

УДК 004.9+504.3.054

Д. О. Шмундяк; В. Б. Мокін, д-р техн. наук, проф.**СИСТЕМНИЙ АНАЛІЗ СТАНУ ПРИРОДНИХ СЕРЕДОВИЩ З
УРАХУВАННЯМ АНОМАЛІЙ**

У статті запропоновано методологію системного аналізу стану природних середовищ з урахуванням аномалій, яка спрямована на покращення процесу аналізу та прогнозування різних природних середовищ для прийняття подальших управлінських кроків щодо контролю та оптимізації їх стану. Протягом останніх десятиліть спостерігається суттєва зміна клімату у світі в цілому та в Україні зокрема. Зміни у різних природних середовищах впливають на інші складові системи. В той же час, антропогенний вплив генерує додаткове навантаження на стан природних середовищ. Усі складові взаємопов'язані та є частиною однієї складної системи і тому необхідно їх розглядати з точки зору системного аналізу складних системи, враховуючи при цьому аномалії як природного, так і антропогенного походження. Відповідно, є потреба в удосконаленні методології системного аналізу такого типу систем, що дозволить приймати правильні рішення щодо управління станом природних середовищ.

У роботі охарактеризовано складну систему, якою є система управління станом довкілля в заданому регіоні. Наведено декомпозицію складної системи за моделлю «DPSIR» та декількома іншими видами декомпозиції. Виділено основні види невизначеності цієї складної системи та її елементів, зокрема й ті, які пов'язані з наявністю аномалій та пікових показників стану природного середовища. Запропоновано алгоритм системного аналізу з послідовно-паралельним розкриття усіх цих видів невизначеності, з використанням як відомих підходів та методів, так і тих, що розроблені авторами статті.

Наведено приклад застосування розробленої методології системного аналізу з урахуванням аномалій щодо різних природних середовищ: стану атмосферного повітря зони «Вінницька» (Вінницької області), стану масивів вод басейну р. Південний Буг та чутливості населення до алергенів (рослин та грибів).

Ключові слова: системний аналіз, штучний інтелект, інформаційні технології, природне середовище, прогнозування часових рядів, якість атмосферного повітря, стан вод, аномалії, сезонність, машинне навчання, інтелектуальний аналіз даних, моделювання, метод аналізу ієрархій.

Вступ

Зміна клімату все більше стає відчутною і це проявляється в появі природних аномалій щодо температурного режиму, опадів. Неочікувані циклони та антициклони спричиняють перенесення як вологи в атмосфері, так і забруднювальних речовин. У свою чергу, це спричиняє непередбачувані зміни екологічного стану довкілля. Крім того, в Україні в останні роки спостерігаються зміни у промисловому і сільськогосподарському секторі, пов'язані з релокацією виробництв. А це спричиняє появу викидів, скидів та іншого антропогенного впливу, який не є характерними для певних місць у певний час. Все більше з'являється таких аномалій. Хоча іноді це вже не аномалії, а – лише звичні пікові (або екстремальні) явища, наприклад, як перенесення пилу Сахари весною у повітрі України. Все це потребує удосконалення методології системного аналізу стану природних середовищ з урахуванням таких аномалій природного та антропогенного походження.

Сучасні інтелектуальні технології передбачення рядів часових даних дають можливість більш гнучко враховувати аномалії, але для цього їх треба спочатку знайти, оцінити характер їх впливу, внести зміни в параметри моделей для опису інших складових цих рядів.

Загалом, системний аналіз застосовується, як правило, в задачах оптимізації та управління складними системами. В цій сфері такою складною системою є система управління станом природних середовищ. В Європі популярною є декомпозиція цієї системи на основі моделі «DPSIR» [1]. Відповідно до методології системного аналізу треба здійснити декомпозицію складної системи, визначити види невизначеності та запропонувати методи їх розкриття [2, 3]. Після цього варто знайти дані, застосувати до них ці методи і тоді можна

ідентифікувати інтелектуальні моделі, здійснити за ними передбачення, провести інтелектуальний аналіз даних, наприклад за алгоритмом з роботи [4], та розробити рекомендації для управління станом природного середовища, з урахуванням виявлених аномалій.

Багато вчених працюють у цьому напрямку, але відсутня дійсно комплексна методологія 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.

Метою статті є розроблення методології системного аналізу стану природних середовищ з урахуванням аномалій для підвищення точності прогнозування цього стану та підвищення ефективності системи управління станом довкілля в заданому регіоні.

Методологія системного аналізу стану природних середовищ

Відповідно до типової методології системного аналізу складних систем [2, 3] та узагальненого алгоритму машинного навчання інтелектуальних моделей та інтелектуального аналізу даних [4] пропонується здійснювати дослідження у такі етапи:

1. Постановка задачі. На цьому етапі необхідно чітко сформулювати, що є складною системою та провести її декомпозицію на підсистеми. Кожна підсистема повинна бути проаналізована індивідуально та проаналізовані її зв'язки і відносини з іншими складовими. Результатом такого аналізу повинні бути виділені різні види невизначеності, які варто розкрити на наступних кроках системного аналізу;

2. Виділення об'єкта з навколишнього середовища. Етап передбачає розкриття невизначеності, отриманої на попередньому кроці. Крім того, важливо знайти різноманітні дані про систему, яка аналізується, та провести їх розвідувальний аналіз, що дасть корисні знання для подальшого моделювання процесів у цій складній системі та її підсистемах;

3. Моделювання процесів. Використовуючи результат розвідувального аналізу, необхідно вибрати тип, архітектуру та параметри оптимальної інтелектуальної моделі. Ідентифікація цих оптимальних параметрів та моделей може вимагати розроблення нових або удосконалення наявних методів аналізу відповідної складної системи;

4. Аналіз з використанням побудованої інтелектуальної моделі. На цьому кроці аналізується інформація, отримана після безпосереднього використання інтелектуальної моделі. З цими новими знаннями корисним може бути повторення попередніх етапів системного аналізу. Очікуваним результатом цього етапу є формування різних висновків про складну систему, яка аналізується, та її поведінку;

5. Оптимізація системи. Маючи чітке розуміння стану складної системи та даних цієї системи, можна запропонувати шляхи її удосконалення, покращення її стану, тощо.

Декомпозиція системи управління станом природних середовищ з урахуванням аномалій

В Європейському союзі існує та активно застосовується модель «DPSIR» [1]—модель, яка використовується для розв'язання різних задач, пов'язаних з управлінням станом природних середовищ (стан атмосферного повітря, екологічний стан водних ресурсів, тощо). Структурно ця модель складається з п'яти взаємопов'язаних компонентів (рис. 1).

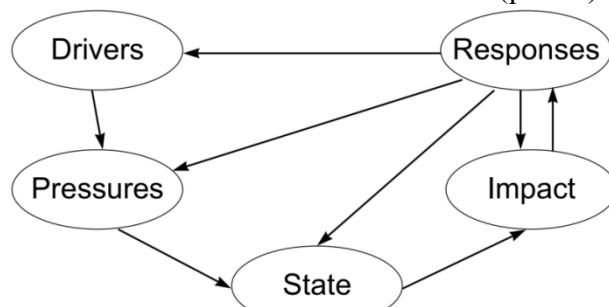


Рис. 1. Структурна схема моделі «DPSIR» [1]

Як видно з рис. 1, прийняття рішень (англ. «Responses» — «R») стосується таких дій [1]:

- вибору місць розташування та вибору складу продуктивних сил (заводів, електростанцій, сільськогосподарських полів тощо) (англ. «Drivers» — «D»);
- 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 довкілля (англ. «Pressures» — «P»);
- вибору місць розташування елементів та параметрів систем моніторингу стану довкілля: станцій спостереження за якістю або екологічним станом води, станцій моніторингу атмосферного повітря, тощо (англ. «State» — «S»);
- 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 довкілля (англ. «Impact» — «I»).

Пропонується застосувати такі 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.

Класифікація видів невизначеності складових складної системи

Виділимо види невизначеності, які мають місце у складній системі управління станом природних середовищ регіону з урахуванням, передусім, впливу аномалій чи пікових значень показників або параметрів.

Проаналізуємо види невизначеності для складових за декомпозицією 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 моніторингу (невизначеність вибору кількості і місць розташування наявних 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 доданням 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, передусім – інформаційну) – 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 доданням 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} .

Методи та підходи розкриття видів невизначеності складових складної системи

Систематизуємо у таблиці 1 усі описані вище види невизначеності та запропонуємо підходи та методи їх розкриття.

Методи розкриття невизначеності 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
U_{I5}	Інформаційні запити щодо системи моніторингу
U_{T3}, U_{T5}	Використання директив ЄС та інтегрованого з ним екологічного законодавства України
U_{T1}	Використання експертних методів або методу аналізу ієрархій [8]
U_{S1}	Метод пошуку можливих джерел забруднення [6, 7] та ін.
U_{S2}	Метод ідентифікації локальних у часі аномалій [12] та ін.
U_{M1}	Методи машинного навчання інтелектуальних моделей прогнозування [4, 5]
U_{P1}, U_{P2}, U_{T6}	Метод ідентифікації параметрів сезонності та аномалій [9 – 11] та ін.
U_{ST1}, U_{ST2}	Байєсівське моделювання [13, 14] 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., приклад у [15, додаток 6]

Алгоритм застосування запропонованої методології системного аналізу

Узагальнюючи наведене вище, запропонуємо алгоритм методології системного аналізу стану природного середовища з послідовним розкриттям усіх видів невизначеності.

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

2. Збирання інформації щодо «DPSR»: зробити запити для отримання інформації про місце 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 довкілля достатньо для формування висновків та управлінських заходів, використовуючи прийоми розвідувального аналізу [18]. 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 довкілля про пости цієї мережі, їх показники, регулярність вимірювань та метрологічні аспекти (це розкриє U_{I5}) та зібрати додаткову інформацію про стан довкілля від суб'єктів державної системи контролю, які здійснюють нерегулярні спостереження (держкоінспекції, органи МОЗ тощо), від мереж громадського моніторингу довкілля, 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». Сформувані системи критеріїв для вибору параметрів місць постів спостережень за станом довкілля, з урахуванням екстремальних значень показників про усіх складові «DPSIR». Для їх систематизації та визначення ваги використати методи аналізу ієрархій, наприклад як це було зроблено у роботі для стану атмосферного повітря у Вінницькій області [8]. Це дозволить визначити де варто розмістити нові станції моніторингу чи перемістити наявні, 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 довкілля, 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 метод, розроблений авторами цієї статті, на основі аналізу аномалій у наявних даних моніторингу стану природного середовища [6, 7]. Це дозволить розкрити невизначеність (U_{S1}). Цей метод дозволяє знайти незареєстровані джерела (чи виявити який обсяг було занижено наявним 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 [5]. Також, дані можуть містити локальні у часі аномалії, тому можна застосувати розроблений одним з авторів метод пошуку локальних аномалій на основі декомпозиції ряду з роботи [12], 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 [19]. Це розкриває невизначеність U_{M1} . При цьому, для ідентифікації параметрів сезонності можна використати як класичні підходи [9, 10], так і авторський метод ідентифікації параметрів гармонік та аномалій періодичного часового ряду, описаний у роботі [11], 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 одного з авторів на базі платформи Kaggle[14].

14. Аналіз даних «S» [17], 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} [15].

Графічно цей алгоритм можна зобразити блок-схемою (рис. 2).

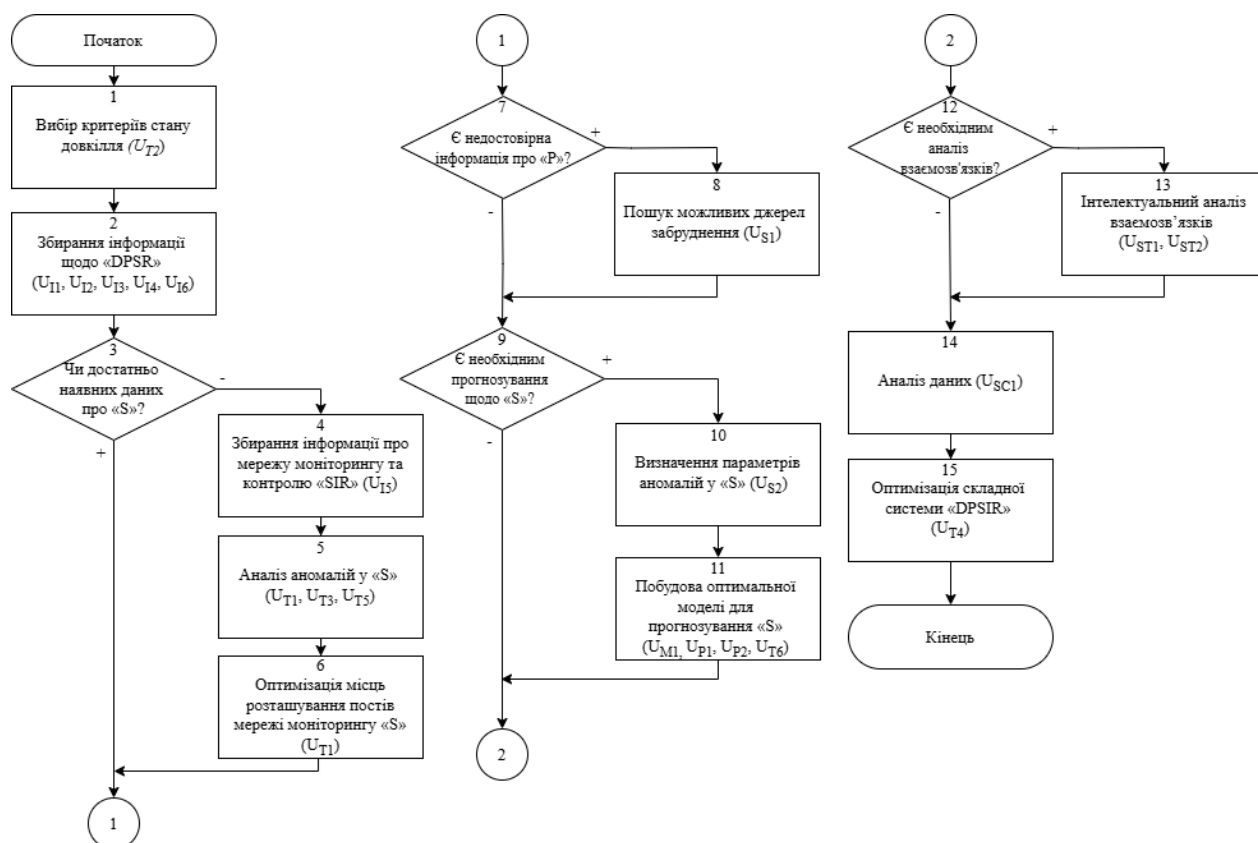


Рис.2. Блок-схема алгоритму системного аналізу стану природного середовища з урахуванням аномалій та з розкриттям основних видів невизначеності

Приклади застосування запропонованої методології системного аналізу

Запропонована методологія системного аналізу була випробувана за активної участі авторів статті в науково-дослідній лабораторії екологічних досліджень та екологічного моніторингу ВНТУ для удосконалення системи управління станом атмосферного повітря зони «Вінницька», межі якої збігаються з межами Вінницької області. Аналізувались 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). Аналіз показав, що 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 за станом атмосферного повітря, з урахуванням екстремальних значень показників. З використанням методу аналізу ієрархій визначено оптимальні місця для розміщення постів моніторингу [8] (крок 6). 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 2022 – 2026 роки [15].

В очікуванні оптимізації мережі державного моніторингу атмосферного повітря і нових рядів даних автори статті аналізували цей стан з використанням даних мережі громадського моніторингу. Зокрема, у ВНТУ 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 [16]. За цими даними побудовано інтелектуальні моделі та вибрано оптимальні за різними критеріями (кроки 9 – 11) (див. роботи [11, 12]), 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 [20]:

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

Результат прогнозування на валідаційних даних наведено на рис. 3.

	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

Рис. 3. 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)

Як видно на рис. 3, модель з використанням власних налаштувань сезонності (варіант 3) дає кращу точність, порівняно з використанням вбудованих параметрів сезонності (варіант 2). Крім того, моделі з використанням методу пошуку локальних аномалій на основі декомпозиції ряду (варіанти 5 та 7) дозволяють покращити точність за усіма метриками

порівняно 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 – «Defaultseasonality»), вдалось зменшити похибки прогнозування за метрикою MAE на 30,2%, а за метрикою RMSE – на 21,7%.

Приклади застосування кроків 7 і 8 описані у роботах [6] (для річок) та [7] (для атмосферного повітря).

Крім того, ця методологія, а саме кроки 9 – 11, була використана під час досліджень членів басейнової ради Південного Бугу та її комісії щодо аналізу та прогнозування якості поверхневих вод річки Південний Буг [7]. Це дозволило розкрити невизначеність U_{S1} , пов'язану з відсутністю достовірної інформації про викиди забруднюючих речовин.

Також, ця методологія (кроки 12 і 13) була використана для розкриття структурної невизначеності складової «I» під час проведення алергологічного дослідження, яке включало детальні симптоми пацієнтів та результати молекулярних тестів населення більшості областей України за 300 видами алергенів (пилки рослин, грибів). Побудовано байєсівські моделі та, за участі вчених з Вінницького національного медичного університету та Національного медичного університету, авторам цієї статті вдалось знайти взаємозв'язки між чутливістю до певних груп грибкових алергенів [13].

Отже, запропонована методологія довела свою ефективність для різних видів природних середовищ у регіонах України.

Висновки

У статті дістала подальший розвиток методологія системного аналізу стану природних середовищ, з урахуванням аномалій, основана на використанні класичних та авторських методів для розкриття основних видів невизначеності, які мають місце в системах управління станом природного середовища різного типу. Розроблено узагальнений алгоритм реалізації цієї методології з 15 етапів. При розкритті невизначеності основна увага приділяється виявленню аномалій (чи екстремальних значень) показників стану природного середовища, оцінюванні їх параметрів та врахуванню впливу.

Наведено приклад прогнозування стану атмосферного повітря у смт Турбів Вінницькій області та доведено, що використання цієї методології та авторських методів розкриття параметричної і модельної невизначеності дозволяє підвищити точність прогнозування на 30,2% за метрикою MAE та на 21,7% за метрикою RMSE.

Крім того, охарактеризовано приклад застосування запропонованої методології для системного аналізу та оптимізації системи управління та моніторингу стану атмосферного повітря зони «Вінницька» (Вінницька область), на основі чого була розроблена, за участі авторів, Програма державного моніторингу в галузі охорони атмосферного повітря зони «Вінницька» на 2022 – 2026 роки, вже затверджена, відповідно до чинного законодавства. Також вона була використана членами басейнової ради Південного Бугу та її комісії для аналізу і прогнозування стану поверхневих вод ділянки річки Південний Буг. Крім того, окремі етапи методології була застосована для інтелектуального аналізу сенсibiliзації населення України до різних грибкових алергенів, в результаті чого вдалось виявити нові закономірності щодо того, наскільки зростає ймовірність того, що чутливість до певних алергенів може бути ознакою чутливості до інших.

Таким чином, було доведено ефективність застосування усіх етапів методології системного аналізу стану природних середовищ різного типу в Україні, з урахуванням аномалій. Це дозволяє підвищити ефективність системи моніторингу та контролю і допомагає приймати необхідні управлінські рішення щодо поліпшення чи стабілізації доброго стану цих природних середовищ.

СПИСОК ЛІТЕРАТУРИ

1. Environmental indicators: Typology and overview. *EdithSmeetsandRobWeterings*. 1999. URL: <https://www.eea.europa.eu/publications/TEC25>.
2. Згуровський М. З., Панкратова Н. Д. Основи системного аналізу. К.: Видавнича група ВНУ, 2007. 546 с.
3. Мокін Б. І., Мокін О. Б., Мокін В. Б. Методологія та організація наукових досліджень : підручник – вид. 3-є, змін. та доп. Вінниця: ВНТУ, 2023. 230 с.
4. Мокін В. Б., Драгований М. В. Наука про дані: машинне навчання та інтелектуальний аналіз даних : електронний навчальний посібник комбінованого (локального та мережевого) використання. Вінниця : ВНТУ, 2024. 258 с. URL: https://pdf.lib.vntu.edu.ua/books/2024/Mokin_2024_263.pdf.
5. Аналіз можливостей Python-бібліотек щодо виявлення аномальних даних у задачі прогнозування стану атмосферного повітря / Д. О. Шмундяк та ін. *Матеріали ЛІІ науково-технічної конференції підрозділів Вінницького національного технічного університету (НТКП ВНТУ–2023)* : збірник доповідей. – Вінниця : ВНТУ, 2023. С. 313 – 316.
6. Інформаційна технологія пошуку можливих джерел підвищеного забруднення річки з використанням моделі Prophet / В. Б. Мокін та ін, *Вісник ВПІ*. Верес. 2020. Вип. 4. С. 15 – 24. <https://doi.org/10.31649/1997-9266-2020-151-4-15-24>.
7. Шмундяк Д. О. Метод пошуку можливих джерел забруднення атмосферного повітря на основі аналізу аномалій. *Матеріали ХХІІІ Міжнародної науково-практичної конференції Математичне моделювання та інформаційно-комунікаційні технології для перемоги та відновлення*, Вінниця, 16-17 жовтня 2024 р.
8. Технологія проєктування мережі спостережень якості атмосферного повітря регіону на основі методу аналізу ієрархій / В. Б. Мокін та ін. *Наукові праці ВНТУ*. 2022. № 4. <https://doi.org/10.31649/2307-5376-2021-4-21-33>.
9. Шмундяк Д. О. Дослідження впливу параметрів моделі Prophet на точність прогнозування якості атмосферного повітря. *Матеріали ЛІІ науково-технічної конференції підрозділів Вінницького національного технічного університету (НТКП ВНТУ–2023)* : збірник доповідей. – Вінниця : ВНТУ, 2023. С. 269 – 272.
10. Огляд підходів до визначення порядку Фур'є у моделі Facebook Prophet для моделювання сезонної складової часового ряду / Д. О. Шмундяк та ін. *Матеріали ЛІІ науково-технічної конференції підрозділів Вінницького національного технічного університету (НТКП ВНТУ–2023)* : збірник доповідей. – Вінниця : ВНТУ, 2023. С. 239 – 241
11. Шмундяк Д. О., Мокін В. Б. Метод ідентифікації параметрів гармонік та аномалій періодичного часового ряду на основі адаптивної декомпозиції. *Вісник ВПІ*. Груд. 2023. Вип. 6. С. 46 – 56. DOI: <https://doi.org/10.31649/1997-9266-2023-171-6-46-56>.
12. Шмундяк Д. О., Копняк В. Є. Метод ідентифікації локальних аномалій значень показників стану довкілля з використанням декомпозиції на півхвилі. *Вісник ВПІ*. Лют. 2024. Вип. 1. С. 88 – 100. <https://doi.org/10.31649/1997-9266-2024-172-1-88-100>.
13. Bayesian analysis suggests independent development of sensitization to different fungal allergens / Rodinkova V. et al. *World Allergy Organization Journal*. 2024. Volume 17, Issue 5. P. 100908. ISSN 1939-4551. <https://doi.org/10.1016/j.waojou.2024.100908>.
14. Shmundiak D. Bayes for air monitoring data Kaggle Notebook. 2024. URL: <https://www.kaggle.com/code/dimashmundiak/bayes-for-air-monitoring-data>.
15. Про затвердження Програми державного моніторингу в галузі охорони атмосферного повітря на 2023–2027 роки для зони «Вінницька». 2021. URL: <https://mepr.gov.ua/wp-content/uploads/2023/01/Natsdopovid-2021-n.pdf>.
16. Mokin V., Shmundiak D., Kopniak V. Air Quality Monitoring from EcoCity” Kaggle Dataset. 2021. URL: <https://www.kaggle.com/datasets/vbmokin/air-quality-monitoring-from-ecocity>.
17. Mokin V., Shmundiak D. WQ SB river : Analysis & Forecasting” Kaggle Notebook. 2022. URL: <https://www.kaggle.com/code/vbmokin/wq-sb-river-analysis-forecasting>.
18. Мокін В. Б., Шмундяк Д. О., Копняк В. Є. Порівняльний розвідувальний аналіз даних про стан атмосферного повітря за даними мережі EcoCity з використанням бібліотеки Sweetviz. *Матеріали ЛІІІ науково-технічної конференції підрозділів Вінницького національного технічного університету (НТКП ВНТУ–2024)* : збірник доповідей. Вінниця : ВНТУ, 2024. С. 948 – 951.
19. Шмундяк Д. О. Порівняльний аналіз моделей прогнозування показника стану атмосферного повітря. *Матеріали ХVІІ Міжнародної конференція «Контроль і управління в складних системах» (КУСС-2024)*, Вінниця, 16-17 жовтня 2024. URL: <https://conferences.vntu.edu.ua/index.php/mccs/mccs2024/paper/view/22120>.
20. Shmundiak D. Comparing time series prediction models” Kaggle Notebook. 2024. URL: <https://www.kaggle.com/code/dimashmundiak/comparing-time-series-prediction-models>.

Стаття надійшла до редакції 19.12.2024.

Стаття пройшла рецензування 23.12.2024.

Шмундяк Дмитро Олександрович – аспірант кафедри системного аналізу та інформаційних технологій ВНТУ, e-mail: dimashmund@gmail.com.

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