

СИСТЕМНИЙ АНАЛІЗ І ТЕОРІЯ ПРИЙНЯТТЯ РІШЕНЬ

SYSTEM ANALYSIS AND DECISION-MAKING THEORY

DOI: 10.20998/2079-0023.2025.01.01

УДК 628.17

О. С. МЕЛЬНИКОВ, кандидат економічних наук (PhD), доцент, Національний технічний університет «Харківський політехнічний інститут», доцент кафедри системного аналізу та інформаційно-аналітичних технологій; м. Харків, Україна, e-mail: Oleg.Melnikov@khpі.edu.ua, ORCID: <https://orcid.org/0000-0002-2409-4983>

Ю. І. ДОРОФЄЄВ, доктор технічних наук, професор, Національний технічний університет «Харківський політехнічний інститут», професор кафедри системного аналізу та інформаційно-аналітичних технологій; м. Харків, Україна, e-mail: Yuri.Dorofiev@khpі.edu.ua, ORCID: <https://orcid.org/0000-0002-7964-1286>

Н. А. МАРЧЕНКО, кандидат технічних наук (PhD), доцент, Національний технічний університет «Харківський політехнічний інститут», заступниця директора навчально-наукового інституту комп'ютерних наук та інформаційних технологій; м. Харків, Україна, e-mail: Natalia.Marchenko@khpі.edu.ua, ORCID: <https://orcid.org/0000-0001-9889-3713>

СТАТИСТИЧНИЙ ПІДХІД ДО ВИЯВЛЕННЯ АНОМАЛІЙ У ВОДОРОЗПОДІЛЬНИХ МЕРЕЖАХ

Робота присвячена вирішенню задачі розробки автоматизованої системи виявлення аномалій у водорозподільних мережах, основними причинами яких є фонові витoki та розриви труб. Для розв'язання задачі використано статистичний підхід, який полягає у перевірці нульової гіпотези про відповідність показань датчиків тиску та/або витрат води, які надходять в реальному часі, стандартним умовам функціонування мережі. В роботі запропоновано триетапну схему виявлення аномалій, яка включає: статистичне профілювання датчиків мережі; калібрацію системи для досягнення бажаного співвідношення між ризиками хибних тривог та пропуску існуючих аномалій; визначення правил формування висновків про наявність аномалій. Розроблено методику статистичного профілювання та калібрації системи на базі імітаційного моделювання в програмному середовищі EPANET за допомогою програмного інтерфейсу WNTR на мові Python. У процесі такого моделювання на основі збурень попиту на воду отримується розподіл значень показань тиску в датчиках мережі. В якості прикладу досліджено модель мережі водопостачання L-Town, яку було використано в змаганні методів виявлення та ізоляції витоків BattLeDIM. Досліджено чутливість результатів виявлення аномалій щодо діапазону значень датчиків, які вважаються нормальними, а також кількості датчиків, задіяних у процедурі виявлення аномалій. Проаналізовано залежність кількості датчиків із аномальними показаннями від розміру витoku. Встановлено, що не існує комбінації параметрів запропонованої системи виявлення аномалій, які забезпечують оптимальні результати для всіх можливих розмірів витоків одночасно, в результаті чого запропоновано виконувати калібрацію системи за допомогою людино-машинної процедури.

Ключові слова: водорозподільна мережа, аномалія, виток, імітаційне моделювання, EPANET, перевірка статистичних гіпотез.

Вступ. Водорозподільні мережі (ВРМ) є вразливими до інфраструктурних збоїв, які можуть призвести до значних втрат води. Згідно з оцінкою, наведеною в [1], у світі в цілому втрачається понад 30 % питної води, що у вартісному еквіваленті становить 39 млрд. доларів США на рік. Значна частина цих втрат пов'язана з фоновими витокami та проривами труб, які можуть статися в будь-якій точці ВРМ. Фонові витoki, як правило, важко виявити через їх порівняно малий розмір на ранніх стадіях. Розриви труб виявити легше, оскільки вони одразу призводять до великих втрат води і можуть проявлятися на поверхні. У будь-якому випадку наявність витоків призводить до аномалій, тобто таких станів або спостережень, які виходять за межі очікуваної поведінки мереж водопостачання. Раннє виявлення та локалізація аномалій у ВРМ є дуже важливою задачею, оскільки це скорочує час, необхідний для усунення їх негативних наслідків та зменшує

ризик подальшої деградації інфраструктури, забруднення води тощо.

Аналіз літератури. Завдання виявлення витоків у системах водопостачання привертає велику увагу з боку як практиків, так і дослідників протягом останніх років [2]. Процес діагностики витоків можна розділити на три етапи:

- виявлення аномалії, тобто встановлення суттєвого відхилення режиму функціонування мережі від операційних норм;
- визначення типу та серйозності аномалії (фоновий витік, розрив труби, несправність датчика, оцінка розміру витoku);
- локалізація витoku, яка має на меті визначення приблизного місця витoku на основі наявних вимірювань.

Проблема ускладнюється тим, що у операторів ВРМ, як правило, немає повної інформації про поточні операційні характеристики системи водопостачання.

© Мельников О. С., Дорофеев Ю. І., Марченко Н. А., 2025



Дослідницька стаття: Цю статтю опубліковано видавництвом *НТУ «ХПІ»* у збірнику «Вісник Національного технічного університету "ХПІ" Серія: Системний аналіз, управління та інформаційні технології». Ця стаття поширюється за міжнародною ліцензією [Creative Commons Attribution \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/). **Конфлікт інтересів:** Автор/и заявив/или про відсутність конфлікту.



Оцінка стану системи виконується за допомогою датчиків витрат води та тиску. На практиці останнім надається перевага з економічних міркувань. Проте, датчики тиску мають порівняно невисоку точність і свідчать про втрати води лише опосередковано. Через високу ціну та експлуатаційні витрати кількість вузлів мережі, в яких встановлені датчики тиску, в реальних ВРМ коливається в межах 3–7 % від загальної кількості вузлів.

В той же час операційні характеристики ВРМ зазнають впливу великої кількості випадкових факторів, серед яких найважливішими є:

- коливання попиту на воду;
- старіння та засмічення труб;
- погодні умови;
- операції з клапанами та вентилями тощо.

Отже, навіть встановлення факту існування аномалії в роботі ВРМ є нетривіальною задачею. Її вирішенню присвячені, зокрема, роботи [3–5].

Практична важливість проблеми обумовила організацію в 2020 р. змагання «Битва методів виявлення та ізоляції витоків» (англ. Battle of the Leakage Detection and Isolation Methods, скорочено BattLeDIM) [6]. Метою змагання було створення відкритого середовища для тестування та порівняння різноманітних підходів до виявлення витоків, використовуючи дані системи диспетчерського контролю та збору даних про тиск та витрати води в мережі. Це дозволяє порівняти ефективність різних методик у стандартизованих умовах. Для змагання було створено реалістичну модель ВРМ для гіпотетичного міста L-Town (прототипом якого було реальне місто на Кіпрі). Модель містила різноманітні сценарії витоків, що дозволяло учасникам тестувати свої методики в умовах, наближених до реальних.

Змагання BattLeDIM викликало значний інтерес як в академічному середовищі, так і серед практиків. В ньому взяли участь 18 команд дослідників, які застосовували різні методики виявлення та локалізації витоків, зокрема аналіз часових рядів, машинне навчання, математичне програмування, евристичні методи тощо. Запропоновані методики оцінювались за допомогою економічних критеріїв, що включали вартість втраченої води та витрати на пошук точного місця витoku в межах зони локалізації. На жаль, навіть у найкращих команд ефективність не перевищувала 50 % від теоретичного оптимуму, що ще раз свідчить про складність проблеми.

Проблему виявлення витоків можна розглядати також в ширшому контексті діагностики аномалій [7]. Зокрема, вона має багато спільних рис із виявленням та локалізацією виробничих неполадок в процесі контролю якості продукції [8].

Постановка задачі. Встановлення факту наявності витoku в системі водопостачання можна розглядати як статистичну перевірку нульової гіпотези про відповідність показань датчиків стандартним умовам функціонування мережі. Зазвичай, при цьому можуть виникати помилки I та II роду:

- хибна тривога – помилкове виявлення аномалії, коли її насправді немає;
- пропуск витoku – класифікація стану системи як нормального, в той час як насправді в системі присутні витoki.

Балансування ймовірностей виникнення цих помилок є важливою з практичної точки зору задачею. Внаслідок складності мереж водопостачання ці ймовірності складно отримати в аналітичній формі. Тому для ефективної роботи автоматизованих систем діагностики витоків необхідна калібрація параметрів на базі комбінації імітаційного моделювання з аналітичними методами.

Інформаційна база дослідження. Для дослідження використовувалась модель ВРМ міста L-Town, розроблена для згаданого вище змагання BattLeDIM. Структура мережі наведена на рис. 1.

Мережа складається з 782 вузлів, поєднаних 905 трубами загальною довжиною 42,6 км. Висота вузлів коливається від 1,5 до 75 м над рівнем моря.

Водорозподільна мережа L-Town отримує воду з двох водосховищ та забезпечує питною водою близько 10 тисяч споживачів. Нормальний робочий тиск у мережі коливається між 20 і 30 м. У нижній частині міста (область В) встановлений редуційний клапан, який забезпечує постійний тиск на вході в область. Редуційні клапани також встановлені нижче за течією від двох основних водосховищ.

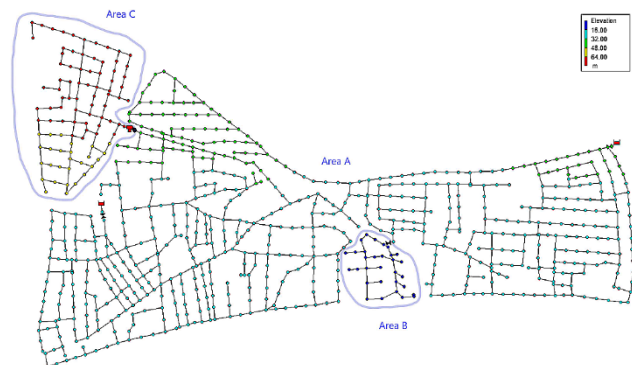


Рис. 1. Структура мережі водопостачання в L-Town

У верхній частині міста (область С) розміщені водонапірна вежа з насосом для забезпечення достатнього тиску для споживачів області. Насос запрограмований так, що вежа наповнюється вночі, а вдень спорожняється до зони С. Мережа має ділянки з різним тиском, а отже, з різною чутливістю до витоків.

Моніторинг стану мережі здійснюється за допомогою 33 датчиків тиску, розташування яких вказано на рис. 2. Покриття мережі датчиками становить, таким чином, біля 4,2 %. На виходах з двох водосховищ та на насосі біля водонапірної вежі розташовані також три датчики витрат води.

Для відновлення значень тиску та інших важливих параметрів у вузлах, які не обладнані датчиками, широко використовується система гідралічного моделювання EPANET [9]. Це програмне забезпечення було розроблене Агентством з охорони навколишнього середовища США (EPA) для моделювання гідраліки

та якості води у ВРМ. Воно є одним із найпопулярніших інструментів для аналізу поведінки систем водопостачання як у сталому стані, так і в динаміці.

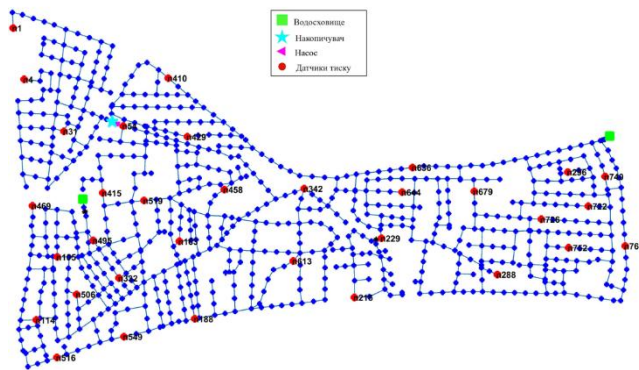


Рис. 2. Розташування датчиків тиску в L-Town

Для побудови моделі в EPANET необхідно надати вихідні дані про конфігурацію мережі: перелік вузлів (споживачів, резервуарів, насосних станцій), трубопроводів (із зазначенням довжин, діаметрів, коефіцієнтів шорсткості), а також графік споживання води в кожному вузлі в різні моменти часу. Також задаються фізичні координати вузлів та їх висота над рівнем моря, параметри насосів, клапанів, джерел тиску тощо.

Розрахунки в EPANET базуються на класичних гідрравлічних рівняннях – законах збереження маси в вузлах та енергії в трубах [10, с.140–148]. Втрати напору в трубах можуть альтернативно описуватись формулами Хейзена – Вільямса, Дарсі – Вейсбаха або Колебрука – Уайта [11, 12]. Для визначення витрат і тисків застосовується метод глобального градієнта Тодіні – Пілаті [13], який є модифікацією метода Ньютона – Рафсона. Передбачено багато способів візуалізації результатів.

Слід зауважити, що на практиці важко побудувати точну модель ВРМ. Попит у вузлах може коливатись у широких межах, шорсткість труб збільшується з плином часу, клапани можуть пропускати воду тощо. Тому для виявлення аномалій спершу слід встановити, який діапазон контрольованих значень можна вважати прийнятним, що потребує статистичних досліджень. Використання EPANET дозволяє виконувати такі дослідження шляхом імітаційного моделювання.

Виклад основного матеріалу. Запропонований підхід до розробки системи виявлення аномалій у ВРМ включає наступні три кроки:

- 1) статистичне профілювання: аналіз статистичних характеристик вимірювань датчиків при варіації вихідних параметрів;
- 2) калібрування параметрів для досягнення бажаних ймовірностей помилок I та II типів;
- 3) розробка правил діагностики аномалій на цій основі.

На першому етапі виконується імітаційне моделювання функціонування ВРМ при варіації попиту та інших частково невизначених параметрів.

Попит у вузлі мережі $i = 1, 2, \dots, N$ в момент часу $t = 1, 2, \dots, T$ моделюється в системі EPANET як:

$$d_{it} = \sum_{k \in D} b_{ik} p_{kt}, \quad (1)$$

де D – множина груп споживачів, p_{kt} – очікуваний попит споживачів k -ї групи в час t , b_{ik} – коефіцієнт навантаження вузла i з боку споживачів k -ї групи.

Для врахування мінливості попиту та визначення чутливості датчиків тиску до його коливань в кожному вузлі на попит було накладено мультиплікативний гаусівський шум. Щоб врахувати поступовість змін попиту в часі, часова послідовність шумів моделювалася як авторегресійний процес першого порядку AR(1), тобто:

$$\begin{aligned} \tilde{d}_{it} &= d_{it} (1 + \varepsilon_{it}), \quad t = 1, 2, \dots, T, \\ \varepsilon_{i0} &\sim N(0, \sigma^2); \quad \varepsilon_{it} = \rho \varepsilon_{i,t-1} + \eta_{it}, \quad t \neq 0, \\ \eta_{it} &\sim N(0, (1 - \rho^2) \sigma^2), \end{aligned} \quad (2)$$

де $\rho \in [0, 1]$ – коефіцієнт автокореляції.

Дисперсія збурювання η_{it} обрана так, щоби процес ε_{it} був стаціонарним із постійною дисперсією σ^2 . При $\rho = 0$ послідовність ε_{it} утворює білий шум; при $\rho = 1$ збурювання попиту в i -му вузлі буде постійним у часі. Наведені нижче результати моделювання відповідають значенням $\sigma = 0,2$; $\rho = 0,8$.

Отримані за схемою (2) реалізації значень попиту передавались для розрахунків параметрів мережі в EPANET за допомогою програмного інтерфейсу WNTR для Python [14]. В результаті проведення серії таких імітаційних експериментів можна отримати вибірку значень тиску p_{it} в усіх вузлах мережі (а також потоки води в трубах, фактично задоволений попит тощо). Серед цих значень найбільший інтерес становлять показники тиску для вузлів, обладнаних датчиками, адже саме на основі їхніх показань здійснюється моніторинг стану мережі в реальному часі. Позначимо множину таких вузлів через S .

Отримані у такий спосіб вибірки були перевірені на узгодженість із нормальним законом розподілу за критеріями Харке – Бера та хі-квадрат Пірсона [15]. За результатами цих тестів не було виявлено суттєвих відхилень від нормальності за виключенням вузлів, суміжних із редуційними клапанами; в останніх значення тиску можна вважати детермінованими в межах точності розрахунків. Типова гістограма розподілу значень тиску наведена на рис. 3.

Опрацювання результатів імітаційних експериментів дозволило отримати середні значення тиску μ_{it} та середні квадратичні відхилення σ_{it} для всіх $i \in S$, $t = 1, \dots, T$. На основі отриманих даних було визначено часовий коридор прийнятних значень для всіх датчиків із врахуванням властивостей нормального розподілу. Надалі будемо називати його зеленим коридором.

На рис. 4 наведено зелений коридор для датчика, встановленого у вузлі 'n410', протягом доби. Результат отримано за правилом трьох сигм на основі 100 імітаційних експериментів для кожного періоду часу.

Очевидно, що дисперсія тиску є неоднорідною: розкид значень уночі значно менше, ніж удень. Отже, аномалії в роботі ВРМ легше виявити вночі через нижчий рівень шумів.

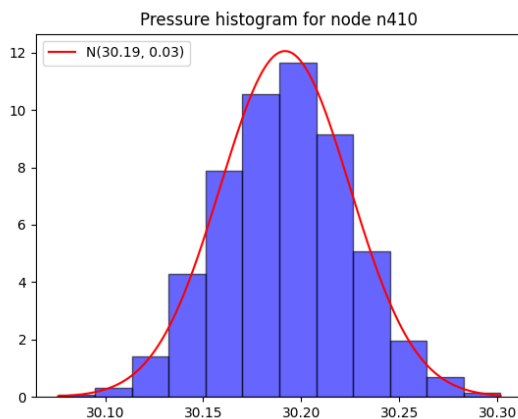


Рис. 3. Гістограма розподілу тиску у вузлі 'n410' опівночі

Хрестиками на рис. 4 позначені екстремальні значення тиску, що спостерігались в серії імітаційних експериментів. Червоні хрестики відповідають випадкам, коли спостережувані значення знаходились за межами зеленого коридору, тобто хибні тривоги. Порівняно висока кількість хибних тривог пояснюється великою кількістю експериментів.

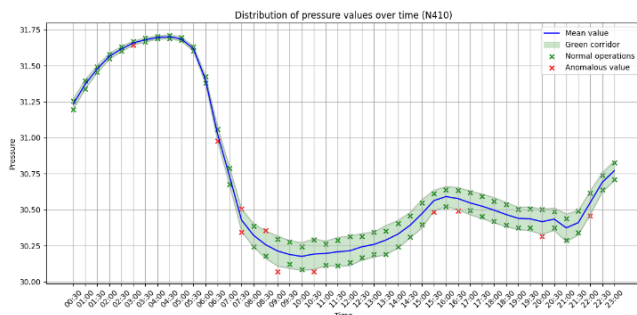


Рис. 4. Розподіл значень тиску у вузлі 'n410' протягом доби

Для побудови графіка з точністю до 30 хвилин потрібна генерація $24 \cdot 2 \cdot 100 = 4800$ нормально розподілених псевдовипадкових величин для кожного вузла. При ймовірності виходу окремого спостереження за межами діапазону $\pm 3\sigma$ $p = 0,003$, можна очікувати біля 14 хибних тривог.

Рис. 4 ілюструє компроміси, пов'язані з побудовою системи виявлення аномалій. При зменшенні ширини зеленого коридору зростає ризик виникнення хибних тривог, а при збільшенні – ризик невиявлення реально існуючої аномалії.

Ситуація ускладнюється тим, що мережа містить багато датчиків, показання яких тісно пов'язані між собою. Тому діагностика наявності витoku на підставі аномальних показань хоча б одного датчика (як це пропонується, наприклад, в роботі [16]) призведе до не виправданого підвищення ризику виникнення хибної тривоги.

Дійсно, якби показання датчиків були статистично незалежними, а ймовірність виходу показань

окремого датчика за межі зеленого коридору в нормальних умовах становила p , то при наявності M датчиків шанс виникнення хибної тривоги складав:

$$p_H(M) = 1 - (1 - p)^M. \quad (3)$$

Для моделі L-Town $M = 33$. Якщо встановлювати межі зеленого коридору за правилом трьох сигм, тоді $p = 0,003$. Підстановка цих значень у формулу (3) наводить до ймовірності хибної тривоги 0,094, що є непрактичним для будь-якої реальної системи.

Насправді, як ілюструє рис. 5, показання датчиків є корельованими між собою. На рисунку градацією кольору показано кореляцію тиску у вузлах мережі з тиском в одному з її центральних вузлів 'n198'. Як можна бачити, значення тиску в зонах А, В та С слабо пов'язані між собою. Проте, в центральній зоні А значення тиску в усіх вузлах є сильно корельованими.

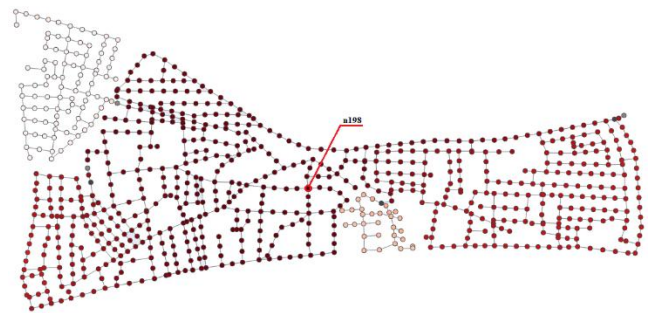


Рис. 5. Кореляція значень тиску у вузлах мережі водопостачання L-Town з тиском у вузлі 'n198'

Тому ймовірність хибної тривоги в реальних умовах буде знаходитись в діапазоні $[p, p_H(M)]$. Аналітичне визначення цієї ймовірності вимагає обчислення багатовимірних інтегралів, тому для розв'язання цієї задачі доцільно звернутись до імітаційного моделювання. Для цього достатньо підрахувати кількість імітаційних експериментів, в яких показання хоча б одного датчика вийшли за межі розрахованого для нього зеленого коридору, і зіставити її із загальною кількістю проведених експериментів.

Аналогічно можна розрахувати ймовірність пропуску аномалії. Для цього за допомогою EPANET було змодельовано ситуацію, коли виток фіксованого розміру по черзі мав місце в кожній з труб ВРМ за наявності шумів, згенерованих за описаною вище схемою. В даному випадку результати розрахунків залежать також від розміру витoku – чим більше виток, тим легше його виявити.

Висока кореляція показань датчиків наводить на думку про те, що найчастіше виток вплине відразу на декілька датчиків. Це підтверджують результати, зображені на рис. 6, де наведено гістограму розподілу кількості датчиків із аномальними показаннями n_a залежно від розміру витoku. Моделювалися три сценарії витоків: малий $2 \text{ м}^3/\text{год}$, середній $5 \text{ м}^3/\text{год}$ та великий $10 \text{ м}^3/\text{год}$.

Для малих витоків існують істотні шанси їх невиявлення. Кількість датчиків із аномальними показаннями буде відносно невеликою, але навіть в цьому

випадку найвірогідніше, що на виток відреагують одразу три датчики. Для середніх та великих витоків є велика ймовірність того, що аномальними виявляться показання всіх датчиків у зоні А. Локальний максимум в точці $n_a = 3$ пояснюється тим, що коливання тиску в зоні С статистично непов'язані з коливаннями тиску в інших частинах мережі, і на виток в цій зоні реагують тільки 3 встановлених там датчики.

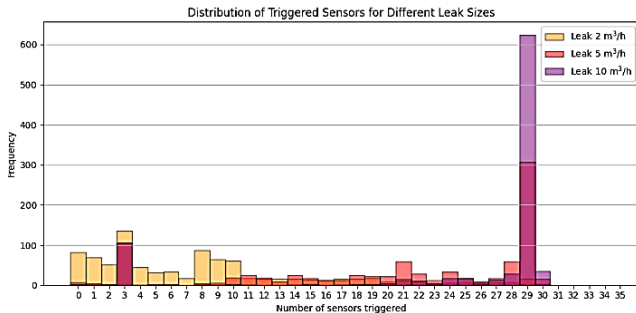


Рис. 6. Розподіл кількості датчиків із аномальними показаннями залежно від розміру витoku

Отримані результати дозволяють стверджувати, що висновок про наявність аномалії надійніше робити, коли за межі зеленого коридорів виходять показання принаймні двох датчиків. На користь цього говорить також той факт, що аномальні показання лише одного датчика можуть пояснюватись його несправністю.

Таким чином, для розробки ефективної системи діагностики аномалій у ВРМ потрібно вирішити (принаймні) два питання:

- вибір ширини зеленого коридору для кожного датчика;
- визначення кількості датчиків, показання яких вийшли за межі зеленого коридору, достатніх для формування висновку про наявність аномалії.

На рис. 7 наведено графік залежності ймовірностей помилок I та II роду від ширини зеленого коридору та кількості датчиків, на базі яких приймається рішення про наявність аномалії. Ширина коридору визначається як кількість середніх квадратичних відхилень від центральних значень, тобто значення k у виразі $\mu_{it} \pm k\sigma_{it}$.

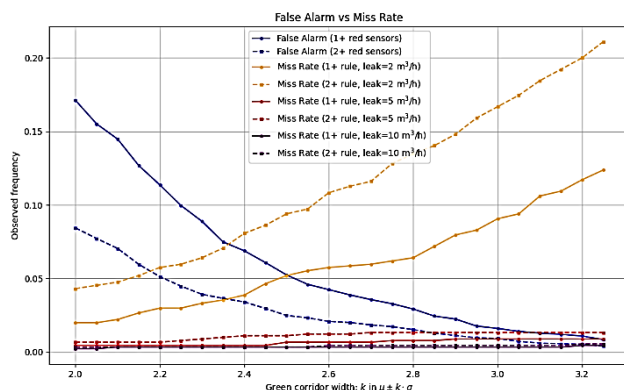


Рис. 7. Ймовірності помилок I та II типів залежно від ширини зеленого коридору та кількості датчиків

Як можна бачити, збільшення ширини зеленого коридору зменшує ймовірність хибних тривог та збільшує ймовірність пропуску витoku. При малих витках величини цих двох ефектів співставні. При середніх та великих витках ризик пропуску витoku зростає зневажливо мало.

Кількість датчиків, дані яких аналізує система, має аналогічний ефект. Діагностика на основі показань двох датчиків значно зменшує ризик хибної тривоги. Для малих витоків цей вигравш нівелюється одночасним збільшенням ризику їх невиявлення. Для середніх та великих витоків ефект зменшення ризику хибної тривоги домінує над ефектом збільшення ризику пропуску аномалії.

Вплив вищезначених параметрів на коректність виявлення аномалій стає більш вираженим, якщо розглянути суму ймовірностей помилок I та II типів. Як свідчать наведені на рис. 8 графіки, для кожного конкретного розміру витoku можна підібрати таку ширину зеленого коридору та кількість датчиків, які призводять до мінімальної ймовірності невірної статистичного висновку. Проте, не існує значення, яке буде оптимальним для всіх розмірів витоків одночасно. Отже, калібрування системи виявлення аномалій повинна проводитись в процесі людино-машинного діалогу з фахівцями, що приймають рішення.

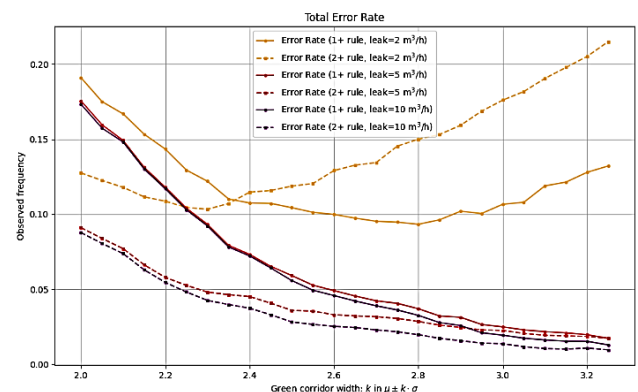


Рис. 8. Ймовірності хибних статистичних висновків залежно від ширини зеленого коридору та кількості індикаторів

Для малих витоків мінімальна ймовірність помилок досягається при виявленні аномалій на основі даних одного датчика. Разом з тим слід зауважити, що ця ймовірність залишається високою, а точна локалізація витoku за показаннями одного датчика практично неможлива. Для середніх та великих витоків використання даних двох датчиків для формування висновків явно домінує над схемою з одним датчиком.

Висновки. В роботі розглянуто задачу проектування та калібрування системи виявлення аномалій у мережах водопостачання. Для розв'язання задачі запропоновано триетапну схему, яка включає: статистичне профілювання датчиків мережі; калібрування системи для досягнення бажаного співвідношення між ризиками хибних тривог та пропуску аномалій; визначення правил формування висновків про наявність аномалій. Розроблено методику статистичного профілювання та калібрування системи на базі імітаційного моделювання в програмному середовищі EPANET/WNTR. Дослід-

жено чутливість результатів до ширини інтервалу значень, які вважаються нормальними, та кількості датчиків, задіяних у процедурі виявлення аномалій.

Встановлено, що не існує комбінації параметрів, які забезпечують оптимальні результати для всіх можливих розмірів витоків одночасно. Тому запропоновано виконувати калібрацію системи виявлення аномалій за допомогою людино-машинної процедури із залученням відповідних фахівців.

Запропонована система має потенціал для подальшого вдосконалення. Зокрема, дослідження фокусувалось на встановленні самого факту наявності аномалій у функціонуванні ВРМ, а не на їх локалізації. У той же час порівняння величин відхилень тиску в різних вузлах мережі від середніх значень може сприяти виявленню місцезнаходження аномалії. Розвиток цієї теми є перспективним напрямком для подальших досліджень.

Фінансування та підтримка. Дослідження виконано за підтримки гранту, отриманого за програмою "NWO Hop-On Call for Researchers Based in Ukraine: NWO-NRFU Partnership Initiative 2023" в рамках проєкту 19454 "DiTEC: Digital Twin for Evolutionary Changes in water networks".

Список використаної літератури

1. Liemberger R., Wyatt A. Quantifying the global non-revenue water problem. *Water Supply*. 2019. Vol. 19 (3), P. 831–837. DOI: 10.2166/ws.2018.129.
2. Chan T. K., Chin C. S., Zhong X. Review of current technologies and proposed intelligent methodologies for water distributed network leakage detection. *IEEE Access*. 2018. Vol. 6 (12), P. 78846–78867. DOI: 10.1109/ACCESS.2018.2885444.
3. Carrasco-Jiménez J. C., Baldaro F., Cucchiatti F. Detection of anomalous patterns in water consumption: an overview of approaches. In: Arai K., Kapoor S., Bhatia R. (eds) *Intelligent Systems and Applications. Advances in Intelligent Systems and Computing*. 2021. Vol. 1250. Springer, Cham. DOI: 10.1007/978-3-030-55180-3_2.
4. Vardhan A. H. et al. Anomaly detection in water distribution systems. *4th Smart Cities Symposium (SCS 2021)*, Online Conference, Bahrain. 2021. P. 219–222. DOI: 10.1049/icp.2022.0344.
5. Gueye K., Boly A., Bame N. Using machine learning for anomaly detection in drinking water from distribution network. *7th International Symposium on Computer Science and Intelligent Control (ISCSIC)*, Nanjing, China. 2023. P. 205–210. DOI: 10.1109/ISCSIC60498.2023.00050.
6. Vrachimis S. G. et al. Battle of the leakage detection and isolation methods. *Journal of Water Resource Planning and Management*. 2022. Vol. 148 (12): 04022068. P. 1–57. DOI: 10.1061/(ASCE)WR.1943-5452.0001601.
7. Chandola V., Banerjee A., Kumar V. Anomaly detection: A survey. *ACM computing surveys (CSUR)*. 2009. Vol. 41 (3). P. 1–58. DOI: 10.1145/1541880.1541882.
8. Bersimis S., Psarakis S., Panaretos J. Multivariate statistical process control charts: an overview. *Quality and Reliability engineering international*. 2007. Vol. 23 (5). P. 517–543. DOI: 10.1002/qre.829.
9. EPANET: Application for modeling drinking water distribution systems. URL: <https://www.epa.gov/water-research/epanet> (дата звернення: 10.05.2025).
10. Любчик Л. М., Дорофеев Ю. І. *Інформаційні технології керування запасами в мережах поставок: монографія*. НТУ «ХПІ». Харків: ФОП Панов А. М., 2019. 180 с.
11. Haktanir T., Ardiclioglu M. Numerical modeling of Darcy-Weisbach friction factor and branching pipes problem. *Advances in Engineering Software*. 2004. Vol. 35 (12). P. 773–779. DOI: 10.1016/j.advengsoft.2001.07.005.
12. Travis Q. B., Mays L. Relationship between Hazen-William and Colebrook-White roughness values. *Journal of Hydraulic Engineering*, 2007. Vol. 133 (11). P. 1270–1273. DOI: 10.1061/(ASCE)0733-9429(2007)133:11(1270).
13. Todini E., Pilati S. A gradient algorithm for the analysis of pipe networks. In: Coulbeck B., Orr C.-H. (eds) *Computer applications in water supply: Vol. 1 - System analysis and simulation*. London, John Wiley & Sons, 1988, pp. 1–20.
14. Klise K. A. et al. Water network tool for resilience (WNTR) user manual: Version 0.2.3. U.S. EPA Office of Research and Development, Washington, DC, EPA/600/R-20/185. 2020. 82 p.
15. Мазманішвілі О. С., Мельников О. С. *Практикум з математичної статистики*. Харків: НТУ «ХПІ», 2025. 288 с.
16. Mazzoni F., Marsili V., Alvisi S., Franchini M. Detection and pre-localization of anomalous consumption events in water distribution networks through automated, pressure-based methodology. *Water Resources and Industry*. 2024. Vol. 31, 100255. DOI: 10.1016/j.wri.2024.100255.

References (transliterated)

1. Liemberger R., Wyatt A. Quantifying the global non-revenue water problem. *Water Supply*. 2019, vol. 19 (3), pp. 831–837. DOI: 10.2166/ws.2018.129.
2. Chan T. K., Chin C. S., Zhong X. Review of current technologies and proposed intelligent methodologies for water distributed network leakage detection. *IEEE Access*. 2018, vol. 6 (12), pp. 78846–78867. DOI: 10.1109/ACCESS.2018.2885444.
3. Carrasco-Jiménez J. C., Baldaro F., Cucchiatti F. Detection of anomalous patterns in water consumption: an overview of approaches. In: Arai K., Kapoor S., Bhatia R. (eds) *Intelligent Systems and Applications. Advances in Intelligent Systems and Computing*. 2021, vol. 1250. Springer, Cham. DOI: 10.1007/978-3-030-55180-3_2.
4. Vardhan A. H. et al. Anomaly detection in water distribution systems. *4th Smart Cities Symposium (SCS 2021)*, Online Conference, Bahrain. 2021, pp. 219–222. DOI: 10.1049/icp.2022.0344.
5. Gueye K., Boly A., Bame N. Using machine learning for anomaly detection in drinking water from distribution network. *7th International Symposium on Computer Science and Intelligent Control (ISCSIC)*, Nanjing, China. 2023, pp. 205–210. DOI: 10.1109/ISCSIC60498.2023.00050.
6. Vrachimis S. G. et al. Battle of the leakage detection and isolation methods. *Journal of Water Resource Planning and Management*. 2022, vol. 148 (12): 04022068, pp. 1–57. DOI: 10.1061/(ASCE)WR.1943-5452.0001601.
7. Chandola V., Banerjee A., Kumar V. Anomaly detection: A survey. *ACM computing surveys (CSUR)*. 2009. Vol. 41 (3), pp. 1–58. DOI: 10.1145/1541880.1541882.
8. Bersimis S., Psarakis S., Panaretos J. Multivariate statistical process control charts: an overview. *Quality and Reliability engineering international*. 2007, vol. 23 (5), pp. 517–543. DOI: 10.1002/qre.829.
9. EPANET: Application for modeling drinking water distribution systems. Available at: <https://www.epa.gov/water-research/epanet> (accessed 10.05.2025).
10. Lyubchik L. M., Dorofiev Yu. I. *Informatsiini tekhnologii keruvannya zapasami v merezhakh postavok: monografii* [Information technologies for inventory control in supply networks]. NTU «KhPI». Kharkiv: FOP Panov A. M., 2019. 180 p.
11. Haktanir T., Ardiclioglu M. Numerical modeling of Darcy-Weisbach friction factor and branching pipes problem. *Advances in Engineering Software*. 2004, vol. 35 (12), pp. 773–779. DOI: 10.1016/j.advengsoft.2001.07.005.
12. Travis Q. B., Mays L. Relationship between Hazen-William and Colebrook-White roughness values. *Journal of Hydraulic Engineering*, 2007, vol. 133 (11), pp. 1270–1273. DOI: 10.1061/(ASCE)0733-9429(2007)133:11(1270).
13. Todini E., Pilati S. A gradient algorithm for the analysis of pipe networks. In: Coulbeck B., Orr C.-H. (eds) *Computer applications in water supply: Vol. 1 - System analysis and simulation*. London, John Wiley & Sons, 1988, pp. 1–20.
14. Klise K.A. et al. Water Network Tool for Resilience (WNTR) User Manual: Version 0.2.3. U.S. EPA Office of Research and Development, Washington, DC, EPA/600/R-20/185, 2020. 82 p.
15. Mazmanishvili O. S., Melnikov O. S. *Praktykum z matematychnoi statystyky* [Workshop on mathematical statistics]. Kharkiv: NTU «KhPI», 2025. 288 p.

16. Mazzoni F., Marsili V., Alvisi S., Franchini M. Detection and pre-localization of anomalous consumption events in water distribution networks through automated, pressure-based methodology. *Water*

Resources and Industry. 2024, vol. 31, 100255. DOI: 10.1016/j.wri.2024.100255.

Надійшла (received) 15.05.2025

UDC 628.17

O. S. MELNIKOV, Candidate of Economic Sciences (PhD), Docent, National Technical University "Kharkiv Polytechnic Institute", Associate Professor at the Department of System Analysis and Information-Analytical Technologies; Kharkiv, Ukraine; e-mail: Oleg.Melnikov@khpi.edu.ua, ORCID ID: <https://orcid.org/0000-0002-2409-4983>.

Yu. I. DOROFIEIEV, Doctor of Technic Sciences, Professor, National Technical University "Kharkiv Polytechnic Institute", Professor at the Department of System Analysis and Information-Analytical Technologies; Kharkiv, Ukraine; e-mail: Yuri.Dorofieiev@khpi.edu.ua, ORCID ID: <https://orcid.org/0000-0002-7964-1286>.

N. A. MARCHENKO, Candidate of Technic Sciences (PhD), Docent, National Technical University "Kharkiv Polytechnic Institute", Deputy Director of the Educational and Scientific Institute of Computer Sciences and Information Technologies; Kharkiv, Ukraine; e-mail: Natalia.Marchenko@khpi.edu.ua, ORCID ID: <https://orcid.org/0000-0001-9889-3713>.

STATISTICAL APPROACH TO DETECTION OF ANOMALIES IN WATER DISTRIBUTION NETWORKS

This paper is devoted to solving the problem of developing an automated system for detecting anomalies in water distribution networks. The main causes of such anomalies are background leaks and pipe breaks. To address this problem, a statistical approach is proposed, which consists in testing the null hypothesis that the readings of pressure and/or water flow sensors received in real time correspond to the standard conditions of the network. The paper proposes a three-stage anomaly detection scheme, which includes: statistical profiling of network sensors; system calibration to achieve the desired ratio between the risks of false alarms and the omission of existing anomalies; determination of rules for drawing conclusions about the presence of anomalies. A methodology for statistical profiling and calibration of the system based on simulation modeling in the EPANET software environment using the WNTR software interface in Python was developed. In the process of such modeling, the distribution of pressure readings in the network sensors is obtained based on water demand fluctuations. As an example, the model of the L-Town water supply network was studied, which was developed for the BattLeDIM leak detection and isolation competition. The sensitivity of the anomaly detection results to the range of sensor values that are considered normal, as well as the number of sensors involved in the anomaly detection procedure, is investigated. The dependence of the number of sensors with anomalous readings on the size of the leak is analyzed. It is established that there is no combination of parameters of the proposed anomaly detection system that provides optimal results for all possible leakage sizes simultaneously, as a result of which it is proposed to calibrate the system using a man-machine procedure.

Keywords: water distribution network, anomaly, leakage, simulation modeling, EPANET, statistical hypothesis testing.

Повні імена авторів / Author's full names

Автор 1 / Author 1: Мельников Олег Станіславович, Melnikov Oleg Stanislavovich

Автор 2 / Author 2: Дорофєєв Юрій Іванович, Dorofieiev Yuri Ivanovich

Автор 3 / Author 3: Марченко Наталя Андріївна, Marchenko Natalia Andriivna