

---

DOI: <https://doi.org/>

УДК: 004.043      СС BY-NC

**БАБАК О.В.**, канд. техн. наук,  
старш. наук. співроб., відд. екологічних цифрових систем  
ORCID: 0000-0002-7451-3314, e-mail: dep115@irtc.org.ua

**СУРОВЦЕВ І.В.**, д-р техн. наук, старш. наук. співроб.,  
зав. відд. екологічних цифрових систем  
ORCID: 0000-0003-1133-6207, e-mail: igorsur52@gmail.com  
Міжнародний науково-навчальний центр інформаційних технологій  
та систем НАН України та МОН України,  
пр. Акад. Глушкова, 40, Київ, 03187, Україна

## **РЕДУКЦІЯ ДАНИХ ЯК МЕТОД ІНТЕЛЕКТУАЛІЗАЦІЇ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ**

---

**Вступ.** Принципи теорії редукції у вигляді методів виявлення прихованих закономірностей у даних, підходів до перетворення систем координат та зниження розмірності вхідної інформації широко використовують в задачах ідентифікації та класифікації об'єктів, пошуку аномальних тенденцій у фінансовій діяльності, технічних системах вимірювання фізичних та хімічних параметрів.

**Мега статті** — огляд теоретичних та практичних результатів застосування теорії редукції в задачах і системах оброблення даних.

**Методи.** Розроблено методи виявлення прихованих закономірностей у даних, класифікації об'єктів за допомогою альфа-процедури, когнітивного моделювання екологічних даних, частотно-часової цифрової фільтрації, перетворення та моделювання багатокомпонентних сигналів, вимірювання концентрацій хімічних речовин в об'єктах довкілля та вмісту парникових газів в земній атмосфері, визначення навантажень на осі автомобіля за сигналом динамічного зважування, оцінювання витрат пилувугільного палива для економії коксу та природного газу доменної печі.

**Результати.** Розроблено принципи теорії редукції, високочутливу аналітичну систему «Аналізатор ІХП» вимірювання концентрацій 20 хімічних речовин у питній воді, харчових продуктах та навколишньому середовищі з використанням трьох нових імпульсних методів хронопотенціометрії для електрохімічного аналізу, інформаційну технологію «Атмосферні гази» визначення концентрацій 38 газів у земній атмосфері, комп'ютерні системи динамічного зважування автомобіля для митного вагового контролю транспортних засобів.

**Висновки.** Для класу задач з монотонними інтегральними даними вимірювання є необхідним перетворення системи координат, що дає змогу розглянути багатокомпонентний сигнал як суму окремих компонентів. Використання методів теорії редукції уможливило розроблення інтелектуальних інформаційних технологій екологічного моніторингу об'єктів біосфери, ефективних технічних систем вимірювання фізичних параметрів, виявлення шахрайських транзакцій у банківській системі.

**Ключові слова:** теорія редукції, приховані закономірності, класифікація об'єктів, перетворення системи координат, частотно-часова фільтрація, визначення концентрації.

© БАБАК О.В., СУРОВЦЕВ І.В., 2022

18      ISSN 2663-2586 (Online), ISSN 2663-2578 (Print). Cyb. and Comp. Eng. 2022. № 1 (207)

**ВСТУП**

Вперше достатньо обґрунтовану логіко-методологічну концепцію стосовно побудови теорії редукції запропонували в 1950-х роках прихильниками логічного емпіризму (Нагель Е., Фейгл Г. та ін.). Редукція (reduction — відсунення назад або повернення до колишнього стану) — логіко-методологічна процедура уявлення складного об'єкта як суми простих елементів, що робить його доступним для аналізу. Принцип редукції в цілому є зведенням складного до простого, вищого до нижчого (початкового), цілого до властивостей частин та частин до специфіки цілого [1].

У науці процес редукції розглядають як метод перетворення даних, пов'язаний з розв'язанням того чи іншого наукового завдання з метою її спрощення. Тому він є невід'ємною частиною основних методів наукового пізнання: ідеалізації, абстракції та моделювання. Особливе значення редукція інформації набуває у вирішенні проблеми великих даних.

На наш погляд, реалізація принципів редукції може сприяти створенню ефективних інтелектуальних інформаційних технологій екологічного моніторингу, в якому до позитивних факторів належать результати вимірювання концентрацій хімічних речовин та атмосферних газів. Розроблення технологій, в яких результати одержують на малих вибірках даних, коли відношення обсягу вибірки до кількості факторів менше за десять, є актуальною науково-технічною проблемою.

**ПОСТАНОВКА ПРОБЛЕМИ**

Будемо виходити з того, що рішенням задачі класифікації образів, ідентифікації об'єктів і непрямих вимірювань фізико-хімічних параметрів є відновлення наявної залежності [2]. При цьому результати таких вимірювань звичайно входять до факторів перших двох задач.

Нехай досліджуваний об'єкт характеризується деякою випадковою вибіркою даних, одержаних у результаті пасивного експерименту

$$\{x_{ij}, y_i\} \quad i = \overline{1, l} \quad j = \overline{1, n} \quad \forall x_{ij} \neq 0, \quad (1)$$

де  $x_{ij}$  — фактори,  $y_i$  — спостереження,  $l$  — кількість спостережень,  $n$  — кількість факторів.

Якщо під час ідентифікації  $y_i$  є натуральними числами, то внаслідок класифікації отримуємо  $y = (+1, -1)$ . Рішення вказаної задачі будемо шукати у вигляді множинної лінійної регресії

$$y = a_0 + \sum_{j=1}^n a_j x_j, \quad (2)$$

де  $a_0, a_j$  — оцінки коефіцієнтів.

У всіх випадках розв'язок задачі, тобто визначення вектору оцінки коефіцієнтів моделі  $a$ , будемо шукати як розв'язання оптимізаційної задачі

$$a = \arg \min_a \left( \sum_{i=1}^l (a_i x + a_0 - y_i)^2 \right). \quad (3)$$

Водночас намагання за даними одержати оптимальне наближення до точного опису об'єкта нерозривно пов'язано зі зменшенням тим чи іншим способом неповноти знань про цей об'єкт, отже, завжди зі зниженням розмірності початкової задачі  $n$  до значення  $n'$  одержимо  $n' < n$ , тобто, в якомусь розумінні редукцію вхідної інформації. Вказану процедуру будемо називати виявленням прихованих закономірностей у вхідних даних.

Потрібно за даними (1) встановити приховані фактори розмірністю  $n'$  та вирішити оптимізаційну задачу (2), тобто знайти вектор  $a'$  оцінки коефіцієнтів моделі прихованої закономірності.

**Мета роботи.** З врахуванням особливостей вирішення проблеми редукції даних, привести у стислому вигляді одержані авторами найвагоміші результати щодо розв'язання задач класифікації та ідентифікації об'єктів, а також у разі непрямих вимірювань. Показати, що ці результати можуть бути використані у процесі інтелектуалізації інформаційних технологій для практичного застосування.

### ОСОБЛИВОСТІ ПРОБЛЕМИ РЕДУКЦІЇ ДАНИХ

Одразу відмітимо, що принцип редукції у моделюванні лежить в основі аналізу даних. Для вирішення задачі виявлення закономірностей майже завжди виникає зайвий клопіт, пов'язаний з наявною мультиколінійністю факторів, тобто їх кореляцією, яка обтяжує аналіз оцінки загального результату. У цьому випадку регресійна модель стає непридатною для прогнозування.

Редукцію даних (Data reduction) використовують у двох різних за змістом задачах: зменшення розмірності або обсягу даних. Або інакше, зведення даних зі складною незручною формою до простішої форми їхнього подання.

У першому випадку основним методом зменшення розмірності даних  $x_j$  є факторний аналіз та виділення основних компонент. Сутність факторного аналізу полягає у поданні вхідних факторів  $x_j, j = \overline{1, n}$  як деяку сукупність латентних факторів  $x'_j, j = \overline{1, n'}$ , де  $n' \ll n$ .

У результаті формується оптимальний простір нових ортогональних (взаємно некорельованих) змінних без істотних втрат змістовної інформації вхідних даних. При цьому, в основу методу головних компонент покладено, що фактори  $x$  є лінійною комбінацією вхідних показників  $x'$

$$x_j = \sum_{k=1}^{n'} a_{kj} x'_{kj},$$

де  $x'$  — головні компоненти,  $a_{kj}$  — факторні навантаження.

Отже, по суті вирішується задача усунення мультиколінійності факторів у вхідній виборці  $X$ , коли  $\det(x^T x) \approx 0$ .

Не зупиняючись детально на виділенні головних компонентів, відзначимо, що всеж є деяка втрата змістовної інформації. Цей недолік повністю усувається, якщо є можливість ортогонального планування експерименту [3]. На жаль, останнє для більшості випадків одержання початкових даних є практично недосяжним і може бути здійснено тільки на евристичній основі.

У другому випадку поширеним є метод переведення результатів вимірювання з одної системи відліку в іншу. При цьому у новій системі може підвищитись точність відліку або скоротитись обсяг даних.

Цілком очевидно, що перший та другий випадок редукції даних методологічно є перетворенням багатомірної системи координат, завдяки якому можливо зменшення розмірності простору або змінення координатних осей, а це може привести до економії обчислювальних ресурсів та підвищення якості моделювання. Більше того, ефективно досягається головна мета комп'ютерного оброблення експериментальних та статистичних даних — машинне навчання (Machine Learning), яке полягає в автоматичному виявленні прихованих закономірностей. Ці закономірності та здобуті з них знання про об'єкт дають змогу зрозуміти сутність процесу і за наявними даними передбачити нові факти та одержати нові знання [4].

Інакше кажучи, у першому випадку розв'язуванню задачі моделювання повинна передувати редукція інформації. Для цього у початковому просторі великої розмірності  $n$  задачу синтезу моделі заміняють на задачу синтезу простору малої розмірності  $n'$ , в якому цю задачу розв'язують як просту, найчастіше лінійну функцію (2).

Якщо редукція полягає у зміні координатних осей, то звичайним розв'язком є зворотна функція, тобто

$$y = f(x) \rightarrow x = \varphi(y). \quad (4)$$

Такий підхід лежить в основі розв'язування цілого ряду задач непрямих вимірювань.

### ЗАДАЧА КЛАСИФІКАЦІЇ ОБ'ЄКТІВ

Різновидом реалізації принципу редукції у розв'язуванні задачі класифікації об'єктів у проблемі розпізнавання образів є запропонована Васильєвим В.І. альфа-процедура [5, 6]. Не вдаючись до подробиць, розглянемо тільки ідею реалізації цієї процедури.

Нехай задано навчальну вибірку ознак  $P_j$ ,  $j = \overline{1, n}$  обсягом  $l$  об'єктів і в ній образи  $V_1$  та  $V_2$ . Спочатку здійснюється попередній відбір і залишаються для подальшого оброблення тільки  $m \leq n$  ознак  $P_j$ , для яких можна знайти два граничних значення  $x_{1j}$  та  $x_{2j}$ , що розділяють вісь значень  $P_j$  на два сегмента  $V_1$ ,  $V_2$  та сегмент їхнього перетинання.

Вводиться поняття розродільної сили для окремої ознаки як  $F(P_j) = w_j / l$ , де  $w_j$  — кількість об'єктів, розташованих зовні сегменту перетинання ознак  $P_j$ .

Алгоритм синтезу оптимального простору  $m'$  полягає у тому, що крок за кроком вибираються ознаки, які мають найкращу розродільну силу. Кожна нова ознака редукує сегмент перетинання та збільшує кількість правильно класифікованих об'єктів. Застосовується таке визначення розродільної сили ознак на  $k$ -му кроці селекції:

$$F(P_{jk}) = \frac{w_k - w_{k-1}}{l} = \frac{\Delta w_k}{l},$$

де  $w_{k-1}$  — кількість правильно класифікованих об'єктів перед вибором  $k$ -ої ознаки,  $w_k$  — кількість правильно класифікованих об'єктів після вибору  $k$ -ої ознаки.

На першому кроці селекції вибирається ознака з найкращою розродільною силою як базова ознака  $f_0$  і надається разом з її числовими значеннями для усіх об'єктів як вісь.

Наступним кроком до системи координат додається друга ознака  $p_k$  і визначаються об'єкти на площині з осями  $f_0$  та  $p_k$ . Далі у цій площині формується нова вісь, яка обертається навколо початку системи координат на величину кута альфа доти, поки проєкції об'єктів на неї не дадуть найкращого поділу.

Ця процедура повторюється для усіх ознак, що залишились, і з них вибирається та, що дає найкращий поділ на відповідній осі  $f_1$ , яка у подальшому застосовується як наступна штучна ознака.

На третьому кроці додається наступна ознака  $p_j$  як третя вісь, визначаються позиції усіх об'єктів на площині з осями  $f_1, p_j$  та шляхом обертання вибирається вісь  $f_2$  з найкращим поділом об'єктів. Цей процес повторюється до повного поділу об'єктів у просторі  $m'$ . Площину, яка є перпендикулярною останньому базовому напрямку  $f_j$  та проходить через оптимальну її межу, приймають як вирішувальне правило. Якщо поділ не досягнуто, то для зупинки процесу застосовують спеціальний критерій [7].

Отже, розглянутий алгоритм альфа-процедури послідовно перетворює початкову систему координат, завдяки чому усувається мультиколінійність ознак у просторі  $m'$ . При цьому досягається висока точність виявлення прихованих закономірностей у даних, що підтверджується чисельними комп'ютерними експериментами. Такий підхід можна порівняти з широко відомим методом опорних векторів Вапника В.Н. [8]. Одну особливість застосування запропонованого підходу розглянуто у [9].

Застосування альфа-процедури дало змогу розв'язати задачу класифікації об'єктів — виявлення підприємств або банків банкрутів, що високо оцінено Податковою службою України.

### ЗАДАЧА ІДЕНТИФІКАЦІЇ ОБ'ЄКТІВ

Як вже було вказано, усунення мультиколінійності факторів  $x$  є найважливою умовою виявлення закономірностей прихованих у даних. При цьому оцінки коефіцієнтів моделі (1) стають стійкими та усувається можливість застосування для них неправильного знаку.

Методи усунення мультиколінійності передбачають попереднє центрування вхідних даних, яке полягає у тому, що для кожного значення обчислюється середня величина в колонці даних:

$$x'_j = x_j - \bar{x}_j, j = \overline{1, n}.$$

Цей простий прийом дає достатньо хороші результати, оскільки надає можливість побудувати гіперповерхні умов методу найменших квадратів (МНК) таким чином, що кути між ними будуть перпендикулярними. У разі однорідних даних можна вважати, що вони стають наче ортогональними один одному.

Це є першим кроком алгоритму виявлення закономірностей.

Другий крок полягає у відновленні залежності (2) за вхідного значення  $n$ . Отже, стають відомими коефіцієнти моделі  $a_j$ , які є оцінками складників градієнта лінійної функції (2). За аналогією з алгоритмом розв'язання задачі класифікації, вводиться поняття сили впливу значень коефіцієнтів  $F(x_j)$  на залежну змінну  $y$  [10].

На третьому кроці здійснюється ранжування за значенням коефіцієнтів  $a_j$ :

$$|a_{z1}| > |a_{z2}| > \dots > |a_{zn}|, z \in \overline{1, \dots, n}, \quad (5)$$

та одержання відповідного ряду  $x_j$  коефіцієнтів  $F(x_j)$ :

$$x_{z1}, x_{z2}, \dots, x_{zn}, z \in 1, \dots, n.$$

Четвертий крок полягає у використанні стратегії відбору позиції пошукової моделі регресії у новому просторі  $n' \leq n$  (3) за деяким визначеним критерієм, наприклад, значенням залишкової суми квадратів (RSS) [11]:

$$y = a_0 + a_1 x_{z1};$$

$$y = a_0 + a_1 x_{z1} + a_2 x_{z2};$$

...

В окремих випадках четвертий крок може бути обмежено тільки початковими позиціями, наприклад, значенням  $z = 2$ . Таке виникає, якщо попередній аналіз (5) показав, що, наприклад, перші два значення оцінок коефіцієнтів  $a_z$  значною мірою перевищують значення інших. Подібна ситуація виникла під час створення моделі непрямих вимірювань витрат пилоугільного палива (ПВП), що вдувається у фурми доменної печі. Проведені дослідження показали, що тільки два фактори з шести: витрата транспортного повітря та перепад тиску на пилопровідах, є головними, а іншими можна знехтувати. Це дало змогу розробити та впровадити відповідну систему непрямих вимірювань ПВП, яка сприяла істотній економії коксу та природного газу [11, 13].

### ЗАДАЧІ НЕПРЯМИХ ВИМІРЮВАНЬ

Серед безлічі непрямих вимірювань фізико-хімічних величин розглянемо клас задач, який в якості корисних (фільтрованих) початкових даних  $y_0(t)$  використовує монотонні зростаючі (спадаючі) значення вимірювання багатокomпонентного процесу. Цей клас задач цікавий тим, що перед побудовою моделей процесу застосовується редуція даних, яка полягає у зміні координатних осей з використанням зворотної функції (4). До цього класу задач можна віднести: визначення концентрацій токсичних речовин в об'єктах довкілля електрохімічними методами інверсійної хронопотенціометрії; визначення навантажень на осі автомобіля за сигналом динамічного зважування під час заїзду транспортного засобу на платформу статичних ваг; визначення маси компонентів у разі неперервного дозування бетонних сумішів або комбікормів та інші задачі. Розглянемо алгоритм редуції даних для цього класу задач детальніше.

Перетворення системи координат дає змогу розглянути інтегральний сигнал  $y(t)$ , спотворений високочастотною завадою  $\xi$ :  $y_0(t) = y(t) + \xi(t)$ , як суму окремих компонентів  $\varphi(y)$  у багатокomпонентному диференційному сигналі інтенсивності  $\tau(y)$ :

$$(t, y) \rightarrow (y, t) \rightarrow (y, dt / dy).$$

Надамо алгоритм редуції початкових даних для цього класу задач:

- відновлення корисного монотонного сигналу  $y(t)$ ;
- формування значень перетвореного сигналу  $t(y)$ ;
- диференціювання перетвореного сигналу  $t(y)$  у сигнал інтенсивності  $\tau(y) = dt / dy$ ;

— ідентифікація окремих компонентів  $\varphi(y)$  у багатокомпонентному сигналі  $\tau(y)$ ;

— визначення пошукових величин компонентів  $\varphi(y)$ .

**Відновлення корисного монотонно зростаючого сигналу.** Дослідження показали, що для таких вхідних даних застосування класичних методів фільтрації гармонічних сигналів не завжди ефективне, бо призводить до порушення форми та інтегральних характеристик у місцях різкої зміни кривизни сигналу, що впливає на точність визначення пошукових величин компонентів.

Одним зі шляхів вирішення проблеми відновлення негармонічних сигналів, спотворених високочастотними завадами, є використання методу частотно-часової фільтрації, оснований на використанні внутрішніх характеристик сигналу: точок екстремуму  $y'(t) = 0$ , перегину  $y''(t) = 0$  та інтервалів часу між ними, які достатньо повно відображають властивості високочастотних перешкод у сигналі.

Метод частотно-часової фільтрації базується на використанні ідей способу визначення надшвидкого спектру аналогового сигналу, запропонованого Скуріхіним В.І., Цепковим Г.В., Пономарьовою І.Д. та ін. [15, 16].

Теоретичною основою методу є адаптивний базис, придатний для аналізу сигналів будь-якого класу, який не застосовує операцій множення. Адаптивність його досягається формуванням системи ортогональних базисних функцій за параметрами екстремумів самого сигналу  $y(t)$ . Приклади застосування методу частотно-часової фільтрації для відновлення вхідних сигналів визначення концентрацій токсичних речовин від високочастотних завад наведено у [17].

**Формування значень перетвореного сигналу.** Після виконання фільтрації високочастотних завад дані  $y(t)$  стають корисним сигналом, що монотонно змінюється, з постійним кроком дискретизації за часом  $\Delta t$ .

Для формування значень перетвореного сигналу  $t(y)$  задаємо крок дискретизації з амплітудою  $\Delta y$ , за допомогою лінійної інтерполяції або ермітових кубічних сплайнів визначаємо значення часу  $t_j$ , які відповідають амплітудам  $y_j = j \Delta y$  в діапазоні зміни амплітуд  $y_{\min} \leq y_j \leq y_{\max}$ .

**Диференціювання перетвореного сигналу у сигнал інтенсивності.** Виконуємо диференціювання перетвореного сигналу  $t(y)$  у сигнал інтенсивності  $\tau(y)$ , застосовуючи центральну різницю

$$\tau(y) = dt / dy = (t_{j+1} - t_j) / \Delta y. \quad (6)$$

Описане диференціювання (6) є взаємно однозначним, що дає змогу відновити сигнал  $t(y)$  шляхом зворотного інтегрування  $\tau(y)$ . Наведемо приклади застосування перетворення системи координат вимірювання багатокомпонентних сигналів двох фізико-хімічних процесів.

*Електрохімічне визначення концентрацій хімічних елементів.* Розглянемо на рис. 1 перетворення системи координат для сигналу вимірювання потенціалів інверсії у часі  $e(t)$  для визначення концентрацій хімічних елементів (цинк, кадмій, свинець, мідь).

Інтегральна сума вимірювань багатокомпонентного процесу інверсії  $e(t)$  (рис. 1.а) перетворюється у суму компонентів сигналу інтенсивності  $\tau(e)$  (рис. 1.в).

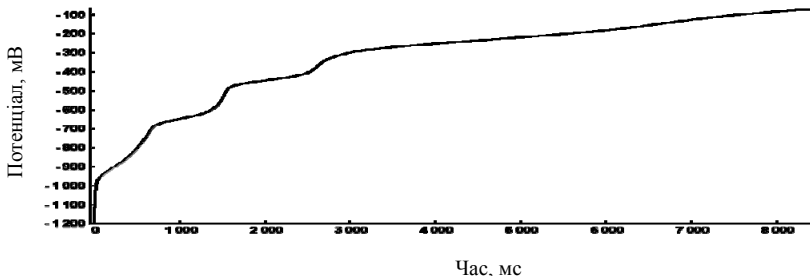
Ідентифікація хімічного елементу здійснюється за потенціалом піку (точка максимуму) та діапазоном вимірювання компонента  $\varphi(e)$  (точки

мінімуму  $e_1, e_2$ ). Виділення окремих компонентів  $\varphi(e)$ , що належать елементу в багатокомпонентному сигналі, виконується за математичною моделлю сигналу інтенсивності  $\tau(e)$  шляхом розв'язання задачі параметричної ідентифікації [17].

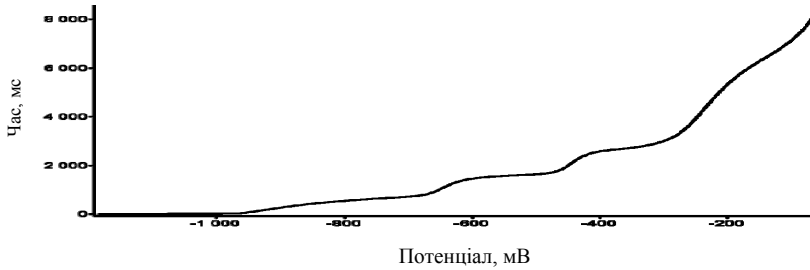
Час інверсії хімічного елемента  $T$ , пропорційний концентрації хімічного елемента, визначається за сигналом інтенсивності  $\tau(e)$  як площа компоненту  $\varphi(e)$  у діапазоні  $[e_1, e_2]$ :

$$T = \int_{e_1}^{e_2} \tau(e) de = \sum_{i=e_1}^{e_2} \tau(e_i) . \quad (7)$$

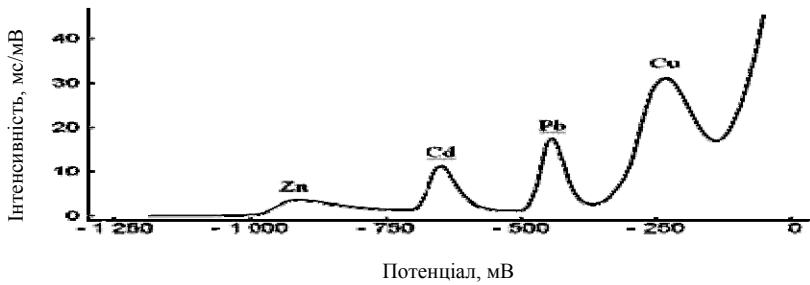
За часом інверсії компонента  $T$ , визначеного в сигналах вимірювання розчинів фону, проби об'єкта та проби з добавкою відомої маси стандартного зразка іонів елемента, визначається концентрація хімічного елемента в об'єкті навколишнього середовища.



а) фільтрований сигнал вимірювання концентрацій  $e(t)$



б) перетворений сигнал  $t(e)$



в) сигнал інтенсивності  $\tau(e)$

Рис. 1. Перетворення структури даних сигналу вимірювання концентрацій



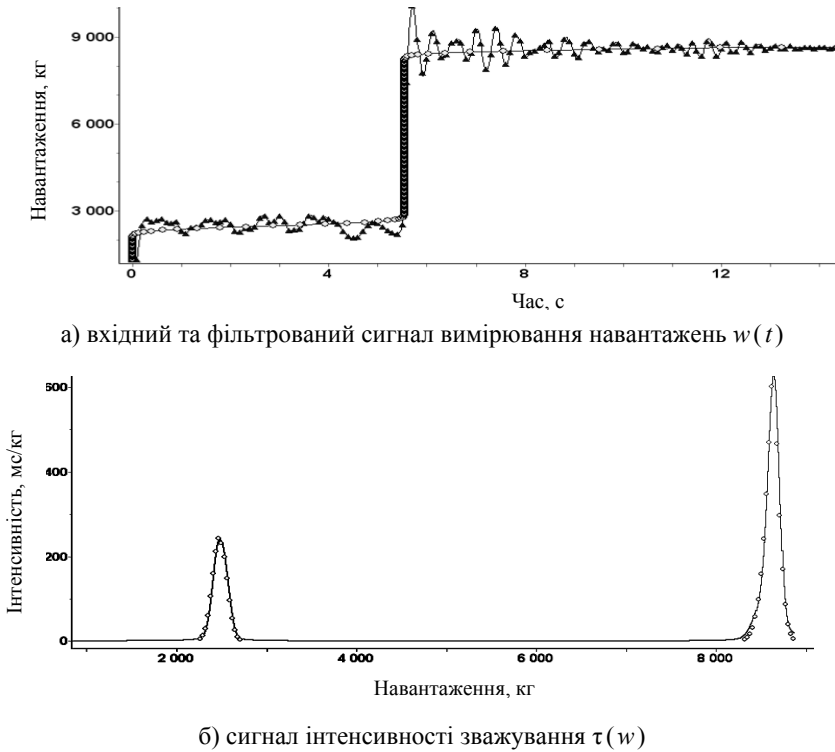


Рис. 2. Перетворення структури даних сигналу динамічного зважування автомобіля

*Динамічне зважування автомобіля.* Застосування редукції вхідних даних дало змогу вирішити важливу технічну проблему визначення навантажень на осі автомобіля за сигналом динамічного зважування під час заїзду на платформу автомобільних ваг [18] для здійснення митного вагового контролю транспортних засобів у Міжнародних автомобільних пунктах пропуску.

Розглянемо на рис. 2 перетворення системи координат для сигналу вимірювання навантажень  $w(t)$  двохосного автомобіля, який одержують з тензометричних давачів під час заїзду транспортного засобу на платформу статичних ваг. Одержані дані є монотонно зростаючим інтегральним сигналом багатокомпонентного поосного зважування, який спотворено високочастотними завадами. Корисний сигнал зважування  $w(t)$  відновлюємо методом частотно-часової фільтрації.

Інтегральна сума вимірювання навантажень багатокомпонентного процесу зважування  $w(t)$  перетворилась у суму компонентів сигналу інтенсивності  $\tau(w)$ .

Ідентифікація компонента навантаження на  $j$  вісь автомобіля здійснюється за абсцисою піку  $w_j$  (точка максимуму компоненту).

Рівень навантаження на вісь автомобіля  $G_j$  обчислюємо як різницю між сусідніми значеннями параметрів  $w_j$  :

$$G_j = w_j - w_{j-1}, \tag{8}$$

контролюючи значення поосних навантажень за загальною масою, визначеною у статичному режимі зважування автомобіля.

Наведені приклади перетворення системи координат багатокомпонентних сигналів фізико-хімічних процесів підтверджують ефективність застосування редукції даних вимірювання для розв’язання технічних завдань.

## ПРИЛАДИ ТА СИСТЕМИ ВИМІРЮВАННЯ

За ініціативи академіка Скуріхіна В.І. у 1987 р. розроблено перший в Україні прилад вимірювання концентрацій важких металів «Аналіз ТМ» з використанням електрохімічного методу інверсійної хронопотенціометрії, запропонованого хіміками Національного університету біоресурсів і природокористування України. Надалі було розроблено пристрої: «Хронопольярограф», «МА-1020», «МХА1000» та аналізатор «М-ХА1000-5», які визначали концентрації п'яти важких металів (свинець, кадмій, цинк, мідь, олово) у харчових продуктах та воді з чутливістю вимірювання концентрацій 1 мікрограм на літр. Усього було вироблено понад 350 приладів, які активно застосовувались в роботі санітарно-епідеміологічної служби МОЗ України, в лабораторіях якості продукції, в наукових та учбових лабораторіях. Але ці прилади не дали змогу визначати такі елементи, як ртуть, миш'як та інші, які нормуються для визначення якості питної води та харчових продуктів.

Було розроблено нову інтелектуальну інформаційну технологію та високочутливу аналітичну систему «Аналізатор ІХП» для визначення екологічного стану об'єктів навколишнього середовища, характеристики якої перевищують міжнародні стандарти високого рівня. Вдалося підвищити чутливість вимірювання концентрацій хімічних елементів до 0,01 мікрограма на літр завдяки застосуванню вперше в світовій практиці трьох імпульсних методів електрохімічного аналізу: імпульсної інверсійної хронопотенціометрії (pulse stripping chronopotentiometry, PSCP); диференційної імпульсної хронопотенціометрії (differential pulse stripping chronopotentiometry, DPSCP); окислювальної імпульсної хронопотенціометрії (oxidative pulse stripping chronopotentiometry, OPSCP). Наукову новизну запропонованих методів підтверджено патентами на винахід [20, 21].

Інформаційна технологія використовує нові цифрові методи оброблення: спосіб перетворення монотонних вхідних даних, методи високочастотної фільтрації та адаптивного згладжування, побудова генеративної моделі багатокomпонентного сигналу [17]. Додання до приладу функцій іономіра дало змогу розширити можливості аналітичної системи [22].

«Аналізатор ІХП» може вимірювати концентрації 14 токсичних елементів методами інверсійної хронопотенціометрії (ртуть, миш'як, свинець, кадмій, цинк, мідь, олово, нікель, кобальт, залізо, марганець, селен, йод, хром) та 6 хімічних речовин за розробленим методом хроноіонометрії (калій, натрій, кальцій, фтор, амоній та нітрати), що дає змогу визначати екологічний стан різних об'єктів довкілля, якість питної води та харчових продуктів. Систему «Аналізатор ІХП» впроваджено в навчальну та наукову діяльність НТУУ «КПІ ім. Ігоря Сікорського» та в Харчову контрольно-виробничу лабораторію Київської обласної спілки споживчих товариств.

Для вирішення важливої екологічної проблеми оцінювання стану забруднення земної атмосфери розроблено інформаційну технологію «Атмосферні гази», яка, використовуючи спектр пропускання атмосферного повітря, одержаного інфрачервоним Фур'є-спектрометром (розробка Інституту фізики напівпровідників ім. В.Є. Лашкарьова НАН України), дає змогу виконувати дистанційне зондування газових викидів промислових підприємств в наземних комплексах, з борту літака або космічного апарату. Після перетворення системи координат вхідного спектру, з використанням банку

даних спектральних ліній газів та ітераційної моделі ідентифікується наявність та визначається концентрація 38 газів атмосфери. Особливу увагу приділяють визначенню концентрації озону, метану, вуглекислого та сірчаного газів, які сприяють виникненню парникового ефекту і негативно впливають на клімат Землі та здоров'я людей. Інформаційну технологію «Атмосферні гази» впроваджено та застосовують в Головній астрономічній обсерваторії НАН України [23].

### **КОМП'ЮТЕРНО-ТЕЛЕКОМУНІКАЦІЙНА ПЛАТФОРМА**

Розроблені методи теорії редукції, передові технології зв'язку 5G, 6G та широке застосування пристроїв Інтернету речей в системах екологічного моніторингу спонукають до побудови сенсорних систем з використанням підключення через Ethernet, WiFi та контролери IOT [24, 25]. Для цього розробляється інформаційна технологія експрес-аналізу екологічної якості питної води з використанням переносних, підключених до хмарних сервісів приладів електрохімічного вимірювання концентрації токсичних елементів. Кінцевим напрямом цих досліджень буде розроблення комп'ютерно-телекомунікаційної платформи для завдань екологічного моніторингу та штучного інтелекту з використанням хмарних та передових технологій зв'язку.

Наразі зусилля співробітників відділу екологічних цифрових системи спрямовано на застосування методів когнітивного моделювання для виконання екологічних завдань: оцінювання стану забруднення ділянок поля важкими металами, детоксикації та відновлювання екологічної якості ґрунтів у точному землеробстві; виявлення небезпечних для здоров'я людини екстремальних чинників у різних об'єктах довкілля для профілактики та прогнозування ризиків соціально значимих захворювань населення; розроблення інформаційних технологій вирощування статистичного обліку продуктів виробництва для фермерських господарств тощо [26].

### **ВИСНОВКИ**

Реалізація принципів теорії редукції уможливило розв'язання важливих практичних завдань класифікації образів, ідентифікації об'єктів, пошуку аномальних тенденцій у фінансовій діяльності підприємств, виявлення шахрайських транзакцій у банківській системі та прихованих закономірностей у великих даних тощо.

Для класу задач з монотонними інтегральними даними вимірювань необхідно застосовувати перетворення системи координат, що дає змогу розглянути багатоконпонентний сигнал як суму окремих компонентів. Такий підхід дав можливість розробити: високочутливу аналітичну систему «Аналізатор ІХП» для визначення екологічного стану об'єктів навколишнього середовища; систему визначення навантажень на осі автомобіля за динамічним сигналом зважування для здійснення митного контролю; інформаційну технологію «Атмосферні гази» визначення вмісту парникових газів в земній атмосфері для дистанційного зондування газових викидів промислових підприємств.

Розглянуті в статті принципи теорії редукції, оригінальні результати розв'язання задач класифікації, ідентифікації та непрямих вимірювань фізико-хімічних параметрів сприятимуть інтелектуалізації інформаційних технологій та систем, особливо для екологічного моніторингу об'єктів біосфери.

## ЛІТЕРАТУРА

1. Ивин А.А., Медведева И.А., Берштейн В.Л. Редукция. Гуманитарный квартал: Концепты. *Центр гуманитарных технологий*, 2002–2021 (Дата звернення: 29.12.2021). URL: <https://market.ru/concepts/6892>
2. Алгоритмы и программы восстановления зависимостей / Под ред. В.Н. Вапника. М.: Наука, 1984. 816 с.
3. Адлер Ю.П., Маркова Е.В., Грановский Ю.В. Планирование экспериментов при поиске оптимальных условий. М.: Наука, 1976. 277 с.
4. Загоруйко Н.Г. Методы анализа данных и знаний. Новосибирск ИМ СО РАН, 1999. 270 с.
5. Vasilyev V.I. The reduction principle in pattern recognition learning (RPL). *Problem. Pattern recognition and Image Analysis*. 1996. N 1. P.23–52.
6. Васильев В.И. Индукция и редукция в проблемах экстраполяции. *Кибернетика и вычисл. техника*. 1998. Вып. 116. с. 65–81.
7. Васильев В.И., Суровцев И.В. Индуктивные методы обнаружения закономерностей, основанные на теории редукции. *Управляющие системы и машины*. 1998. № 5. С. 3–13.
8. Vapnik V. *Statistical learning theory*. New York: John-Wiley Sans. Inc., 1998. 286 p.
9. Бабак О.В. Об одном подходе к оптимизации решения задач обучения распознаванию образов на основе метода опорных векторов. *Кибернетика и системный анализ*. 2004. № 2. С. 179–185.
10. Бабак О.В. Об одном принципе самоорганизации математических моделей. *Проблемы управления и информатики*. 2001. № 2. С. 98–107.
11. Бабак О.В., Васильев В.И., Чернов В.А. и др. Особенности измерения расхода пилеугольного топлива при вдувании его в фурмы доменной печи. *Сталь*. 1988. № 1. С. 18–20.
12. Пристрій для комбінаторного моделювання фізичних об'єктів: патент на винахід 124909, Україна : МПК (2006). G05B 17/00, G06G 7/48 (2006.01). а202003852; заявл. 26.06.2020; опубл. 08.12.2021, Бюл. № 49.
13. Пристрій регулювання витрати пилевогільного палива : пат. 82758 Україна : МПК C21B 7/24. а200608430; заявл. 27.07.2006; опубл. 12.05.2008, Бюл. № 9.
14. Пристрій виміру витрати пилевогільного палива : пат. 83106 Україна : МПК G01F 9/00, G01F 25/00, G05B 17/00. а200608431; заявл. 27.07.2006; 10.06.2008, Бюл. № 11.
15. Способ определения спектра аналогового сигнала: А. С. 845600 СССР. / Скурихин В.И., Пономарева И.Д., Сиверский П.М., Цепков Г.В. 1981.
16. Гриценко В.І., Скурихин В.І., Цепков Г.В. Інформаційні технології цифрової обробки сигналів: нові підходи і перспективи впровадження. *Вісник НАН України*. 2005. № 12. С. 33–41.
17. Суровцев І.В., Галімов С.К., Татарінов О.Е. Інформаційна технологія визначення концентрації токсичних елементів в об'єктах навколишнього середовища. *Кибернетика и вычисл. техника*. 2018. № 191. С. 5–31.
18. Система поосового зважування на платформних вагах : пат. 106013 Україна : МПК G01G 19/02. а201309799; заявл. 06.08.2013; опубл. 10.07.2014, Бюл. №13.
19. Суровцев И.В. Новая информационная технология поосного взвешивания автомобилей на платформенных весах. *Управляющие системы и машины*. 2015. №3. С. 77–81.
20. Пристрій для вимірювання концентрації токсичних елементів : пат. 107412 Україна : МПК G01N 27/48. а201306295; заявл. 21.05.2013; опубл. 25.12.2014. Бюл. №24.
21. Пристрій для вимірювання концентрації хімічних елементів методами імпульсної хронопотенціометрії: пат. 123459, Україна : МПК G01N 27/48 (2006.01). а201902429; заявл. 12.03.2019; опубл. 07.04.2021, Бюл. №14.
22. Surovtsev I.V., Velykyi P.Y., Galimova V.M., Sarkisova M.V. Ionometric method for determination of concentrations of microelements in research of digital medicine. *Cyb. and comp. eng.* 2020. №. 4 (202). С. 25–43.
23. Венгер Е.Ф., Бабак О.В., Суровцев И.В. и др. Алгоритм оперативного косвенного измерения концентрации газовых примесей в атмосфере по спектрам Фурье-спектрометра. *Управляющие системы и машины*. 2007. № 2. С. 33–38, 81.
24. Гриценко В.І., Суровцев І.В., Бабак О.В. Система бездротового зв'язку 5G. *Кибернетика и вычисл. техника*. 2019. № 3 (197). С. 5–19.
25. Гриценко В.І., Бабак О.В., Суровцев І.В. Особливості взаємозв'язку мереж 5G, 6G з великими даними, інтернетом речей та штучним інтелектом. *Cyb. and comp. eng.* 2021. №. 2 (204). С. 5–18.

26. Пристрій для передпосівної обробки насіння : пат. 122085, Україна : МПК (2020.01) A01C 1/00, A01G 7/04 (2006.01). а201809456; заявл. 19.09.2018; опубл. 10.09.2020, Бюл. № 17.

Отримано 8.02.2022

#### REFERENCES

1. Ivin A.A., Medvedeva I.A., Bershtein V.L. Reduction. Humanitarian Quarter: Concepts. *Center for Humanitarian Technologies*, 2002–2021 (Last accessed: 12/29/2021), URL: <https://market.ru/concepts/6892> (in Russian).
2. Algorithms and programs for recovering dependencies / Ed. V.N. Vapnik. M.: *Nauka*, 1984, 81p. (in Russian).
3. Adler Yu.P., Markova E.V., Granovsky Yu.V. Planning experiments in search of optimal conditions. M.: *Nauka*, 1976, 277 p. (in Russian).
4. Zagoruiko N.G. Methods of data and knowledge analysis. Novosibirsk *IM SO RAN*. 1999, 270p. (in Russian).
5. Vasilyev V.I. The reduction principle in pattern recognition learning (RPL). *Problem. Pattern recognition and Image Analysis*. 1996, No 1, pp. 23–52.
6. Vasilyev V.I. Induction and reduction in extrapolation problems. *Kibernetika i vychislitel'naâ tehnika*. 1998, Vol. 116, pp.65–81 (in Russian).
7. Vasilyev V.I., Surovtsev I.V. Inductive methods for pattern detection based on reduction theory. *Control System and Computers*. 1998, No 5, pp. 3–13 (in Russian).
8. Vapnik V. Statistical learning theory. New York: *John-Wiley Sans. Inc.*, 1998, 286.
9. Babak O.V. On one approach to optimizing the solution of problems of learning pattern recognition based on the support vector machine method. *Cybernetics and System Analysis*. 2004, No 2, pp. 179–185 (in Russian).
10. Babak O.V. On one principle of self-organization of mathematical models. *Problems of control and informatics*. 2001, No 2, pp. 98–107 (in Russian).
11. Babak O.V., Vasilyev V.I., Chernov V.A. Peculiarities of measuring the consumption of pilecoal fuel when it is blown into the blast furnace tuyeres. *Steel*. 1988, No 1, pp.18–20 (in Russian).
12. Device for combinatorial modeling of physical objects: patent 124909, Ukraine: IPC (2006). G05B 17/00, G06G 7/48 (2006.01). а202003852; claimed 26.06.2020; published 08.12.2021 (in Ukrainian).
13. Device for adjusting of pulverized coal fuel consumption: patent 82758 Ukraine: IPC C21B 7/24/. а200608430; claimed 27.07.2006; published 12.05.2008 (in Ukrainian).
14. Method for measurement of flow rate of dust-coal fuel: patent 83106 Ukraine: IPC G01F 9/00, G01F 25/00, G05B 17/00. а200608431; claimed 27.07.2006; published 10.06.2008 (in Ukrainian).
15. Method for determining the spectrum of an analog signal: A.S. 845600 USSR. 1981. (in Russian).
16. Gritsenko VI, Skurikhin VI, Tsepkov GV Information technologies of digital signal processing: new approaches and prospects of implementation. *Visnyk NAS of Ukraine*. 2005, No 12, pp. 33–41 (in Ukrainian).
17. Surovtsev I.V., Galimov S.K., Tatarinov O.E. Information technology for determining the concentration of toxic elements in environmental objects. *Kibernetika i vychislitel'naâ tehnika*. 2018, No.1(191), pp. 5–33. DOI: <https://doi.org/10.15407/kvt191.01.005> (in Ukrainian).
18. System for axle-by-axle weighing on platform scales : patent 106013 Ukraine: IPC G01G 19/02, а201309799; claimed 06.08.2013; published 10.07.2014 (in Ukrainian).
19. Surovtsev I.V. New information technology for axial weighing of cars on platform scales. *Control System and Computers*. 2015, No 3, pp. 77–81 (in Russian).
20. Device for measuring the concentration of toxic elements: patent 107412, Ukraine: IPC (2006) G01N 27/48. а201306295; claimed 21.05.13; published 25.12.14 (in Ukrainian).
21. Device for measuring the concentration of chemical elements by pulsed chronopotentiometry: patent 123459, Ukraine: IPC G01N 27/48 (2006.01). а201902429; claimed 12.03.2019; published 07.04.2021 (in Ukrainian).
22. Surovtsev I.V., Velykyi P.Y., Galimova V.M., Sarkisova M.V. Ionometric method for determination of concentrations of microelements in research of digital medicine. *Cyb. and comp. eng.*, 2020, No 4 (202), pp. 25–43.

23. Wenger E.F, Babak O.V., Surovtsev I.V. et al. Algorithm for operational indirect measurement of the concentration of gaseous impurities in the atmosphere by Fourier spectrometer spectra. *Control System and Computers*, 2007, No 2, pp.33–38, 81 (in Russian).
24. Gritsenko V.I., Surovtsev I.V., Babak O.V. 5G wireless communication system. *Cyb. and comp. eng.* 2019, No 3 (197), pp. 5-19. DOI: 10.15407/kvt197.03.005 (in Ukrainian).
25. Gritsenko V.I., Babak O.V., Surovtsev I.V. Peculiarities of interconnection 5G, 6G networks with big data, internet of things and artificial intelligence. *Cyb. and comp. eng.* 2021, No 2 (204), pp. 5–19 (in Ukrainian).
26. Device for pre-sowing seed treatment: patent 122085, Ukraine: IPC (2020.01) A01C 1/00, A01G 7/04 (2006.01). a201809456; claimed 19.09.2018; published 10.09.2020 (in Ukrainian).

Received 8.02.2022

*Babak O.V.*, PhD (Engineering),  
Senior Researcher of the Ecological Digital Systems Department  
Ecological Digital Systems Department  
ORCID: 0000-0002-7451-3314  
e-mail: dep115@irtc.org.ua

*Surovtsev I.V.*, DSc (Engineering), Senior Researcher,  
Head of the Ecological Digital Systems Department  
ORCID: 0000-0003-1133-6207  
e-mail: igorsur52@gmail.com  
International Research and Training Center for Information Technologies  
and Systems of the National Academy of Sciences of Ukraine  
and Ministry of Education and Science of Ukraine,  
40, Acad. Glushkov av., Kyiv, 03187, Ukraine

#### DATA REDUCTION AS A METHOD OF INTELLECTUALIZATION OF INFORMATION TECHNOLOGIES

**Introduction.** *The principles of reduction theory in the form of methods for detecting hidden patterns in data, approaches to transforming coordinate systems and reducing the dimensionality of input information are widely used in identifying and classifying objects, finding abnormal trends in financial activities, technical systems for measuring physical and chemical parameters.*

**The purpose** of the paper is to review the theoretical and practical results of the application of the theory of reduction in problems and data processing systems.

**Methods.** *Methods for detecting hidden patterns in data, classification of objects using alpha procedure, cognitive modeling of environmental data, frequency-time digital filtering, conversion and modeling of multicomponent signals, measuring concentrations of chemicals in the environment and the content of greenhouse gases in the earth atmosphere, determining the loads on the axis of the car on the signal of dynamic weighing, estimating the consumption of pulverized coal to save coke and natural gas blast furnace.*

**Results.** *Principles of the theory of reduction, highly sensitive analytical system "Analyzer SCP" for measuring concentrations of 20 chemicals in drinking water, food and the environment using three new pulsed chronopotentiometry methods for electrochemical analysis, information technology "Atmospheric gases" for determining concentrations of 38 gases atmosphere, computer systems of dynamic weighing of the car for customs weight control of vehicles.*

**Conclusions.** *For a class of problems with monotonic integrated measurement data, it is necessary to transform the coordinate system, which allows us to consider a multicomponent signal as the sum of individual components. The use of reduction theory methods has made it possible to develop intelligent information technologies for environmental monitoring of biosphere objects, effective technical systems for measuring physical parameters, and detecting fraudulent transactions in the banking system.*

**Keywords:** *reduction theory, hidden patterns, object classifications, coordinate system transformations, frequency-time filtering, concentration determination.*