

НЕПАРАМЕТРИЧНИЙ ПОКАЗНИК ВАРІАБЕЛЬНОСТІ ГІДРОЕКОЛОГІЧНИХ ЧАСОВИХ РЯДІВ

NONPARAMETRIC FACTOR OF VARIABILITY OF HYDROECOLOGICAL TIME SERIES



Артеменко Владислав Андрійович, магістр екології, Український гідрометеорологічний інститут Державної служби України з надзвичайних ситуацій та Національної академії наук України, науковий співробітник відділу гідрохімії, e-mail: artemenko@uhmi.org.ua, тел. 380936011250, Україна, 03028, м. Київ, просп. Науки 37, к.34.

<https://orcid.org/0000-0003-0536-5415>



Петрови́ч Володи́мир Васи́льович, кандидат технічних наук, професор, старший науковий співробітник, професор кафедри транспортного будівництва та управління майном Національного транспортного університету. e-mail: petrovichvv60@ukr.net, тел. +380442807338, Україна, 01010, м. Київ, вул. М. Омеляновича-Павленка, 1, к. 138.,

<https://orcid.org/0000-0003-0422-2535>

Анотація. З метою вдосконалення оцінки гідроекологічних часових рядів запропоновано визначати непараметричний показник (коефіцієнт) варіабельності.

Визначені зміни непараметричного показника варіабельності гідроекологічного ряду за сезонами календарного року.

Метод дозволяє отримувати значення непараметричного коефіцієнта варіабельності не тільки для гідроекологічних, але також для всіх інших природних часових рядів.

Ключові слова: гідроекологічні часові ряди, обробка вихідних даних, аномальні значення ряду, непараметричний коефіцієнт варіабельності ряду, розчинений у воді кисень.

Вступ

Однією з важливих характеристик екологічного часового ряду є показник його варіабельності [1].

Як відомо, параметричний показник (коефіцієнт) варіабельності ряду припускає, що вихідні дані досить добре описуються нормальним законом розподілу, та виражається формулою

$$PCV = \frac{Std(V)}{Mean(V)},$$

де V - вектор даних, які аналізуються,

$Std(V)$ - стандартне відхилення (скаляр),

$Mean(V)$ - середнє значення (скаляр).

При цьому також вважають, що всі значення (складові) вектору даних невід'ємні, $Mean(V) \neq 0$.

Однак більшість гідроекологічних часових рядів мають статистичні розподілення, які не є нормальними [2].

Тому, строго кажучи, параметричний коефіцієнт варіабельності, що часто використовується для аналізу ряду в даному випадку не підходить.

У цьому зв'язку пропонується використовувати непараметричний показник (коефіцієнт) варіабельності, придатний для досліджень різних гідроекологічних часових рядів.

Непараметричний показник варіабельності

У зв'язку з тим, що застосовувати для аналізу параметричний показник варіабельності в даному випадку некоректно, пропонується використовувати непараметричний показник варіабельності, придатний для дослідження будь-яких природних часових рядів.

Непараметричний показник варіабельності обчислюється за формулою

$$NPCV = \frac{(Quantile(V, 0.95) - Quantile(V, 0.05))}{Median(V)},$$

де *Median (V)* - значення медіани даних.

При цьому, як і для випадку використання параметричного показника, повинна виконуватися умова $Min(V) \geq 0$, а також умова $Median(V) \neq 0$.

Для реалізації непараметричного показника був написаний код на мові програмування *MATLAB*. Код основних функцій наводиться нижче.

```
function [ ANS ] = IS_VECTOR( V );ANS = false;
if isempty( V ) == true;return
end

if isnumeric( V ) == false;return
end

if isreal( V ) == false;return
end

if isvector( V ) == false;return
end

if any( isnan( V ) ) == true;return
end

if any( isinf( V ) ) == true;return
end

ANS = true;end
function [ ANS ] = IS_MATRIX( M );ANS = false;
if isempty( M ) == true;return
end

if isnumeric( M ) == false;return
end

if ndims( M ) ~= 2;return
end

if any( any( isnan( M ) ) ) == true;return
end

if any( any( isinf( M ) ) ) == true;return
end
```

```

if any( any ( isreal( M ) ) ) == false return
end

ANS = true; end
function [ OUT ] = PCV( IN ) OUT = [ ];
if IS_MATRIX( IN ) == false return
end

A = std( IN ); B = mean( IN );

OUT = A ./ B;

if IS_MATRIX( OUT ) == false OUT = [ ];
end end
function [ OUT ] = NPCV( IN ) OUT = [ ];
L = 0.05; H = 1 - L;

if IS_MATRIX( IN ) == false return
end

A = quantile( IN , H ); B = quantile( IN , L ); C = median( IN );

OUT = ( A - B ) ./ C;

if IS_MATRIX( OUT ) == false OUT = [ ];
end end

```

При цьому функції *IS_VECTOR* () та *IS_MATRIX* () є допоміжними (службовими). Вони необхідні для перевірки правильності вхідних аргументів основних (обчислюваних) функцій. Функція *IS_VECTOR* () перевіряє - чи є вхідний аргумент вектором, тобто матрицею розміру $1 \times SZ2$ або $SZ1 \times 1$. Крім того, функція *IS_VECTOR* () перевіряє, щоб всі елементи вектору були дійсними числами і не містили *NAN*, - *Inf*, + *Inf* значень і комплексних чисел. Відповідно, *IS_MATRIX* () перевіряє, чи є вхідний аргумент матрицею, як це прийнято на мові програмування *MATLAB* (тобто матрицею розміру $SZ1 \times SZ2$, де *SZ1* і / або *SZ2* можуть бути одиницею).

Функція *IS_MATRIX* () перевіряє, щоб всі елементи також були дійсними числами (діє за аналогією з *IS_VECTOR* ()). Відповідно обчислювальні функції - це *PCV* () і *NPCV* ().

В основу даних обчислювальних функцій покладено формули, що наведені вив тексті статті. Вхідні аргументи (*IN*) цих функцій - матриці, вихідні аргументи (*OUT*) цих функцій - "горизонтальні вектори", тобто матриці розміру $1 \times SZ2$, де *SZ2* - число стовпців вхідної матриці *IN*.

Якщо в деяких випадках неможливо провести відповідні розрахунки (нульові середні значення, нульові медіанні значення, неадекватні вхідні аргументи і ін.), тоді наведені функції повертають порожню матрицю.

Це легко проконтролювати, використовуючи службову функцію *isempty* ().

Непараметричний коефіцієнт варіабельності розчиненого у воді кисню

Розглянемо як приклад два водних об'єкта з істотно різним режимом функціонування.

У роботі були використані середньодобові дані по концентраціям розчиненого у воді кисню за 16-річний період (з 1995 по 2010 роки) для крупної рівнинної річки (*RIVER I*) і крупного водосховища (*RIVER II*) України.

Довжина ряду даних склала $16 \times 365 = 5840$ (точок).

Раніше ми вже розглядали аномалії часового ряду концентрації розчиненого кисню, що мають місце при температурі річкової води, близької до температури замерзання [3].

У разі наявності льоду на поверхні водного об'єкта і завдяки процесам окислення ряду речовин у водному середовищі, створюється суттєвий недолік розчиненого кисню.

Таким чином, при температурі води, близької до температури замерзання, принципово можлива гранично низька концентрація кисню.

У зв'язку з цим при температурі, близькій до 0°C , слід очікувати також найбільший ступінь варіабельності розчиненого кисню, оскільки при цьому лід на поверхні водойми може як бути (дуже малий вміст розчиненого кисню), так льоду може і не бути (вміст розчиненого кисню досить великий).

Оскільки вихідні дані не підкоряються нормальному розподілу, в якості "середнього" була використана "медіана" [2].

На рис.1 представлений хід медіани для значень (*RIVER I*) і (*RIVER II*) по днях календарного року.

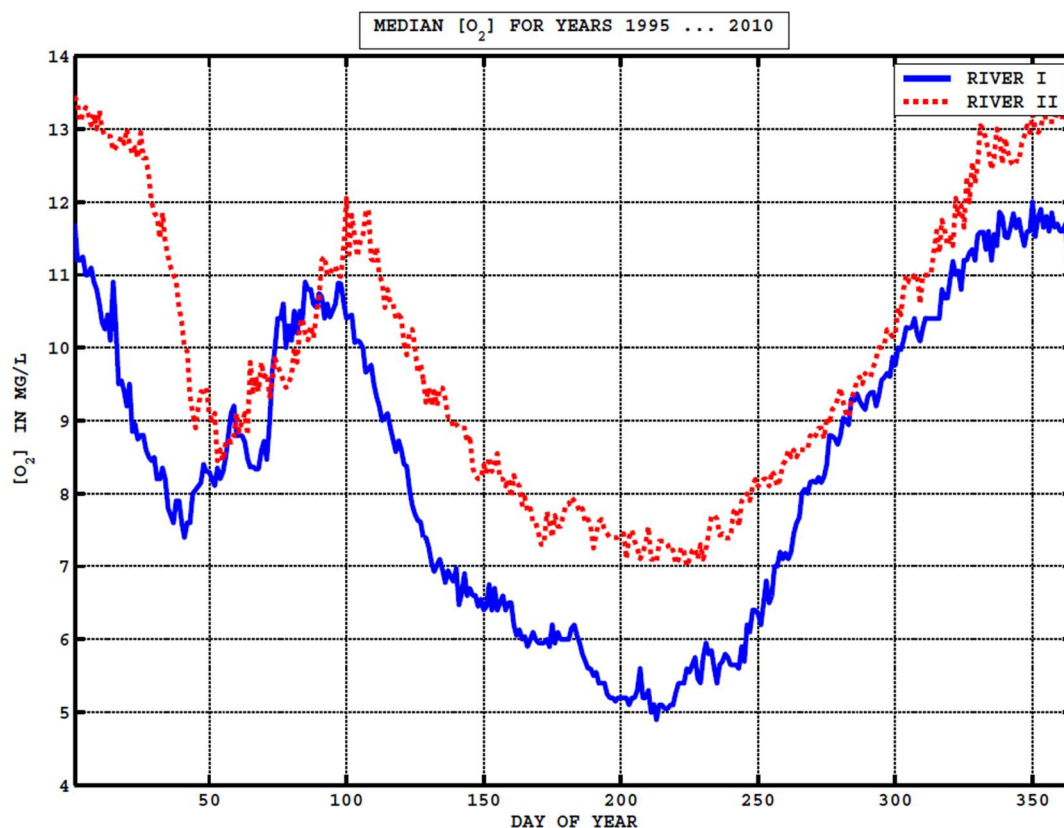


Рисунок 1 – Хід медіани концентрації розчиненого кисню для крупної річкової річки (*RIVER I*) і крупного водосховища (*RIVER II*).

Figure 1 – Move median of concentration of the dissolved oxygen for large flat river (*RIVER I*) and for large water reservoir (*RIVER II*).

При побудові графіка був використаний той же підхід, як це прийнято в метеорології при оцінці "метеорологічної норми".

Як видно, графіки для (*RIVER I*) і (*RIVER II*) йдуть досить злагоджено, проте концентрація розчиненого кисню у водосховищі помітно більше, ніж у річці.

Методика оцінки відмінностей базувалась на методі непараметричного бутстрепа [4]. Факт наявності відмінностей був встановлений статистично. Використовували рівень значущості 0.05 (5%), зазвичай прийнятий для задач гідроекології.

На рис. 2 наведено графік ходу непараметричного коефіцієнта варіабельності. При цьому вихідні дані оброблялися відповідно до методики, аналогічної побудови медіан. Зрозуміло, що в цьому випадку замість значень медіан знаходили значення *NPCV*.

Як видно з рисунка, мінімальні значення варіабельності розчиненого кисню спостерігаються для 25 ... 75-х чисел календарного року, тобто припадають на кінець січня - початок (або навіть середину) березня.

Крім того, пік варіабельності розчиненого у воді кисню для водосховища спостерігається помітно пізніше, ніж для річки. При цьому слід зазначити, що відстань між пунктами спостережень для річки і водосховища було невеликим.

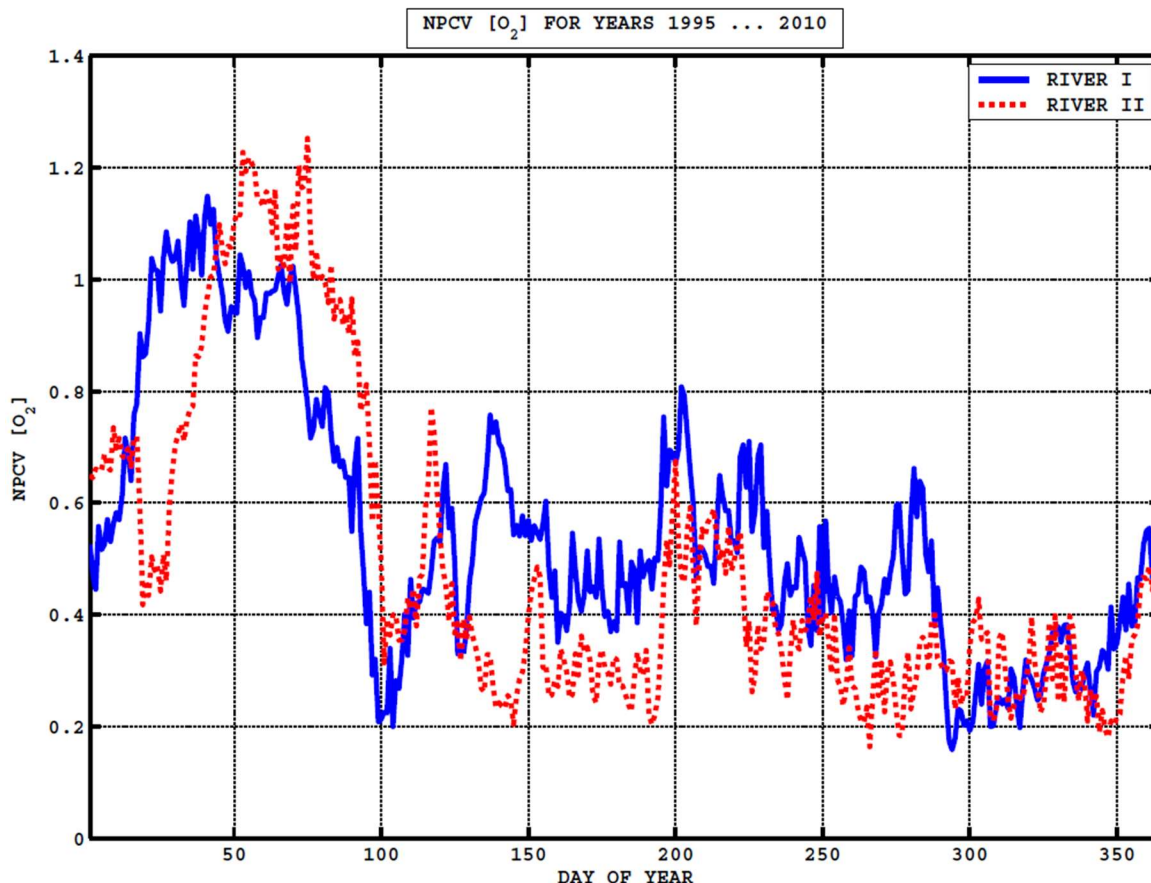


Рисунок 2 – Хід величини непараметричного коефіцієнта варіабельності концентрації розчиненого кисню для крупної рівнинної річки (*RIVER I*) і крупного водосховища (*RIVER II*).

Figure 2 – Move of the value of nonparametric factor of variability concentration of the dissolved oxygen for large flat river (*RIVER I*) and for large water reservoir (*RIVER II*).

Як відомо, для екологічних процесів абсолютно звичайним явищем є наявність досить великого числа аномальних значень в рядах даних.

У зв'язку з цим ми спеціально врахували цей факт, зберігши більшу частину аномалій. З цієї метою використовували квантілі рівнів 0.95 і 0.05, а не квантілі, наприклад, рівнів 0.75 і 0.25 відповідно.

Оскільки більшість природних явищ і процесів не описуються нормальним законом розподілу, має сенс у всіх випадках використовувати метод оцінки варіабельності, запропонований в даній роботі.

Висновки

1. З урахуванням особливостей природних часових рядів (не підкоряються закону нормального розподілу, містять значну кількість аномальних значень та ін.), розроблений метод оцінки варіабельності таких рядів - метод непараметричної варіабельності.

2. Практичне використання методу оцінки непараметричної варіабельності показаний на прикладі оцінки часових рядів середньодобових концентрацій розчиненого у воді кисню (як для крупної рівнинної річки, так і крупного водосховища України)

3. В результаті дослідження встановлено, що значно більшою інформативністю володіє не варіабельність часового ряду в цілому, а варіабельність з урахуванням сезонності (по днях, місяцях, кварталах і ін.).

4. Максимальні значення варіабельності розчиненого кисню спостерігаються переважно для 25 ... 75-х чисел календарного року, тобто на кінець січня - початок (або навіть середину) березня. При цьому пік варіабельності розчиненого у воді кисню для водосховища спостерігається значно пізніше, ніж для річки.

Перелік посилань

1. Шитиков В.К. Количественная гидроэкология: методы, критерии, решения/ В.К. Шитиков. Г.С. Розенберг, Т.Д. Зинченко. –М.: Наука, 2005. –Кн. 1, 281с.; Кн. 2, 337 с.

2. Артеменко В.А. Сучасні статистичні методи обробки екологічних даних/В.А. Артеменко, В.В. Петрович// Автомобільні дороги і дорожнє будівництво, вип. 105. –К.: Вид-во Націон. трансп. ун-ту. – 2019. – С. 35-43.

3. Артеменко В.А. Сезонна динаміка біогенних речовин крупних водних об'єктів/В.А. Артеменко, В.В. Петрович// Автомобільні дороги і дорожнє будівництво, вип. 104. –К.: Вид-во Націон. трансп. ун-ту. – 2018. – С. 31-43.

4. Шитиков В.К. Рандомизация и бутстреп: статистический анализ в биологии и экологии с использованием R / В.К. Шитиков, Г.С. Розенберг. - Тольятти: Кассандра, 2013. - 314 с.

NONPARAMETRIC FACTOR OF VARIABILITY OF HYDROECOLOGICAL TIME SERIES

Artemenko Vladislav A., Master of Ecology, Ukrainian Hydrometeorological Institute, State Service on Emergencies of Ukraine and National Academy of Science of Ukraine, Hydrochemical Research, Scientific Employee, e-mail: artemenko@uhmi.org.ua, tel. 380936011250, Nauki avenue, 37, Kyiv, Ukraine, 03028, room 34, <https://orcid.org/0000-0003-0536-5415>

Petrovych Volodymyr V., Candidate of Technical Sciences, Professor, Senior Researcher, Professor of the Transportation Construction and Property Management Department, National Transport University. e-mail: petrovichvv60@ukr.net, tel. +380442807338, Ukraine, 01010, Kyiv, street M. Omelyanovicha-Pavlenka, 1, room 138, <https://orcid.org/0000-0003-0422-2535>

Abstract. Nonparametric factor of variability is offered for improvement of the estimation of variability of time series. By means of this nonparametric factor in this article is considered change of variability of hydroecological time series depending on season of the year. Considered nonparametric method is allows to find factor of variability practically for any natural time series.

Keywords: hydroecological time series, processing the original data, anomal values, nonparametric factor of variability of time series, dissolved oxygen in water.

References

1. Shitikov V.K. Kolichestvennaya gidroekologiya: metody, kriterii, resheniya/ V.K. Shitikov. G.S. Rozenberg, T.D. Zinchenko. –М.: Nauka, 2005. –Кн. 1, 281s.; Кн. 2, 337 s.

2. 2.Artemenko V.A. Suchasni statystychni metody obrobky ekolohichnykh danykh/V.A. Artemenko, V.V. Petrovych// Avtomobil'ni dorohy i dorozhnye budivnytstvo, vyp. 105. –К.: Vyd-vo Natsion. transp. un-tu. – 2019. – S. 35-43.

3. Artemenko V.A. Sezonna dynamika biohennykh rehovyn krupnykh vodnykh ob'yektiv/V.A. Artemenko, V.V. Petrovych// Avtomobil'ni dorohy i dorozhnye budivnytstvo, vyp. 104. –К.: Vyd-vo Natsion. transp. un-tu. – 2018. – С. 31-43.

4. Shytykov V.K. Randomyzatsyya y but-strep: statystycheskyy analiz v byolohyy y ékolohyy s yspol'zovanyem R / V.K. Shytykov, H.S. Rozenberh. - Tol'yatty: Kassandra, 2013. - 314 s.