

УДК 556.11

КОЛИЧЕСТВЕННАЯ ОЦЕНКА КАЧЕСТВА ПРИРОДНОЙ ВОДЫ: ЭКСПЕРТНО-СТАТИСТИЧЕСКИЙ АНАЛИЗ

© 2022 г. Член-корреспондент РАН В. И. Данилов-Данильян^{1,*}, О. М. Розенталь¹

Поступило 10.10.2021 г.
После доработки 13.10.2021 г.
Принято к публикации 15.10.2021 г.

Информация о составе и свойствах воды, прежде всего в природных водных объектах, необходима для принятия водохозяйственных решений, поэтому она должна быть достаточно достоверна. Между тем водный контроль, оперирующий номинальной шкалой сравнения порогового норматива качества с вариабельными контролируруемыми показателями, необходимой точности результатов не дает. Для совершенствования контроля предложена модель оценивания вероятностной меры соответствия/несоответствия качества воды нормативным требованиям в непрерывной шкале. Показано, что при этом альтернативой статистическому подходу могут быть экспертно-статистические методы аппроксимации изменения искомой меры: простейший линейный, а также на основе шкал Тагути и Харрингтона. Для практически важных случаев малых объемов выборки и высокой вариации контролируемых показателей обоснована приемлемость использованных подходов. Это позволяет рекомендовать использование экспертно-статистических методов в системе водно-экологического контроля.

Ключевые слова: показатели качества воды, установление соответствия, зона неопределенности, доверительный интервал, вероятностная мера соответствия качества, строгая приемка, нестрогая приемка

DOI: 10.31857/S2686739722020049

ВВЕДЕНИЕ

С развитием промышленного производства мировая статистика отмечает ускоренное снижение качества пресных вод в зонах активного техногенеза [1]. Органы власти противодействуют этой опасной тенденции, ограничивая выдачу разрешений на сбросы сточных вод и пытаясь совершенствовать контрольно-надзорную деятельность. Такие попытки, однако, останутся безуспешными до тех пор, пока сохраняется оценка качества воды в номинальной шкале [2], по которой соответствие качества установленным требованиям признается, если концентрация C загрязняющего вещества по единичному измерению (C_i) или по среднему значению ($\bar{C} = \frac{1}{n} \sum_{i=1}^n C_i$) суммы n измерений не превышает порогового норматива C_N (например ПДК – предельно допустимой концентрации), а несоответствие – если эти величины больше C_N . Таким образом, недоучитывается “недопустимый риск, связанный с

причинением вреда” [3], и вода считается “соответствующей” при $\bar{C} \leq C_N$ или даже $C_i \leq C_N$, и “несоответствующей” при $\bar{C} > C_N$ или даже $C_i > C_N$.

Неприемлемость такой оценки следует из мониторинговой статистики. В частности, концентрация загрязнений в воде может быть соответствующей нормативу или несоответствующей в разные моменты времени, как это видно для цинка на рис. 1. Даже если концентрация остается по одну сторону норматива, как для никеля на рис. 1, использовать требования [2] невозможно, так как выборочные данные не гарантируют, что изменение состава воды происходит пропорционально приращению времени в межконтрольные (по терминологии [4]) периоды. Соответственно, штриховые линии на этом рисунке лишь соединяют точки измерений, но не характеризуют качество воды.

Таким образом, “стохастичность гидрохимического режима водных объектов” [5] плохо сочетается с практикой деления воду на соответствующую и не соответствующую нормативам. Это не отменяет существующим реалиям и порождает метрологические коллизии [6]. Было бы правильнее определять меру искомого соответствия/несоответствия результатов исследования, напри-

¹ Институт водных проблем Российской академии наук, Москва, Россия

*E-mail: vidd38@yandex.ru

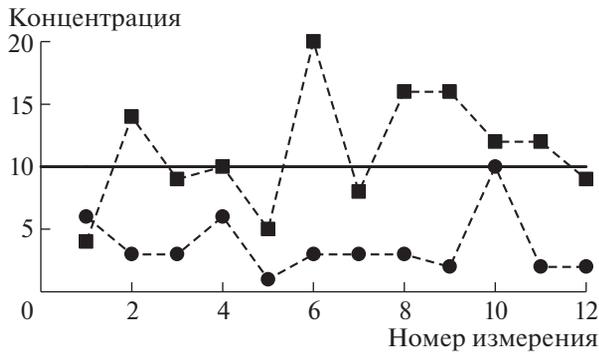


Рис. 1. Концентрация цинка (квадратные маркеры) и никеля (круглые маркеры), мкг/дм³, в воде р. Исети, створ г. Екатеринбург, ежемесячные измерения, 2011 г. Горизонтальная линия – ПДК (=10 мкг/дм³).

мер, с использованием статистических методов. С другой стороны, классические методы статистического анализа чувствительны к нарушению исходных предпосылок, а реальные процессы в природных водных объектах имеют нестационарный характер при неизвестных и изменяющихся значениях распределения вероятностей, что является сильным ограничением использования традиционных статистических методов и приводит к необходимости поиска альтернативных вариантов. Одним из возможных направлений для этого является использование так называемых экспертно-статистических методов, позволяющих обрабатывать статистическую информацию на основе экспертно задаваемых алгоритмов. Подобные методы в настоящее время нашли широкое применение в допусковом контроле для снятия неопределенности в пределах допуска [7]. Они также все более широко применяются в слабо формализованных задачах [8].

Целью статьи является сопоставление классического статистического метода принятия решения “соответствует – не соответствует” в том частном случае, когда это решение возможно, и экспертно-статистического метода при введении непрерывной функции потерь качества в пределах допуска.

СТАТИСТИЧЕСКИЙ ПОДХОД

Значительная вариация результатов измерений C_i в общем случае вызывается как случайным изменением концентрации в рассматриваемые периоды, так и погрешностями измерений. В условиях многих соизмеримых естественных и антропогенных факторов оценка \bar{C} , в соответствии с центральной предельной теоремой теории вероятностей, асимптотически нормальна с математическим ожиданием $\mathbf{M}[C]$ и дисперсией

$\frac{\sigma^2}{n} = \frac{\sigma_c^2}{n} + \frac{\sigma_0^2}{n}$, где σ_c^2 – дисперсия генеральной совокупности, из которой производится выборка n измерений, σ_0^2 – дисперсия случайной погрешности измерений.

Если σ_0^2 для аттестованных измерительных приборов установлена, то составляющая σ_c^2 как правило неизвестна. Вследствие этого вместо σ^2 используется ее оценка, произведенная по той же выборке, что и \bar{C} : $S^2 = \frac{1}{n-1} \sum_{i=1}^n (C_i - \bar{C})^2$. При этом центрированная и нормированная статистика

$$\frac{\bar{C} - \mathbf{M}[\bar{C}]}{S/\sqrt{n}} \quad (1)$$

будет иметь уже не нормальное, а t -распределение вероятностей Стьюдента с $(n-1)$ степенями свободы. Что же касается отклонения распределения концентрации C от нормального закона распределения вероятностей, то оно мало влияет на t -статистику [9].

Учет случайного характера статистики \bar{C} приводит к трансформации условий приемки и браковки воды, так что имеем $\gamma\%$ доверительный интервал, являющийся универсальной характеристикой точности оценки:

$$\bar{C}_H = \bar{C} - t_{\frac{1+\gamma}{2}} \frac{S}{\sqrt{n}}, \quad \bar{C}_B = \bar{C} + t_{\frac{1+\gamma}{2}} \frac{S}{\sqrt{n}},$$

где $t_{\frac{1+\gamma}{2}}$ – квантиль распределения Стьюдента, индексы H и B здесь и далее обозначают нижнюю и верхнюю границы доверительного интервала. При достаточном числе измерений доля отрицательной части левого “хвоста” функции распределения ничтожна и может не учитываться, как в приведенных ниже примерах.

Тогда проверяемое условие соответствия среднего значения концентрации нормативу запишется в виде:

для приемки – $\bar{C} < C_N - t_{\frac{1+\gamma}{2}} \frac{S}{\sqrt{n}}$ (решение $\bar{C} < C_N$ с вероятностью $\frac{1+\gamma}{2}$),

для браковки – $\bar{C} > C_N + t_{\frac{1+\gamma}{2}} \frac{S}{\sqrt{n}}$ (решение $\bar{C} > C_N$ с вероятностью $\frac{1+\gamma}{2}$).

Отсюда при $C_N - t_{\frac{1+\gamma}{2}} \frac{S}{\sqrt{n}} \leq \bar{C} \leq C_N + t_{\frac{1+\gamma}{2}} \frac{S}{\sqrt{n}}$ имеем зону неопределенности с вероятностью γ , в ко-



Рис. 2. Шкала значений норматива, где областью приемки является зона соответствия показателя качества нормативу, а областью браковки – зона несоответствия.

торой истинное значение $M[\bar{C}]$ может быть как меньше, так и больше норматива (рис. 2).

В математической статистике зону неопределенности обычно считают областью принятия решения $\bar{C} = C_N$. В других случаях ее с вероятностью $\left(\frac{1+\gamma}{2} + \frac{1+\gamma}{2}\right) - 1 = \gamma$ объединяют, в зависимости от конкретной постановки задачи, либо с зоной приемки ($\bar{C} \leq C_N$), либо с зоной браковки ($\bar{C} \geq C_N$).

Из-за малого объема выборки и больших значений среднеквадратического отклонения S зона неопределенности может быть значительной, например превышать 30% величины норматива. Поэтому возникает задача выяснения степени соответствия воды нормативу при попадании в эту зону.

Для решения этой задачи необходимо учесть, что имеется большое количество загрязняющих воду веществ с различными значениями нормативов – от долей мкг/дм^3 до десятков мг/см^3 , так что в качестве оценки степени соответствия целесообразно выбрать некоторую относительную единую величину, изменяющуюся в диапазоне 1–0, которая при приближении \bar{C} к нижней границе

$$C_{NH} = C_N - t_{1+\gamma} \frac{S}{\sqrt{n}} \tag{2}$$

стремится к единице, а при приближении к верхней границе

$$C_{NB} = C_N + t_{1+\gamma} \frac{S}{\sqrt{n}} \tag{3}$$

к нулю.

Наличие известного t -распределения позволяет выбрать в качестве меры, удовлетворяющей приведенным выше условиям, вероятностную характеристику, полученную при сравнении среднего значения \bar{C} с нормативом C_N . Действительно, приравнявая $M[\bar{C}] = C_N$, переходим к задаче проверки статистической гипотезы о равенстве

математического ожидания заданному значению. В этом случае при точном равенстве $\bar{C} = C_N$ из t -распределения имеем $t_\phi = 0$ для функции распределения $\phi = 0.5$. При $\bar{C} = C_{NB}$ из (1) получим $t_\phi = t_{1+\gamma}$, $\phi_B = \frac{1+\gamma}{2}$; при $\bar{C} = C_{NH}$ $t_\phi = -t_{1+\gamma}$, $\phi_H = 1 - \frac{1+\gamma}{2}$.

Из условий (2, 3) в качестве искомой меры следует выбрать $F_M = 1 - \phi$, т.е. $F_{M,B} = 1 - \phi_B = \frac{1+\gamma}{2}$;

$F_{M,H} = \frac{1-\gamma}{2}$. В частности, при $\gamma = 0.95$ получим $F_{M,B} = 0.975$, $F_{M,H} = 0.025$.

Пример 1. По данным рис. 1 оценить качество воды по цинку.

Решение. Здесь $n = 12$, $\bar{C} = 11.3 \text{ мкг/дм}^3$, $S = 4.7 \text{ мкг/дм}^3$, $C_N = 10 \text{ мкг/дм}^3$. Отсюда $C_{NH} = 6.6 \text{ мкг/дм}^3$, $C_{NB} = 13.4 \text{ мкг/дм}^3$, так что среднее значение находится в зоне неопределенности.

При этом $\frac{C_N - \bar{C}}{S/\sqrt{n}} = 0.95$, следовательно, $\phi = 0.82$,

а уровень меры соответствия воды установленным требованиям $F_M = 0.18$.

Примечание: здесь левый “хвост” t -распределения составляет $\frac{0-10}{4.7/\sqrt{12}} = -7.36$, т.е. доля отрицательной части функции ϕ ничтожна ($< 10^{-7}$).

Пример 2. По данным рис. 1 оценить качество воды по никелю.

Решение. Здесь $\bar{C} = 3.7 \text{ мкг/дм}^3$, $S = 2.5 \text{ мкг/дм}^3$, $n = 12$. Отсюда $C_{NH} = 8.2 \text{ мкг/дм}^3$, $C_{NB} = 11.8 \text{ мкг/дм}^3$, следовательно, \bar{C} находится в зоне строгой приемки, $F_M = 1$.

Отметим, что в целом статистический метод может быть применен при наличии известного распределения анализируемой статистики (здесь – оценки среднего значения). В других прикладных

Таблица 1. Результаты экспертно-статистического анализа по данным примера 1

Мера	C_{NH}	C_N	\bar{C}	C_{NB}
F_T	1	0.5	0.31	0
F_X	1	0.37	0.25	0.135

задачах, например при построении толерантного интервала (оценки доли распределения исследуемой выборки, отсекаемой нормативом), такая информация может отсутствовать. При этом альтернативой статистического подхода может стать экспертно-статистический с экспертным выбором функции снижения меры степени соответствия от области приемки к области браковки.

ЭКСПЕРТНО-СТАТИСТИЧЕСКИЙ МЕТОД

Простейшей аппроксимацией изменения меры оценки степени соответствия является линейная. В этом случае при $\gamma = 0.95$ нормативу будет соответствовать среднее значение меры $F_{M_{cp}} = \frac{0.975 - 0.025}{2} + 0.025 = 0.5$, а среднему значению \bar{C} из простейших геометрических построений – $F_M = \frac{0.975 - 0.025}{2t_{(1+\gamma)/2} \frac{S}{\sqrt{n}}} (C_{NB} - \bar{C}) + 0.025$. В частности,

в вышеприведенном примере 1: $F_M = \frac{0.95}{6.8} (13.4 - 11.3) + 0.025 = 0.325$, а в случае граничных условий $F_{M,B} = 1, F_{M,H} = 0: F_{M_{cp}} = 0.5, F_M = 0.31$.

Другими возможными экспертно-статистическими методами являются широко используемые в допусковом контроле функции Тагути и Харрингтона [10–12]. Это функции с двусторонним допуском, который устанавливают, например, для температуры и водородного показателя (рН) воды.

Оценка потерь качества по подобным показателям внутри поля допуска может быть выполнена с помощью функции Тагути:

$$F_T = \left(\frac{C - C_0}{C_m} \right)^2, \quad (4)$$

где $C_m = \frac{C_{NB} - C_{NH}}{2}$, $C_0 = \frac{C_{NB} + C_{NH}}{2}$ – номинал, начало отсчета потери качества.

В соответствии с моделью (4) вне допустимого интервала (при $C \geq C_{NB}, C \leq C_{NH}$) потеря качества максимальна ($F_T = 1$), тогда как фактически признаки соответствия воды нормативу могут присутствовать не только в допустимом интервале, но и вне его. Оценить уровень такого “присут-

ствия” позволяет функция желательности Харрингтона: $F_X = \exp\{-F_T(C)\}$.

Выбор в (4) степени $k = 2$ применяется в двусторонних функциях для симметризации отклонений в обе стороны от желаемых значений; такой выбор k не обязательно оптимален [13], потому что квадратичная форма функции (4) может слишком резко усиливать влияние больших отклонений C от C_0 на $F_X(C)$, занижая роль малых отклонений. Учитывая это, ниже в качестве конкурента квадратичной форме использована также линейная, позволяющая сгладить исследуемые зависимости.

В общем случае модифицированные односторонние функции имеют вид: $F_T = \left[\frac{\bar{C} - C_{NH}}{C_{NB} - C_{NH}} \right]^k$ (Тагути), $F_X = \exp\left\{-\frac{1}{2} \left[\frac{\bar{C} - C_{NH}}{C_{NB} - C_{NH}} \right]^k\right\}$ (Харрингтон).

В настоящей задаче выбор $k = 1$ обусловлен, во-первых, односторонностью функций, не требующих симметризации, а во-вторых, условием обеспечения при $\bar{C} = C_N$ значения $\phi = 0.5$, необходимого для сравнения рассматриваемого подхода со статистическим и логично вытекающего из практической интерпретации результатов расчета (табл. 1).

Как видно, значения функции Тагути совпадают с полученными для линейной функции потерь при том, что в целом все три рассмотренные экспертно-статистические функции обеспечивают близкие результаты, превышающие полученные статистическим методом. Однако при использовании функции Харрингтона количественному значению $\bar{C} = C_N$ соответствует не вероятность 0.5, что характерно для симметричных распределений, а меньшее значение 0.37. Это связано с тем, что реальное распределение вероятностей концентрации загрязняющего вещества несимметрично, характеризуется “тяжелым” правым “хвостом”.

При выборе степени $k = 2$ в выражении для функции Харрингтона получим: $F_X(\bar{C}) = 0.14$ при $\bar{C} = 11.3$; $F_X(C_{NB}) = 0.02$ при $C_{NB} = 13.4$, т.е. имеем более полное совпадение со статистической оценкой. Совпадение можно еще повысить, если нормировать статистическую оценку, т.е. перевести ее к граничному значению $\phi_B = 1, \phi_H = 0$ по формуле усеченных вероятностей [14]: $\phi = \frac{\phi_{yc} - \phi_H}{\phi_B - \phi_H} = \frac{0.835 - 0.025}{0.975 - 0.025} = 0.85$ (здесь $\phi_{yc} = \phi_B - F_X(\bar{C})$), так что $F_X = 1 - \phi = 0.15$.

Таблица 2. Оценки качества воды по модифицированной шкале Харрингтона

Качество воды	Очень плохое	Пониженное	Удовлетворительное	Хорошее	Очень хорошее
Балл	1	2	3	4	5
F_X	0–0.2	0.2–0.37	0.37–0.63	0.63–0.8	0.8–1.0

Ненулевое значение функции Харрингтона $F_X(C_{NB}) = 0.02$ изменяет результат расчета только в третьем знаке, т.е. может не приниматься во внимание.

В результате проведенного анализа получаем, что функция Харрингтона в квадратичной форме может быть рекомендована в качестве экспертной оценки искомой меры, близкой, но более жесткой, чем при статистическом подходе. При этом в шкале желательности Харрингтона [11, 12], характеризующей степень приближения показателей качества к некоторому идеалу и состоящей из пяти интервалов единичного отрезка, рассматриваемая вода “очень плохая” по цинку как при статистическом, так и при экспертно-статистическом оценивании (табл. 2).

В [13] предложено рассматривать функцию Харрингтона как субъективную вероятность. Исходя из такого подхода совместная вероятность соответствия нескольких (R) загрязняющих веществ определится как $\varphi_\Sigma = \prod_{i=1}^R F_i$. Так, в приведенном примере при $F_{X_1} = 0.16$ для цинка и $F_{X_2} = 1$ для никеля: $\varphi_\Sigma = 0.16 \times 1 = 0.16$ – определяется в соответствии с принципом “слабого звена”, когда наибольшее влияние оказывает загрязняющее вещество с максимальной концентрацией.

ЗАКЛЮЧЕНИЕ

Существующая классификация показателей качества воды в номинальной шкале неприемлема, когда дело касается их определения не в конкретный момент, а за определенный период, поскольку контролируемые показатели могут оказываться как соответствующими, так и не соответствующими нормативу в разные моменты времени, к тому же невозможно характеризовать их поведение в межконтрольные (между забором проб) периоды. Необходима более строгая шкала, выделяющая зоны соответствия и несоответствия с заданной доверительной вероятностью; при этом появляется зона неопределенности, которая из-за малого объема выборки и больших значений среднеквадратического отклонения показателей часто превышает 30% величины норматива.

Для выяснения качества воды в зоне неопределенности предложена вероятностная мера соответствия контролируемых показателей установленным требованиям, изменяющаяся в диапазоне 1–0 так, что при приближении средней

концентрации к нижней границе она стремится к единице, а при приближении к верхней – к нулю. Это позволило оценить качество воды в конкретных случаях, когда контролируемые показатели в разные моменты времени пересекают значение норматива или когда остаются по одну сторону от него.

Показано, что экспертно-статистические методы оценивания, обычно рассматриваемые как оценочные, в задачах количественной оценки качества воды являются сильными конкурентами статистических методов, чувствительных к нарушению исходных предпосылок. В условиях малых выборок и высокой вариации контролируемых показателей экспертно-статистические методы могут быть отнесены к классу робастных. Проведенный анализ показал приемлемость таких методов с позиций традиционного статистического оценивания (сложность аппарата которого препятствует его внедрению), что позволяет дать рекомендации к их широкому использованию в практически важных случаях принятия решений в условиях неопределенности, тем более что они позволяют оценивать общую вероятность соответствия качества воды нормативу по нескольким загрязняющим веществам.

ИСТОЧНИК ФИНАНСИРОВАНИЯ

Работа выполнена в рамках Государственного задания ФГБУН Института водных проблем РАН (тема № 0147-2019-0004, государственная регистрация № АААА-А19-119040990079-3).

СПИСОК ЛИТЕРАТУРЫ

1. Данилов-Данильян В.И. Водные ресурсы – стратегический фактор долгосрочного развития экономики России // Вестник РАН. 2009. Т. 79. № 9. С. 789–796.
2. СанПиН 2.1.3684-21 Санитарно-эпидемиологические требования к содержанию территорий городских и сельских поселений, к водным объектам, питьевой воде и питьевому водоснабжению населения, атмосферному воздуху, почвам, жилым помещениям, эксплуатации производственных, общественных помещений, организации и проведению санитарно-противоэпидемических (профилактических) мероприятий // <https://www.garant.ru/products/ipo/prime/doc/400189764/>
3. ФЗ 184 О техническом регулировании // http://www.consultant.ru/document/cons_doc_LAW_40241/

4. ГОСТ 8.586.2-2005 (ИСО 5167-2:2003) Измерение расхода и количества жидкостей и газов с помощью стандартных сужающих устройств. Часть 2 // http://www.peterburgregiongaz.ru/assets/files/metrologi/gost_8_586_2_2005.pdf
5. *Ленихин А.П., Возняк А.А., Тиунов А.А., Богомолов А.В.* К проблеме корректности методов расчетов и задания исходной гидрологической и гидрохимической информации при регламентации техногенных воздействий на водные объекты // Водное хозяйство России: проблемы, технологии, управление. 2017. № 1. С. 32–46.
6. *Авербух А.И., Розенталь О.М.* Анализ правил аналитического контроля качества вод в нормативно-правовых актах // Аналитика и контроль. 2015. Т. 19. № 4. С. 381–386.
7. Decision Rules Applied to Conformity Assessment. EUROLAB Technical Report 2017. № 1. 14 p.
8. *Pawari M.J., Gavande S.M.* Assessment of Water Quality Parameters: A Review // Int. J. of Science and Research. 2015. V. 4. Iss. 7. P. 1427–1431.
9. *Райфа Г., Шлейфер Р.* Прикладная теория статистических решений. М.: Статистика, 1977. 360 с.
10. *Taguchi G., Chowdhury S.* Taguchi's Quality Engineering Handbook. Wiley, 2004. 1696 p.
11. *Tyagi S., Sharma B., Singh P., Dobhal R.* Water Quality Assessment in Terms of Water Quality Index // American J. of Water Resources. 2013. 1 (3). P. 34–38.
12. *Voitenko L., Kopilevich V., Strokal M.* The Conception of Water Quality Assessment Used Harrington's Desirability Function for Different Kinds of Water Consumption // Біоресурси і природокористування. 2015. Т. 7. №1–2. P. 25–36.
13. *Розенталь О.М., Александровская Л.Н.* Оценка степени соответствия воды нормативным требованиям // Водные ресурсы. 2018. Т. 45. № 3. С. 289–297.
14. Р 50.1.037–2002. Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Часть 2. Непараметрические критерии // <https://files.stroyinf.ru/Data2/1/4294846/4294846286.pdf>

METHODOLOGY FOR QUANTITATIVE ASSESSMENT OF WATER QUALITY

Corresponding Member of the RAS V. I. Danilov-Danilyan^{a,#} and O. M. Rosenthal^a

^a *Institute of Water Problems of the Russian Academy of Sciences, Moscow, Russian Federation*

[#] *E-mail: vidd38@yandex.ru*

Information on the composition and properties of water, primarily in natural water bodies, is necessary for making water management decisions, therefore it must be sufficiently reliable. Meanwhile, water control, operating on a nominal generally accepted scale for comparing the threshold quality standard with variable controlled indicators, does not give the required accuracy of the results. To improve control, a model for assessing the probabilistic measure of compliance/non-compliance of water quality with regulatory requirements on a continuous scale is proposed. It is shown that, in this case, an alternative to the statistical approach can be expert-statistical methods for approximating the change in the desired measure: the simplest linear, as well as on the basis of the Taguchi and Harrington scales. For practically important cases of small sample sizes and high variation of controlled indicators, the acceptability of the approaches used is justified. This makes it possible to recommend the use of expert-statistical methods in the system of water-ecological control.

Keywords: water quality indicators, establishment of compliance, zone of uncertainty, confidence interval, probabilistic measure of quality compliance, strict acceptance, non-strict acceptance