

---

---

# ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В СФЕРЕ ИССЛЕДОВАНИЙ ОКРУЖАЮЩЕЙ СРЕДЫ

---

---

УДК 556.01  
DOI: 10.17586/0021-3454- 2021-64-5-392-397

## МЕТОДИКА НЕПАРАМЕТРИЧЕСКОГО СТАТИСТИЧЕСКОГО АНАЛИЗА ОДНОРОДНОСТИ ЭКОЛОГИЧЕСКИХ ПОКАЗАТЕЛЕЙ ВОДНЫХ ОБЪЕКТОВ

В. Н. ЗАВГОРОДНИЙ

*Российский государственный гидрометеорологический университет,  
192007, Санкт-Петербург, Россия  
E-mail: zavgor@list.ru*

Рассматривается возможность применения методов непараметрического анализа к результатам измерений экологических показателей водных объектов. Показана целесообразность использования непараметрических критериев при малых объемах выборок при анализе данных в гидрологии и экологии. Предложена методика расчета непараметрических критериев для оценки однородности временных рядов экологических показателей водных ресурсов.

*Ключевые слова:* статистические непараметрические критерии, оценка статистической однородности, экологические показатели воды

**Введение.** Появление первых систем мониторинга водных ресурсов относится ко второй половине XX века [1]. В настоящее время статистический анализ данных, получаемых системами мониторинга, позволяет обнаружить причины загрязнения водных объектов и спрогнозировать дальнейшие изменения их состояния. Статистические методы являются ключевыми при анализе данных измерений, поскольку эти данные должны соответствовать определенным статистическим критериям случайности, независимости, однородности и стационарности [2—5].

**Проверка однородности.** Сходство статистических параметров выборок оценивается однородностью. В настоящей статье однородность оценивается по различию математических ожиданий выборок и проверяется как равенство или неравенство медиан с помощью непараметрических критериев Вилкоксона и Вилкоксона — Манна — Уитни, а также критерия Колмогорова — Смирнова для двух выборок [6—8]. Тест Вилкоксона для связанных выборок и тест Вилкоксона — Манна — Уитни являются непараметрическими аналогами теста Стьюдента.

В системе мониторинга водных объектов, в частности р. Охта (Санкт-Петербург), используются в основном результаты измерений растворенных веществ вблизи водной поверхности и дна [9, 10]. Частота измерений — один раз в год, в первой половине июля. Данные измерений экологических показателей р. Охта, произведенных студентами Российского государственного гидрометеорологического университета, являются репрезентативными. Для представленных данных непараметрические методы анализа однородностей и связей имеют преимущество по сравнению с параметрическими, поскольку размеры выборок в наборе измерений относительно небольшие, а функции распределения физических данных не обязательно нормально распределены.

Итак, имеются две выборки результатов измерений жесткости воды на уровне поверхности и дна (табл. 1), полученных на двух станциях в 2006—2015 гг.

Таблица 1

Номер станции	Данные измерений жесткости воды, о.е.									
	1	2,9	1,5	4,2	1,7	2,7	1,4	2,0	1,79	1,82
2	3,0	1,8	2,2	1,74	2,7	1,4	1,6	1,76	1,15	2,48

За нулевую гипотезу принимается утверждение об отсутствии смещения проверяемого параметра при альтернативной гипотезе о его наличии (двусторонний тест).

*Критерий Вилкоксона — Манна — Уитни.* Критерий Вилкоксона — Манна — Уитни используется для проверки однородности двух независимых выборок непрерывных величин  $x_1, x_2, x_3, \dots, x_n$  и  $y_1, y_2, y_3, \dots, y_n$  [11]. Нулевая гипотеза заключается в том, что разность медиан двух функций распределения равна нулю:  $\text{Med } X - \text{Med } Y = 0$ . Конкурирующие гипотезы:  $\text{Med } X - \text{Med } Y \neq 0$ ,  $\text{Med } X - \text{Med } Y < 0$ ,  $\text{Med } X - \text{Med } Y > 0$ .

Формула для расчета критерия:

$$U = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - T,$$

где  $T$  — ранговая сумма ( $T$ -статистика Вилкоксона),  $n_1, n_2$  — объемы соответствующих выборок.

Для проверки нулевой гипотезы используется меньшее из двух значений критерия:  $U = \min(U_1, U_2)$ . Если полученное значение меньше табличного или равно ему, то признается наличие существенного различия между уровнем признака в рассматриваемых выборках. Нулевая гипотеза может приниматься при равенстве медиан.

Критерий имеет следующие ограничения [12]:

— при отсутствии или небольшом сдвиге медиан наличие положительной последовательной корреляции увеличивает вероятность ошибки первого рода; наличие отрицательной последовательной корреляции, в свою очередь, увеличивает вероятность ошибки второго рода;

— положительная последовательная корреляция в больших выборках также снижает вероятность обнаружения сдвига медиан, при этом отрицательная корреляция вероятность обнаружения сдвига незначительно увеличивает.

Проверим нулевую гипотезу. Представим обе выборки в виде одного пронумерованного вариационного ряда (табл. 2), присваивая одинаковым значениям признака среднее арифметическое значений их рангов.

Таблица 2

Ранг	Значение	Ранг	Значение
1	1,15	11	1,82
2,5	1,4	12	1,92
2,5	1,4	13	2
4	1,5	14	2,2
5	1,6	15	2,48
6	1,7	16,5	2,7
7	1,74	16,5	2,7
8	1,76	18	2,9
9	1,79	19	3
10	1,8	20	3,2

Найдем сумму порядковых номеров вариантов выборок:

$$T_1 = 2,5 + 4 + 6 + 9 + 11 + 12 + 13 + 16,5 + 18 + 20 = 112;$$

$$T_2 = 1 + 2,5 + 5 + 7 + 8 + 10 + 14 + 15 + 16,5 + 19 = 98.$$

Затем найдем значение статистики  $U$ :

$$U_1 = 10 \cdot 10 + \frac{10 \cdot (10 + 1)}{2} - 112 = 43;$$

$$U_2 = 10 \cdot 10 + \frac{10 \cdot (10 + 1)}{2} - 98 = 57,$$

$$U = \min(U_1, U_2) = 43.$$

При уровне значимости  $\alpha = 0,05$  значение  $U_{кр} = 17$ , т.е.  $U > U_{кр}$ . Следовательно, нет оснований отклонить нулевую гипотезу  $\text{Med } X = \text{Med } Y = 0$ .

*Критерий Вилкоксона для связанных выборок.* Критерий Вилкоксона служит для проверки однородности двух связанных выборок непрерывных величин  $x_1, x_2, x_3, \dots, x_n$  и  $y_1, y_2, y_3, \dots, y_n$ . Нулевая гипотеза заключается в том, что функции распределения равны при любом значении аргумента  $x$ :  $F(x) = G(x)$ . Конкурирующие гипотезы:  $F(x) \neq G(x)$ ,  $F(x) < G(x)$ ,  $F(x) > G(x)$ .

Критерий Вилкоксона — один самых часто применяемых непараметрических тестов для связанных выборок [12, 13]. Данный критерий используется при выполнении следующих условий:

- выборки связаны;
- парные данные выбраны случайным образом независимо друг от друга;
- данные измерены в порядковой или числовой шкале.

Порядок применения критерия [14].

1. Рассчитываются разности значений с одинаковым порядковым номером в выборках и их модули (табл. 3).

Таблица 3

Порядковый номер	Станция		Разность	Ранг разности
	1	2		
1	2,9	3	-0,1	5
2	1,5	1,8	-0,3	6
3	4,2	2,2	2	10
4	1,7	1,74	-0,04	4
5	2,7	2,7	0	1,5
6	1,4	1,4	0	1,5
7	2	1,6	0,4	7
8	1,79	1,76	0,03	3
9	1,82	1,15	0,67	9
10	1,92	2,48	-0,56	8

2. Ранжируются разности согласно полученным в п. 1 значениям.

3. Вычисляются отдельно суммы рангов положительных  $T^+$  и отрицательных  $T^-$  значений.

4. По таблице вычисляется критическое значение  $U_{кр}$  при заданном уровне значимости  $\alpha$  и размере выборок  $n$ :  $T_\alpha, n$ .

5. Нулевая гипотеза отклоняется, если хотя бы одна из полученных сумм меньше критического значения  $T$ :

$$T^+ = 10 + 7 + 3 + 9 = 29;$$

$$T^- = 5 + 6 + 4 + 8 = 23.$$

$T_{0,05;10} = 8 < 23$ , следовательно, при уровне значимости  $\alpha = 0,05$  оснований отклонить нулевую гипотезу нет.

*Критерий Колмогорова — Смирнова для двух выборок.* Главное назначение теста Колмогорова — Смирнова — определение равенства или неравенства распределений двух выборок. Основная его особенность заключается в том, что тест чувствителен не только к смещению медиан, но и к смещениям других параметров, например дисперсий. В отличие от представленных выше критериев, критерий Колмогорова — Смирнова позволяет проверить однородность функций распределения в более строгом смысле [15—17]. Также этот критерий можно использовать для проверки одной выборки, если заранее задано множество распределений, к которому, как предполагается, принадлежит выборка.

При проверке сложных гипотез критерий перестает быть свободным от распределения, поэтому в данной работе тест рассматривается только для простых гипотез.

Нулевая гипотеза —  $H_0: F(t) = G(t)$  для любого  $t$ ; альтернативная гипотеза —  $H_1: F(t) \neq G(t)$  для некоторого  $t$ .

Для проверки гипотезы используется статистика  $J$  Колмогорова — Смирнова:

$$J = \sqrt{\frac{n_1 \cdot n_2}{n_1 + n_2}} \cdot \max |F_m(t) - G_n(t)|,$$

где  $F_m(t)$  и  $G_n(t)$  — эмпирические функции распределения для проверяемых выборок.

Для расчета критерия вычислим накопленные частоты выборок, рассчитаем их долю от полной суммы для каждой выборки и найдем наибольшее значение модуля соответствующих разностей (табл. 4).

Таблица 4

Год	Станция		Сумма		$F_1^*(x)$	$F_2^*(x)$	$ F_1^*(x) - F_2^*(x) $
	1	2	1	2			
2006	2,9	3,0	2,9	3	0,13	0,15	0,019
2007	1,5	1,8	4,4	4,8	0,2	0,24	0,041
2008	4,2	2,2	8,6	7	0,39	0,35	0,039
2009	1,7	1,74	10,3	8,74	0,47	0,44	0,029
2010	2,7	2,7	13	11,44	0,59	0,58	0,016
2011	1,4	1,4	14,4	12,84	0,66	0,65	0,009
2012	2,0	1,6	16,4	14,44	0,75	0,73	0,02
2013	1,79	1,76	18,19	16,2	0,83	0,82	0,013
2014	1,82	1,15	20,01	17,35	0,91	0,87	0,038
2015	1,92	2,48	21,93	19,83	1	1	0

Значение статистики:

$$\lambda = \max |F_1^*(x) - F_2^*(x)| \sqrt{\frac{n_1 n_2}{n_1 + n_2}} = 0,041 \sqrt{\frac{100}{20}} = 0,092.$$

При уровне значимости  $\alpha = 0,05$   $\lambda_{0,05} = 1,36$ . Следовательно, можно считать, что измерения, произведенные на первой и второй станциях, описываются одной функцией распределения.

**Заключение.** Главное отличие непараметрических критериев заключается в том, что для их применения не требуется знать функцию распределения выборки. Поэтому непараметрические критерии в основном используются для малых выборок. Поскольку для выборки достаточно большого размера, состоящей из случайных слабозависимых величин, функция распределения согласно центральной предельной теореме близка к нормальной, то в качестве тестов чаще используются параметрические. Большинство непараметрических критериев являются ранговыми, т.е. для проверки гипотезы используются порядковые номера величин, а не сами величины, поэтому можно использовать данные, представленные в любых шкалах измерения. Непараметрические критерии имеют меньшую мощность по сравнению с параметрическими.

Учитывая малые объемы выборок и то, что в гидрологии и экологии функции распределения переменных зачастую отличаются от нормального распределения, для анализа данных в районах морской деятельности целесообразно использовать непараметрические критерии.

#### СПИСОК ЛИТЕРАТУРЫ

1. Герасимов И. П. Научные основы современного мониторинга окружающей среды // Изв. АН СССР. Сер. География. 1975. № 3. С. 13—25.
2. Wilk D. S. Statistical Methods in the Atmospheric Sciences. Elsevier, 2019.

3. Кутузов О. И., Татарникова Т. М. Экологический мониторинг как объект автоматизации управления // Ученые записки РГГМУ. 2011. № 21. С. 165—171.
4. Завгородний В. Н., Зарукина М. А. Статистические методы анализа экологических явлений // Метеорологический вестник. 2018. Т. 10, № 4. С. 1—6.
5. Завгородний В. Н., Зарукина М. А. Роль и место статистических методов в исследовании экологических явлений // Информационные технологии и системы: управление, экономика, транспорт, право. 2017. № 3 (21). С. 16—19.
6. Кобзарь А. И. Прикладная математическая статистика. Для инженеров и научных работников. М.: ФИЗМАТЛИТ, 2006.
7. Критерий Вилкоксона для проверки однородности выборок [Электронный ресурс]: <<https://docplayer.ru/35735873-Kriteriy-vilkoksona-w-dlya-proverki-odnorodnosti-vyborok-v-1-2.html>>, 15.04.2020.
8. Кузнецов Л. А., Журавлева М. Г. Построение карт контроля качества с помощью непараметрического критерия Вилкоксона — Манна — Уитни // Заводская лаборатория. Диагностика материалов. 2009. Т. 75, № 1. С. 70—75.
9. Урсова Е. С. Оценка загрязненности реки Охта в пределах Санкт-Петербурга на основе применения интегральных кривых // Общество. Среда. Развитие. 2015. № 4 (37). С. 171—175.
10. Станскова А. Г. Оценка экологического состояния р. Охта в Санкт-Петербурге и Ленинградской области. СПб: РГГМУ, 2016.
11. Гмурман В. Е. Теория вероятностей и математическая статистика. М.: Высш. школа, 2003.
12. Hollander M., Wolfe D. A., Chicken E. Nonparametric Statistical Methods. Wiley, 2014.
13. Davis C. S. Statistical Methods of the Analysis of Repeated Measurements. N. Y.: Springer-Verlag, 2002.
14. Орлов А. И. Нечисловая статистика. М.: МЗ-Пресс, 2004.
15. Кутузов О. И., Татарникова Т. М. Из практики применения метода Монте-Карло // Заводская лаборатория. Диагностика материалов. 2017. Т. 83, № 3. С. 65—70.
16. Орлов А. И. Непараметрическое точечное и интервальное оценивание характеристик распределения // Заводская лаборатория. Диагностика материалов. 2004. Т. 70, № 5. С. 65—70.
17. Орлов А. И. Состоятельные критерии проверки абсолютной однородности независимых выборок // Заводская лаборатория. Диагностика материалов. 2012. Т. 78.

**Сведения об авторе**

**Владимир Николаевич Завгородний** — д-р техн. наук, профессор; РГГМУ, кафедра морских информационных систем; E-mail: zavgor@list.ru

Поступила в редакцию  
23.01.2021 г.

**Ссылка для цитирования:** Завгородний В. Н. Методика непараметрического статистического анализа однородности экологических показателей водных объектов // Изв. вузов. Приборостроение. 2021. Т. 64, № 5. С. 392—397.

**METHOD OF NONPARAMETRIC STATISTICAL ANALYSIS  
OF HOMOGENEITY OF ECOLOGICAL INDICATORS OF WATER BODIES**

**V. N. Zavgorodniy**

*Russian State Hydrometeorological University,  
192007, St. Petersburg, Russia  
E-mail: zavgor@list.ru*

The possibility of applying the methods of nonparametric analysis to results of measuring the ecological indicators of water bodies is considered. The expediency of using nonparametric criteria for samples of small size in analysis of hydrological and ecological data is shown. A method is proposed for

calculating nonparametric criteria for assessing homogeneity of time series of the environmental indicators of water resources.

**Keywords:** statistical nonparametric criteria, assessment of statistical homogeneity, environmental indicators of water

#### REFERENCES

1. Gerasimov I.P. *Izvestiya Akademii Nauk SSSR. Ser. Geograficheskaya*, 1975, no. 3, pp. 13–25. (in Russ.)
2. Wilk D.S. *Statistical methods in the atmospheric sciences*, 4th Edition, 2019.
3. Kutuzov O.I., Tatarnikova T.M. *Uchenyye zapiski Rossiyskogo gosudarstvennogo gidrometeorologicheskogo universiteta*, 2011, no. 21, pp. 165–171. (in Russ.)
4. Zavgorodniy V.N., Zarukina M.A. *Meteorologicheskij vestnik*, 2018, no. 4(10), pp. 1–6. (in Russ.)
5. Zavgorodniy V.N., Zarukina M.A. *Informatsionnyye tekhnologii i sistemy: upravleniye, ekonomika, transport, pravo*, 2017, no. 3(21), pp. 16–19. (in Russ.)
6. Kobzar' A.I. *Prikladnaya matematicheskaya statistika. Dlya inzhenerov i nauchnykh rabotnikov* (Applied Mathematical Statistics. For engineers and scientists), Moscow, 2006. (in Russ.)
7. <https://docplayer.ru/35735873-Kriteriy-vilkoksona-w-dlya-proverki-odnorodnosti-vyborok-v-1-2.html>. (in Russ.)
8. Kuznetsov L.A., Zhuravleva M.G. *Zavodskaya Laboratoriya. Diagnostika Materialov*, 2009, no. 1(75), pp. 70–75. (in Russ.)
9. Urusova E.S. *Obshchestvo. Sreda. Razvitiye*, 2015, no. 4(37), pp. 171–175. (in Russ.)
10. Stanskova A.G. *Otsenka ekologicheskogo sostoyaniya reki Okhta v Sankt-Peterburge i Leningradskoy oblasti* (Assessment of the Ecological State of the Okhta River in St. Petersburg and the Leningrad Region), St. Petersburg, 2016. (in Russ.)
11. Gmurman V.E. *Teoriya veroyatnostey i matematicheskaya statistika* (Theory of Probability and Mathematical Statistics), Moscow, 2003. (in Russ.)
12. Hollander M., Wolfe D.A., Chicken E. *Nonparametric Statistical Methods*, 3th Edition, 2014.
13. Davis C.S. *Statistical methods of the analysis of repeated measurements*, NY, Springer-Verlag, 2002.
14. Orlov A.I. *Nechyslovaya statystyka* (Non-Numerical Statistics), Moscow, 2004. (in Russ.)
15. Kutuzov O.I., Tatarnikova T.M. *Zavodskaya Laboratoriya. Diagnostika Materialov*, 2017, no. 3(83), pp. 65–70. (in Russ.)
16. Orlov A.I. *Zavodskaya Laboratoriya. Diagnostika Materialov*, 2004, no. 5(70), pp. 65–70. (in Russ.)
17. Orlov A.I. *Zavodskaya Laboratoriya. Diagnostika Materialov*, 2012, vol. 78. (in Russ.)

#### Data on author

**Vladimir N. Zavgorodniy** — Dr. Sci., Professor; Russian State Hydrometeorological University, Department of Marine Information Systems; E-mail: zavgor@list.ru

**For citation:** Zavgorodniy V. N. Method of nonparametric statistical analysis of homogeneity of ecological indicators of water bodies. *Journal of Instrument Engineering*. 2021. Vol. 64, N 5. P. 392–397 (in Russian).

DOI: 10.17586/0021-3454-2021-64-5-392-397