

УДК 519.688

О. П. Луценко, О. Г. Байбуз, Н. П. Шерстюк

Дніпропетровський національний університет імені Олеся Гончара

Структура інформаційної системи аналізу часових рядів даних гідрохімічного моніторингу

Розроблено структуру інформаційної системи підтримки прийняття рішень у процесі обробки результатів гідрохімічного моніторингу. Система включає засоби ймовірнісно-статистичного аналізу точок розладнання, а також визначення та аналізу рівноважних-нерівноважних станів у хімічному складі води водних об'єктів.

Ключові слова: *ймовірнісно-статистичний аналіз, гідрохімічний моніторинг, розладнання, марківський процес.*

Разработана структура информационной системы поддержки принятия решений в процессе гидрохимического мониторинга. Система включает средства вероятностно-статистического анализа точек разладки, а также определения и анализа равновесных-неравновесных состояний процесса мониторинга.

Ключевые слова: *вероятностно-статистический анализ, гидрохимический мониторинг, разладка, марковский процесс.*

The structure of information support decision support system in the process of hydrochemical monitoring is developed. The system consists of components for probabilistical change-point analysis, as well as detection and analysis of equilibrium and non-equilibrium states of the monitoring process.

Keywords: *probabilistical statistic analysis, hydrochemical monitoring, change-point, Markov process .*

Вступ. Забруднення навколишнього середовища у районах видобутку корисних копалин є загальносвітовою проблемою, оскільки негативний вплив здійснюється на всі природні компоненти: поверхневі та ґрунтові води, атмосферу, ґрунт. У таких районах сформувалися природно-техногенні системи, що охоплюють джерела забруднення, поверхневі води, підземні води, процеси в яких характеризуються багатопараметричними залежностями, що ускладнює їх моніторинг і потребує розробки додаткових принципів аналізу інформації.

Аналіз літературних даних. Процеси формування хімічного

складу поверхневих вод досить детально розглянуто у монографії В. І. Осадчого, Б. Й. Набиванця та ін. [1]. Вивчено особливості зв'язку між водним стоком і хімічним складом води водних об'єктів, гідродинамічні процеси; оцінено вплив гідрофізичних факторів на вміст у воді розчиненого кисню, каламутність і кольоровість води, а також роль донних відкладень у формуванні хімічного складу вод.

Останні роботи з вивчення гідролого-гідрохімічних умов водних об'єктів Криворіжжя пов'язані з іменами В. К. Хільчевського, Р. Л. Кравчинського [2]. У монографіях визначено, оцінено та науково доведено співвідношення антропогенних та природних факторів, під впливом яких формується гідрохімічний режим водних об'єктів.

Проблема виявлення стрибкоподібних змін станів випадкових процесів розглядалася у працях Б. Є. Бродського та Б. С. Дарховського, І. В. Нікіфорова [3].

В роботі R. P. Addams та D. J. C. McKay наведено метод байєсівського прогнозування зміни статистичних характеристик нестационарного процесу [4].

Авторами даної статті в попередніх працях були запропоновані перетворення методів виявлення розладнання, що дозволяють працювати з часовими рядами гідрохімічних спостережень [5] та модель ймовірно-статистичного аналізу точок розладнань на часовому ряді.

Постановка цілей статті. Мета даної статті – опис запропонованої авторами інформаційної системи ймовірно-статистичного аналізу гідрохімічних процесів, які формують хімічний склад води водних об'єктів районів видобутку корисних копалин. Інформаційна система заснована як на розробках, запропонованих у попередніх працях авторів даної статті, так і на нових напрацюваннях, що включають аналіз рівноважних-нерівноважних станів гідрохімічних процесів.

Основний матеріал. Інформаційна система працює за принципом клієнт-сервер, де на сервері зберігається база даних і виконуються хранимі процедури. На стороні клієнта проводиться решта розрахунків.

В якості СУБД використовується MS SQL Server, клієнтська частина проекту написана на C# і потребує бібліотек NET Framework.

До складу системи підтримки прийняття рішень входять такі підсистеми:

База даних гідрохімічного моніторингу, в якій містяться дані про показники стану води, а також хранимі процедури додавання, видалення та вибірки (з перетворенням до різних масштабів часу) інформації;

Підсистема управління даними на рівні програмного додатка, що включає методи отримання даних, розпаковки, читання і внесення даних до бази;

Підсистема графічного виводу включає засоби відрисовки кривих різноманітних типів та забезпечує правильну перерисовку кривих у разі зміни масштабу або часового інтервалу (рис. 1). Бібліотека включає дерево класів, призначених для відрисовки кривих та точок розладнання на кривих.

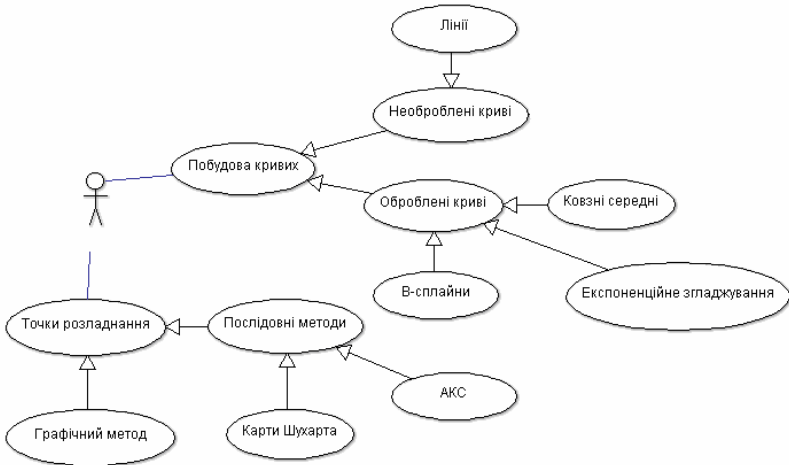


Рисунок 1 – UML діаграма варіантів використання підсистеми графічного виводу

Підсистема первинного статистичного аналізу включає засоби статистичного аналізу часових рядів, а також специфічні для предметної області гідрохімічного моніторингу обчислювальні схеми відповідних термодинамічних розрахунків хімічного поділу, масопереносу у підземних водах, хімічної термодинаміки обчислювальні схеми ймовірнісного прогнозування хімічного складу води водних об'єктів.

Підсистема ймовірнісно-статистичного аналізу точок розладнань включає методи виявлення точок розладнання статистичних характеристик часового ряду, зокрема алгоритм кумулятивних сум [6], тест Бродського – Дарховського і тест Манна – Уїтні [3; 6].

На кожному кроці алгоритму кумулятивних сум накопичується сума:

$$S_t = \max(0, S_{t-1} + g_t) .$$

Сигнал про розладнання подається у момент часу:

$$\tau = \inf(t \geq 1 : S_t > b) ,$$

де b – бар'єр чутливості методу.

Для випадку збільшення математичного очікування m :

$$g_t = z(t_i) - m_1 - k \quad \text{при } m_2 > m_1 ,$$

для випадку зменшення m :

$$g_t = -z(t_i) + m_1 + k \quad \text{при } m_2 < m_1 ,$$

де m_1, m_2 – математичне очікування величини у до і після розладнання відповідно,

$k \geq 0$ – поріг чутливості методу для відхилення від m_1 .

Щоб виключити вплив наявності тренду, у складі процесу використовувалося перетворення вихідного ряду до ряду кутів нахилу апроксимуючих відрізків, знайдених шляхом лінійної апроксимації методом найменших квадратів, до горизонтальної осі, описане авторами даної статті у [6].

В якості методів послідовного виявлення було використано апостеріорний тест Бродського – Дарховського і тест Манна – Уїтні [3; 6].

Обчислювальна схема методу Бродського – Дарховського:

$$Y(\tau) = \frac{\tau}{N} \left(1 - \frac{\tau}{N} \right) \left(\frac{1}{\tau} \sum_1^{\tau} y_i - \frac{1}{N - \tau} \sum_{\tau+1}^N y_i \right) ,$$

$$G = \max_{1 \leq \tau \leq N-1} |Y(\tau)| ,$$

$$G < h \Rightarrow \text{дефекта немає} ,$$

$$G \geq h \Rightarrow \text{дефект} ,$$

$$T(\tau) = Y(\tau + [\varepsilon N]) - Y(\tau) , \quad 0 < \varepsilon < \frac{\delta}{4} .$$

$$\tau_1 = \begin{cases} \min A_1, A_1 \neq \emptyset \\ N, A_1 = \emptyset \end{cases},$$

$$A_1 = \{\tau \geq 1: \text{sign}(T(\tau)) \neq \text{sign}(T(\tau+1))\},$$

$$\tau_i = \begin{cases} \min A_i, A_i \neq \emptyset \\ N, A_i = \emptyset \end{cases},$$

$$A_i = \{\tau \geq \tau_{i+1} + [\delta N / 2]: \text{sign}(T(\tau)) \neq \text{sign}(T(\tau+1))\},$$

$$i = 2.. \tilde{k},$$

$$\tilde{k} = \min.$$

Обчислювальна схема тесту Манна – Уїтні:

$$u_{ik} = \begin{cases} 1, y_i \geq y_k \\ 0, y_i < y_k \end{cases},$$

$$G(\tau) = \frac{\sum_{k=\tau+1}^N \sum_{i=1}^n u_{ik}}{\tau(N-\tau)},$$

$$\tau_0 = \arg \min_{[\alpha N] \leq \tau \leq [N-\alpha N]} G(\tau), \mu_1 < \mu_2,$$

$$\tau_0 = \arg \max_{[\alpha N] \leq \tau \leq [N-\alpha N]} G(\tau), \mu_1 > \mu_2,$$

де τ_0 – момент розладнання.

Для тесту Манна – Уїтні для виявлення повторюваних розладнань був застосований рекурсивний порядок виклику алгоритму на ділянках зліва і справа від розладнання: $(0; \tau_0)$, (τ_0+1, N) . Умови завершення рекурсії: на ділянці відсутній локальний максимум $G(\tau)$ або локальний максимум $G(\tau)$ не перевищує встановлений бар'єр.

Більш детальний опис модифікацій та налаштувань алгоритмів підсистеми аналізу точок розладнань наведено в [5].

Підсистема моделювання рівноважних-нерівноважних станів включає алгоритми марківського моделювання гідрохімічних процесів. Формування хімічного складу води відбувається під впливом значного числа випадкових факторів. У зв'язку із цим багато процесів у поверхневих водах мають імовірнісний характер, що визначає велику роль теорії імовірності в їхньому дослідженні. Зазначена сукупність

факторів утворює багатовимірну систему часових рядів, комбінація значень яких у кожен момент часу утворює певний стан системи, схильний до утримання рівноваги впродовж певного проміжку часу. У випадкові моменти часу стрибкоподібне збурення, обумовлене внутрішніми чинниками взаємодії між компонентами системи або зовнішнім втручанням, спричиняє вихід системи з рівноважного стану з наступним встановленням нової точки рівноваги, що може відрізнятися від попередньої.

Робочою гіпотезою для побудови даної підсистеми стала умова статистичної стійкості хімічного складу води одного об'єкта, що дозволяє використовувати в процесі аналізу встановлення динамічних гідрохімічних рівноваг ефективні математичні методи теорії випадкових процесів та марківських ланцюгів. При виконанні цієї умови за допомогою апарату ймовірно-статистичного аналізу часових рядів можна оцінювати час збереження поточного стану, при цьому шляхом марківського моделювання прогнозувати ймовірності переходу процесу з одного рівноважного стану в інший.

З математичної точки зору, рівноважні гідрохімічні стани утворюють марківський процес з кінцевою кількістю станів $\{S_1, S_2, \dots, S_K\}$, що переходять один в одний у випадкові моменти часу $\{\tau_1, \tau_2, \dots, \tau_n\}$ з інтенсивностями:

$$\lambda = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \dots & \lambda_{1K} \\ \lambda_{21} & \lambda_{22} & \dots & \lambda_{2K} \\ \dots & \dots & \dots & \dots \\ \lambda_{K1} & \lambda_{K2} & \dots & \lambda_{KK} \end{bmatrix}.$$

Для побудови початкової вибірки визначимо моменти часу $\{\tau_1, \tau_2, \dots, \tau_n\}$ як моменти розладнання процесу.

У загальному випадку, належність відрізка до певного стану визначається кластерною приналежністю відрізка у багатовимірному просторі за вектором параметрів моніторингу. З практичної точки зору, так як гідрохімічні показники, що розглядалися, є сильно корельованими між собою, що добре видно як з обчислень, так і з графіків процесу (рис. 2), на попередньому етапі досліджень стани виділялися за напрямом тренду досліджуваних часових рядів (зростаючий, спадний, відсутність тренду).

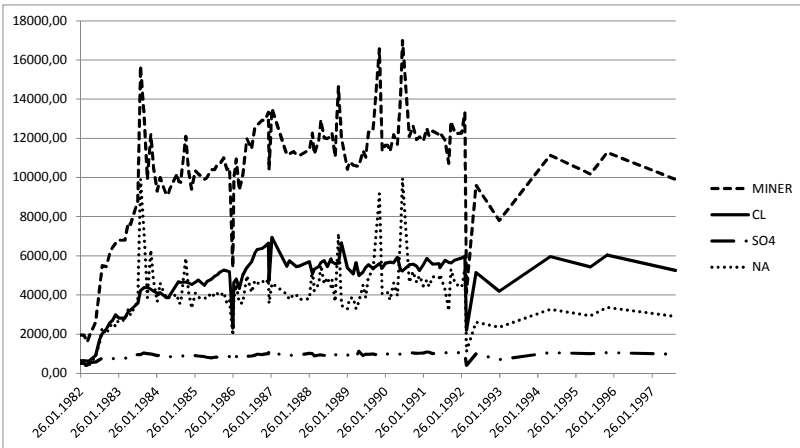


Рисунок 2 – Ілюстрація взаємної корельованості гідрохімічних характеристик у воді хвостосховища Північного ГЗК

Маючи дані про відрізки між розладнаннями, на яких процес перебуває в деякому одному стані, а також дані про приналежність кожного відрізка до одного зі станів, можемо побудувати вибірки відрізків часу τ_{kj} , де k – номер стану, в якому система перебувала перед тим, як перейти до стану j . За вибіркою відрізків чисельно встановлюються інтенсивності переходу з k -го стану в j -й.

Знаючи всі інтенсивності переходу, маємо можливість обчислити ймовірність перебування системи в тому чи іншому стані за рівняннями Колмогорова.

Сфера застосування запропонованої інформаційної системи – мінімізація фінансових втрат, пов'язаних зі справдженням гіпотези про розладнання, шляхом здійснення керування, виходячи з мінімізації величини математичного очікування таких втрат.

Висновки. Запропонована інформаційна система дозволить значно підвищити екологічну безпеку у регіонах з підвищеним технологічним навантаженням. Додатковий соціальний ефект полягає у зниженні ризиків загрози здоров'ю та життю населення, попередженні захворювань людей в зонах поблизу водних об'єктів з підвищеним технологічним навантаженням.

Математична модель ймовірнісної оцінки стану нестационарних процесів, заснована на відтворенні щільності розподілу точок розладнань, а також обчислювальні схеми функцій ризику

розладнання, що впливають з моделі, розроблені і отримали практичну реалізацію в галузі гідрохімічного моніторингу вперше. Математичний апарат обчислення функцій ризику виникнення розладнання для процесу моніторингу з множинними розладнаннями відсутній в жодному з програмних продуктів аналізу гідрохімічних даних, представлених на ринку програмного забезпечення, і таким чином, інформаційна система характеризується новизною.

Напрямок подальших досліджень: узагальнення алгоритму виділення рівноважних-нерівноважних станів на випадок з довільною кількістю станів.

Бібліографічні посилання

1. Осадчий В. І., Набиванець Б. Й., Линник П. М. та ін. Процеси формування хімічного складу поверхневих вод. Київ: Ніка-Центр, 2013. 240 с.

2. Хільчевський В. К., Кравчинський Р. Л., Чунарьов О. В. Гідрохімічний режим та якість води Інгульця в умовах техногенезу. Київ: Ніка-Центр, 2012. 180 с.

3. Brodsky B. E., Darkhovsky B. S. Nonparametric Methods in Change-Point Problems. Dordrecht: Kluwer Academic Publishings, 1993. 210 p.

4. Adams R. P., McKay D. J. C. Bayesian Online Changepoint Detection. URL: <http://arxiv.org/pdf/0710.3742v1.pdf>

5. Луценко О. П., Байбуз О. Г., Чорна А. О. Особливості застосування методів виявлення розладнань у процесі дослідження даних екологічного моніторингу в регіонах з техногенним навантаженням // Актуальні проблеми автоматизації та інформаційних технологій. Дніпро. 2016. Т. 20. С. 31–41.

6. Бродский Б. Е., Дарховский Б. С. Асимптотический анализ некоторых оценок в апостериорной задаче о разладке // Теория вероятностей и ее применения. 1990. Т. 35. № 3. С. 551–557.

Надійшла до редколегії 01.08.17