програмне забезпечення таке як мова програмування Python та різноманітні бібліотеки до неї значно допомагають в цьому.

Було побудовано різні регресійні моделі концентрацій забруднювачів повітря для України на основі точкових наземних вимірів та даних з CAMS, а саме: лінійні, поліноміальні, ізотонічні, для різних забруднювачів повітря: PM_{10} , $PM_{2.5}$, CO, O_3 , NO_2 . Якість цих моделей було порівняно за коефіцієнтом детермінації R^2 . Виявилось, що найкращі результати показують саме поліноміальні моделі різних ступенів, а саме: для PM_{10} та $PM_{2.5}$ - четвертого, з R^2 , що дорівнює 0.788 та 0.829 відповідно, для CO - шостого, з $R^2 = 0.611$, для O_3 - п'ятого, з $R^2 = 0.58$, для NO_2 - восьмого, з $R^2 = 0.929$. Ці ступені поліномів було обрано тому, що при їх збільшенні R^2 зростав мало, а моделі ускладнювались. Лінійні регресійні моделі показали доволі низькі результати. Це свідчить про нелінійний характер зв'язку між даними. Ізотонічні моделі показали результати, кращі ніж лінійні, але гірші ніж поліноміальні.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАНЬ

1. Rybarczyk Y., Zalakeviciute R. Regression Models to Predict Air Pollution from Affordable Data Collection [Електронний ресурс] / Y. Rybarczyk, R. Zalakeviciute – Режим доступа: <https://www.intechopen.com/books/machine-learning-advanced-techniques-and-emerging-applications/regression-models-to-predict-air-pollution-from-affordable-data-collections#B6> - 3.06.2018 р.
2. Slini Th., Karatzas K., Papadopoulos A. Regression analysis and urban air quality forecasting: an application for the city of athens [Електронний ресурс] / Th. Slini, K. Karatzas, A. Papadopoulos – Режим доступа: <https://journal.gnest.org/sites/default/files/Journal%20Papers/karatzas.pdf> - 3.06.2018 р.
3. Moustiris K.P., Nastos P.T., Larissi I.K., Paliatsos A.G. Application of Multiple Linear Regression Models and Artificial Neural Networks on the Surface Ozone Forecast in the Greater Athens Area, Greece [Електронний ресурс] / K.P. Moustiris., P.T. Nastos, I.K. Larissi, A.G. Paliatsos – Режим доступа: <https://www.hindawi.com/journals/amete/2012/894714/> - 3.06.2018 р.
4. Mang Lin, Jun Tao, Chuen-Yu Chan¹, Jun-Ji Cao, Zhi-Sheng Zhang, Li-Hua Zhu, Ren-Jian Zhang. Regression Analyses between Recent Air Quality and Visibility Changes in Megacities at Four Haze Regions in China [Електронний ресурс] / Mang Lin, Jun Tao, Chuen-Yu Chan¹, Jun-Ji Cao, Zhi-Sheng Zhang, Li-Hua Zhu, Ren-Jian Zhang – Режим доступа: http://www.aaqr.org/files/article/1035/2_AAQR-11-11-OA-0220_1049-1061.pdf - 3.06.2018 р.
5. Путренко В.В., Назаренко С.Ю. Визначення якості повітря на основі інтелектуального аналізу даних дистанційного зондування [Текст] / В.В. Путренко, С.Ю. Назаренко // Математичне моделювання в економіці – 2016. – №3-4 – с.176-187.
6. Садыхов А.Б., Гасанов Р.Э., Исмаилова А.М. Методы прогноза загрязнения атмосферы на основе линейной регрессии [Електронний ресурс] /

А.Б. Садыхов, Р.Э. Гасанов, А.М. Исмаилова – Режим доступа: https://www.researchgate.net/publication/312968157_Metody_proгноза_zagraznenia_atmosfery_na_osnove_linejnoj_regressii - 3.06.2018 р.

7. Шелестов А.Ю., Куссуль Н.М., Скакун С.В., Міронов А.І., Остапенко В.А., Яйлимов Б.Я. Верифікація глобальних продуктів на основі наземних досліджень в рамках проекту JESAM [Електронний ресурс] / А.Ю. Шелестов, Н.М. Куссуль, С.В. Скакун, А.І. Міронов, В.А. Остапенко, Б.Я. Яйлимов – Режим доступа: <http://geoss-conf-2014.ikd.kiev.ua/wp-content/uploads/2014/06/98.pdf> - 3.06.2018 р.

8. N. Kussul, S. Skakun, A. Shelestov, O. Kravchenko, J.F. Gallego, O. Kussul. Crop area estimation in Ukraine using satellite data within the MARS project 2012 IEEE International Geoscience and Remote Sensing Symposium - 22-27 July, 2012, - pp. 3756-3759.

9. Khaled Ahmad Ali Abdulla Al Koas GIS-based mapping and statistical analysis of air pollution and mortality in Brisbane, Australia / Khaled Ahmad Ali Abdulla Al Koas – the Queensland University of Technology – April 2010 – 67 p.

10. Richa Rai Gaseous air pollutants: a review on current and future trends of emissions and impact on agriculture / Richa Rai, Madhu Rajput, Madhoolika Agrawal and S.B. Agrawal - Journal of Scientific Research - BanarasHinduUniversity, Varanasi – pp. 77–102.

11. Wang S.W. Growth in NO_x emissions from power plants in China: bottom-up estimates and satellite observations / S.W. Wang, Q. Zhang, D. G. Streets, K. B. He, R. V. Martin, L. N. Lamsal, D. Chen, Y. Lei, and Z. Lu – Published in Atmos. Chem. Phys. Discuss.: 2 January 2012 – pp. 4429–4447.

12. Atmosphere monitoring services [Електронний ресурс] / Режим доступа:<https://atmosphere.copernicus.eu/> - 3.06.2018 р.

13. Дрейпер Н., Смит Г. Прикладной регрессионный анализ, 2-е издание. – Пер. с англ. – М.: Финансы и статистика, 1986. – 577 с.

14. Лутц М. Изучаем Python, 4-е издание. – Пер. с англ. – СПб.: Символ-Плюс, 2011. – 1280 с.