

## РЕАЛІЗАЦІЯ КОНЦЕПЦІЇ ІНФОРМАЦІЙНОЇ ЕНТРОПІЇ ПРИ СТВОРЕННІ МЕРЕЖ МОНІТОРИНГУ ВОД

Безсонний В.Л.

Харківський національний економічний університет імені Семена Кузнеця  
пр. Науки, 9А, 61166, м. Харків,

Харківський національний університет імені В.Н. Каразіна  
майд. Свободи, 4, 61022, м. Харків  
[bezsonny@gmail.com](mailto:bezsonny@gmail.com)

У статті наводиться огляд використання концепції інформаційної ентропії при здійсненні проектування мереж моніторингу вод. Базовими принципами проектування мереж моніторингу вод стали: певна кількість станцій моніторингу, місця розташування станцій і період передачі даних або частота дискретизації. Теорія інформації була спочатку розроблена для вимірювання вмісту інформації в наборі даних та з часом набула застосування для вирішення проблем раціонального використання водних ресурсів. На разі її застосування поширюється на проектування мереж моніторингу вод. Приймається концепція, згідно з якою ентропія зможе пояснити властиву для станції моніторингу або мережі моніторингу інформацію. Основна мета при цьому – отримання максимальної кількості інформації. На сьогодні зниження числа станцій у мережах моніторингу вод стало загальною тенденцією через фінансові обмеження і зміни пріоритетів моніторингу. Тому визначення адекватної кількості станцій моніторингу та їх розташування стало критично важливим для проектування мереж моніторингу вод. Важливість мереж моніторингу якості вод полягає в їх здатності допомогти у визначенні тих параметрів, які перевищують нормативи якості води. Однак стандартизована методологія процесу проектування робіт для належної мережі моніторингу води ще не була розроблена у зв'язку з практичною і соціально-економічною складністю в різних умовах проектування. Унікальна перевага цього підходу з використанням інформаційної ентропії полягає в тому, що мережу моніторингу вод можна оцінити або спроектувати на основі інформації, яку мережа контролює, що відрізняється від заданої щільності станцій, запропонованої в керівних принципах Всесвітньої метеорологічної організації. Крім того, у поєднанні з методами багатоцільової оптимізації в оптимальний процес проектування мережі можна включити специфічні критерії водокористувачів. *Ключові слова:* екологічний стан поверхневих вод, концепція інформаційної ентропії, моніторинг вод.

### Implementation of the concept of information entropy when creating water monitoring networks. Bezsonnyi V.

The article provides an overview of the use of the concept of information entropy in the design of water monitoring networks. The basic principles of designing water monitoring networks have become: a certain number of monitoring stations, the location of the stations and the period of data transmission or sampling frequency. Information theory was originally developed to measure the content of information in a data set and over time gained application to solve the problems of rational use of water resources. In the case of its application, it applies to the design of water monitoring networks. The concept is adopted, according to which entropy will be able to explain the information specific to the monitoring station or monitoring network. The main goal is to obtain the maximum amount of information. Today, the decrease in the number of stations in water monitoring networks has become a general trend due to financial constraints and changes in monitoring priorities. Therefore, determining the adequate number of monitoring stations and their locations has become critically important for the design of water monitoring networks. The importance of water quality monitoring networks lies in their ability to help identify those parameters that exceed water quality standards. However, a standardized methodology for the design process of works for a proper water monitoring network has not yet been developed due to the practical and socio-economic complexity in different design conditions. A unique advantage of this information entropy approach is that a water monitoring network can be evaluated or designed based on the information that the network monitors, which differs from the given station density suggested by the World Meteorological Organization guidelines. In addition, in combination with multi-objective optimization methods, specific criteria of water users can be included in the optimal network design process. *Key words:* ecological state of surface waters, concept of information entropy, water monitoring.

**Постановка проблеми.** Незважаючи на те, що Україна на сьогодні має внутрішні запаси водних ресурсів, важливими є питання їх доступності до 100% міського та сільського населення, очистки, зменшення впливу дисбалансу між посушливими регіонами півдня та іншими більш водоемними територіями півночі та північного заходу, звести до мінімуму та ліквідувати скидання відходів, небезпечних речовин, збільшити частку повторного використання

води та підняти ефективність використання водних ресурсів. Стандартизованого процесу проектування мереж моніторингу води не існує, за винятком загальних рекомендацій Всесвітньої метеорологічної організації (ВМО) щодо мінімальної щільності мережі. Хоча однією з основних проблем при проектуванні оптимальних гідрометричних мереж було встановлення цілей проектування, теорія інформації була успішно прийнята до проблем проектування мереж

шляхом надання заходів інформаційного вмісту, який може бути доставлений зі станції або мережі.

**Актуальність дослідження.** Адекватні мережі моніторингу води та дані про якість з них є одним з перших і первинних кроків на шляху до ефективного управління водними ресурсами. Базовими принципами проектування мереж моніторингу вод стали: певна кількість станцій моніторингу, місця розташування станцій і період передачі даних або частота дискретизації [1, 2]. Останні технологічні досягнення дозволили поступово перейти від ручного відбору проб до автоматизованих спостережень, в той час як деякі параметри якості води все ще вимагають польового та/або лабораторного аналізу води або інших проб навколишнього середовища. Теорія інформації була спочатку розроблена Шенноном в 1948 році [3] для вимірювання вмісту інформації в наборі даних і була застосована для вирішення проблем водних ресурсів. На разі її застосування поширюється на проектування мереж моніторингу вод, прийнявши концепцію, згідно з якою ентропія зможе пояснити властиву для станції моніторингу або мережі моніторингу інформації. Зрозуміло, що основна ціль при цьому – мати максимальну кількість інформації.

**Зв'язок авторського доробку із важливими науковими та практичними завданнями.** Тема дослідження відповідає основним напрямкам Концепції Загальнодержавної цільової соціальної програми “Питна вода України” на 2022–026 роки, схваленої розпорядженням Кабінету Міністрів України від 28 квітня 2021 р. № 388-р. У вересні 2015 року в рамках 70-ї сесії Генеральної Асамблеї ООН у Нью-Йорку відбувся Саміт ООН зі сталого розвитку та прийняття Порядку денного розвитку після 2015 року, на якому було затверджено нові орієнтири розвитку. Підсумковим документом Саміту «Перетворення нашого світу: порядок денний у сфері сталого розвитку до 2030 року» було затверджено 17 Цілей Сталого Розвитку та 169 завдань.

Ціллю № 6 є «Чиста вода та належні санітарні умови», серед основних завдань якої – підвищення ефективності водокористування та забезпечення впровадження інтегрованого управління водними ресурсами, що не є можливим без ефективно організованої мережі моніторингу вод.

**Аналіз останніх досліджень і публікацій.** У наявних роботах досліджується широкий спектр підходів до проектування мереж моніторингу вод, таких як методи оптимізації, просторова інтерполяція, статистичний аналіз, застосування теорії інформації, фізико-географічний аналіз, опитування користувачів або рекомендації експертів і комбінації декількох методів [6–16]. У роботі [15] розглянуто положення щодо зниження гідрометричної щільності мережі, підкреслено важливість якісних даних з добре спроектованих мереж та розглянуто ряд підходів до спроектованих мереж. Автори також порів-

няли статистичний, просторовий інтерполяційний, фізіографічний, вибірковий та ентропійний підходи до проектування гідрометричних мереж.

**Виділення не вирішених раніше частин загальної проблеми, котрим присвячується означена стаття.** На сьогодні зниження мереж моніторингу води стало загальною тенденцією через фінансові обмеження і зміни пріоритету моніторингу [13–15]. Тому визначення адекватної кількості станцій моніторингу та їх розташування стало критично важливим для проектування мереж. Однак стандартизована методологія процесу проектування робіт для належної мережі моніторингу води ще не була розроблена у зв'язку з практичною і соціально-економічною складністю в різних умовах проектування [6, 14].

**Новизна.** Підхід – чим більше даних ми збираємо, тим більше проблем з водними ресурсами вирішується ефективно – не є завжди вірним, оскільки нерелевантні, неадекватні або неефективні дані в неправильному місці або в невідповідний час можуть перешкоджати якості набору даних [4, 14, 15]. Одним з найбільш перспективних підходів до проектування мереж є застосування методів ентропії з використанням принципу максимальної ентропії і передачі інформації.

**Виклад основного матеріалу.** У термодинаміці ентропія розуміється як міра неупорядкованості або випадковості системи. Шеннон [16] розширив концепцію ентропії до теорії інформації, визнавши, що невизначеність у системі буде зменшена, коли інформація додається до системи. Тому термін ентропія в теорії інформації, введений Шенноном [16] у 1948 році, описує кількість інформаційного вмісту у випадковій величині. Імовірність події зазвичай визначається ймовірністю  $p$ . Якщо ймовірність події дуже висока, наприклад 0,9999 або одиниця, це не буде несподіванкою і можна точно передбачити результат. З іншого боку, будь-яка подія з низькою ймовірністю характеризується високим ступенем невизначеності, тому у випадку її виникнення може бути отримано значну кількість інформації. Отже, інформація про подію, що сталася, обернено пропорційна її ймовірності,  $1/p$  [17]. Припустимо, що є дві незалежні події  $A$  і  $B$  з їх імовірностями  $p_A$  і  $p_B$  відповідно. Імовірність спільного виникнення подій  $A$  і  $B$  може бути  $p_A p_B$ , а інформація, отримана спільною подією, тоді дорівнює  $1/(p_A p_B)$ . Однак сума інформації від кожної окремої події не дорівнює інформації від спільної події, тобто:

$$\frac{1}{p_A p_B} \neq \frac{1}{p_A} + \frac{1}{p_B}. \quad (1)$$

Єдиним переходом, який зробить обидві частини рівняння (1) рівними, є логарифм [17–19], який можна записати так:

$$\log \frac{1}{p_A p_B} = \log \frac{1}{p_A} + \log \frac{1}{p_B} = -\log p_A - \log p_B. \quad (2)$$

Подібним чином показано [20], що невизначеність події з імовірністю  $p$  дорівнює  $-\log p$ , що стало основою ентропії Шеннона.

Коли інформація надається в системі, можна очікувати, що невизначеність системи буде зменшена; тому кількість інформації, яка була надана системі при умові знання змінної, називається граничною ентропією. Якщо очікується, що випадкова змінна  $X$  матиме  $N$  результатів із розподілом ймовірностей  $P = \{p_1, p_2, \dots, p_N\}$ , (зважена) середня інформація, представлена  $N$  спільними подія, визначається як:

$$H(X) = -p_1 \log p_1 - p_2 \log p_2 - \dots - p_N \log p_N = \quad (3)$$

$$= -\sum_{i=1}^N p_i \log p_i, \quad \sum_{i=1}^N p_i = 1, \quad p_i \geq 0.$$

де  $H(X)$  – гранична ентропія випадкової величини  $X$ .

У рівнянні (3) можна використовувати будь-яку основу логарифма, вибір залежить від поставленої проблеми. У бінарних запитаннях (тобто питаннях «так» або «ні») слід використовувати двійку, а відповідну одиницю ентропії – біт. Подібним чином, одиниця трит для основи 3, одиниця нат для основи  $e$  та одиниця децибел або децит для основи 10 є прикладами одиниць інформації. Оскільки ми розглядаємо застосування ентропії для проектування гідрометричної мережі, то і очікувана відповідь процесу проектування може бути або «використовувати / включати / встановлювати станцію» або «не використовувати / включати / встановлювати станцію», щоб мережа була оптимальною. Таким чином, логарифм у розрахунку ентропії Шеннона для проектування гідрометричної мережі є найбільш прийнятним з основою два. Тоді значення  $H(X)$  з рівняння (3) буде розумітися як інформаційний вміст станції  $X$ , який може бути отриманий у разі її встановлення.

Якщо змінна  $X$  має відоме значення, ймовірність події буде дорівнювати одиниці, а всі інші альтернативні ймовірності будуть рівні нулю. Інформаційний вміст у змінній  $X$ ,  $H(X)$  дорівнюватиме нулю за рівнянням (3), що означає відсутність невизначеності або певного результату. З іншого боку, якщо змінна  $U$  має рівномірний розподіл (тобто ймовірність кожної події дорівнює  $1/N$ ), ентропія змінної  $U$  буде:

$$H(U) = \log N \quad (4)$$

Значення рівняння (4) часто називають максимальною ентропією або насиченою ентропією. Ці дві ентропії,  $H(X)$  і  $H(U)$ , визначають мінімальну та максимальну межі значень ентропії, тобто:

$$0 \leq H(X) \leq \log N \quad (5)$$

Важливість мереж моніторингу якості вод полягає в їх здатності допомогти у визначенні тих параметрів, які перевищують нормативи якості води. Кілька стратегій моніторингу якості вод, в тому числі два методи, які використовують вимірювання ентропії [21, 22], були розглянуті у роботі [10]. Побудова єдиного підходу до розробки мережі моніторингу якості вод є практично неможливим.

Незважаючи на це, різні застосування методів трансінформації показали перспективність оптимального перепроєктування та зменшення мереж моніторингу якості води [21–22]. У роботі [24] виявлено, що шляхом максимізації багатовимірної передачі інформації між вибраними та невибраними станціями, використовуючи модель керування зливовими водами для імітації загального вмісту завислих речовин і генетичний алгоритм для оптимізації, можна спроектувати мережу оптимальної якості води для каналізаційної системи. Дослідники [25] порівнювали теорію інформації, час виявлення та показники надійності для проектування мережі моніторингу каналізаційної системи за допомогою підходів оптимізації як з одним, так і з кількома цілями. Було показано, що для невеликої мережі моніторингу методи мали однакову продуктивність, тоді як метод виявлення на основі одного об'єкта, заснований на часі, мав дещо кращі результати, коли кількість станцій моніторингу більша. У роботі [26] використовували ймовірності перевищення для визначення ентропії порушення розчиненого кисню та хлорофілу-а в гирлі річки Нойс. Поряд з ентропією порушення, загальна ентропія системи використовувалася як міра для визначення областей важливості моніторингу. Багатоцільова схема оптимізації, заснована на вагових коефіцієнтах, призначених експертом, була використана для розробки компромісного рішення з трьох мір ентропії. Зрештою, метод дозволив ідентифікувати зони з високою невизначеністю, які виграють від майбутнього моніторингу якості води. Доступність даних є проблемою при використанні ентропійних методів, особливо при спробі використати їх у проєкті мережі моніторингу в басейні, де не проводилося вимірювань. Щоб вирішити це питання, у дослідженні [27] розроблено метод, який використовує міру, аналогічну граничній ентропії. Цей метод використовує такі характеристики басейну, як довжина та кількість ділянок у річкової мережі у якості функції витрат, яка потім оптимізується за допомогою комбінованого генетичного алгоритму та фільтрації. Було показано, що це обчислювально ефективний метод для використання в оптимальному проєктуванні мережі басейну річки без вимірювання.

Гідрологічні процеси взаємопов'язані кругообігом води, існують причини та наслідки між гідрологічними змінними. Наприклад, якщо випадає значна кількість опадів, ймовірно, збільшиться стік або рівень ґрунтових вод; отже, інформаційний вміст змінної може впливати на вміст інших змінних. Дослідниками [28] розроблено метод проектування багатовимірної мережі, взявши умовну ентропію як міру інформації, яка не залежить від даної змінної. У їхньому дослідженні метод передбачає одночасне проектування мережі моніторингу опадів і потоків. Зокрема, метод дотримувався традиційного багатоцільового підходу, який максимізує спільну ентропію.

піню та мінімізує загальну кореляцію, але додає ще одну задачу, яка максимізує умовну ентропію мережі потоку з урахуванням мережі опадів, щоб імітувати напрямок кругообігу води, оскільки річковий стік може коливатися через опади. Після порівняння інтегрованого плану з планом однієї змінної, їхні результати показали, що ефективність мережевої інтеграції здебільшого досягається від зменшення кількості додаткових станцій для випадіння опадів. Було також виявлено, що інтегрований підхід до проектування мережі дозволяє додати станцію моніторингу опадів в місці, яке буде корисним для водонісної мережі.

**Головні висновки.** Успішного управління водними ресурсами неможливо досягти без належних мереж моніторингу вод. Незважаючи на значний прогрес у методах проектування мереж, стандартизована методологія проектування ще не напрацьована. Після розробки теорії інформації в 1940-х роках концепції ентропії були застосовані в різних програмах щодо проблем проектування мереж моні-

рингу вод. Унікальна перевага цього підходу полягає в тому, що мережу моніторингу вод можна оцінити або спроектувати на основі інформації, яку мережа контролює, що відрізняється від заданої щільності станцій, запропонованої в керівних принципах ВМО. Крім того, у поєднанні з методами багатоцільової оптимізації в оптимальний процес проектування мережі можна включити специфічні критерії водокористувачів.

**Перспективи подальших досліджень.** Майбутні дослідження мають бути зосереджені на порівняльних дослідженнях методів проектування множинної ентропії, підходів до дискретизації та характеристик даних. Сучасна література надає багато нових підходів до проектування ентропії та еволюції концепцій, але ретельні порівняння мають вирішальне значення для надання загальних рекомендацій для проектування мережі. Незважаючи на виявлені потенційні джерела суб'єктивності, ентропійні методи залишаються одним із найбільш об'єктивних підходів до проектування мережі.

### Література

1. Nemes J., Askew A.J. Mean and variance in network-design philosophies. *In Integrated Design of Hydrological Networks (Proceedings of the Budapest Symposium); International Association of Hydrological Sciences Publication*: Washington, DC, USA, 1986; P. 123–131.
2. Rodda J.C., Langbein W.B. Hydrological Network Design-Needs. *Problems and Approaches*; World Meteorological Organization: Geneva, Switzerland, 1969.
3. Shannon C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* 1948, 27, P. 379–423.
4. World Meteorological Organization. Casebook on Hydrological Network Design Practice. World Meteorological Organization: Geneva, Switzerland, 1972.
5. Mishra A.K., Coulibaly P. Developments in Hydrometric Network Design: A Review. *Rev. Geophys.* 2009.
6. Chacon-hurtado J.C., Alfonso L., Solomatine D.P. Rainfall and streamflow sensor network design: A review of applications, classification, and a proposed framework. *Hydrol. Earth Syst. Sci.* 2017, 21, P. 3071–3091.
7. Moss M.E. Concepts and Techniques in Hydrological Network Design. World Meteorological Organization: Geneva, Switzerland, 1982.
8. Van der Made J.W., Schilperoort T., van der Schaaf S., Buishand T.A., Brouwer G.K., van Duyvenbooden W., Becinsky P. Design Aspects of Hydrological Networks. *Commissie voor Hydrologisch Onderzoek TNO: The Hague*. The Netherlands, 1986.
9. Pyrcie R.S. Review and Analysis of Stream Gauge Networks for the Ontario Stream Gauge Rehabilitation Project. *2nd ed.; Watershed Science Centre*: Peterborough, ON, Canada, 2004.
10. Behmel S., Damour M., Ludwig R., Rodriguez M.J. Water quality monitoring strategies – A review and future perspectives. *Sci. Total Environ.* 2016, 571, P. 1312–1329.
11. Безсонний В.Л. Методика оцінки екологічного стану водойми на основі ентропійно зваженого індексу якості води. *Екологічні науки : науково-практичний журнал*, 2023. № 2(47). С. 44–48.
12. Pilon P.J., Yuzik T.R., Hale R.A., Day T.J. Challenges Facing Surface Water Monitoring in Canada. *Can. Water Resour. J.* 1996, 21, P. 157–164.
13. U.S. Geological Survey. Streamflow Information for the Next Century – A Plan for the National Streamflow Information Program of the U.S. *Geological Survey; U.S. Geological Survey*: Denver, CO, USA, 1999.
14. Langbein W.B. Overview of Conference on Hydrologic Data Networks. *Water Resour. Res.* 1979, 15, P. 1867–1871.
15. Davis D.R., Duckstein L., Krzysztofowicz R. The Worth of Hydrologic Data for Nonoptimal Decision Making. *Water Resour. Res.* 1999, 15, P. 1733–1742.
16. Shannon C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* 1948, 27, P. 379–423.
17. Batty M. Space, scale, and scaling in entropy maximizing. *Geogr. Anal.* 2010, 42, P. 395–421.
18. Singh V.P. Entropy Theory in Hydrologic Science and Engineering. *McGraw-Hill Education*. New York, NY, USA, 2015.
19. Lathi B.P. An Introduction to Random Signals and Communication Theory. *International Textbook Company*. Scranton, PA, USA, 1968.
20. Tribus M. Rational Descriptions, Decisions and Designs. *Pergamon Press*: Oxford, UK, 1969.
21. Mahjouri N., Kerachian R. Revising river water quality monitoring networks using discrete entropy theory: The Jajrood River experience. *Environ. Monit. Assess.* 2011, 175, P. 291–302.
22. Memarzadeh M., Mahjouri N., Kerachian R. Evaluating sampling locations in river water quality monitoring networks: Application of dynamic factor analysis and discrete entropy theory. *Environ. Earth Sci.* 2013, 70, P. 2577–2585.

23. Boroumand A., Rajaei T. Discrete entropy theory for optimal redesigning of salinity monitoring network in San Francisco bay. *Water Sci. Technol. Water Supply* 2017, 17, P. 606–612.
24. Lee J.H. Determination of optimal water quality monitoring points in sewer systems using entropy theory. *Entropy*. 2013, 15, P. 3419–3434.
25. Banik B.K., Alfonso L., di Cristo C., Leopardi A., Mynett A. Evaluation of Different Formulations to Optimally Locate Sensors in Sewer Systems. *J. Water Resour. Plan. Manag.* 2017, P. 143.
26. Alameddine I., Karmakar S., Qian S.S., Paerl H.W., Reckhow K.H., Optimizing an estuarine water quality monitoring program through an entropy-based hierarchical spatiotemporal Bayesian framework. *Water Resour. Res.* 2013, 49, P. 6933–6945.
27. Lee C., Paik K., Yoo D.G., Kim J.H. Efficient method for optimal placing of water quality monitoring stations for an ungauged basin. *J. Environ. Manag.* 2014, 132, P. 24–31.
28. Keum J., Coulibaly P. Information theory-based decision support system for integrated design of multi-variable hydrometric networks. *Water Resour. Res.* 2017, 53, P. 6239–6259.