

ВИЗНАЧЕННЯ СЕРЕДНЬОРІЧНОГО ТРЕНДУ ПОКАЗНИКІВ ЯКОСТІ АТМОСФЕРНОГО ПОВІТРЯ РЕГІОНУ ЗА ДАНИМИ ГРОМАДСЬКОГО МОНІТОРИНГУ

¹Вінницький національний технічний університет

Анотація

Запропоновано алгоритм визначення середньорічного тренду показників якості атмосферного повітря регіону за даними громадського моніторингу, з урахуванням фільтрування станцій з аномальними даними, окремих аномалій в даних, заповнення пропущених по сусідніх значеннях заданої станції і по сусідніх станціях, з подальшим агрегуванням. Продемонстровано ефективність застосування алгоритму на прикладі міста Вінниці за даними 2019-2025 рр. з «Кабінету дослідника якості повітря України» Української мережі громадського моніторингу якості повітря «Eco City».

Ключові слова: прогнозування часових рядів, громадський моніторинг якості атмосферного повітря, місто Вінниця, EcoCity, розвідувальний аналіз даних, інтелектуальне передоброблення даних.

Abstract

An algorithm for determining the average annual trend of air quality indicators in a region based on public monitoring data is proposed. The algorithm takes into account the filtering of stations with anomalous data, the detection of individual anomalies in the dataset, the imputation of missing values using neighboring values of the given station and nearby stations, followed by aggregation. The effectiveness of the algorithm is demonstrated using the example of Vinnytsia city based on data from 2019–2025, obtained from the "Cabinet of Researchers on Air Quality of Ukraine" of the Ukrainian Public Air Quality Monitoring Network "Eco City."

Keywords: time series forecasting, public air quality monitoring, Vinnytsia city, EcoCity, exploratory data analysis, intelligent data preprocessing.

Вступ

В задачах середньо- і довгострокового прогнозування якості атмосферного повітря регіону головне правильно передбачати тренд, який, як правило, є основною складовою. Крім того, під час коротко- і середньострокового прогнозування, навіть за наявності суттєвих аномалій та гетероскедастичного характеру часового ряду, треба точно знати яким мав би бути тренд, щоб правильно виділяти ті аномалії та моделювати їх динаміку з використанням спеціальних моделей [1, 2].

В Україні широко сформувалась добре розгалужена мережа громадського моніторингу атмосферного повітря, доступ до даних якої є через портали «ЕкоЗагроза» Міндовкілля [3], «Eco City» [4] та SaveEcoBot [5]. Причому, мережа «Eco City» для зареєстрованих користувачів надає безоплатний доступ до гарно підготовлених рядів даних (без частково зіпсованих даних, коли зникає зв'язок чи світло під час їх збереження, що потім унеможливило їх автоматизоване оброблення) по усіх показниках якості атмосферного повітря із заданою періодичністю [6]. Саме такий доступ мають автори цієї роботи.

На жаль, є чимало проблем. Вимірювальні прилади атмосферного повітря не є повіреними, більше того, їх ніхто не обслуговує, хоча мережа «Eco City» (ГО “Фрі Ардуіно”, м. Івано-Франківськ) надає і радить таку послугу (періодичне діагностування та ремонт пристроїв), але власники станцій (як правило, звичайні громадяни) не квапляться нею користуватись. Крім того, бувають випадки, коли станція встановлена у місці, де часто паркується чи проїждить стара вантажівка або якісь люди влаштовують стихійне місце для паління, в порушення усіх правил, тоді такі дані не є релевантними

ситуації в цілому по регіону. А ще в багатьох станціях, з різних причин, може зникати світло та/чи перериватись зв'язок. Однак, завдяки розгалуженості мережі, можна виявляти основний тренд та ігнорувати дані зі старих станцій чи станцій зі значними аномаліями та доповнювати пропуски значень.

Мета дослідження – розробити алгоритм визначення середньорічного тренду показників якості атмосферного повітря регіону за даними громадського моніторингу та випробувати його на прикладі міста Вінниці.

Алгоритм розв'язання задачі

Для розв'язання поставленої задачі пропонуємо такий алгоритм:

1. Завантажити ряди даних з усіх станцій моніторингу за задану кількість років (чи за весь доступний період часу) із заданою періодичністю одним файлом з Кабінету дослідника «Есо Сіті» (або з порталу SaveEcoBot – по кожній станції це треба робити окремо і потім ще самому перевіряти на відсутність недозаповнених даних, коли кортежи (рядки таблиці) зберігаються частково, крім того, по багатьох станціях такі ряди, з незрозумілих причин, відсутні). Для визначення середньорічного тренду доцільно брати до уваги середньодобові дані, оскільки середні за хвилину чи годину будуть надто зашумленими.

2. Перевірити розмірність усіх показників. «Есо Сіті», як і SaveEcoBot і самі мають свої мережі спостережень, і агрегують дані з інших мереж, наприклад «ЛІУН Місто» (див. список на порталі [5] - <https://www.saveecobot.com/platforms>). Але аналіз показав, що, наприклад атмосферний тиск, на одних станціях вимірюється у «мм.рт.ст.», а на інших – в кПа. А отже, їх не можна просто додавати. Треба або якось перераховувати, або брати до уваги ту розмірність, по якій є більше рядів даних.

3. За заданий період згенерувати послідовність усіх кроків з регулярним інтервалом і на цю послідовність накласти усі зібрані ряди станцій. Це дозволить вибрати показники та піддіапазони часу, найбільш забезпечені даними.

4. Проаналізувати місця розташування станцій та тривалість їх функціонування. Можливо, деякі не варто враховувати, оскільки вони розташовані біля об'єктів з аномальним забруднювальним впливом, або вони є старшими за 2 роки і не проходили діагностування, що робить їхні дані малодостовірними, передусім, щодо концентрації пилу (показники PM2.5, PM10) та забруднювальних речовин у повітрі. Прибрати такі пости з вибірки.

5. Проаналізувати кількість пропущених даних – якщо їх не забагато, тоді можна спробувати застосувати імпутинг, тобто інтерполювання по сусідніх значеннях цього показника, або – усунення пропущених з використання інтелектуальної моделі, коли виявляється зв'язок між рядами різних станцій і тоді пропущені на одних вираховуються по значеннях інших [8].

6. Провести системний аналіз аномалій рядів спостережень та відфільтрувати їх [9] – це важливо для виділення саме тренду ряду, а не для аналізу розкиду значень у регіоні.

7. Усереднити дані по усіх станціях по кожній даті року. Важливо, щоб у вибірці були дані за 29 лютого, інакше, їх треба інтерполювати (виконати імпутинг).

8. Проаналізувати отриманий середньорічний ряд на пропущені, аномалії, наявність періодичності. За результатами аналізу повторити якісь з етапів цього алгоритму.

Отриманий у такий спосіб ряд буде гарним джерелом інформації і для виділення сезонних складових для прогнозування даних, і для виділення аномалій та моделювання їх динаміки, і для створення текстового датасету для донавчання великої мовної моделі для створення чат-боту про стан атмосферного повітря в регіоні.

Приклад застосування алгоритму

Випробуємо запропонований алгоритм на прикладі визначення середньорічного тренду за 2019-2025 рр. (5 повних років 2020-2024 р., оскільки перша станція «Есо Сіті» № 194 з'явилась у Вінниці у вересні 2019 р. у ТОВ «ІННОВІННПРОМ», судячи з рядів даних по ній). Використаємо дані сервісу «Кабінет дослідника» [6], до якого мають доступ автори, завдяки угоді між «Есо Сіті» і ВНТУ. На рис. 1 відображено мапу усіх станцій, дані по яких є у сервісі [6], а на рис. 2 – дані з цих станцій за показником «PM2.5», який, як правило, вимірюють усі станції громадського моніторингу повітря. На рис. 1 деякі номери зливаються, тому в описі до рисунку наведено коментар для станцій, координати яких збігаються або стосуються однієї організації. Чимало станцій є на об'єктах Вінницької міськради, але вони рознесені у просторі.

Як видно з рис. 1 та 2, усього рядів даних – 23. Але № 194 = № 400 = № 848 (хоча координати – інші, оскільки офіс переїхав), № 1427 = № 1488, тому самих станцій як приладів – 20, а не 23. Аналіз повних рядів щодобових даних показав, що багато з них в даний час не функціонують, деякі (старші 2 років) демонструють аномальний характер (№№337, 400, 767, 848, 1315, 1427, 1864), що свідчить або про необхідність діагностування, або про наявність біля них аномально потужних джерел забруднення.

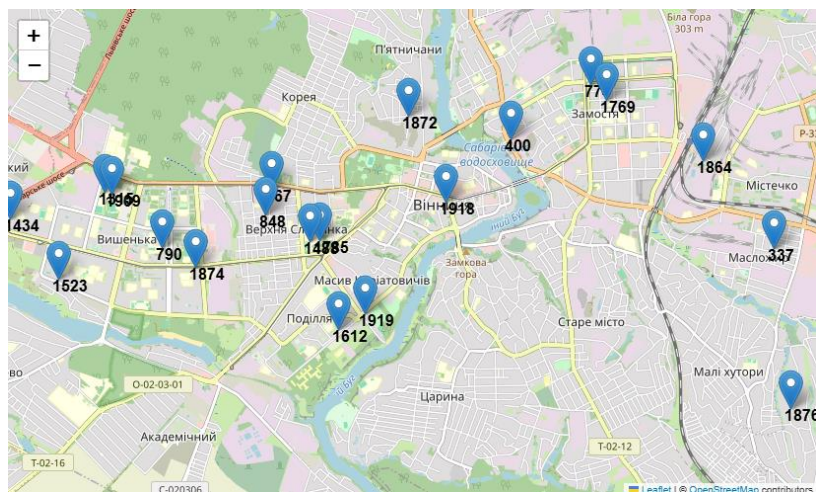


Рисунок 1. Станції громадського моніторингу міста Вінниці, які працювали чи працюють протягом 2019-2025 рр. (на території Вінницького національного технічного університету – 2 станції, але 4-річна станція № 1315 (2021-2025 рр.) демонтована і відправляється на діагностику, № 1909 (2024-2025 рр.) функціонує і встановлюється ще одна нова, на території ВНМУ ім. М.І.Пирогова – 3 станції, з яких № 1427 працювала у 2022-2023 рр, № 1488 (2023-2024 рр.), № 1785 (2024-2025 рр.), у ТОВ «ІННОВІНПРОМ» - 1 станція, яка раніше мала номер № 194, потім там же - № 400, а після їх переїзду отримала № 848; станція № 194 збігається з місцем станції № 400, а № 1427 – з № 1488, тому на мапі не показані)

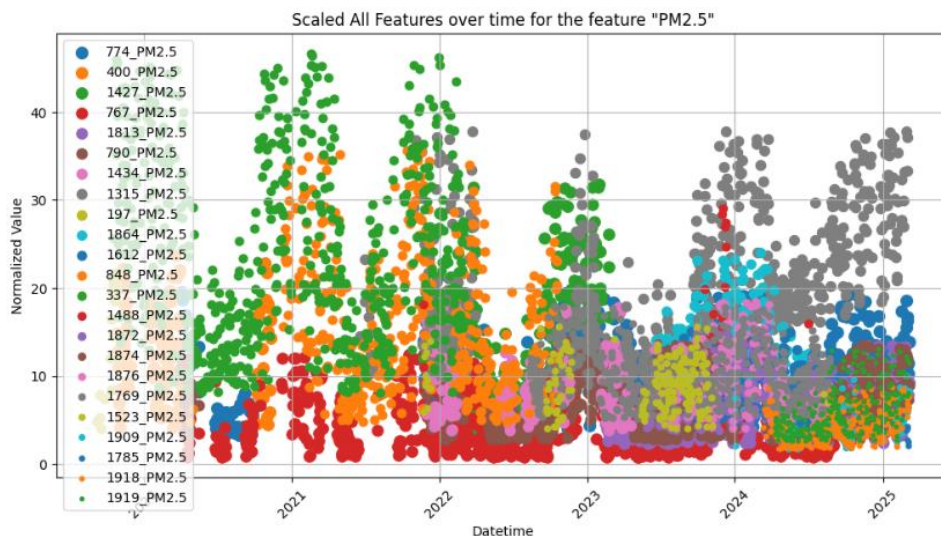


Рисунок 2. Усі ряди даних показника «PM2.5» по усіх станціях м. Вінниці за всю їх історію спостережень

Аналіз мапи показує, що біля станцій з великими показниками дійсно розташовані одні з найбільших транспортних шляхів чи розв'язок у місті, тому великі значення можуть бути виправданими, але доцільно регулярно станцію відправляти на діагностику, як це зараз робиться у ВНТУ зі станцією № 1315 (вона є однією з найбільш інформативних в Україні, маючи 15 датчиків, деякі дублюються для підвищення достовірності спостережень).

До даних на рис. 2 застосовано запропонований алгоритм з використанням програми на Python на базі платформи Kaggle: прибрано станції з аномальними даними (рис. 3а), але це призвело до втрати 2-х років, де були тільки такі станції, тоді обчислено ще варіант з цими станціями (рис. 3б); здійснено імпутинг та усунення пропущених з використанням інтелектуальної моделі RandomForestRegressor [9]; відфільтровано аномальні дані, які перевищують 90-й квантиль і є меншими за 10-й квантиль; усереднено дані по кожній даті (у т.ч. за 29.02.2020 р.) по усіх станціях (рис. 4).

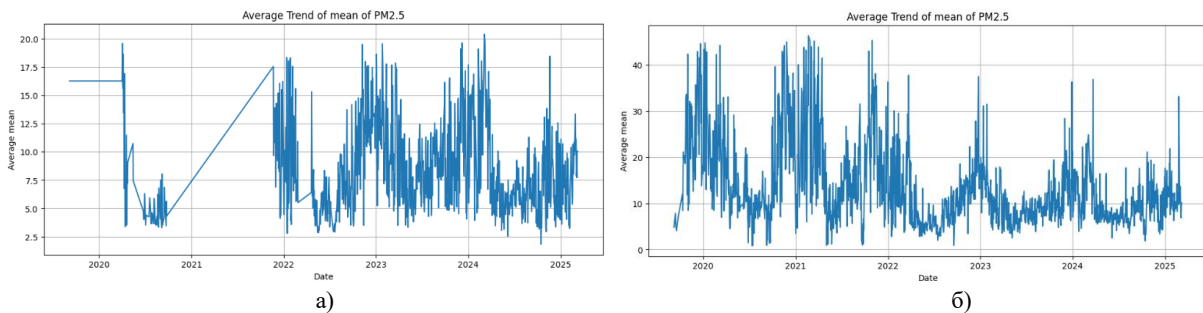


Рисунок 3. Значення усереднених по усіх станціях м. Вінниці даних показника «PM2.5» за всю історію їх спостережень: а) з фільтруванням усіх даних станцій з аномальними даними; б) без фільтрування

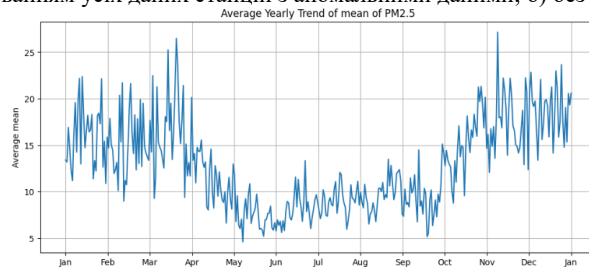


Рисунок 4. Значення усереднених по усіх станціях м. Вінниці даних показника «PM2.5» за всю історію їх спостережень

Висновки

Робота присвячена визначенню середньорічного тренду показників якості атмосферного повітря регіону за даними громадського моніторингу. Запропоновано алгоритм визначення цього тренду, з урахуванням фільтрування станцій з аномальними даними, окремих аномалій в даних, заповнення пропущених по сусідніх значеннях заданої станції і по сусідніх станціях, з подальшим агрегуванням.

Продемонстровано працездатність алгоритму на прикладі міста Вінниці за даними 2019-2025 рр. з «Кабінету дослідника якості повітря України» Української мережі громадського моніторингу якості повітря «Еко Сіті». Побудовано середньорічний тренд показника «PM2.5» міста, який продемонстрував ефективність запропонованого алгоритму.

Отриманий у такий спосіб ряд середньорічного тренду заданого показника якості атмосферного повітря буде гарним джерелом інформації для виділення сезонних складових для прогнозування даних, для виділення аномалій та моделювання їх динаміки, для створення датасету для донавчання великої мовної моделі для створення чат-боту про стан атмосферного повітря в регіоні тощо.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Копняк В. Є., Мокін В. Б., Жуков С. О., Варчук І. В., Скринник Т. В., Метод бустингу гетероскедастичних моделей для прогнозування концентрацій пилу Сахари в атмосферному повітрі України, *Наукові праці ВНТУ* [Електронний ресурс]. Вип. 2, Лип 2024. Режим доступу: [doi https://doi.org/10.31649/2307-5376-2024-2-28-38](https://doi.org/10.31649/2307-5376-2024-2-28-38)
2. Копняк В. Є., Мокін В. Б. Дослідження проблем із гетероскедастичністю даних моніторингу якості атмосферного повітря // Матеріали ЛП Науково-технічної конференції факультету інтелектуальних інформаційних технологій та автоматизації Вінницького національного технічного університету, Вінниця, 21 – 23 червня 2023 р. – Електрон. текст. дані. – 2023. – Режим доступу: <https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2023/paper/view/18962/15728>
3. Еко-Загроза. Офіційний ресурс Міністерства захисту довкілля та природних ресурсів України, 2025 [Електронний ресурс] – Режим доступу до ресурсу: https://ecozagroza.gov.ua/map?layer=air_pollution
4. «Еко Сіті» Громадський моніторинг стану якості повітря, 2025 [Електронний ресурс] – Режим доступу до ресурсу: <https://reborn.eco-city.org.ua/>
5. SaveEcoBot. Громадський моніторинг стану довкілля, 2025 [Електронний ресурс] – Режим доступу до ресурсу: <https://www.saveecobot.com/maps>
6. Інформаційний сервіс «Кабінет дослідника якості повітря України» Української мережі громадського моніторингу якості повітря «Еко Сіті», 2025 [Електронний ресурс] – Режим доступу до ресурсу: <https://archive.eco-city.org.ua/>
7. Mokin Vitalii, Shmundiak Dmytro, Kopniak Volodymyr. Air Quality Monitoring from EcoCity, Kaggle Dataset, Mar 2025 [Електронний ресурс] – Режим доступу: <https://www.kaggle.com/datasets/vbmkokin/air-quality-monitoring-from-ecocity>
8. Наука про дані: машинне навчання та інтелектуальний аналіз даних : електронний навчальний посібник комбінованого (локального та мережевого) використання [Електронний ресурс] / В. Б. Мокін, М. В. Дратований – Вінниця : ВНТУ, 2024. – 258 с. – Режим доступу: <https://docs.vntu.edu.ua/card.php?id=8163>
9. Системний аналіз стану природних середовищ з урахуванням аномалій / Шмундяк Д. О., Мокін В. Б. // *Наукові праці ВНТУ*, вип. 4, Груд. 2024. <https://doi.org/10.31649/2307-5376-2024-4-63-73>

Копняк Володимир Євгенович — аспірант кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, vkopnyak@gmail.com

Мокін Віталій Борисович — д-р техн. наук, проф., завідувач кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, e-mail: vbmokin@vntu.edu.ua

Варчук Ілона Вячеславівна — канд. техн. наук, доцент кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, e-mail: ilona.varchuk@gmail.com

Kopniak Volodymyr Y. — Postgraduate student of the Department of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, vkopnyak@gmail.com

Mokin Vitalii B. — Dr. Tech. Sciences, Prof., Head of the Department of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, e-mail: vbmokin@vntu.edu.ua

Varchuk Ilona V. — Cand. Sc. (Eng), Department of Systems Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, e-mail: ilona.varchuk@gmail.com