

## Теоретические основы вероятностно-статистического разделения величины показателей химического состава водных объектов на природную и антропогенную составляющие

*На основе аксиоматики теории вероятностей, формулы свертки, симметричной экспоненциальной модели статистических распределений показателей химического состава водных объектов положено теоретическое обоснование и разработан алгоритм разделения величин концентраций растворимых веществ на природную и антропогенную составляющие путем идентификации законов из статистических распределений. Достоверность вычисления концентраций антропогенной составляющей колеблется от 0,88 до 0,97.*

**Состояние проблемы.** Важнейшим направлением современной гидрохимии является исследование физических, химических и гидробиологических процессов, определяющих трансформацию растворенных химических веществ и элементов. Количественные характеристики показателей химического состава природных водных объектов представляют собой интегральные величины, сформированные в результате комплексного действия природных и антропогенных факторов. Оценки антропогенной составляющей необходимы для прогноза загрязнения вод и разработки мер по управлению качеством воды. Применяемые в настоящее время подходы основаны на расчетах стока химических элементов с водосборных бассейнов. Погрешности расчета ионного стока составляют 30–50%, а погрешности оценок стока биогенных элементов и органических веществ могут достигать 100%.

Нами предложено теоретическое обоснование и разработан алгоритм достоверного вероятностно-статистического разделения количественных показателей содержания отдельных ингредиентов на природную и антропогенную составляющую.

**Материал и методика.** В основу исследования положены материалы ежедневных гидрохимических наблюдений за содержанием аммония ( $\text{NH}_4^+$ ,  $\text{мгN/дм}^3$ ) в замыкающем створе р. Десны в течение последних 10 лет. Для статистического анализа использованы аномалии ежедневных величин, вычисленные путем вычитания из фактических концентраций среднемноголетних значений (нормы) соответствующих суток.

В качестве методологической основы для разделения концентраций химических компонентов воды принята аксиоматика теории вероятностей. Исходя из ее эмпирических предпосылок [1], формирование природного содержания веществ хотя и протекает под влиянием постоянного комплекса условий, но величины концентраций оказываются различными. Фактические концентрации отдельных показателей химического состава воды ( $Y$ ) представлены суммой двух независимых составляющих, первая из которых обусловлена влиянием природного фактора ( $X_1$ ), а вторая — является следствием деятельности человека ( $X_2$ ):

$$Y = X_1 + X_2. \quad (1)$$

Величины отмеченных выше составляющих пропорциональны их вероятностям, что позволило вычислить значения  $X_1, X_2$  на основе законов их статистических распределений. Сами законы идентифицированы по параметрам статистических распределений показателей посредством модели, описывающих обширный класс симметричных экспоненциальных распределений [2]:

$$p(x) = \frac{\alpha}{2\lambda\sigma\Gamma(1/\alpha)} \exp\left(-\left|\frac{x-m}{\lambda\sigma}\right|^\alpha\right), \quad (2)$$

где  $\lambda = \sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}}$ ;  $\Gamma(z)$  — гамма-функция;  $\sigma = \sigma(x)$  — стандартное отклонение;  $m = m(x)$  — центр распределения;  $\alpha = \alpha(x)$  — некоторая характеристика распределения, однозначно определяющая параметр формы экспоненциальных распределений — эксцесс ( $\varepsilon$ ):

$$\varepsilon = \frac{\Gamma(1/\alpha)\Gamma(5/\alpha)}{(\Gamma(3/\alpha))^2}. \quad (3)$$

Если  $\alpha < 1$ , то модель (1) описывает распределения, близкие по своим свойствам распределению Коши;  $\alpha = 1$  — модель представляет распределение Лапласа, имеющее более островершинный вид, чем у нормального распределения;  $\alpha = 2$  — модель соответствует нормальному распределению;  $\alpha > 2$  — модель идентифицирует распределения, близкие по своим свойствам к трапецидальным распределениям;  $\alpha \rightarrow \infty$  — модель отражает равномерное распределение.

Параметры ( $m_Y, \sigma_Y, \varepsilon_Y$ ) законов статистических распределений фактических концентраций  $p_Y(y)$  получены на основе их временных рядов наблюдений ( $Y$ ).

Параметры ( $m_{X_1}, \sigma_{X_1}$ ) законов  $p_{X_1}(x)$  статистических распределений содержания веществ, обусловленного природными процессами ( $X_1$ ), как известно, близки распределению Гаусса, и получены из ранжированных временных рядов фактических концентраций, но только из той их части, которой свойственно нормальное распределение, показателем чего служили выборочные близкие к нулю значения эксцесса и эксцентриситета.

Параметры ( $m_{X_2}, \sigma_{X_2}$ ) законов статистических распределений содержания веществ, обусловленного деятельностью человека  $p_{X_2}(y-x)$ , определены на основе свертки [3]:

$$p_Y(y) = \int_{-\infty}^{+\infty} p_{X_2}(y-x)p_{X_1}(x) dx. \quad (4)$$

В частности, свертка распределений Гаусса дает распределение Гаусса. Следует подчеркнуть, что первые два момента симметричного экспоненциального распределения по формуле (2) и распределения Гаусса идентичны. В этой связи, первые два момента ( $m_{X_2}, \sigma_{X_2}$ ) статистических распределений концентраций антропогенного происхождения вычислены по известным соотношениям свертки нормальных распределений:

$$m_{X_2} = m_Y - m_{X_1}, \quad (5)$$

$$\sigma_{X_2} = (\sigma_Y^2 - \sigma_{X_1}^2)^{0,5}, \quad (6)$$

где  $m_{X_2}$  — центр статистического распределения концентраций, обусловленных деятельностью человека;  $m_Y$  — центр статистического распределения фактических концентраций;

$m_{X1}$  — центр статистического распределения концентраций природного происхождения;  $\sigma_{X2}$  — стандартное отклонение статистического распределения концентраций антропогенного происхождения;  $\sigma_Y$  — стандартное отклонение статистического распределения фактических концентраций;  $\sigma_{X1}$  — стандартное отклонение статистического распределения концентраций природного происхождения. В качестве верхней границы множества фактических концентраций гидрохимического показателя принимался их супремум (supremum).

На основе вычисленных значений параметров  $m_{X1}$  и  $\sigma_{X1}$ , посредством модели (2) при  $\alpha_{X1} = 2$ , идентифицировался закон статистических распределений концентраций природного происхождения. После этого с помощью функции распределения рассчитывались их вероятности. Аналогично, на основе вычисленных значений параметров  $m_{X2}$  и  $\sigma_{X2}$ , посредством модели (2) при  $\alpha_{X2} = 2$ , идентифицировался закон статистических распределений концентраций антропогенного происхождения, и рассчитывались их вероятности. Идентификация законов статистических распределений со значениями параметра формы распределения  $\alpha_{X2}$ , равными параметру формы статистического распределения  $\alpha_Y$  фактических концентраций, показала, что достоверность определения антропогенной составляющей колебалась в пределах 2%.

Верификация алгоритма разделения фактических аномалий концентраций на природную и антропогенную составляющую осуществлялась посредством численных экспериментов. Наблюденные значения концентраций аммония выборочно были увеличены экспертом; модифицированные временные ряды концентраций  $\text{NH}_4^+$  рассматривались в качестве аналога временных рядов содержания аммония, сформированного совокупным влиянием природных и антропогенных факторов. После завершения процедур разделения аномалий концентраций на природную и антропогенную составляющую были вычислены коэффициенты корреляции между выявленными концентрациями антропогенного происхождения и внесенными добавками.

**Обсуждение результатов исследования.** Главная проблема, возникающая при разделении концентраций показателей химического состава воды на природную и антропогенную составляющую, заключается в оценке параметров их статистических распределений, что осложняется нерегулярными циклическими трендами исходных временных рядов. Корректное исключение годового хода и многолетних природных циклическостей невозможно без наличия ежедневных долгосрочных наблюдений.

В случае отсутствия тренда во временном ряде аномалий концентраций  $\text{NH}_4^+$  (рис. 1, а), линия выборочных значений эксцесса ранжированного ряда аномалий (своей преобладающей частью) прижата к оси абсцисс (см. б на рис. 1). Частоты аномалий концентраций сосредоточены в первой трети диапазона колебаний (см. в), а законы статистических распределений концентраций  $\text{NH}_4^+$  природного и антропогенного происхождения имеют близкие центры и вариабельности (см. г). Достоверность выделения концентраций антропогенного происхождения с исключением годового хода — 0,954, без исключения годового хода — 0,953. Для пяти подобных временных рядов достоверность выделения концентраций антропогенного происхождения с исключением годового хода в среднем составила — 0,962, без исключения годового хода — 0,968.

Когда временной ряд аномалий концентраций  $\text{NH}_4^+$  содержит в себе циклический нерегулярный тренд, то линия выборочных значений эксцесса ранжированного ряда аномалий удалена от оси абсцисс. Частоты аномалий концентраций рассредоточены на значительной части диапазона колебаний с двумя модами. После удаления тренда аномалий законы статистических распределений концентраций природного и антропогенного происхождения

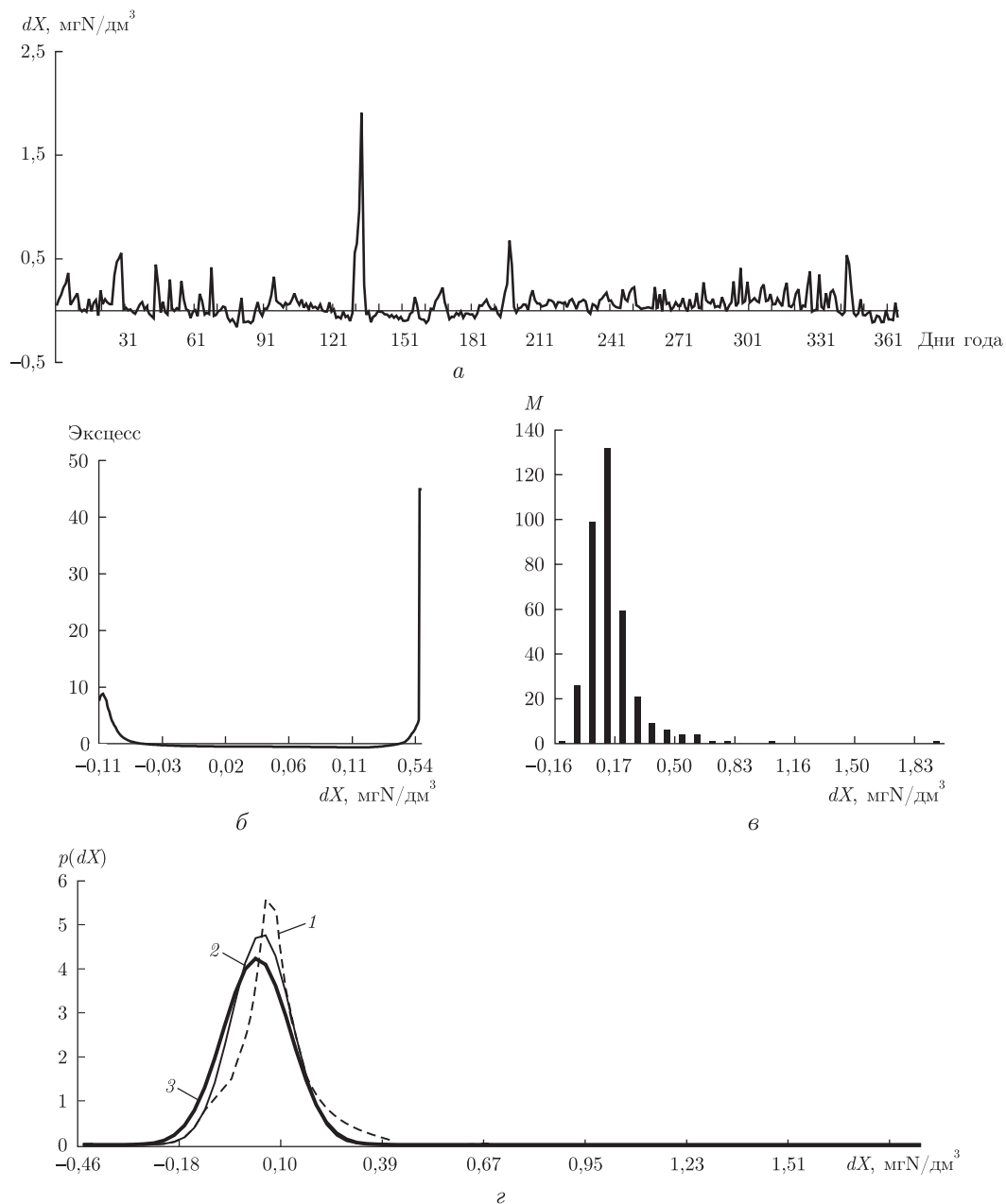


Рис. 1. Отклонения ежедневных концентраций аммония от нормы (аномалии концентраций  $dX$ , мгN/дм<sup>3</sup>, вариант 1) за годовой экзаменационный период (а); эксцесс ранжированного анализируемого временного ряда аномалий концентраций (б); гистограмма частот ( $M$ ) аномалий концентраций (в); законы статистических распределений аномалий концентраций: 1 — фактическое распределение, 2 — природная составляющая, 3 — антропогенная составляющая (г)

имеют смещенные центры, но близкие вариабельности. Достоверность выделения концентраций антропогенного происхождения с исключением годового хода — 0,879, без исключения годового хода — 0,782.

Наличие во временном ряде аномалий концентраций  $\text{NH}_4^+$  циклического нерегулярного тренда (рис. 2, а), может приводить к тому, что линия выборочных значений эксцесса

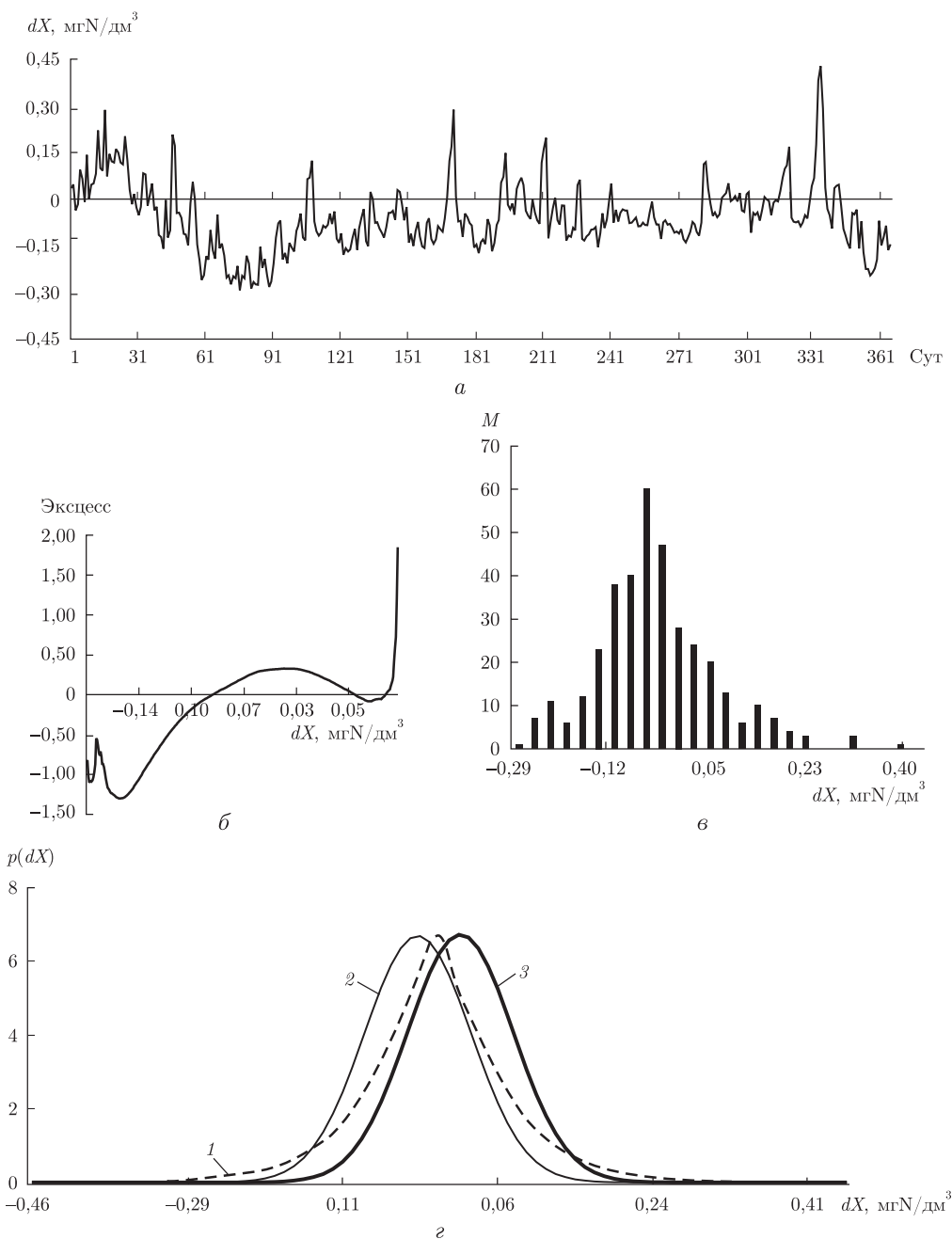


Рис. 2. Отклонения ежедневных концентраций аммония от нормы (аномалии концентраций  $dX$ ,  $\text{mgN}/\text{dm}^3$ , вариант 3) за годовой экзаменационный период (а); эксцесс ранжированного анализируемого временного ряда аномалий концентраций (б); гистограмма частот ( $M$ ) аномалий концентраций (в); законы статистических распределений аномалий концентраций: 1 — фактическое распределение, 2 — природная составляющая, 3 — антропогенная составляющая (г)

ранжированного ряда аномалий пересекает ось абсцисс (см. б на рис. 2). Частоты аномалий концентраций рассредоточены на преобладающей части диапазона колебаний (см. в); законы статистических распределений концентраций природного и антропогенного происхождения, после удаления тренда аномалий, имеют существенно смещенные центры и раз-

личные вариабельности (см. *z*). Достоверность выделения концентраций антропогенного происхождения с исключением годового хода — 0,848, без исключения годового хода — 0,776.

Подобный временной ряд отклонений концентраций аммония от нормы с циклическим трендом также сопровождается линией эксцесса, пересекающей ось абсцисс; слабой двух-модальностью частот аномалий концентраций; смещенными центрами статистических распределений концентраций природного и антропогенного происхождения, но близкой вариабельностью. Достоверность выделения концентраций антропогенного происхождения с исключением годового хода — 0,858, без исключения годового хода — 0,822.

Итак, теоретическая основа разделения наблюдаемых концентраций гидрохимических показателей на природную и антропогенную составляющую включает *вероятностно-статистический и гидрохимический аспекты*, заключающиеся в экспериментальных предположениях вероятностной аксиоматики и симметричных экспоненциальных законах статистических распределений гидрохимических компонент. Расчет параметров статистических распределений природной и антропогенной составляющей содержания показателей химического состава воды осуществляется посредством соотношений их параметров в свертке распределений, аппроксимируемых нормальными законами. Предложенный подход позволяет минимизировать процесс получения достоверной информации о влиянии природного и антропогенного факторов, определяющих трансформацию веществ в водной среде.

1. Гизман И. И., Скороход А. В. Теория случайных процессов. Том 1. – Москва: Наука, 1971. – 664 с.
2. Новицкий П. В., Зограф И. А. Оценка погрешностей результатов измерений. – Ленинград: Энергоатомиздат, 1985. – 248 с.
3. Яноши Л. Теория и практика обработки результатов измерений. – Москва: Мир, 1968. – 462 с.

Украинский научно-исследовательский  
гидрометеорологический институт  
НАН Украины и МЧС Украины, Киев

Поступило в редакцию 12.10.2012

Член-корреспондент НАН України В. І. Осадчий, Л. А. Ковальчук

### **Теоретичні основи ймовірнісно-статистичного розділення величини показників хімічного складу водних об'єктів на природну і антропогенну складові**

*На основі аксіоматики теорії ймовірностей, формули згортки, симетричної експоненціальної моделі статистичних розподілів показників хімічного складу водних об'єктів зроблено теоретичне обґрунтування та розроблено алгоритм розділення величин концентрацій речовин на природну і антропогенну складові шляхом ідентифікації законів їх статистичних розподілів. Достовірність обрахування концентрацій антропогенної складової коливається від 0,88 до 0,97.*

Corresponding Member of the NAS of Ukraine **V. I. Osadchiy, L. A. Kovalchuk**

**Theoretical bases for the probabilistic-statistical separation of the concentrations of dissolved substances in surface water bodies into natural and anthropogenic components**

*The theoretical explanation of the separation of the concentrations of dissolved substances in water into natural and anthropogenic components is given, by basing on the axiomatics of probability theory, the convolution formula, and the symmetric exponential model of statistical distributions of these components. The calculation algorithm of the separation of dissolved substances is developed by the identification of laws of their statistical distributions. Reliability of calculated anthropogenic component varies from 0.88 to 0.97.*