

ВИКОРИСТАННЯ МЕТОДІВ ДИСКРЕТНОГО ВЕЙВЛЕТ- ПЕРЕТВОРЮВАННЯ ТА БІБЛІОТЕК PYTHON ДЛЯ ОТРИМАННЯ МАТЕМАТИЧНИХ МОДЕЛЕЙ ЕКОЛОГІЧНИХ ДАНИХ

Ольга Тузенко¹, Наталія Сідун², Єгор Волобуєв³

^{1, 2, 3} Державний вищий навчальний заклад «Приазовський державний технічний університет»,

кафедра інформатики, Дніпро, Україна

E-Mail: tuzenkoaa@gmail.com, ORCID: 0000-0002-4920-9417

E-Mail: natalya.sidun@gmail.com, ORCID: 0009-0001-8343-1273

E-Mail: egorwww7@gmail.com, ORCID: 0009-0006-5221-8134

© Ольга Тузенко, Наталія Сідун, Єгор Волобуєв, 2025

У статті розглянуто актуальну проблему комп'ютерного моделювання великих масивів даних екологічного моніторингу із застосуванням дискретних вейвлет-перетворень. Об'єкт дослідження становлять часові ряди концентрацій шкідливих домішок в атмосферному повітрі – зокрема оксидів азоту, бензолу, сірчистого газу, які зібрано з автоматизованих станцій у Центральній та Східній Європі. Вхідні дані характеризуються високим рівнем стохастичності, наявністю шумів, втрачених значень і часових зсувів, що суттєво ускладнює виділення трендів і закономірностей, необхідних для прогнозування. Запропоновано ефективний підхід до їх обробки, який поєднує багаторівневу вейвлет-декомпозицію з адаптивною фільтрацією на основі soft-thresholding та механізмом автоматичного вибору рівня декомпозиції, що ґрунтується на енергетичному аналізі коефіцієнтів.

Для моделювання використано екосистему бібліотек Python: PyWavelets для побудови хвильових моделей, Matplotlib – для візуалізації динаміки, Pandas – для структурування великих вхідних масивів. У рамках експериментального дослідження проведено серію комп'ютерних експериментів на прикладі польських, чеських і німецьких міст, що підтвердили дієвість застосованого методу. Продемонстровано, що поєднання вейвлет-перетворень Добеші (db4) з методами ансамблевого навчання (Random Forest, XGBoost) та нейронними мережами типу LSTM дозволяє досягати високої точності прогнозування навіть за умов нестаціонарних викидів та погодних флуктуацій. Запропонований підхід може бути використаний як основа для побудови адаптивних систем екологічного моніторингу нового покоління, зокрема у містах з високим рівнем індустріального навантаження.

Ключові слова – навколишнє середовище, забруднення повітря, концентрація домішок, часовий ряд, вейвлет-функція, вейвлет-перетворення, математична модель, бібліотеки Python.

Постановка проблеми

Забруднення атмосферного повітря дедалі частіше стає об'єктом занепокоєння на глобальному рівні, адже воно справляє істотний негативний вплив як на здоров'я людини, так і на стан довкілля та функціонування господарських систем. Надійний і точний моніторинг якості повітря

відіграє ключову роль у впровадженні дієвих механізмів контролю за забрудненням та у пом'якшенні його згубних наслідків. Наразі існує велика кількість чинників, що впливають на стан атмосфери, і комплексне врахування кожного з них ускладнюється з кожним роком. У контексті стрімкої індустріалізації, зростання транспортного навантаження та воєнних конфліктів, обсяги викидів в атмосферу досягають критичних показників. Одним із найефективніших інструментів оцінки якості атмосферного повітря є системи безперервного моніторингу концентрацій основних забруднюючих речовин (Kimon Hadjibiros, 2013). Саме автоматизоване безперервне спостереження дозволяє забезпечити найповніше накопичення даних для подальшого аналітичного опрацювання. Проте специфіка такого підходу полягає в тому, що він генерує надзвичайно великі обсяги екологічної інформації. З одного боку, доступ до даних у режимі реального часу створює можливості для оперативного виявлення потенційних екологічних ризиків і більш глибокого розуміння еволюції екосистем, що, своєю чергою, підтримує прийняття обґрунтованих управлінських рішень. З іншого боку, масштаб і складність таких даних створюють істотне навантаження на інформаційні системи обробки. До того ж, зібрані дані мають стохастичний характер і часто містять шум, що значно ускладнює виявлення сталих закономірностей і тенденцій у часових рядах.

У процесі аналізу результатів спостережень важливо не лише інтерпретувати власне числові значення концентрацій окремих речовин, а й інтегрувати дані щодо потенційних джерел забруднення, характеру впливу зовнішніх чинників, їхньої інтенсивності та змін у часі. Проте витягнення значущої інформації з великих і складних екологічних масивів даних залишається складним завданням, що потребує застосування спеціалізованих методів математичного моделювання та інтелектуального аналізу.

Аналіз останніх досліджень та публікацій

Засновані на математичних перетвореннях підходи традиційно застосовуються для аналізу часових рядів, зокрема для виявлення закономірностей глобального рівня, таких як характерні частоти зміни концентрацій забруднюючих речовин у довготривалих спостереженнях. Проте ці методи мають обмеження, коли йдеться про вивчення більш динамічних і локалізованих процесів. Наприклад, ситуативні викиди, зумовлені раптовим зростанням транспортного потоку, локальними пожежами або вибухами внаслідок бойових дій, а також змінами погодних умов, не завжди коректно відображаються за допомогою класичних трансформацій. Крім того, такі підходи є надто узагальненими і не пристосовані до відображення швидкоплинних змін у роботі джерел забруднення. Їх практична реалізація у сфері екологічного моніторингу ускладнюється значними обчислювальними витратами та труднощами у побудові високоточної моделі (Serov, 2017).

На відміну від цього, вейвлет-аналіз дозволяє отримати не строго точне, але інформативне наближене представлення сигналу, з можливістю локалізації аномальних значень у даних. Завдяки цьому вдається виявити короткочасні явища, як-от зміни вітрового режиму, короткотривале функціонування окремих джерел забруднення або стохастичні коливання, що притаманні навіть стабільним джерелам викидів. Вейвлети ефективно працюють як інструмент багаторівневої декомпозиції, що дозволяє розкласти складні атмосферні часові ряди на простіші компоненти. Це, у свою чергу, сприяє виявленню фундаментальних закономірностей, прихованих у необроблених масивах даних, таких як сезонні коливання температури повітря чи довготривалі кліматичні тенденції (Ekstrom & Hales, 2000). Зважаючи на це, для дослідження даних, отриманих у межах екологічного моніторингу, було прийнято рішення застосувати математичний апарат вейвлет-перетворень та побудувати відповідну модель, яка включає як дискретні, так і безперервні варіанти перетворень (Tuzenko & Sidun, 2023).

Формулювання цілі статті

Метою цього дослідження є розроблення комп'ютерних моделей для аналізу даних, що надходять у процесі моніторингу концентрацій забруднюючих речовин у приземному шарі атмосфери. Особлива увага приділяється вивченню впливу супутніх факторів, які можуть істотно змінювати хід коливань

концентрацій: таких як метеорологічні умови, географічне положення станцій спостереження, добові цикли, особливості рельєфу, інтенсивність транспортного потоку та наявність промислових джерел викидів. У рамках дослідження аналізується складна стохастична структура часового ряду, яка виникає в результаті взаємодії великої кількості змінних і джерел даних.

З метою якісного моделювання таких складних і неоднорідних даних було застосовано математичний апарат дискретного вейвлет-перетворення, що дає змогу виконати багаторівневу декомпозицію сигналу, локалізувати аномалії, згладити шум і виокремити структурні патерни. Дискретні вейвлети особливо ефективні при роботі з нестационарними процесами, властивими для екологічних часових рядів, оскільки дозволяють одночасно проводити аналіз як у часовій, так і у частотній області. Використання такого інструменту моделювання спрямоване на отримання надійних аналітичних результатів, що можуть бути застосовані для прогнозування стану повітряного середовища в режимі реального часу або для підтримки прийняття екологічних рішень.

Виклад основного матеріалу

У процесі аналізу даних, що характеризують стан атмосферного повітря в різних містах Європейського регіону, було встановлено, що сучасні системи екологічного моніторингу формують великі обсяги неструктурованої інформації. Зокрема, кожна окрема станція вимірює не лише концентрації певних забруднюючих речовин, але й фіксує низку додаткових параметрів, які можуть прямо чи опосередковано впливати на кінцевий результат вимірювань. До таких параметрів належать географічна віддаленість від потенційних джерел викидів, тип самого джерела (автомобільний, промисловий, побутовий тощо), метеорологічні умови на момент спостереження (температура повітря, напрям і швидкість вітру), висота розташування сенсора над рівнем моря, відстань до меж міської забудови, а також положення станції в межах щільно забудованої або відкритої частини міста. Усі ці змінні формують складне багатовимірне середовище, в якому деякі параметри мають визначальний вплив, інші ж – можуть виявитися статистично незначущими або навіть завадити аналізу, ускладнюючи побудову узагальненої моделі.

Для встановлення впливовості кожного з факторів було проведено попередню оцінку значущості за допомогою критерію Стюдента. Це дало можливість виокремити ті змінні, які мають безпосередній статистичний вплив на рівень зафіксованої концентрації. Подальший аналіз уточнення взаємозв'язків між групами змінних здійснювався із використанням дисперсійного аналізу (ANOVA) (IBM SPSS Statistics, 2021). Зокрема, для станцій, що розташовані в межах одного населеного пункту або перебувають у зоні впливу одного домінантного джерела забруднення, вдалося підтвердити істотність таких параметрів, як тип джерела, відстань до нього, а також час доби, протягом якого здійснювалося спостереження. Отримані висновки було покладено в основу моделі, що поєднує математичний апарат вейвлет-перетворень (Tuzenko & Sidun, 2023) із емпіричними даними, отриманими зі станцій моніторингу.

Для реалізації комп'ютерної моделі аналізу використовувалося середовище програмування Python, зокрема бібліотеки PyWavelets (Wasilewski, n.d.) та PyTorch Wavelets (Introduction – Pytorch Wavelets 0.1.1 Documentation, 2020). Вони надали інструментарій для реалізації дискретних та комплексних вейвлет-перетворень. Як джерело вхідних даних використовувалась платформа UTD Air Quality Data (European Environment Agency, 2024) підтримувана Європейською Екологічною Агенцією (European Environment Agency, 2023), яка забезпечує доступ до актуальних і історичних даних про стан атмосферного повітря у вигляді часових рядів. Спостереження охоплюють широкий спектр забруднювачів (наприклад, NO_2 , SO_2 , C_6H_6 , PM_{10}), що фіксуються у різних містах залежно від локальних умов, структури промисловості, інтенсивності трафіку тощо.

Після проведення первинної очистки та попереднього структурування даних було здійснено моделювання за допомогою дискретного вейвлет-перетворення із використанням вейвлетів Добеші (Daubechies wavelets). Це сімейство функцій особливо ефективно для роботи з нестационарними сигналами, які мають випадкову природу та високу щільність шуму. За рахунок функції масшта-

бування стало можливим застосування "віконних" фільтрів до окремих сегментів часового ряду, що дало змогу виявляти локальні тренди, повторювані закономірності та нетипові відхилення. Добре відома здатність вейвлетів Добеші зберігати структурну цілісність сигналу за мінімального викривлення дозволила провести багатоетапну декомпозицію екологічних рядів з високою точністю.

Ефективність процесу денойзингу оцінювалася з використанням кількох стандартних метрик – середньої абсолютної похибки (MAE), кореня середньоквадратичної похибки (RMSE) та структурного індексу подібності (SSIM), адаптованого для одномірних даних. У ході порівняння було встановлено, що комбінація вейвлета db4 з методом soft-thresholding забезпечує зменшення MAE приблизно на 15 % порівняно з класичним median filter. У складніших умовах, коли вхідний ряд містив потужні флуктуації, спричинені змінами погодних умов або хвилеподібною транспортною активністю, було досягнуто вищої точності при використанні PyTorch-Wavelets і реалізації Dual Tree Complex Wavelet Transform.

Процедура комп'ютерного моделювання просторово-часових змін у концентраціях забруднюючих речовин у повітряному середовищі реалізується у декілька етапів. Першим етапом є збирання вихідних даних, які надходять у вигляді первинних, неструктурованих масивів, зазвичай представлених у форматі CSV-файлів. Ці таблиці містять ключову інформацію: числові значення концентрацій окремих забруднювачів, супутні параметри, такі як тип і назва джерела потенційного викиду, географічні координати точки вимірювання, а також точний час фіксації кожного показника. Такі дані формують основу для подальшої обробки та аналітичного моделювання.

Часто трапляється, що декілька станцій, розташованих у межах однієї агломерації або регіону, здійснюють вимірювання одних і тих самих параметрів в один і той самий часовий момент. Для усунення надмірностей та уточнення структури вибірки виконується об'єднання (злиття) відповідних таблиць з метою виявлення дублікатів, а також виокремлення тих значень, які можуть бути використані для подальшого дисперсійного аналізу. Цей етап є критично важливим для забезпечення достовірності статистичних висновків у процесі дослідження впливу супутніх факторів.

Інформація, отримана з відкритих джерел, зокрема з платформ екологічного моніторингу, зазвичай має табличну структуру з декількох взаємозалежних компонентів. Основна таблиця зазвичай містить часовий ряд виміряних концентрацій, отриманих із кроком у 5 або 10 хвилин. Додаткові таблиці включають дані щодо геопросторового розташування потенційних джерел забруднення, їх класифікації, а також розрахункової відстані від цих джерел до пунктів вимірювання. Крім того, фіксуються координати самих станцій і, за можливості, – характеристики локального середовища (щільність забудови, тип ландшафту, інтенсивність транспортного потоку тощо).

Після синхронізації вхідних масивів даних і проведення перевірки на повноту й узгодженість, було сформовано уніфіковані часові ряди для кожного забруднювача окремо. Об'єднання базових таблиць дозволило відновити послідовності змін концентрацій у прив'язці до факторних параметрів, що відкриває шлях до подальшого вейвлет-аналізу.

Для аналізу було використано часові ряди концентрацій забруднювачів повітря з відкритого джерела (European Environment Agency, 2024) що охоплюють станції у міських і приміських зонах Центральної та Східної Європи. Встановлено, що пікові концентрації мають регулярний характер, пов'язаний із виробничими циклами, транспортною активністю в години пік та впливом статичних джерел викидів. Особливо виражені ці закономірності у таких урбанізованих регіонах, як Острава, Краків, Ченстохова, а також західні промислові регіони Німеччини.

На основі відкритих даних було проведено ряд експериментів з вейвлет перетвореннями для даних із різних міст Європи, індустріальних центрів та великих міст і агломерацій, зокрема Кракова та його приміських регіонів Глівіце та ін., так званої розширеної агломерації. Були досліджені коливання основних забруднювачів повітря, які характерні для даної зони через індустріальне навантаження, великий рух транспорту тощо.

На рис. 1 представлено приклад графічної візуалізації часового ряду концентрацій оксиду азоту (NO_2), створений за допомогою засобів візуалізації бібліотеки Matplotlib. Побудований графік наочно демонструє характер флуктуацій у реальних вимірюваннях, що містять як регулярні патерни, так і випадкові піки, й потребують подальшого згладжування та денойзингу.

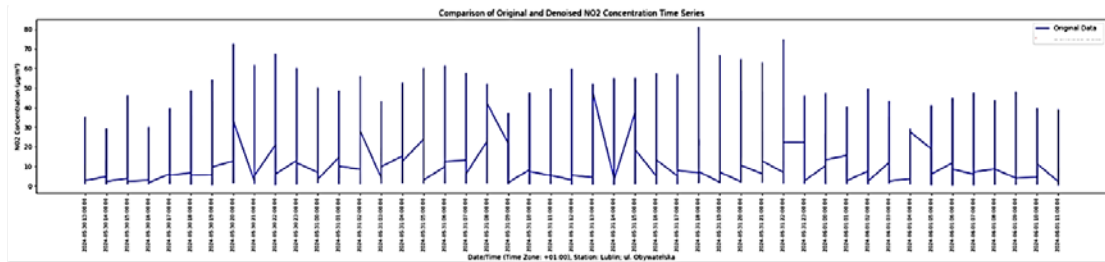


Рис. 1. Графічне представлення необроблених даних про концентрацію NO_2

Для виявлення трендів зміни концентрацій забруднюючих речовин у часі було застосовано вейвлет-перетворення до стохастичних часових рядів, з метою локалізації помилкових значень, згладжування флуктуацій та побудови узагальнених траєкторій зміни концентрації. Такий підхід дозволяє ефективно відфільтровувати шум і виокремлювати структурні закономірності в нестационарних даних.

Первинні дані, отримані з відкритих джерел, пройшли етап попередньої фільтрації, під час якого були видалені явно аномальні значення – зокрема, непропорційно високі або випадкові нульові спостереження, які не мають фізичного обґрунтування. Після очищення було виконано дискретне вейвлет-перетворення, що дозволило виокремити основні трендові компоненти, які на графіку відображено червоним кольором. На рис. 2 представлено приклад трансформованого часового ряду для концентрації оксиду азоту (NO_2) у повітряному басейні міста Краків, отриманого після застосування вейвлет-аналізу.

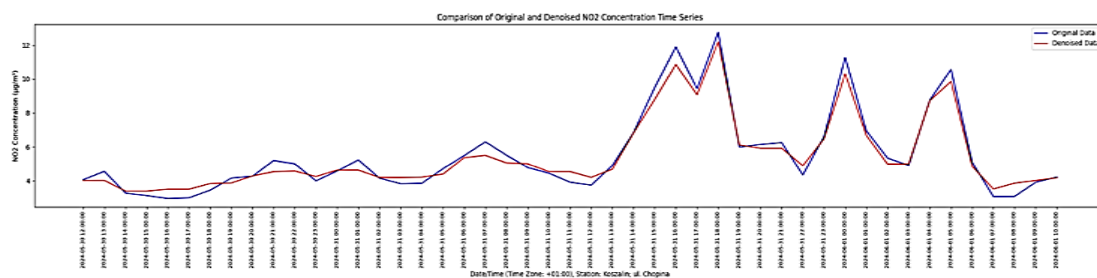


Рис. 2. Часовий ряд даних про концентрацію NO_2 , очищений за допомогою дискретного вейвлет-перетворення

Вейвлет-перетворення широко використовуються для очищення часових рядів від шуму та виявлення локальних закономірностей. Завдяки локалізації у просторі та частоті, вейвлети дозволяють деталізовано аналізувати сигнали, у яких поєднуються регулярні та стохастичні коливання. В основі лежить масштабованість і здатність до зсуву, що дає змогу виявляти особливості сигналу на різних ділянках інтервалу спостереження. У роботі застосовано вейвлети Добеші першого типу для багаторівневого дискретного розкладу на базі бібліотеки PyWavelets.

Було підтверджено, що коливання в концентраціях мають часовий зсув, який залежить від метеофакторів – вітру, температури тощо. Попереднє очищення даних здійснювалося за допомогою функції «м'якого» порогового відсікання: замість видалення аномалій значення адаптувалися, щоб зберегти загальну структуру сигналу. Повторне вейвлет-перетворення дозволило прибрати локальні піки, які не несуть інформаційного навантаження, але заважають оцінці довготривалих трендів.

Проведений кратномасштабний аналіз допоміг виявити часові закономірності змін, недоступні при звичайному спектральному аналізі. На прикладі Кракова проаналізовано рівні фенолу, бензолу та сірчистого газу. Дані фіксувалися з інтервалом 10 хвилин протягом кількох діб. Після згладжування часових рядів виявлено повторювані ранкові піки концентрації. Додаткове порівняння зі станцією в місті Глівіце (відстань 114 км) підтвердило просторову кореляцію трендів, що свідчить про регіональну узгодженість екологічної динаміки. На рис. 3 показано типовий очищений від шуму ряд для центральної частини Кракова (вул. Bulwarowa), в якості прикладу використано згладжений ряд концентрації бензолу C_6H_6 ,

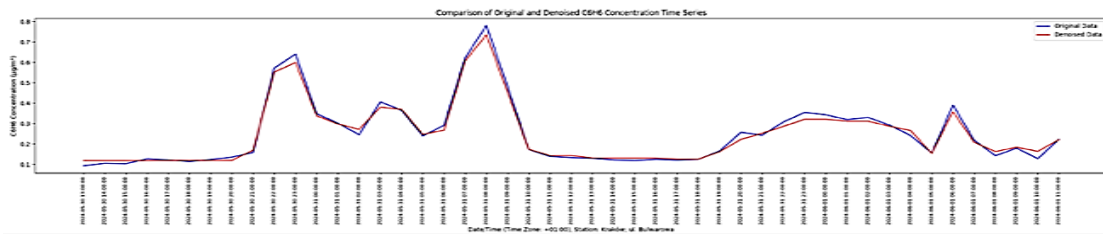


Рис. 3. Часовий ряд концентрацій C_6H_6 від станції Kraków; ul. Bulwarowa після дискретних вейвлет-перетворень

Цей підхід дозволяє дослідити механізми поширення забруднення у просторі та часі, враховуючи як локальні джерела, так і фонові регіональні впливи, відкриваючи перспективи для точнішого прогнозування стану повітря.

На рис. 4 зображено часовий ряд концентрації бензолу, виміряний на станції Gliwice; ul. Mewy у той самий час, коли проводились вимірювання у Кракові.

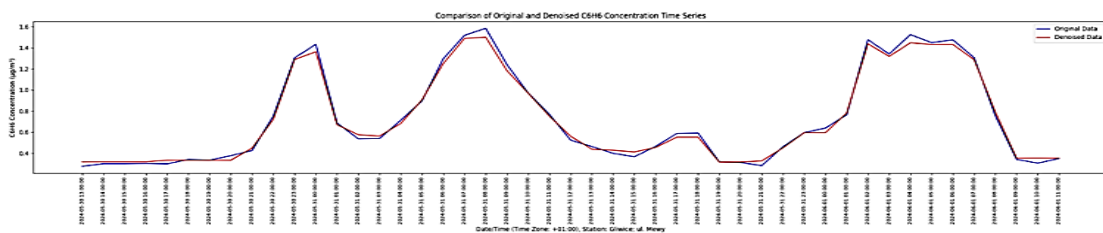


Рис. 4. Часовий ряд концентрацій C_6H_6 від станції Gliwice; ul. Mewy після дискретних вейвлет-перетворень.

Незважаючи на різницю в масштабах населених пунктів, характер зміни концентрації бензолу у місті Глівіце загалом відповідає тенденціям, зафіксованим у Кракові, хоча й демонструє певне часове запізнення. Окрім впливу локальних факторів, таких як години пік і підвищене транспортне навантаження, на якість повітря в Глівіце можуть впливати викиди з Кракова, які за відповідних метеоумов транспортуються повітряними потоками до віддалених районів протягом кількох годин.

Застосування високоточних вейвлет-перетворень дозволяє виявити подібні локальні геопросторові аномалії (Mallat, 2008), і, відповідно, сформулювати висновки про фізичні механізми, що визначають поведінку концентрацій шкідливих речовин у повітрі на міжміському рівні. У разі необхідності фокусування на глобальних закономірностях – наприклад, при побудові довгострокових трендів – точність перетворення може бути навмисне знижена. Це дає змогу приглушити одиничні сплески, зменшити вплив випадкових подій, помилкових значень та залишкового шуму, що забезпечує стабільність прогнозової моделі (Percival & Walden, 2000).

Підсумовуючи результати моделювання, можна стверджувати, що вейвлети є ефективним інструментом аналізу екологічної ситуації як на локальному рівні, так і в регіональному контексті. У подальших дослідженнях планується вдосконалити модель з урахуванням турбулентності атмосфери, впливу другорядних джерел викидів, а також екстремальних або непередбачуваних подій, таких як вибухи, техногенні аварії чи стихійні явища, що має на меті підвищити точність прогнозування та адаптивність заходів із захисту навколишнього середовища до умов конкретної ситуації.

Із зростанням обсягів екологічної інформації, що надходить з автоматизованих сенсорних систем у режимі реального часу, виникає потреба у використанні більш гнучких, адаптивних та обчислювально ефективних методів аналізу (Addison, 2017). У цьому контексті перспективним є поєднання вейвлет-декомпозиції (I Daubechies, 1992) з інструментами машинного та глибокого навчання.

Довгострокові пам'яттєві нейронні мережі (LSTM) широко застосовуються у задачах прогнозування часових рядів, особливо в умовах нерівномірної структури даних і наявності стохастичних коливань. У контексті моделювання екологічних даних ми апробували поєднання багаторівневої вейвлет-декомпозиції (на основі вейвлетів типу db4 та sym5) з подальшим використанням

апроксимуючих коефіцієнтів як вхідних ознак для LSTM. Деталізуючі коефіцієнти водночас застосовувались як вхідні параметри для ансамблевих моделей типу Random Forest і XGBoost. Такий підхід забезпечує ефективне виявлення регулярних шаблонів, зумовлених ритмікою добових коливань, погодними умовами або транспортною активністю. Застосування зваженого об'єднання моделей (ensemble averaging) дозволяє досягти більшої узагальнюваності результату, що особливо важливо у випадках часткової втрати або зашумленості вхідних даних.

У перспективі планується інтеграція scattering-мереж на основі вейвлетів – структур, які дозволяють формувати стабільні ознаки без потреби у глибокому навчанні на великих масивах даних (Shumway & Stoffer, 2017). Окрему увагу буде приділено автоматизованому вибору параметрів декомпозиції за допомогою засобів оптимізації, зокрема бібліотеки Optuna, що дозволить гнучко адаптувати модель до конкретних задач прогнозування (Akiba et al., 2019).

Таким чином, результати дослідження підтверджують доцільність застосування комплексного підходу, який об'єднує можливості вейвлет-аналізу з сучасними методами штучного інтелекту для екологічного прогнозування, підтримки прийняття рішень і розробки превентивних заходів у сфері охорони довкілля.

Висновки

У ході дослідження було проаналізовано ключові фактори, що впливають на якість атмосферного повітря, зокрема концентрації таких поширених забруднювачів, як оксиди азоту та бензол. Встановлено, що найбільш значущими чинниками, які визначають зміну рівнів цих речовин, є інтенсивність транспортного потоку, а також метеорологічні умови – температура, напрям і швидкість вітру. Особливо чітко виявляється вплив цих факторів у години підвищеного навантаження (так звані години пік) та під час активного функціонування промислових підприємств, коли концентрації забруднювачів значно зростають.

Виявлена потреба в ефективній обробці великих масивів даних, що надходять із численних екологічних станцій. Для реалізації моделювання було обрано мову програмування Python, зокрема бібліотеки Pandas, Matplotlib та PyWavelets, які забезпечують широкий функціонал для роботи з часовими рядами. Зокрема, Pandas використовувалась для структурування, очищення та попереднього аналізу даних; Matplotlib – для візуалізації динаміки концентрацій; PyWavelets – для реалізації дискретних вейвлет-перетворень та багаторівневої декомпозиції сигналів.

У межах дослідження було проведено серію комп'ютерних експериментів із моделювання змін концентрацій забруднювачів, що дало змогу оцінити ефективність застосування вейвлет-аналізу для задач денойзингу та виокремлення трендів. Результати показали, що дискретне вейвлет-перетворення дозволяє з високою точністю виявляти просторово-часові закономірності у зміні екологічних параметрів, що є надзвичайно важливим для оперативного реагування на зміни якості повітря та прийняття обґрунтованих управлінських рішень.

У процесі дослідження було враховано можливість автоматизованого підбору оптимального рівня декомпозиції сигналу. Зокрема, перспективним є використання підходів, заснованих на аналізі енергетичних характеристик вейвлет-коефіцієнтів на кожному рівні розкладу. Такі методи дозволяють уникнути як надмірної фільтрації, що може спричинити втрату інформативних компонентів сигналу, так і недостатнього згладжування, коли залишковий шум спотворює результати аналізу. Застосування подібних критеріїв у майбутньому дасть змогу досягти балансу між точністю виявлення закономірностей і стабільністю обробки вхідних даних.

Попри високу ефективність запропонованого підходу, варто відзначити низку обмежень, які потребують подальшого опрацювання. Зокрема, застосована модель поки що не враховує складну топографію території, сезонні особливості метеорологічних процесів, а також не охоплює деякі специфічні типи забруднювачів, такі як тверді частинки PM_{2.5} та вуглеводні складної структури. Перспективним напрямом подальших досліджень є розширення моделі для інтеграції багатофакторного аналізу з геопросторовими даними, що дозволить застосовувати запропоновану методологію в зада-

чах регіонального планування, екологічного зонування та моделювання впливу зміни клімату на якість повітря. Іншим важливим кроком може стати розробка адаптивних моделей прогнозування для конкретних типів міських агломерацій або промислових зон із урахуванням індивідуальних профілів забруднення.

СПИСОК ЛІТЕРАТУРИ

- Addison, P. S. (2017). *The illustrated wavelet transform handbook: Introductory theory and applications in science, engineering, medicine and finance*. CRC Press.
- Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 2623–2631). <https://doi.org/10.1145/3292500.3330701>
- Daubechies, I. (1992). *Ten lectures on wavelets*. Society for Industrial and Applied Mathematics.
- Ekstrom, P. A., & Hales, J. M. (2000). A wavelet-based approach for atmospheric pollution modeling: Algorithm development. *Monthly Weather Review*, 128 (9), 3169–3186. [https://doi.org/10.1175/1520-0493\(2000\)128<3169:ODDWFS>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<3169:ODDWFS>2.0.CO;2)
- European Environment Agency. (2023, November). Up-to-date air quality data. <https://www.eea.europa.eu/data-and-maps/explore-interactive-maps/up-to-date-air-quality-data>
- European Environment Agency. (2024). Data source for online air pollution data from EEA. <https://discomap.eea.europa.eu/map/fme/AirQualityUTDExport.htm>
- Hadjibiros, K. (2013). *Ecology and applied environmental science*. OAPEN Foundation. <https://doi.org/10.1201/b14609>
- IBM SPSS Statistics. (2021, December 7). Dispersion analysis with repeated measurements. <https://www.ibm.com/docs/ru/spss-statistics/beta?topic=statistics-repeated-measures-anova>
- Mallat, S. (2008). *A wavelet tour of signal processing: The sparse way* (3rd ed.). Academic Press.
- Percival, D. B., & Walden, A. T. (2000). *Wavelet methods for time series analysis*. <https://doi.org/10.1017/CBO9780511841040>
- Pytorch Wavelets. (2020). *Pytorch Wavelets 0.1.1 documentation*. <https://pytorch-wavelets.readthedocs.io/en/latest/readme.html>
- Serov, V. (2017). *Fourier series, Fourier transform and their applications to mathematical physics*. Springer Nature. <https://doi.org/10.1007/978-3-319-65262-7>
- Shumway, R. H., & Stoffer, D. S. (2017). *Time series analysis and its applications: With R examples* (4th ed.). Springer.
- Tuzenko, O., & Sidun, N. (2023). Mathematical modeling of ecological observations data using time series analysis methods. In *2023 IEEE 18th International Conference on Computer Science and Information Technologies (CSIT)* (pp. 1–4). IEEE. <https://doi.org/10.1109/CSIT61576.2023.10324166>
- Wasilewski, F. (n.d.). *PyWavelets – Wavelet transforms in Python*. PyWavelets Documentation. <https://pywavelets.readthedocs.io/en/latest/>

EMPLOYING DISCRETE WAVELET TRANSFORM METHODS AND PYTHON LIBRARIES TO DERIVE MATHEMATICAL MODELS OF ENVIRONMENTAL DATA

Olga Tuzenko¹, Nataliia Sidun², Yehor Volobuyev³

^{1, 2, 3} Pryazovskyi State Technical University,

Department of Informatics, Dnipro, Ukraine

E-Mail: tuzenkooa@gmail.com, ORCID 0000-0002-4920-9417

E-Mail: natalya.sidun@gmail.com, ORCID 0009-0001-8343-1273

E-Mail: egorwww7@gmail.com, ORCID 0009-0006-5221-8134

© Olga Tuzenko, Nataliia Sidun, Yehor Volobuyev, 2025

The article addresses the urgent problem of computer modeling of large-scale environmental monitoring datasets using discrete wavelet transforms. The research object consists of time series of harmful pollutant concentrations in the atmosphere, including nitrogen oxides, benzene, and sulfur

dioxide, collected from automated stations in Central and Eastern Europe. The input data are characterized by high stochasticity, noise, missing values, and temporal shifts, which significantly complicate the extraction of trends and patterns required for forecasting. An efficient processing approach is proposed, combining multilevel wavelet decomposition with adaptive filtering based on soft-thresholding and an automatic decomposition-level selection mechanism grounded in energy-based coefficient analysis.

For modeling, the Python ecosystem was employed: PyWavelets for wavelet modeling, Matplotlib for dynamics visualization, and Pandas for structuring large input datasets. A series of computational experiments was conducted on data from Polish, Czech, and German cities, confirming the effectiveness of the proposed method. It is demonstrated that the integration of Daubechies (db4) wavelet transforms with ensemble learning methods (Random Forest, XGBoost) and LSTM neural networks enables highly accurate forecasting, even under non-stationary emissions and meteorological fluctuations. The proposed approach can serve as the basis for developing next-generation adaptive environmental monitoring systems, particularly in industrially intensive urban areas.

Keywords – environment, air quality, pollutant concentration, time series, wavelet function, wavelet transform, mathematical modeling, Python libraries.