

Вінницький національний технічний університет
Міністерство освіти і науки України

Кваліфікаційна наукова
праця на правах рукопису

ШМУНДЯК ДМИТРО ОЛЕКСАНДРОВИЧ

УДК 303.732.4+004.9:502/504

ДИСЕРТАЦІЯ
СИСТЕМНИЙ АНАЛІЗ СТАНУ ПРИРОДНИХ СЕРЕДОВИЩ З
УРАХУВАННЯМ АНОМАЛІЙ

Спеціальність 124 – «Системний аналіз»

Галузь знань 12 – «Інформаційні технології»

Подається на здобуття ступеня доктора філософії

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

_____ Д. О. Шмундяк

Науковий керівник: Мокін Віталій Борисович,
доктор технічних наук, професор

Вінниця – 2025

АНОТАЦІЯ

Шмундяк Д.О. Системний аналіз стану природних середовищ з урахуванням аномалій – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 124 «Системний аналіз». – Вінницький національний технічний університет, Вінниця, 2025.

Метою дисертаційного дослідження є підвищення точності прогнозування стану природних середовищ з урахуванням аномалій шляхом створення методів та технології системного аналізу.

Дисертація присвячена вирішенню актуальної науково-технічної задачі підвищення точності аналізу та прогнозування даних стану природного середовищ з урахуванням аномальних значень в системах управління їх станом за рахунок використання системної методології.

У вступі даної роботи наведено: актуальність та обґрунтування вибору теми дослідження; описано мету, об'єкт, предмет та завдання даного дослідження; наукова новизна, отримана в рамках виконання дослідження; зв'язок дисертаційної роботи з науково-дослідними програмами та темами; практична цінність результатів дослідження; апробація отриманих результатів та наукові публікації за тематикою дисертаційного дослідження.

У першому розділі проведено аналіз різних природних середовищ та систем управління їх станом. Проаналізовано основні показники стану природних середовищ та їх сучасний рівень, на прикладі України. Описано державні та громадські механізми моніторингу стану природних середовищ. Проаналізовано системи управління станом природних середовищ. Оглянуто інтелектуальні моделі аналізу та прогнозування часових рядів, а також моделі пошуку аномалій в цих рядах. Визначено подальші задачі дослідження.

У другому розділі охарактеризовано складну систему, що досліджується, з використанням методології системного аналізу. Проведено декомпозицію системи за декількома видами декомпозиції з урахуванням впливу аномалій та

визначено які види невизначеності мають місце в таких системах. Проаналізовано методи та підходи для розкриття визначених видів невизначеності, формалізовано задачі, які варто розв'язати для розкриття цих видів невизначеності. Описано удосконалення методу пошуку можливих джерел забруднення, що базується на пошуку аномалій. Запропоновано підхід формування системи критеріїв для вибору параметрів місць постів спостережень стану природного середовища, з урахуванням екстремальних значень показників та використанням методу аналізу ієрархій.

Третій розділ присвячено прогнозуванню стану природних середовищ та розробленню алгоритму їх системного аналізу. Запропоновано і описано метод ідентифікації параметрів сезонності та аномалій періодичного часового ряду на основі адаптивної декомпозиції. Удосконалено метод пошуку локальних аномалій в часових рядах, що базується на аналізі його напівхвиль. Описано пошук взаємозв'язків показників стану природного середовища на основі байєсівського моделювання. Розроблено алгоритм технології системного аналізу стану природних середовищ з урахуванням аномалій, що базується на використанні класичних та авторських методів для розкриття видів невизначеності, що мають місце в таких природних середовищах.

У четвертому розділі продемонстровано апробацію результатів дослідження та прикладне застосування розроблених методів та технології системного аналізу стану природних середовищ з урахуванням аномалій. Описано застосування кроків алгоритму технології до систем управління станом річкових вод та впровадження їх результатів у роботу басейнової ради річки Південний Буг. Наведено приклад застосування кроків алгоритму запропонованої технології для оптимізації мережі моніторингу атмосферного повітря у зоні «Вінницька» (що відповідає межах Вінницької області), результати якого стали основою Програми державного моніторингу в галузі охорони атмосферного повітря зони «Вінницька» на 2023 – 2027 роки, затвердженої Міндовкілля. Наведено приклад використання розроблених методів в рамках алергологічного дослідження, де проводилося моделювання взаємозв'язків чутливості людей до різних груп

грибкових алергенів, які є біологічними забруднювачами атмосферного повітря України. Описано впровадження технології у навчальний процес ВНТУ.

У висновках подані основні результати дисертаційного дослідження.

Основні наукові та практичні результати полягають в наступному:

1. Проведено порівняльний аналіз інтелектуальних моделей аналізу та прогнозування часових рядів: Prophet, ARIMA та LSTM. Проаналізовано переваги та недоліки кожної з цих моделей. Виходячи з даного аналізу та власного досвіду, модель Prophet була пріоритетним варіантом щодо використання в подальших дослідженнях. Крім того, проведено порівняльний аналіз відомих методів пошуку аномалій в часових рядах та засобів автоматизації використання цих методів на мові програмування Python. Було порівняно методи Isolation Forest, LOF, k-nn та SESD. В ході аналізу було встановлено, що не існує єдиного методу, що може задовільнити усі потреби щодо аналізу часових рядів, в тому числі рядів станів природних середовищ. Перспективним може бути комбінування різних методів та інтерпретація результатів обох моделей, що дозволить підвищити точність пошуку аномальних значень.

2. Охарактеризовано систему управління станом природного середовища як складну систему, проведено декомпозицію системи за різними критеріями та виділено види невизначеності, які мають місце у таких системах. Проаналізовано методи та підходи розкриття невизначеності. Визначено, що частину видів невизначеності можна вирішити класичними підходами та різними інформаційними запитами до відповідних органів. Решта ж видів невизначеності вимагають розробки та удосконалення методів.

3. Дістав подальший розвиток метод пошуку можливих джерел підвищеного антропогенного навантаження на стан природного середовища за даними мережі моніторингу стану цього природного середовища. Метод дозволяє локалізувати в часі та просторі початок забруднюючого впливу на природне середовище за рахунок кусково-лінійної апроксимації тренду. Метод базується на пошуку моментів максимальних значень приросту, тобто приросту 2-го порядку, що дозволяє виявляти моменти, коли з'явилося нове джерело забруднення. Цей

метод є корисним в ситуаціях коли існує недостовірні інформація щодо викиди забруднювальних речовин, про які звітують підприємства та установи, що здійснюють викиди забруднювачів в природне середовище. Разом із поєднанням з даними аналізу супутникових знімків або даних аерофотозйомки, дозволяє перевірити гіпотезу про те, що саме цей скид є винуватцем погіршення стану середовища.

4. Запропоновано підхід щодо формування ієрархічної системи критеріїв вибору місць розташування постів мережі моніторингу стану природного середовища. При формуванні інтегральних критеріїв використовуються складові моделі DPSIR і враховуються аномальні та пікові значення показників стану природного середовища. Використовуючи метод аналізу ієрархій, можна комплексно враховувати критерії вибору місць для розташування станцій моніторингу, що може дозволити ефективно використовувати матеріальні ресурси при проектуванні або вдосконаленні мережі моніторингу та отримувати корисну інформацію щодо стану природного середовища після впровадження цих змін.

5. Розроблено метод ідентифікації параметрів сезонності та аномалій періодичного часового ряду на основі адаптивної декомпозиції. Метод дозволяє ідентифікувати різні варіанти сезонності, що мають місце в часовому ряді стану природного середовища та враховувати їх при моделюванні даних стану природного середовища, що, в свою чергу, дозволяє побудувати адекватну модель та підвищити точність прогнозування цього ряду. Метод було перевірено на даних показника пилу PM10 зі станції моніторингу якості атмосферного повітря, розташованої у Вінницькій області, від мережі громадського моніторингу EcoCity. В ході тестування, з'ясувалось, що використання запропонованого методу дозволили покращити точність роботи оптимальної моделі на основі Prophet за метрикою R^2 – у 1,7 рази, а за метрикою MSE – у 2 рази у порівнянні з моделлю Prophet з параметрами і видами сезонності за замовчуванням.

6. Удосконалено метод ідентифікації локальних аномалій значень показників стану природного середовища за рахунок декомпозиції ряду на умовні півхвилі. Метод дозволяє розбивати часовий ряд на півхвилі та виконувати пошук

аномалій на кожному окремому відрізку ряду послідовним та паралельним комбінуванням декількох відомих методів пошуку аномалій і, тим самим ідентифікувати аномалії, що мають місце в певному проміжку часового ряду та, в цілому, підвищувати точність прогнозування цього ряду у різні інтервали часу. Метод було застосовано до часового ряду показник пилу PM2.5 за період 2019-2022 рр за даними мережі громадського моніторингу EcoCity та використано для побудови моделі Prophet. За метриками RMSE та MAE у порівнянні з моделлю Prophet з параметрами налаштування за замовчуванням вдалось підвищити точність на 11% та 8% відповідно.

7. Описано аналіз взаємозв'язків показників стану природних середовищ з використанням байєсівського моделювання. Визначено як зміни значень показника стану атмосферного повітря впливають на відповідні показники на інших станціях моніторингу. Крім того, даний підхід аналізу застосовано для моделювання взаємозв'язків чутливості людей до різних груп грибкових алергенів в рамках проведення алергологічного дослідження.

8. Удосконалено технологію системного аналізу стану природного середовища з урахуванням аномалій, що базується на використанні класичних та авторських методів для розкриття основних видів невизначеності, які мають місце в системах управління станом природного середовища різного типу та дозволяє більш точно прогнозувати стан цього середовища за рахунок використання більш достовірних даних. Технологія описується алгоритмом, що складається з 15 етапів. Дана технологія може використовуватися для комплексного аналізу системи управління станом природного середовища та підвищення точності аналізу та прогнозування даних стану цього середовища.

Результати дисертаційного дослідження опубліковані у 14 роботах, у т.ч. у 5 статтях у наукових фахових періодичних виданнях, 1 стаття в міжнародному виданні, що належить до списку видань першого квартилю реєстрів Scopus і Web of Science Core Collection, та 8 тезах доповідей на науково-практичних конференціях.

Наукова новизна дисертаційної роботи полягає в наступних положеннях:

1. Вперше розроблено метод ідентифікації параметрів сезонності та аномалій періодичного часового ряду даних про стан природного середовища на основі його адаптивної декомпозиції, що дозволяє підвищити точність прогнозування цього ряду з використанням інтелектуальної моделі Prophet.

2. Удосконалено метод ідентифікації локальних аномалій значень часового ряду показника стану природного середовища за рахунок декомпозиції цього ряду на умовні півхвилі між найменшими значеннями та об'єднання результатів послідовного та паралельного застосування різних комбінацій відомих методів автоматичного інтелектуального пошуку аномалій у цих півхвилях, що дозволяє більш точно визначити параметри усіх аномалій та, в цілому – підвищити точність прогнозування цього ряду у різних інтервалах часу.

3. Дістав подальший розвиток метод пошуку можливих джерел підвищеного антропогенного навантаження на стан природного середовища за даними мережі моніторингу цього стану з локалізацією в часі та просторі початку систематичного забруднюючого впливу на природне середовище, за рахунок пошуку моментів максимальних значень приросту з подальшим порівнянням з даними аналізу супутникових знімків або даних аерофотозйомки, що дозволяє більш точно визначити та формалізувати опис регулярної появи надмірного забруднення, на відміну епізодичних аномалій, що в цілому дозволяє підвищити точність прогнозування стану природного середовища.

4. Удосконалена технологія системного аналізу стану природного середовища, що дозволяє більш точно прогнозувати цей стан з використанням розроблених методів ідентифікації параметрів моделі, байєсівського моделювання зв'язків між даними та за рахунок удосконалення підходів і методів усунення чи мінімізації різних видів невизначеності для заданого регіону, оснований на DPSIR-декомпозиції та методі аналізу ієрархій.

Практична цінність одержаних результатів полягає в наступному:

1. Розроблено комплекс алгоритмів та програмно-інформаційних засобів, які реалізують основні елементи запропонованої технології системного аналізу стану природних середовищ з урахуванням аномалій. Більшість програм є

публічно доступними у вигляді програм-ноутбуків у Google-платформі Kaggle експертів у сфері Data Science.

2. На базі платформи Kaggle, завдяки укладеному договору між ВНТУ та проектом EcoCity, створено датасет «Air Quality Monitoring from EcoCity» з даними показників стану атмосферного повітря з станцій громадського моніторингу у Вінницькій області за 2019-2025 роки.

3. Запропонована технологія системного аналізу стану природних середовищ з урахуванням аномалій була застосована для аналізу якості поверхневих вод річки Південний Буг з відкритих даних Державного агентства водних ресурсів України. Результати були розглянуто на засіданні Басейнової ради Південного Бугу.

4. Запропонована технологія використовувалась під час розроблення Програми державного моніторингу у галузі охорони атмосферного повітря зони «Вінницька» на 2023-2027 роки, що була затверджена згідно чинного законодавства.

5. Продемонстровано ефективність застосування окремих кроків запропонованої технології під час проведення алергологічного дослідження за участі вчених з Вінницького національного медичного університету ім. М. І. Пирогова та Національного медичного університету імені О. О. Богомольця, яке передбачало моделювання взаємозв'язків чутливості людей до різних груп грибкових алергенів.

Отримані результати є корисними для:

- системних аналітиків, що займаються аналізом та прогнозуванням даних стану природного середовища в системах управління цим станом;
- навчального процесу Вінницького національного технічного університету і наукову діяльність кафедри системного аналізу та інформаційних технологій для студентів, які навчаються за освітньою програмою «Системний аналіз» рівня «доктора філософії» спеціальності 124 «Системний аналіз».

Ключові слова: системний аналіз, система підтримки прийняття рішень, нейронна мережа, штучний інтелект, інформаційні технології, природне

середовище, прогнозування часових рядів, якість атмосферного повітря, аномалії, сезонність, машинне навчання, інтелектуальний аналіз даних, моделювання, екологічне здоров'я, байєсівське моделювання.

ABSTRACT

Shmundiak D.O. System analysis of the state of natural environments considering anomalies – Qualification scientific work on the rights of manuscript.

Dissertation for obtaining the Doctor of Philosophy degree in 124 «Systems Analysis». – Vinnytsia National Technical University, Vinnytsia, 2025.

The dissertation aims to enhance the accuracy of forecasting the state of natural environments, considering anomalies, by developing methods and technologies of system analysis.

The dissertation addresses relevant scientific and technical problems, aiming to enhance the accuracy of analysis and forecasting of data on the state of natural environments by considering anomalies that take place in the system of environment management through applying system analysis methodology.

The introduction of this work presents: the relevance and justification for selecting the research topic; an outline of the study's goal, object, subject, and objectives; the scientific novelty achieved within the research; the connection of the work with scientific and various research programs; the practical significance of the research findings; validation of the obtained results; scientific publications related to the research topic.

The first chapter analyzes various natural environments and systems for managing their condition. It examines key indicators of the state of natural environments and their current levels in Ukraine, as well as state and public mechanisms for monitoring natural environments. Additionally, intellectual models for analyzing and forecasting time series are analyzed and compared, along with models for detecting anomalies in these series. Finally, further research tasks are identified.

The second chapter examines a complex system by applying system analysis methodology. The system is decomposed using various types of decomposition, and the uncertainty present in such systems is identified. Methods and approaches for reducing uncertainty have been analyzed, and the tasks that need to be solved to achieve this have been formalized. The need for improving or developing specific methods is established.

The chapter also describes the enhancement of a method for identifying potential pollution sources based on anomaly detection. Additionally, an approach for developing a system of criteria for selecting environmental monitoring post locations is described, incorporating extreme indicator values and employing the analytic hierarchy process.

The third section focuses on forecasting the state of natural environments and developing an algorithm for their systematic analysis. It introduces a method for identifying seasonality parameters and anomalies in periodic time series using adaptive decomposition. An improved method for detecting local anomalies in the time series is presented, based on the analysis of half-waves. The section also explores the discovery of relationships between environmental indicators through Bayesian modeling. Furthermore, an algorithm for a systematic analysis technology that accounts for anomalies, incorporating both classical and original methods for reducing uncertainty present in natural environments, is developed.

The fourth section presents the validation of the research results and the practical implementation of the developed methods and technology for the system analysis of the state of natural environments, with consideration of anomalies. It describes the application of the technology algorithm's steps to river water management systems, including their implementation in the activities of the Southern Bug River Basin Council. An example of how the proposed technology was applied to optimize the atmospheric air monitoring network in the Vinnytsia zone (aligned with the administrative boundaries of the Vinnytsia region) is provided. The outcomes of this application were taken as a basis for the State Monitoring Program for Atmospheric Air Protection in the Vinnytsia Zone for 2023–2027, approved by the Ministry of Environmental Protection. The section also includes an example of the use of the developed methods in an allergological study, which involved modeling the relationships between human sensitivity to various groups of fungal allergens—biological pollutants of Ukraine's atmospheric air. In addition, the implementation of the proposed technology in the educational process at Vinnytsia National Technical University is presented.

The conclusions summarize the key findings of the dissertation research.

The main scientific and practical results are as follows:

1. A comparative analysis of intelligent models for time series analysis and forecasting (Prophet, ARIMA, and LSTM) was carried out. The strengths and limitations of each model were examined. Based on the results of this analysis and practical experience, the Prophet model was selected as the preferred option for further research. Additionally, a comparative evaluation of widely used anomaly detection methods in time series, as well as tools for their automation using the Python programming language, was performed. The methods compared included Isolation Forest, LOF, k-NN, and SESD. The analysis revealed that no single method fully meets all the requirements for time series analysis, particularly in the context of natural environment data. Therefore, combining multiple methods and interpreting their results jointly appears to be a promising approach for improving the accuracy of anomaly detection.

2. The system for analyzing and managing the state of the natural environment is characterized by using the system analysis methodology. It is decomposed according to various criteria, and the types of uncertainty inherent in such systems are identified. Methods and approaches for addressing these uncertainties are analyzed. It is established that certain types of uncertainty can be resolved using classical methods and by obtaining additional information from relevant authorities. However, other types of uncertainty require the development and enhancement of specialized methods.

3. The method for identifying potential sources of increased anthropogenic impact on the state of the natural environment, based on monitoring network data, has been further developed. This method enables the localization of the onset of pollutant influence through piecewise linear trend approximation. It is based on detecting points of maximum second-order increases, which makes it possible to identify the moments when a new pollution source likely emerged. The method is particularly valuable in cases where reported data on pollutant emissions from enterprises and institutions are unreliable. When combined with satellite imagery or aerial photography analysis, it facilitates the verification of hypotheses regarding the origin of environmental degradation and supports the identification of specific discharges as potential causes.

4. An approach to developing a hierarchical system of criteria for selecting locations for monitoring network posts is presented. In forming the integral criteria, the components of the DPSIR model are utilized, with consideration of anomalous and peak values of environmental indicators. By applying the analytic hierarchy process, it is possible to comprehensively evaluate the criteria for selecting monitoring station locations. This approach facilitates the effective use of material resources in the design or improvement of the monitoring network, ensuring the acquisition of valuable information about the state of the natural environment following the implementation of these changes.

5. A method for identifying seasonal parameters and anomalies in periodic time series based on adaptive decomposition has been developed. This method enables the identification of various seasonal patterns in the time series of natural environment indicators, allowing them to be incorporated into data modeling. This, in turn, facilitates the construction of an accurate model and enhances the forecasting accuracy of the series. The method was tested on PM10 data from the air quality monitoring station in the Vinnytsia region, part of the EcoCity public monitoring network. During testing, it was found that the application of the proposed method improved the accuracy of the optimal model based on Prophet, with the R2 metric increasing by 1.7 times and the MSE metric improving by 2 times, compared to the default Prophet model with standard parameters and seasonal patterns.

6. The method for detecting local anomalies in environmental indicator values has been enhanced by decomposing the time series into conditional half-waves. This method allows the series to be split into half-waves, with anomalies identified in each segment through the sequential and parallel combination of several anomaly detection methods. As a result, anomalies occurring within specific time intervals can be identified, which in turn improves the overall accuracy of forecasting the series across different periods. The method was applied to the time series of the PM2.5 indicator data for the period of 2019–2022, provided by the EcoCity public monitoring network, and was used to build a Prophet model. The results showed that, compared to the Prophet

model with default configuration parameters, the proposed method improved accuracy by 11% according to the RMSE metric and by 8% according to the MAE metric.

7. The analysis of relationships between indicators of the state of natural environments using Bayesian modeling is described. This analysis determines how changes in atmospheric air quality indicators at one monitoring station affect corresponding indicators at other stations. Additionally, this approach is applied to model the relationships between human sensitivity to various groups of fungal allergens as part of an allergology study.

8. The technology for systematic analysis of the state of the natural environment, accounting for anomalies, has been enhanced. This technology utilizes both classical and novel methods to identify the primary types of uncertainty present in various natural environment management systems, enabling more accurate predictions of the environmental state using more reliable data. The technology is described by an algorithm consisting of 15 stages. It can be applied for a comprehensive analysis of natural environment management systems, significantly improving the accuracy of both the analysis and prediction of environmental data.

The results of the dissertation research have been published in 14 works, including 5 articles in scientific professional periodicals in specialty 124, 1 article in an international journal indexed in the Scopus and Web of Science databases, and 8 abstracts presented at scientific and practical conferences.

The scientific novelty of the dissertation lies in the following provisions:

1. For the first time, a method has been developed for identifying seasonality parameters and anomalies in a periodic time series of natural environment data, based on adaptive decomposition. This method enhances the accuracy of forecasting the series using the Prophet intelligent model.

2. The method for identifying local anomalies in the time series of natural environment indicators has been improved by decomposing the series into conditional half-waves between the smallest values. The results of sequential and parallel application of various combinations of known automated anomaly detection methods in these half-waves are then combined. This approach allows for more accurate

determination of anomaly parameters and, overall, enhances the forecasting accuracy of the series across different time intervals.

3. The method for identifying potential sources of increased anthropogenic impact on the natural environment, based on monitoring network data, has been further developed. It involves localizing the onset of systematic pollution by detecting moments of maximum growth values, followed by comparison with satellite imagery or aerial photography data. This approach allows for more precise determination and formalization of the regular occurrence of excessive pollution, distinguishing it from episodic anomalies. As a result, it enhances the accuracy of predicting the state of the natural environment.

4. The technology for systematic analysis of the state of the natural environment has been enhanced, enabling more accurate predictions of this state. This improvement is achieved through the development of methods for identifying model parameters, Bayesian modeling of relationships between data, and the refinement of approaches to eliminate or minimize various types of uncertainty for a given region. These advancements are based on DPSIR decomposition and the analytic hierarchy process.

The practical value of the results:

1. A set of algorithms and software tools has been developed to implement the core components of the proposed technology for systematic analysis of natural environments, accounting for anomalies. Most of the programs are publicly available as notebook programs on the Kaggle platform, aimed at Data Science professionals.

2. Based on the Kaggle platform and through an agreement between VNTU and the EcoCity project, the “Air Quality Monitoring from EcoCity” dataset was created. This dataset includes atmospheric air quality data from public monitoring stations in the Vinnytsia region for the period 2019-2025.

3. The proposed technology for systematic analysis of the state of natural environments, incorporating anomaly detection, was applied to analyze the quality of surface waters in the Southern Bug River, using open data from the State Agency of Water Resources of Ukraine. The results were presented at a meeting of the Southern Bug Basin Council.

4. The technology was employed during the development of the State Monitoring Program for atmospheric air protection in the Vinnytsia zone for 2023-2027, which was officially approved per current legislation.

5. The effectiveness of the technology was demonstrated during an allergological study conducted in collaboration with scientists from Vinnytsia National Medical University (M. I. Pirogov) and the National Medical University (O. O. Bogomolets). The study involved modeling the relationships between human sensitivity to various groups of fungal allergens.

The results of this work provide benefits for:

- System analysts engaged in the analysis and enhancement of environmental monitoring networks.

- The educational process at Vinnytsia National Technical University, as well as the scientific activities of the Department of System Analysis and Information Technologies. The results are integrated into the System Analysis educational program at the Doctor of Philosophy level, specialty 124 System Analysis.

Keywords: systems analysis, decision support system, neural network, artificial intelligence, information technology, natural environment, time series forecasting, air quality, anomalies, seasonality, machine learning, intelligent data analysis, modeling, environmental health, Bayesian modeling.

СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА ЗА ТЕМОЮ ДИСЕРТАЦІЇ

[1] В. Б. Мокін, О. В. Слободянюк, О. М. Давидюк, і Д. О. Шмундяк, «Інформаційна технологія пошуку можливих джерел підвищеного забруднення річки з використанням моделі Prophet», *Вісник Вінницького політехнічного інституту*, № 4, с. 15–24, (Верес. 2020). DOI: <https://doi.org/10.31649/1997-9266-2020-151-4-15-24>. (**Index Copernicus, Наукове фахове видання України, категорія «Б» зі спеціальності 124**)

Запропоновано удосконалення методу пошуку джерел підвищеного антропогенного впливу за рахунок пошуку моментів максимальних значень приросту, тобто приросту 2-го порядку.

[2] В. Б. Мокін, Є. М. Крижановський, А. Р. Ящолт, і Д. О. Шмундяк, «Технологія проєктування мережі спостережень якості атмосферного повітря регіону на основі методу аналізу ієрархій», *Наукові Праці ВНТУ*, № 4, (Лют. 2022). DOI: <https://doi.org/10.31649/2307-5376-2021-4-21-33>. (**Наукове фахове видання України, категорія «Б» зі спеціальності 124**)

Запропоновано підхід формування ієрархічних критеріїв з урахуванням екстремальних даних моніторингу стану атмосферного повітря та моделі DPSIR, автоматизовано метод аналізу ієрархій, використовуючи мову програмування Python.

[3] Д. О. Шмундяк і В. Б. Мокін, «Метод ідентифікації параметрів гармонік та аномалій періодичного часового ряду на основі адаптивної декомпозиції», *Вісник Вінницького політехнічного інституту*, № 6, с. 46–56, (Груд. 2023). DOI: <https://doi.org/10.31649/1997-9266-2023-171-6-46-56>. (**Index Copernicus, Наукове фахове видання України, категорія «Б» зі спеціальності 124**)

Розроблено метод ідентифікації параметрів сезонності за результатами адаптивної декомпозиції часового ряду. Запропоновано варіант комбінування різних параметрів сезонності при їх одночасному застосуванні при побудові моделі прогнозування.

[4] Victoria Rodinkova, Serhii Yuriev, Vitalii Mokin, Mariia Kryvopustova, Dmytro Shmundiak, Mykyta Bortnyk, Yevhenii Kryzhanovskyi, Andrii Kurchenko, «Bayesian analysis suggests independent development of sensitization to different fungal allergens», *World Allergy Organization Journal*, Volume 17, (Issue 5, 2024) . DOI: <https://doi.org/10.1016/j.waojou.2024.100908>. (**Scopus i Web of Science**)

Здійснено статистичний аналіз даних чутливості людей до алергенів та розроблено алгоритм аналізу взаємозв'язків з використанням байєсівського моделювання.

[5] Д. О. Шмундяк і В. Є. Копняк, «Метод ідентифікації локальних аномалій значень показників стану довкілля з використанням декомпозиції на півхвилі», *Вісник Вінницького політехнічного інституту*, № 1, с. 88–100, (Лют. 2024). DOI: <https://doi.org/10.31649/1997-9266-2024-172-1-88-100>. (**Index Copernicus, Наукове фахове видання України, категорія «Б» зі спеціальності 124**)

Запропоновано удосконалення методу пошуку локальних аномалій за результатами декомпозиції часового ряду на умовні півхвилі з подальшим аналізом аномалій на різних інтервалах ряду та проведено аналіз ефективності застосування ансамблів методів пошуку аномалій.

[6] Д. О. Шмундяк, В. Б. Мокін, «Системний аналіз стану природних середовищ з урахуванням аномалій», *Наукові Праці ВНТУ*, № 4, (Грудень 2024)/ DOI: <https://doi.org/10.31649/2307-5376-2024-4-63-73/>. (**Наукове фахове видання України, категорія «Б» зі спеціальності 124**)

Запропоновано алгоритм системного аналізу стану природних середовищ з урахуванням аномалій, оснований на функціональній DPSIR-декомпозиції та на використанні класичних і авторських методів для розкриття основних видів невизначеності.

[7] В. Б. Мокін, Є. М. Крижановський, А. Р. Яцолт, Д. О. Шмундяк, Б. В. Паєвський, «Оптимізація моніторингової мережі стану атмосферного повітря вінницької зони за даними 2016-2020 рр.», *матеріали LI Науково-технічної конференції факультету інтелектуальних інформаційних технологій та автоматизації*, Вінниця, 31 травня - 1 червня 2022 р. – Електрон. текст. дані. –

2022. – Режим доступу: <https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2022/paper/view/16118/13531>.

Описано процес оптимізації моніторингової мережі стану атмосферного повітря зони «Вінницька» з використанням кроків технології системного аналізу стану природних середовищ.

[8] Д. О. Шмундяк, Н. С. Іжаковська, Д. О. Литвиненко, А. О. Судець, «Аналіз можливостей Python-бібліотек щодо виявлення аномальних даних у задачі прогнозування стану атмосферного повітря», *матеріали ЛІІ Науково-технічної конференції факультету інтелектуальних інформаційних технологій та автоматизації*, 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/18957/15722>

Проведено порівняльний аналіз ефективності використання деяких Python-бібліотек для автоматизації процесу пошуку аномалій у часових рядах.

[9] Д. О. Шмундяк, «Дослідження впливу параметрів моделі Prophet на точність прогнозування якості атмосферного повітря», *матеріали ЛІІ Науково-технічної конференції факультету інтелектуальних інформаційних технологій та автоматизації*, 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/18210/15056>

Проведено дослідження ефективності роботи моделі прогнозування часових рядів Prophet в залежності від різних налаштувань сезонності.

[10] Д. О. Шмундяк, А. В. Лосенко, В. Б. Мокін, «Огляд підходів до визначення порядку Фур'є у моделі Facebook Prophet для моделювання сезонної складової часового ряду», *матеріали ЛІІ Науково-технічної конференції факультету інтелектуальних інформаційних технологій та автоматизації*, 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/17200/14329>

Запропоновано підхід визначення порядку Фур'є для налаштувань сезонності моделі прогнозування часових рядів.

[11] В. Б. Мокін, Д. О. Шмундяк, В. Є. Копняк, «Порівняльний розвідувальний аналіз даних про стан атмосферного повітря за даними мережі EcoCity з використанням бібліотеки Sweetviz», *матеріали LIII Науково-технічної конференції факультету інтелектуальних інформаційних технологій та автоматизації*, Вінниця, 20 - 22 березня 2024 р. – Електрон. текст. дані. – 2024. – Режим доступу: <https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2024/paper/view/20846/17251>

Проведено розвідувальний аналіз даних стану атмосферного повітря з використанням бібліотеки для автоматизованого розвідувального аналізу Sweetviz.

[12] Порівняльний аналіз моделей прогнозування показника стану атмосферного повітря / Д. О. Шмундяк // Матеріали XVII Міжнародної конференція «Контроль і управління в складних системах» (КУСС-2024) [Електронний ресурс], Вінниця, 16-17 жовтня 2024 року] – Режим доступу: <https://conferences.vntu.edu.ua/index.php/mccs/mccs2024/paper/view/22120>

Наведено порівняльний аналіз точності роботи моделі Prophet для прогнозування показника стану атмосферного повітря з використання авторських методів ідентифікації параметрів сезонності та аномалій.

[13] Метод пошуку можливих джерел забруднення атмосферного повітря на основі аналізу аномалій / Д. О. Шмундяк // Матеріали XXIII Міжнародної науково-практичної конференції Математичне моделювання та інформаційно-комунікаційні технології для перемоги та відновлення, Вінниця, 12-13 листопада 2024 р. – Режим доступу: https://itgip.org/wp-content/uploads/2024/11/2024-11-24_zbirka_all_07_11_2024_148x210.pdf

Наведено приклад застосування методу пошуку можливих джерел забруднення за даними мережі моніторингу до даних стану атмосферного повітря.

[14] Аналіз взаємовпливів рядів даних моніторингу стану атмосферного повітря з використанням методу байєсівського моделювання / Д. О. Шмундяк // *матеріали LIV Всеукраїнської науково-технічної конференції підрозділів Вінницького національного технічного університету, Вінниця, 24 - 27 березня 2025 р. – Електрон. текст. дані. – 2025. – Режим доступу: <https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2025/paper/view/23870/19675>.*

Запропоновано і описано алгоритм застосування Байєсівського моделювання для визначення як дані моніторингу стану атмосферного повітря залежать від змін на сусідніх станціях в межах регіону.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ ТА СКОРОЧЕНЬ.....	25
ВСТУП.....	26
РОЗДІЛ 1 АНАЛІЗ СТАНУ ПРИРОДНИХ СЕРЕДОВИЩ З УРАХУВАННЯМ АНОМАЛІЙ.....	35
1.1 Аналіз стану природних середовищ та механізмів моніторингу	35
1.1.1. Показники забруднення стану природних середовищ.....	36
1.1.2. Інтегральні показники стану природних середовищ	38
1.1.3. Сучасний рівень забруднення природних середовищ в Україні	40
1.1.4. Державні механізми моніторингу та контролю стану природних середовищ.....	44
1.1.5. Громадські проекти моніторингу стану природних середовищ	47
1.2. Аналіз систем управління станом природного середовища.....	52
1.3. Огляд інтелектуальних моделей прогнозування часових рядів.....	54
1.4. Аналіз методів та засобів пошуку аномалій в часових рядах	61
1.4.1. Аналіз відомих методів пошуку аномалій	61
1.4.2. Огляд засобів для автоматизації пошуку аномалій в часових рядах.....	68
1.5. Висновки до розділу та постановка задач дослідження.....	72
РОЗДІЛ 2 СИСТЕМНИЙ АНАЛІЗ СИСТЕМИ УПРАВЛІННЯ СТАНОМ ПРИРОДНИХ СЕРЕДОВИЩ	74
2.1. Характеристика складної системи.....	74
2.2.1. Декомпозиція складної системи за різними критеріями	76
2.2.2. Класифікація видів невизначеності у складній системі	76
2.2. Підходи та методи щодо розкриття видів невизначеності у складній системі	79
2.3. Удосконалення методу пошуку можливих джерел забруднення на основі аналізу аномалій.....	85
2.4. Усунення цільової невизначеності щодо системи моніторингу стану природного середовища.....	98

2.5. Висновки до розділу	108
РОЗДІЛ 3 ПРОГНОЗУВАННЯ СТАНУ ПРИРОДНИХ СЕРЕДОВИЩ ТА РОЗРОБЛЕННЯ АЛГОРИТМУ ЇХ СИСТЕМНОГО АНАЛІЗУ З УРАХУВАННЯМ ПІКОВИХ ЗНАЧЕНЬ	110
3.1. Розроблення методу ідентифікації параметрів сезонності часового ряду	110
3.1.1. Дослідження впливу параметрів моделі Prophet на точність прогнозування стану природного середовища.....	111
3.1.2. Ідентифікація порядку ряду Фур'є часового ряду показника стану природного середовища.....	114
3.1.3. Розроблення нового методу	116
3.2. Удосконалення методу ідентифікації локальних аномалій в часових рядах	126
3.3. Пошук взаємозв'язків показників стану природного середовища на основі байєсівського моделювання	138
3.4. Розроблення алгоритму технології системного аналізу стану природного середовища.....	144
3.5. Висновки до розділу	148
РОЗДІЛ 4 ПРИКЛАДНЕ ЗАСТОСУВАННЯ РОЗРОБЛЕНИХ МЕТОДІВ ТА ТЕХНОЛОГІЇ СИСТЕМНОГО АНАЛІЗУ ПРИРОДНИХ СЕРЕДОВИЩ З УРАХУВАННЯМ АНОМАЛІЙ.....	150
4.1. Застосування методів та технології системного аналізу до системи управління стану річкових вод	150
4.2. Застосування методів та технології системного аналізу до системи управління стану атмосферного повітря.....	153
4.3. Застосування методів та технології системного аналізу до моделювання поширення алергенних речовин	157
4.4. Впровадження в навчальний процес	159
4.5. Висновки до розділу	160
ВИСНОВКИ.....	162
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	167

ДОДАТКИ.....	178
ДОДАТОК А СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА ЗА ТЕМОЮ ДИСЕРТАЦІЇ.....	179
ДОДАТОК Б ПОКАЗНИКИ ЗАБРУДНЕННЯ СТАНУ ПРИРОДНИХ СЕРЕДОВИЩ.....	182
ДОДАТОК В РОЗВІДУВАЛЬНИЙ АНАЛІЗ ДАНИХ.....	186
ДОДАТОК Г РОЗКРИТТЯ ВИДІВ НЕВИЗНАЧЕНОСТІ	192
ДОДАТОК Д АКТ ПРО ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ У ПРОГРАМІ ДЕРЖАВНОГО МОНІТОРИНГУ У ГАЛУЗІ ОХОРОНИ АТМОСФЕРНОГО ПОВІТРЯ ЗОНИ «ВІННИЦЬКА».....	193
ДОДАТОК Е АКТ ПРО ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ У НАВЧАЛЬНИЙ ПРОЦЕС ВНТУ	194
ДОДАТОК Є СВДОЦТВО ПРО РЕЄСТРАЦІЮ АВТОРСЬКОГО ПРАВА НА ТВІР	195

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ ТА СКОРОЧЕНЬ

ВНТУ – Вінницький національний технічний університет

Кафедра САІТ – кафедра системного аналізу та інформаційних технологій

ВНТУ

Prophet – модель для моделювання і прогнозування часових рядів

AQI – Air Quality Index

ARIMA – Autoregressive Integrated Moving Average

DAG – Directed Acyclic Graph

ЕЕА – European Environment Agency

ГДК – Гранично допустима концентрація

ЄС – Європейський союз

КМУ – Кабінет Міністрів України

МОЗ – Міністерство охорони здоров'я України

БСК – Біохімічне споживання кисню

БПЛА – Безпілотний літальний апарат

PM – Particulate matter

ВСТУП

Обґрунтування вибору теми дослідження. Стан природного середовища, прикладами якого є атмосферне повітря, поверхневі води тощо, має прямий і досить суттєвий вплив не тільки на життя та здоров'я людей, але й на усі інші живі організми, що проживають на Землі. Протягом останніх десятиліть у світі в цілому та в Україні зокрема з багатьох причин спостерігається значне погіршення якості довкілля, через все більше забруднення та інші види негативного впливу на нього та на здоров'я населення, особливо на його найбільш вразливі верстви: дітей, літніх людей, людей з хронічними захворюваннями тощо. Велика кількість існуючих захворювань, серед яких захворювання серця, респіраторні та онкологічні захворювання, спричинені, в першу чергу, поганим станом навколишнього середовища та побічними ефектами цього стану. Забруднення атмосферного повітря, щорічно, є причиною мільйонів людських смертей.

Забруднення повітря та інших природних середовищ відбувається як природним шляхом, так і через діяльність людей. До природних джерел забруднення відносяться виверження вулканів, лісові пожежі тощо. Антропогенне ж забруднення є наслідком індустріалізації та урбанізації: хімічні та токсичні викиди об'єктів промисловості, викиди від транспортних засобів, скидання стічних і забруднених зворотних вод тощо. Для зменшення цього забруднюючого впливу необхідно здійснювати відповідне регулювання діяльності людей, моніторинг, потрібна спеціальна система управління станом довкілля та підтримка прийняття оптимальних рішень, відповідно до концепції сталого розвитку.

Система управління станом природного середовища – це складна система, складові якої взаємопов'язані і тому зміни в одній з підсистем ведуть до змін в інших складових системи [1]. Останнім часом спостерігається зростаючий вплив антропогенного характеру, що створює величезне навантаження на стан природних середовищ, через все більший рівень забруднювальних речовин, що викидаються різними підприємствами у різні природні середовища. Крім того, у

цьому десятиріччі відчувається вплив змін клімату, аномально теплі зими, аномальні засушливі періоди, аномально сильні вітри і зливи у різних регіонах планети. Це провокує появу аномалій, пошук та аналіз яких є дуже важливим для досягнення цілей сталого розвитку. Деякі захворювання та їх переносники є чутливими до клімату, а аномальні природні явища та зміна цього клімату дозволяє їм розповсюджуватися на набагато більших територія, що, в свою чергу, ставить під ризик людей, що проживають на цій території. Крім того, аномальна зміна температури може впливати на діяльність населення щодо виробництва продуктів харчування та може провокувати міграцію населення, що чинить тиск на середовище.

Виявлення та врахування впливу аномалій дозволяє підвищити адекватність моделей аналізу та прогнозування даних, а отже – це є однією з найголовніших задач в системному аналізі. Ефективність аналізу системи управління природного середовища та даних його стану дозволяє реагувати на екологічні проблеми та приймати обґрунтовані управлінські рішення. Пошук аномалій – це процес оброблення наборів даних з метою ідентифікації подій або елементів, які не відповідають певним шаблонам поведінки або іншим елементам даного набору. Машинне навчання використовується для автоматизації та, тим самим, підвищення ефективності виявлення подібних аномалій. Найбільш поширеним є пошук аномалій на нерозмічених даних, який відомий як метод виявлення аномалій без нагляду (або неконтрольовані алгоритми) [2, 3]. При використанні цього методу до уваги береться тільки внутрішня структура даних і припускається, що переважна більшість усіх наявних елементів є нормальною. Іншим методом є метод контрольованого виявлення аномалій, який в свою чергу потребує набір даних, в якому визначено нормальні або аномальні елементи та включають навчання класифікатора. Також існують методи напівконтрольованого пошуку аномалій. Вони передбачають створення моделі на основі нормальних навчальних даних, яка задовольняє очікувану поведінку даних. Потім цю модель застосовують на тестовому елементі для перевірки правдоподібності, що даних елемент був створений розробленою моделлю.

Сучасні інтелектуальні технології аналізу та передбачення часових рядів дозволяють враховувати аномалії [4, 5], але для цього їх треба спочатку знайти та оцінити їх вплив. Відповідно до методології системного аналізу треба здійснити декомпозицію складної системи, визначити види невизначеності та запропонувати методи їх розкриття [1]. Після цього варто знайти дані, застосувати до них ці методи і тоді можна ідентифікувати інтелектуальні моделі, здійснити за ними передбачення, провести інтелектуальний аналіз даних та розробити рекомендації для управління станом природного середовища. При цьому, необхідно все розглядати, враховуючи аномалії як природного, так і антропогенного походження.

Багато вчених працюють у цьому напрямку, але відсутня дійсно комплексна технологія розв'язання цієї задачі, з розкриттям усіх видів невизначеності на прикладі стану довкілля в Україні та декомпозиції на основі моделі «DPSIR» з використанням методів системного аналізу та інтелектуальних моделей передбачення та методів інтелектуального аналізу даних. Відповідно, є потреба в удосконаленні системного аналізу такого типу систем, що дозволить підвищити точність аналізу та прогнозування даних в таких системах. Це, в свою чергу, дозволить комплексно аналізувати існуючі системи управління станом природного середовища та приймати обґрунтовані рішення щодо покращення стану цих середовищ.

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконувалась у межах НДР кафедри САІТ 28-К7 «Інформаційні технології та системи моніторингу, моделювання та системного інтелектуального аналізу й оптимізації даних у складних об'єктах» протягом 2021-2023 рр. та НДР кафедри САІТ 28-К8 «Системний аналіз та створення інформаційних систем й інтелектуальних технологій процесів у складних об'єктах» у 2024 р. Крім того, у листопаді-грудні 2021 року здобувач брав участь у виконанні госпдоговірної науково-дослідної роботи на замовлення ТОВ "ІННОВІНПРОМ" (генеральний замовник – Вінницька облдержадміністрація) на тему "Розробка розділів програми державного моніторингу у галузі охорони

атмосферного повітря Вінницької області» № 2859 (номер держреєстрації 0121U114085); згідно із завданням, на Python здійснювалось визначення пріоритетності пунктів державного моніторингу якості атмосферного повітря, які проєктуються у програмі державного моніторингу, за допомогою формування ієрархічної системи критеріїв з урахуванням екстремальних значень даних моніторингу та застосування методу аналізу ієрархій.

Мета і завдання дослідження. *Метою дослідження є підвищення точності прогнозування стану природних середовищ, з урахуванням аномалій, шляхом створення методів та технології системного аналізу.*

Задачами дослідження є:

– провести аналіз стану природних середовищ і механізмів моніторингу та контролю цього стану на прикладі України, провести аналіз інтелектуальних моделей і засобів прогнозування часових рядів стану природного середовища та аналіз відомих підходів і методів пошуку аномалій у цих рядах;

– охарактеризувати складну систему управління станом природних середовищ, здійснити її декомпозицію, виділити види невизначеності, що мають місце у таких системах, і сформулювати задачі дослідження, спрямовані на розкриття цих видів невизначеності;

– систематизувати відомі чи розробити нові методи розкриття видів невизначеності складної системи;

– запропонувати технологію системного аналізу природних середовищ з урахуванням аномалій;

– випробувати розроблені методи і технологію на реальних прикладах аналізу та прогнозування стану природних середовищ, здійснити їх впровадження і апробацію.

Об'єктом дослідження є процеси аналізу та прогнозування стану природних середовищ в системах управління цим станом.

Предметом дослідження є методи та технології системного аналізу стану природних середовищ з урахуванням аномалій.

Методи дослідження містять загальнонаукову методологію проведення досліджень і принципи системного підходу: аналіз літературних джерел, відкритих даних, експериментальні дослідження, системний аналіз, аналіз даних. Крім того, в дослідженнях використовувалися метод аналізу ієрархій та методи, що базуються на нейронних мережах, машинному навчанні, алгоритмах аналізу та прогнозування часових рядів, байєсівському моделюванні, аналізі даних дистанційного зондування Землі.

Наукова новизна отриманих результатів.

1. Вперше розроблено метод ідентифікації параметрів сезонності та аномалій періодичного часового ряду даних про стан природного середовища на основі його адаптивної декомпозиції, що дозволяє підвищити точність прогнозування цього ряду з використанням інтелектуальної моделі Prophet.

2. Удосконалено метод ідентифікації локальних аномалій значень часового ряду показника стану природного середовища за рахунок декомпозиції цього ряду на умовні півхвилі між найменшими значеннями та об'єднання результатів послідовного та паралельного застосування різних комбінацій відомих методів автоматичного інтелектуального пошуку аномалій у цих півхвилях, що дозволяє більш точно визначити параметри усіх аномалій та, в цілому – підвищити точність прогнозування цього ряду у різних інтервалах часу.

3. Дістав подальший розвиток метод пошуку можливих джерел підвищеного антропогенного навантаження на стан природного середовища за даними мережі моніторингу цього стану з локалізацією в часі та просторі початку систематичного забруднюючого впливу на природне середовище, за рахунок пошуку моментів максимальних значень приросту з подальшим порівнянням з даними аналізу супутникових знімків або даних аерофотозйомки, що дозволяє більш точно визначити та формалізувати опис регулярної появи надмірного забруднення, на відміну епізодичних аномалій, що в цілому дозволяє підвищити точність прогнозування стану природного середовища.

4. Удосконалена технологія системного аналізу стану природного середовища, що дозволяє більш точно прогнозувати цей стан з використанням

розроблених методів ідентифікації параметрів моделі, байєсівського моделювання зв'язків між даними та за рахунок удосконалення підходів і методів усунення чи мінімізації різних видів невизначеності для заданого регіону, оснований на DPSIR-декомпозиції та методі аналізу ієрархій.

Практичне значення отриманих результатів.

Розроблено комплекс алгоритмів та програмно-інформаційних засобів, які реалізують основні елементи запропонованих методів та технології системного аналізу стану природних середовищ з урахуванням аномалій. Більшість програм є публічно доступними у вигляді програм-ноутбуків у Google-платформі Kaggle експертів у сфері Data Science.

На базі платформи Kaggle, завдяки укладеному договору між ВНТУ та проектом EcoCity, створено датасет «Air Quality Monitoring from EcoCity» з даними показників стану атмосферного повітря зі більше як 30 станцій громадського моніторингу у Вінницькій області за 2019-2025 роки.

Запропонована технологія системного аналізу стану природних середовищ з урахуванням аномалій була застосована для аналізу якості поверхневих вод річки Південний Буг з відкритих даних Державного агентства водних ресурсів України. Результати були розглянуто на засіданні Басейнової ради Південного Бугу.

Запропонована технологія використовувалась під час розроблення Програми державного моніторингу у галузі охорони атмосферного повітря зони «Вінницька» на 2023-2027 роки, що була затверджена згідно чинного законодавства.

Продемонстровано ефективність застосування окремих кроків запропонованої технології під час проведення алергологічного дослідження за участі вчених з Вінницького національного медичного університету ім. М. І. Пирогова та Національного медичного університету імені О. О. Богомольця, яке передбачало моделювання взаємозв'язків чутливості людей до різних груп грибкових алергенів.

Особистий внесок здобувача. Усі теоретичні результати, що виносяться на захист, отримані здобувачем особисто. У роботі здобувачеві належать усі теоретичні результати. У роботі [6] запропоновано удосконалення методу пошуку джерел підвищеного антропогенного впливу за рахунок пошуку моментів максимальних значень приросту, тобто приросту 2-го порядку. У [7] запропоновано підхід формування ієрархічних критеріїв з урахуванням екстремальних даних моніторингу стану атмосферного повітря та моделі DPSIR, автоматизовано метод аналізу ієрархій, використовуючи мову програмування Python. У [8] розроблено метод ідентифікації параметрів сезонності за результатами адаптивної декомпозиції часового ряду. Запропоновано варіант комбінування різних параметрів сезонності при їх одночасному застосуванні при побудові моделі прогнозування. У [9] здійснено статистичний аналіз даних чутливості людей до алергенів та розроблено алгоритм аналізу взаємозв'язків з використанням байєсівського моделювання. У роботі [10] запропоновано удосконалення методу пошуку локальних аномалій за результатами декомпозиції часового ряду на умовні півхвилі з подальшим аналізом аномалій на різних інтервалах ряду та проведено аналіз ефективності застосування ансамблів методів пошуку аномалій. У [11] запропоновано алгоритм системного аналізу стану природних середовищ з урахуванням аномалій, оснований на функціональній DPSIR-декомпозиції та на використанні класичних і авторських методів для розкриття основних видів невизначеності. У [12] описано процес оптимізації моніторингової мережі стану атмосферного повітря зони «Вінницька» з використанням кроків технології системного аналізу стану природних середовищ. У [13] проведено порівняльний аналіз ефективності використання деяких Python-бібліотек для автоматизації процесу пошуку аномалій у часових рядах. У [14] проведено дослідження ефективності роботи моделі прогнозування часових рядів Prophet в залежності від різних налаштувань сезонності. У [15] запропоновано підхід визначення порядку Фур'є для налаштувань сезонності моделі прогнозування часових рядів. У [16] проведено розвідувальний аналіз даних стану атмосферного повітря з використанням бібліотеки для автоматизованого

розвідувального аналізу Sweetviz. У [17] наведено порівняльний аналіз точності роботи моделі Prophet для прогнозування показника стану атмосферного повітря з використання авторських методів ідентифікації параметрів сезонності та аномалій. У [18] наведено приклад застосування методу пошуку можливих джерел забруднення за даними мережі моніторингу до даних стану атмосферного повітря. У [19] запропоновано і описано алгоритм застосування байєсівського моделювання для визначення як дані моніторингу стану атмосферного повітря залежать від змін на сусідніх станціях в межах регіону.

Апробація матеріалів дисертації. Усі описані наукові результати та здобутки, які є частиною дисертаційної роботи пройшли апробацію на 6 наукових конференціях, де загалом було опубліковано 8 тез як особисто, так і у співавторстві:

- LI Науково-технічна конференція факультету інтелектуальних інформаційних технологій та автоматизації, Вінниця, 31 травня - 1 червня 2022 р;
- LII Науково-технічна конференція факультету інтелектуальних інформаційних технологій та автоматизації, Вінниця, 21 - 23 червня 2023 р.;
- LIII Науково-технічна конференція факультету інтелектуальних інформаційних технологій та автоматизації, Вінниця, 20 - 22 березня 2024 р.;
- XVII Міжнародна конференція «Контроль і управління в складних системах» (КУСС-2024), Вінниця, 16-17 жовтня 2024 р.;
- XXIII Міжнародна науково-практична конференція «Математичне моделювання та інформаційно-комунікаційні технології для перемоги та відновлення», Київ, 12-13 листопада 2024 р.;
- LIV Всеукраїнська науково-технічна конференція підрозділів Вінницького національного технічного університету, Вінниця, 23 – 27 березня 2025 р.

Публікації. Загалом за тематикою дисертаційного дослідження було опубліковано 14 наукових праць. Серед них було опубліковано 5 статей у фахових журналах України зі спеціальності 124 – Системний аналіз, 1 стаття в міжнародному фаховому журналі, яка належить до списку видань першого

квартилю реєстрів Scopus і Web of Science Core Collection, 8 тез доповідей у збірниках науково-практичних конференцій.

Документи інтелектуальної власності, що отримані за результатами дисертації.

Отримано свідоцтво (№134476 від 17.03.2025 р.) про реєстрацію авторського права на твір (комп'ютерну програму) «Python-реалізація методу ідентифікації параметрів гармонік та аномалій інтелектуальної моделі часового ряду на основі Prophet» [20]. Свідоцтво наведено в додатку Є.

Структура та обсяг дисертації. Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел та додатків. Повний обсяг дисертації становить 195 сторінок, у тому числі: 141 сторінки основного тексту, 80 рисунків, 11 таблиць, список використаних джерел із 88 найменувань, кількість додатків – 7.

РОЗДІЛ 1

АНАЛІЗ СТАНУ ПРИРОДНИХ СЕРЕДОВИЩ З УРАХУВАННЯМ АНОМАЛІЙ

1.1 Аналіз стану природних середовищ та механізмів моніторингу

Забруднення довкілля та зміна клімату є одними з найбільших викликів сучасності. З кожним роком ці зміни стають все більше відчутними, що проявляється в появі природних аномалій щодо температурного режиму, опадів тощо. Стрімка індустріалізація, урбанізація та багато інших причин призвели до значних екологічних змін, багато з яких, на жаль, мають незворотній характер [11].

Неочікувані циклони та антициклони спричиняють перенесення як вологи в атмосфері, так і забруднювальних речовин. У свою чергу, це спричиняє непередбачувані зміни екологічного стану довкілля [21]. Крім того, в Україні в останні роки спостерігаються зміни у промисловому і сільськогосподарському секторі, пов'язані з переміщенням виробництв. Це, у свою чергу, спричиняє появу викидів, скидів та іншого антропогенного впливу, який не є характерними для конкретного регіону та у певний час. Крім того, прогнозується, що в найближчі роки ситуація продовжить погіршуватися та матиме катастрофічні наслідки для усього населення. Серед основних наслідків зміни клімату, що проявляються по всьому світу, можна виділити наступні [22]:

- танення льодовиків, що призводить до збільшення рівня морів та океанів та ставить під загрозу затоплення прибережні регіони багатьох країн;
 - збільшення дефіциту води, що призводить до зниження врожайності.
- Крім того, за рахунок меншої кількості води зростає і концентрація забруднювальних речовин, які скидаються у воду;
- збільшення кількості захворювань, що частково спровоковано тим, що переносники різних захворювань можуть тепер проживати на набагато ширшій території [21];

– збільшення кількості ураганів, повені, штормів, які можуть нести як економічні наслідки, так і ставити під загрозу життя людей.

Виходячи з усього цього, важливим є контроль за станом природних середовищ з ціллю зменшення негативного на нього впливу. Задовільний стан природного середовища, в свою чергу, допоможе убезпечити людей від негативного впливу, зазначеного вище. Це можна віднести до концепції «Екологічне здоров'я» (англійською «Environmental health») – галузі охорони людського здоров'я, спрямовану на аналіз усунення природних чинників підвищеної захворюваності та смертності людей [21].

Оскільки системи управління станом природного середовища є складними системами, що містять багато взаємопов'язаних підсистем та елементів, має сенс підходити до цієї задачі з позиції системного аналізу.

Одним з найперших кроків проведення системного аналізу систем управління станом природних середовищ є детальний аналіз актуальної ситуації щодо показників та рівнів забруднення цих природних середовищ з акцентом, в першу чергу, на ситуацію в Україні. Крім того, потрібно визначити які механізми контролю та управління за станом природних середовищ зараз існують, щоб зрозуміти які саме проблеми мають місце у даній сфері. Цьому присвячені наступні пункти цього підрозділу.

1.1.1. Показники забруднення стану природних середовищ

Один з кроків системного аналізу стану природного середовища – це аналіз показників відповідного середовища. Дані про стан природного середовища, зазвичай, містять декілька показників, що є часовими рядами – певні хімічні сполуки або елементи, що знаходяться у відповідному природному середовищі та впливають на його стан або якість [23]. Варто охарактеризувати основні показники деяких природних середовищ, адже саме значення цих показників будуть в подальшому використовуватися для аналізу та прогнозування стану відповідного природного середовища.

Забруднення атмосферного повітря відбувається природним та антропогенним шляхом. Прикладами природних джерел забруднення повітря є лісові пожежі та виверження вулканів [24]. Штучне забруднення, у свою чергу, є, переважно, наслідком індустріалізації та урбанізації суспільства: побічні ефекти роботи різних об'єктів промисловості, транспортних засобів тощо. В результаті цього у повітря викидаються величезні обсяги різних шкідливих речовин, наприклад, у формі газів та твердих частинок. Перелік основних показників забруднення атмосферного повітря, що будуть використовуватися в подальших дослідженнях, наведено в таблиці 1.1, а їх детальніший опис наведено в додатку Б.

Таблиця 1.1 – Основні показники забруднення атмосферного повітря [22]

Позначення	Опис
PM1.0/PM2.5/PM10	Дрібні частинки пилу розміром 1/2,5/10 мкм
CO	Оксид вуглецю
CO ₂	Діоксид вуглецю
VOC (H ₂ CO)	Формальдегід
O ₃	Озон
NO	Оксид азоту
NO ₂	Діоксид азоту
SO ₂	Діоксид сірки

Іншим важливим для людства природним середовищем є водні ресурси. Як і атмосферне повітря, забруднення водних ресурсів є одним з найбільших екологічних викликів сучасності. Подібно до атмосферного повітря, забруднення води відбувається через потрапляння забруднювальних речовин хімічного або біологічного характеру, що, в основному, спричинене індустріалізацією [25]. Прикладами джерел забруднення є скидання погано очищених стічних вод у природні водойми, використання хімічних сполук у сільському господарстві тощо. Джерелами забруднення води можуть бути або точкові, або дифузійні

джерела. Точковими називають просторово-зосереджені джерела, наприклад, певна очисна споруда. Дифузійні ж джерела є просторово-розподіленими, наприклад, стік з сільськогосподарських полів. Перелік основних показників, за якими здійснюється моніторинг стану водних ресурсів, наведено в таблиці 1.2, а їх детальніший опис наведено в додатку Б.

Таблиця 1.2 – Основні показники забруднення водних ресурсів [22]

Позначення	Опис
БСК5	Біохімічне споживання кисню за 5 діб
SO ₄	Сульфат-іони
NH ₄	Амоній-іони
NO ₃	Нітрат-іони
NO ₂	Нітрит-іони
PO ₄	Фосфат-іони

Як можна побачити з таблиць 1.1 та 1.2, стан кожного природного середовища характеризується великим переліком різноманітних показників. Проте, окрім них, також існують різні інтегральні показники та індекси, що комбінують в собі описані вище та дозволяють оцінити екологічну ситуацію і представити її у спрощеному та зрозумілому для населення вигляді. Розглянемо приклади таких індикаторів стану декількох природних середовищ.

1.1.2. Інтегральні показники стану природних середовищ

Окрім згаданих в попередньому пункті показників, стан природного середовища часто характеризується певними інтегральними індикаторами – кількісними або якісними оцінками стану середовища, що враховують показники забруднення та відображають рівень забруднення в більш спрощеному та зрозумілому для ширшого переліку людей форматі. Подібні оцінки використовуються, в першу чергу, для інформування населення про поточний

стан середовища, ризики, пов'язані з відповідним рівнем забруднення, та для формування рекомендацій щодо необхідних дій щодо уникання контакту із забрудниками.

Загальноприйнятим методом інформування населення щодо якості атмосферного повітря є індекс якості повітря [26] (англійською «Air Quality Index» або скорочено «AQI»). На рівні різних країн визначаються різні категорії та підходи щодо обрахунку і категоризації цих показників до відповідного рівня забруднення, але, зазвичай, індекс якості повітря – це певне значення (деякі країни використовують числові значення, а деякі – слова), яке відповідає рівню ризику для населення та вразливих його верств (люди з хронічними хворобами, діти, літні люди). До кожного рівня індексу якості повітря також додаються рекомендації для населення, які можуть містити як обмежену, так і повну заборону перебування на відкритому повітрі, оскільки цей рівень забруднення може призвести до важких наслідків для здоров'я людини.

Як було зазначено, різні країни мають дещо різні підходи до обрахунку цих значень, але, в загальному, методологія вимагає збору інформації по певних показниках забруднення та їх усереднення за певний часовий проміжок. В переважній більшості, країни беруть до уваги показники діоксиду сірки (SO₂), великих та дрібних часток пилу (PM₁₀ та PM_{2.5} відповідно), діоксиду азоту (NO₂) та оксиду вуглецю (CO). Потім, обраховані значення порівнюються з граничними, що дозволяє віднести поточний стан забруднення до відповідного діапазону.

Прикладом такого індексу якості повітря є Європейський індекс якості повітря (англійською «European Air Quality Index»), розроблений Європейським агентством з навколишнього середовища (англійською «European Environment Agency» або скорочено «EEA»), який використовується усіма країнами-членами Європейського союзу [27]. Також, варто згадати і про Український індекс якості повітря, розроблений в рамках міжнародної програми «Чисте повітря для України» і який реалізовано та застосовується мережею громадського моніторингу «EcoCity» [28].

Подібно до AQI існують індекси і для визначення якості води. Індекс якості води (англійською «Water Quality Index», або «WQI») є одним із таких інструментів – він заснований на фізичних, хімічних і біологічних факторах, які об'єднані в одне значення в діапазоні від 0 до 100. Його загальна концепція складається з чотирьох кроків: вибір параметрів якості води, приведення параметрів до загальної шкали, визначення ваг для параметрів, агрегація значень для отримання остаточного значення WQI.

Прикладами реалізації цієї концепції є індекс якості води Національного санітарного фонду [29] (англійською «National Sanitation Foundation Water Quality Index», або «NSFWQI»), індекс якості води Канадської ради міністрів навколишнього середовища [30] (англійською «Canadian Council of Ministers of Environment Water Quality Index», або «CCMEWQI»).

Проаналізуємо сучасний рівень забруднення природних середовищ в Україні, в тому числі, значення згаданих вище показників стану цих середовищ.

1.1.3. Сучасний рівень забруднення природних середовищ в Україні

Згідно законодавства та відповідних постанов [31-35], Міністерство захисту довкілля та природних ресурсів України щороку готує аналітичний звіт щодо стану навколишнього середовища в Україні. У цьому звіті наводиться різна статистична та аналітична інформація щодо рівнів забруднення основних природних середовищ, інформація про основні джерела забруднення тощо. Серед багатьох напрямків, цей звіт, у тому числі, містить інформацію про атмосферне повітря та водні ресурси. В умовах воєнного стану, дані звіти не публікують, тому розглянемо останній на даний момент публічний звіт, що був опублікований про стан навколишнього середовища за 2021 рік [36].

Згідно цього звіту, головними джерелами забруднення повітря були транспорт, підприємства теплоенергетики та переробна промисловість [36]. В той же час, головними проблемами та причинами того, що існуючий стан атмосферного повітря небезпечний для населення, були:

- низькі темпи впровадження технологій для очищення викидів;
- невиконання необхідних заходів для зменшення загальних обсягів викидів;
- наближеність підприємств до місць проживання населення.

Річні об'єми кількості забруднювальних речовин за 2019-2021 роки, згідно цього звіту, наведено в таблиці 1.3.

Таблиця 1.3 – Обсяги викидів в атмосферу забруднювальних речовин по Україні за 2019-2021 роки [36]

	Викиди забруднювальних речовин, т		
	2019	2020	2021
SO ₂	19803	18169	19643
CO	1255225	1084752	1164873
NO ₂	178778	161781	174420
NO	798	799	896
NH ₃	6	7	8

Як можна помітити в таблиці 1.3, рівень забруднення за основними показниками або залишились на попередньому рівні, або зростали протягом останніх років. Проте, важливим є не тільки абсолютні значення викидів забруднювальних речовин, але і частота перевищення гранично допустимих значень забруднення. Розглянемо відсоток випадків за 2019-2021 роки, коли аналіз атмосферного повітря дозволяє зробити висновок про перевищення гранично допустимих значень. Результат наведено в таблиці 1.4.

Таблиця 1.4 – Випадки перевищення гранично-допустимих значень при вимірюванні стану атмосферного повітря по Україні за 2019-2021 роки [36]

	Частота перевищень гранично-допустимим рівнів, %		
	2019	2020	2021
Сумарний показник	4%	4.2%	4.2%
Пил	6.5%	6.7%	6.5%
Окис вуглецю	5.6%	5.2%	5.8%

Згідно даних цього ж звіту, значну роль у забрудненні водних ресурсів відіграло потрапляння шкідливих речовин у воду через поверхневий стік із забудованих територій та сільськогосподарських угідь. Найбільшими забруднювачами були сільське господарство, видобувна та переробна промисловість.

Обсяги скидів зворотних вод, згідно цього звіту, за 2020-2021 роки наведено в таблиці 1.5.

Таблиця 1.5 – Обсяги скидів зворотних вод по Україні за 2020-2021 роки [36]

	Обсяги скинутих зворотних вод, млн. м ³	
	2020	2021
Усього	5159	4684,6
Забруднені	518	541,6
Без очищення	100	119,3
Недостатньо очищені	418	422,2
Нормативно-очищені	1425	1430,2
Нормативно-чисті	3216	2712,9

Подібно до загально-національного звіту стану навколишнього середовища, щорічно готуються і регіональні звіти для кожної адміністративної одиниці України. Аналізуючи останній публічно доступний звіт щодо екологічного стану Вінницької області [37], можна сказати, що рівень забруднення по багатьох показниках загалом має певні позитивні зміни. Основними речовинами-забрудниками повітря є сполуки вуглецю та сірки. Основними ж джерелами, подібно до загальноукраїнського стану, є об'єкти промисловості, сільського господарства та транспорт.

В таблиці 1.6 наведено середнє річне значення та максимальне зафіксоване за рік значення перевищення граничних рівнів забруднення для деяких основних показників забруднення атмосферного повітря.

Таблиця 1.6 – Середнє річне значення викидів забруднювальних речовин у Вінницькій області [37]

	Середнє значення забруднювальної речовини за рік	Зафіксоване перевищення граничного рівня забруднення
SO ₂	0,215 мг/м ³	0,54 мг/м ³
CO	2,9 мг/м ³	6,5 мг/м ³
PM	0,36 мг/м ³	4,15 мг/м ³
NH ₃	0,13 мг/м ³	0,36 мг/м ³
CO ₂	0,7571 мг/м ³	-

З таблиці 1.6 видно, що за основними показниками існує значне перевищення граничних показників, що може мати негативний вплив на населення та стан регіону, що, у свою чергу, вимагає посилення заходів контролю та зменшення викидів.

В таблиці 1.7 наведемо річні обсяги забруднювальних речовин, виявлених у воді протягом 2021-2023 років.

Таблиця 1.7 – Обсяги забруднювальних речовин води у Вінницькій області [37]

Забруднювальна речовина	2021	2022	2023
БСК5	159,6	141,4	148,2
Нітрати	481,3	508,5	442,3
Нітрити	52,9	34,9	49,4
Фосфати	37,7	31,7	32,3
Амоній	56,5	40,4	44,6
Сульфати	1765,4	1537,3	1395,8

Аналізуючи дані в таблиці 1.7, видно що присутня певна тенденція до зменшення концентрації більшості забруднювальних речовин. У той же час, концентрація нітратів і нітритів показує нестабільну ситуацію, коли значення спочатку зменшуються, а потім знову зростають. Тому важливим є подальший моніторинг та заходи щодо контролю за станом середовища.

Проаналізуємо які державні механізми функціонують в Україні та наскільки ефективно вони дозволяють контролювати стан природних середовищ.

1.1.4. Державні механізми моніторингу та контролю стану природних середовищ

Державна політика та системи моніторингу стану природних середовищ довгий час базувались на застарілих нормах, проте, протягом останніх років, Україна стрімко рухається в напрямку всебічної інтеграції з Європейським союзом. Серед багатьох норм та моделей, які діють на заході, важливим місцем є екологічний контроль і Україна, як офіційний кандидат на вступ до Європейського союзу, зобов'язана адаптуватися до них. Так, наприклад, згідно Угоди про асоціацію з ЄС [38] та її додатками, пов'язаними з контролем за навколишнім середовищем, Україна зобов'язується створювати та розвивати системи моніторингу за якістю повітря, запроваджувати плани дій на випадок

перевищення різних показників забруднення та зменшувати викиди небезпечних речовин у повітря.

Україна вже має великий прогрес в питанні адаптації до зазначених норм та директив. Європейські підходи щодо моніторингу якості атмосферного повітря та водних ресурсів вже затверджені на законодавчому рівні переліком постанов та наказів різних державних органів та міністерств [31-33]. Тобто, можна говорити, що вже є певна законодавча основа для подальшого розвитку мережі державного моніторингу та певні законодавчі механізми контролю та боротьби з джерелами антропогенного забруднення. Суб'єкти господарювання ж зобов'язані дотримуватись відповідних норм та скорочувати об'єми шкідливих речовин, які вони викидають в природне середовище.

Згідно закону України [33] передбачається наявність державної системи моніторингу довкілля, до складу якої входять спостереження за станом і рівнем забруднення навколишнього середовища: якість повітря, стан вод, стан ґрунтів тощо. Відповідальним за впровадження і розвиток системи моніторингу є Міністерство захисту довкілля та природних ресурсів України. За інформацією міністерства, Державна гідрометеорологічна служба проводить моніторинг забруднення атмосферного повітря у 53 містах України на 162 стаціонарних станціях моніторингу. Крім того, мережа також налічує 2 маршрутних поста спостереження та 2 станції транскордонного переносу [35]. Обов'язкове спостереження проводиться за 7-ма речовинами: пил, двоокис азоту (NO_2), двоокис сірки (SO_2), оксид вуглецю, формальдегід (H_2CO), свинець та бензопірен.

Державна екологічна інспекція (Міндовкілля) здійснює відбір проб на джерелах викидів та перевіряє їх на понад 65 параметрів забруднення [35]. Міністерство охорони здоров'я України здійснює спостереження за якістю атмосферного повітря, в першу чергу у житлових районах та санітарно-захисних зонах (зони розміщення житлових будинків, навчальні та медичні заклади тощо) [35]. Також проводиться аналіз якості повітря у житловій зоні за скаргами мешканців відповідної зони.

На території Вінницької області, дослідження стану та рівню забруднення атмосферного повітря проводиться державною лабораторією Вінницького обласного лабораторного центру Міністерства охорони здоров'я України [35]. Також, на 4 пунктах спостереження Вінницьким центром з гідрометеорології ведеться моніторинг атмосферних опадів.

Головним виконавчим органом України, який реалізує політику управління та використання водних ресурсів, є Державне агентство водних ресурсів. Дане агентство проводить заходи, спрямовані на покращення екологічного стану вод, виконує державний моніторинг поверхневих вод та впроваджує плани управління річковими басейнами (ПУРБ) [39]. У межах кожного річкового басейну функціонує басейнова рада, основними завданнями якої є проведення на локальному рівні заходів щодо узгодження співпраці місцевих органів влади, підприємств і організацій для забезпечення, як мінімум, доброго екологічного стану водних ресурсів та контроль за виконанням ПУРБ.

Важливим механізмом екологічного контролю водних ресурсів є надання регулярної звітності від усіх водокористувачів. Водокористувачі зобов'язані подавати річний звіт за формою 2-ТП (водгосп) до відповідного локального органу управління Держводагентства [40]. Одна з таблиць даної форми наведена на рисунку 1.1.

№ рядка	Назва джерела водопостачання або водокористувача	КОДИ			Відстань від гирла, км	Забрано або одержано води за рік												тис. куб. м	
		типу джерела та водокористувача, що передає	поверхневого водного об'єкта (джерела постачання)	категорії якості води		усього	у тому числі за місяцями												
							I	II	III	IV	V	VI	VII	VIII	IX	X	XI		XII
A	B	V	Г	Д	Е	1	2	3	4	5	6	7	8	9	10	11	12	13	
1		/																	
2		/																	
3		/																	
4		/																	
5		/																	

№ рядка	Ліміт використання	Фактично використано води за рік							Передано іншим водокористувачам за рік				Втрати води за рік	тис. куб. м
		усього	у тому числі на потреби			на інші потреби		без використання		після використання				
			питні і санітарно-гігієнічні	виробничі (технологічні)	прошених	питні і санітарно-гігієнічні користувачів, що не звітують	код виду використання	об'єм	код категорії якості води	об'єм	код категорії якості води	об'єм		
A	14	15	16	17	18	19	20	21	22	23	24	25	26	
1														
2														
3														
4														
5														

Рисунок 1.1 – Форма подачі звітності щодо водокористування [40]

У цьому звіті необхідно вказати перелік місць звідки проводиться водозабір та зазначити фактичне місце, звідки проводиться забір води, тип водокористування, річні і місячні об'єми отриманої води та спеціальні коди типів джерел водопостачання, які затверджені у Додатку № 1 наказу Міністерства екології та природних ресурсів України про порядок ведення обліку водокористування. Наступною важливою частиною звіту є інформація про водовідведення, яка містить назви приймачів стічних вод, річні об'єми відведених вод з поділом на кількість води, яка була очищена перед відведенням та ту, яку було викинуто у забрудненому стані. Обов'язково необхідно вказати вміст забруднювальних речовин у зворотних водах. Також додаються різні додаткові показники, такі як: кількість робочих днів, річні ліміти забору, потужності очисних споруд тощо.

Окрім описаних та інших різноманітних механізмів моніторингу та управління станом природними середовища, важливе місце мають громадські ініціативи та проекти, що намагаються розв'язати різні екологічні задачі. Тому також варто їх окремо розглянути.

1.1.5. Громадські проекти моніторингу стану природних середовищ

Як було згадано в попередньому пункті, сучасний стан моніторингу природних середовищ в Україні суттєво покращився протягом останніх років, а системи моніторингу, загалом, мають тенденцію до подальшого розвитку та модернізації. Проте, це не завжди було так і це давало поштовх до створення та розвитку громадських проектів та ініціатив, які додатково намагалися шукати певні рішення щодо контролю та покращення природних середовищ та екологічної ситуації в Україні загалом.

Одним з таких проектів є мережа громадського моніторингу EcoCity. Цей проект був створений і продовжує функціонувати з 2019 року за підтримки міжнародної програми «Чисте повітря для України», яка була ініційована громадською організацією «Арніка» (Чехія) [28, 41]. Проект займається

розробленням та розповсюдженням станцій моніторингу, які можуть бути встановлені компаніями або звичайними людьми для збору інформації про стан атмосферного повітря. Крім того, проєкт розробив інтерактивну мапу забруднення, доступну як веб-додаток, що дозволяє будь-кому переглядати щохвилинну інформацію про забруднення повітря майже в режимі реального часу. Станом на початок 2025 року, мережа нараховує понад 350 станцій, розташованих у понад 100 населених пунктах у межах всієї України [28]. Дві такі станції розташовані на території Вінницького національного технічного університету та доступні через згаданий веб-додаток EсоCity.

Окрему увагу варто приділити досить важливому та корисному елементу мережі моніторингу EсоCity – веб-сервіс «Кабінет дослідника» [42]. Як було зазначено раніше, інтерактивна мапа EсоCity дозволяє переглядати актуальну інформацію щодо якості атмосферного повітря, проте не надає можливості перегляду та завантаженню архівів даних, які є корисними для проведення різного роду досліджень. «Кабінет дослідника» вирішує цю проблему, адже сервіс дає можливість користувачам робити запити на отримання архівних даних за усіма наявними показниками за заданий проміжок часу. Доступ до даного сервісу відкривається за умови укладання договору з проєктом EсоCity. ВНТУ уклав цю угоду (договір про приєднання до Програми підтримки дослідників якості повітря України та реєстрацію Кабінету дослідника якості повітря України від 24.03.2023 р), що відкрило доступ до даних з більш як 30-и станцій моніторингу, які розташовані у м. Вінниця та Вінницькій області, з даними за період від 2019 року.

Приклад форми створення запиту на завантаження даних та налаштування часового проміжку, за який необхідно завантажити дані, наведено на рисунках 1.2 та 1.3, відповідно.

Запит /Request

Архівні дані доступні від **2019-02-20** до **2024-11-11**

Оберіть адміністративну одиницю /Select an administrative unit

Вінницька область / Vinnytska oblast

Оберіть параметри моніторингу якості повітря /Select air quality monitoring parameters

[Усі параметри](#) /All parameters

<input checked="" type="checkbox"/>	PM1.0 ug/m ³	Пил 1.0 мкм мкг/м ³
<input checked="" type="checkbox"/>	PM2.5 ug/m ³	Пил 2.5 мкм мкг/м ³
<input checked="" type="checkbox"/>	PM10 ug/m ³	Пил 10 мкм мкг/м ³
<input type="checkbox"/>	Temperature °C	Температура °C
<input type="checkbox"/>	Humidity %	Вологість %
<input type="checkbox"/>	Pressure hPa,Pa	Атмосферний тиск мм.рт.ст.
<input type="checkbox"/>	Noise db	Шум дб.
<input type="checkbox"/>	RAD uR/h	Радіаційний фон мкР/год
<input type="checkbox"/>	Solar Radiation Vat/m ²	Сонячна радіація Вт/м ²

Рисунок 1.2 – Форма на створення запиту на отримання даних з сервісу «Кабінет дослідника» [42]

Масштаб усереднення результатів /The averaging results scale

20 хвилин / 20 minutes

Дата від /Date from

02/20/2019

Дата до /Date to

11/11/2024

Сформувати запит /Create the request

Рисунок 1.3 – Налаштування часового проміжку та масштабу усереднення даних з сервісу «Кабінет дослідника» [42]

Як видно на рисунках 1.2 та 1.3, система дозволяє обрати показники забруднення з списку усіх наявних (важливе уточнення, що зазначений перелік містить абсолютно усі показники, які теоретично можуть збиратися станціями, проте деякі більш простіші по комплектації станцій збирають дані по обмеженому переліку показників), часовий проміжок, за який необхідно отримати набір даних та масштаб усереднення цих даних. Система надає можливість отримати усередненні дані за 20 хвилин, за 1 годину або за 1 добу. Приклад отриманих даних наведено на рисунку 1.4.

# id_station	datetime	indicator_name	indicator_dimens...	# value
650	2020-01-22 2024-03-11	Temperature	ug/m3	-8.68
650	2020-01-22 00:00:00	Humidity	°C	102k
650	2020-01-22 00:00:00	Other (5517)	Other (4201)	
650	1/22/2020 00:00:00	NH ₃	ppm	0
650	1/22/2020 00:00:00	NO ₂	ppm	5.2887
650	1/22/2020 00:00:00	CO	ppm	0.002
650	1/23/2020 00:00:00	NH ₃	ppm	0
650	1/23/2020 00:00:00	NO ₂	ppm	7.0416
650	1/23/2020 00:00:00	CO	ppm	0.002
650	1/24/2020 00:00:00	NH ₃	ppm	0
650	1/24/2020 00:00:00	NO ₂	ppm	9.4682

Рисунок 1.4 – Приклад даних отриманих за допомогою сервісу «Кабінет дослідника» [23]

Іншим проектом є система моніторингу якості атмосферного повітря YourAirTest. Проект займається збиранням та обробленням інформації про якість атмосферного повітря. Користувачі системи можуть переглядати інформацію в режимі, близькому до реального часу, через відповідний веб-портал (рис. 1.5).

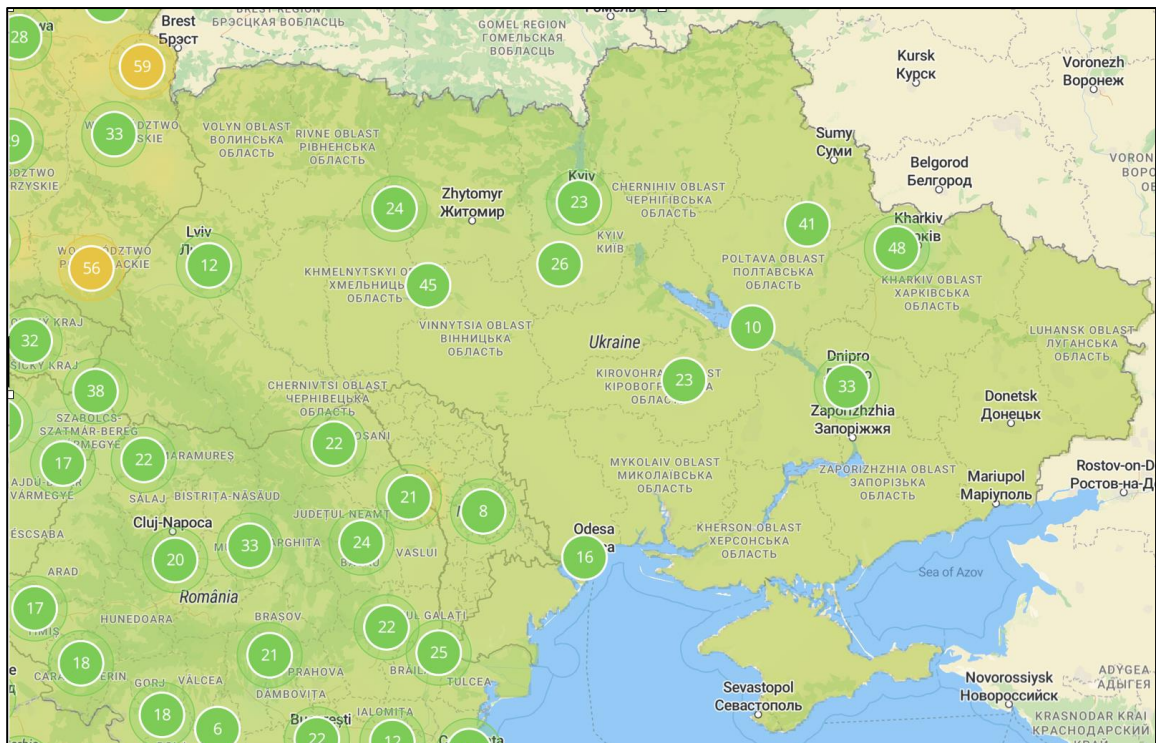


Рисунок 1.5 – Веб-додаток системи «YourAirTest» (<https://yourairtest.com/>)

Крім того, серед напрямків діяльності системи є аналіз та моделювання пожеж, дослідження та прогнозування погоди, тощо.

Також варто згадати інший громадський проєкт, що займається екологічними питаннями – SaveEcoBot [43]. Проєкт є розробкою громадської організації «SaveDnipro», яка займається аналізом екологічних проблем та пошуком їх рішень. SaveEcoBot — це екологічний чат-бот та інтерактивна веб-мапа, яка дозволяє переглядати дані про забруднення та забруднювачів, а також отримувати сповіщення в разі зміни рівнів забруднення. Інтерактивна мапа містить дані моніторингу якості повітря, надані самою ГО «SaveDnipro», так і дані від партнерів даного проєкту. Раніше серед таких партнерів був і проєкт EcoCity, проте, на жаль, в останні роки EcoCity не передає інформацію в SaveEcoBot, відповідно є можливість перегляду тільки застарілих даних і станцій.

Варто зазначити, що дані з цих двох мереж було агреговано Міндовкілля – вони доступні через веб-сервіс «ЕкоЗагроза», приклад вікна якого наведено на рисунку 1.6.

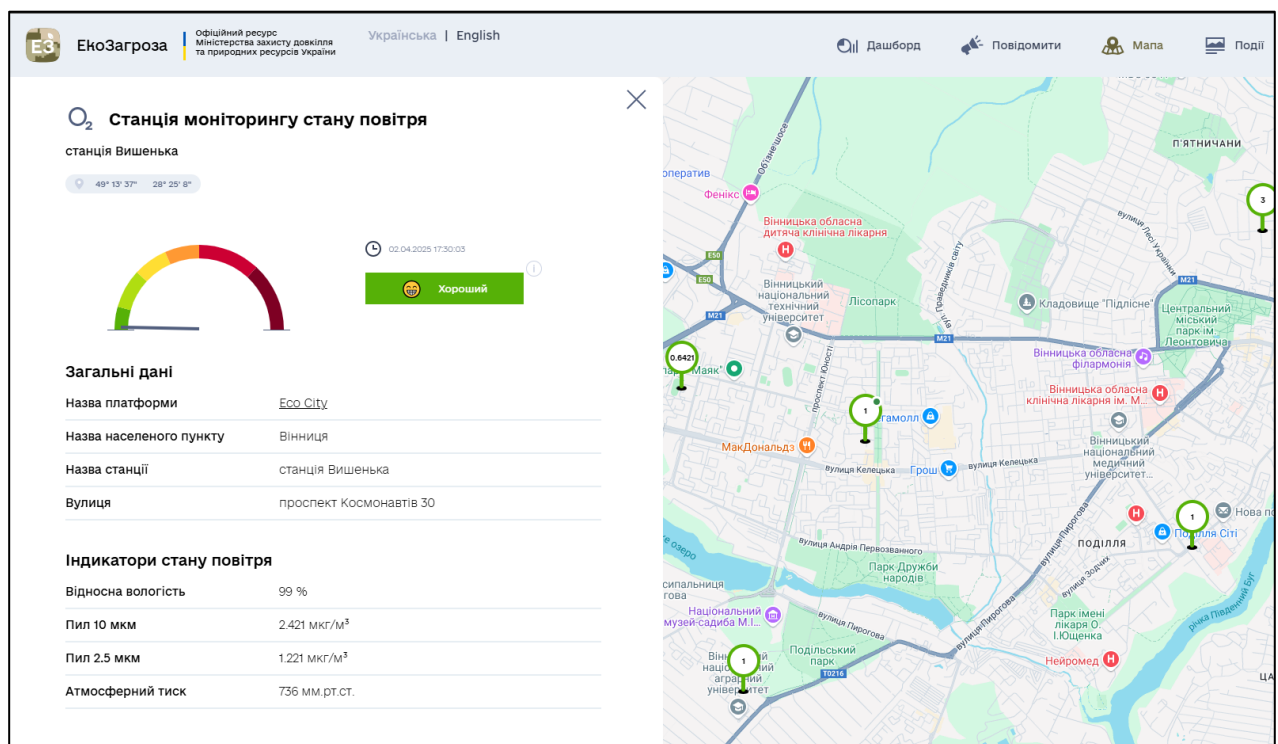


Рисунок 1.6 – Веб-додаток сервісу «ЕкоЗагроза»

1.2. Аналіз систем управління станом природного середовища

Системою управління станом природного середовища називають сукупність методів, технологій та підходів моніторингу стану природного середовища, аналізу відповідних показників, прогнозування та прийняття необхідних управлінських рішень для покращення стану цього природного середовища [44].

Стан природних середовищ та системи управління ними є розповсюдженою і актуальною проблемою у всьому світі протягом останніх кількох десятиліть. В кінці 20-го століття було розроблено багато міжнародних стандартів екологічного управління. Прикладом такого є низка стандартів сімейства ISO 14000, які визначають загальні критерії оцінювання відповідності систем управління навколишнім середовищем [45]. Наприклад, ISO 14001 – це стандарт систем екологічного менеджменту, який встановлює вимоги до системи управління навколишнім середовищем, яку можуть застосовувати будь-які організації для покращення стану природних середовищ та дотримання законодавчих вимог. Стандарт не вказує конкретні вимоги, а визначає концептуальну структуру, якої варто дотримуватися задля ефективного функціонування системи управління станом природного середовища.

На основі цього стандарту, Європейським Союзом було розроблено стандарт EMAS (Eco-Management and Audit Scheme) [46]. EMAS це – система управління екологічного стану, спрямована на покращення та універсалізація звітування про екологічні показники організацій. Це дає можливість організаціям оцінювати, керувати та постійно покращувати свою екологічну ефективність.

Крім того, існує багато підходів для комплексного оцінювання стану природних середовищ та впливу на нього. Одним з таких є застосування моделі DPSEEA [47]. Назва моделі є аббревіатурою перших літер назв 6-ти її взаємопов'язаних складових (рис 1.7): рушійні сили (англійською «Driving Forces» – «D»), тиск (англійською «Pressure» – «P»), стан (англійською «State» –

«S»), вплив (англійською «Exposure» – «E»), наслідки (англійською «Effect» – «E») та дії (англійською «Actions» – «A»).

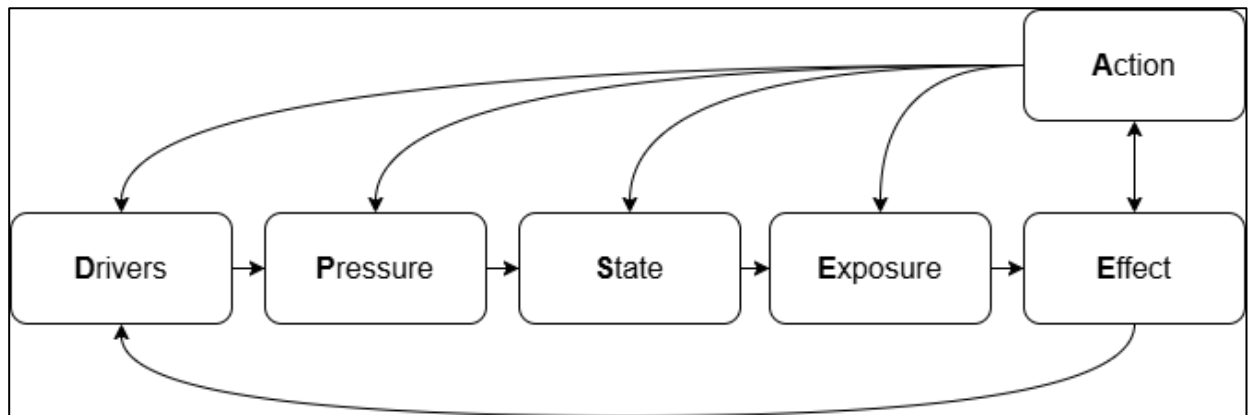


Рисунок 1.7 – Структурна схема моделі «DPSEEA»

Модель була розроблена Всесвітньою організацією охорони здоров'я і призначена, в першу чергу, для вимірювання впливу змін довкілля на здоров'я людини. Часто використовується в Європі для моніторингу впливу зміни клімату на здоров'я населення. Хоча, вплив на здоров'я і є досить важливим фактором при оптимізації систем управління станом природних середовищ, але ця модель має більш вузьку сферу та можливості для застосування.

Більш узагальненим підходом комплексного оцінювання стану природного середовища є концепція екологічного управління на основі моделі DPSIR [48]. Дана модель була розроблена в кінці минулого століття Європейським агентством з навколишнього середовища. DPSIR використовується для опису взаємодії між суспільством і навколишнім середовищем, дозволяє структурувати різні екологічні проблеми, виділити зв'язки між елементами середовища та розробляти ефективні підходи покращення стану середовища. Подібно до DPSEEA, модель виділяє п'ять елементів системи, що закладені в назві самої моделі (рис. 1.8): продуктивні сили (англійською «Drivers» — «D»), тиск на середовище (англійською «Pressures» — «P»), стан середовища (англійською «State» — «S»), вплив (англійською «Impact» — «I») та рішення (англійською «Responses» —

«R»). Складова «Impact» відповідає складовим «Exposure» та «Effect» у моделі DPSEEA.

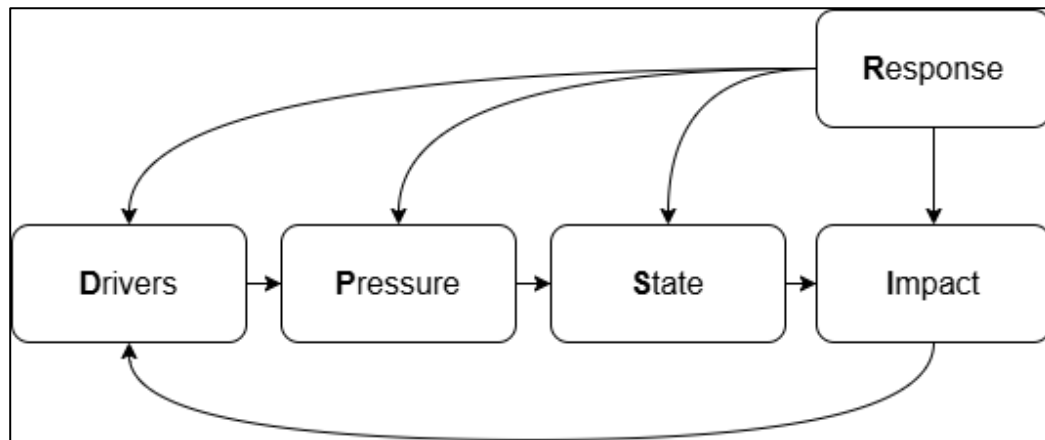


Рисунок 1.8 – Структурна схема моделі «DPSIR»

З моменту створення ця модель знайшла широке застосування у різних сферах контролю стану різних природних середовищ. Це досить гнучкий інструмент, який можна успішно застосовувати в соціальних та економічних сферах.

Виходячи з аналізу, модель DPSIR може бути основою для подальшого системного аналізу систем управління та стану відповідних природних середовищ, оскільки вона дозволяє виділити елементи складної системи та аналізувати взаємозв'язки між ними та є більш гнучким підходом для всебічного аналізу природного середовища.

Системний аналіз, у свою чергу, вимагає проведення аналізу та прогнозування часових рядів даних стану природних середовищ, тому розглянемо та проаналізуємо інструменти та моделі для роботи з часовими рядами.

1.3. Огляд інтелектуальних моделей прогнозування часових рядів

Аналіз та прогнозування часових рядів знаходять практичне застосування у багатьох сферах, в тому числі і у сфері системного аналізу стану природних середовищ. До задач, які розв'язуються з використанням методів аналізу та

прогнозування часових рядів, можна віднести ідентифікацію різних параметрів сезонності ряду та побудову моделей передбачення з використанням цих параметрів.

Результати аналізу дозволяють побудувати ефективну модель передбачення та, тим самим, дозволяє прогнозувати зміни показників відповідного природного середовища. Розуміння того, як розвивається ситуація щодо рівня забруднення, можна приймати певні обґрунтовані рішення, спрямовані на покращення тенденцій цих змін у кращу сторону. При цьому, результати аналізу часового ряду самі по собі мають велику цінність, оскільки це дозволяє знайти багато цікавих закономірностей щодо самого ряду.

Зазвичай, аналіз часового ряду містить декілька можливих етапів. Ці етапи та їх порядок обираються системним аналітиком у залежності від потреб щодо конкретного часового ряду. Прикладами таких етапів є [49]:

1. Аналіз пропусків даних. Цей крок передбачає пошук та заповнення існуючих пропусків певними очікуваними даними, які визначаються за рахунок існуючих. Необхідність цього кроку, зазвичай, визначається в залежності від того, яка модель буде будуватися далі, оскільки деякі моделі ефективно працюють і за наявності пропусків у даних.

2. Пошук та аналіз аномалій. Аномалії, зазвичай, суттєво погіршують ефективність моделей передбачення, тому важливим є знайти ці аномалії та проаналізувати їх вплив. Аномалії можуть мати як короткотривалий, так і довготривалий вплив. Детальніше про цей етап буде згадано в наступних розділах – він є ключовим у цьому дослідженні.

3. Аналіз законів розподілу. Досить велика кількість методів аналізу даних базуються на припущенні про нормальний розподіл даних. Виходячи з цього, корисним може бути перевірка цієї гіпотези і якщо вона не підтверджується, то необхідно розглядати можливість нормалізації даних.

4. Аналіз стаціонарності ряду. Стаціонарним називають ряд, статистичні показники якого є незмінними в часі. А отже, статистичні показники різних вибірок ряду будуть однаковими чи зіставними у межах заданої похибки.

Класичним прикладом нестационарного ряду є температура повітря протягом року. Якщо взяти вибірку взимку чи влітку, звичайно, середнє буде суттєво відрізнятися. За можливості, ряд варто стаціонаризувати, що буде особливо корисно при використанні моделей на основі нейронних мереж.

5. Аналіз сезонності ряду. Моделі прогнозування часових рядів, на відміну від інших моделей машинного навчання, зазвичай, враховують сезонність цього ряду, яка, як правило, описується рядами Фур'є [49]:

$$\begin{aligned}
 a_0 &= \frac{1}{T} \int_0^T y(t) dt, \\
 a_i &= \frac{2}{T} \int_0^T y(t) \cos\left(\frac{2\pi}{T} it\right) dt, \\
 b_i &= \frac{2}{T} \int_0^T y(t) \sin\left(\frac{2\pi}{T} it\right) dt, \\
 y(t) &= \frac{a_0}{2} + \sum_{i=1}^n a_i \cos\left(\frac{2\pi}{T} it\right) + \sum_{i=1}^n b_i \sin\left(\frac{2\pi}{T} it\right), \tag{1.1}
 \end{aligned}$$

де n – кількість гармонік ряду, T – період ряду, a_0 , a_i , b_i , – коефіцієнти Фур'є, а $i = 1 \dots n$.

Сучасні бібліотеки та інструменти для навчання моделей часових рядів, як правило, самі ідентифікують параметри Фур'є. Для цього їм потрібно лише 2 гіперпараметри: період T і кількість гармонік ряду Фур'є n .

В додатку В наведено детальніший приклад та декілька відомих інструментів проведення розвідувального аналізу даних на прикладі даних стану атмосферного повітря.

Результати усіх застосованих етапів розвідувального аналізу застосовуються для прийняття рішення про те, які далі будувати моделі, як правило інтелектуальні, та які саме параметри відповідного набору даних використовувати. Процес вибору значущих параметрів називають Інженерія ознак (англійською «Feature engineering» або скорочено «FE») [49]. У разі, якщо далі планується використання класичних багатofакторних моделей машинного

навчання (регресійні, дерева рішень, звичайні нейронні мережі тощо), потрібно спробувати знайти або синтезувати побільше ознак, навіть, якщо задано тільки одна ознака. Існують Python-бібліотеки, які дозволяють синтезувати велику кількість ознак для заданого часового ряду, наприклад, бібліотека tsfresh [50]. Обчислюється велика кількість різних статистичних характеристик заданого часового ряду (не тільки дисперсія чи середньоквадратичне відхилення, а й енергія та ін.) із різним «вікном».

Наступним етапом після аналізу часового ряду, зазвичай, є прогнозування його значень, суть якого полягає в розрахунку майбутніх значень показників цього ряду на основі існуючих спостережень [49]. Прогнозування дозволяє зрозуміти тенденцію можливих змін стану природного середовища та вжити певні заходи задля його покращення.

Безпосередньо для прогнозування часових рядів можна використовувати рекурентні нейронні мережі, зокрема GRU та LSTM, але, частіше, використовуються специфічні для часових рядів моделі ARIMA та Prophet [49].

Математичний апарат ARIMA [49, 51] (англійською «Autoregressive Integrated Moving Average» – модель авторегресії та проінтегрованого ковзного середнього). Ця модель є узагальненням моделі авторегресійного ковзного середнього (ARMA) для нестационарних рядів. Модель є різновидом регресійного аналізу, яка вимірює силу однієї залежної змінної відносно інших змінних. Вона поєднує в собі три компоненти:

- AR – компонент авторегресії;
- I – інтегруючий компонент;
- MA – компонент ковзного середнього.

Нехай часовий ряд в момент часу t позначається Y_t . Якщо часовий ряд нестационарний, то його потрібно перетворити шляхом диференціювання за що відповідає інтегруючий компонент [49, 51]:

$$\nabla^d y_t = (1 - B)^d y_t, \quad (1.2)$$

де B - оператор зсуву, а d – порядок диференціювання.

Авторегресійна модель для Y_t порядку p [49, 51]:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t, \quad (1.3)$$

де $\phi_1, \phi_2, \dots, \phi_p$ - коефіцієнти авторегресії, а ε_t - білий шум (з постійною дисперсією та нульовим середнім).

Компонент ковзного середнього порядку q позначається [49, 51]:

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q}, \quad (1.4)$$

де $\theta_1, \theta_2, \dots, \theta_q$ - коефіцієнти ковзного середнього, а ε_t - білий шум.

Об'єднавши всі компоненти отримуємо узагальнений запис моделі ARIMA [49, 51]:

$$y'_t = \phi_1 y'_{t-1} + \phi_2 y'_{t-2} + \dots + \phi_p y'_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t, \quad (1.5)$$

де d - порядок диференціювання, p - порядок авторегресійного компонента, а q - порядок компонента ковзного середнього.

Модель Prophet - доволі популярний метод аналізу та прогнозування часових рядів, розроблений компанією Facebook (Meta) [52]. Модель є зручним інструментом для роботи з даними, в яких присутні аномалії та різні сезонні зміни і тренди. Математично, модель Prophet для моделювання та прогнозування значень ряду $y(t)$, залежно від часу t , записується наступним чином [49, 52]:

- для адитивного випадку:

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t, \quad (1.6)$$

- для мультиплікативного випадку:

$$y(t) = g(t) \cdot s(t) \cdot h(t) \cdot \varepsilon_t, \quad (1.7)$$

де $g(t)$ - тренд ряду (логістична або кусково-лінійна апроксимація даних), $s(t)$ - сезонна складова, апроксимована рядом Фур'є, $h(t)$ - складова, яка враховує вплив свят чи інших аномалій, які відбуваються нерегулярно протягом одного або кількох днів і діють з певним «вікном», тобто, в діапазоні певних дат, ε_t - похибка («шум») з нульовим середнім, розподілена за нормальним законом.

Модель Prophet враховує аномалії та ефективно працює з пропущеними даними (для даних моніторингу, за умов відключення світла чи Інтернету, це – актуальна проблема). З власного досвіду відомо, що модель може досить ефективно працювати з параметрами за замовчуванням [8, 14], але все ж кращої ефективності можна досягти, за рахунок використання багатьох параметрів, передусім, щодо аномалій даних та для опису сезонних складових.

Популярними для прогнозування часових рядів, також, є рекурентні нейромережеві моделі, зокрема їх підвид LSTM (Long Short-Term Memory) [53]. Головна перевага LSTM полягає в її здатності запам'ятовувати довгострокові залежності, завдяки спеціальній архітектурі. Основними компонентами LSTM є комірка пам'яті (елемент, що зберігає інформацію), приховані стани (вихідні дані) та вентиля (механізми, які контролюють процес обробки інформації). Існують три типи вентилів [53]:

– забувальний вентиль f_t , що керує тим як довго значення залишається в пам'яті:

$$f_t = \sigma(W_f \times |h_{t-1}, x_t| + b_f), \quad (1.8)$$

де σ – сигмоїдна функція, W_f – матриця ваг, h_{t-1} – прихований стан попереднього кроку, x_t – поточний вхідний вектор, а b_f – вектор зміщення.

– вхідний вентиль i_t , що керує тим, як нове значення потрапляє в пам'ять:

$$\begin{aligned} i_t &= \sigma(W_i \times |h_{t-1}, x_t| + b_i), \\ \tilde{C}_t &= \text{tahn}(W_c \times |h_{t-1}, x_t| + b_c), \end{aligned} \quad (1.9)$$

де \tilde{C}_t – кандидат на оновлення комірки, W_i та W_c – матриці ваг, h_{t-1} – прихований стан попереднього кроку, x_t – поточний вхідний вектор, а b_i та b_c – вектори зміщення.

– вихідний вентиль o_t , що керує використанням значення для активування виходу блоку [53]:

$$o_t = \sigma(W_o \times |h_{t-1}, x_t| + b_o), \quad (1.10)$$

де σ – сигмоїдна функція, W_o – матриця ваг, h_{t-1} – прихований стан попереднього кроку, x_t – поточний вхідний вектор, а b_o – вектор зміщення.

Модель може запам'ятовувати інформацію з довгих послідовностей, що є важливим для прогнозування часових рядів. За рахунок механізму вентилів модель досить стійка до зашумлених даних, оскільки вентиля дозволяють ігнорувати нерелевантну інформацію.

В таблиці 1.8 наведено узагальнений результат порівняльного аналізу трьох згаданих вище моделей.

Таблиця 1.8 – Порівняльний аналіз моделей прогнозування часових рядів

Параметр	ARIMA	Prophet	LSTM
Принцип роботи	Статистична модель, що відокремлює сигнал від білого шуму, та робить прогноз для сигналу.	Модель, яка розкладає часовий ряд на тренд, сезонність та святкові дні	Рекурентна нейронна мережа, яка може навчатися довгостроковим залежностями
Переваги	Мала обчислювальна потужність	Стійкість до аномалій, невелика обчислювальна потужність	Ефективна з нелінійними часовими рядами та з довгостроковим залежностями
Недоліки	Вимагає стаціонарність даних, залежить від параметрів	Менш ефективна при складних нелінійних залежностях	Вимагає великих обсягів даних, схильна до перенавчання

Продовження таблиці 1.8

Параметр	ARIMA	Prophet	LSTM
Ефективна коли	Коли ряд лінійний та стаціонарний	Коли присутня сильна сезонність	Коли ряд складний та нелінійний

Як видно в таблиці 1.8, усі розглянуті моделі мають свої переваги та недоліки. Виходячи з проведеного аналізу і власного досвіду, модель Prophet є оптимальним варіантом для прогнозування даних про стан природних середовищ.

Важливим елементом аналізу часових рядів, та загалом даної роботи, є аналіз та врахування аномалій, відповідно, необхідно провести аналіз методів та засобів пошуку аномалій.

1.4. Аналіз методів та засобів пошуку аномалій в часових рядах

Як було зазначено в попередньому підрозділі, одним з етапів аналізу часових даних є пошук аномальних значень – подій або елементів, які не відповідають певним шаблонам поведінки або іншим елементам даного набору. Процес виявлення аномалій є фундаментальною дослідницькою проблемою, якою займаються науковці різних країн, відповідно, не дивним є те, що наукова література надає досить широкий перелік існуючих методів та підходів [54]. Вдале врахування аномалій, як правило, суттєво підвищує точність аналізу та прогнозування даних.

Розглянемо відомі методи пошуку аномалій та шляхи автоматизації процесів пошуку, за рахунок використання мови програмування Python.

1.4.1. Аналіз відомих методів пошуку аномалій

Популярним методом пошуку аномалій в задачах аналізу часових рядів є Isolation Forest. Isolation Forest – це алгоритм виявлення аномалій, який працює шляхом ізоляції аномалій у даних [55, 56]. Алгоритм роботи даного методу виглядає наступним чином [56]:

- 1) З вхідного набору даних випадковим чином обирається підвибірка;
- 2) Обирається випадкова ознака та поріг – випадкове значення між мінімальними та максимальними значеннями обраної ознаки;
- 3) Точка набору даних порівнюється з пороговим значенням. Якщо воно менше порогу, то точка додається у ліву гілку дерева, якщо ж більше – у праву;
- 4) Повторюємо крок 3, поки усі точки набору даних не будуть розділені в окремі гілки дерева або поки не буде досягнуто заданої максимальної глибини дерева;
- 5) Після створення дерев, для кожного елементу рахується середня довжина шляху до гілок дерева;
- 6) Визначаються аномалії за рахунок перевірки значень середньої довжини шляху з певним порогом.

Приклад графічного подання результатів використання методу наведено на рисунку 1.9.

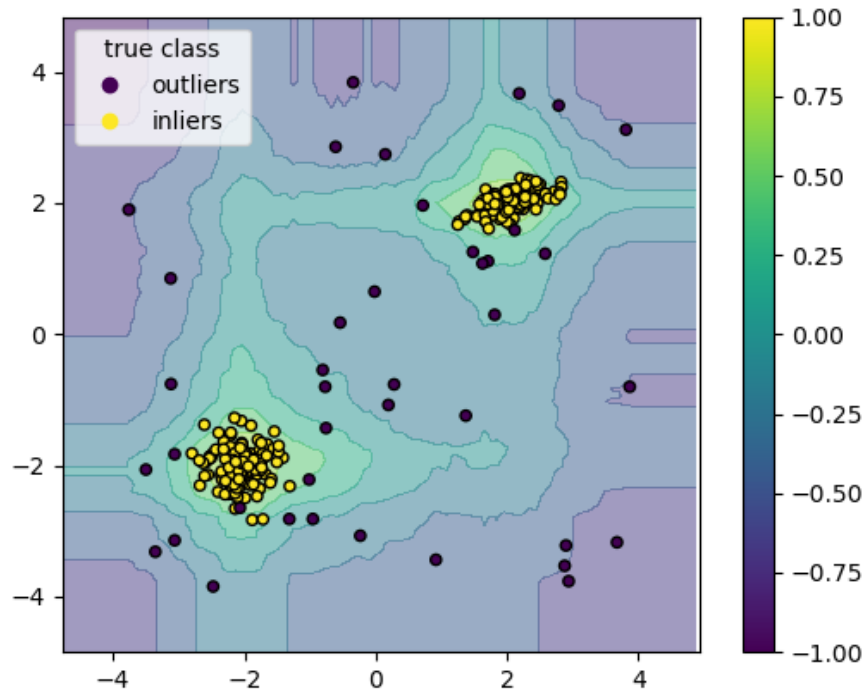


Рисунок 1.9 – Результати роботи методу Isolation Forest (світлими точками позначені нормальні значення, а темними – аномалії) [57]

На рисунку 1.8 видно, що метод дозволив ізолювати світлі точки в два кластери за рахунок довжини шляху в дереві, а темні точки, в свою чергу, віддалені від цих кластерів і можуть бути аномаліями.

Цей метод може ефективно працювати з великими наборами даних, стійкий до шуму та може досить ефективно знаходити аномальні значення. В той же час, метод може бути схильний до перенавчання за великої кількості дерев, що у свою чергу, може призвести до зниження точності роботи методу.

Іншим досить розповсюдженим методом є метод k -найближчих сусідів. Метод досить поширений в сфері машинного навчання, в першу чергу, в задачах класифікації та регресії, проте може використовуватися і для пошуку аномалій [58]. Основний принцип методу полягає в знаходженні відстані від точки до k -найближчих сусідів. Для визначення відстані, зазвичай, застосовують евклідову відстань, хоча в певних випадках застосовують інші метрики, наприклад, відстань Геммінга [58]. Знаходження евклідової відстані d для двох точок можна записати наступною відомою формулою [58]:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}, \quad (1.11)$$

де x_1 та y_1 - координати першої точки, а x_2 та y_2 – відповідні координати другої точки.

Перевірка чи є певна точка аномальною полягає в порівнянні середнього значення відстаней з певним пороговим значенням. Приклад результату застосування методу наведено на рисунку 1.10.

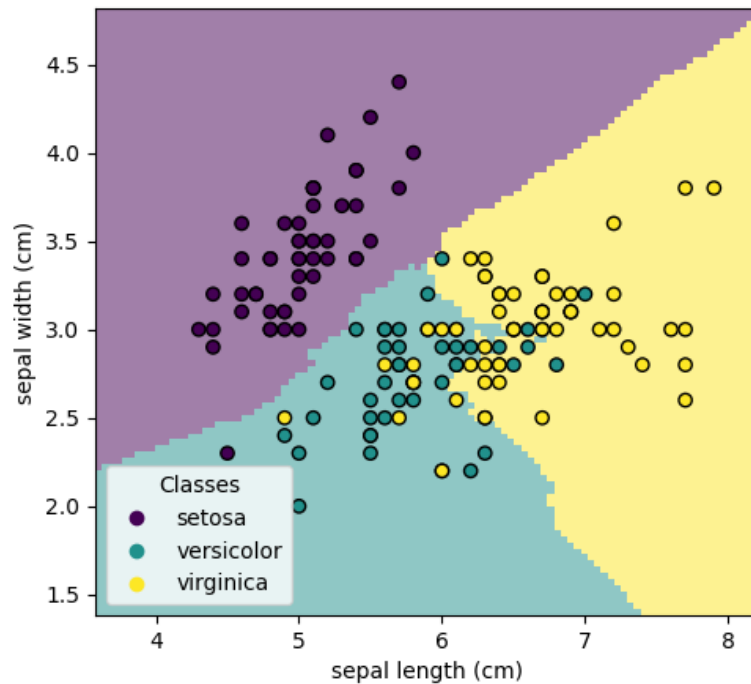


Рисунок 1.10 – Результати роботи методу k -найближчих сусідів (точка певного кольору позначає приналежність до відповідного класу цього самого кольору) [57]

Цей метод може працювати з різними типами даних та при неоднорідному розподілі даних. Головним недоліком є залежність якості роботи від налаштувань, а саме параметру k та порогового значення. Вибір оптимального значення параметру k вимагає гарного знання предметної області та, потенційно, застосування таких технік як перехресна валідація. Також в літературі рекомендують для визначення параметру k знаходити квадратний корінь від розміру усього набору даних або перебирати усі варіанти і вибрати найкращий.

Широкого застосування для пошуку аномалій мають і методи, які базуються на концепції щільності. Серед таких методів можна виділити коефіцієнт локального відхилення (англійською «Local outlier factor» або просто «LOF») – алгоритм, який базується на наведеному вище методі та дозволяє знайти локальну густину точок в наборі [58, 59]. Суть методу полягає в наступному:

1) Для кожної точки набору знаходить k -distance – відстань до k -найближчого сусіда (параметр k задається при створенні моделі);

2) Отримане значення використовується для знаходження відстані досяжності RD_k за формулою [59]:

$$RD_k(A, B) = \max \{k - \text{distance}(B), d(A, B)\}, \quad (1.12)$$

де $d(A, B)$ – відстань від об'єкту A до об'єкту B , а $k - \text{distance}$ – відстань об'єкту B до його k -найближчого сусіда;

3) Обраховується локальна щільність досяжності об'єкту lrd_k , яка є оберненою до середньої відстані досяжності відповідного об'єкту до його сусідів. Це позначається за допомогою [59]:

$$lrd_k(A) = 1 / \left(\frac{\sum_{B \in N_k(A)} RD_k(A, B)}{|N_k(A)|} \right), \quad (1.13)$$

де $N_k(A)$ – набір k -найближчих сусідів об'єкту A , а $RD_k(A, B)$ – відстань досяжності від об'єкту A до об'єкту B ;

4) Знаходиться коефіцієнт локального відхилення (LOF_k) шляхом відношення середнього значення lrd від k -кількості сусідів точки до lrd цієї точки. Це можна позначити за допомогою наступного виразу [59]:

$$LOF_k(A) = \frac{\sum_{B \in N_k(A)} \frac{lrd_k(B)}{lrd_k(A)}}{|N_k(A)|} = \frac{\sum_{B \in N_k(A)} lrd_k(B)}{|N_k(A)| * lrd_k(A)}, \quad (1.14)$$

де $lrd_k(B)$ - локальна щільність досяжності об'єкту B , а $lrd_k(A)$ - локальна щільність досяжності об'єкта A ;

5) Отримане значення, що дорівнює приблизно 1, вказує на те, що об'єкт можна порівняти зі своїми сусідами. Значення нижче 1 вказує на більш щільну область (що вказує на нормальну точку), тоді як значення, значно більші за 1, вказують на аномальні точки.

Метод дозволяє знаходити аномалії з урахуванням локальних властивостей набору даних і при цьому може працювати з різними законами розподілу даних. Загалом, він дає кращі результати за глобальний підхід до пошуку, проте немає загального порогового значення LOF , тому вибір допустимого відхилення – не

найпростіша задача. Приклад результатів застосування методу наведено на рисунку 1.11.

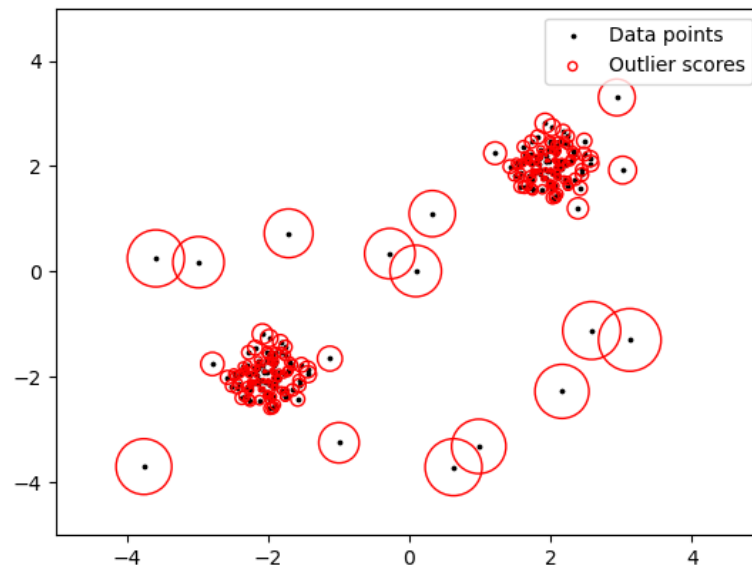


Рисунок 1.11 – Результати роботи методу LOF [57]

Досить цікавим методом є Seasonal Hybrid ESD, який дозволяє знаходити аномалії в часових рядах з сезонністю [60]. Тест Грабса, також відомий як екстремальний студентизований тест відхилень (ESD), є тестом, який використовується для виявлення аномалій в одномірному наборі даних, підпорядкованому нормальному розподілу. Тест Грабса полягає в обчисленні G [60]:

$$G = \frac{\max_{i=1 \dots N} |Y_i - \bar{Y}|}{s}, \quad (1.15)$$

де \bar{Y} - це середнє значення вибірки, а s – стандартне відхилення.

Проблема тесту Грабса в тому, що він передбачає нормальний розподіл, проте дані далеко не завжди розподілені по такому закону. Метод Seasonal Hybrid ESD пропонує використовувати медіану та середнє абсолютне відхилення замість середнього та стандартного відхилення, оскільки вони дуже чутливі до великих та чисельних аномалій. Він комбінує сезонну адаптацію з експоненційно зваженою середньою (Exponential Smoothing) і використовує Extreme Studentized Deviate

(ESD) для виявлення відхилень від очікуваного розподілу даних. Перевагами методу такі:

- урахування сезонності: враховує сезонні зміни в часових рядів, що дозволяє ефективно виявляти аномалії, які повторюються у певних періодах часу;
- гнучкість та налаштування: метод може бути налаштований залежно від потреб користувача (період сезонності та порогові значення);
- підтримка широкого спектра даних: може бути застосований до різних типів даних, в тому числі, часових рядів.

Серед недоліків можна виділити такі:

- обмежена робота з незвичайними аномаліями: метод краще працює з аномаліями, які досить схожі на звичайні сезонні зміни;
- вимоги до тривалості обчислень: алгоритм має не найбільшу швидкість, особливо для великих обсягів даних або довгих часових рядів.

В таблиці 1.9 наведено узагальнення результатів порівняльного аналізу згаданих методів.

Таблиця 1.9 – Порівняльний аналіз моделей прогнозування часових рядів

Метод	Isolation Forest	k-NN	LOF	SESD
Принцип роботи	Використовує дерева для ізоляції точок набору	Обчислює відстань до k найближчих сусідів; точки, далекі від інших, є аномаліями	Визначає аномалії за локальною щільністю точок	Використовує сезонну декомпозицію з ESD-тестом
Переваги	Ефективний для багатовимірних даних	Простота реалізації, добре працює на невеликих вибірках	Враховує локальні особливості	Ефективний щодо виявлення сезонних аномалій

Продовження таблиці 1.9

Метод	Isolation Forest	k-NN	LOF	SESD
Недоліки	Менш ефективний при складних залежностях у даних	Чутливість до вибору k	Чутливий до вибору параметрів, велика обчислювальна потужність	Може бути менш ефективним для несезонних даних

Як видно з таблиці 1.9 та за результатами аналізу загалом, існує велика кількість різноманітних методів та підходів до пошуку аномалій в часових рядах, в тому числі даних стану природного середовища. Кожен з методів має як свої переваги, так і свої недоліки. При цьому, не існує якогось універсального методу, який буде ефективним в будь-якій ситуації. Тому має сенс розробляти нові методи або застосовувати ансамблювання методів різними стратегіями комбінування, що дозволить усувати недоліки одного методу за рахунок використання іншого.

Розглянемо засоби автоматизації описаних методів пошуку аномалій засобами мови програмування Python.

1.4.2. Огляд засобів для автоматизації пошуку аномалій в часових рядах

Як було зазначено вище, існує досить великий перелік метод та підходів для пошуку аномалій в часових рядах. В переважній більшості випадків науковці не проводять необхідні розрахунки вручну, а застосовують автоматизовані шляхи. Мова програмування Python беззаперечно є найбільш популярною сучасною мовою для вирішення задач системного аналізу. Однією з основних причин цієї популярності є наявність широкого набору бібліотек та інструментів, які, в свою чергу, програмно реалізують усі раніше згадані методи. Виходячи з цього, варто розглянути декілька подібних бібліотек та проаналізувати їх ефективність.

Для огляду буде використано дані моніторингу стану атмосферного повітря за даними мережі EcoCity. Дані для дослідження отримані за допомогою сервісу «Кабінет дослідника» [42], який був детально описаний в пункті 1.1.4. В цьому конкретному випадку було обрано показник «PM10» (пил, розміром 10 мкм і менше), та дані, отримані з однієї зі станцій у Вінницькій області. На рисунку 1.12 зображено графік даного часового ряду.

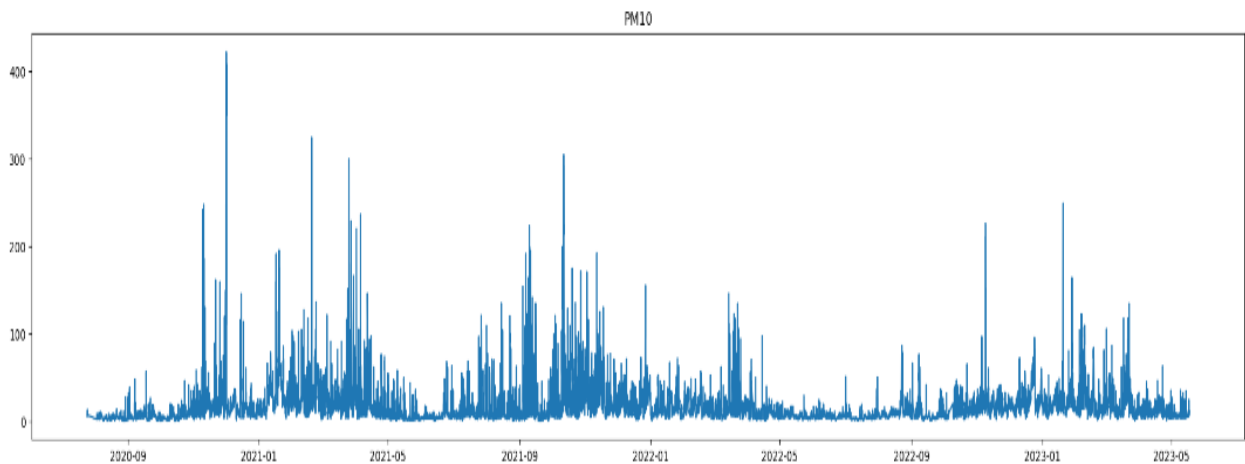


Рисунок 1.12 – Вхідні дані для передбачення аномалій за показником PM10 у Вінницькій області [13]

Проаналізуємо можливості Python-бібліотеки `sesd`. Дана бібліотека реалізує Метод Seasonal Hybrid ESD, який був детальніше описаний в попередньому пункті. Приклад коду для виявлення аномалій, використовуючи бібліотеку `sesd` (Seasonal Hybrid ESD), наведено на рисунку 1.13, а результат його роботи – на рисунку 1.14.

```
anomaly_df = df['y']
outliers_indices = sesd.seasonal_esd(anomaly_df,
                                     hybrid=True,
                                     max_anomalies=2500,
                                     alpha = 3)
anomalies = df.loc[outliers_indices]
```

Рисунок 1.13 – Приклад використання бібліотеки `sesd` [13]

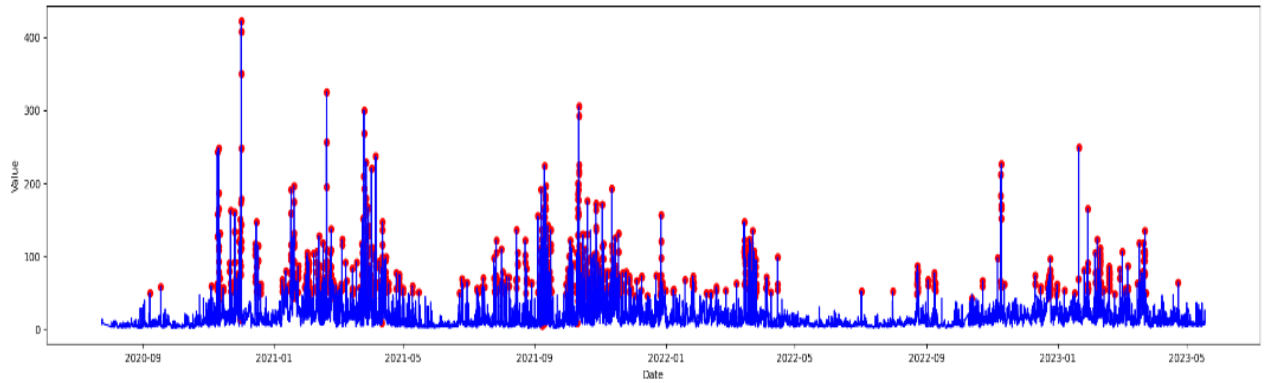


Рисунок 1.14 – Результат роботи бібліотеки `sesd` [13]

Наступною розглянемо бібліотеку `scikit-learn` [57]. Дана бібліотека містить велику кількість методів і моделей машинного навчання як для задач регресії і кластеризації, так і для пошуку аномалій. В даному огляді обмежимося модулем `Isolation Forest`, який реалізує однойменний метод пошуку аномалій, який був детально описаний в попередньому пункті. Приклад коду застосування даного методу наведено на рисунку 1.15, а результат роботи – на рисунку 1.16.

```
model = IsolationForest(contamination=0.05)
model.fit(df['y'].values.reshape(-1, 1))
y_pred = model.predict(df['y'].values.reshape(-1, 1))
```

Рисунок 1.15 – Приклад використання модуля `Isolation Forest` Python- бібліотеки `scikit-learn` [13]

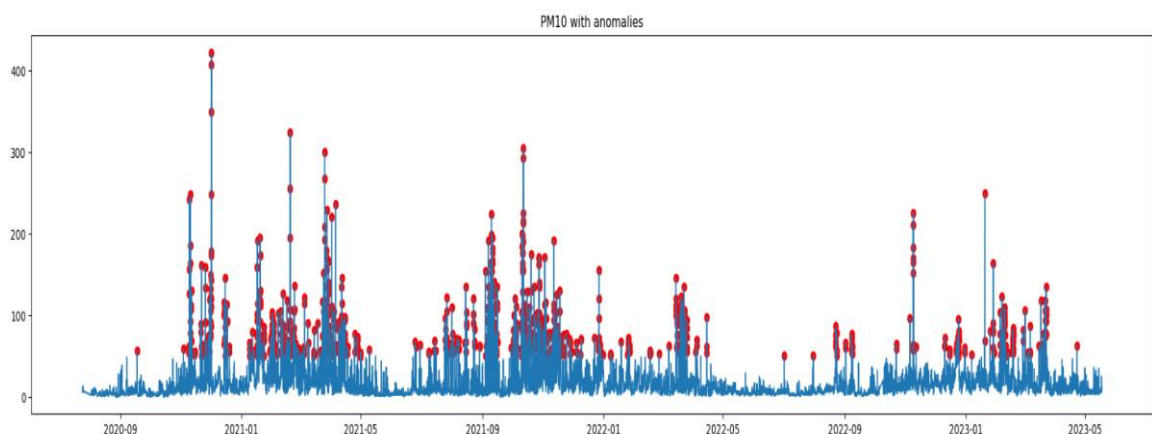


Рисунок 1.16 – Результат роботи методу `Isolation Forest` [13]

Розглянемо бібліотеку statsmodels [61]. У statsmodels немає вбудованих функцій для прямого виявлення та фільтрації аномалій у часових рядах. Однак, можна використовувати інші функції та методи з бібліотеки для реалізації такого аналізу. Один з підходів - це аналіз залишків. Можна використовувати статистичні моделі, такі як ARIMA або ETS, для прогнозування часового ряду. Порівнюючи фактичні значення з прогнозованими, можна виявити аномалії, коли значення значно відрізняються від прогнозу. Інший підхід – це використання статистичних методів. Прикладами таких є встановлення порогових значень, використання стандартного відхилення, Z-перетворення та інших методів для виявлення значень, які перевищують певний поріг.

Серед переваг використання statsmodels наявність статистичних моделей та гнучкість у виборі підходу для виявлення аномалій. Однак, для ефективного використання бібліотеки потрібні знання статистики та часових рядів. Спробуємо знайти аномалії, використовуючи дачі часового ряду (рис. 1.17) –результати наведено на рисунку 1.18.

```

period = 24
result = seasonal_decompose(df_anomaly['y'], model='additive', period=period)
residual = result.resid
threshold = 200.0
anomalies = df_anomaly[residual.abs() > threshold]
df_filtered = df_anomaly.drop(anomalies.index)

```

Рисунок 1.17 – Приклад використання Python-бібліотеки statsmodels [13]

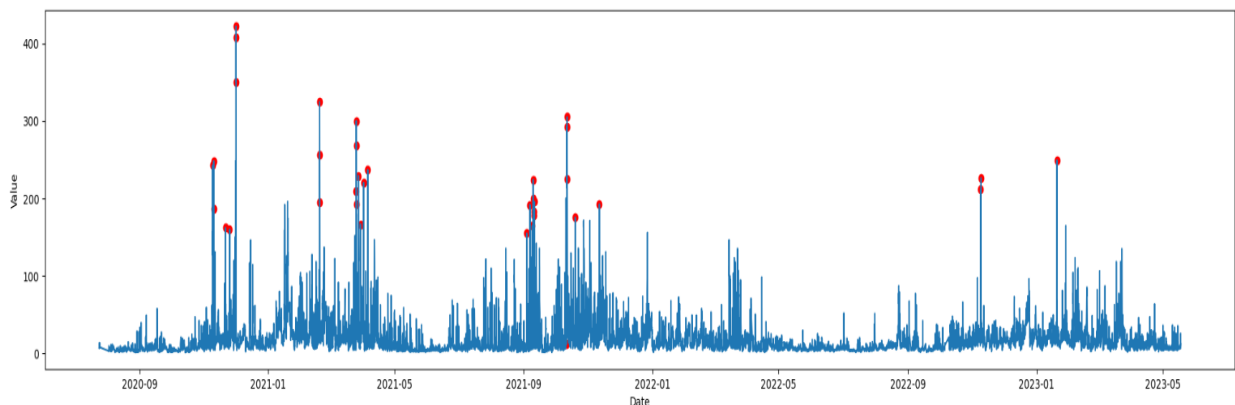


Рисунок 1.18 – Результати, отримані за допомогою Python-бібліотеки statsmodels [13]

Як видно з аналізу, мова програмування Python має великий перелік бібліотек, серед яких ті, що були безпосередньо описані в даному пункті. Це дозволить швидше проводити дослідження щодо пошуку аномалій відомими методами, оскільки усуває потребу в програмній реалізації відповідного методу.

1.5. Висновки до розділу та постановка задач дослідження

В розділі, на прикладі України, проаналізовано сучасний стан природних середовищ, а саме стану атмосферного повітря та водних ресурсів, проаналізовано існуючі директиви, державні та громадські механізми моніторингу та контролю за їх станом. Аналіз показав, що в Україні успішно функціонують різні державні і громадські мережі моніторингу. Показники стану проаналізованих природних середовищ ж демонструють певну позитивну тенденцію, але загалом є нестабільними і вимагають подальшого аналізу.

Проведено аналіз систем управління станом природного середовища, що використовуються у світі. Визначено переваги та недоліки застосування відповідних підходів та було визначено, що модель DPSIR може бути основою для подальшого системного аналізу, оскільки вона є гнучкою для всебічного аналізу природного середовища. Проведено порівняльний аналіз методів та моделей, у тому числі інтелектуальних, прогнозування часових рядів стану природних середовищ з урахуванням методів пошуку аномалій в цих рядах. Крім того, проаналізовано шляхи автоматизації пошуку аномалій в часових рядах за рахунок використання мови програмування Python. Аналіз показав, що відсутній універсальний метод пошуку аномалій і необхідно розробляти нові або певним чином покращувати існуючі.

Отже, для досягнення мети дисертаційного дослідження, яка полягає у підвищенні точності прогнозування стану природних середовищ з урахуванням аномалій шляхом створення методів та технології системного аналізу, необхідно розв'язати наступні задачі:

1) Охарактеризувати систему управління станом природного середовища з позиції системного аналізу: провести декомпозицію системи, виділити невизначеності, що мають місце;

2) Проаналізувати можливі методи та підходи розкриття описаних видів невизначеності, що дозволить з'ясувати, які методи потребують удосконалення або які нові авторські методи необхідно розробити;

3) Розробити або удосконалити методи та підходи для розкриття основних видів невизначеності системи управління станом природних середовищ різного типу з урахуванням впливу аномалій;

4) Узагальнити усі методи розкриття видів невизначеності та розробити технологію системного аналізу стану природних середовищ з урахуванням аномалій, що базується на використанні класичних та авторських методів для розкриття основних видів невизначеності та дозволяє підвищити точність прогнозування цього стану.

5) Перевірити ефективність та впровадити розроблену технологію системного аналізу на реальних прикладах аналізу та прогнозування даних у система управління станом природних середовищ.

Розв'язанню цих задач будуть присвячені наступні розділи та підрозділи дисертаційної роботи.

Описаний в розділі матеріал висвітлений в таких роботах: [13, 16].

РОЗДІЛ 2

СИСТЕМНИЙ АНАЛІЗ СИСТЕМИ УПРАВЛІННЯ СТАНОМ ПРИРОДНИХ СЕРЕДОВИЩ

2.1. Характеристика складної системи

Охарактеризуємо складну систему, яка досліджується, з використанням методології системного аналізу. Системний аналіз, передусім, застосовується до складних систем, тому необхідно спочатку сформулювати що являє собою складна система у цій задачі [1, 62]. Після цього необхідно провести декомпозицію цієї складної системи та виділити які види невизначеності мають місце в таких системах і визначити як саме ці види невизначеності можна розкрити. Цьому присвячені наступні пункти цього підрозділу.

Пропонується розглянути систему управління стану природного середовища, відштовхуючись від згаданої в підрозділі 1.2 моделі «DPSIR – модель, що активно застосовується країнами-членами Європейському союзу та безпосередньо використовується для розв’язання різних задач, пов’язаних з управлінням стану природних середовищ [48] (стан атмосферного повітря, екологічний стан водних ресурсів, тощо). При цьому, як зазначено раніше, ця робота присвячена системному аналізу системи, яка зосереджена, в першу чергу, на врахуванні аномалій або пікових значень показників у системі, тому необхідно враховувати це на наступних етапах аналізу.

Модель DPSIR складається з п’яти взаємопов’язаних компонентів, як це графічно наведено на рисунку 2.1.

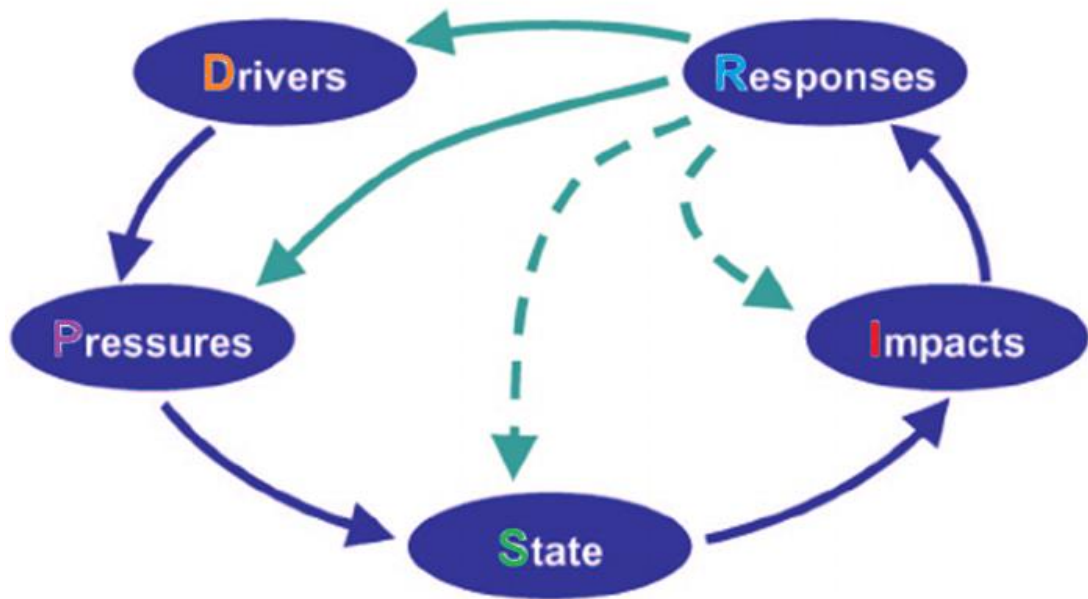


Рисунок 2.1 – Структурна схема моделі «DPSIR»

Як видно, з рисунку 2.1, прийняття рішень впливає та залежить від [48]:

- вибору місць розташування та вибору складу продуктивних сил (заводів, електростанцій, сільськогосподарських полів тощо);
- вибору місць розташування, очисних систем та способів викидів газів або скидання вод та обґрунтування гранично допустимих дозволів щодо їх обсягу і допустимих забруднювальних речовин тощо, які характеризують безпосередній тиск продуктивних сил на природне середовище;
- вибору місць розташування елементів та параметрів систем моніторингу стану природного середовища: станцій спостереження за якістю або екологічним станом води, станцій моніторингу атмосферного повітря тощо;
- вибору рекомендацій для органів влади, екоінспекцій та інших контрольних установ для підвищення ефективності та оперативності контролю, рекомендацій для населення щодо зменшення негативного впливу на них від поганого стану природного середовища.

Базуючись на даній структурі складної системи, проведемо декомпозицію системи за різними критеріями декомпозиції.

2.2.1. Декомпозиція складної системи за різними критеріями

Наступним етапом системного аналізу є декомпозиція складної системи на підсистеми та елементи [1, 62]. Загалом, існує велика кількість різних видів декомпозиції, але пропонується застосувати наступні види декомпозиції системи управління станом природного середовища з урахуванням, передусім, впливу аномалій чи пікових значень показників або параметрів:

- *функціональна* (виділення підсистем за функціями чи їх призначенням): D_{DPSIR} (D_D – джерела-забруднювачі, D_P – види навантаження, D_S – стан системи, D_I – наслідки, D_R – реакція);

- *просторову* (за просторовими зв'язками та географічними координатами): D_G – зон (Вінницька область), агломерацій (м. Вінниця або агломерація «Вінницька»), локальну (мікрорайон Вишенька), організацій (ВНТУ), ділянок річки – у цій роботі основна увага буде приділятися рівню зон (областей) D_{G-O} ;

- *часову*, оскільки різні аномалії можуть бути різними у різні інтервали часу: $D_T(\Delta T)$ – характеристика різних підсистем та вибірки спостережень різних показників з узагальненням за інтервал часу ΔT – 1 хв., 20 хв., годину, добу, тиждень, місяць, рік тощо;

- *сезонну*, оскільки повторюваність інтервалів часу дозволяє зрозуміти що це – аномалія чи лише сезонне пікове значення: $D_{SEC}(\Theta)$ – з сезоном, тривалість Θ – добова, тижнева, місячна, квартална, річна тощо.

Виділимо та охарактеризуємо види невизначеності, які мають місце для цих складових.

2.2.2. Класифікація видів невизначеності у складній системі

Для кожного виду декомпозиції, описаних у попередньому пункті, визначимо та коротко охарактеризуємо які види невизначеності мають місце у складних системах управління станом природних середовищ регіону з

урахуванням, передусім, впливу аномалій чи пікових значень показників або параметрів.

Проаналізуємо види невизначеності для складових декомпозиції D_{DPSIR} :

- D_D – джерела-забруднювачі:
 - *інформаційна* (неповна чи неточна інформація про місця розташування підприємств-забруднювачів – U_{I1} або про найбільші обсяги гранично допустимих значень забруднювальних речовин у викидах чи скидах) – U_{I2} ;
- D_P – види навантаження:
 - *інформаційна* (неповна чи неточна інформація про місця розташування джерел викидів підприємств-забруднювачів (скидань) та про фактичні викиди (скидання) з форми 2-ТП (повітря) (2-ТП (водгосп))) – U_{I3} ;
 - *ситуаційна* (відсутність достовірної інформації про перелік усіх забруднювачів, через те, що якесь підприємство здійснює викид (скидання), але не подає про це звітність, чи фальсифікує інформацію про реальні викиди (скидання)) – U_{S1} ;
- D_S – стан системи:
 - *інформаційна* про дані моніторингу (відсутні чи неточні дані щодо значень показників стану природного середовища) – U_{I4} ;
 - *інформаційна* про мережу системи моніторингу (відсутні чи неточні дані про пости мережі спостережень, їх показники, регулярність вимірювань та метрологічні аспекти) – U_{I5} ;
 - *цільова* для системи збирання даних спостережень (невизначеність вибору кількості і місць розташування існуючих чи нових постів моніторингу) – U_{T1} ;
 - *цільова* для системи комплексного оцінювання якості (екологічного стану) природного середовища (невизначеність щодо критеріїв якості природного середовища в цілому та їх складових) – U_{T2} ;

- *цільова* для аналізу аномальних та пікових значень показників стану природного середовища (як правило, аномалії фільтрують за доланням порогу – має місце невизначеність щодо вибору варіантів цього порогу) – U_{T3} ;
- *модельна* (невизначеність щодо варіантів моделей складної системи) – U_{M1} ;
- *параметрична* (невизначеність щодо значень параметрів моделей) – U_{P1} ;
- D_I – наслідки або вплив:
- *структурна* (неповне знання про елементи системи або неточність уявлень про зв'язки між ними) – U_{ST1} ;
- *сценарна* (відсутність чи неточність прогнозів динаміки системи для різних сценаріїв розвитку) – U_{SC1} ;
- D_R – реакція або прийняття управлінських рішень:
- *інформаційна* (неповна чи неточна інформація про вже прийняті рішення) – U_{I6} ;
- *цільова* (невизначеність вибору рішень) – U_{T4} .

Види невизначеності для складової D_{G-O} , тобто – для системи управління станом природного середовища зони (наприклад, адміністративної області України або ділянки річки Південний Буг) на додаток до наведених вище:

- *структурна* (неповне знання про взаємозв'язок між різними показниками стану природного середовища на різних постах спостережень – наявність цих знань дозволила б за даними сусідніх постів здійснювати прогнозування одних показників по інших, що дозволило б розкрити інші види невизначеності: інформаційну невизначеність щодо зростання кількості хворих на алергію у регіоні на основі знань про алергічну сенсibilізацію до різних груп алергенів в атмосферному повітрі, коли зростання концентрації одного алергену збільшує ймовірність захворювання і щодо інших алергенів) – U_{ST2} .

Види невизначеності для складових $D_T(\Delta T)$, на додаток до наведених вище:

- *ситуаційна* (у різні інтервали часу можуть бути різні аномалії чи пікові значення, які важко виявляти) – U_{S2} ;
- *цільова* (як правило, аномалії фільтрують за доланням порогу – має місце невизначеність щодо вибору варіантів цього порогу) – U_{T5} .

Види невизначеності для складових $D_{SEC}(\Theta)$, на додаток до наведених вище:

- *цільова* (вибір того, яка саме сезонна складова або їх сукупність присутні) – U_{T6} ;
- *параметрична* (невизначеність щодо параметрів моделей, які описують вплив сезонної складової: період, вікно впливу, коефіцієнт регуляризації або сила впливу) – U_{P2} .

Необхідно з'ясувати та описати яким чином виділені види невизначеності можуть бути розкриті. Це дасть розуміння які саме методи потребують удосконалення або розробки. Цьому присвячений наступний підрозділ роботи.

2.2. Підходи та методи щодо розкриття видів невизначеності у складній системі

У підрозділі 2.1 здійснено декомпозицію складної системи, яка досліджується, за різними критеріями та виділено види невизначеності для усіх елементів декомпозиції. Опишемо яким чином можна мінімізувати або навіть повністю усунути види невизначеності, що були виділені у попередньому пункті.

Для розкриття інформаційної невизначеності про джерела-забруднювачі D_D та види навантаження D_P , достатньо подати запит до Державної служби статистики України та органів влади щодо інформації із форми 2ТП (повітря) або 2ТП (водгосп) про джерела забруднення та інформацію про розташування їх джерел викиду [63]. На жаль, там буде тільки офіційна інформація, яка потребує

уточнення щодо незареєстрованих джерел забруднення або уточнення фактичних викидів. Для цього варто розробити метод, який дозволить оцінювати можливий вплив додаткових джерел забруднення та здійснити їх локалізацію у просторі та часі, точніше – визначити координати та в які роки (чи місяці) їх додатковий вплив був відчутним. Це дозволить розкрити ситуаційну невизначеність щодо D_p .

Крім того, інформаційну невизначеність можна розкрити за рахунок використання даних дистанційного зондування Землі та даних супутникової зйомки, що дозволяє покрити великі території (рис 2.2). Проте, варто зауважити, що в сучасних умовах можуть бути обмеження на використання подібних ресурсів інформації. Проте, це не відмінняє корисність даного підходу і, за можливості, його варто розглядати.

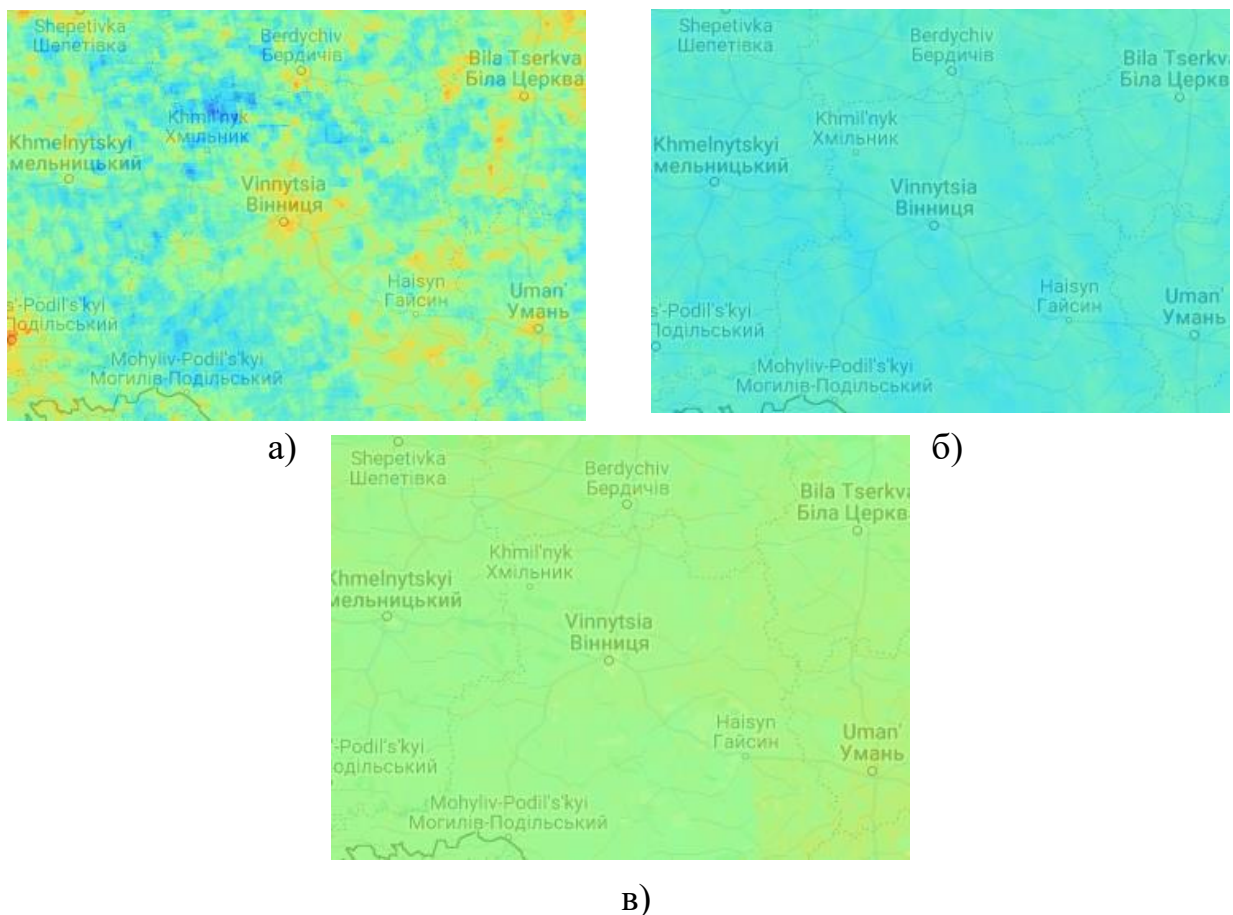


Рисунок 2.2 – Мапи шарів даних щодо стану атмосфери, отриманих супутником у тропосфері Вінницької зони: а) діоксид азоту, б) оксид вуглецю, в) озон [63]

Додатково можна залучати безпілотні літальні апарати, що дозволить доволі швидко перевірити певну ділянку з метою пошуку можливих джерел забруднення атмосферного повітря, водних ресурсів тощо. Подібно до супутникових знімків, зараз можуть бути певні обмеження на їх використання, але за можливості аналітики можуть залучати і такі засоби аналізу.

Щодо розкриття інформаційної невизначеності для D_S , достатньо отримати інформацію з систем моніторингу [63]. Краще використовувати інформацію з державної системи моніторингу, але, за відсутності такої, у першому наближенні, можна використовувати й дані громадської системи моніторингу, які в Україні доступні на вище згаданих порталах SaveEcoBot [43] та EcoCity [28]. Однак, цієї інформації може бути недостатньо, тоді варто розширяти мережу, передусім державного моніторингу, з виведенням інформації на порталах державних органів влади та місцевого самоврядування. Такі пости (станції) є доволі дорогими, тому потрібна система критеріїв для визначення їх місць розташування, показників спостережень, регулярності вимірювань та пріоритетності їх монтажу. Для цього потрібно розробити підхід формування такої системи критеріїв, орієнтовану на аномальні та максимальні показники складових моделі DPSIR [48]: де вже є чи будується найбільша кількість (щільність у просторі) джерел забруднення, де найбільше проживає населення, зокрема й вразливого, де мають місце найбільші викиди тощо. Це дозволить розкрити й цільову невизначеність для системи збирання даних спостережень для складової D_S . Застосування такої системи критеріїв тісно пов'язано з розкриттям інших двох підвидів цільової невизначеності для цієї складової:

1) цільова невизначеність для системи комплексного оцінювання якості природного середовища – пропонується як комплексну оцінку використовувати загально прийнятий в ЄС індекс атмосферного повітря або індекс якості води;

2) цільова невизначеність для аналізу аномальних та пікових значень показників – пропонується використовувати директиву 2008/50/ЄС Європейського парламенту та ради [64] та постанову Кабінету Міністрів України №827 [31] для аналізу найбільших та середніх значень, але треба деталізувати цей

підхід.

Для розкриття модельної невизначеності щодо складової D_S , тобто побудови оптимальної моделі для моделювання і прогнозування даних моніторингу стану природного середовища, потрібно проаналізувати яку модель прогнозування доцільно застосовувати в даній задачі. Особливу увагу слід приділити питанням розкриття параметричної невизначеності щодо врахування аномалій, визначення видів та параметрів складових для урахування сезонності цих моделей. Причому, під час розкриття цієї параметричної невизначеності важливо приділити увагу розкриттю ще й таких видів невизначеності:

- 1) ситуативної та цільової невизначеності для складових $D_T(\Delta T)$ щодо того як виявляти аномалії та пікові значення на різних інтервалах часу;
- 2) цільової та параметричної невизначеності для складових $D_{SEC}(\Theta)$ щодо того яка саме сезонна складова або їх сукупність мають місце та які їх параметри.

З метою кращого розуміння підходів щодо розкриття усіх цих видів невизначеності, пов'язаних з параметричною невизначеністю, варто вибрати регіон для аналізу, знайти для нього дані та провести їх розвідувальний аналіз, на якому відпрацювати усі методи і підходи.

Для більш повного розкриття сценарної невизначеності для складової D_I (та для складової D_{G-O} в цілому) потрібно більше інформації про усі взаємозв'язки в системі, але, як правило, така інформація є недостатньою. Мережа моніторингу в Україні, на жаль, досі не є достатньо розвиненою. Крім того, метеорологічні показники, як правило, не є відкритими даними і спостерігаються теж не дуже точно, через брак фінансування всієї галузі. А тому недостатність інформації можна спробувати компенсувати використанням інтелектуальних моделей та засобів для аналізу даних складової D_S і для прогнозування окремих показників стану природного середовища, і для оцінювання їх взаємовпливу, наприклад на основі байєсівського моделювання. Наявність моделей для прогнозування стану природного середовища окремих регіонів та моделей для оцінювання їх сукупного

впливу на інші регіони дозволить програвати сценарії зміни стану в цілому, що дозволить розкрити сценарну невизначеність для складової D_I .

Запит до органів влади дозволить розкрити інформаційну невизначеність щодо складової D_R про вже прийняті управлінські рішення, відображені у різного роду звітах, оглядах, прес-релізах. Цьому може допомогти й аналіз матеріалу на сайтах та сторінках у соцмережах органів влади і місцевого самоврядування, які опікуються екологічною політикою, зокрема й у сфері екоінспекційного контролю. А технології сучасного генеративного штучного інтелекту дозволять на основі цієї текстової інформації сформуванати множину можливих управлінських рішень, із подальшою валідацією групою експертів серед цих рішень можна налагодити вибір одним із відомих методів оптимізації. Це все дозволить розкрити й цільову невизначеність щодо складової D_R .

Систематизуємо описані підходи та методи у таблиці 2.1 та визначимо які варіанти розкриття невизначеності ще потребують удосконалення, а які методи необхідно розробити чи удосконалити.

Таблиця 2.1 – Методи розкриття невизначеності для різних видів декомпозиції системи управління станом природного середовища

Невизначеність	Підхід чи метод розкриття невизначеності
U_{T2}	Використання загально прийнятих критеріїв для тих мереж моніторингу, звідки надаються дані, наприклад, для мереж громадського моніторингу стану атмосферного повітря, це – індекс AQI
$U_{I1}, U_{I2}, U_{I3}, U_{I4}, U_{I6}$	Інформаційний запит щодо даних. Відкриті дані супутникових знімків або дистанційного зондування землі
U_{I5}	Інформаційні запити щодо системи моніторингу
U_{T3}, U_{T5}	Використання директив ЄС та інтегрованого з ним екологічного законодавства України
U_{T1}	Використання підходу оптимізації системи критеріїв на основі методу аналізу ієрархій
U_{S1}	Метод пошуку можливих джерел забруднення та ін.

Продовження таблиці 2.1

Невизначеність	Підхід чи метод розкриття невизначеності
U_{S2}	Метод ідентифікації локальних у часі аномалій та ін.
U_{M1}	Методи машинного навчання інтелектуальних моделей прогнозування
U_{P1}, U_{P2}, U_{T6}	Метод ідентифікації параметрів сезонності та аномалій та ін.
U_{ST1}, U_{ST2}	Байєсівське моделювання та інші методи
U_{SC1}	Програвання сценаріїв з використанням прогнозів побудованих оптимальних моделей та ін.
U_{T4}	Методи теорії прийняття рішень, генеративний штучний інтелект з використанням великих мовних моделей та ін.

Виходячи з усього описаного, можна зробити висновок, що для розкриття деяких видів невизначеності достатньо використати інформацію з відкритих джерел, таких як інформація мереж моніторингу, звіти відповідних статистичних органів тощо. Проте, для переважної більшості елементів необхідно розробити або удосконалити наступні методи:

1) розробити метод пошуку можливих незареєстрованих джерел підвищеного забруднення за даними точкового моніторингу стану природного середовища на основі аналізу аномалій, який повинен надавати можливість локалізувати у часі та просторі джерела сталого негативного ефекту на рівень забруднення, аналізуючи наявні дані моніторингу стану природного середовища;

2) запропонувати підхід формування системи інтегральних критеріїв для вибору параметрів місць нових постів спостережень державної системи моніторингу стану природного середовища регіону та визначення пріоритетності їх монтажу, з урахуванням аномалій та максимальних значень різних показників, які будуть братись до уваги;

3) розробити метод визначення параметрів інтелектуальних моделей, у тому числі параметрів сезонності та аномалій, для прогнозування даних про стан природного середовища, з урахуванням аномалій та сезонних особливостей. Метод повинен дозволити знайти та ефективно комбінувати декілька варіантів сезонності при роботі з інтелектуальними моделями прогнозування стану

природного середовища та підвищувати точність прогнозування;

4) удосконалити метод ідентифікації локальних у часі аномалій часового ряду показників природного середовища, що має дозволити ідентифікувати окремі інтервали часового ряду та виконувати пошук аномалій на усіх таких інтервалах. Використання цього методу повинно підвищити точність прогнозування цього ряду у різні інтервали часу;

5) запропонувати підхід до аналізу взаємозв'язків між різними показниками стану природного середовища різних регіонів на основі інтелектуальних моделей, наприклад, методу байєсівського моделювання;

Виконанню цих задач та розкриттю визначених видів невизначеності буде присвячено наступні підрозділи та розділи даної роботи.

2.3. Удосконалення методу пошуку можливих джерел забруднення на основі аналізу аномалій

Як було зазначено в пункті 2.2.2, один з видів невизначеності пов'язаний з відсутністю достовірної інформації про перелік усіх джерел забруднення, через те, що підприємства, які здійснюють викид забруднювальних речовин, або фальсифікують інформацію про реальні викиди, або взагалі не звітують щодо цього (невизначеність U_{S1}). Це, у свою чергу, ускладнює процес боротьби з підвищенням рівнів забруднення, оскільки немає можливості визначити що та з якого моменту часу є безпосередньою причиною зростання рівня забруднення.

Для вирішення цієї проблеми необхідно розробити метод, який дозволить за наявними даними моніторингу локалізувати у просторі та часі джерела забруднення. Отриману інформацію в подальшому можна використовувати для порівняння з наявними дозволами на викиди забруднювальних речовин і, за потреби, вжити обґрунтовані управлінські рішення.

Пропонується метод, що базується на використанні моделі Prophet [53], яку було описано у підрозділі 1.4. Дані моніторингу природних середовищ, зазвичай, мають різну періодичність та мають місце пропуски моментів часу в рядах

спостереження. Однією з переваг моделі Prophet є саме можливість ефективно працювати за таких умов, що робить доцільним її використання для даної задачі. Звичайно, ця модель не позбавлена і недоліків, наприклад не дозволяє застосовувати експоненціальне згладжування даних у вибраному ковзному вікні, але для поставленої задачі цей недолік не є суттєвим, а важливими є саме переваги моделі.

Пропонується такий алгоритм розв'язання задачі [6, 18]:

1. Обираються дані моніторингу, які будуть використовуватися і як тренувальні для налаштування параметрів моделі Prophet, і як контрольні, для перевірки ефективності роботи методу.

2. Вибираються параметри моделі Prophet (варіанти сезонності, порядок коефіцієнтів рядів Фур'є для їх опису) та ін.

3. Встановлюються точки зміни тренду (англійською «changepoints» або «CP») на 1 число кожного місяця і будується лінійна модель Prophet між цими CP для показника якості повітря на різних постах спостережень окремо, таким чином, щоб вона дозволяла виявляти початок забруднюючого впливу джерел із тренувальної вибірки. Вплив визначається за рахунок аналізу нахилу тренду для кожного такого випадку (між кожними CP для кожного показника на кожному пості) та знаходяться серед них найбільші прирости, тобто найбільші нахили, але сходящого типу, тобто, коли мало місце збільшення навантаження по заданому показнику, виявлене на певному пості, а потім воно не зменшувалось, а тільки наростало аж по останній день, тобто ті випадки коли, наприклад, підприємство розпочало роботу та, відповідно, забруднюючий вплив, так і продовжує це робити. Результатом аналізу є приріст забруднення у дату, з якої все почалось, решта приростів, де ця умова не виконується, прирівнюються до нуля.

4. Виявлені дати та забруднюючий вплив порівнюється із контрольними даними, відібраними у пункті 1, за певною метрикою. Змінюємо структуру і параметри моделі та повторюємо пункти 2 та 3, поки похибка не стане прийнятною, або робимо повний перебір допустимих варіантів моделі і вибираємо модель з найменшою похибкою.

В результаті отримуються місця та дати початку забруднюючого впливу. Цю інформацію можна порівняти з відомими даними щодо викидів забруднювальних речовин та відповідними дозволами на їх викиди [65, 66]. Потім інформація про локалізовані у часі та просторі незареєстровані джерела забруднення передається відповідним контролюючим органам для перевірки.

Для підвищення точності роботи даного методу варто на етапі 2 використати «holidays» – набір дат відомих аномальних впливів, таких як природні або антропогенні катастрофи, та їх характеристики. Аномалії можуть суттєво впливати на рівень забруднення у відповідний момент часу та, тим самим, погіршувати точність роботи моделі. Важливим є також те, що вплив таких забруднень може бути довготривалим, тобто у наборі «holidays» слід правильно вказувати «вікно» тривалості такого впливу.

Важливо зазначити, що запропонований метод не розрізняє причини забруднювальних впливів (антропогенні чи природні), а тільки робить співставлення з наявними даними. Проте, подальше вивчення результатів використання методу дозволяє ідентифікувати їх природний чи антропогенний характер.

Метод є дуже чутливим і дозволяє зафіксувати навіть незначне зростання тренду, але таке, що значно відрізняється від тренду у попередні моменти часу.

Ключовим у цьому алгоритмі, окрім застосування моделі Prophet та правильного налаштування її параметрів, є алгоритм виявлення імпульсів, які характеризують початок появи сталого забруднення, тобто такого, що має місце тривалий час після появи, та актуального, тобто такого, що має місце і по цей час, тому зупинимось на ньому більш детально.

Модель Prophet повертає початковий тренд k_0 ряду до першої точки зміни d_0 і m приростів $\Delta k_i, i = \overline{1, m}$ тренду між наступними m точками аж до кінця тренувального інтервалу даних. За необхідності, потім по цій моделі можуть передбачатись дані на наступні періоди. При цьому, дати $d_i, i = \overline{1, m}$ розташування цих точок змін вже не зазнають.

Головним є знаходження моментів приростів, тому початковий тренд значення не має і слід аналізувати тільки прирости тренду відносно нього. Результатом роботи моделі Prophet є [6, 18]:

$$y_i = (k + \Delta k_i) d_i, i = \overline{1, m}, \quad (2.1)$$

де y_i – стандартизоване значення тренду в i -й точці зміни d_i , тобто масштабоване між мінімальним і максимальним значеннями та поділене на середньоквадратичне відхилення – це стандартне позначення пакету програм Prophet, яке у результатах так і називається: “yhat” (рис. 2.3).

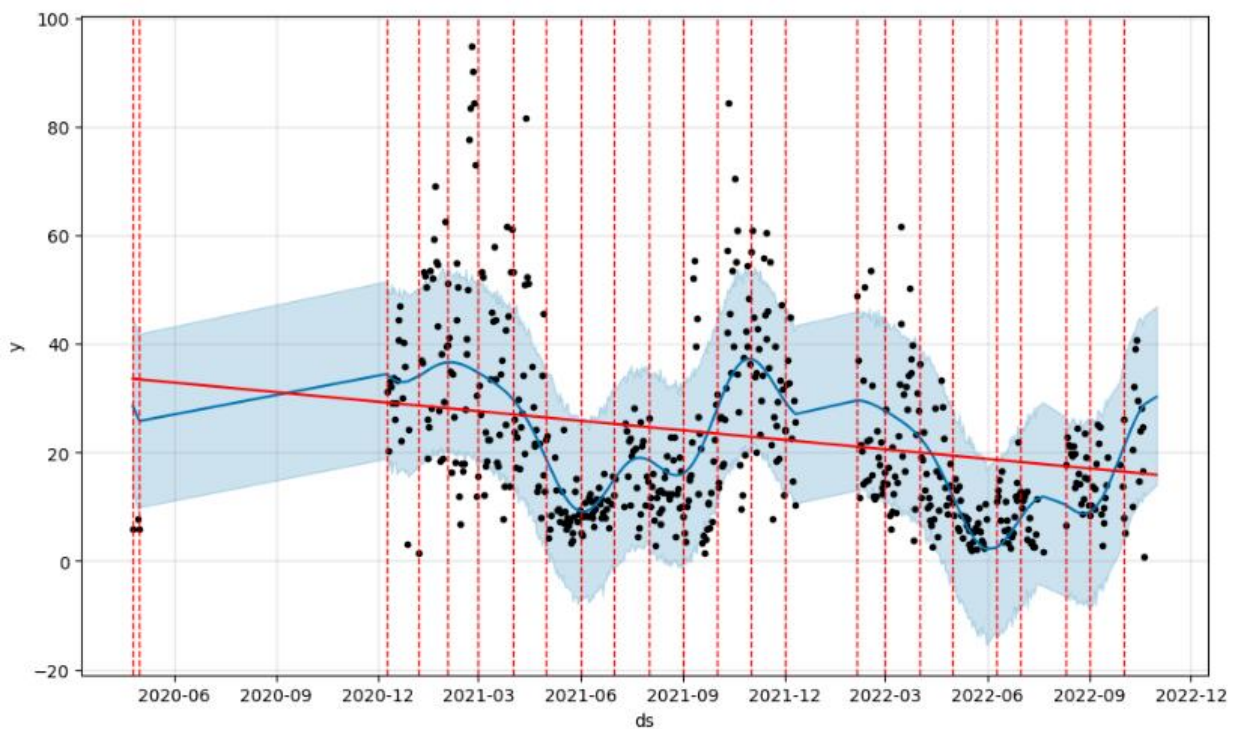


Рисунок 2.3 – Значення частинок PM10 на посту моніторингу у місті Вінниця за 2020-2022 рр. та результат роботи моделі [6, 18]

На рисунку 2.4 показано прирости Δk_i тренду між точками зміни d_i .

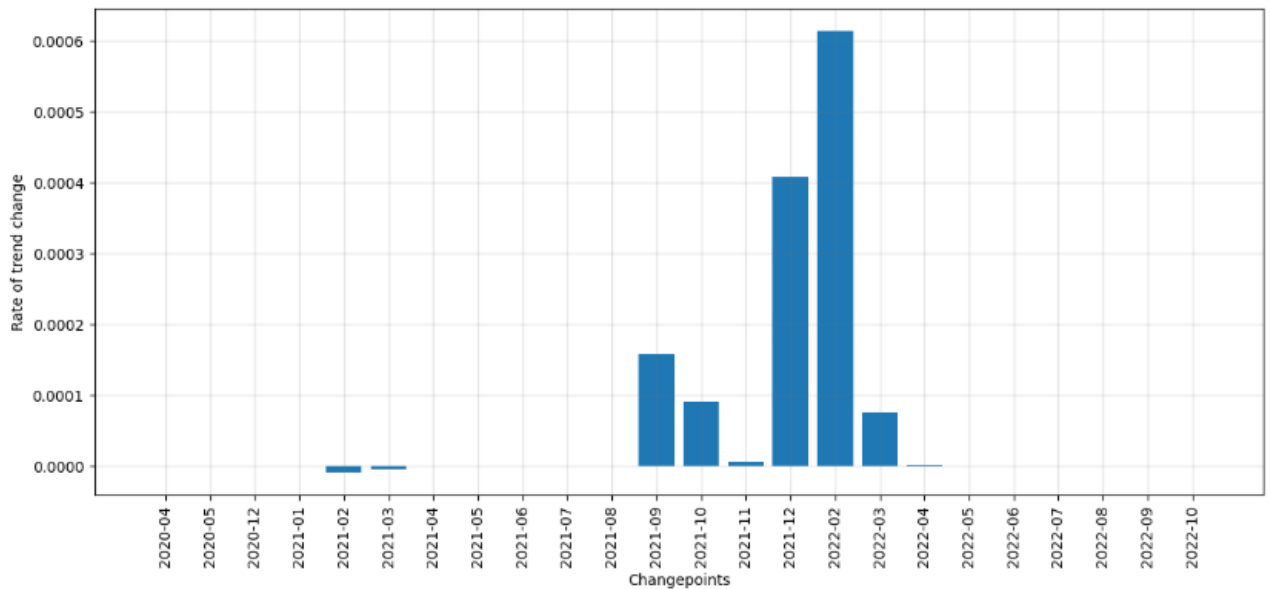


Рисунок 2.4 – Значення приростів Δk_i тренду показника PM10 на посту в місті Вінниця за 2020-2022 рр. між точками зміни (changepoint) d_i [6, 18]

Необхідно сформулювати умову для виявлення ділянок, де прирости трендів є суттєвими (наприклад, більшими, ніж на 5% від максимального значення). Було здійснено інтегрування, тобто знайдено S суми приростів тренду з накопиченням [6, 18]:

$$S = \sum_{i=1}^m \Delta k_i. \quad (2.2)$$

На рисунку 2.5 наведено результат оброблення даних рисунку 2.4 за формулою (2.2).

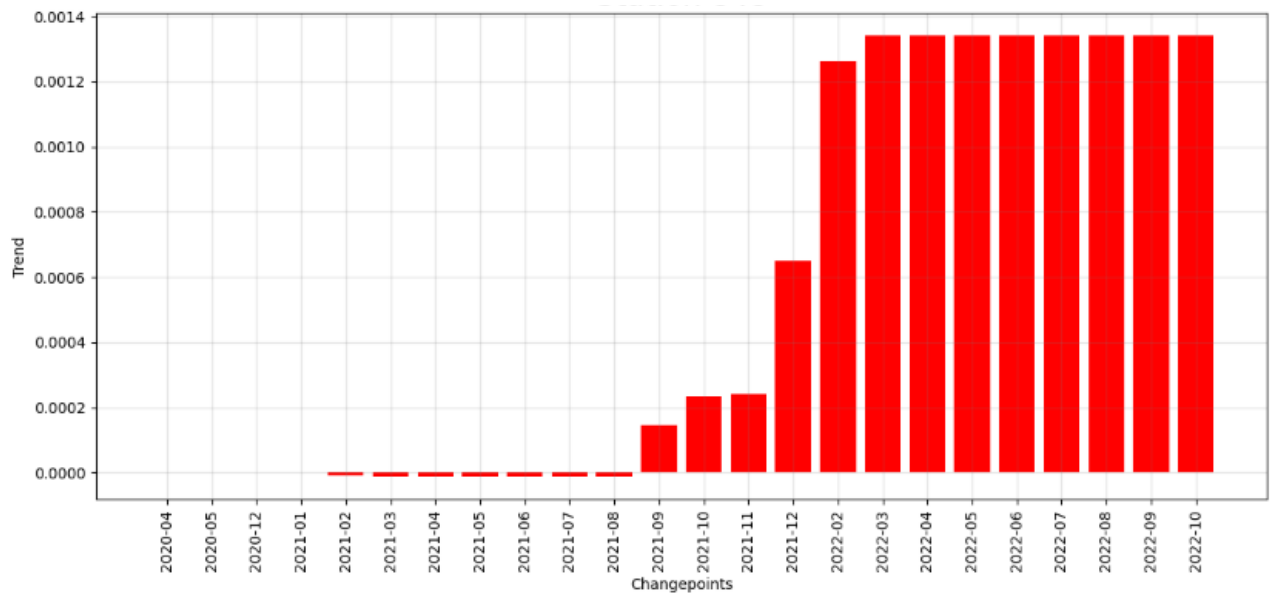


Рисунок 2.5 – Значення сум приростів тренду з накопиченням показника PM10 на посту в місті Вінниця за 2020-2022 рр. між точками зміни d_i [6, 18]

Але аналіз показав, що ряд приростів Δk_i тренду може бути знакозмінним. У той же час, важливо знайти саме такі імпульси, вплив яких є і досі актуальним, а не епізодичним. Тобто треба знайти ситуації, коли після такого імпульсу концентрація далі весь час тільки наростала. Значення S можуть змінюватись синусоїдально і по них важко буде однозначно виявити місця постійного наростання концентрації. Для цього пропонується [6, 18] здійснити ще одне інтегрування:

$$J = \sum_{j=1}^m S_j = \sum_{q=j}^m \sum_{j=1}^m \Delta k_{qj}, \quad (2.3)$$

а тоді слід лише знайти такі імпульси на графіку на рис. 2.4, після яких функціонал J далі лише монотонно наростала (рис. 2.6).

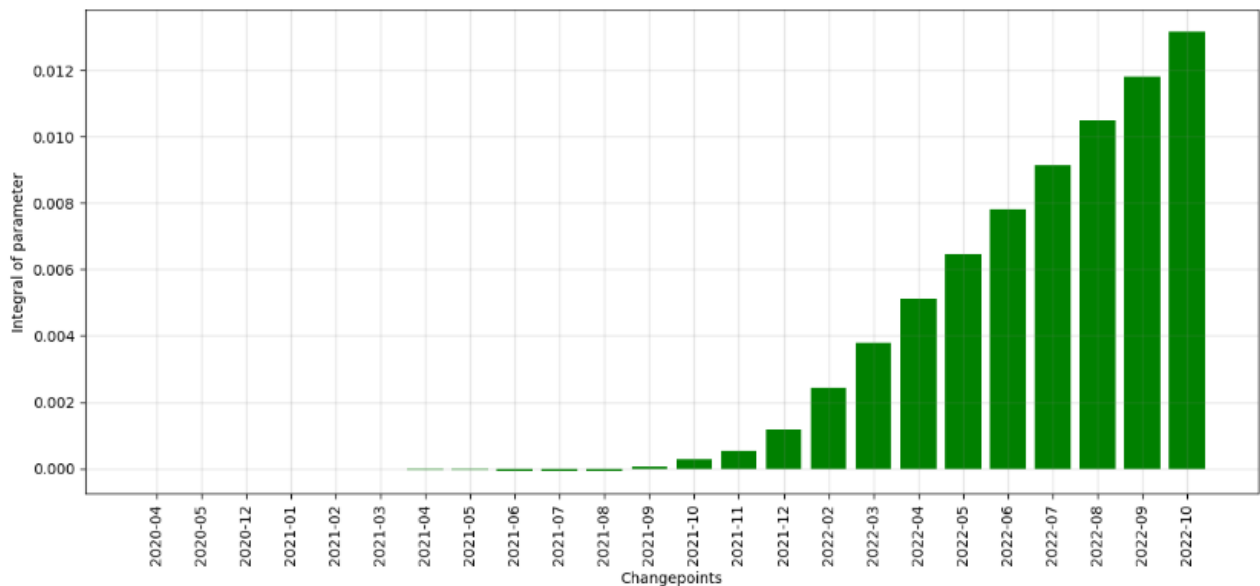


Рисунок 2.6 – Значення сум J приростів тренду з накопиченням показника PM_{10} на посту в місті Вінниця за 2020-2022 рр. між точками зміни d_i [6, 18]

Таким чином, алгоритм методу пошуку можливих джерел підвищеного забруднення на основі аналізу аномалій містить такі кроки:

1. Вибрати показники моніторингу якості повітря, для яких часто має місце перевищення гранично допустимих значень у контрольній точці та по яких немає впевненості щодо джерел їх походження.
2. Сформувати та очистити датасети від помилкових та пропущених даних.
3. Сформувати набір даних з датами відомих зафіксованих викидів забруднювальних речовин (наприклад, антропогенні або природні катастрофи), в яких можливі перевищення показника, що досліджуються.
4. Виділити з кожного датасету тренувальні дані (75-80%), на яких будуть ідентифікуватись параметри моделі, і тестові дані, на яких буде вибиратись оптимальна модель з числа ідентифікованих на тренувальних даних.
5. Задати початкове наближення точок зміни тренду *changepoints*, встановивши їх на першу дату кожного місяця з числа тренувальних даних.
6. Здійснити налаштування моделі на основі Prophet з лінійним трендом, з урахування даних типу «holidays» та врахуванням сезонності, яка мінімізує

похибку на тестових даних. Побудувати різні моделі, варіюючи такі параметри:

- Адитивний чи мультиплікативний варіант врахування тренду і сезонних компонент моделі;
- Різне врахування сезонності – тільки річна, яка є обов’язковою, чи ще й по порах року (період 365,25/4);
- Різні значення коефіцієнтів рядів Фур’є, якими апроксимуються сезонні компоненти ряду;
- Різна кількість точок зміни тренду.

7. Вибрати оптимальну за структурою і параметрами модель, ідентифіковану на етапі 6, яка забезпечить мінімальну похибку на тестових даних, відібраних на етапі 4.

8. Для кожного i -го показника для кожної ділянки між j -ми постами спостережень і для усіх дат між k -ми точками changepoints показника якості повітря x_{ijk} , масштабованого по цьому показнику, визначити бальну y_{ijk} оцінку приросту тренду x цього показника, яка більшим значенням дає значно більшу вагу, наприклад, шляхом взяття квадрату від їх значень [6, 18]:

$$y_{ijk} = \left(\frac{x_{ijk} - \min_i x_{ijk}}{\max_i x_{ijk} - \min_i x_{ijk}} \right)^2 \quad (2.4)$$

або більш простий варіант для випадку, коли мінімальне значення береться таким, що дорівнює нулю [6, 18]:

$$y_{ijk} = \left(\frac{x_{ijk}}{\max_i x_{ijk}} \right)^2. \quad (2.5)$$

9. Для кожної k -ї дати кожної ділянки між j -ми постами спостережень слід додати усі ці бальні оцінки та ранжувати їх за зменшенням [6, 18]:

$$J_{jk} = \sum_i y_{ijk}. \quad (2.6)$$

10. Проаналізувати місця та дати з найбільшими значеннями J_{jk} з

урахуванням даних звітності, дозволами на викиди забруднювальних речовин, результатами перевірок, даними супутникових спостережень, даними громадського екологічного моніторингу та іншою інформацією, здійснити пошук, коли у визначені дати на визначених ділянках могли з'явитись джерела забруднення, та рекомендувати відповідним органам провести більш ретельну перевірку виявлених фактів.

Блок-схема алгоритму даного методу наведена на рисунку 2.7.

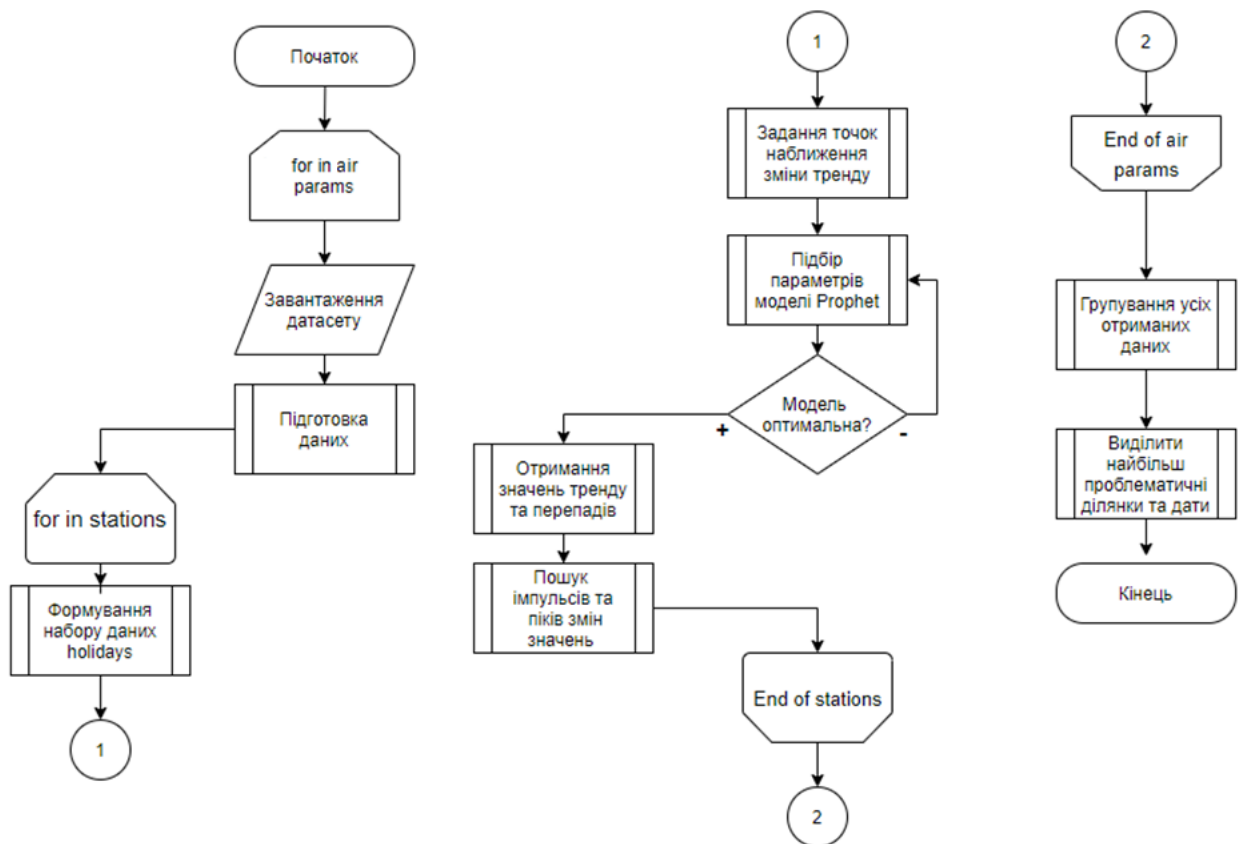


Рисунок 2.7 – Блок-схема алгоритму методу пошуку можливих джерел забруднення на усіх станціях (stations) [6, 18]

Метод було перевірено для даних моніторингу якості атмосферного повітря від мережі громадського моніторингу EcoCity [23, 42], отримані з декількох станцій моніторингу атмосферного повітря, розташованих у Вінницькій області (рис. 2.8).

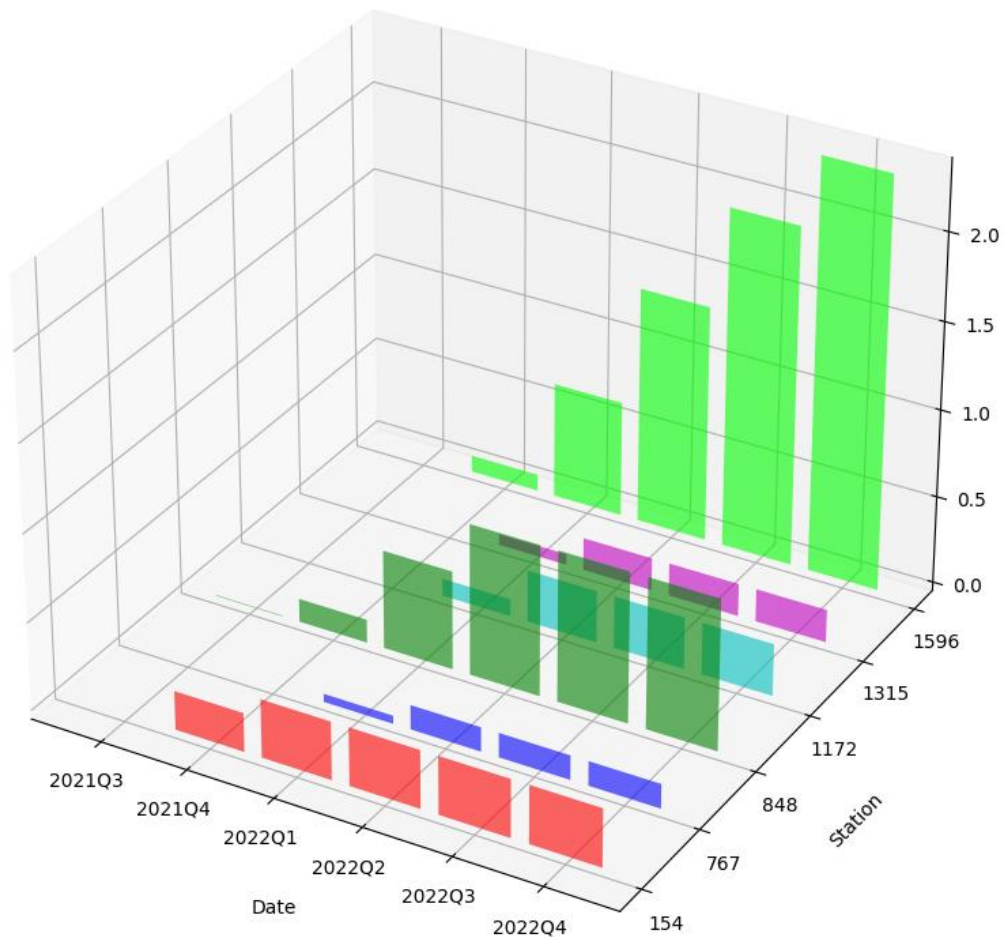


Рисунок 2.8 – Результат пошуку аномальних впливів у часі у Вінницькій області за середньоквартальними даними 2020-2022 рр. [6, 18]

З рисунку 2.8 видно, що на станції № 1596 (м. Хмільник), починаючи з 4-го кварталу 2021 року, розпочався сталий вплив забруднення, що продовжувався місяцями.

Метод є універсальним і може застосовуватися для інших природних середовищ, наприклад, для пошуку можливих джерел підвищеного забруднення річки за даними мережі спостереження за її станом. Це було підтверджено за рахунок застосування запропонованого методу для даних системи моніторингу якості вод Держводагентства за 2002-2019 рр. для постів ділянки р. Південний Буг від витoku до м. Вінниці. Приклад даних за показником азоту амонійного однієї з станцій моніторингу наведено на рисунку 2.9.

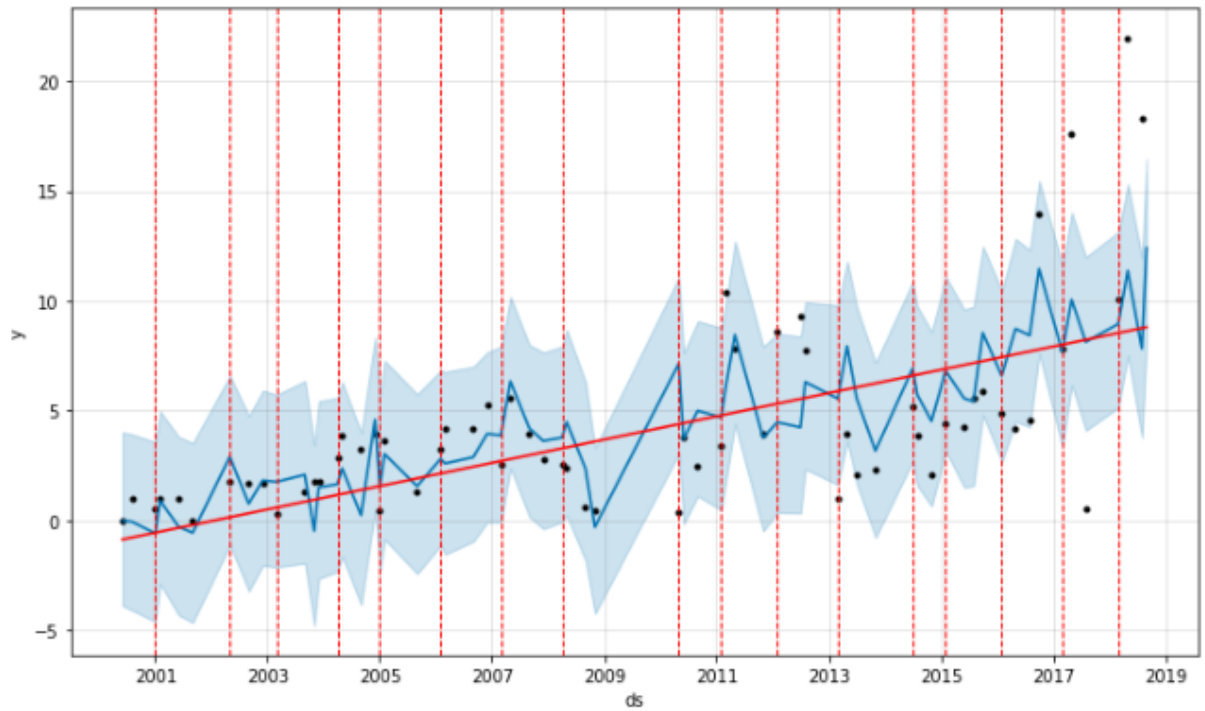


Рисунок 2.9 – Значення азоту амонійного на посту «Меджибіж» за 1994-2019 рр. та результат роботи моделі Prophet [6]

Згідно алгоритму запропонованого методу, для цих даних, було знайдено прирости Δk_i тренду між точками зміни d_i (рис 2.10).

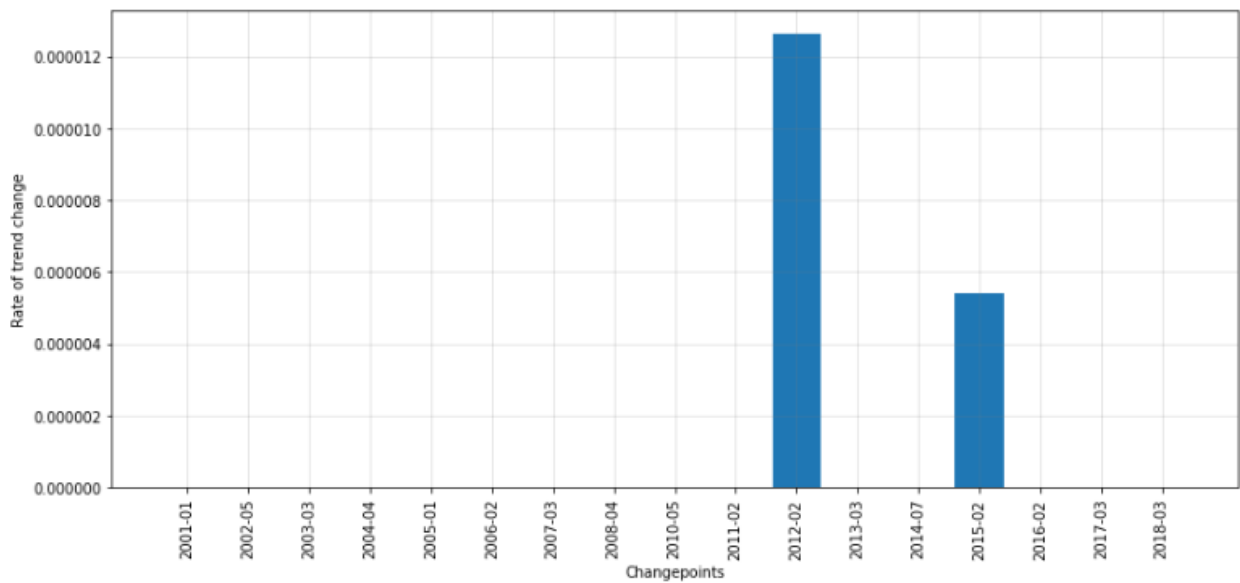


Рисунок 2.10 – Значення приростів Δk_i тренду азоту амонійного на посту «Меджибіж» за 1994-2019 рр. між точками зміни (changepoint) d_i [6]

Використовуючи формулу (2.2), було знайдено значення сум приростів тренду з накопиченням (рис. 2.11).

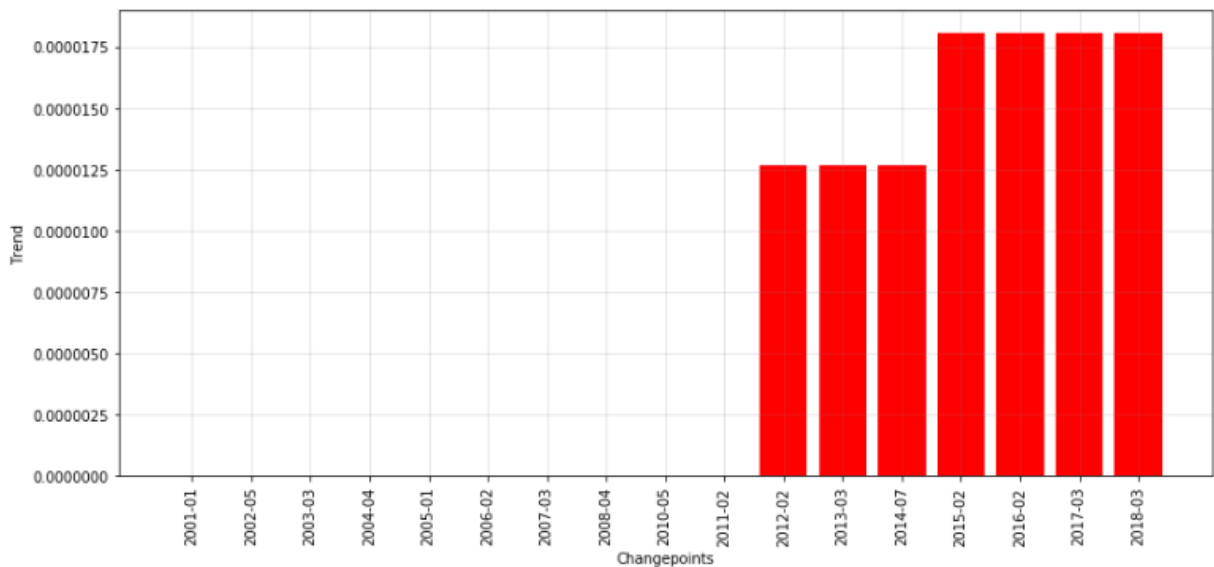


Рисунок 2.11 – Значення сум приростів тренду з накопиченням азоту амонійного на посту «Меджибіж» за 1994-2019 рр. між точками зміни d_i [6]

Застосовано формулу (2.3) та знайдено такі імпульси з рис. 2.11, які монотонно наростали (рис. 2.12).

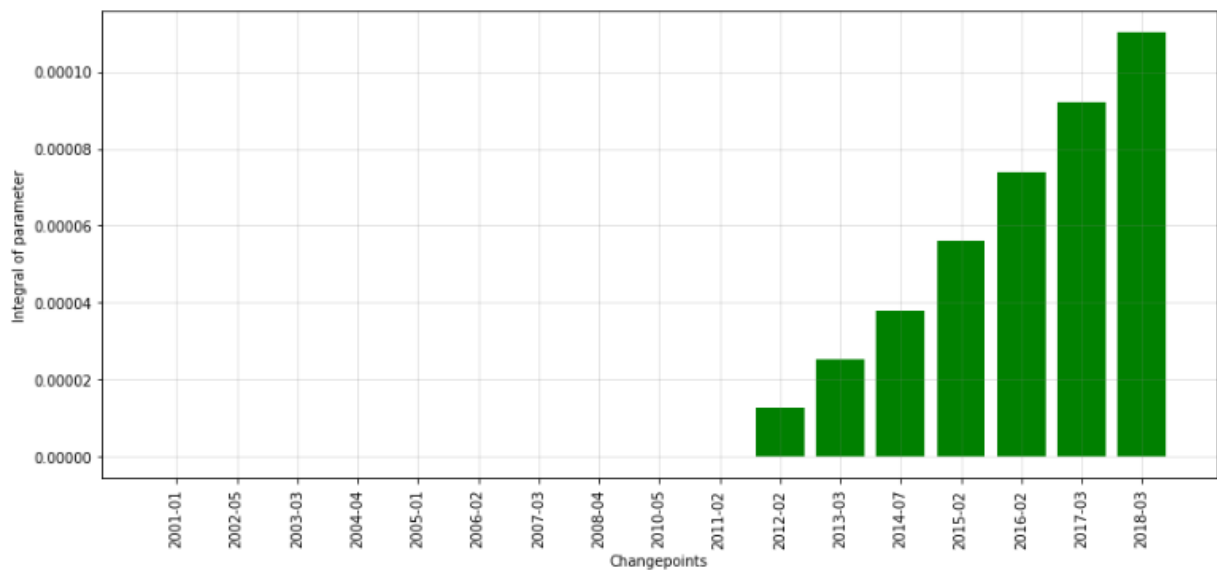


Рисунок 2.12 – Значення сум J приростів показника азоту амонійного на посту «Меджибіж» за 1994-2019 рр. між точками зміни d_i [6]

Як видно на рисунку 2.12, для поста моніторингу «Меджибіж» за показником азоту амонійного спостерігається підвищений вплив з початку 2012 року, що потенційно може свідчити про незареєстроване джерело забруднення та що вимагає додаткової перевірки з наявними дозволами на скидання вод та реальні значення звітування щодо цієї діяльності.

Отриману інформацію можна застосувати для подальшого і більш детального аналізу з використанням технології, запропонованій у роботі [67] під час пошуку джерел забруднення водних ресурсів. В роботі пропонується аналізувати та виявляти незареєстровані джерела забруднення за даними супутникових знімків, наприклад OpenStreetMap (рис. 2.13) та, за відсутності такої інформації, використовувати знімки з безпілотних літальних апаратів.

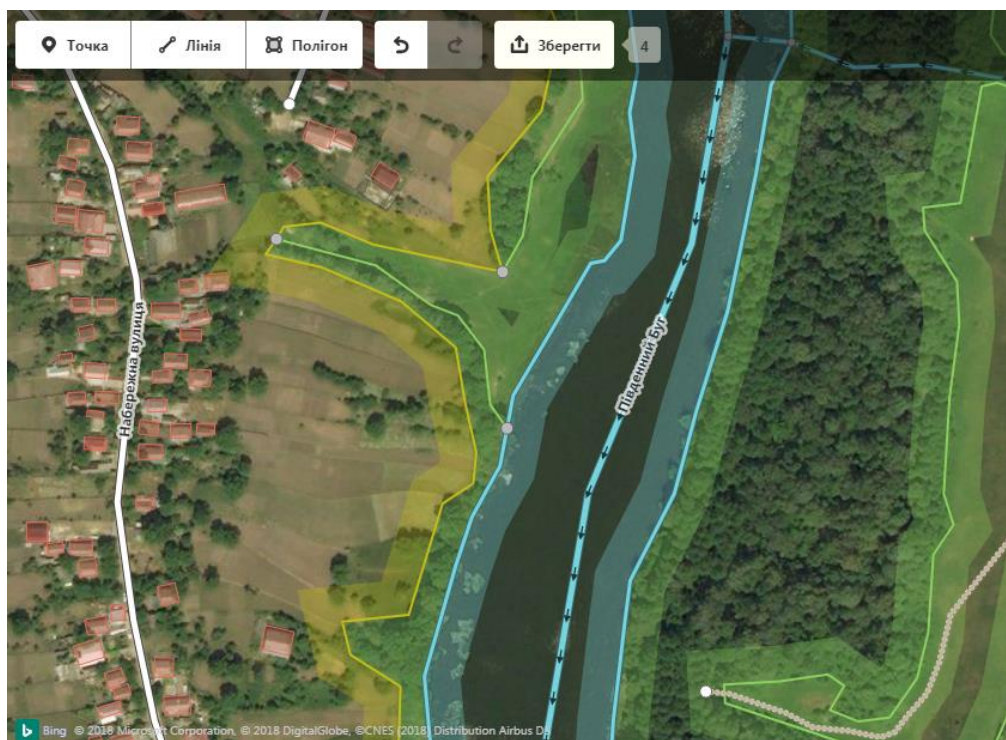


Рисунок 2.13 – Супутникові та векторні дані з сервісу OpenStreetMap [67]

Дана технологія, в першу чергу, базується на даних супутникових знімків, які досить часто можуть бути неналежної якості або взагалі відсутні. Застосування безпілотних літальних апаратів (БПЛА) дозволяє усунути цей недолік [68], але в той же час застосування БПЛА на великих площах відповідного річкового

басейну може займати дуже багато часу і бути, через це, менш ефективним методом [69]. Проте, використовуючи запропонований метод пошуку можливих джерел забруднення на основі аналізу аномалій, можна локалізувати у просторі незареєстровані джерела забруднення і використовувати технологію з роботи [67] саме на цих ділянках. Це може слугувати як підходом перевірки результатів так і уточнення інформації про незареєстрований викид забруднювальних речовин у річку або інші водні ресурси.

Узагальнені результати, як після використання запропонованого методу пошуку можливих джерел забруднення, так і комбінованого підходу з використанням технології з роботи [67], можуть бути передані в Держекоінспекцію відповідного регіону або області для порівняння з інформацією про дозволи на викиди та, за потреби, прийняття відповідних рішень щодо усунення джерел забруднення.

Інформація про сталий вплив негативний вплив на стан природного середовища також може допомогти ідентифікувати саме епізодичні аномалії, що в цілому дозволяє підвищити якість і точність аналізу та прогнозування.

Отже, метод пошуку можливих джерел забруднення на основі аналізу аномалій є ефективним інструментом, який можна успішно застосовувати для вирішення поставленої задачі і який дозволяє розкрити невизначеність U_{S1} .

2.4. Усунення цільової невизначеності щодо системи моніторингу стану природного середовища

В підрозділі 2.1 було визначено, що існує проблема з вибором місць розташування постів (як при побудові нової мережі моніторингу, так і при оптимізації вже існуючої) моніторингу стану природного середовища (невизначеність U_{T1}), оскільки існуючі системи можуть збирати недостатню кількість даних для проведення ефективного аналізу стану природного середовища [70]. Відповідно, потрібно покращити та оптимізувати відповідну систему, але цей процес є довгим і досить дорогим, тому необхідно сформулювати

та описати підхід, яка дозволить визначити пріоритетні місця для встановлення станцій, що в далекій перспективі дозволить системним аналітикам точніше аналізувати та прогнозувати стан природного середовища в межах розташування відповідної мережі моніторингу.

Пропонується підхід формування ієрархічної системи критеріїв для вибору місць розташування постів спостережень стану природного середовища, з урахуванням екстремальних значень показників, які є характерними для регіону і наведені вище [7, 12].

Пропонується відштовхуватися від згаданої в підрозділах 1.3 та 2.1 моделі «DPSIR» [48]. При визначенні місць розташування станцій необхідно враховувати особливості відповідного регіону. Базуючись на цій моделі, важливим є врахування трьох ключових груп факторів:

1) **D-P** - продуктивні сили, у тому числі різні антропогенні джерела (**D**), які впливають на стан природного середовища через викидання різного роду забруднювальних речовин (**P**);

2) поточний стан природного середовища (**S**), що визначається за даними спостереження мережі моніторингу за відповідним природним середовищем (станції моніторингу атмосферного повітря, станції моніторингу вод тощо);

3) вплив на населення та на довкілля (**I**), що робить необхідним прийняття відповідних керівних рішень (**R**).

Виходячи з цього, важливо розташовувати станції моніторингу з урахуванням трьох основних видів критеріїв [7, 71]:

1. Знаходження неподалік від місць відомих аномально великих антропогенних джерел викидів – підприємства машинобудування, хімічного виробництва, металургійні підприємства тощо;

2. Розташування у місцях, де раніше було виявлено аномальні дані або перевищення граничних показників стану цього природного середовища;

3. Розташування неподалік від місць проживання населення, передусім, особливо вразливих верств цього населення. До таких місць можуть відноситися санаторії, дитячі лікарні, дитячі табори, школи тощо.

Ці критерії можна застосувати з відомим методом аналізу складних рішень – методом аналізу ієрархій [72]. Є декілька алгоритмів його застосування, але пропонується використати алгоритм зі статті [73]. Як відомо, за цим методом знаходять максимальне дійсне додатне власне число λ_{\max} кожної C -матриці та власні вектори w^j для нього та розв'язують систему матричних рівнянь виду [7, 73]:

$$\left[C^{(j)} - \lambda_{\max}^{(j)} I \right] \left[w^{(j)} \right] = 0, \quad (2.7)$$

де C – матриця, λ_{\max} – власне число матриці, w^j – власні вектори матриці $j=1..n$.

При цьому, необхідно замінити одне з рівнянь на таке [7, 73]:

$$w_1^j + w_2^j + \dots + w_n^j = 1. \quad (2.8)$$

Розв'язавши систему рівнянь, отримуємо ваги критеріїв $w_1^{(j)}$, де $j=1..n$ та можемо використати їх для обчислення значень інтегральних факторів K_j та кожного базового фактору K_{ji} за допомогою наступних формул [7, 73]:

$$\begin{aligned} K_j &= w_j^{(0)} K_{nn}^*, \\ K_{ji} &= w_j^{(0)} w_i^{(j)} K_{nn}^*, \end{aligned} \quad (2.9)$$

де $j=1..n$, $i=1..n$

Таким чином, необхідно виділити набір інтегральних критеріїв, які базуються на моделі DPSIR та дозволяють врахувати особливості конкретного регіону і враховувати аномальні або граничні значення показників стану природного середовища.

Згідно алгоритму методу аналізу ієрархій, для кожного критерію необхідно визначити декілька базових критеріїв, як це показано на рисунку 2.14.

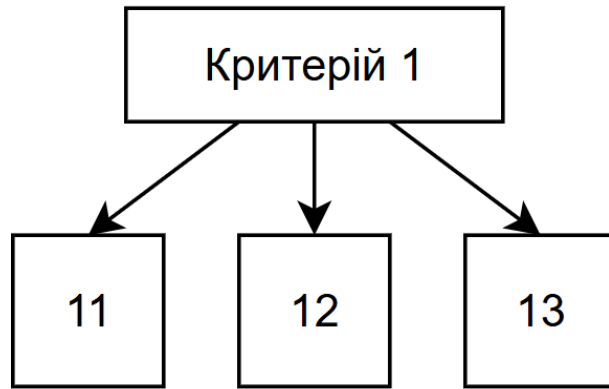


Рисунок 2.14 – Структура ієрархічного критерія вибору місця розміщення станції

Для визначених ієрархічних критеріїв, експерти визначають оцінки, що, в тому числі, базуються на наявних даних моніторингу стану природного середовища. Після застосування методу аналізу ієрархій можна отримати список пріоритетних місць розташування станцій моніторингу стану природного середовища. За наявності проблем або зауважень до отриманих результатів, експерти можуть переглянути раніше виставлені експертні оцінки та повторно застосувати метод аналізу ієрархій, щоб обрахувати нові значення інтегральної оцінки кожної станції.

В цілому, блок-схема алгоритму запропонованого підходу наведена на рисунку 2.15.

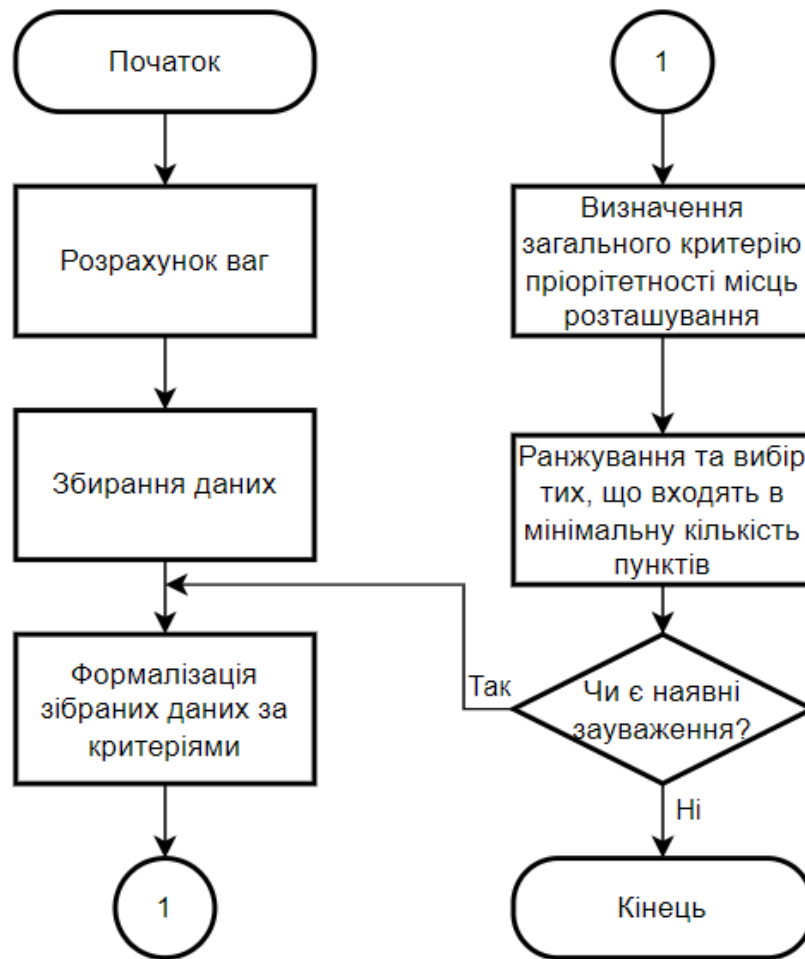


Рисунок 2.15 – Блок-схема алгоритму проєктування мережі спостережень стану природного середовища на основі методу аналізу ієрархій [7, 12]

Розглянемо приклад застосування даного підходу для оптимізації мережі моніторингу стану атмосферного повітря у Вінницькій області. Залучення експертів та аналіз рекомендацій нормативної бази Міндовкілля, з урахуванням наведених вище критеріїв, показав, що важливими є такі 5 критеріїв [7]:

Φ_1 – близькість до найбільш впливових джерел забруднення атмосферного повітря (група Р-критеріїв – «Pressures»);

Φ_2 – частота та рівень перевищень усереднених та максимальних разових показників стану атмосферного повітря та атмосферних опадів (група S-критеріїв – «States»);

Φ_3 – близькість до стейкхолдерів (населення, у т.ч. вразливого, інвестиційних об'єктів тощо) (група I-критеріїв – «Impacts», інвестиції можна

умовно розглядати як один з непрямих позитивних «наслідків» – у більш чистих регіонах більше інвестують, оскільки меншими будуть витрати на покращення здоров'я працівників, на фільтри у вентиляційних системах та ін.);

Φ_4 – рівень репрезентативності запроваджених керівних заходів (встановлені очисні системи, введений контроль за викидами пересувних транспортних заходів тощо) (група R-критеріїв – «Responses»);

Φ_5 – зручність інфраструктури, умов для обслуговування пунктів спостережень, цінність для рівномірного охоплення регіону, що спростить побудову мап забруднення в майбутньому та ін. (група допоміжних критеріїв).

У 1-му інтегральному критерії пріоритетного вибору місць розташування пунктів спостережень якості атмосферного повітря регіону базовими критеріями є такі:

Φ_{11} – сумарний обсяг стаціонарних джерел викидів підприємств в останній із наданих років зі списку найбільших забруднювачів повітря регіону у певному околі (наприклад у колі, площею 1/100 чи 1/1000 від площі регіону);

Φ_{12} – сумарний обсяг гранично допустимих викидів підприємств в останній із наданих років зі списку найбільших забруднювачів повітря регіону у певному околі;

Φ_{13} – кількість центроїдів виробничих майданчиків у певному околі;

Φ_{14} – близькість до автостради міжнародного чи регіонального значення або дороги з високою завантаженістю пересувними транспортними засобами у години «пік», передусім вантажними;

Φ_{15} – ступінь близькості від пункту до геометричного центра сусіднього населеного пункту або території з найбільшою кількістю стаціонарних джерел викидів.

У 2-му інтегральному критерії пріоритетного вибору місць розташування пунктів спостережень якості атмосферного повітря регіону базовими критеріями є такі [7]:

Φ_{21} – відсоток перевищень середньодобової ГДК показника у точках у певному околі (наприклад у колі, площею 1/100 чи 1/1000 від площі регіону);

Φ_{22} – відсоток перевищень максимально разової ГДК показника у точках у певному околі;

Φ_{23} – відсоток перевищень нижнього порогу оцінювання показника у точках у певному околі;

Φ_{24} – відсоток перевищень верхнього порогу показника у певному околі;

Φ_{25} – відсоток перевищень ГДК за супутниковими даними та/або даними громадського моніторингу атмосферного повітря у певному околі.

У 3-му інтегральному критерії пріоритетного вибору місць розташування пунктів спостережень якості атмосферного повітря регіону базовими критеріями є такі [7]:

Φ_{31} – близькість до інтенсивної забудови багатопверхових житлових масивів міст, км;

Φ_{32} – близькість до місць розташування найбільш вразливого населення: санаторіїв, лікарень, шкіл, дитсадків тощо;

Φ_{33} – близькість до наявних інвестиційних об'єктів, км;

Φ_{34} – близькість до потенційних інвестиційних об'єктів;

Φ_{35} – близькість до органів самоврядування (міськрад, сільрад) та інших місць, де часто зустрічають різні делегації, у т.ч. іноземні, км.

У 4-му інтегральному критерії пріоритетного вибору місць розташування пунктів спостережень якості атмосферного повітря регіону базовими критеріями є такі [7]:

Φ_{41} – близькість до об'єктів, де нещодавно були встановлені очисні споруди та проведено інші заходи зі зниження забруднення атмосферного повітря, у т.ч. асфальтування доріг та інших дорожньо-транспортних робіт, заходів з оптимізації дорожнього руху тощо;

Φ_{42} – близькість до об'єктів, де планується встановити очисні споруди чи провести інші заходи зі зниження забруднення атмосферного повітря, у т.ч. асфальтування доріг та інших дорожньо-транспортних робіт, заходів з оптимізації дорожнього руху тощо;

Φ_{43} – близькість до місць, де запроваджено контроль за викидами пересувних джерел, за вагою вантажних засобів та ін., км;

Φ_{44} – близькість до об'єктів природно-заповідного фонду або сільської місцевості (для зон);

Φ_{45} – близькість до паркових зон чи зон з обмеженим рухом для вантажного транспорту.

У 5-му інтегральному критерії пріоритетного вибору місць розташування пунктів спостережень якості атмосферного повітря регіону базовими критеріями є такі [7]:

Φ_{51} – зручність для обслуговування вимірювального обладнання пункту;

Φ_{52} – зручність інфраструктури для встановлення вимірювального обладнання пункту;

Φ_{53} – цінність для рівномірного охоплення регіону, що спростить побудову карт забруднення в майбутньому;

Φ_{54} – цінність для моніторингу за якістю атмосферного повітря регіону на думку органу, відповідального за моніторинг якості атмосферного повітря;

Φ_{55} – цінність для контролю за регулярними порушниками умов, які прописані в дозволах підприємств даного регіону, на думку органу, відповідального за екоінспекційний контроль якості атмосферного повітря.

Таким чином, отримуємо ієрархічну структуру критеріїв, що можна зобразити схемою, наведеною на рисунку 2.16.

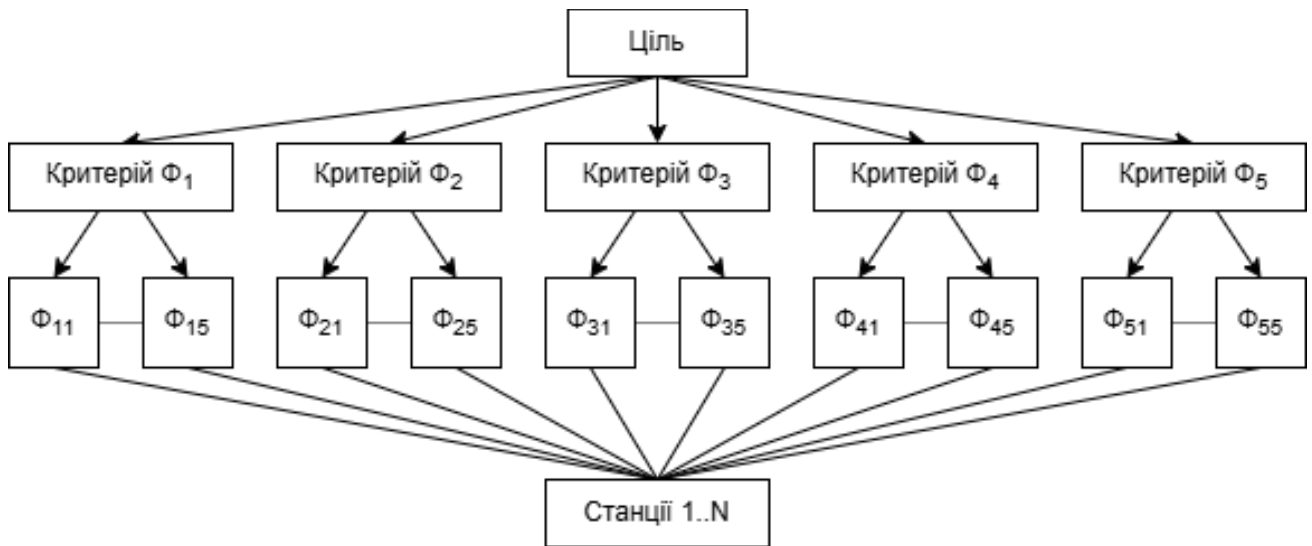


Рисунок 2.16 – Діаграма інтегральних критеріїв для вибору оптимального місця розташування станції моніторингу

Наведемо приклад обрахунків згідно описаного підходу. Експерти провели попарне порівняння критеріїв та визначили за допомогою оцінок на скільки один критерій переважає над іншим. З цих оцінок формується матриця, як це показано в таблиці 2.2.

Таблиця 2.2 – Приклад експертних оцінок по визначених інтегральних критеріях [7, 12]

Критерії	Φ_1	Φ_2	Φ_3	Φ_4	Φ_5
Φ_1	1	2	5	6	6
Φ_2	$\frac{1}{2}$	1	3	4	4
Φ_3	$\frac{1}{5}$	$\frac{1}{3}$	1	4	2
Φ_4	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{4}$	1	2
Φ_5	$\frac{1}{6}$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{2}$	1

Використовуючи формули (2.7) та (2.8), була сформована система рівнянь (рис 2.17)

$$\begin{aligned}
 -4.258*w_1 + 2.0*w_2 + 5.0*w_3 + 6.0*w_4 + 6.0*w_5 &= 0 \\
 0.5*w_1 + -4.258*w_2 + 3.0*w_3 + 4.0*w_4 + 4.0*w_5 &= 0 \\
 0.2*w_1 + 0.333*w_2 + -4.258*w_3 + 4.0*w_4 + 2.0*w_5 &= 0 \\
 0.167*w_1 + 0.25*w_2 + 0.25*w_3 + -4.258*w_4 + 2.0*w_5 &= 0 \\
 0.167*w_1 + 0.25*w_2 + 0.5*w_3 + 0.5*w_4 + -4.258*w_5 &= 0
 \end{aligned}$$

Рисунок 2.17 – Приклад системи рівнянь при застосуванні методу аналізу ієрархій [7, 12]

Розв'язавши цю систему за допомогою формули (2.9), було отримано ваги кожного критерію (рис 2.18).

$$\begin{aligned}
 K_1 &= w_{\theta_1} * K_{nn} = 0.466 * K_{nn} \\
 K_2 &= w_{\theta_2} * K_{nn} = 0.27 * K_{nn} \\
 K_3 &= w_{\theta_3} * K_{nn} = 0.136 * K_{nn} \\
 K_4 &= w_{\theta_4} * K_{nn} = 0.069 * K_{nn} \\
 K_5 &= w_{\theta_5} * K_{nn} = 0.058 * K_{nn}
 \end{aligned}$$

Рисунок 2.18 – Приклади обрахованих ваг інтегральних критеріїв [7, 12]

Запропонований алгоритм застосовується для кожного показника якості атмосферного повітря та атмосферних опадів окремо [7, 12]. Спочатку збираються дані, потім за кожним критерієм визначаються 3-5 оптимальних місць (наприклад, за Φ_{11} – де найбільше стаціонарних джерел викидів, за Φ_{32} – де найбільше лікарень, санаторіїв, шкіл і дитсадків, за Φ_{51} та Φ_{52} – у пунктах, де вже є стаціонарні пости, наприклад ДСНС України, або місця, де регулярно проводить вимірювання МОЗ, та ін.), формується єдиний список з усіх цих місць, проводяться розрахунки, ранжуються усі місця за даним показником,

відбираються з них першочергові для встановлення. Потім можна додати ранги цих місць по усіх показниках і вибрати найбільш пріоритетні місця одночасно за усіма показниками, але можна і не додавати. Іноді краще в одному місці вимірювати пил, в другому – важкі метали, у третьому – озон (давач на озон – порівняно дешевий, його можна й в усіх пріоритетних місцях вимірювати), у четвертому – якість атмосферних опадів. Це вже – на розсуд проєктувальників та керівництву зон чи агломерації.

В результаті, створена або покращена мережа моніторингу стану природного середовища дозволить системним аналітиками отримати корисні дані, що в подальшому можна буде використовувати для аналізу та прогнозування.

Таким чином, застосування запропонованого підходу формування критеріїв оптимізації мережі спостереження дозволяє розкрити невизначеність U_{T1} . Під час формування цих критеріїв використовуються порогові значення показників, згідно офіційних директив, що дозволяє розкрити невизначеність U_{T3} та U_{T5} .

2.5. Висновки до розділу

У розділі було охарактеризовано систему управління станом природного середовища як складну систему. Проведено декомпозицію системи за різними критеріями та визначено які види невизначеності мають місце для кожного елементу декомпозиції. Було проаналізовано методи та підходи для розкриття зазначених видів невизначеності. Аналіз показав, що частина видів невизначеності розкривається інформаційними запитами, але для більшості видів невизначеності необхідно створити або удосконалити методи аналізу та прогнозування даних стану природного середовища з урахуванням аномальних та пікових значень.

Описано подальший розвиток методу пошуку можливих джерел забруднення з використанням методу пошуку аномалій. Метод дозволяє ідентифікувати у часі та просторі початок незареєстрованого підвищеного антропогенного впливу на стан природного середовища, що дозволить,

наприклад, разом з поєднанням з даними аналізу супутникових знімків або даних аерофотозйомки локалізувати в часі та просторі ці джерела і використовувати цю інформацію для прогнозування часового ряду цих даних. Це, у свою чергу, дозволило розкрити один з видів невизначеності.

Запропоновано підхід формування ієрархічної системи критеріїв для оптимізації мережі моніторингу стану природного середовища, яка містить етап розроблення інтегральних критеріїв, що, в першу чергу, базуються на аномальних та пікових значеннях стану відповідного природного середовища. Далі, застосовуючи метод аналізу ієрархій, проводиться вибір оптимальних місць для встановлення станцій моніторингу, що дозволяє ефективно використовувати ресурси та отримувати при цьому корисну інформацію, необхідну для прийняття оптимальних керівницьких рішень, щодо стану природного середовища та підвищення якісної і кількісної характеристик існуючих даних моніторингу, що в подальшому може використовуватися аналітиками для аналізу та прогнозування.

Наведені в розділі результати були висвітлені в таких роботах: [6, 7, 11, 12, 18].

РОЗДІЛ 3

ПРОГНОЗУВАННЯ СТАНУ ПРИРОДНИХ СЕРЕДОВИЩ ТА РОЗРОБЛЕННЯ АЛГОРИТМУ ЇХ СИСТЕМНОГО АНАЛІЗУ З УРАХУВАННЯМ ПІКОВИХ ЗНАЧЕНЬ

3.1. Розроблення методу ідентифікації параметрів сезонності часового ряду

В підрозділі 2.2 було визначено, що існують декілька видів невизначеності, які пов'язані з побудовою ефективної інтелектуальної моделі прогнозування стану природного середовища:

- невизначеність щодо значень параметрів моделей (U_{P1});
- невизначеність щодо параметрів моделей, які описують вплив сезонної складової: період, вікно впливу, коефіцієнт регуляризації або сила впливу (U_{P2});
- вибір яка саме сезонна складова або їх сукупність мають місце (U_{T6}).

Для розкриття цих видів невизначеності варто спочатку розглянути класичні підходи та методи ідентифікації параметрів сезонності часового ряду стану природного середовища [74]. Також, необхідно запропонувати метод або набір методів, які дозволять проаналізувати наявні дані моніторингу стану природного середовища та визначити основні параметри інтелектуальної моделі прогнозування. Важливими параметрами періодичних часових рядів є параметри їх тренду, сезонних складових та аномалій. Задача визначення тренду часового ряду має багато універсальних методів розв'язання, проте, ідентифікація одночасно параметрів різних видів сезонності та аномалій різної природи у різні часові проміжки є складною задачею, яка не має універсального вирішення.

Розглянемо декілька класичних підходів до вирішення задачі.

3.1.1. Дослідження впливу параметрів моделі Prophet на точність прогнозування стану природного середовища

Модель Prophet [53], яку було детально описано в підрозділі 1.3, дозволяє налаштувати багато параметрів і, тим самим, моделювати часовий ряд, враховуючи експертні знання та певні особливості самого ряду. Одним з таких параметрів є набір «holidays», що вже згадувався у підрозділі 2.3. За рахунок його використання, можна зменшити вплив відповідних аномальних дат на якість роботи моделі і в результаті отримати більшу точність передбачення часового ряду даних природного середовища.

Іншим важливим параметром моделі Prophet є параметр регуляризації «prior_scale», який дозволяє контролювати наскільки сильно ці аномальні дати впливають на прогноз, що робиться моделлю [14]. При великих значеннях даного параметру, модель буде більше реагувати на аномальні дати, що збільшує їх вплив на прогноз.

У пункті 1.4.2 було здійснено пошук аномалій з використанням бібліотек `sesd`, `scikit-learn` та `statsmodels`. Як видно на рисунках 1.10, 1.12 та 1.14, вони знаходять не усі локальні аномалії, які там є. Тому пропонуємо їх фільтрувати за певним пороговим значенням, яке визначається певним квантилем. На рисунку 3.1 наведено часовий ряд показника пилу PM10, отриманого від мережі громадського моніторингу ECoCity, яка була згадана раніше в підрозділі 1.1.5. На рисунку червоними точками відмічені аномальні дати, які були визначені за рахунок порівняння усіх точок ряду з квантилем 95%.

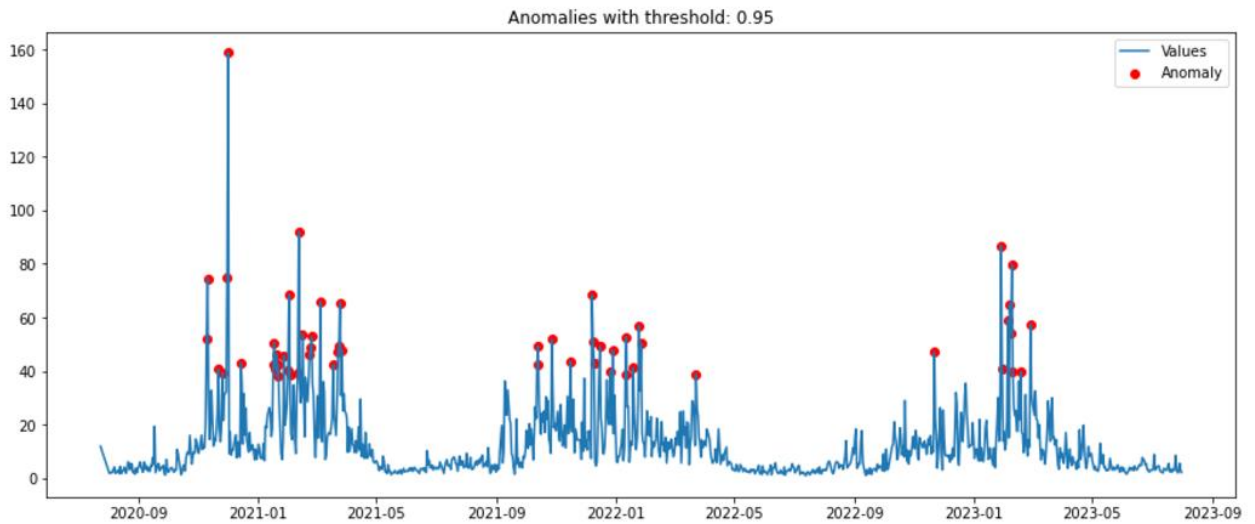


Рисунок 3.1 – Приклад ряду показника пилу PM10, отриманого від мережі громадського моніторингу EcoCity [14]

Обираємо декілька значень для параметру «prior_scale» та декілька значень квантилів. Для часового ряду обраного параметру відсікаються аномальні дані, використовуючи відповідний квантиль, і будується модель Prophet. Кожна модель будується з налаштуванням денної сезонності за замовчуванням та набору «holidays», який є переліком аномальних даних, обраних раніше та відповідним параметром prior_scale. Ці кроки повторюються для усіх можливих комбінацій використання цих параметрів. Для кожного набору тестових даних робиться прогноз та обраховується показник точності. Приклад результату роботи моделі Prophet для часового ряду з аномаліями, визначеними по квантилю 95% та зі значенням prior_scale 10, наведено на рисунку 3.2.

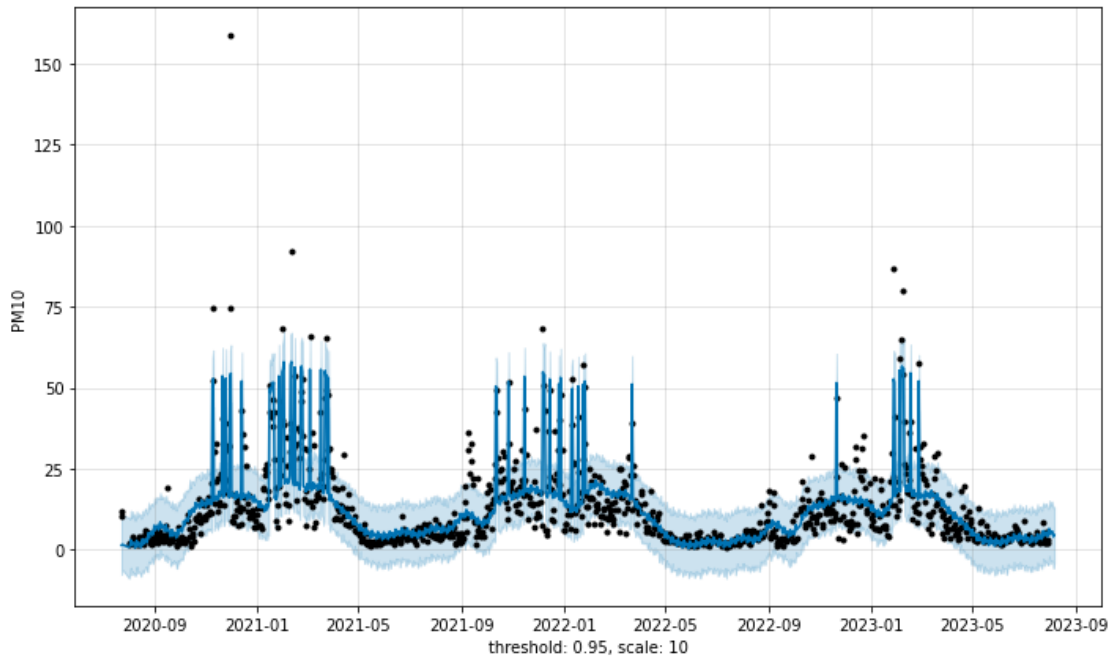


Рисунок 3.2 – Результат роботи моделі Prophet для даних, наведених на рис 3.1 [14]

Результат роботи моделі Prophet було порівняно з реальними значеннями ряду, використовуючи коефіцієнт детермінації (зазвичай, позначається як «r2_score») Результати для усіх комбінацій вхідних параметрів показані на рисунку 3.3.

	threshold	scale	r2
0	0.50	1.0	51.72
1	0.50	5.0	51.84
2	0.50	10.0	51.91
3	0.50	15.0	51.88
4	0.50	20.0	51.89
5	0.50	30.0	51.85
6	0.70	1.0	64.38
7	0.70	5.0	64.48
8	0.70	10.0	64.51
9	0.70	15.0	64.49
10	0.70	20.0	64.49
11	0.70	30.0	64.50
12	0.85	1.0	70.37
13	0.85	5.0	70.43
14	0.85	10.0	70.44
15	0.85	15.0	70.43
16	0.85	20.0	70.43
17	0.85	30.0	70.46
18	0.95	1.0	66.20
19	0.95	5.0	66.27
20	0.95	10.0	66.27
21	0.95	15.0	66.27
22	0.95	20.0	66.26
23	0.95	30.0	66.26

Рисунок 3.3 – Результати перевірки впливу вхідних параметрів на точність передбачення [14]

Аналізуючи отримані результати, можна зробити висновок про те, що зміна параметру `prior_scale` не мала істотного впливу на прогнозування певного квантиля, а сама точність передбачення збільшувалась в межах використання квантиля P70-P85, але при подальшому збільшенні точність починала погіршуватися.

3.1.2. Ідентифікація порядку ряду Фур'є часового ряду показника стану природного середовища

У моделі Prophet для моделювання сезонних компонентів використовується ряд Фур'є, згаданий раніше в підрозділі 1.4. Для пошуку порядку ряду Фур'є зазвичай використовують наступні підходи [15, 49]: частотний аналіз, аналіз спектра та перебір усіх можливих значень, наприклад від 3 до 12 чи 20, причому останній підхід є більш популярним. Разом із ним, ще визначаються й інші параметри моделі, а тоді кількість комбінацій для повного перебору зростає дуже сильно – часто їх взагалі не можна перебрати усі, коли інші параметри є дробовими числами, а тому потрібні більш точні методи, які, на відміну від інших, добре працюють з даними, що містять сильний шум.

Відповідно, корисним є розроблення підходу до визначення порядку Фур'є у моделі Prophet, що дозволить моделювати сезонну складову часового ряду за наявності значного шуму.

Для вирішення поставленої задачі пропонується наступний підхід [15]:

1) Синтезувати ряд Фур'є для заданого порядку n (решта змінних задається фіксованими) [15, 49]:

$$f(x) = \sum_{i=1}^n a_i \cos ix + b_i \sin ix. \quad (3.1)$$

2) На отриманий набір значень накласти шум із певною дисперсією d_1 ;

3) Використовуючи синтезовані на кроці 2 дані, побудувати модель Prophet. При цьому, відключити налаштування сезонності за замовчуванням та

замінити на заданий порядок n ряду Фур'є;

4) Зробити прогноз для створеної моделі, а на отримані значення додатково накласти шум із заданою дисперсією d_2 ;

5) Повторити кроки 1-4 для різних наборів вхідних параметрів (порядок ряду і дисперсія шумів) для генерування набору тренувальних даних.

6) Використати тренувальні дані, отримані на кроці 5, для ідентифікації порядку ряду з використанням методів машинного навчання.

Було реалізовано функцію, яка, перебираючи в циклі різні значення порядку ряду Фур'є та дисперсії шуму, генерувала дані моделлю Prophet. Приклад роботи даної функції у вигляді графіка зображено на рисунку 3.4.

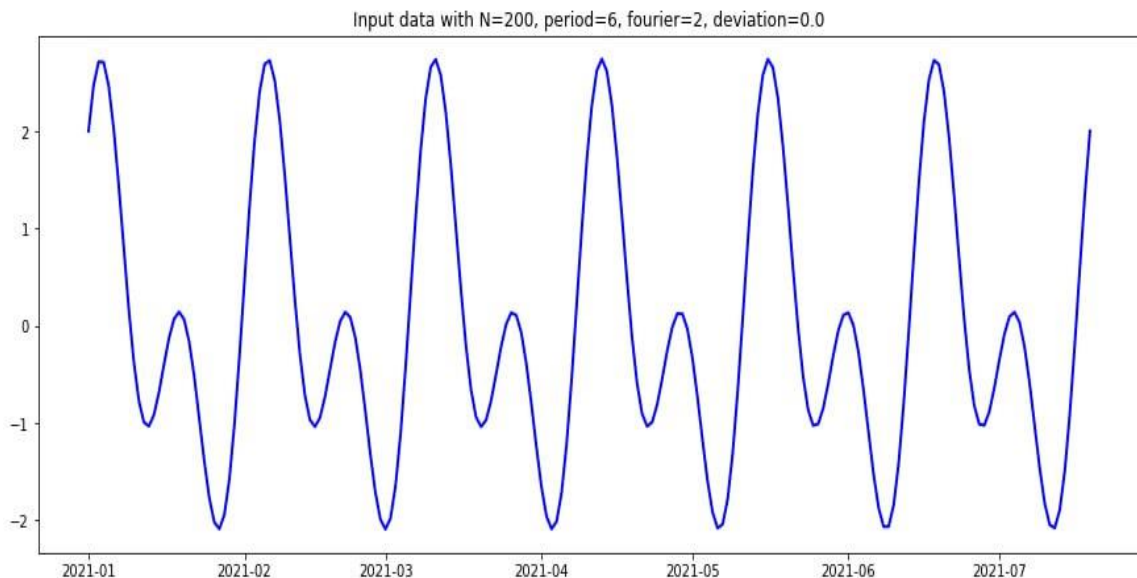


Рисунок 3.4 – Приклад роботи функції для генерування даних за допомогою моделі Prophet [15]

Використовуючи описану функцію, було отримано набір тренувальних даних із більше як 800 різних комбінацій порядку ряду Фур'є та дисперсії шуму (рис. 3.5).

value_1	value_2	value_3	value_4	value_5	value_6	...	value_1
1.212929	1.033774	1.490677	1.588684	1.327723	1.247872	...	-0.438
0.839207	1.214710	1.414985	1.412930	1.583367	1.651833	...	-0.542
0.881902	1.257847	1.499206	0.962251	1.067706	0.983035	...	-0.245
1.235705	0.767356	0.820297	1.253692	1.766596	1.389924	...	-0.892
0.903866	1.106160	1.049172	1.790335	1.210122	1.248544	...	-0.434
...
12.785682	5.139152	0.473676	2.031674	-0.133421	-0.055842	...	-1.828
13.039057	4.370853	-0.426115	1.719395	-0.604905	-0.593807	...	-0.983
12.363012	4.422465	0.239665	2.585668	0.533538	-0.348935	...	0.1629
13.145853	4.489238	-0.905436	0.917066	1.230329	-0.708511	...	-0.987
12.532139	4.957577	-1.328651	1.285472	-0.185101	-0.434669	...	-0.967

Рисунок 3.5 – Сформовані тренувальні дані [15]

Тренувальні дані були використані для ідентифікації порядку ряду Фур'є з використанням методу Random Forest Regressor. Для тестування було використано вибірково взяті 20% даних, які не використовувались для тренування моделі. Відносна похибка визначення порядку Фур'є склала 2-3%, що є дуже гарним показником. Зауважимо, що у разі збільшення дисперсії шуму, ця похибка суттєво зростає, тому цей метод має обмежене застосування.

Загалом, аналіз показав, що традиційні підходи не є достатньо ефективними і варто розробити новий метод, що дозволить вирішити поставлену задачу.

3.1.3. Розроблення нового методу

Для розв'язання поставленої задачі пропонується новий метод, в якому максимально автоматизується кожен етап традиційного розв'язання з використанням автоматичних процедур та ряду критеріїв.

Етап 1. Визначення кількості та періоду можливих видів сезонності.

У роботі [75, 76] запропоновано будувати графік усіх варіантів відношень амплітуд $S_y(P)$ з періодом P від 1 до 50% від кількості усіх значень ряду:

$$S_y(P) = \frac{S(P)}{Y}, \quad (3.2)$$

де $P = 1 \dots 0.5N$, N – кількість значень ряду.

Приклад такого графіку (назвемо його «декомпозиційною кривою») $S_y(P)$ наведено на рис. 3.6.

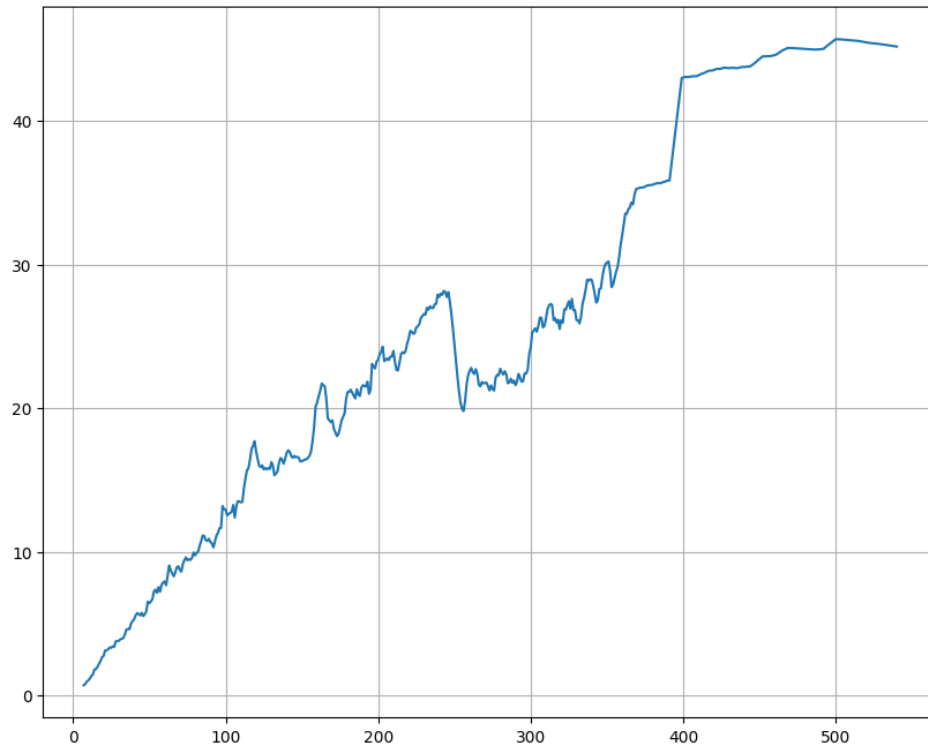


Рисунок 3.6 – Декомпозиційна крива часового ряду значення РМ10 за даними станції № 650 мережі громадського моніторингу атмосферного повітря EcoCity за 2020-2023 рр. [8, 76]

Пропонується згладити декомпозиційну криву та знайти усі номери кроків, яким відповідають локальні максимуми, тобто місця, де зростання значення $S_y(P)$ змінюється на зменшення, це і будуть періоди P_i [8, 76], тобто мають місце такі умови щодо першої та другої похідних:

$$P_i : S'_Y(P_i) = 0, \\ S''_Y(P_i) < 0, \text{ де } i = 1 \dots \phi. \quad (3.3)$$

Етап 2. Визначення параметрів видів сезонності.

Власний досвід роботи з моделлю Prophet протягом останніх декількох років показав, що у випадку, коли у моделі поєднується декілька видів сезонності з різним періодом, для мінімізації ризику перенавчання ефективними є такі прийоми [8]:

- сезонність з найменшим періодом варто описувати рядом Фур'є найбільшого порядку n_{\max} і з найменшою регуляризацією $R_{s\max}$, тобто вона має бути найбільш негладкою – найкраще припасовуватись до ряду;
- порядок n_i ряду Фур'є варто брати між $n_{\min} = 2$ і $n_{\max} = 12$, інші межі (наприклад, 0, 1 чи 20) – менш ефективні;
- регуляризацію сезонності варто брати між $R_{s\min} = 0.1$ і $R_{s\max} = 0.5$, інші межі (наприклад, 0,01, 0,05 чи 0,9) – менш ефективні;
- у разі вибору між різними варіантами сезонних складових першими слід враховувати складові з більшим періодом, послідовно додаючи ті, в яких він – менший.

Ці прийоми дають можливість вивести універсальні співвідношення для розрахунку параметрів усіх видів сезонності. Зокрема, пропонуються такі нові співвідношення для i -ої складової, враховуючи, що період P_0 є найбільшим значенням [8]:

$$n_i = \frac{n_{\min} - n_{\max}}{\phi - 1} i + n_{\max},$$

$$R_{si} = \frac{R_{s\max} - R_{s\min}}{\phi - 1} i + R_{s\min}, \text{ де } i = 1 \dots (\phi - 1). \quad (3.4)$$

Етап 3. Визначення параметрів аномалій.

Для забезпечення більшої універсальності методу пропонується спрощення формули (3.4) шляхом відмови від різних видів аномалій. Крім того, пропонується під видами аномалій розрізняти не різні за формалізацією чи природою аномалії, а аномалії у різні періоди чи частини даних ряду [8]:

$$y(t) = g(R_g, t) + \prod_{i=1}^{\phi} s(R_{si}, P_i, n_i, t) + h\left(\bigcup_{i=1}^{\phi} \bigcup_{j=1}^{\phi} H_{ij}(R_h, t)\right) + \varepsilon_t, \quad (3.5)$$

тобто для кожного періоду, визначеного на попередньому етапі, пропонується робити послідовні вибірки по одному періоду, скільки їх має місце у ряді (1-й, 2-й, 3-й..., останній може бути неповним) і в кожному окремо шукати аномалії, а потім усі об'єднувати в єдину множину. Усі аномалії брати з нульовим вікном, тобто вважати, що вони не впливали до чи після, а якщо впливали, то це – інші аномалії, які треба враховувати окремо. Регуляризацію для усіх аномалій пропонується брати однакову R_h , наприклад, 1 або 10, або враховувати як гіперпараметр.

Однак, крім цього, потрібний критерій для перевірки чи дійсно вибрані у такий спосіб точки є аномаліями. Тому пропонується для кожної вибірки для другої складової виразу (3.5) перевіряти нульову гіпотезу [78, 79]:

H_0 – нульова гіпотеза про те, що відібрана множина значень не є аномаліями відносно інших значень вибірки;

H_1 – альтернативна гіпотеза про те, що відібрана множина значень є аномаліями відносно інших значень вибірки.

У разі, якщо нульова гіпотеза підтверджується, тоді варто не враховувати знайдені точки як аномальні.

Етап 4. Ідентифікація гіперпараметрів.

На етапах 1-3 визначаються більшість параметрів, але частина має залишатись гіперпараметрами для формування множини можливих моделей і вибору оптимальної за точністю. Критерієм точності може бути відносна похибка або відомі метрики: середня абсолютна похибка (англійською «Mean absolute error» або скорочено – MAE), середньоквадратична похибка (англійською «Mean squared error» або скорочено – MSE) та коефіцієнт детермінації (зазвичай, позначається як «r2_score», або «R2» чи «r2») [55].

Гіперпараметрами моделі (9) можуть бути такі:

1) регуляризація тренду R_g (має бути побільше – на рівні $R_{s \min}$);

- 2) кількість ϕ видів сезонності, які варто брати до уваги з числа відібраних на етапі 1;
- 3) граничні значення порядку Фур'є n_{\min} і n_{\max} та регуляризації $R_{s \min}$ і $R_{s \max}$;
- 4) регуляризація аномальної складової R_h .

Якщо ж усі їх задати числами, тоді варто вибирати між моделями, які враховують різну кількість комбінацій видів сезонності.

Блок-схема алгоритму запропонованого методу наведена на рис. 3.7.

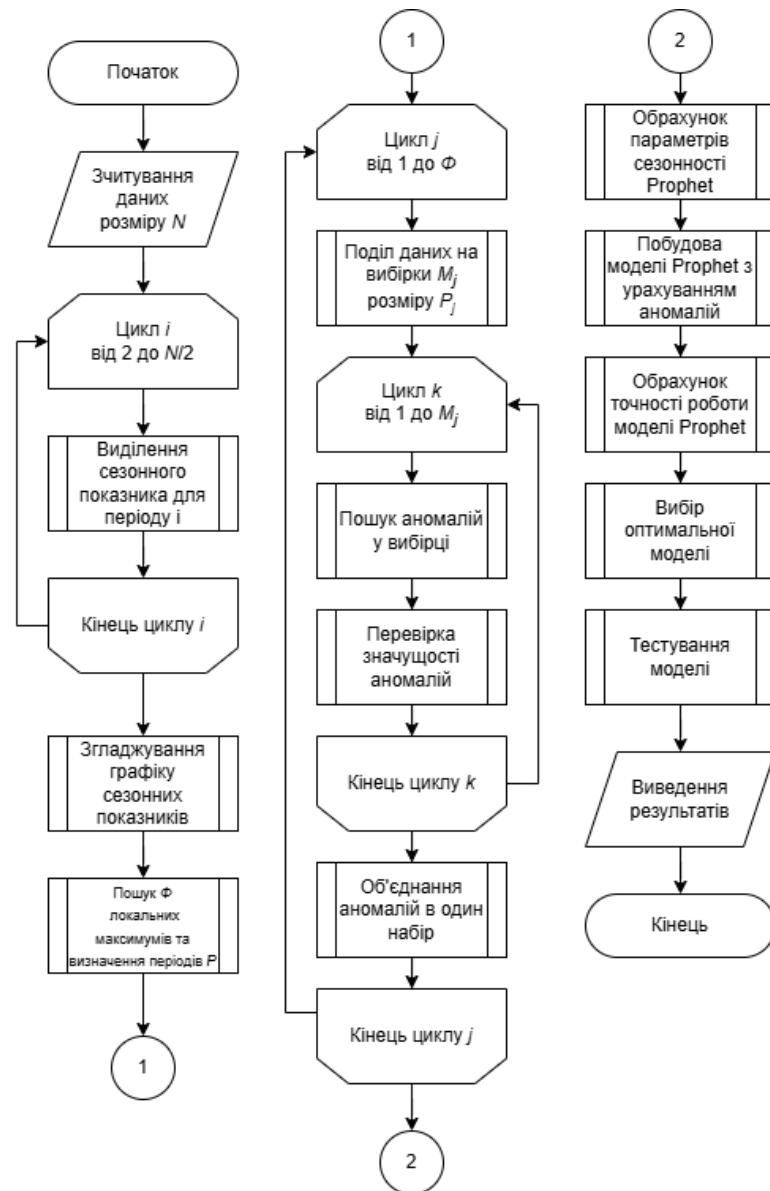


Рисунок 3.7 – Блок-схема алгоритму методу ідентифікації параметрів сезонності та аномалій періодичного часового ряду [8, 77]

Для випробування даного методу було використано дані показник пилу PM10 (мікроскопічні тверді частинки) за 3 роки (рис. 3.8), отримані з мережі громадського моніторингу EcoCity, описаної раніше у пункті 1.1.4.

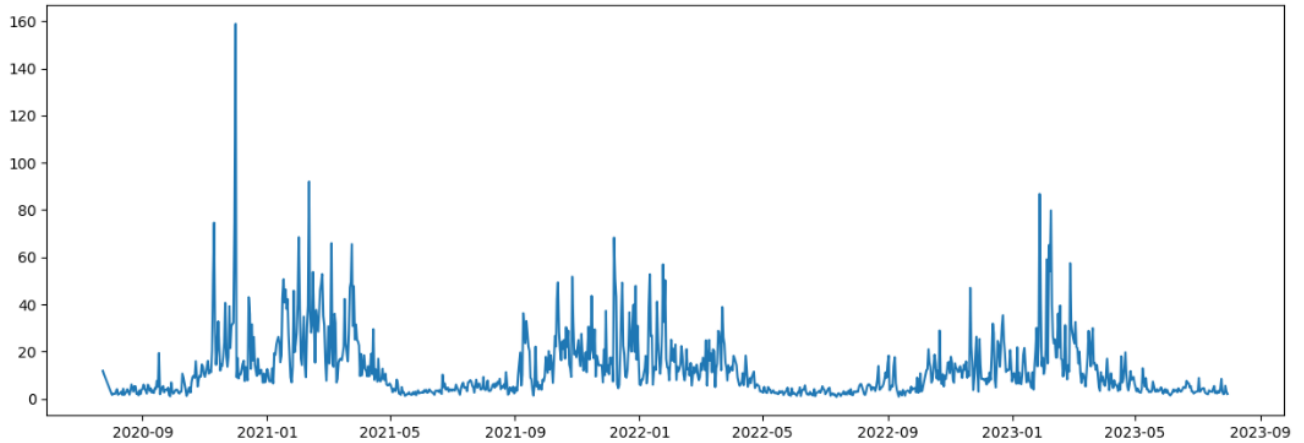


Рисунок 3.8 – Дані показника PM10 станції у смт Турбів Вінницького району [8, 77]

Застосуємо запропонований метод.

Етап 1. Декомпозиційний графік для ряду на рис. 3.8 наведено на рис. 3.6. На рис. 3.9 наведено результат його згладжування фільтром Баттерворта з використанням Python-бібліотеки `scipy.signal`. Пошук локальних максимумів методом `find_peaks` бібліотеки `scipy.signal` дозволив з'ясувати, що $\phi = 3: P_1 = 230, P_2 = 419, P_3 = 507$ діб.

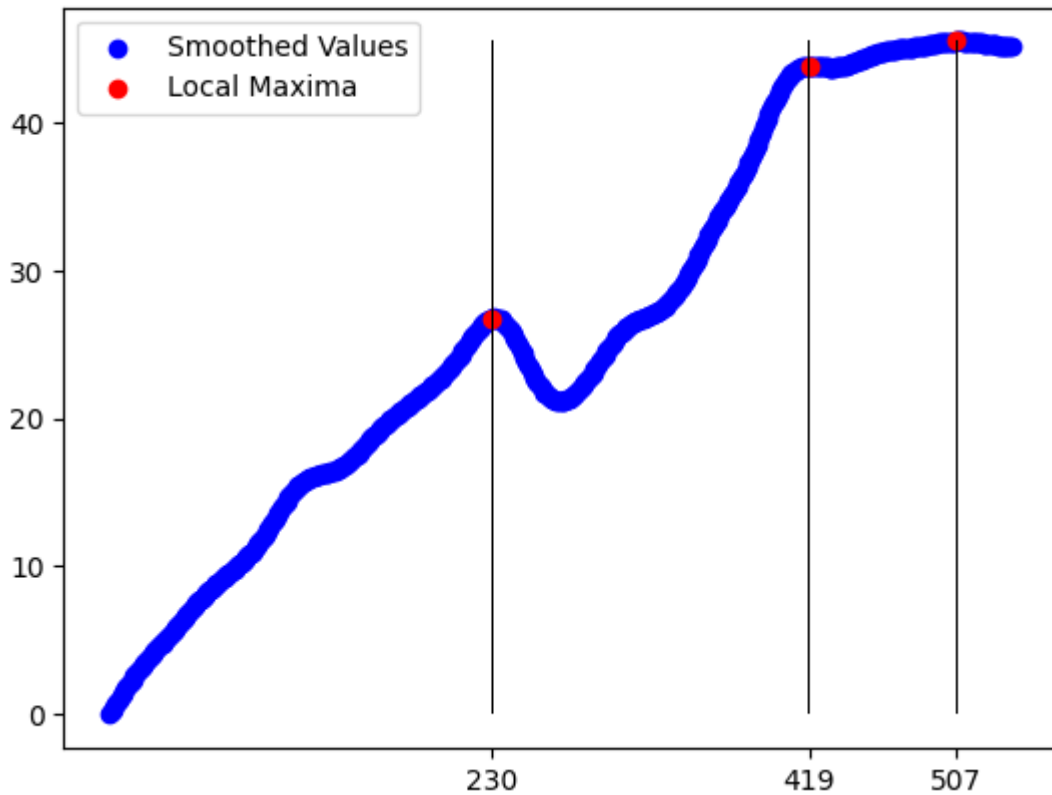


Рисунок 3.9 – Локальні максимуми для графіка з рис. 3.6, згладженого методом Баттерворта [8, 77]

Етап 2. Визначено параметри сезонності з 3-ма різними періодами. Використовуючи формули (3.4), обчислено інші параметри цих сезонних складових (рис. 3.10).

```

period: 507, prior_scale: 0.1, fourier_order: 2
period: 419, prior_scale: 0.3, fourier_order: 7
period: 230, prior_scale: 0.5, fourier_order: 12

```

Рисунок 3.10 – Налаштування сезонності для моделі Prophet [8, 77]

Етап 3. Визначено параметри аномалій. Вхідний набір даних було розділено на рівні частини, де розмір частини відповідає визначеному на етапі 1 періоду. І ця операція повторюється для кожного виду сезонності окремо. Розглянемо для прикладу сезонну складову з періодом P_2 . Оскільки $P_2 = 419$, а загальна кількість

даних складає 1080, то отримуємо дві вибірки розміром 419 та одну неповну вибірку з 186 даними. Для кожної з вибірок, застосовуючи метод Isolation Forest на основі дерев рішень, визначаємо аномалії, а тоді – перевіряємо нульову гіпотезу щодо того, що ці значення не є аномальними. В нашому випадку для перших двох вибірок нульова гіпотеза не підтвердилась і, відповідно, вважаємо, що аномалії на цих двох вибірках є статистично значущими. Відповідно, як було зазначено вище, ці аномалії додаються в загальну вибірку усіх аномалій для відповідної періодичної складової. Результат пошуку аномалій зображено на рисунку 3.11, де червоними точками позначені відповідні елементи графіку показника РМ10, які було ідентифіковано як аномалії. Для частини знайдених потенційних аномалій нульова гіпотеза підтвердилась, а тому вони не були додані до множини аномальних.

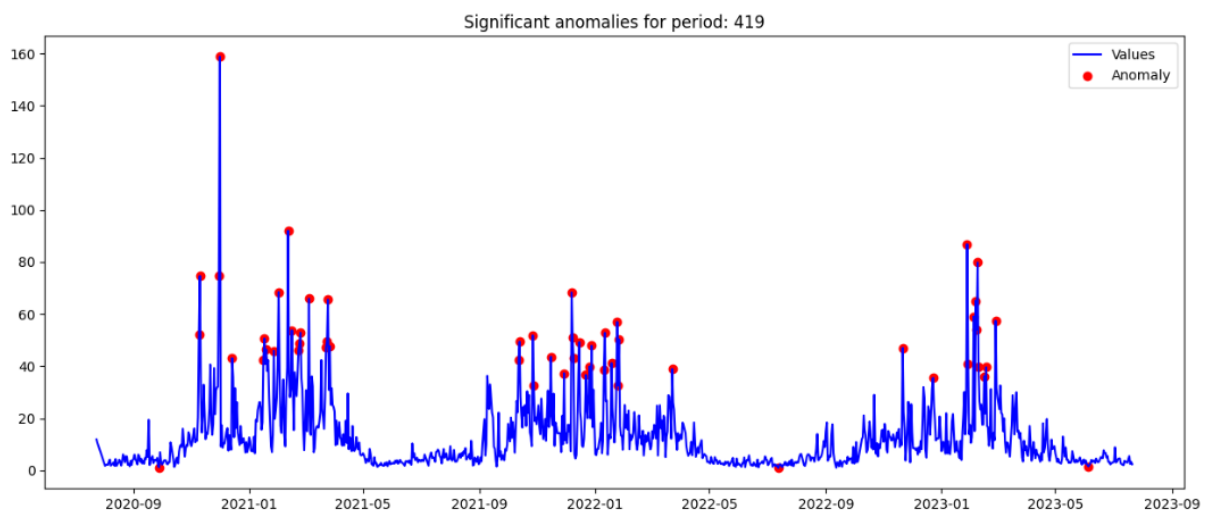


Рисунок 3.11 – Результат пошуку аномалій показника РМ10 для періоду 419 [8, 77]

Етап 4. Сформувавши усі параметри на попередніх етапах, побудовано модель Prophet та, використовуючи метрики, зазначені раніше, обраховано точність роботи моделі Prophet для різних варіантів (рис. 3.12):

Варіант 0 «Without default seasonality». Модель FB Prophet з параметрами за замовчуванням, але без сезонних складових.

Варіант 1. Варіант 0 з 3-ма видами сезонності (англійською «default seasonality»), які ідентифікуються за замовчуванням (добова, тижнева та річна).

Варіант 2. Варіант 0 з 3-ма визначеними на етапі 1 сезонними складовими одночасно ($\phi = 3$ з періодами $P_1 = 230, P_2 = 419, P_3 = 507$ діб) (англійською «custom seasonality»), але без аномалій («no holidays», оскільки в моделі FB Prophet аномалії прийнято умовно називати «свята»).

Варіант 3. Варіант 2 з усіма аномаліями («all anomalies»), визначеними у вибірках, розміром 230 діб.

Варіант 4. Варіант 3, але тільки зі статистично значущими аномаліями (англійською «significant anomalies»).

Варіант 5. Варіант 2 з усіма аномаліями, визначеними у вибірках, розміром 419 діб.

Варіант 6. Варіант 5, але тільки зі статистично значущими аномаліями.

Варіант 7. Варіант 2 з усіма аномаліями, визначеними у вибірках, розміром 507 діб.

Варіант 8. Варіант 7, але тільки зі статистично значущими аномаліями.

	name	r2	mean_squared_error	mean_absolute_error
0	Without default seasonality	0.264846	126.149253	7.053936
1	Default seasonality	0.401954	102.622188	6.125823
2	Custom seasonality + no holidays	0.431548	97.543824	5.901587
3	Period: 230 + all anomalies + custom seasonality	0.673193	56.078709	4.751458
4	Period: 230 + significant anomalies + custom s...	0.673193	56.078709	4.751458
5	Period: 419 + all anomalies + custom seasonality	0.699386	51.584003	4.647670
6	Period: 419 + significant anomalies + custom s...	0.699386	51.584003	4.647670
7	Period: 507 + all anomalies + custom seasonality	0.651996	59.715948	4.904736
8	Period: 507 + significant anomalies + custom s...	0.669741	56.671101	4.824506

Рисунок 3.12 – Таблиця порівняння точності апроксимації моделі Prophet для різних варіантів її налаштування (кількості сезонних складових) і параметрів [8, 77]

З рис. 3.12 видно, що для двох із трьох видів сезонності точність між статистично значущими аномаліями та усіма наявними аномаліями однакова. Причиною цього стало те, що перевірка нульової гіпотези класифікувала на усіх вибірках цих періодів аномалії як суттєві, відповідно для них ці два набори аномалій – абсолютно однакові. В той же час, для періоду 507 днів перевірка нульової гіпотези відкинула деякі з аномалій і в результаті це дозволило досягти кращої точності роботи моделі Prophet.

Отже, оптимальною моделлю була модель зазначена як Варіант 5 – модель для періоду 419 днів зі статистично значущими аномаліями. З порівняння точності оптимальної моделі з точністю моделі Prophet з параметрами і видами сезонності за замовчуванням (добова, тижнева та річна) видно, що оптимальна модель була значно точнішою, зокрема, за метрикою R^2 (чим ближче до 1, тим – краще) – в 1,7 рази більше, а за метрикою MSE (чим ближче до нуля, тим – краще) – у 2 рази менше. Результат роботи оптимальної моделі наведено на рис. 3.13.

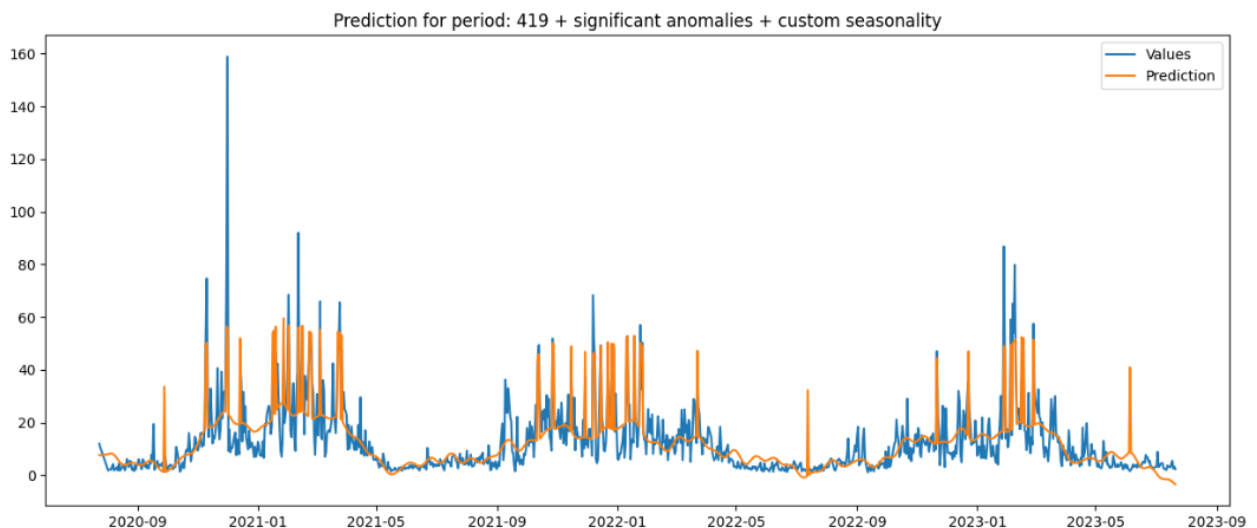


Рисунок 3.13 – Результат роботи оптимальної моделі FB Prophet, ідентифікованої за запропонованим методом [8, 77]

Таким чином, запропонований метод дозволяє вирішити поставлену задачу та розкрити зазначені на початку підрозділу 3.1 види невизначеності (U_{P1} , U_{P2} та U_{T6}).

3.2. Удосконалення методу ідентифікації локальних аномалій в часових рядах

Для побудови інтелектуальної моделі часового ряду на основі моделі Prophet потрібно, передусім, ідентифікувати аномалії. Як було згадано в пункті 1.3.4 та підрозділі 1.6, ці аномалії можуть бути локальними для певного інтервалу часового ряду, що відповідає визначеному в пункті 2.2.3 виду невизначеності U_{S2} .

Для розв'язання задачі пропонується метод, основна ідея якого це гіпотеза про те, що часовий ряд складається з окремих фрагментів між локальними мінімумами тренду – точками, де графік змінює свій напрям зі спадання на зростання, і на кожному з цих фрагментів ряд поводить себе по різному [10]. Відповідно замість застосування певних методів пошуку аномалій до усього ряду, пропонується аналізувати кожен з фрагментів окремо, а потім об'єднувати усі отримані дати аномалії в один набір [10]. Як було сказано раніше, показники стану природного середовища мають циклічність, обумовлену впливом природних факторів. Як приклад подібного ряду, наведемо графік значень показника пилу PM10 за 2020-2023 рр. у Вінницькій області. Графік зображено на рис. 3.14.



Рисунок 3.14 – Часовий ряд значень показника PM10 з приблизним поділом на окремі фрагменти [10]

Як схематично зображено на рис. 3.14, часовий ряд можна розділити на декілька фрагментів, які мають схожу поведінку, обумовлену певними природними факторами. Попередній досвід показує, що розгляд усього ряду як одне ціле не завжди дає гарний результат і доцільним є пошук локальних аномалій в кожному окремому фрагменті декомпозиції ряду, а потім – об'єднати результати по кожному фрагменту в одну вибірку.

Опишемо етапи роботи запропонованого методу.

Етап 1. Виділення тренду часового ряду.

Для виділення тренду скористаємося відомим у Python-бібліотеці «pandas» прийомом – методом «rolling». Цей метод дозволяє згладити коливання даних та виокремити тренд часового ряду. Щоб обмежитися меншою кількістю сегментів та тим самим спробувати захопити більше подібних точок часового ряду, отриману криву тренду додатково згладжуємо певним алгоритмом. Результатом цього етапу для прикладу, наведеного на рис 3.14, є графік зображений на рис. 3.15.

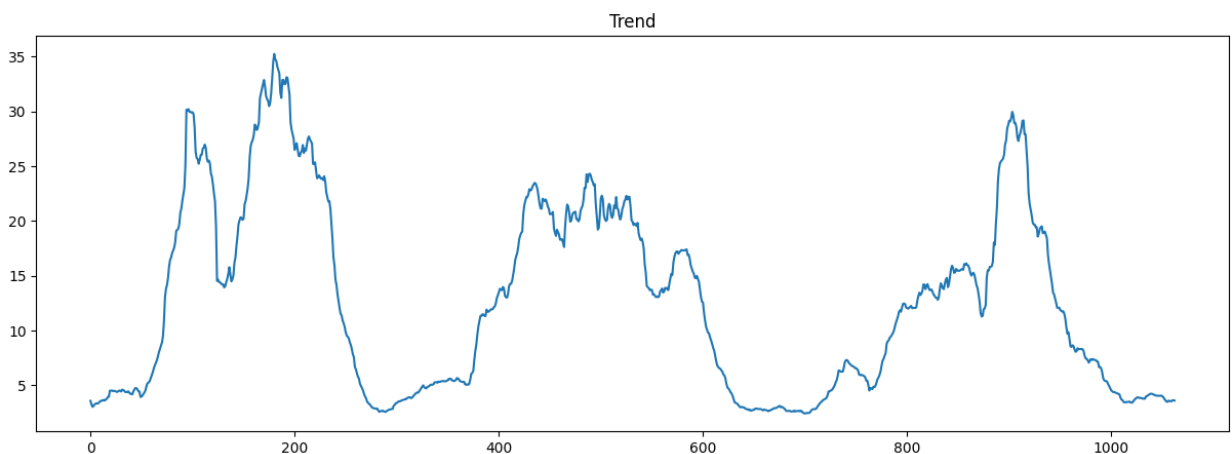


Рисунок 3.15 – Згладжена складова тренду для часового ряду показника РМ10, наведеного на рис. 3.14 [10]

Етап 2. Пошук локальних мінімумів та максимумів і виділення окремих фрагментів ряду.

На отриманій згладженій кривій тренду часового ряду необхідно знайти локальні мінімуми та локальні максимуми – точки кривої, де спадання змінюється на зростання та навпаки. Крім того, варто знайти та проігнорувати ті локальні максимуми, де відстань до наступного локального максимуму є малою. Таким чином, зможемо відкинути незначні коливання та не брати до уваги невеликі фрагменти, як окремі складові часового ряду, оскільки пошук аномалій на малих часових відрізках не є доцільним. Приклад локальних мінімумів та локальних максимумів для ряду, зображені на рис. 3.16.

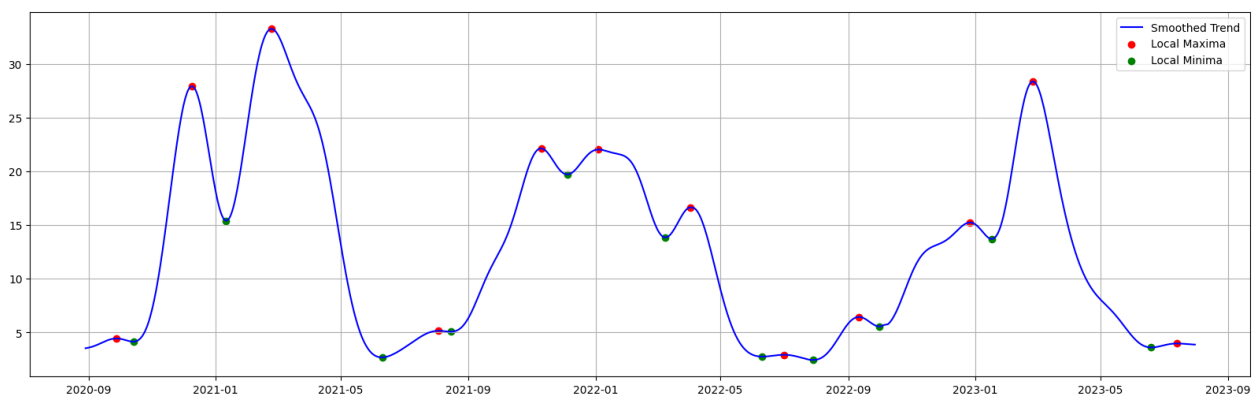


Рисунок 3.16 – Локальні мінімуми та локальні максимуми на кривій тренду часового ряду з рис. 3.15 [10]

Виділивши з загального списку вагомі локальні мінімуми, отримуємо точки по яких розділяємо наш часовий ряд на декілька фрагментів. Кількість цих фрагментів залежить від самого часового ряду, параметрів згладжування та визначення вагомості. Останні два значення можуть бути винесені як параметри налаштування методу, задля більш гнучких можливостей аналізу відповідного часового ряду.

Етап 3. Пошук аномалій на окремих сегментах часового ряду, використовуючи комбінації методів.

Пропонується одночасне використання декількох методів за різними стратегіями їх комбінування. Подібне поєднання може дозволити досягти більшої точності, оскільки один метод може знайти ті аномалії, які були пропущені іншим

методом або ж – навпаки, два методи будуть валідувати один одного і відкидати хибні аномалії.

Для цього пропонуються три стратегії поєднання методів:

Варіант 1. Одночасне застосування двох методів та об'єднання їх результатів. Нехай вхідний набір даних це X , тоді аномальні значення, отримані даним комбінованим методом можна позначити як [10]:

$$V_1(X) = M_1(X) \cup M_2(X), \quad (3.6)$$

де $M_1(X)$ – аномальні дані, отримані першим методом для даних X , а $M_2(X)$ – аномальні дані, отримані другим методом.

Варіант 2. Одночасне застосування двох методів та знаходження перетину їх результатів. Нехай вхідний набір даних це X , тоді аномальні значення, отримані даним комбінованим методом, можна позначити як [10]:

$$V_2(X) = M_1(X) \cap M_2(X), \quad (3.7)$$

Варіант 3. Покрокове застосування обох методів, де після застосування першого методу, з набору вхідних даних видаляються аномальні дані і до отриманого набору застосовується другий метод. Результат роботи обох методів об'єднується в один набір. Нехай вхідний набір даних це X , а A_1 – аномальні значення, отримані за допомогою першого методу пошуку аномалій, тоді аномальні значення, отримані даним комбінованим методом, можна позначити як [10]:

$$V_3(X) = A_1 \cup M_2(X \setminus A_1). \quad (3.8)$$

Усі можливі комбінації методів пошуку аномалій потрібно застосовувати на кожному фрагменті ряду, виділених на попередньому кроці. Таким чином, слідуючи гіпотезі, що часовий ряд показників стану природного середовища поводить себе по-різному на кожному окремому відрізьку ряду, зможемо проаналізувати фрагмент окремо від інших частин часового ряду та ідентифікувати локальні аномалії, властиві безпосередньо цьому відповідному фрагменту ряду.

Блок-схема алгоритму запропонованого методу наведена на рис. 3.17.

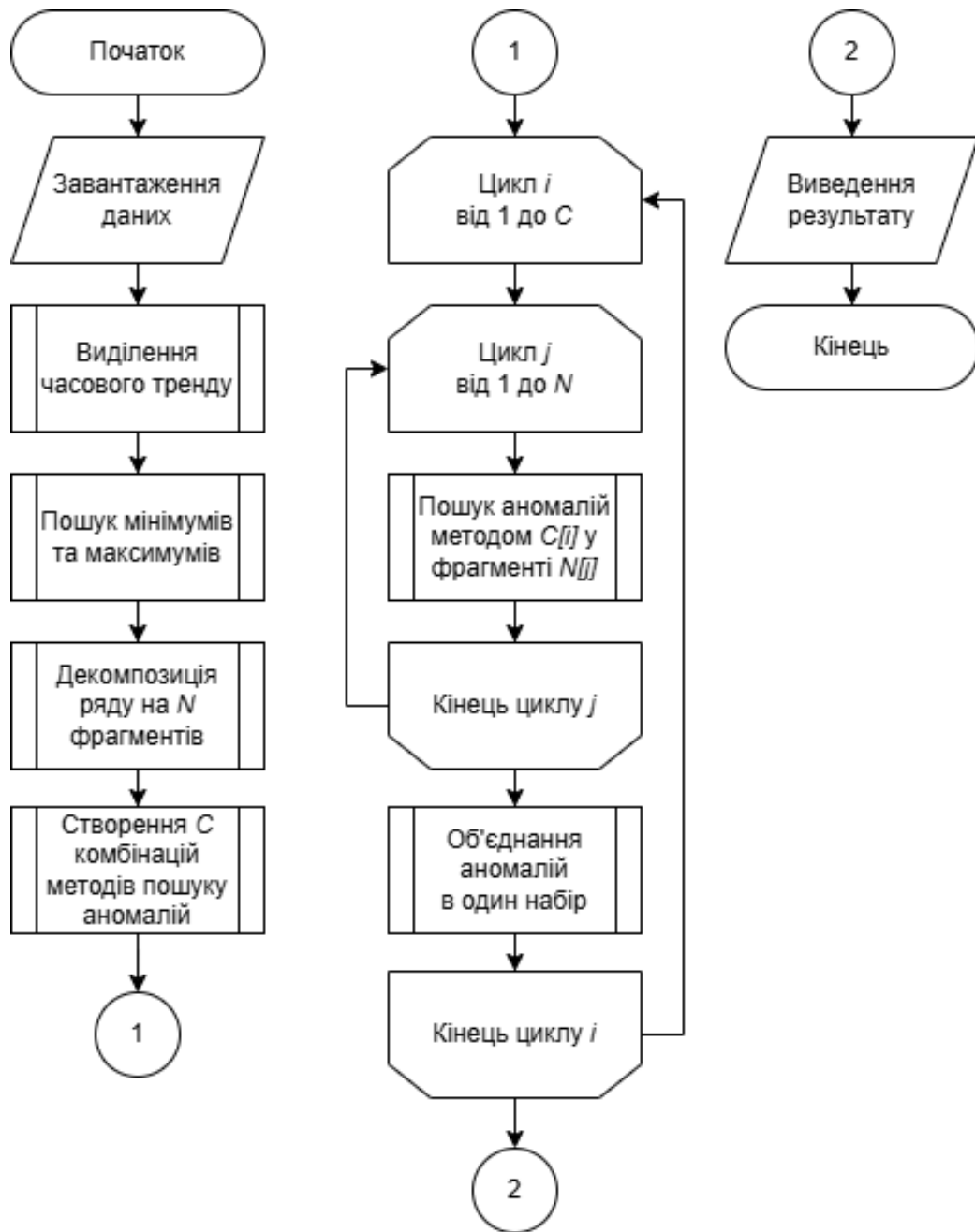


Рисунок 3.17 – Блок-схема алгоритму запропонованого методу ідентифікації локальних аномалій [10]

Для перевірки методу знову використаємо дані показників якості атмосферного повітря від мережі громадського моніторингу EcoCity. Було обрано одну зі станцій, розташованих у місті Вінниця, та показник пилу PM2.5 за період 2019-2022 рр. Графік цих даних наведено на рис. 3.18.

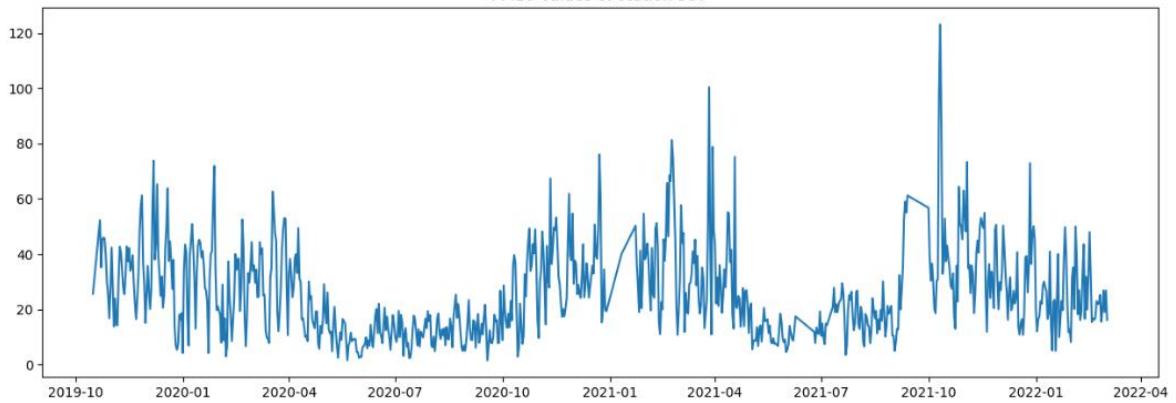


Рисунок 3.18 – Дані показника PM2.5 станції у місті Вінниця [10]

Покроково застосуємо метод ідентифікації.

Етап 1. Знайдемо складову тренду для даних, зображених на рис. 3.17. та згладимо їх за допомогою фільтру Баттерворта – одного з модулів Python-бібліотеки `scipy.signal`. Крива тренду та її згладжений варіант наведено на рис. 3.19.

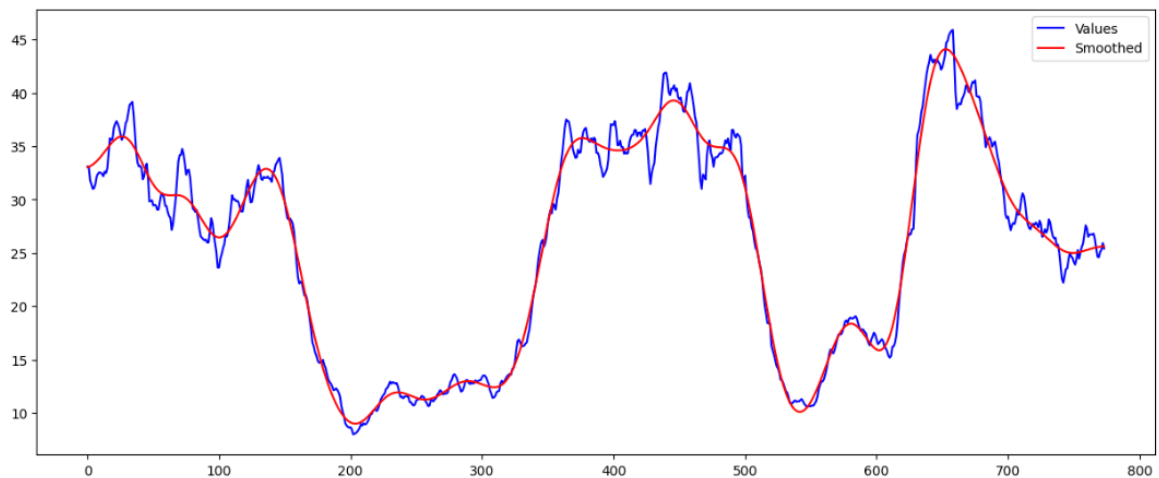


Рисунок 3.19 – Складова тренду та її згладжена версія для часового ряду, зображеного на рис. 3.17 [10]

Етап 2. Визначимо локальні мінімуми та локальні максимуми на згладженій кривій тренду. Для цього застосуємо методи з раніше згаданої Python-бібліотеки `scipy.signal`. Отримані значення зображено на рис. 3.20.

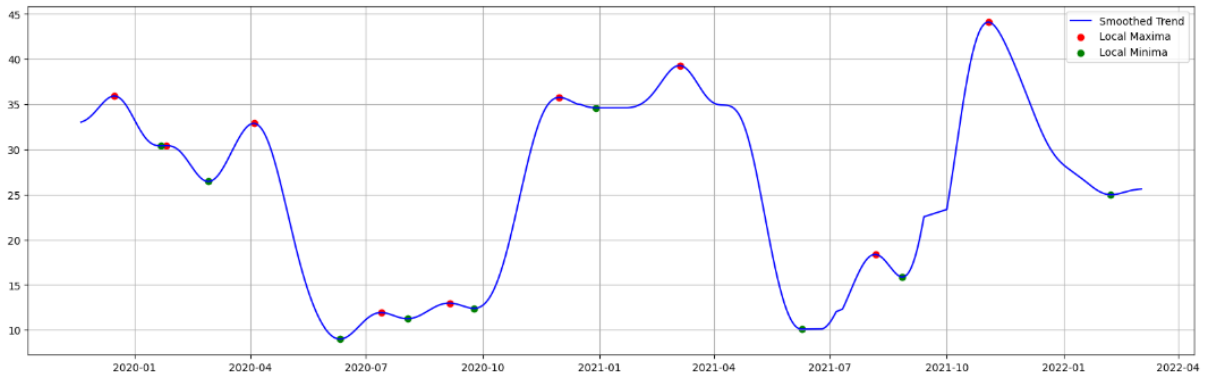


Рисунок 3.20 – Локальні максимуми та локальні мінімуми кривої тренду ряду з рис. 3.19 [10]

Порівнюючи сусідні мінімуми та максимуми, необхідно знайти ті, які не суттєво відрізняються один від одного. Такі точки локальних мінімумів, як було описано в попередньому розділі, будуть ігноруватися та не будуть використовуватися для визначення сегментів часового ряду [10]. Важливі локальні мінімуми зображено на рис. 3.21.

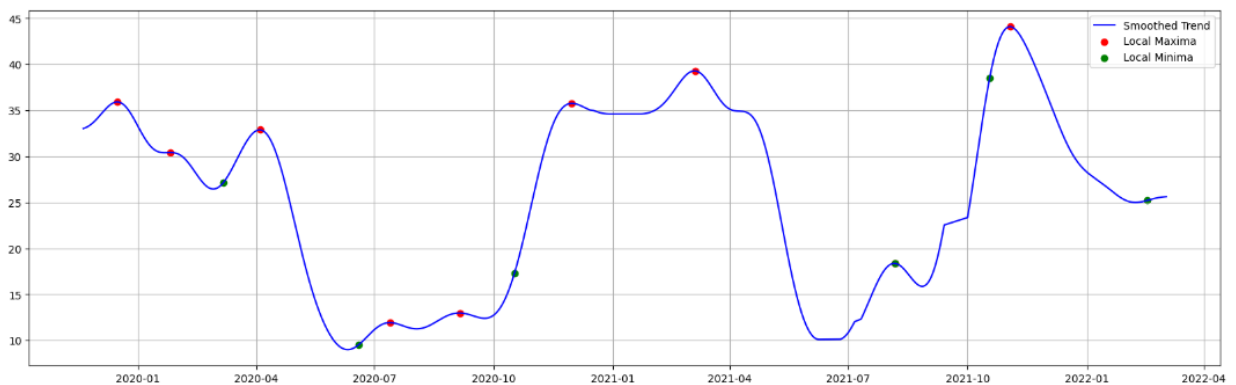


Рисунок 3.21 – Значущі локальні мінімуми кривої тренду ряду з рис. 3.19 [10]

Розділимо часовий ряд на окремі сегменти, використовуючи отримані точки значущих локальних мінімумів. Результат зображено на рис. 3.22.

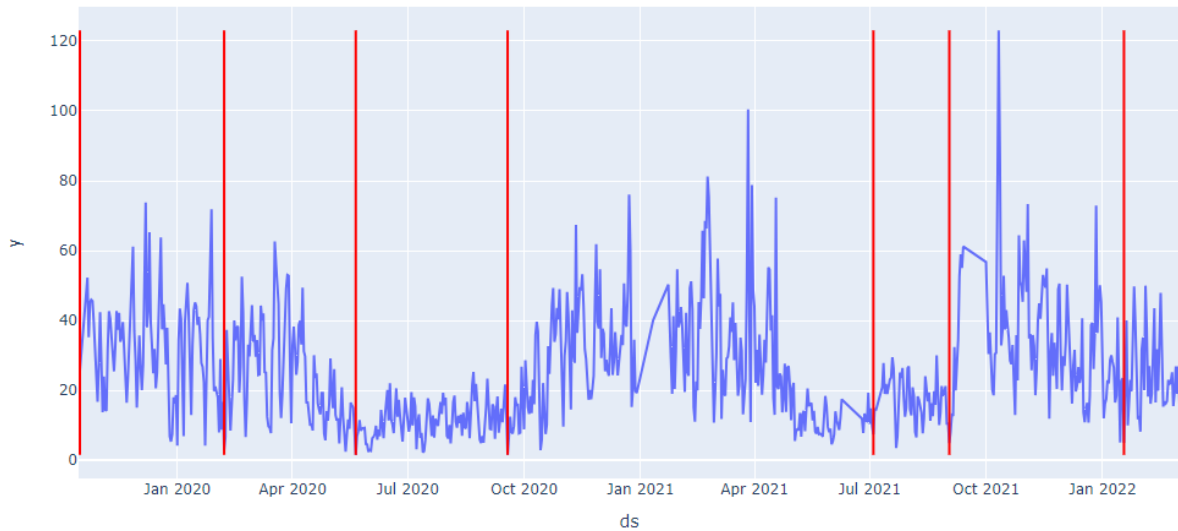


Рисунок 3.22 – Окремі сегменти часового ряду розділені червоними лініями – навпроти точок локальних мінімумів [10]

Етап 3. Виконаємо пошук аномалій на кожному окремому фрагменті часового ряду, використовуючи запропонований комбінований підхід. Для цього використаємо 7 методів пошуку аномалій, описаних раніше в порівняльному аналізі. Крім використання різних комбінацій цих методів, варто також спробувати застосувати кожен з них окремо і тим самим порівняти чи дає комбінування методів якусь перевагу, порівнюючи з окремими методами пошуку аномалій. Як було описано, комбінування методів буде виконуватися за трьома стратегіями. Перші дві стратегії комбінування не залежать від порядку використання методу пошуку, що відповідає вибірці, яка називається комбінація без повторень [10]:

$$N_1 = \frac{n!}{r!(n-r)!}, \quad (3.9)$$

де n – загальна кількість методів, r – кількість методів, які застосовуються разом, а N_1 – загальна кількість можливих унікальних комбінацій.

В нашому випадку є усього 7 методів пошуку аномалій і кожна комбінація містить 2 методи при застосуванні першої стратегії комбінування методів пошуку аномалій. Підставивши ці числа в формулу (3.9), отримуємо 21 можливу

комбінацію. Така сама кількість комбінацій є і у разі застосування другої стратегії комбінування методів пошуку аномалій.

В той же час, у разі застосування третьої стратегії комбінування методів пошуку аномалій, результат буде різним у залежності від порядку застосування методу. Таким чином, загальну кількість комбінацій можна знайти за формулою [10]:

$$N_2 = n \times (n - 1), \quad (3.10)$$

де n – загальна кількість методів, а N_2 – загальна кількість комбінацій з можливими перестановками.

Оскільки для перевірки використовуються 7 різних методів пошуку аномалій, то загальна кількість комбінацій згідно формули (3.10) дорівнює 42. Крім того, також застосовуємо кожен з 7 методів окремо, що додає ще 7 варіантів пошуку аномалій. Таким чином, загалом цей приклад має 91 варіант. Реалізувавши цей перебір за допомогою Python, одержуємо результат, зображений на рис. 3.23.

	name	anomaly_values
0	Hampel	[13, 49, 75, 76, 93, 130, 170, 215, 239, 260, ...
1	Moving Average	[46, 100, 110, 125, 148, 152, 153, 161, 260, 2...]
2	Isolation Forest	[46, 98, 148, 149, 305, 316, 332, 465, 497, 59...]
3	LOF	[1, 34, 35, 36, 45, 46, 49, 57, 58, 66, 71, 79...]
4	k-neighbors	[1, 34, 35, 36, 45, 46, 49, 58, 79, 97, 98, 10...]
...
86	SVM → Moving Average anomalies	[1, 6, 28, 36, 45, 46, 47, 49, 50, 54, 61, 63,...]
87	SVM → Isolation Forest anomalies	[1, 6, 28, 45, 46, 47, 49, 50, 61, 63, 71, 77,...]
88	SVM → LOF anomalies	[1, 6, 28, 32, 33, 45, 46, 47, 49, 50, 56, 61,...]
89	SVM → k-neighbors anomalies	[1, 6, 27, 28, 31, 32, 33, 45, 46, 47, 49, 50,...]
90	SVM → SESD anomalies	[1, 6, 13, 23, 27, 28, 32, 33, 42, 45, 46, 47,...]

Рисунок 3.23 – Результат пошуку аномалій кожним з 91-го варіанту комбінування декількох методів [10]

Для перевірки точності роботи експертним шляхом знайдемо аномальні записи на кожному з фрагментів часового ряду. В ході експертного аналізу було виділено 41 запис, який візуально відрізняється від інших записів відповідного фрагменту часового ряду. Перелік аномалій, знайдених експертним шляхом, зображено на рис. 3.24.

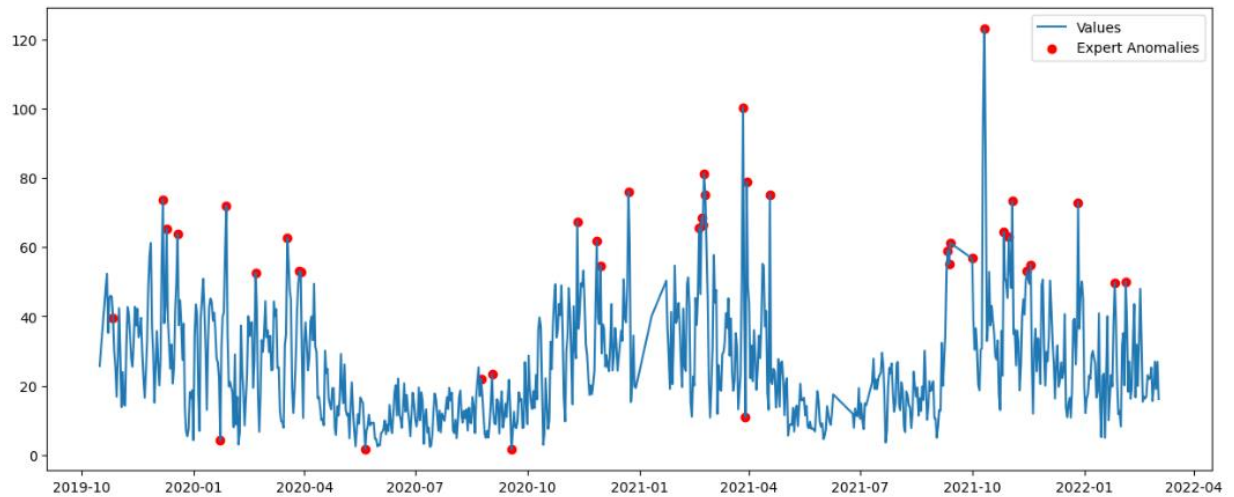


Рисунок 3.24 – Аномалії часового ряду, визначені експертним шляхом [10]

Для перевірки точності кожної з 91 комбінації методів пошуку аномалій, перевіримо наскільки список експертних аномалій відрізняється від відповідного списку аномалій, знайдених програмним шляхом. Для цього створимо список значень, який дорівнює розміру даного часового ряду. Кожен елемент списку це 1, якщо відповідний елемент ряду вважається аномалією або 0, якщо елемент вважається нормальним. Як критерій точності використаємо декілька статистичних метрик: влучність (англійською «precision» або «precision score»), повнота (англійською «recall» або «recall score») та f-міра – показник, який обчислюється через значення повноти та влучності [57]. Обрахувавши ці метрики для кожного зі списків, отримуємо результат, як це показано на рис. 3.25.

	name	anomaly_values	precision_score	recall_score	f1_score
43	LOF \cap k-neighbors anomalies	[1, 34, 35, 36, 45, 46, 49, 58, 79, 97, 98, 10...	0.737571	0.830996	0.775075
18	Isolation Forest + LOF anomalies	[1, 34, 35, 36, 45, 46, 49, 57, 58, 66, 71, 79...	0.705752	0.836943	0.751830
3	LOF	[1, 34, 35, 36, 45, 46, 49, 57, 58, 66, 71, 79...	0.705752	0.836943	0.751830
69	LOF \rightarrow Isolation Forest anomalies	[1, 34, 35, 36, 45, 46, 49, 57, 58, 66, 71, 79...	0.702652	0.836288	0.749021
45	LOF \cap SVM anomalies	[1, 45, 46, 49, 71, 93, 98, 108, 148, 149, 158...	0.737516	0.748739	0.742974
4	k-neighbors	[1, 34, 35, 36, 45, 46, 49, 58, 79, 97, 98, 10...	0.687918	0.844201	0.737270
19	Isolation Forest + k-neighbors anomalies	[1, 34, 35, 36, 45, 46, 49, 58, 79, 97, 98, 10...	0.687918	0.844201	0.737270
47	k-neighbors \cap SVM anomalies	[1, 45, 46, 49, 98, 108, 148, 149, 158, 159, 2...	0.736894	0.736894	0.736894
75	k-neighbors \rightarrow Isolation Forest anomalies	[1, 34, 35, 36, 45, 46, 49, 50, 58, 79, 97, 98...	0.685411	0.843545	0.734748
44	LOF \cap SESD anomalies	[34, 35, 36, 45, 46, 49, 57, 58, 66, 71, 79, 9...	0.750709	0.702670	0.723725
63	Isolation Forest \rightarrow LOF anomalies	[1, 34, 35, 36, 45, 46, 48, 56, 57, 65, 70, 78...	0.681653	0.797477	0.722049
22	LOF + k-neighbors anomalies	[1, 34, 35, 36, 45, 46, 49, 57, 58, 66, 71, 79...	0.669843	0.850147	0.719860
64	Isolation Forest \rightarrow k-neighbors anomalies	[1, 34, 35, 36, 45, 46, 48, 56, 57, 78, 92, 96...	0.668928	0.816579	0.714111
14	Moving Average + LOF anomalies	[1, 34, 35, 36, 45, 46, 49, 57, 58, 66, 71, 79...	0.665431	0.837647	0.713570
76	k-neighbors \rightarrow LOF anomalies	[1, 34, 35, 36, 45, 46, 49, 50, 58, 63, 79, 84...	0.665431	0.837647	0.713570
46	k-neighbors \cap SESD anomalies	[34, 35, 36, 45, 46, 49, 58, 79, 148, 149, 237...	0.742520	0.690170	0.712649

Рисунок 3.25 – Результати обрахунку точності кожного з методів пошуку аномалій у порівнянні з аномаліями, отриманими експертним шляхом, відсортовані за показником f1_score [10]

Як можна помітити, серед найкращих методів часто зустрічається метод LOF. Також високу точність мають методи з використанням Isolation Forest та k-neighbors. Якщо аналізувати які стратегії комбінування методів були найефективнішими, тоді видно, що серед найкращих є усі три підходи.

Як додатковий етап перевірки, спробуємо використати найточніший метод для апроксимації часового ряду. Для цього побудуємо модель Prophet з параметрами за замовчуванням та з набором «holidays», який дозволяє передати в модель список аномальних дат і тим самим покращити роботу моделі. Результати роботи моделі Prophet з параметрами за замовчуванням та з використанням аномалій зображено на рис. 3.26 та рис. 3.27, відповідно.

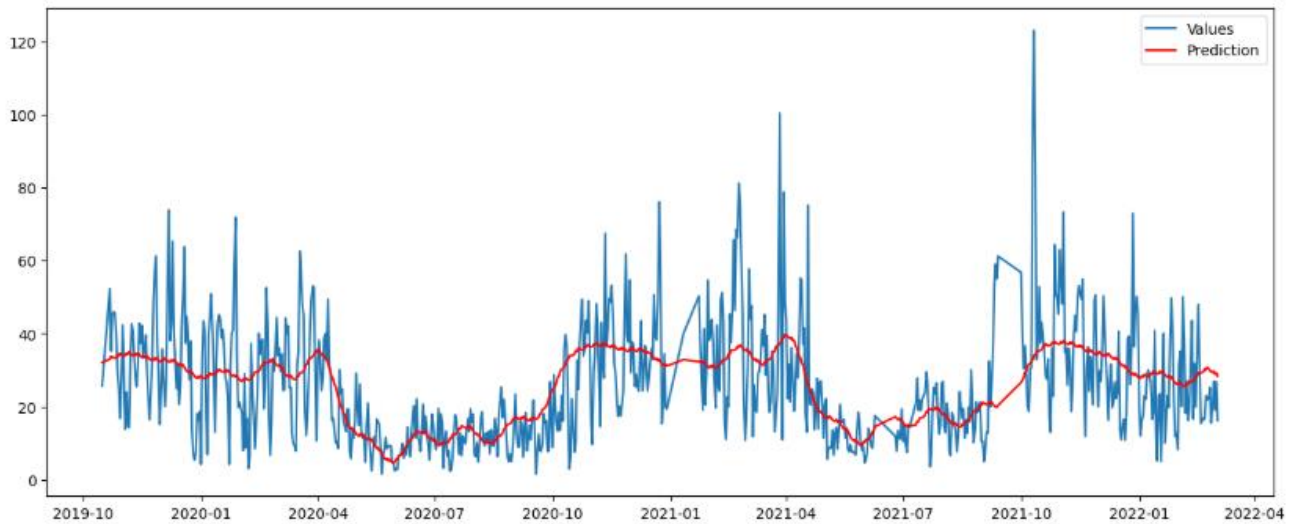


Рисунок 3.26 – Результати роботи моделі Prophet з параметрами налаштування за замовчуванням [10]

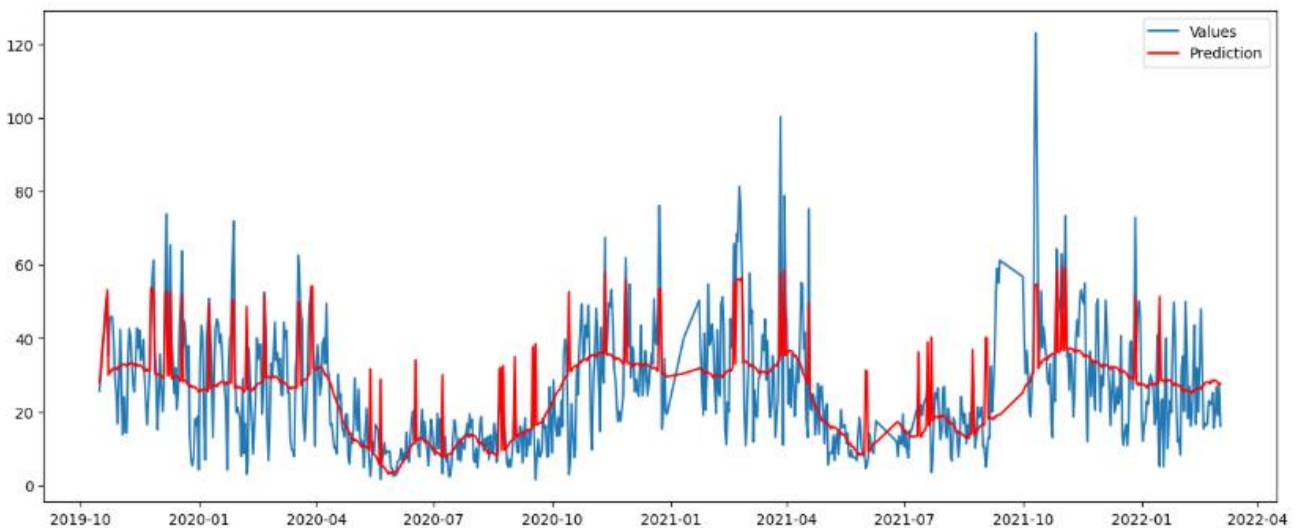


Рисунок 3.27 – Результати роботи моделі Prophet та аномалій, отриманих найточнішим комбінованим методом [10]

Точність апроксимації було обраховано за допомогою середньої абсолютної похибки (англійською «Mean absolute error» або скорочено – MAE) та середньоквадратичої похибки (англійською «Root mean squared error» або скорочено – RMSE) [10]. За цим показником похибка зменшилась на 11% за RMSE та на 8% за MAE, у разі використання ідентифікованих аномалій у порівнянні з моделлю Prophet з параметрами за замовчуванням. Варто зазначити, що обраний

часовий ряд досить важко аналізувати та передбачувати, але все одно вдалось підвищити точність роботи моделі.

Отже, запропонований метод є корисним інструментом системного аналізу стану природного середовища та дозволяє розкрити невизначеність U_{S2} .

3.3. Пошук взаємозв'язків показників стану природного середовища на основі байєсівського моделювання

Однією із проблем аналізу якості атмосферного повітря та інших природних середовищ є складність механізмів розповсюдження забруднювальних речовин, які можуть залежати від різних природних та географічних умов. Традиційні методи, наприклад моделі дифузії забруднення (Ейлерові, Гауссові або Лагранжеві моделі), не завжди дозволяють отримати точну інформацію щодо того як зміна значень показників на одній станції моніторингу може впливати на зміну відповідних показників на сусідніх станціях [80]. Ця проблема відповідає видам невизначеності U_{ST1} та U_{ST2} , що були визначені в пункті 2.2.3.

Наявність певного аномального показника забруднення атмосферного повітря в певній географічній точці, зазвичай, повинна впливати та проявлятися в інших місцях, наприклад, підвищення рівня пилу на одній зі станцій моніторингу може передувати подібному підвищенню рівня забруднення на іншій станції, за рахунок напрямку вітру, який переносить пил.

Одним із можливих підходів для аналізу залежності між різними чинниками забруднення або між різними станціями моніторингу цього забруднення є байєсівські мережі, які дозволяють ідентифікувати причинно-наслідкові моделі на основі існуючих даних [81, 82]. Метод дозволяє визначити яким чином зміни значень показника впливають на зміну цього або інших показників в межах певної географічної зони, що, у свою чергу, може допомогти локалізувати джерело забруднення та вжити відповідні заходи для зниження ризиків для здоров'я населення та захисту довкілля.

В роботі [9] було застосовано байєсівський аналіз для моделювання чутливості людей до певних грибкових алергенів, які присутні в атмосферному повітрі. Байєсівська мережа застосовувалась для визначення ймовірності розвитку комбінованої чутливості в окремих пацієнтів до молекулярних компонентів грибів. Цей метод може бути застосований і для аналізу взаємовпливу показників стану природного середовища, а саме для моделювання імовірностей зміни значення показника стану природного середовища за умов зміни значень відповідного показника на сусідніх станціях мережі моніторингу.

Пропонується наступний алгоритм [9, 83]:

1. *Підготовка даних.* Дані моніторингу, зазвичай, подаються в абсолютних величинах, але для байєсівського моделювання, переважно, використовують бінарні значення. Набір даних стану атмосферного повітря необхідно привести до бінарного формату, за рахунок порівняння значення показника у певній момент часу з його попереднім спостереженням – якщо поточне значення більше попереднього, то воно заміняється на «1», якщо ж значення менше або дорівнює – «0».

2. *Побудова спрямованого ациклічного графа.* Для отриманих даних проводиться вивчення структури, що дозволяє виявити залежності між існуючими змінними та сформувавши спрямований ациклічний граф.

3. *Вивчення параметрів.* Вивчення параметрів даних – це оцінювання значень умовних розподілів ймовірностей, яке проводиться на графі, отриманому на попередньому етапі. Результатом є набір ймовірностей зміни значення певної вершини графа, що залежать від її батьківських вершин.

Для перевірки роботи цього методу було вирішено застосувати його на даних моніторингу якості атмосферного повітря мережі громадського моніторингу EcoCity з декількох станцій моніторингу у Вінницькій області за період 2022-2024 рр. [23]. Для тестування використовувався показник пилу PM10. Приклад таких даних наведено на рисунку 3.28.

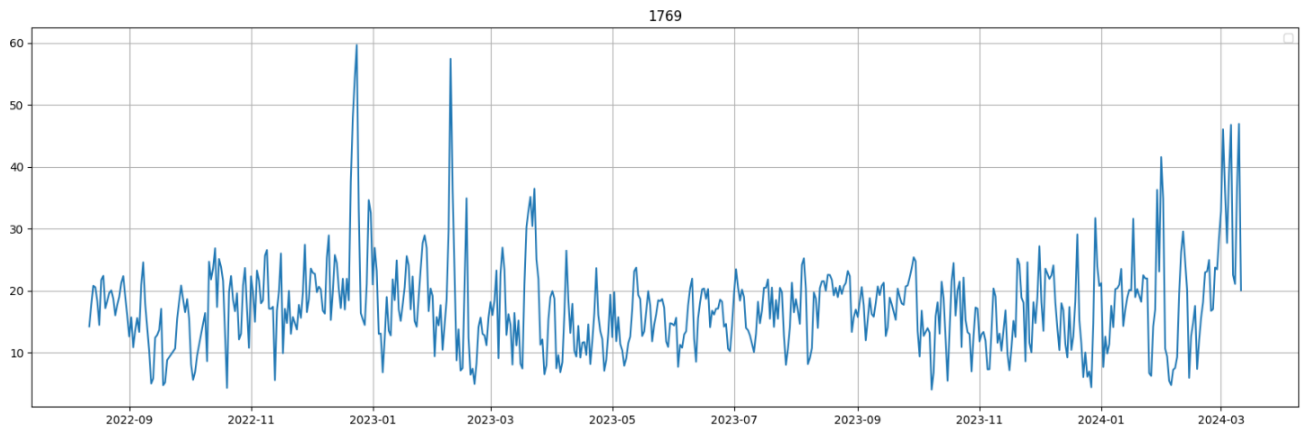


Рисунок 3.28 – Дані показника якості атмосферного повітря однієї з станцій моніторингу у Вінницькій області за 2022-2024 рр [83]

Як було описано раніше, для застосування методу необхідно зібрати дані з декількох станцій та створити один набір, де кожен рядок – це дата проведення вимірювання, а значення в колонках – значення показника PM10 для відповідної станції моніторингу. Загалом, було побудовано набір даних з семи станцій моніторингу (рис 3.29).

	st_650	st_1769	st_1811	st_1813	st_1825	st_1830	st_1834
0	9.4885	7.4381	11.8504	8.0747	12.7674	9.7900	9.9807
1	31.2040	8.3401	10.1477	8.9014	13.2750	22.9919	10.7449
2	11.8139	14.1653	13.9522	22.2361	18.0257	10.8561	19.7283
3	8.3002	15.6867	19.0788	27.1034	27.0721	18.1392	20.3049
4	14.4049	13.0627	23.9908	15.5545	22.7843	12.5923	15.0530
...
279	48.6737	22.5754	39.2001	30.8495	30.3635	29.2202	33.4889
280	17.6561	21.1076	47.1072	32.4196	39.6464	43.1531	32.4889
281	26.1567	36.6581	68.6969	68.4406	74.4583	38.1823	36.8276
282	19.1412	46.9558	56.4339	65.3547	51.9435	91.6873	52.6314
283	7.0339	20.1094	15.9684	16.2481	17.2579	12.3558	13.9910

Рисунок 3.29 – Набір даних для байєсівського аналізу [83]

Наступним кроком є перетворення отриманих даних у бінарний формат, використовуючи описане раніше правило, де кожне значення для відповідної станції порівнюється з попереднім спостереженням для визначення того чи рівень забруднення зростає, чи спадає (рис 3.30).

	st_650	st_1769	st_1811	st_1813	st_1825	st_1830	st_1834
0	0	0	0	0	0	0	0
1	1	1	0	1	1	1	1
2	0	1	1	1	1	0	1
3	0	1	1	1	1	1	1
4	1	0	1	0	0	0	0
...
279	1	0	0	0	0	0	0
280	0	0	1	1	1	1	0
281	1	1	1	1	1	0	1
282	0	1	0	0	0	1	1
283	0	0	0	0	0	0	0

Рисунок 3.30 – Дані після перетворення, згідно кроку 2 алгоритму [83]

Для отриманого набору даних будемо спрямований ациклічний граф (рис 3.31), використовуючи Python-бібліотеку `bnlearn`. Ця бібліотека дозволяє моделювати байєсівські мережі та використовувати широкий набір допоміжних методів для рисування графіків тощо.

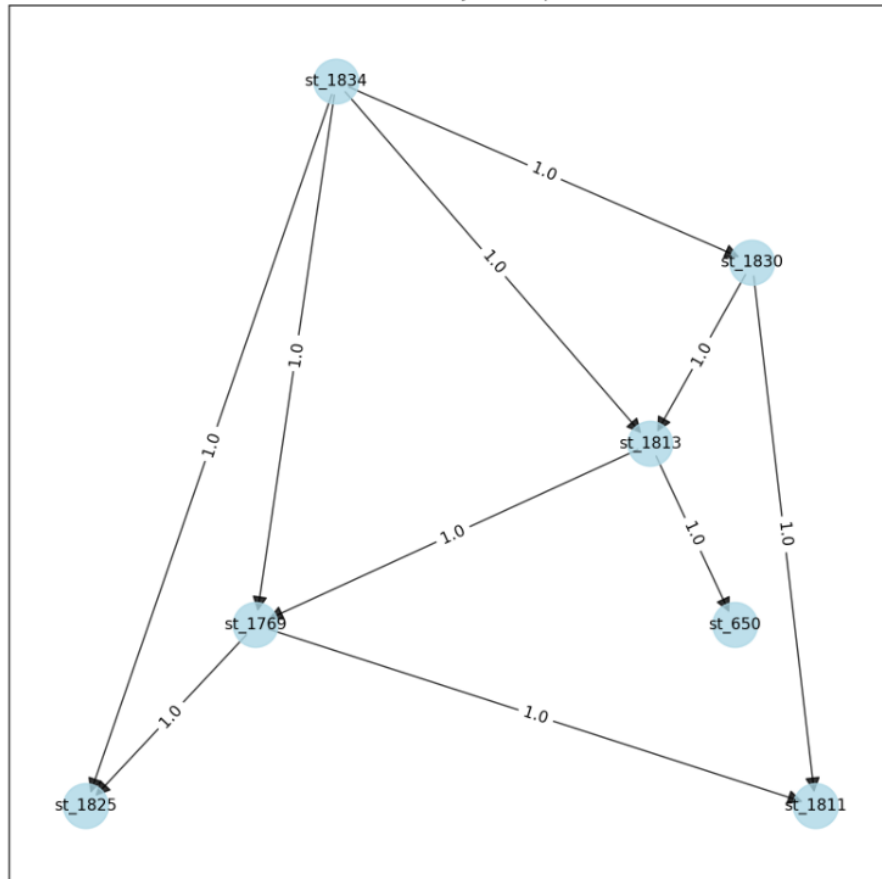


Рисунок 3.31 – Результат роботи бібліотеки bnlearn [83]

Як можна побачити, використовуючи бібліотеку для отриманих раніше даних, вдалось встановити зв'язки між різними станціями моніторингу та побудувати DAG. Використаємо його для оцінювання значення умовних розподілів ймовірностей (CPD). Для цього можна застосувати вже згадану бібліотеку bnlearn, використовуючи модуль «parameter_learning» Результат роботи даного модулю наведено на рисунках 3.32 та 3.33.

```

[bnlearn] >CPD of st_1769:
+-----+-----+-----+-----+
| st_1813 | st_1813(0) | ... | st_1813(1) |
+-----+-----+-----+-----+
| st_1834 | st_1834(0) | ... | st_1834(1) |
+-----+-----+-----+-----+
| st_1769(0) | 0.5813253012048193 | ... | 0.3994252873563218 |
+-----+-----+-----+-----+
| st_1769(1) | 0.4186746987951807 | ... | 0.6005747126436781 |
+-----+-----+-----+-----+
  
```

Рисунок 3.32 – Значення умовних розподілів ймовірностей для станції № 1769

```
[bnlearn] >CPD of st_1811:
```

st_1769	st_1769(0)	...	st_1769(1)	
st_1830	st_1830(0)	...	st_1830(1)	
st_1811(0)	0.5772870662460567	...	0.4057971014492754	
st_1811(1)	0.4227129337539432	...	0.5942028985507246	

Рисунок 3.33 – Значення умовних розподілів ймовірностей для станції № 1811

Як можна побачити на рисунку 3.29, у разі зростання забруднення на станціях № 1813 (с. Якушинці) та № 1834 (с. Стрижавка), імовірність зростання забруднення на станції № 1769 (м. Вінниця) зростає в 1.5 рази. Подібний результат можна побачити і на рисунку 3.30, а саме імовірність зростання забруднення на станції № 1811 (м. Хмільник) збільшується майже в 1.5 рази за умови зростання забруднення на станціях № 1769 (м. Вінниця) та № 1830 (с. Війтівці).

Було перевірено точність роботи даної моделі, зробивши прогноз та порівнявши результат з реальними даними. На рисунку 3.34 наведено результат даної перевірки, який свідчить про задовільну точність.

```
[bnlearn]> Remaining columns for inference: 6
100%|██████████| 30/30 [00:00<00:00, 430.56it/s]
Accuracy - 0.7719298245614035
```

Рисунок 3.34 – Точність роботи моделі [83]

В підрозділі проаналізовано ефективність застосування методу байєсівського моделювання для аналізу взаємовпливів між рядами даних показників стану природного середовища. Описаний алгоритм застосування методу байєсівського моделювання є корисним інструментом для аналізу та попередньої обробки даних, що дозволяє усувати пропуски в існуючих даних за

рахунок їх обчислення з сусідніх станцій моніторингу стану природного середовища. Його застосовуватися для системного аналізу природних середовищ та дозволяє розкрити види невизначеності U_{ST1} та U_{ST2} .

3.4. Розроблення алгоритму технології системного аналізу стану природного середовища

Узагальнюючи наведене в попередніх підрозділах та пунктах, пропонується технологія системного аналізу стану природного середовища з послідовним розкриттям усіх видів невизначеності [11]:

1. Вибір критеріїв стану природного середовища J : обрати критерій та його складові, за якими буде аналізуватися стан природного середовища, що розкриває невизначеність U_{T2} .

2. Збирання інформації щодо «DPSIR»: зробити запити для отримання інформації про місця розташування джерел забруднення (для розкриття невизначеності U_{I1}), про розташування джерел викидів чи скидань та фактичні обсяги і концентрації забруднювальних речовин у цих викидах (скиданнях) (U_{I3}), про найбільші обсяги гранично допустимих значень цих обсягів та концентрацій (U_{I2}); щодо надання рядів даних спостережень за певний період часу (U_{I4}); про вже прийняті управлінські рішення щодо удосконалення моніторингу, контролю та охорони стану довкілля, боротьби з рівнем забруднення (U_{I6}).

3. Аналіз даних спостережень про «S»: проаналізувати *чи наявних даних про стан природного середовища достатньо* для формування висновків та управлінських заходів, використовуючи прийоми розвідувального аналізу [16]. Якщо – так, тоді слід перейти до кроку 7, а якщо – ні, тоді – слід оптимізувати систему моніторингу.

4. Збирання інформації про мережу моніторингу та контролю «SIR». Слід зібрати інформацію від суб'єктів державної системи моніторингу

природного середовища про пости цієї мережі, їх показники, регулярність вимірювань та метрологічні аспекти (це розкриє U_{I5}) та зібрати додаткову інформацію про стан природного середовища від суб'єктів державної системи контролю, які здійснюють нерегулярні спостереження (держкоінспекції, органи МОЗ тощо), від мереж громадського моніторингу природних середовищ, непрямі дані про щільність населення, зокрема вразливого, про захворюваність людей в регіоні тощо.

5. Аналіз аномалій у «S»: відповідно до чинного законодавства (наприклад, для атмосферного повітря – за Постановою КМУ від 14.08.2019 р. № 827) або за результатами статистичного аналізу, визначити верхнє і нижнє порогові значення відповідних показників та кількість фактів виходу за межі цих порогових значень, які потім будуть застосовуватися для пошуку аномалій чи екстремальних значень показників стану природного середовища – це розкриває невизначеності U_{T3} та U_{T5} . А потім на основі аналізу кількості таких аномалій за цією ж Постановою визначається мінімальна кількість постів мережі моніторингу для кожної групи показників окремо, що дозволяє розкрити невизначеність U_{T1} в частині кількості постів.

6. Оптимізація місць розташування постів мережі моніторингу «S». Сформувати систему критеріїв для вибору параметрів місць постів спостережень стану природного середовища, з урахуванням екстремальних значень показників про усіх складові «DPSIR». Для їх систематизації та визначення ваги використати метод аналізу ієрархій, наприклад як це було зроблено у роботі для стану атмосферного повітря у Вінницькій області [7]. Це дозволить визначити де варто розмістити нові станції моніторингу чи перемістити існуючі, що розкриє невизначеність U_{T1} в частині місць розташування цих станцій чи постів.

7. Проаналізувати достовірність даних про «P». Якщо аналіз покаже певну невідповідність, наприклад, антропогенні впливи (наприклад, скидання забруднень у річку) не змінюються, а якість води біля місця впливу неухильно погіршується, тоді варто висунути та перевірити гіпотезу щодо недостовірності

або даних про стан природного середовища, або даних про антропогенні впливи. Якщо немає підстав вважати дані недостовірними, тоді – перейти до кроку 9.

8. Пошук можливих джерел забруднення. Для цього можна застосувати авторський метод на основі аналізу аномалій у наявних даних моніторингу стану природного середовища [6, 18]. Це дозволить розкрити невизначеність (U_{S1}). Цей метод дозволяє знайти незареєстровані джерела (чи виявити який обсяг було занижено існуючим джерелом), які можуть мати місце, і в просторі, і в часі одночасно, шляхом зіставлення результатів роботи методу з наявною інформацією про складові моделі «DPSIR» заданого регіону. А це, у свою чергу, дозволить ще зменшити рівень невизначеності для U_{I3} .

9. Чи є необхідним прогнозування щодо «S»? Якщо – ні, тоді – перейти до кроку 12.

10. Визначення параметрів аномалій у «S». Для побудови ефективної моделі прогнозування, необхідно знайти аномалії в наявних даних стану природного середовища. Це можна робити, використовуючи відомі методи пошуку аномалій, наприклад, описані у роботі [13]. Також, дані можуть містити локальні у часі аномалії, тому можна застосувати метод пошуку локальних аномалій на основі декомпозиції ряду з роботи [10], що, в свою чергу, дозволить розкрити невизначеність U_{S2} .

11. Побудова оптимальної моделі для прогнозування «S». Це передбачає вибір можливих інтелектуальних моделей, ідентифікацію усіх їх параметрів, з урахуванням аномалій та можливих видів сезонності, їх порівняльний аналіз та вибір оптимальної за певною метрикою моделі. Приклад такого вибору наведено у роботі [17]. Це розкриває невизначеність U_{M1} . При цьому, для ідентифікації параметрів сезонності можна використати як класичні підходи [14, 15], так і авторський метод ідентифікації параметрів сезонності та аномалій періодичного часового ряду, описаний у роботі [8], що дозволить знайти які сезонності мають місце для відповідного набору даних та яким чином ці сезонності можна поєднати при моделюванні даних. Усе це дозволить розкрити такі

невизначеності: U_{P1} , U_{P2} , U_{T6} .

12. Чи є необхідним аналіз взаємозв'язків? Якщо – ні, тоді – перейти до кроку 14.

13. Інтелектуальний аналіз взаємозв'язків. Застосувати метод аналізу взаємовпливу показників стану природного середовища на основі байєсівського моделювання, що дозволить, хоча б частково, розкрити невизначеності U_{ST1} та U_{ST2} . В результаті можна буде отримати інформацію про те, який взаємовплив існує між різними показниками на сусідніх станціях моніторингу, що, у свою чергу, дозволить компенсувати проблеми з даними на різних станціях (наприклад, наявність пропусків через відсутність в певні проміжки часу зв'язку або електроенергії). Приклад такого аналізу наведено у Python програмі-ноутбуці на базі платформи Kaggle [82].

14. Аналіз даних «S» [84], у т.ч. інтелектуальний, за різними сценаріями, що дозволить розкрити U_{SC1} .

15. Оптимізація складної системи «DPSIR». Виходячи з усіх отриманих результатів, провести оптимізацію різних складових моделі «DPSIR» та розробити рекомендації щодо можливих управлінських заходів системи і вибрати оптимальні серед них одним із відомих методів, що розкриє невизначеність U_{T4} [85].

Графічно алгоритм можна зобразити блок-схемою (рис. 3.35).

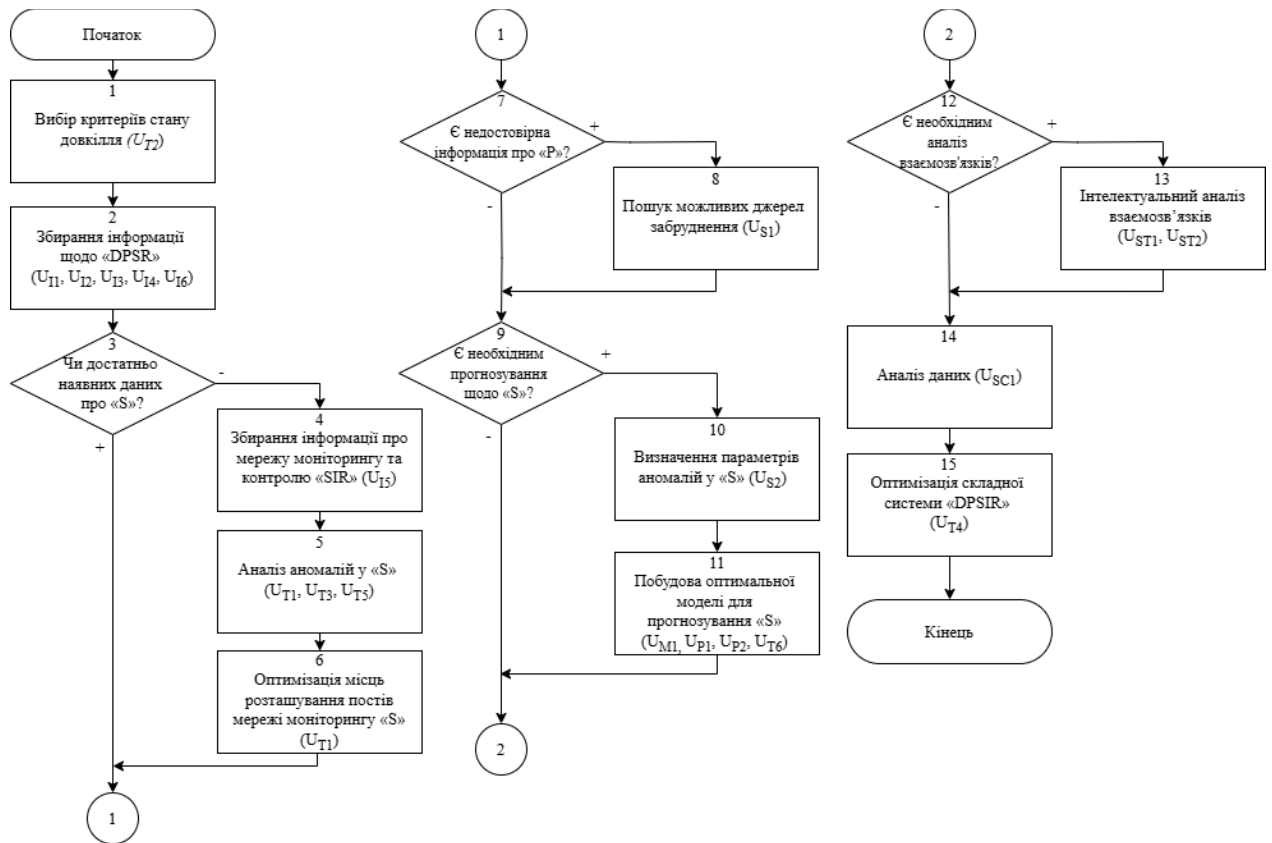


Рисунок 3.35 – Блок-схема алгоритму системного аналізу природного середовища з розкриттям невизначеності [11]

3.5. Висновки до розділу

Запропоновано метод ідентифікації параметрів сезонності та аномалій періодичного часового ряду, що базується на адаптивній декомпозиції та дозволяє комбінувати налаштування декількох варіантів сезонності при роботі з інтелектуальними моделями прогнозування стану атмосферного повітря. Наведено приклад застосування на реальних даних моніторингу стану атмосферного повітря від мережі громадського моніторингу EcoCity. Тестування показало, що використання цього методу дозволяє покращити точність оптимальної моделі на основі Prophet за метрикою R2 – у 1,7 рази, а за метрикою MSE – у 2 рази у порівнянні з моделлю Prophet з параметрами і видами сезонності за замовчуванням.

Удосконалено метод ідентифікації локальних аномалій значень показників стану природного середовища з використанням декомпозиції на півхвилі, який

базується на розбитті часового ряду на півхвилі та інтелектуального пошуку аномалій на кожному окремому відрізку ряду. Використання методу на даних моніторингу атмосферного повітря від мережі EcoCity продемонструвало збільшення точності роботи моделі Prophet за метриками RMSE та MAE на 11% та 8% відповідно у порівнянні з моделлю Prophet з параметрами налаштування за замовчуванням.

Описано алгоритм аналізу взаємозв'язків показників стану природного середовища, що базується на використанні байєсівського аналізу даних та дозволяє аналізувати взаємовпливи між рядами значень показників природного середовища.

Систематизовано усі розроблені методи розкриття видів невизначеності та запропоновано технологію системного аналізу стану природних середовищ з урахуванням аномалій. Сформовано алгоритм застосування даної технології, що складається з 15 кроків та дозволяє комплексно проаналізувати систему управління станом природного середовища.

Розкриття видів невизначеності, які були описані в підрозділі 2.2 наведено в додатку Г.

Описані в розділі результати були висвітлені в таких роботах автора: [8-11, 14, 15, 17, 18].

РОЗДІЛ 4

ПРИКЛАДНЕ ЗАСТОСУВАННЯ РОЗРОБЛЕНИХ МЕТОДІВ ТА ТЕХНОЛОГІЙ СИСТЕМНОГО АНАЛІЗУ ПРИРОДНИХ СЕРЕДОВИЩ З УРАХУВАННЯМ АНОМАЛІЙ

4.1. Застосування методів та технологій системного аналізу до системи управління стану річкових вод

Запропонована технологія системного аналізу стану природних середовищ з урахуванням аномалій, а саме кроки 9-11 алгоритму, була застосовано для аналізу якості поверхневих вод річки Південний Буг [6]. Вхідними даними для дослідження були відкриті дані Державного агентства водних ресурсів України, отримані на Єдиному державному веб-порталі відкритих даних [86], а також на веб-сайті «Моніторинг та екологічна оцінка водних ресурсів України» [87] (рис 4.1).

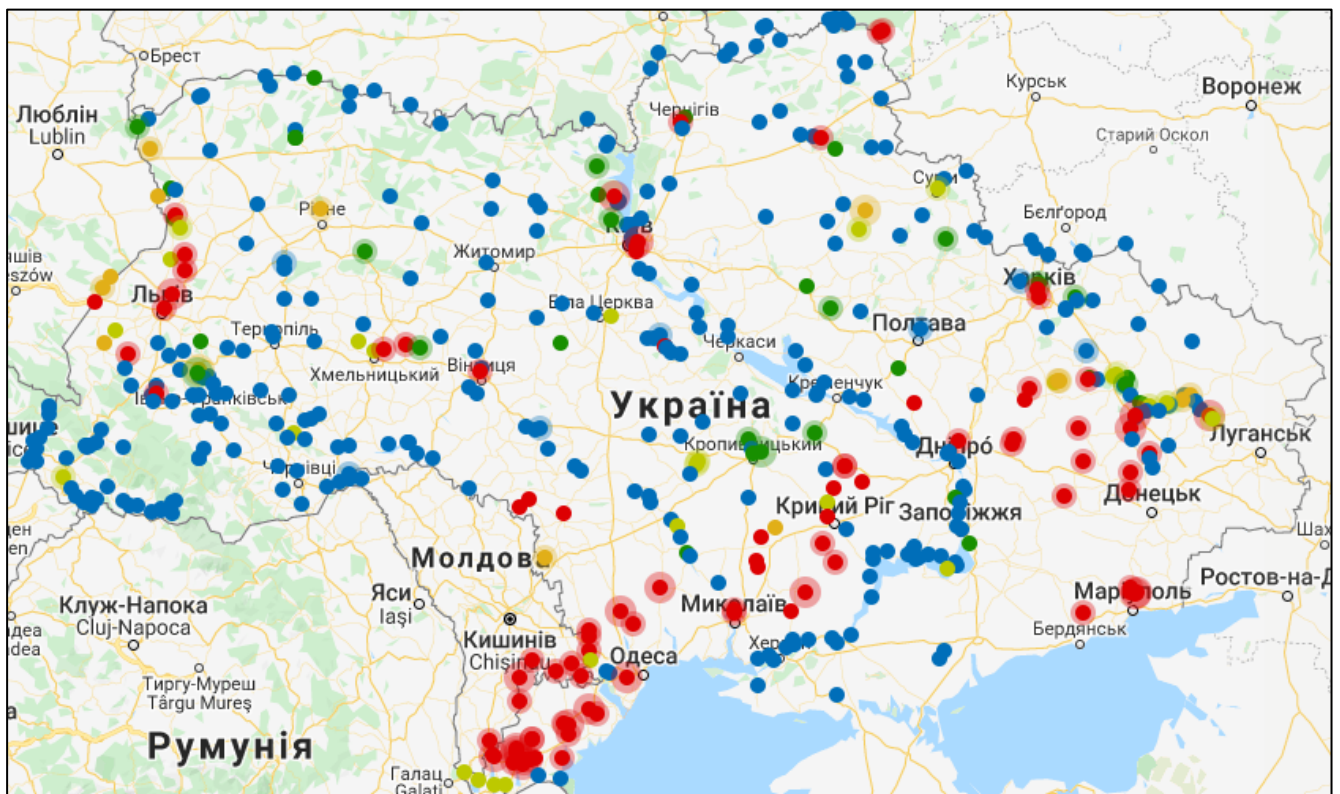


Рисунок 4.1 – Портал моніторингу водних ресурсів

Ці дані містили інформацію про стан середовища, починаючи з 2002 року, а саме значення наступних параметрів (рис 4.2):

- азот загальний, мг/дм³;
- біохімічне споживання кисню за 5 діб, мгО/дм³;
- завислі (суспендовані) речовини, мг/дм³;
- кисень розчинений, мгО₂/дм³;
- сульфат-іони, мг/дм³;
- хлорид-іони, мг/дм³;
- амоній-іони, мг/дм³;
- нітрат-іони, мг/дм³;
- нітрит-іони, мг/дм³;
- фосфат-іони (поліфосфати), мг/дм³.

<p>Повна назва ПС: р. Південний Буг, 570 км, смт Сабарів, (Сутиське водосховище), 500 м нижче скиду ВОКВП ВКП "Вінницяводоканал"</p> <p>Район річкового басейну: Південний Буг</p> <p>Назва головної лабораторії, що виконує ІЛВ: Лабораторія моніторингу вод та ґрунтів БУВР річок Причорномор'я та нижнього Дунаю</p> <p>Остання дата спостереження: 14.08.2024</p> <p>Дані моніторингу</p>			
Показник	Фактичне значення	ГДК (ОБУВ)	Перевищення нормативу, раз
Азот загальний, мг/дм ³			
Біохімічне споживання кисню за 5 діб, мгО/дм ³		3	
Завислі (суспендовані) речовини, мг/дм ³		15	
Кисень розчинений, мгО ₂ /дм ³	8,04	4	Немає
Сульфат-іони, мг/дм ³		100	
Хлорид-іони, мг/дм ³		300	
Амоній-іони, мг/дм ³		0,5	
Нітрат-іони, мг/дм ³		40	
Нітрит-іони, мг/дм ³		0,08	
Фосфат-іони (поліфосфати), мг/дм ³			

Рисунок 4.2 – Значення моніторингу якості води на порталі «Моніторинг та екологічна оцінка водних ресурсів України»

Застосування методів технології до 8 постів спостереження ділянки річки Південний Буг від витоків до м. Вінниці за даними системи моніторингу дозволило розкрити невизначеність U_{S1} , що пов'язана з відсутністю достовірної інформації про викиди забруднювальних речовин.

Результати використання кроків для однієї з станцій було продемонстровано в підрозділі 2.3, а саме на рисунках 2.9-2.12. Узагальнений результат застосування для усіх станцій моніторингу наведено на рисунку 4.3.

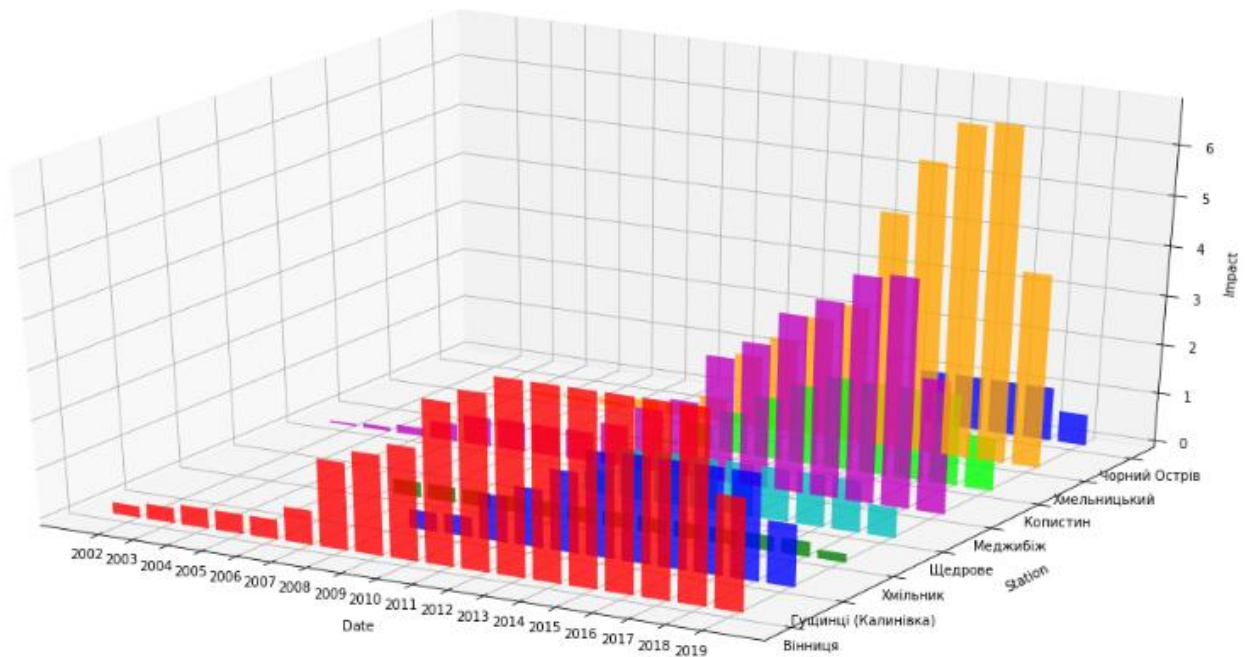


Рисунок 4.3 – Результат застосування методів технології для пошуку незареєстрованих джерел забруднення вод [6]

Як видно на рисунку 4.4, чітко ідентифікуються ділянки і роки, де мав місце наростаючий забруднюючий вплив на стан води у річці Південний Буг. Результати застосування алгоритму були розглянуті на засіданні Басейнової ради Південного Бугу (протокол № 4 від 22 травня 2020 р.). Розроблений метод та технологія була включена до переліку заходів щодо залучення потенціалу науки та освіти для вирішення проблем басейну річки Південний Буг. Членами ради було затверджено запропонований на засіданні перелік і рекомендовано комісії з освіти

та науки та іншим комісіям басейнової ради розвивати співпрацю у цьому напрямку.

4.2. Застосування методів та технологій системного аналізу до системи управління стану атмосферного повітря

Запропонована технологія системного аналізу активно застосовувалась в процесі роботи науково-дослідній лабораторії екологічних досліджень та екологічного моніторингу ВНТУ, а саме в рамках роботи щодо удосконалення системи управління станом атмосферного повітря зони «Вінницька», межі якої збігаються з межами Вінницької області [7, 12]. Під час проведення цієї роботи було проаналізовано показники, згідно вимог Постанови КМУ від 14.08.2019 р. № 827 (крок 1), для чого було надіслано запити до органів влади і місцевого самоврядування (крок 2). Було з'ясовано, що наявних даних недостатньо (крок 3). Зібрано інформацію про наявну мережу моніторингу (крок 4). Проаналізовано наявні дані та їх аномалії і визначено мінімальну кількість постів для різних груп показників (крок 5). Обґрунтовано систему критеріїв для вибору параметрів місць постів спостережень стану атмосферного повітря, з урахуванням екстремальних значень показників. Обраховані ваги усіх критеріїв наведено на рисунку 4.4.

Інтегральні критерії	Ваги	Критерії	Ваги	Критерії	Ваги	Критерії	Ваги	Критерії	Ваги	Критерії	Ваги
Ф1	0,466	Ф11	0,25	Ф21	0,116	Ф31	0,062	Ф41	0,026	Ф51	0,018
Ф2	0,27	Ф12	0,123	Ф22	0,074	Ф32	0,043	Ф42	0,018	Ф52	0,018
Ф3	0,136	Ф13	0,048	Ф23	0,049	Ф33	0,02	Ф43	0,014	Ф53	0,012
Ф4	0,069	Ф14	0,035	Ф24	0,025	Ф34	0,008	Ф44	0,007	Ф54	0,008
Ф5	0,058	Ф15	0,011	Ф25	0,006	Ф35	0,003	Ф45	0,004	Ф55	0,002

Рисунок 4.4 – Ваги обраних інтегральних критеріїв [12]

З використанням методу аналізу ієрархій визначено оптимальні місця для розміщення постів моніторингу [7] (крок 6). Перелік оптимальних місць та відповідні результати проведених обчислень наведено на рисунку 4.5.

Критерії та пункти	Ф1	Ф2	Ф3	Ф4	Ф5	J
Ваги критеріїв	0,5	0,27	0,14	0,069	0,058	1,0
Центр м. Ладижин	4,43	2,70	1,29	0,5	0,45	9,365
На в'їзді в м. Гайсин біля траси	4,24	2,70	0,75	0,6	0,55	8,821
смт. Іллінці	3,55	2,69	0,56	0,4	0,24	7,402
Центр м. Хмільник	0,95	0,27	1,15	0,4	0,42	3,152
м. Калинівка біля траси	0,78	1,36	0,25	0,4	0,37	3,131
с. Немія біля м. Могилів-Подільський	1,06	0,81	0,54	0,5	0,21	3,110
м. Могилів-Подільський	1,04	1,07	0,43	0,3	0,24	3,082
У центрі м. Немирів	0,96	0,53	0,69	0,3	0,30	2,728
м. Жмеринка, біля залізничного вокзалу	0,92	0,53	0,26	0,3	0,42	2,395
с. Іванів	0,91	0,32	0,16	0,2	0,22	1,790

Рисунок 4.5 – Перелік пріоритетних місць розташування станцій моніторингу [12]

Проведено аналіз усіх цих даних (крок 14) та розроблено управлінські заходи для оптимізації системи моніторингу (крок 15). На основі цього була розроблена і затверджена, відповідно до чинного законодавства, Програма державного моніторингу в галузі охорони атмосферного повітря зони «Вінницька» на 2023 – 2027 роки (підтверджується актом впровадження управління розвитку територій та інфраструктури Вінницької обласної військової адміністрації № 08-20/1134 від 14 квітня 2025 року, що наведено в додатку Д) [85].

В очікуванні оптимізації мережі державного моніторингу атмосферного повітря і нових рядів даних аналізувався стан з використанням даних мережі громадського моніторингу. Зокрема, у ВНТУ, як згадувалось в підрозділі 1.4, підписано договір про доступ до сервісу «Кабінет дослідника якості повітря України» з мережею EcoCity, що відкрило доступ до даних моніторингу станцій, розташованих у м. Вінниця та Вінницькій області [23, 42]. За цими даними побудовано інтелектуальні моделі та вибрано оптимальні за різними критеріями (кроки 9-11) (див. роботи [8, 10]), за цими моделями здійснено прогнозування та проведено інтелектуальний аналіз даних (крок 14). Ефективність розроблених інтелектуальних моделей для прогнозування часового ряду перевірялась на прикладі прогнозування показника PM_{2.5} (тверді частинки до 2.5 мкм) за даними

станцій №650 («Турбів») мережі громадського моніторингу атмосферного повітря EcoCity. Для розкриття параметричної та модельної невизначеності проводилось порівняння прогнозів на основі моделі Prophet з різними параметрами для опису сезонності та аномалій [88]:

1. Без використання вбудованих параметрів сезонності (англійською «Without default seasonality»);
2. Використання вбудованих параметрів сезонності (англійською «Default seasonality»);
3. Обчислення параметрів за авторським методом ідентифікації параметрів сезонності та аномалій періодичного часового ряду (англійською «Custom seasonality») (див. п.11 алгоритму);
4. Використання варіанту 2 та параметрів аномалій, знайдених класичним підходом щодо пошуку аномалій у п.10 алгоритму (англійською «Series anomalies + default seasonality»);
5. Використання налаштувань з варіанту 2 та аномалій, знайдених за авторським методом пошуку локальних аномалій на основі декомпозиції ряду теж у п.10 алгоритму (англійською «Fragment anomalies + default seasonality»);
6. Використання налаштувань з варіанту 3 та аномалій, знайдених класичним підходом пошуку аномалій (англійською «Series anomalies + custom seasonality»);
7. Використання налаштувань з варіанту 3 та аномалій, знайдених за авторським методом пошуку локальних аномалій на основі декомпозиції ряду (англійською «Fragment anomalies + custom seasonality»), тобто поєднання двох авторських методів з пп. 10 і 11 алгоритму.

Результат прогнозування на валідаційних даних наведено на рисунку 4.6.

	name	mae	rmse	mape
0	Without default seasonality	7.300028	8.724954	0.498087
1	Default seasonality	7.331930	8.725626	0.506034
2	Custom seasonality	5.444506	7.056464	0.452396
3	Series anomalies + default seasonality	7.178978	8.569961	0.498134
4	Fragment anomalies + default seasonality	7.043577	8.399435	0.487132
5	Series anomalies + custom seasonality	5.472565	7.071537	0.452011
6	Fragment anomalies + custom seasonality	5.152847	6.838839	0.435090

Рисунок 4.6 – Результати порівняння точності прогнозу з використанням розроблених інтелектуальних моделей та 3-х різних метрик (похибок) MAE (середня абсолютна похибка), RMSE (корінь з середньоквадратичної похибки) та MAPE (середня відносна похибка) [88]

Як можна побачити на рисунку 4.3, модель з використанням власних налаштувань сезонності (варіант 3) дозволила отримати кращу точність, у порівнянні з використанням вбудованих параметрів сезонності (варіант 2). Крім того, моделі з використанням методу пошуку локальних аномалій на основі декомпозиції ряду (варіанти 5 та 7) дозволили покращити точність за усіма метриками у порівнянні з класичним підходом до пошуку аномалій (варіанти 4 та 6). Найменшої ж похибки прогнозування було досягнуто моделлю з варіанту 7, що відповідає застосуванню обох авторських методів з п. 10 і 11 алгоритму. Порівнюючи результати, отримані з використанням цієї оптимальної моделі (варіант 7) з результатами, отриманими за рахунок використання параметрів за замовчуванням (варіант 2 – «Default seasonality»), вдалось зменшити похибки прогнозування за метрикою MAE на 30,2%, а за метрикою RMSE – на 21,7%.

4.3. Застосування методів та технології системного аналізу до моделювання поширення алергенних речовин

Запропоновану технологію системного аналізу стану природних середовищ, а саме кроки 12 і 13 її алгоритму, було використано під час проведення алергологічного дослідження за участі вчених з Вінницького національного медичного університету та Національного медичного університету, яке передбачало моделювання взаємозв'язків чутливості людей до різних груп грибкових алергенів [9]. Вхідними даними даного дослідження був набір даних, що містив детальні симптоми пацієнтів та результати молекулярних тестів населення більшості областей України за 300 видами алергенів (пилки рослин, грибів). Відсотковий розподіл чутливості пацієнтів до певних груп алергенів наведено на рисунку 4.7.

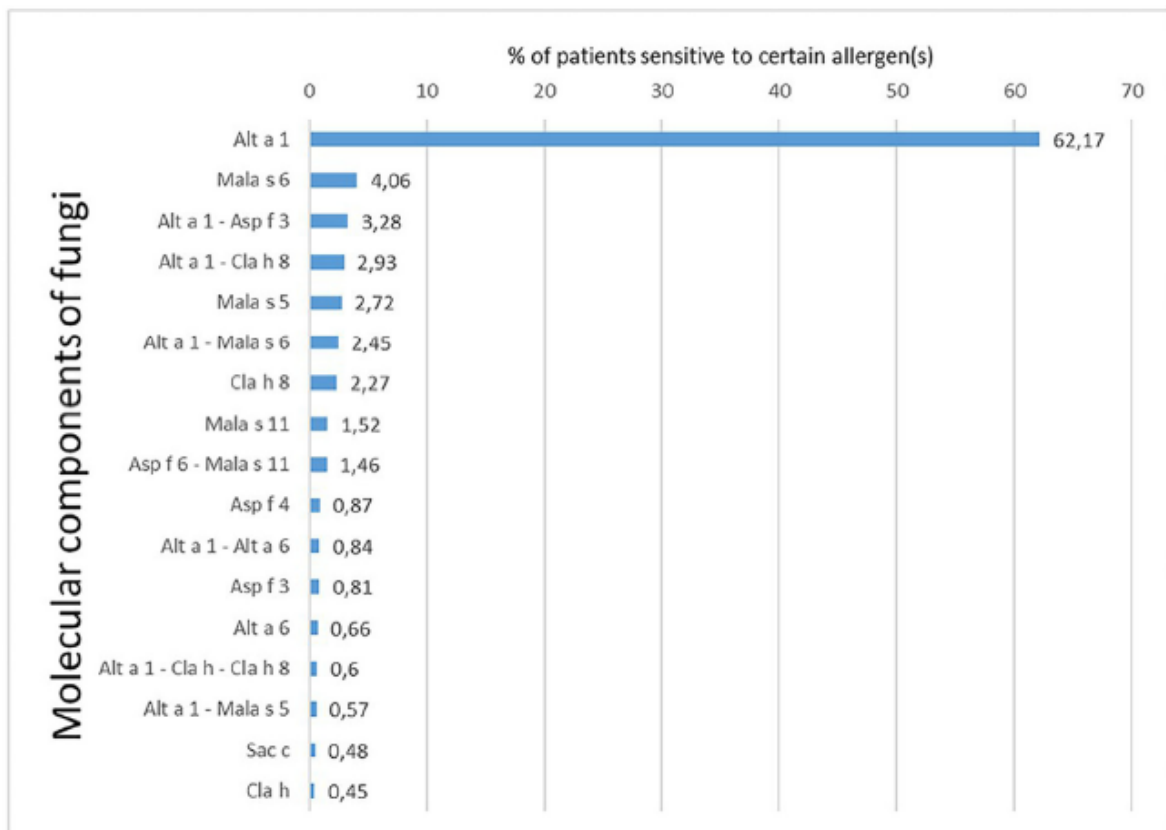


Рисунок 4.7 – Чутливість пацієнтів до алергенів [9]

Таким чином, байєсівське моделювання дозволяє усунути невизначеність на кроках 12 і 13 алгоритму запропонованої технології системного аналізу у сфері дослідження впливу алергенних рослин на здоров'я населення.

4.4. Впровадження в навчальний процес

Результати дослідження впроваджені у навчальний процес і наукову діяльність кафедри системного аналізу та інформаційних технологій Вінницького національного технічного університету для аспірантів, які навчаються за освітньою програмою «Системний аналіз» рівня «доктора філософії» спеціальності 124 для навчання знанням і навичкам застосування методів і моделей машинного навчання для аналізу та прогнозування часових рядів значень показників стану природних середовищ. Зокрема, результати дисертаційного дослідження застосовуються під час викладання дисципліни «Інформаційні технології моніторингу та аналізу стану складних систем». Результати використовуються як додатковий матеріал для навчання так і для поглиблення розуміння концепцій, пов'язаних із цією областю.

Зокрема, при викладанні використовуються такі результати досліджень, отримані у дисертаційній роботі здобувача (підтверджується актом впровадження ВНТУ від 24 березня 2025 року, що наведено в додатку Е):

- аналіз та прогнозування часових рядів стану природного середовища з урахуванням аномалій та пікових значень на основі моделі прогнозування часових рядів Prophet;
- використання різних методів пошуку аномалій (Isolation Forest, SESD) за рахунок використання відомих бібліотек на мові програмування Python (statsmodels, sklearn);
- ідентифікація параметрів аномалій та різних варіантів сезонності часового ряду на основі моделі Prophet з використанням як класичних підходів, так і авторських підходів.

Використання зазначених результатів дозволило підвищити якість навчального процесу.

4.5. Висновки до розділу

В розділі описано впровадження результатів, отриманих в рамках виконання дисертаційного дослідження.

Описано застосування кроків 9-11 алгоритму запропонованої технології системного аналізу стану природних середовищ для аналізу стану басейну річки Південний Буг за даними державної системи моніторингу якості вод за 2002-2019 роки. Отримані результати були розглянуті на засіданні Басейнової ради Південного Бугу та включені до переліку заходів щодо залучення потенціалу науки та освіти для вирішення проблем басейну річки Південний Буг, в тому числі вирішення проблем ідентифікації джерел незареєстрованих викидів вод, що є причиною погіршення якості води в географічних межах басейну річки Південний Буг.

Наведено застосування кроків алгоритму технології в рамках роботи щодо удосконалення системи управління станом атмосферного повітря зони «Вінницька», межі якої збігаються з межами Вінницької області. Результати даної роботи стали основою для розробки Програми державного моніторингу в галузі охорони атмосферного повітря зони «Вінницька» на 2023 – 2027 роки, яка затверджена, відповідно до чинного законодавства. Крім того, в очікуванні результатів оптимізації мережі моніторингу, певні кроки алгоритму технології системного аналізу використовувалися для аналізу та прогнозування стану атмосферного повітря у Вінницькій області за даними мережі громадського моніторингу «EcoCity».

Описано застосування технології, а саме кроків 12 і 13 її алгоритму під час проведення алергологічного дослідження, яке містило детальні симптоми пацієнтів та результати молекулярних тестів населення більшості областей України за 300 видами алергенів (пилки рослин, грибів). За участі вчених з

Вінницького національного медичного університету та Національного медичного університету, в рамках дослідження вдалось знайти взаємозв'язки між чутливістю до певних груп грибкових алергенів.

Охарактеризовано впровадження у навчальний процес і наукову діяльність кафедри системного аналізу та інформаційних технологій Вінницького національного технічного університету під час викладання для аспірантів кафедри, які навчаються за освітньою програмою «Системний аналіз».

Описані в розділі результати були висвітлені в таких роботах автора: [6-12, 17]

ВИСНОВКИ

У дисертаційній роботі розв'язано актуальну науково-прикладну задачу підвищення точності прогнозування стану природних середовищ з урахуванням аномалій шляхом створення методів та технології системного аналізу. Завдання, поставлені в ході дослідження, було виконано. Мету дисертаційного дослідження досягнуто за рахунок проведення наступних досліджень та розв'язанню наступних задач:

1. Проведено аналіз систем управління станом природних середовищ. Проаналізовано сучасний стан забруднення природних середовищ на території України, державні та громадські механізми контролю та управління їх станом. Аналіз показав, що на території України доволі успішно функціонують як державні, так і громадські механізми моніторингу та контролю стану природних середовищ. Рівні забруднення, хоча і мають деякі позитивні тенденції, все ж переважно показують нестабільну ситуацію та вимагають більш детального аналізу та контролю. Також проведено аналіз інтелектуальних моделей та методів прогнозування часових рядів даних стану природного середовища, а також засобів та способів автоматизації процесу пошуку аномалій в даних часових рядах, який показав, що існуючі методи мають низку недоліків. Має сенс розробляти нові методи або використовувати різні підходи комбінування існуючих. А головне – потрібно розробити універсальну технологію системного аналізу стану природних середовищ з урахуванням аномалій з усуненням різного роду видів невизначеності щодо цього стану.

2. Охарактеризовано систему управління станом природного середовища як складної системи, проведено її декомпозицію за різними критеріями та виділено види невизначеності, які мають місце. Запропоновано методи та підходи для розкриття цих видів невизначеності. В результаті з'ясувалось, що частина видів невизначеності може бути розкрита, використовуючи інформацію з відкритих джерел або через інформаційні запити до органів влади. Проте, переважна більшість видів невизначеності вимагають удосконалення існуючих

або розробки нових методів. Це дозволило формалізувати задачі, які треба розв'язати в рамках дослідження;

3. Розроблено методи розкриття видів невизначеності складної системи:

– удосконалено метод пошуку можливих джерел підвищеного антропогенного навантаження на стан природного середовища за даними мережі моніторингу стану цього природного середовища, що дозволяє, разом з поєднанням з даними аналізу супутникових знімків або даних аерофотозйомки, локалізувати в часі та просторі початок забруднюючого впливу на природне середовище за рахунок кусково-лінійної апроксимації тренду і в цілому – це дозволяє розкрити невизначеність, що пов'язана з фальсифікацією інформації про реальні викиди;

– запропоновано підхід формування ієрархічної системи критеріїв для оптимізації мережі спостережень стану природних середовищ, що базується на методі аналізу ієрархій та комплексному урахуванні критеріїв вибору місць для розташування станцій моніторингу, а також урахування аномальних значень або перевищення показників стану відповідного природного середовища, що дозволило розкрити невизначеність вибору кількості і місць розташування існуючих чи нових постів моніторингу.

– запропоновано метод ідентифікації параметрів сезонності та аномалій періодичного часового ряду на основі адаптивної декомпозиції кривої, що будується з усіх варіантів відношень амплітуд з періодом від 1 до половини кількості усіх значень ряду. Пошук локальних максимумів кривої дозволяє ідентифікувати різні варіанти сезонності, що мають місце в часовому ряді стану природного середовища – це дозволяє розкрити декілька видів невизначеності, пов'язаних з побудовою ефективної інтелектуальної моделі прогнозування стану природних середовищ;

– удосконалено метод ідентифікації локальних аномалій значень показників стану довкілля з використанням декомпозиції на півхвилі, який базується на розбитті часового ряду на півхвилі та інтелектуального пошуку аномалій на кожному окремому відрізку ряду, що дозволяє розкрити

невизначеність та, в цілому – підвищити точність прогнозування цього ряду у різні інтервали часу;

– удосконалено метод аналізу взаємозв'язків показників стану природного середовища, що дозволяє аналізувати взаємовпливи між рядами значень показників природних середовищ, отриманих з різних станцій моніторингу в межах певного регіону і в цілому дозволяє розкрити невизначеність про неповне знання про елементи системи або неточність уявлень про зв'язки між ними та невизначеність неповне знання про взаємозв'язок між різними показниками стану природного середовища на різних постах спостережень.

4. Систематизовано усі розроблені методи розкриття видів невизначеності та, базуючись на проведеному аналізі складної системи, запропоновано технологію системного аналізу природних середовищ з урахуванням аномалій. При розкритті невизначеності увага приділяється виявленню аномалій (чи екстремальних значень) показників стану природного середовища, оцінюванні їх параметрів та врахуванню впливу, а для розкриття невизначеності використовуються як класичні, так і авторські методи. Розроблено узагальнений алгоритм реалізації цієї технології з 15 етапів. Запропоновану технологію можна застосовувати для аналізу наявної системи управління, її оптимізації та побудови інтелектуальних моделей прогнозування її стану, що дозволяє комплексно аналізувати дану складну систему та підвищувати точність аналізу та прогнозування даних стану цього середовища.

5. Здійснено випробування розробленої технології та створених методів на реальних прикладах систем управління станом природних середовищ, перевірено її ефективність та здійснено впровадження:

– метод пошуку можливих джерел забруднення на основі аналізу аномалій випробувано під час досліджень членів басейнової ради Південного Бугу та її комісії щодо аналізу та прогнозування якості поверхневих вод річки Південний Буг та оприлюднено на засіданні цієї комісії (протокол № 4 від 22 травня 2020 р.) і рекомендовано для більш широкого впровадження;

– підхід формування ієрархічної системи критеріїв для вибору координат місць розташування постів спостережень застосовано для системного аналізу та оптимізації системи управління та моніторингу стану атмосферного повітря зони «Вінницька» (Вінницька область), на основі чого була розроблена Програма державного моніторингу в галузі охорони атмосферного повітря зони «Вінницька» на 2023 – 2027 роки та затверджена, відповідно до чинного законодавства (наказ Вінницької обласної військової адміністрації № 722 від 22 травня 2023 року);

– метод ідентифікації параметрів сезонності та аномалій періодичного часового ряду на основі його адаптивної декомпозиції перевірявся з використанням реальних даних показника пилу PM10 зі станції моніторингу якості атмосферного повітря, розташованої у Вінницькій області, від мережі громадського моніторингу EcoCity. В ході тестування, з'ясувалось, що використання запропонованого методу дозволило покращити точність оптимальної моделі на основі Prophet за метрикою R^2 – у 1,7 рази, а за метрикою MSE – у 2 рази у порівнянні з моделлю Prophet з параметрами і видами сезонності за замовчуванням;

– метод пошуку локальних аномалій значень часового ряду показника стану довкілля за рахунок декомпозиції цього ряду на умовні півхвилі застосувався для апроксимації часового ряду показник пилу PM2.5 за період 2019-2022 рр моделлю Prophet. За метриками RMSE та MAE у порівнянні з моделлю Prophet з параметрами налаштування за замовчуванням вдалось підвищити точність на 11% та 8% відповідно;

– метод пошуку взаємозв'язків показників стану природного середовища застосовувався під час проведення алергологічного дослідження даних, що містили детальні симптоми пацієнтів та результати молекулярних тестів майже 3500 пацієнтів Вінницької, Дніпропетровської, Закарпатської, Івано-Франківської, Київської, Луганської, Львівської, Миколаївської, Одеської, Полтавської, Рівненської, Сумської, Харківської, Херсонської, Хмельницької, Черкаської областей України за 300 видами алергенів рослинного походження, за

участі вчених з Вінницького національного медичного університету та Національного медичного університету, що дозволило знайти взаємозв'язки між чутливістю до певних груп грибкових алергенів. Це довело універсальність запропонованого методу у різних типах навколишнього середовища.

б. Результати дисертаційного дослідження були впровадженні під час розроблення Програми державного моніторингу в галузі охорони атмосферного повітря зони «Вінницька» на 2023 – 2027 роки, що підтверджується актом впровадження управління розвитку територій та інфраструктури Вінницької обласної військової адміністрації № 08-20/1134 від 14 квітня 2025 року (додаток Д) Також проведено впровадження у навчальний процес Вінницького національного технічного університету під час викладання дисциплін «Інформаційні технології моніторингу та аналізу стану складних систем» для аспірантів, які навчаються за освітньою програмою «Системний аналіз», що підтверджується актом впровадження ВНТУ від 24 березня 2025 року (додаток Е).

Завдяки укладеному договору з між ВНТУ та проєктом EcoCity (договір про приєднання до Програми підтримки дослідників якості повітря України та реєстрацію Кабінету дослідника якості повітря України від 24.03.2023 р), було створено та активно підтримується набір даних на платформі Kaggle, який, в тому числі, використовувався в ході цього дослідження.

Розроблено комплекс алгоритмів та програмно-інформаційних засобів, які реалізують основні елементи запропонованої технології системного аналізу стану природних середовищ з урахуванням аномалій. Більшість програм є публічно доступними у вигляді програм-ноутбуків у Google-платформі Kaggle експертів у сфері Data Science.

Результати дисертаційного дослідження опубліковані у 14 роботах, у т.ч. у 5 статтях у наукових фахових періодичних виданнях, 1 стаття в міжнародному виданні, що належить до списку видань першого квартилю реєстрів Scopus і Web of Science Core Collection, та 8 тезах доповідей на науково-практичних конференціях.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] Згуровський М. З., Панкратова Н. Д. Основи системного аналізу. К. : Видавнича група ВНУ, 2007. 546 с.
- [2] Bolton RJ, Hand DJ. Unsupervised Profiling Methods for Fraud Detection. *Statistical Science*. 2002; 17(3) :235–255. Режим доступу: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.24.5743&rep=rep1&type=pdf>
- [3] Goldstein, M., & Uchida, S. (2016). A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data. *PloS one*, 11(4), e0152173. Режим доступу: <https://doi.org/10.1371/journal.pone.0152173>
- [4] Thudumu, S., Branch, P., Jin, J. et al. A comprehensive survey of anomaly detection techniques for high dimensional big data. *J Big Data* 7, 42 (2020). Режим доступу: <https://doi.org/10.1186/s40537-020-00320-x>
- [5] Aggarwal C.C. (2017) An Introduction to Outlier Analysis. In: *Outlier Analysis*. Springer, Cham. Режим доступу: https://doi.org/10.1007/978-3-319-47578-3_1
- [6] В. Б. Мокін, О. В. Слободянюк, О. М. Давидюк, і Д. О. Шмундяк, «Інформаційна технологія пошуку можливих джерел підвищеного забруднення річки з використанням моделі Prophet», *Вісник Вінницького політехнічного інституту*, № 4, с. 15–24, (Верес. 2020). DOI: <https://doi.org/10.31649/1997-9266-2020-151-4-15-24>.
- [7] В. Б. Мокін, Є. М. Крижановський, А. Р. Яцолт, і Д. О. Шмундяк, «Технологія проектування мережі спостережень якості атмосферного повітря регіону на основі методу аналізу ієрархій», *Наукові Праці ВНТУ*, № 4, (Лют. 2022). DOI: <https://doi.org/10.31649/2307-5376-2021-4-21-33>.
- [8] Д. О. Шмундяк і В. Б. Мокін, «Метод ідентифікації параметрів гармонік та аномалій періодичного часового ряду на основі адаптивної декомпозиції», *Вісник Вінницького політехнічного інституту*, № 6, с. 46–56, (Груд. 2023). DOI: <https://doi.org/10.31649/1997-9266-2023-171-6-46-56>.

[9] Victoria Rodinkova, Serhii Yuriev, Vitalii Mokin, Mariia Kryvopustova, Dmytro Shmundiak, Mykyta Bortnyk, Yevhenii Kryzhanovskyi, Andrii Kurchenko, «Bayesian analysis suggests independent development of sensitization to different fungal allergens», World Allergy Organization Journal, Volume 17, (Issue 5, 2024) . DOI: <https://doi.org/10.1016/j.waojou.2024.100908>.

[10] Д. О. Шмундяк і В. Є. Копняк, «Метод ідентифікації локальних аномалій значень показників стану довкілля з використанням декомпозиції на півхвилі», Вісник Вінницького політехнічного інституту, № 1, с. 88–100, (Лют. 2024). DOI: <https://doi.org/10.31649/1997-9266-2024-172-1-88-100>.

[11] Д. О. Шмундяк, В. Б. Мокін, «Системний аналіз стану природних середовищ з урахуванням аномалій», Наукові Праці ВНТУ, № 4, (Грудень 2024). DOI: <https://doi.org/10.31649/2307-5376-2024-4-63-73>

[12] В. Б. Мокін, Є. М. Крижановський, А. Р. Ящолт, Д. О. Шмундяк, Б. В. Паєвський, «Оптимізація моніторингової мережі стану атмосферного повітря вінницької зони за даними 2016-2020 рр.», матеріали ІІ Науково-технічної конференції факультету інтелектуальних інформаційних технологій та автоматизації, Вінниця, 31 травня - 1 червня 2022 р. – Електрон. текст. дані. – 2022. – Режим доступу: <https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2022/paper/view/16118/13531>

[13] Д. О. Шмундяк, Н. С. Іжаковська, Д. О. Литвиненко, А. О. Судець, «Аналіз можливостей Python-бібліотек щодо виявлення аномальних даних у задачі прогнозування стану атмосферного повітря», матеріали ІІ Науково-технічної конференції факультету інтелектуальних інформаційних технологій та автоматизації, 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/18957/15722>

[14] Д. О. Шмундяк, «Дослідження впливу параметрів моделі Prophet на точність прогнозування якості атмосферного повітря», матеріали ІІ Науково-технічної конференції факультету інтелектуальних інформаційних технологій та автоматизації, 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/18210/15056>

[15] Д. О. Шмундяк, А. В. Лосенко, В. Б. Мокін, «Огляд підходів до визначення порядку Фур'є у моделі Facebook Prophet для моделювання сезонної складової часового ряду», матеріали LII Науково-технічної конференції факультету інтелектуальних інформаційних технологій та автоматизації, 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/17200/14329>

[16] В. Б. Мокін, Д. О. Шмундяк, В. Є. Копняк, «Порівняльний розвідувальний аналіз даних про стан атмосферного повітря за даними мережі EcoCity з використанням бібліотеки Sweetviz», матеріали LIII Науково-технічної конференції факультету інтелектуальних інформаційних технологій та автоматизації, Вінниця, 20 - 22 березня 2024 р. – Електрон. текст. дані. – 2024. –

Режим доступу: <https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2024/paper/view/20846/17251>

[17] Порівняльний аналіз моделей прогнозування показника стану атмосферного повітря / Д. О. Шмундяк // Матеріали XVII Міжнародної конференції «Контроль і управління в складних системах» (КУСС-2024) [Електронний ресурс], Вінниця, 16-17 жовтня 2024 року] – Режим доступу: <https://conferences.vntu.edu.ua/index.php/mccs/mccs2024/paper/view/22120>

[18] Метод пошуку можливих джерел забруднення атмосферного повітря на основі аналізу аномалій / Д. О. Шмундяк // Матеріали XXIII Міжнародної науково-практичної конференції Математичне моделювання та інформаційно-комунікаційні технології для перемоги та відновлення, Вінниця, 12-13 листопада 2024 р. – Режим доступу: https://itgip.org/wp-content/uploads/2024/11/2024-11-24_zbirka_all_07_11_2024_148x210.pdf

[19] Аналіз взаємовпливів рядів даних моніторингу стану атмосферного повітря з використанням методу байєсівського моделювання / Д. О. Шмундяк // Матеріали LIV Всеукраїнської науково-технічної конференції підрозділів

Вінницького національного технічного університету, Вінниця, 24 - 27 березня 2025 р. – Електрон. текст. дані. – 2025. – Режим доступу: <https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2025/paper/view/23870/19675>.

[20] Шмундяк Д.О., Мокін В.Б. Авторське свідоцтво на комп'ютерну програму «Python-реалізація методу ідентифікації параметрів гармонік та аномалій інтелектуальної моделі часового ряду на основі Prophet», № 134476 від 17.03.2025 р

[21] Environmental health. [Електронний ресурс]. Режим доступу: <https://www.who.int/health-topics/environmental-health>.

[22] Моніторинг довкілля: підручник / [Боголюбов В.М., Клименко М.О., Мокін В. Б. та ін.]; за ред. проф. В.М. Боголюбова. Вид. 2-ге, переробл. і доповн. – Київ: НУБіПУ, 2018. – 435 с.

[23] Vitalii Mokin, Dima Shmundiak, Volodymyr Kopniak “Air Quality Monitoring from EcoCity” Kaggle Dataset, 2021. [Електронний ресурс]. Режим доступу: <https://www.kaggle.com/datasets/vbmokin/air-quality-monitoring-from-ecocity>.

[24] Impacts of air pollution on ecosystems in Europe. [Електронний ресурс] – Режим доступу: <https://www.eea.europa.eu/en/analysis/publications/impacts-of-air-pollution-on-ecosystems-in-europe>.

[25] Сапко О.Ю., Кур'янова С.О. Вплив антропогенних джерел забруднення на якість річки Південний Буг. Екологічні науки : науково-практичний журнал / Головний редактор Бондар О.І. – К. : Видавничий дім «Гельветика», 2023. – № 2(47). – 250 с. DOI <https://doi.org/10.32846/2306-9716/2023.eco.2-47.10>

[26] Air Quality Index (AQI) Basics. – Режим доступу: <https://www.airnow.gov/aqi/aqi-basics/>

[27] European Air Quality Index». [Електронний ресурс] – Режим доступу: <https://airindex.eea.europa.eu/AQI/index.html>

[28] «EcoCity». [Електронний ресурс] – Режим доступу: <https://ecocity.org.ua/>.

[29] Javid, Amir & Yaghmaeian, Kamyar & Abbasi, Ebrahim & Roodbari, Ali. (2014). An evaluation of water quality from Mojen River, by NSFWQI index. Journal of Ecological Engineering. 15. 1-6. 10.12911/22998993.1125451.

[30] Canadian Water Quality Guidelines for the Protection of Aquatic Life. [Електронний ресурс] – Режим доступу: <https://ccme.ca/en/res/wqimanualen.pdf>.

[31] Постанова Кабінету Міністрів України від 14.09.2019 № 827 «Деякі питання здійснення державного моніторингу у галузі охорони атмосферного повітря», якою затверджено Порядок здійснення державного моніторингу в галузі охорони атмосферного повітря [<https://zakon.rada.gov.ua/laws/show/827-2019-%D0%BF#Text>];

[32] Наказ Міністерства захисту довкілля та природних ресурсів України від 25.02.2021 № 147 «Про затвердження форми Програми державного моніторингу у галузі охорони атмосферного повітря», зареєстрований у Міністерстві юстиції України 21.04.2021 за № 543/36165 [<https://zakon.rada.gov.ua/laws/show/z0543-21#n14>];

[33] Наказ Міністерства внутрішніх справ України від 21.04.2021 № 300 «Про затвердження Порядку розміщення пунктів спостережень за забрудненням атмосферного повітря в зонах та агломераціях», зареєстрований у Міністерстві юстиції України 13.05.2021 за № 635/36257 [<https://zakon.rada.gov.ua/laws/show/z0635-21#Text>];

[34] Закон України про навколишнє середовище [<https://zakon.rada.gov.ua/laws/show/1264-12#Text>]

[35] Екологічний моніторинг довкілля. [Електронний ресурс] – Режим доступу: <https://mepr.gov.ua/diyalnist/napryamky/ekologichnyj-monitoryng/ekologichnyj-monitoryng-dovkillya/>.

[36] Національна доповідь про стан навколишнього природного середовища в Україні у 2021 році. [Електронний ресурс] – Режим доступу: <https://mepr.gov.ua/wp-content/uploads/2023/01/Natsdopovid-2021-n.pdf>.

[37] Екологічний паспорт Вінницької області за 2024 рік. [Електронний ресурс] – Режим доступу: https://www.vin.gov.ua/images/UPRTER/2024/ogoloshennya/Ekologichnuu%20pasport%202024_%20.pdf.

[38] Угода про асоціацію з ЄС. [Електронний ресурс] – Режим доступу: https://zakon.rada.gov.ua/laws/show/984_011#Text.

[39] Моніторинг та екологічна оцінка водних ресурсів України [Електронний ресурс] – Режим доступу до ресурсу: <https://www.davr.gov.ua/>

[40] Постанова про затвердження порядку здійснення державного моніторингу вод [Електронний ресурс] – Режим доступу до ресурсу: <https://zakon.rada.gov.ua/laws/show/758-2018-п>

[41] «Чисте повітря для України». [Електронний ресурс] – Режим доступу: <https://cleanair.org.ua/>.

[42] Кабінет дослідника «EcoCity». [Електронний ресурс] – Режим доступу: <https://archive.eco-city.org.ua/>

[43] «SaveEcoBot». [Електронний ресурс] – Режим доступу: <https://www.saveecobot.com/>.

[44] Юрій Мацелюх, Василь Литвин (2024) «Моделювання систем аналізу пасажиропотоків низько-вуглецевих перевезень у розумному місті», Вісник Національного університету «Львівська політехніка» «Інформаційні системи та мережі», Випуск 15, 2024, с. 430 - 448. doi: 10.23939/sisn2024.15.430.

[45] ISO 14000 family. [Електронний ресурс] – Режим доступу: <https://www.iso.org/standards/popular/iso-14000-family>.

[46] Eco-Management and Audit Scheme (EMAS). [Електронний ресурс] – Режим доступу: https://green-business.ec.europa.eu/emas_en.

[47] Environmental health indicators: framework and methodologies. [Електронний ресурс] – Режим доступу: https://iris.who.int/bitstream/handle/10665/66016/WHO_SDE_OEH_99.10.pdf.

[48] Typology of indicators and the DPSIR framework. [Електронний ресурс] – Режим доступу: <https://www.eea.europa.eu/publications/TEC25>.

[49] 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.

[50] tsfresh [Електронний ресурс]. Режим доступу: <https://tsfresh.readthedocs.io/en/latest/>

[51] Shumway, R.H., Stoffer, D.S. (2017). ARIMA Models. In: Time Series Analysis and Its Applications. Springer Texts in Statistics. Springer, Cham. https://doi.org/10.1007/978-3-319-52452-8_3

[52] Taylor SJ, Letham B. 2017. Forecasting at scale. *PeerJ Preprints* 5:e3190v2 <https://doi.org/10.7287/peerj.preprints.3190v2>

[53] Staudemeyer, Ralf C., and Eric Rothstein Morris. "Understanding LSTM--a tutorial into long short-term memory recurrent neural networks." arXiv preprint arXiv:1909.09586 (2019). <https://doi.org/10.48550/arXiv.1909.09586>

[54] Vladov, S.; Vysotska, V.; Sokurenko, V.; Muzychuk, O.; Nazarkevych, M.; Lytvyn, V. Neural Network System for Predicting Anomalous Data in Applied Sensor Systems. *Appl. Syst. Innov.* 2024, 7, 88. <https://doi.org/10.3390/asi7050088>

[55] Omar, Salima & Ngadi, Md & Jebur, Hamid & Benqdara, Salima. (2013). Machine Learning Techniques for Anomaly Detection: An Overview. *International Journal of Computer Applications.* 79. 10.5120/13715-1478.

[56] Julien Lesouple, Cédric Baudoin, Marc Spigai, Jean-Yves Tourneret, Generalized isolation forest for anomaly detection, *Pattern Recognition Letters*, Volume 149, 2021, Pages 109-119, ISSN 0167-8655, <https://doi.org/10.1016/j.patrec.2021.05.022>.

[57] Sklearn. API Reference – Available: <https://scikit-learn.org/stable/modules/classes.html>. Accessed: 07.12.2023

[58] Yumin Chen, Duoqian Miao, Hongyun Zhang, Neighborhood outlier detection, *Expert Systems with Applications*, Volume 37, Issue 12, 2010, Pages 8745-8749, ISSN 0957-4174, <https://doi.org/10.1016/j.eswa.2010.06.040>.

[59] Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, and Jörg Sander. 2000. LOF: identifying density-based local outliers. SIGMOD Rec. 29, 2 (June 2000), 93–104. <https://doi.org/10.1145/335191.335388>

[60] VIEIRA, Rafael G.; LEONE FILHO, Marcos A.; SEMOLINI, Robinson. An Enhanced Seasonal-Hybrid ESD Technique for Robust Anomaly Detection on Time Series. In: SIMPÓSIO BRASILEIRO DE REDES DE COMPUTADORES E SISTEMAS DISTRIBUÍDOS (SBRC), 36. , 2018, Campos do Jordão. Anais. Porto Alegre: Sociedade Brasileira de Computação, 2018 . p. 281-294. ISSN 2177-9384. DOI: <https://doi.org/10.5753/sbrc.2018.2422>.

[61] Statsmodels. [Електронний ресурс] – Режим доступу: <https://statsmodels.org/stable/index.html>

[62] Методологія та організація наукових досліджень: підручник – вид. 3-є, змін. та доп. [Електронний ресурс] / Б.І.Мокін, О.Б.Мокін, В.Б. Мокін. – Вінниця: ВНТУ, 2023. – (PDF, 230 с.)

[63] Програма державного моніторингу у галузі охорони атмосферного повітря зони «Вінницька» на 2023-2027 роки [Електронний ресурс] – Режим доступу: https://sait.vntu.edu.ua/wp-content/uploads/2023/02/V_atmo_m.pdf

[64] Директива 2008/50/ЄС Європейського Парламенту та Ради від 21 травня 2008 року про якість атмосферного повітря та чистіше повітря для Європи. [https://zakon.rada.gov.ua/laws/show/994_950#Text]

[65] В. Б. Мокін, Л. М. Скорина, та А. Р. Ящолт, Удосконалення технології аналізу даних дозвільної документації зі спеціального водокористування в системі Держводагенства, 4-те вид. Вінниця, Україна: Вісник Вінницького політехнічного інституту, 2017, 22-31 с.

[66] В. Б. Мокін, А. Р. Ящолт, Комп'ютеризовані регіональні системи державного моніторингу поверхневих вод: моделі, алгоритми, програми. Монографія. Вінниця, Україна: ВНТУ, 2005, 78-85 с.

[67] Мокін В. Б. Інформаційна технологія пошуку незареєстрованих місць надходження забруднених вод у річку за супутниковими і векторними даними геопорталів / В. Б. Мокін, Л. М. Скорина, Є. М. Крижановський, М. А. Гораш //

Наукові праці Донецького національного технічного університету. Серія: «Інформатика, кібернетика та обчислювальна техніка». – Покровськ, 2018. – № 2 (27). – С. 30-35.

[68] V. Hnatushenko, T. Bulana, B. Molodets and D. Boldyriev, "Development of UAV Image Processing Algorithms for Early Detection of Fires in Natural Ecosystems," 2024 IEEE 7th International Conference on Actual Problems of Unmanned Aerial Vehicles Development (APUAVD), Kyiv, Ukraine, 2024, pp. 219-223, doi: 10.1109/APUAVD64488.2024.10765845.

[69] Wójcik, W., Kalizhanova, A., Kulyk, Y. A., Knysh, B. P., Kvyetnyy, R. N., Kulyk, A. I., Yussupova, G. (2022). The Method of Time Distribution for Environment Monitoring Using Unmanned Aerial Vehicles According to an Inverse Priority. *Journal of Ecological Engineering*, 23(11), 179-187. <https://doi.org/10.12911/22998993/153458>

[70] Людмила Левенчук, Оксана Тимощук, Віра Гуськова, Петро Бідюк (2023) «Невизначеності в обробленні даних, прогнозування і прийняття рішень», Системні дослідження та інформаційні технології, № 3 (2023), с. 66-80. doi: 10.20535/SRIT.2308-8893.2023.3.05.

[71] Ломазов П. К., Павличенко А. В., Бучавий Ю. В. Удосконалення методологічних підходів до розвитку системи спостереження за забрудненням атмосферного повітря в агломераціях. Збірник наукових праць Національного гірничого університету. 2023. № 73. С. 240-252. <https://doi.org/10.33271/crpnmu/73.240>

[72] Saaty T. L. *Multicriteria Decision Making: The Analytic Hierarchy Process*. – Pittsburgh: RWS Publ., 1996. – 287 p.

[73] Mokin B. I. Decision Support System For The Use Of Funds Received From Higher Education Institution Paid Services / Mokin, Borys I; Mokin, Vitalii B; Mokina, Yuliya V. *Actual Problems in Economics*, # 3(177), 2016, Pages 372-383.

[74] Василь Литвин, Олексій Хобор (2024) «Розшифровка Цезію-137: підхід глибинного навчання до екологічного прогнозування», Вісник Національного університету «Львівська політехніка» «Інформаційні системи та мережі», Випуск 15, 2024, с. 200 - 209. doi: 10.23939/sisn2024.15.200.

[75] А. В. Лосенко, «Інформаційна технологія прогнозування часового ряду кількості хворих на коронавірус на основі моделі Facebook Prophet», Вісник Вінницького політехнічного інституту, вип. 5, с. 50–59, Жовт. 2023.

[76] Vitalii Mokin, and Arsen Losenko, “COVID-19 Ukraine daily cases - EDA” Kaggle Notebook. [Електронний ресурс]. Режим доступу: <https://www.kaggle.com/code/vbmokin/covid-19-ukraine-daily-cases-eda>.

[77] Dmytro Shmundiak and Vitalii Mokin, “Adaptive decomposition for harmonics and anomalies” Kaggle Notebook. [Electronic resource]. Available: <https://www.kaggle.com/code/dimashmundiak/adaptive-decomposition-for-harmonics-and-anomalies>. Accessed:20.12.2023.

[78] Woolson, R.F. (2008). Wilcoxon Signed-Rank Test. In Wiley Encyclopedia of Clinical Trials (eds R.B. D'Agostino, L. Sullivan and J. Massaro). <https://doi.org/10.1002/9780471462422.eoct979>

[79] McKnight, P.E. and Najab, J. (2010). Mann-Whitney U Test. In The Corsini Encyclopedia of Psychology (eds I.B. Weiner and W.E. Craighead). <https://doi.org/10.1002/9780470479216.corpsy0524>

[80] Адаменко Микола, Кацман Михайло, Білецька Євгенія. Аналіз існуючих математичних моделей і комп'ютерних програм для прогнозування розповсюдження забруднюючих речовин в атмосфері. Системи обробки інформації. 2018. № 1(152). С. 155-162. <https://doi.org/10.30748/soi.2018.152.22>.

[81] Nagarajan R, Scutari M, Lèbre S. Bayesian network inference algorithms. In: Bayesian networks in R. New York: Springer, 2013. doi: 10.1007/978-1-4614-6446-4

[82] Трофимчук, О. М., Бідюк, П. І., Просянкіна-Жарова, Т. І. і Терентьєв, О. М. (2023) «Байєсівський аналіз даних у моделюванні та прогнозуванні нелінійних нестационарних фінансово-економічних процесів», Міжнародний науково-технічний журнал "Проблеми керування та інформатики", 68(4), с. 71–83. doi: 10.34229/1028-0979-2023-4-6.

[83] Dmytro Shmundiak, “Bayes for air monitoring data” Kaggle Notebook. [Електронний ресурс]. Режим доступу: <https://www.kaggle.com/code/dimashmundiak/bayes-for-air-monitoring-data>.

[84] Vitalii Mokin, Dima Shmundiak, “WQ SB river : Analysis & Forecasting” Kaggle Notebook, 2022. [Електронний ресурс]. Режим доступу: <https://www.kaggle.com/code/vbmokin/wq-sb-river-analysis-forecasting>.

[85] Про затвердження Програми державного моніторингу в галузі охорони атмосферного повітря на 2023–2027 роки для зони «Вінницька», 2021. [Електронний ресурс] – Режим доступу: <https://mepr.gov.ua/wp-content/uploads/2023/01/Natsdopovid-2021-n.pdf>

[86] Єдиний державний веб-портал відкритих даних. Дані державного моніторингу поверхневих вод // Державне агентство водних ресурсів України. – Режим доступу: <https://data.gov.ua/dataset/ee2bc3b0-42d4-4f19-8d96-913cd9d1f02a>

[87] Моніторинг та екологічна оцінка водних ресурсів України // Державне агентство водних ресурсів України, Інститут розробки інформаційних систем, 2020. – Режим доступу: <http://monitoring.davr.gov.ua/EcoWaterMon/GDKMap/Index>

[88] Shmundiak D., «Comparing time series prediction models» Kaggle Notebook, 2024. [Електронний ресурс]. Режим доступу: <https://www.kaggle.com/code/dimashmundiak/comparing-time-series-prediction-models>.

ДОДАТКИ

ДОДАТОК А

СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА ЗА ТЕМОЮ ДИСЕРТАЦІЇ

[1] В. Б. Мокін, О. В. Слободянюк, О. М. Давидюк, і Д. О. Шмундяк, «Інформаційна технологія пошуку можливих джерел підвищеного забруднення річки з використанням моделі Prophet», *Вісник Вінницького політехнічного інституту*, № 4, с. 15–24, (Верес. 2020). DOI: <https://doi.org/10.31649/1997-9266-2020-151-4-15-24>. (**Index Copernicus, Наукове фахове видання України, категорія «Б» зі спеціальності 124**)

[2] В. Б. Мокін, Є. М. Крижановський, А. Р. Ящолт, і Д. О. Шмундяк, «Технологія проєктування мережі спостережень якості атмосферного повітря регіону на основі методу аналізу ієрархій», *Наукові Праці ВНТУ*, № 4, (Лют. 2022). DOI: <https://doi.org/10.31649/2307-5376-2021-4-21-33>. (**Наукове фахове видання України, категорія «Б» зі спеціальності 124**)

[3] Д. О. Шмундяк і В. Б. Мокін, «Метод ідентифікації параметрів гармонік та аномалій періодичного часового ряду на основі адаптивної декомпозиції», *Вісник Вінницького політехнічного інституту*, № 6, с. 46–56, (Груд. 2023). DOI: <https://doi.org/10.31649/1997-9266-2023-171-6-46-56>. (**Index Copernicus, Наукове фахове видання України, категорія «Б» зі спеціальності 124**)

[4] Victoria Rodinkova, Serhii Yuriev, Vitalii Mokin, Mariia Kryvopustova, Dmytro Shmundiak, Mykyta Bortnyk, Yevhenii Kryzhanovskyi, Andrii Kurchenko, «Bayesian analysis suggests independent development of sensitization to different fungal allergens», *World Allergy Organization Journal*, Volume 17, (Issue 5, 2024) . DOI: <https://doi.org/10.1016/j.waojou.2024.100908>. (**Scopus i Web of Science**)

[5] Д. О. Шмундяк і В. Є. Копняк, «Метод ідентифікації локальних аномалій значень показників стану довкілля з використанням декомпозиції на півхвилі», *Вісник Вінницького політехнічного інституту*, № 1, с. 88–100, (Лют. 2024). DOI: <https://doi.org/10.31649/1997-9266-2024-172-1-88-100>. (**Index Copernicus, Наукове фахове видання України, категорія «Б» зі спеціальності 124**)

[6] Д. О. Шмундяк, В. Б. Мокін, «Системний аналіз стану природних середовищ з урахуванням аномалій», *Наукові Праці ВНТУ*, № 4, (Грудень 2024)/ DOI: <https://doi.org/10.31649/2307-5376-2024-4-63-73/>. (**Наукове фахове видання України, категорія «Б» зі спеціальності 124**)

[7] В. Б. Мокін, Є. М. Крижановський, А. Р. Яцолт, Д. О. Шмундяк, Б. В. Паєвський, «Оптимізація моніторингової мережі стану атмосферного повітря вінницької зони за даними 2016-2020 рр.», *матеріали LI Науково-технічної конференції факультету інтелектуальних інформаційних технологій та автоматизації*, Вінниця, 31 травня - 1 червня 2022 р. – Електрон. текст. дані. – 2022. – Режим доступу: <https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2022/paper/view/16118/13531>

[8] Д. О. Шмундяк, Н. С. Іжаківська, Д. О. Литвиненко, А. О. Судець, «Аналіз можливостей Python-бібліотек щодо виявлення аномальних даних у задачі прогнозування стану атмосферного повітря», *матеріали LII Науково-технічної конференції факультету інтелектуальних інформаційних технологій та автоматизації*, 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/18957/15722>

[9] Д. О. Шмундяк, «Дослідження впливу параметрів моделі Prophet на точність прогнозування якості атмосферного повітря», *матеріали LIII Науково-технічної конференції факультету інтелектуальних інформаційних технологій та автоматизації*, 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/18210/15056>

[10] Д. О. Шмундяк, А. В. Лосенко, В. Б. Мокін, «Огляд підходів до визначення порядку Фур'є у моделі Facebook Prophet для моделювання сезонної складової часового ряду», *матеріали LIII Науково-технічної конференції факультету інтелектуальних інформаційних технологій та автоматизації*, 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/17200/14329>

[11] В. Б. Мокін, Д. О. Шмундяк, В. Є. Копняк, «Порівняльний розвідувальний аналіз даних про стан атмосферного повітря за даними мережі EcoCity з використанням бібліотеки Sweetviz», *матеріали LIII Науково-технічної конференції факультету інтелектуальних інформаційних технологій та автоматизації*, Вінниця, 20 - 22 березня 2024 р. – Електрон. текст. дані. – 2024. – Режим доступу: <https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2024/paper/view/20846/17251>

[12] Порівняльний аналіз моделей прогнозування показника стану атмосферного повітря / Д. О. Шмундяк // *Матеріали XVII Міжнародної конференції «Контроль і управління в складних системах» (КУСС-2024) [Електронний ресурс]*, Вінниця, 16-17 жовтня 2024 року] – Режим доступу: <https://conferences.vntu.edu.ua/index.php/mccs/mccs2024/paper/view/22120>

[13] Метод пошуку можливих джерел забруднення атмосферного повітря на основі аналізу аномалій / Д. О. Шмундяк // *Матеріали XXIII Міжнародної науково-практичної конференції Математичне моделювання та інформаційно-комунікаційні технології для перемоги та відновлення*, Вінниця, 12-13 листопада 2024 р. – Режим доступу: https://itgip.org/wp-content/uploads/2024/11/2024-11-24_zbirka_all_07_11_2024_148x210.pdf.

[14] Аналіз взаємовпливів рядів даних моніторингу стану атмосферного повітря з використанням методу байєсівського моделювання / Д. О. Шмундяк // *Матеріали LIV Всеукраїнської науково-технічної конференції підрозділів Вінницького національного технічного університету*, Вінниця, 24 - 27 березня 2025 р. – Електрон. текст. дані. – 2025. – Режим доступу: <https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2025/paper/view/23870/19675>.

ДОДАТОК Б

ПОКАЗНИКИ ЗАБРУДНЕННЯ СТАНУ ПРИРОДНИХ СЕРЕДОВИЩ

До основних показників забруднення повітря належать:

1. Тверді частинки (англійською «Particulate Matter», або просто «PM») – дрібний пил, який може проникати в легеневу тканину або кров і викликати серйозні наслідки для функціонування організму. Прийнято поділяти на декілька груп, що відповідають розмірам частинок. Частинки, розміром менше 1 мкм, називаються PM1 (або PM1.0), частинки, розміром менше 2,5 мкм – PM2.5, а частинки розміром до 10 мкм, відповідно, – PM10;

2. Озон (хімічна формула «O₃») – газ, що утворюється в результаті реакції сонячного світла та оксиди азоту, який, в свою чергу, утворюються в результаті роботи транспортних засобів, об'єктів промисловості, хімічних заводів та електростанцій;

3. Діоксид азоту (хімічна формула «NO₂») і оксид азоту (хімічна формула «NO»). Оксиди азоту утворюються в промисловості, шляхом спалювання палива при високій температурі або з автомобільних вихлопних газів. Також можуть проникати в побут із зовнішнього повітря при палінні або через користування газовими плита;

4. Діоксид сірки (хімічна формула «SO₂») — безбарвний газ з різким запахом, який утворюється при спалюванні вугілля або нафти. Джерелами значних обсягів даного газу є побутове опалення, виробництва електроенергії та автотранспорт;

5. Оксид вуглецю (хімічна формула «CO») — безбарвний отруйний газ, який утворюється при спалюванні будь-якого виду палива, що робить його дуже поширеним: ТЕС, ТЕЦ, металургія, нафтопереробна промисловість, транспортні засоби.

До основних показників забруднення водних ресурсів відносять:

1. Біохімічне споживання кисню (БК5) - це кількість кисню в міліграмах, потрібна для окиснення органічних речовин, що містяться в 1 л води, аеробними

бактеріями до CO_2 і H_2O впродовж 5 діб без доступу повітря і світла. Це важливий екологічний показник стану природних водойм. За високого вмісту органічних речовин у воді швидко розмножуються аеробні бактерії, для життєдіяльності яких необхідний кисень. Це може зумовити зниження вмісту розчиненого кисню, створити гіпоксичні умови і загибель окремих видів гідробіонтів.

2. Нітрати (NO_3) – сполуки, що утворюються внаслідок окислення аміаку (NH_3) в процесі нітрифікації. Основними джерелами нітратів у водних ресурсах є сільське господарство (зокрема, використання азотних добрив), стічні води, розкладання органічних речовин і промислові відходи. Високий рівень нітратів у питній воді може призвести до метгемоглобінемії - стан підвищеного рівня метгемоглобіну в крові, що викликає зниження здатності крові переносити кисень. Крім того надлишок нітратів може викликати евтрофікацію водних екосистем, що призводить до масового росту водоростей, зниження концентрації розчиненого кисню і загибелі риби та інших водних організмів.

3. Нітрити (NO_2) – сполуки, що утворюються як проміжний продукт в процесі нітрифікації, коли аміак окислюється до нітратів. Джерела нітритів аналогічні джерелам нітратів, хоча їх концентрація у водних ресурсах зазвичай значно нижча через швидке перетворення в нітрати. Нітрити є більш токсичними для людей, ніж нітрати, оскільки вони безпосередньо впливають на гемоглобін, що призводить до утворення метгемоглобіну. Вони також можуть реагувати з амінами, утворюючи нітрозаміни, які є потенційними канцерогенами. Як і нітрати, нітрити можуть спричинити евтрофікацію водних ресурсів, але цей вплив менш значний через нижчу стабільність нітритів у воді.

4. Фосфати (PO_4) – солі, які надходять у водні ресурси в основному з миючих засобів, добрив, стічних вод і природного вивітрювання гірських порід. Вони є основною формою фосфору, що легко засвоюється водоростями і рослинами. Фосфати є основним лімітуючим фактором для росту водоростей. Навіть невелика кількість фосфатів у воді може викликати евтрофікацію, що призводить до цвітіння водоростей, зменшення рівня кисню та загибелі водних організмів. Цей процес також може вплинути на естетичні та рекреаційні

властивості водних ресурсів. Прямий вплив фосфатів на здоров'я людини менш значний порівняно з нітратами та нітритами, однак надлишок фосфатів у водних ресурсах може негативно вплинути на загальну екологічну якість води, що непрямо може вплинути на здоров'я людини через забруднення джерел питної води.

5. Амоній (NH_4) – хімічна сполука, що утворюється внаслідок розкладання органічних речовин, таких як рослинні залишки, тваринні відходи та мікроорганізми. Основні джерела амонію у водних ресурсах: стічні води (побутові та промислові стоки, які можуть містити аміак або амонійні сполуки), сільське господарство (використання добрив, багатих на азот, а також розкладання органічних добрив, таких як гній), атмосферні опади (амоній може потрапляти у воду разом з атмосферними опадами, особливо в регіонах з високим рівнем забруднення повітря). Амоній сам по собі є менш токсичним для водних організмів, ніж аміак. Проте за умов високого рН і температури може утворюватися аміак, який є високотоксичним для риби та інших водних організмів. Аміак може викликати пошкодження жабр, нервової системи та може бути смертельним при високих концентраціях. Амоній, як і інші форми азоту, може сприяти евтрофікації водних екосистем, що призводить до надмірного росту водоростей і погіршення якості води. У процесі нітрифікації амоній окислюється до нітритів, а потім до нітратів, при цьому споживається велика кількість розчиненого кисню, що може призвести до зниження кисню у воді і створення анаеробних умов. Висока концентрація амонію в питній воді може свідчити про наявність інших забруднювачів, таких як органічні речовини або патогенні мікроорганізми. Хоча амоній сам по собі не є дуже токсичним для людини, його присутність може бути ознакою недостатньої очистки води. При хлоруванні води, що містить амоній, можуть утворюватися хлорорганічні сполуки, такі як хлораміни, які мають неприємний запах і смак, а також можуть бути шкідливими для здоров'я.

6. Сульфати (SO_4) – клас мінералів, можуть потрапляти у водні ресурси з різних джерел: природне вивітрювання гірських порід (сульфати природно

утворюються при вивітрюванні мінералів, таких як гіпс ($\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$) та ангідрит (CaSO_4), що містять сірку), атмосферні опади (сульфати можуть утворюватися в атмосфері внаслідок окислення сірчаних газів, таких як сірчистий газ (SO_2), і потім потрапляти в воду з дощем або снігом), промислові викиди (викиди від спалювання вихопного палива, особливо вугільних електростанцій і металургійних підприємств, можуть бути джерелом сульфатів у водних ресурсах), сільське господарство (використання сульфатних добрив може призводити до забруднення ґрунтових вод і поверхневих вод сульфатами), каналізаційні стоки (стоки з промислових підприємств і очисних споруд можуть містити підвищені концентрації сульфатів). Сульфати самі по собі зазвичай не є токсичними для водних організмів у звичайних концентраціях, що зустрічаються у природних умовах. Проте високі концентрації сульфатів можуть впливати на водні рослини і тварин, викликаючи осмотичний стрес або зміну рН води. Сульфати можуть впливати на кислотність води, особливо у випадках, коли вони асоціюються з кислотними дощами. Це може призвести до зниження рН води, що є несприятливим для багатьох водних організмів. У деяких умовах сульфати можуть бути відновлені до сульфідів (S^{2-}) анаеробними бактеріями, що може призвести до утворення токсичного сірководню (H_2S), який є дуже небезпечним для водних організмів. Вода з високим вмістом сульфатів має гіркий або солоний смак. Сульфати можуть спричиняти проносний ефект у людей, особливо у дітей та людей, не звиклих до води з високим вмістом сульфатів. Це особливо важливо в районах, де джерела питної води мають природно високі концентрації сульфатів. Сульфати можуть сприяти корозії трубопроводів, що може призвести до забруднення води продуктами корозії.

ДОДАТОК В

РОЗВІДУВАЛЬНИЙ АНАЛІЗ ДАНИХ

Розвідувальний аналіз даних є важливим етапом дослідження даних, що дозволяє ідентифікувати основні характеристики та закономірності, які присутні в цих даних. Знання, отриманні в результаті цього аналізу, дозволяють створювати ефективні моделі для передбачення даних.

Найпростішим, але у той же час теж корисним, є звичайний статистичний звіт по набору даних. Це дозволяє побачити кількість пропущених даних та їх приблизний розподіл, що, у свою чергу, допоможе правильно обробити дані перед подальшим аналізом. Найпростішим методом створення описової статистики є метод «describe» з бібліотеки pandas. Результат виклику цього методу містить статистику, яка підсумовує центральну тенденцію, дисперсію та форму розподілу набору даних (рис. В.1).

indicator_name	ds	Humidity	PM10	PM2.5	Pressure	Temperature	VOC (H ₂ CO)
count	1409	1316.000000	1316.000000	1316.000000	1316.000000	1316.000000	1290.000000
mean	2022-03-29 07:00:02.555003392	57.393446	12.142296	8.068391	98755.962347	13.078716	9.737088
min	2020-01-22 00:00:00	22.581800	0.907300	0.657500	95829.707100	-8.678400	9.000000
25%	2021-04-16 00:00:00	47.091075	4.100875	3.219300	98273.677000	5.080300	9.377750
50%	2022-04-03 00:00:00	57.984750	8.205900	6.042800	98760.952400	12.026750	9.995900
75%	2023-03-25 00:00:00	67.971600	15.327775	10.571200	99269.527700	21.460075	10.000000
max	2024-03-11 00:00:00	99.064600	158.840700	77.698800	101507.842900	34.762000	12.005900
std	NaN	13.001544	12.388034	6.970500	787.119241	9.193796	0.380268

Рисунок В.1 – Описові статистичні дані набору

Іншим корисним інструментом розвідувального аналізу є обчислення попарної кореляції стовпців. Цей метод дозволяє зрозуміти залежності між парами стовпців – висока кореляція між двома стовпцями може говорити про надлишковість (дані є дуже схожими і один з подібних стовпців може бути викинуто з загального набору). Дуже високі або низькі значення кореляції можуть

бути сигналами щодо можливих проблем або помилок в даних, що вимагатиме більш детального аналізу.

Побудувати кореляційну матрицю можна через метод «corr» з бібліотеки pandas (рис. В.2).

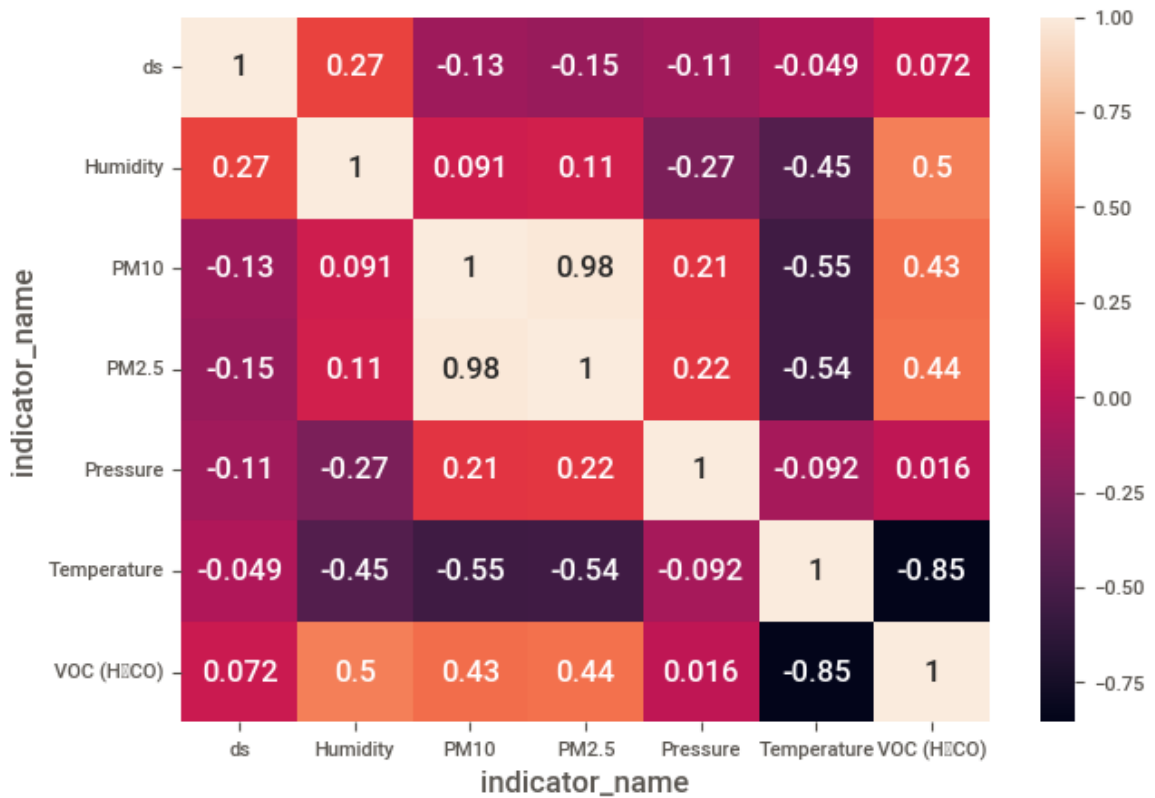


Рисунок В.2 – Кореляційна матриця для попарних комбінацій усіх параметрів часового ряду

Як можна побачити на рисунку В.2, існує велика кореляція між показником пилу PM2.5 та PM10, що є доволі очевидним, оскільки PM2.5 фактично є підмножиною показника PM10, хоча вони і відрізняються джерелами походження. Також існує залежність між значенням показника формальдегіда та значеннями вологості повітря.

Наступним методом аналізу даних часового ряду є аналіз його сезонних компонентів, який проводиться за рахунок сезонної декомпозиції, що дозволяє виділити тренд, сезонність та залишки. Виявлення сезонних коливань забруднень допомагає зрозуміти, як певні природні або антропогенні фактори впливають на

якість повітря у різні пори року. Декомпозиція допомагає побудувати більш точні моделі прогнозування якості повітря, оскільки кожна з компонент може бути змодельована окремо. Одним з методів для виділених сезонних компонентів є бібліотека `statsmodel` та її модуль «`seasonal_decompose`». Застосуємо цей метод для даних показника пилу $PM_{2.5}$ з значенням періоду 30 (рис. В.3).

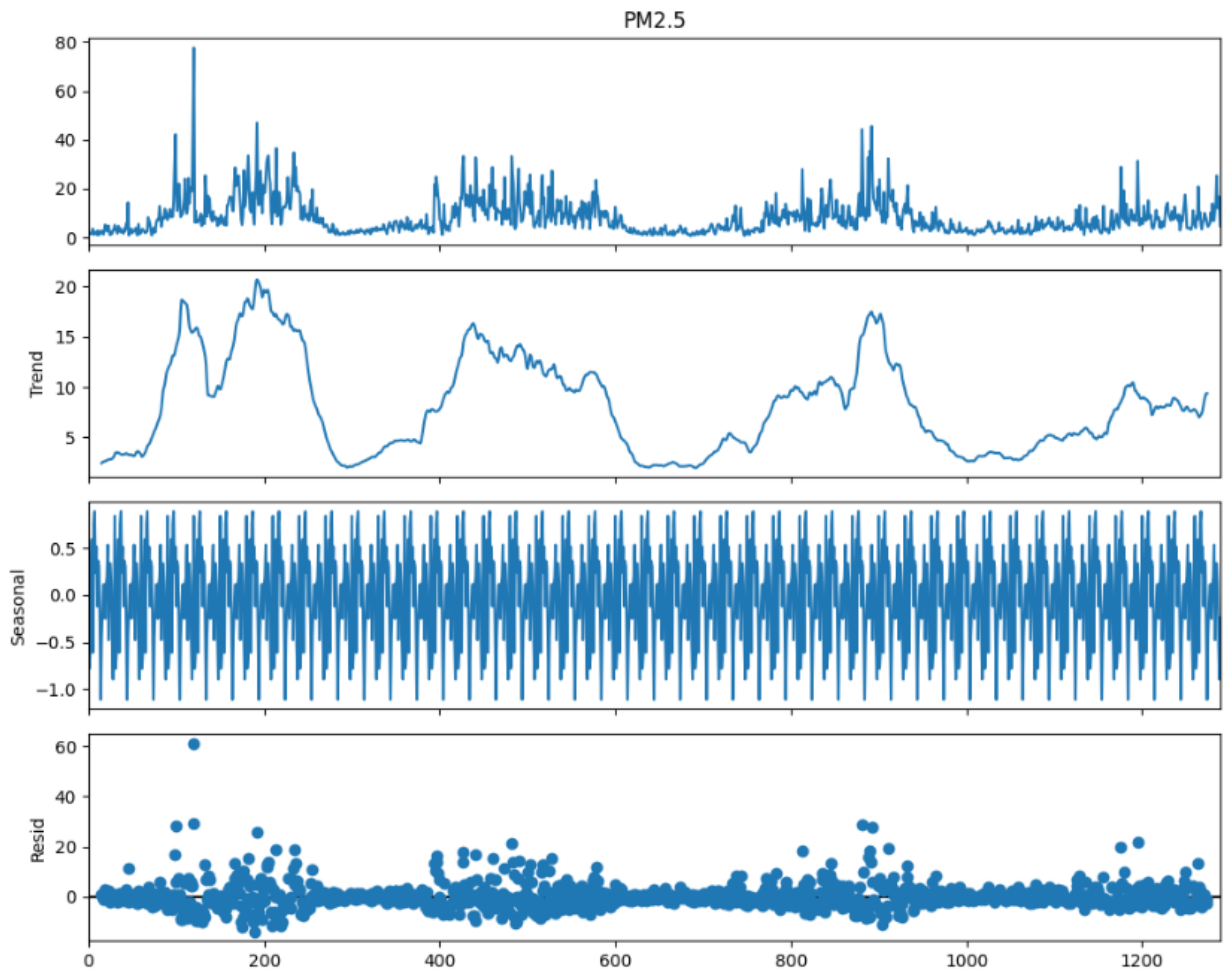


Рисунок В.3 – Сезонна декомпозиція показника пилу $PM_{2.5}$ з періодом 30

Також корисним є використання бібліотек для автоматизації розвідувального аналізу. Такі бібліотеки допомагають швидко зрозуміти основні характеристики наявних даних та виявити потенційні проблеми перед подальшим моделюванням та прогнозуванням.

Прикладом таких бібліотек є `SweetViz` – бібліотека для візуалізації та аналізу даних в середовищі `Python`. Вона автоматично будує гістограми для всіх

числових та категоріальних змінних. Є можливість аналізу зв'язку між числовою та категоріальною ознаками, будуючи графіки відносності. Це дозволяє визначити, як числові характеристики змінюються в залежності від категоріальних ознак. SweetViz надає зручні інтерфейси для порівняння двох різних наборів даних. Це може бути корисно для аналізу різниці між наборами даних, наприклад, навчальним та тестовим, або даними, зібраними у різний час.

Застосуємо дану бібліотеку для показників однієї станції моніторингу якості атмосферного повітря за 2022 та 2023 роки (рис. В.4).

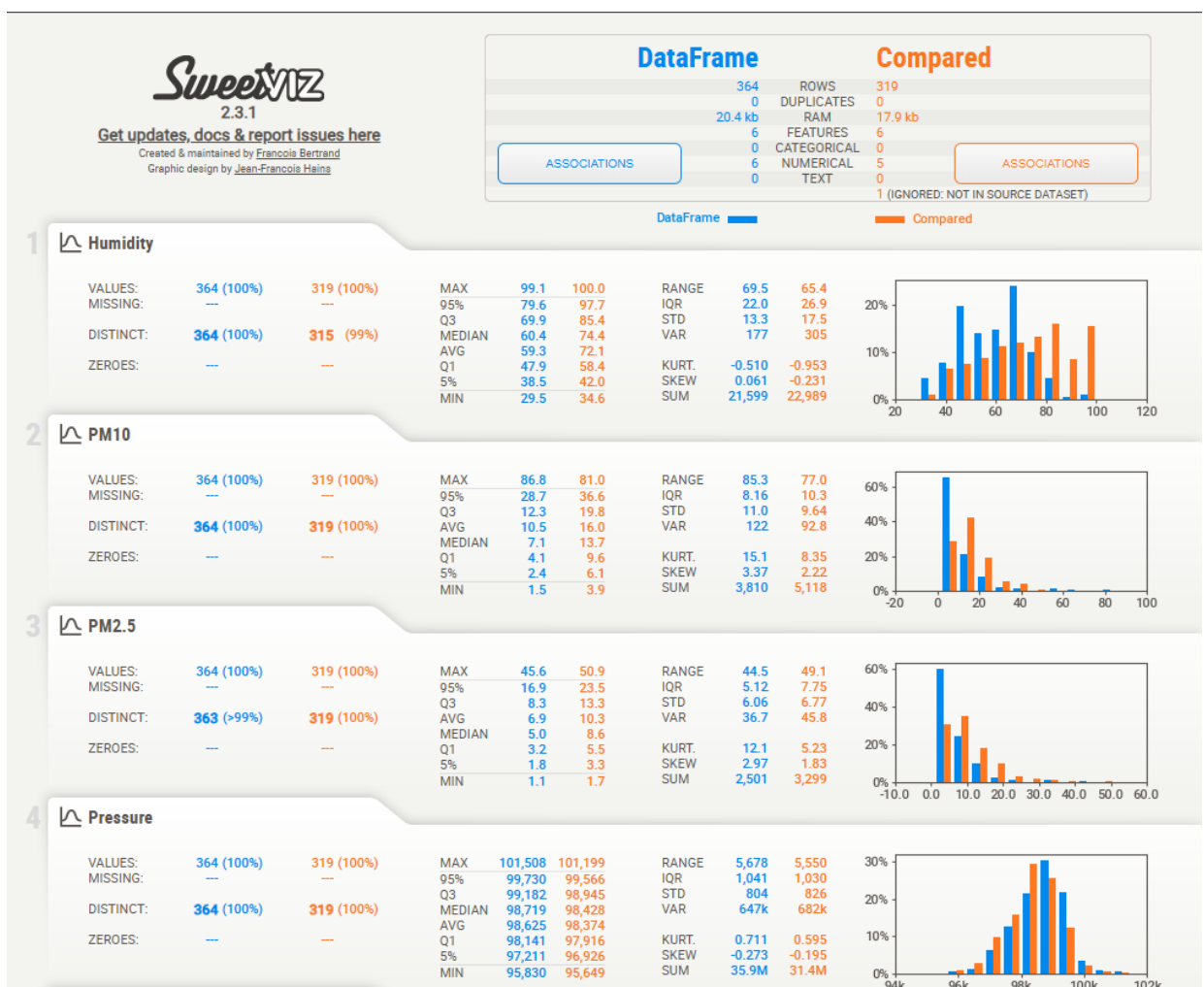


Рисунок В.4 – Результат роботи бібліотеки SweetViz

Як видно на рисунку В.4, бібліотека створює інтерактивний звіт з результатами порівняння наявних показників у відповідних наборах даних. Для

кожного показника можна переглянути деталі порівняння, як це зображено на рисунку В.5.



Рисунок В.5 – Звіт щодо показника пилу PM10, отриманий через бібліотеку SweetViz

Аналізуючи отримані результати, можна зробити висновок, що ті самі показники однієї станції мало відрізняються по роках. Крім того, показники однієї станції загалом мають подібні динаміку.

Іншою бібліотекою для автоматизованого розвідувального аналізу є AutoViz. AutoViz автоматично визначає тип графіка для кожної змінної в

залежності від її характеристик. Наприклад, числові змінні можуть бути відображені у вигляді гістограм, діаграм розсіювання або лінійних графіків, а категоріальні – у вигляді кругових діаграм чи стовпчикових графіків. Крім того, бібліотека може робити згруповані аналізи, наприклад, враховуючи залежності між змінними або розподіл значень змінних за певною категорією та генерувати інтерактивні графіки.

Застосувавши бібліотеку для наявних даних моніторингу якості повітря, отримуємо декілька графіків та різних аналітичних результати. Одна з частин результату роботи бібліотеки це аналіз якості даних (рис. В.6).

indicator_name	Data Type	Missing Values%	Unique Values%	Minimum Value	Maximum Value	DQ Issue
Humidity	float64	0.000000	NA	22.581800	99.064600	No issue
PM10	float64	0.000000	NA	0.907300	86.797800	Column has 36 outliers greater than upper bound (26.46) or lower than lower bound(-9.93). Cap them or remove them.
PM2.5	float64	0.000000	NA	0.657500	45.575400	Column has 32 outliers greater than upper bound (17.68) or lower than lower bound(-5.87). Cap them or remove them., Column has a high correlation with [PM10]. Consider dropping one of them.
Pressure	float64	0.000000	NA	95829.707100	101507.842900	Column has 16 outliers greater than upper bound (100658.95) or lower than lower bound(96748.51). Cap them or remove them.
Temperature	float64	0.000000	NA	-4.606800	31.648800	No issue
VOC (H ₂ CO)	float64	0.000000	NA	9.012200	11.569500	Column has 2 outliers greater than upper bound (10.96) or lower than lower bound(8.40). Cap them or remove them., Column has a high correlation with [Temperature]. Consider dropping one of them.
class	int64	0.000000	0	0.000000	1.000000	Target column

Рисунок В.6 – Аналітична інформація, отримана за допомогою бібліотеки AutoViz

Проведений аналіз показує, що для деяких з показників присутні певні аномалії – перевищення граничних значень. На ці дані варто звернути увагу та опрацювати перед початком моделювання моделі передбачення.

ДОДАТОК Г
РОЗКРИТТЯ ВИДІВ НЕВИЗНАЧЕНОСТІ

Таблиця Г.1 – Розкриття видів невизначеності запропонованими та удосконаленими методами системного аналізу стану природного середовища

№ п/п	Вид декомпозиції	Вид невизначеності	Підрозділ чи пункт роботи, де було розкрито невизначеність
1.	D_{DPSIR}	U_{S1}	Підрозділ 2.3
2.	D_{DPSIR}	U_{T3}	Підрозділ 2.4
3.	D_{DPSIR}	U_{T5}	Підрозділ 2.4
4.	D_{DPSIR}	U_{T1}	Підрозділ 2.4
5.	D_{DPSIR}	U_{P2}	Пункт 3.1.1
6.	D_{DPSIR}	U_{P1}	Пункт 3.1.2
7.	D_{G-O}	U_{T6}	Пункт 3.1.3
8.	$D_T(\Delta T)$	U_{S2}	Підрозділ 3.2
9.	$D_T(\Delta T)$	U_{ST1}	Підрозділ 3.3
10.	$D_{SEC}(\Theta)$	U_{ST2}	Підрозділ 3.3
11.	$D_{SEC}(\Theta)$	U_{SC1}	Підрозділ 3.4

ДОДАТОК Д

АКТ ПРО ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ У ПРОГРАМІ
ДЕРЖАВНОГО МОНІТОРИНГУ У ГАЛУЗІ ОХОРОНИ
АТМОСФЕРНОГО ПОВІТРЯ ЗОНИ «ВІННИЦЬКА»

ЗАТВЕРДЖУЮ
 В.о. начальника управління розвитку територій
 та інфраструктури Вінницької обласної
 військової адміністрації
 Владислав ВАЛЕНДЮК
 № 08-1139 від 14.04 2025 р.


АКТ
про впровадження результатів дисертаційної роботи
«Системний аналіз стану природних середовищ з урахуванням аномалій»
Шмундяка Дмитра Олександровича

Цей акт підтверджує, що результати дисертаційної роботи на здобуття наукового ступеня доктора філософії Шмундяка Д. О. було впроваджено під час розроблення Програми державного моніторингу в галузі охорони атмосферного повітря на 2023–2027 роки для зони «Вінницька», яка затверджена Міністерством захисту довкілля та природних ресурсів України, Обласною радою Вінницької області та наказом голови Вінницької обласної військової адміністрації (Наказ №722 від 22.05.2023 р.). Зокрема цінними є наступні отримані результати дослідження Шмундяка Д. О.:

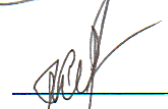
використання методології оптимізації формування ієрархічної системи критеріїв для вибору координат місць розташування постів спостережень за станом атмосферного повітря з урахуванням екстремальних значень даних моніторингу, оснований на DPSIR-декомпозиції («Drivers – Pressures – State – Impact – Responses»). Це дозволило враховувати особливості регіону під час проєктування мережі моніторингу за станом атмосферного повітря;

використання методології системного аналізу стану природних середовищ з урахуванням аномалій, основу на функціональній DPSIR-декомпозиції, на використанні класичних та авторських методів для розкриття основних видів невизначеності, що дозволило підвищити ефективність розроблення Програми державного моніторингу в галузі охорони атмосферного повітря на 2024–2028 роки для зони «Вінницька».

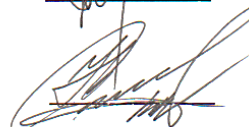
Заступник начальника управління
розвитку територій та інфраструктури
обласної військової адміністрації

 Дмитро ДЯДИК

Начальник відділу охорони земель,
лісової та заповідної справи

 Ігор ФЕДЧИШИН

Головний спеціаліст відділу охорони
земель, лісової та заповідної справи

 Сергій КРАВЧУК

ДОДАТОК Е
АКТ ПРО ВПРОВАДЖЕННЯ РЕЗУЛЬТАТІВ
У НАВЧАЛЬНИЙ ПРОЦЕС ВНТУ



ЗАТВЕРДЖУЮ

Проректор з науково-педагогічної роботи та організації освітнього процесу Вінницького національного технічного університету

доц. Олександр ПЕТРОВ

« 03 » 2025 р.

АКТ

про впровадження результатів дисертаційної роботи
Шмундяка Дмитра Олександровича

«Системний аналіз стану природних середовищ з урахуванням аномалій»,
представленої до захисту на здобуття наукового ступеня доктора філософії зі
спеціальності 124 Системний аналіз, в навчальному процесі

Комісія Вінницького національного технічного університету у складі:

голова комісії – декан Факультету інтелектуальних інформаційних технологій та автоматизації (ФІТА), к.т.н, доцент Севастьянов В. М. та члени комісії – заступник декана з навчально-методичної роботи ФІТА, ст. викладач каф. САІТ Присяжнюк В. В., заступник завідувача кафедри системного аналізу та інформаційних технологій (САІТ) ФІТА, к.т.н., доцент Крижановський Є. М., склала цей акт про підтвердження того, що дійсно у Вінницькому національному технічному університеті під час викладання дисципліни «Інформаційні технології моніторингу та аналізу стану складних систем» для аспірантів, які навчаються за освітньою програмою «Системний аналіз» рівня «доктора філософії» цієї ж спеціальності 124, впроваджено результати дисертаційної роботи на здобуття доктора філософії аспіранта кафедри САІТ Шмундяка Дмитра Олександровича у низці лекцій та практичних занять з питань аналізу та пошуку аномалій часових рядів.

При викладанні цієї дисципліни використовуються такі результати досліджень, отримані у дисертаційній роботі Шмундяка Д. О.:

- методологія системного аналізу стану природних середовищ з урахуванням аномалій, основу на функціональній DPSIR-декомпозиції, на використанні інтелектуальних моделей та класичних і авторських методів для розкриття основних видів невизначеності, що дозволяє підвищити точність прогнозування цього стану;

- спосіб викладення методології системного аналізу шляхом структурування результатів декомпозиції, видів невизначеності і методів щодо їх розкриття;

- використання різних методів пошуку аномалій (Isolation Forest, SESD) за рахунок використання відомих Python-бібліотек (statsmodels, sklearn) та запропонованих автором підходів;

- ідентифікація параметрів аномалій та різних варіантів сезонності часового ряду на основі інтелектуальної моделі Prophet з використанням як класичних, так і авторських підходів.

Використання зазначених результатів дозволило підвищити якість навчального процесу із згаданої дисципліни.

Голова комісії

Володимир СЕВАСТЬЯНОВ

Члени комісії

Василь ПРИСЯЖНЮК

Євгеній КРИЖАНОВСЬКИЙ

ДОДАТОК Є
СВІДОЦТВО ПРО РЕЄСТРАЦІЮ АВТОРСЬКОГО ПРАВА НА ТВІР

УКРАЇНА



СВІДОЦТВО

про реєстрацію авторського права на твір

№ 134476

Комп'ютерна програма «Python-реалізація методу ідентифікації параметрів гармонік та аномалій інтелектуальної моделі часового ряду на основі Prophet» («Парам гармонік»)

(вид, назва твору)

Автор (співавтори) Шмундяк Дмитро Олександрович, Мокін Віталій Борисович

(прізвище, ім'я, по батькові (за наявності), псевдонім (за наявності))

Авторські майнові права належать спільно Шмундяк Дмитро Олександрович, вул. Виноградна, 19, с. Немія, Могилів-Подільський р-н, Вінницька обл., 24006; Мокін Віталій Борисович, вул. Скалецького, 35/194, кв. 111, м. Вінниця, 21018

(прізвище, ім'я, по батькові (за наявності) фізичної особи / найменування юридичної особи, адреса)

Дата реєстрації 17 березня 2025 р.

Директор Державної організації
«Український національний
офіс інтелектуальної власності
та інновацій»


Олена ОРЛЮК

