

УНИФИКАЦИЯ ЭМПИРИЧЕСКИХ РАСПРЕДЕЛЕНИЙ ПОКАЗАТЕЛЕЙ КАЧЕСТВА ВОДЫ НА ОСНОВЕ НОРМАЛЬНОГО И СТЕПЕННОГО ЗАКОНОВ

© 2014 г. А.И. Авербух¹, О.М. Розенталь²

¹ АНО «Институт независимых экспертиз», г. Екатеринбург

² Институт водных проблем Российской академии наук, Москва

Ключевые слова: законы распределения, показатели качества воды, степенной закон, унификация, усеченное нормальное распределение.



А.И. Авербух



О.М. Розенталь

Предложена унификация эмпирических распределений концентрации посторонних веществ в воде путем описания их левой и центральной частей усеченной функцией Гаусса, а правой, нередко «тяжелой» ветви – функцией степенного закона распределения. Показано, что подобная унификация естественна в важнейших на практике случаях повышенной

изменчивости состава вод, обусловленной сочетанием случайных (штатных) и неслучайных (нештатных) факторов. На примерах речных и сточных вод показана справедливость предложенной унификации, позволяющей свести многообразие форм исследуемых функций к рациональному минимуму, тем самым способствуя выработке единых правил управления устойчивым водопользованием.

Показатели качества воды, такие как концентрация посторонних веществ – величины случайные, непредвиденным образом меняющиеся во времени и пространстве [1–3]. Отношения полученных по результатам измерений стандартных отклонений (СКО) этих показателей к их средним значениям C^- нередко намного превышают общепризнанную границу однородности данных, равную 0,33 [4]. Поэтому при принятии водохозяйственных решений недостаточно учитывать одну из упомянутых величин C^- , хотя только она, как правило, приводится в статистических сборниках [5]. Корректное водное управление предполагает учет изменчивости контролируемых показателей, что на практике затруднительно вследствие многообразия законов распределения как различных веществ в одном выбранном водном объекте, так и каждого конкретного вещества в раз-

ных объектах [6, 7]. Построение и исследование эмпирических функций распределения показателей состава и свойств воды могут проводиться для решения широкого круга практических задач. Таковы оценка корректности и обоснованности использования методов статистического анализа при изучении гидрохимической информации, прогноз экстремальных событий, связанных с ухудшением качества воды, установление «статистических» фильтров отбраковки некорректных результатов анализа проб воды и т. д. [6, 7]. С целью упрощения подобных работ ниже рассмотрена возможность унификации эмпирических законов распределения путем анализа формирующих их причин – факторов случайной (штатной) и неслучайной (специальной, нештатной) природы.

Вода – динамичная, химически активная среда, контролируемые показатели состава и свойств которой формируются под влиянием огромного числа независимых случайных факторов, среди которых в обычных условиях ни один не является «решающим». В таких условиях следует ожидать нормального распределения контролируемых показателей. Другие распределения возникают в ситуациях, несвойственных указанным «обычным» условиям. Например, часто используемый закон Вейбулла формируется при наличии индукционного периода (задержки «отклика» системы на воздействие), а логнормальный закон – при мультипликативном воздействии влияющих факторов, тогда как в водной матрице это воздействие явно аддитивно. Поэтому описывать эмпирические распределения следует в первую очередь нормальным законом.

На первый взгляд подобная унификация не очевидна. Это следует, например, из рис. 1, демонстрирующего многообразие распределений показателей качества природных и сточных вод. Многообразие теоретических распределений (нормального, логнормального, бимодального и т. п.), которые согласуются с эмпирическими распределениями показателей состава и свойств вод, описано в [6, 8, 9]. По приведенным в табл. 1 значениям критерия Пирсона χ^2 сложно выбрать какой-либо теоретический закон (нормальный $F'(c)$, логарифмически-нормальный $F''(c)$ или Вейбулла $F'''(c)$) в качестве наиболее приемлемого для описания кривых рис. 1. Даже если при какой-либо доверительной вероятности P такое описание и было бы признано приемлемым, это оказалось бы малополезно для управления водопользованием, поскольку из-за высокой вариации (соизмеримости СКО и \bar{C}) удастся сопоставлять лишь отрезки распределений, ограниченные со стороны малых концентраций (слева). Например, концентрация сульфат-ионов ограничена наименьшим значением $6,99 \text{ мг/дм}^3$, так что теоретические и экспериментальные распределения можно сопоставить только в диапазоне концентраций, превышающих $\bar{C} - 1,1 \text{ СКО}$. Для других случайно выбранных веществ (см. рис. 1) наименьший результат измерений контролируемых показателей оказался меньше \bar{C} всего на $1,1 \text{ СКО}$ для нефтепродуктов (наименьший экспериментальный результат

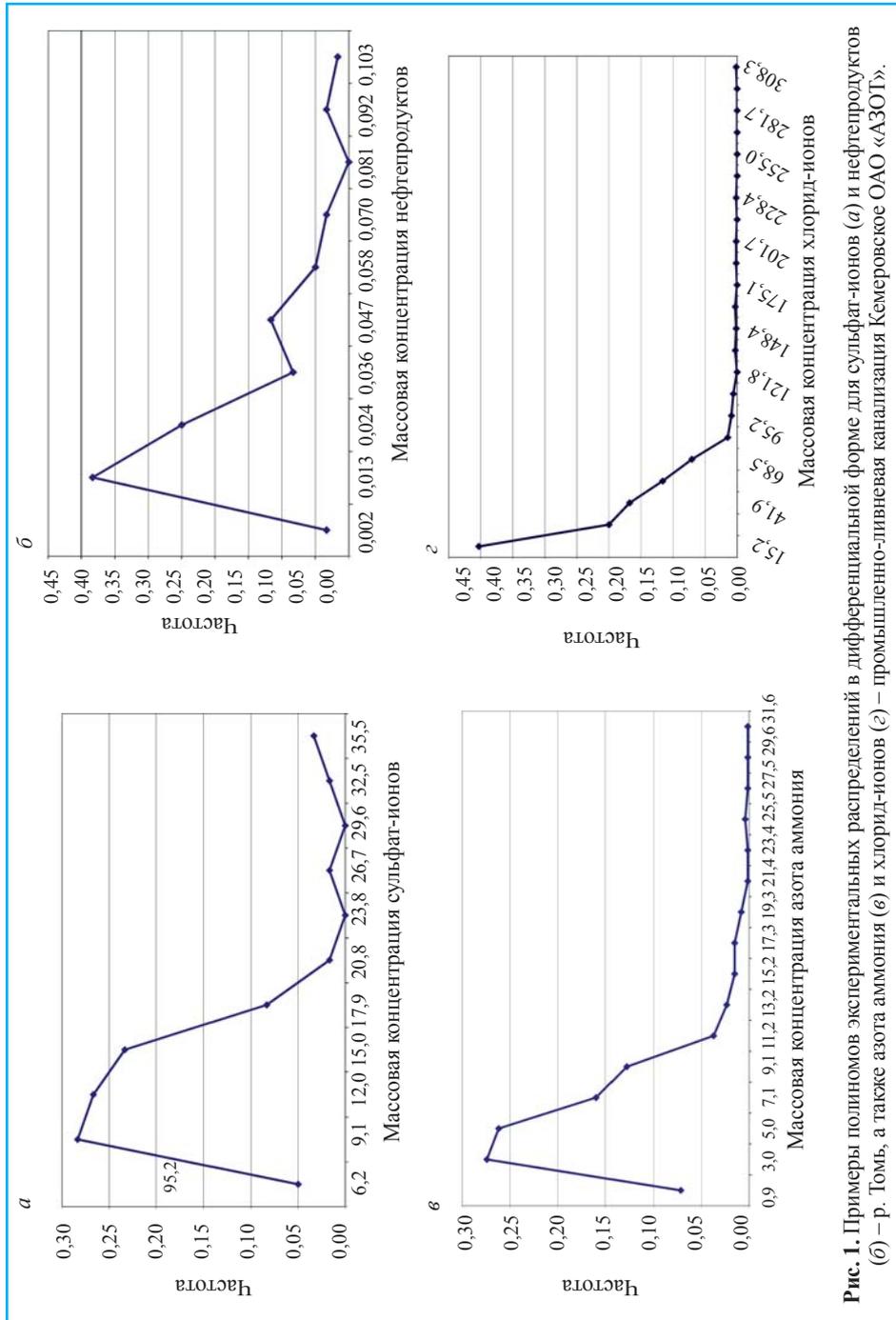


Рис. 1. Примеры полиномов экспериментальных распределений в дифференциальной форме для сульфат-ионов (а) и нефтепродуктов (б) – р. Томь, а также азота аммония (в) и хлорид-ионов (г) – промышленно-ливневая канализация Кемеровское ОАО «АЗОТ».

Таблица 1. Проверка соответствия эмпирических (рис. 1) и теоретических распределений

Показатель	Расчетное значение χ^2			Табличное значение χ^2 ($P = 0,99$)
	$F'(x)$	$F''(x)$	$F'''(x)$	
Сульфат-ионы	8,3	1,5	73,9	9,21 ($f = 2$)
Нефтепродукты	13,0	3,0	3,1	9,21 ($f = 2$)
Азот аммония	290 309	17,5	83,6	20,1 ($f = 8$)
Хлорид-ионы	399,5	62,7	111,8	15,1 ($f = 5$)

Примечание: f – число степеней свободы, $f = m - k$, где m – число классов разбиения, количество точек попадания в которые больше или равно 5; k – число параметров, необходимых для описания выборки.

0,0045 мг/дм³), на 1,4 СКО – для азота аммония (наименьший результат 0,36 мг/дм³) и на 0,99 СКО – для хлорид-ионов (8,78 мг/дм³). Так что значительная часть левой ветви реальных распределений переходит в область практически не измеряемых или вообще физически невозможных отрицательных значений контролируемых показателей. Само же распределение оказывается уже не симметричным и не асимптотическим, а усеченным [10], что необходимо учитывать при анализе его формы.

В свою очередь «тяжелая» правая ветвь должна рассматриваться как естественный результат влияния факторов неслучайной (нештатной) природы [11, 12], возможных «залповых» сбросов загрязняющих воду веществ [13], смывов, снеготаяния или других нежелательных природных или техногенных явлений [14].

Изложенное выше позволяет унифицировать рассматриваемые функции распределения путем сведения их многообразия к обоснованному рациональному минимуму. В частности, представленные на рис. 1 кривые следует рассматривать как результат суммарного влияния факторов случайной и неслучайной природы. Первые ответственны за гауссову составляющую распределения, усеченную со стороны малых значений показателей и имеющую плотность

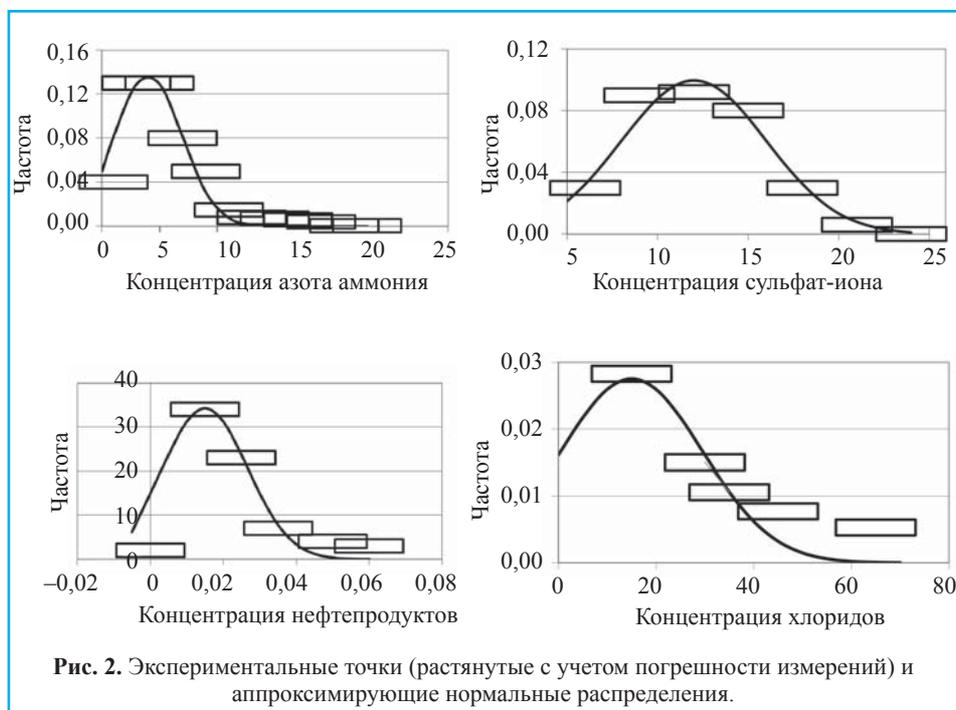
$$p(C) = \frac{\sqrt{\frac{2}{\pi}}}{\sigma \left(\operatorname{erf} \left(\frac{C}{\sqrt{2}\sigma} \right) + 1 \right)} \exp \left(-\frac{(C - \bar{C})^2}{2\sigma^2} \right),$$

где C , \bar{C} и σ – концентрация вещества, ее среднее значение и среднеквадратическое отклонение соответственно, а предэкспоненциальный множитель обеспечивает условие нормировки, при котором интеграл плотности рас-

предела от нуля до бесконечности равен единице. Факторы неслучайной природы ответственны за «тяжелый» хвост распределения справа.

Дополнительно действуют субъективные факторы, такие как ограниченная возможность химико-аналитических лабораторий оценивать слишком малые содержания посторонних веществ в воде, вследствие чего в химико-аналитических методиках установлена нижняя граница диапазона измерений, например, на уровне половины норматива качества [15]. Использование различных методик измерений для одного и того же показателя также определяет различие в получаемых результатах измерений [14]. Немаловажен и вклад погрешности измерений, для приоритетных показателей качества природных вод составляющей в среднем $\pm 40\%$ при концентрации, близкой к предельно допустимой, и $P = 0,95$ [3]. Это «размывает» эмпирические распределения в интервале значений показателя, накрывающим с известной (указываемой) доверительной вероятностью действительное значение измеряемой величины [16], шириной 80 % от результата измерений.

Итоги унификации распределений действительных значений контролируемых показателей, представленных на рис. 1 с выделением их нормальной (гауссовой) составляющей, даны на рис. 2. Показано, что распределения сульфат-ионов и азота аммония удовлетворительно описываются усеченным нормальным распределением даже без учета их «размывания» из-за погреш-



ности измерений вероятно потому, что исследование этих веществ особенно хорошо отработано на практике. Для хлоридов и нефтепродуктов такое описание приемлемо в области менее $\bar{C} + 2$ СКО уже с учетом «размывания».

Для прогноза нештатных ситуаций, ответственных за «хвосты» справа, полезно использование степенных законов [17], основанных на гипотезе о линейной зависимости логарифмов значений «выбросов» от логарифмов частоты их появления, как это было отмечено для геологических процессов. При этом незаметные оператору или малозаметные изменения, постепенно накапливаясь и достигая некоторого предельного уровня, создают нештатные ситуации, каковыми могут быть обрушения берегов природных водных объектов, залповые сбросы сточных вод [13], снеготаяние, смывы [14] и т. д. с «аномальным» увеличением содержания посторонних веществ в воде и появлением указанных «хвостов». Динамичность действительных значений показателей качества воды описывается преимущественно нормальным законом распределения с возможным правосторонним «хвостом». Такое распределение представляет функцию, состоящую из двух частей:

- описываемую законом Гаусса или усеченным законом Гаусса (для значений показателя, не превышающих $\bar{C} + 2$ СКО);
- описываемую степенным законом (для значений показателя, превышающих $\bar{C} + 2$ СКО) при наличии правостороннего тяжелого «хвоста».

И хотя количество «аномальных» результатов измерений значений, образующих «хвост», может быть невелико (для азота аммония на рис. 1в только 5 %, для хлоридов на рис. 1г – 3 %), все же именно на них следует в первую очередь указать специалисту, ответственному за устойчивое водопользование.

Для проверки подчиненности «хвостов» исследуемых эмпирических распределений степенному закону были рассчитаны и сопоставлены с табличными значениями коэффициенты корреляции $|r|$ между логарифмами значений контролируемых показателей и логарифмами частот, с которыми эти значения появляются (табл. 2). Гипотеза о подчиненности «хвостов»

Таблица 2. Результаты расчета коэффициента корреляции

Показатель	$ r $	Табличное $r_{\text{табл}}$ при $P = 0,99$
Сульфат-ионы	0,73	0,96 ($f = 3$)
Нефтепродукты	0,72	0,99 ($f = 2$)
Азот аммония	0,89	0,77 ($f = 8$)
Хлорид-ионы	0,89	0,59 ($f = 16$)

Примечание: f – число степеней свободы, $f = m - 2$; m – число интервалов разбиения диапазона значений, превышающих $\bar{C} + 2$ СКО.

исследуемых эмпирических распределений степенному закону принимает, если рассчитанное значение коэффициента корреляции превышает табличное значение на заданном уровне вероятности $P = 0,99$.

Как видно, лучшая корреляция, а следовательно, подчиненность степенному закону, наблюдается для концентрации азота аммония и хлорид-ионов, распределения которых характеризуются наиболее «тяжелыми хвостами». Например, при средних колебаниях значений концентрации хлорид-ионов в диапазоне 20–60 мг/дм³ появляются значения на уровне 100–160 г/дм³ (до 10,5 СКО) фактически залповые сбросы. То же для азота аммония: при преобладающей концентрации не выше 10 мг/дм³ наблюдаются значения на уровне 30 мг/дм³ (до 6,5 СКО). При этом максимальные концентрации сульфат-ионов и нефтепродуктов ограничиваются 4 СКО и 3,5 СКО соответственно, т. е. эмпирические распределения здесь не имеют «тяжелого хвоста», не соответствуют степенному закону и, с учетом «размывания» распределения за счет погрешности измерений могут быть описаны усеченным законом Гаусса. Параметры степенного закона распределения могут быть найдены методом наименьших квадратов или одним из способов, описанных в [17].

Подход, основанный на частичном соответствии эмпирического распределения контролируемого показателя нормальному закону распределения, а точнее – соответствия в ограниченном диапазоне значений измеряемой величины, рассмотрен, например, в [18].

В [19] показано, что при решении задач оценки соответствия экстраполяция результатов статистического моделирования целесообразна путем состыковки нормального распределения с распределением Парето в ситуации, когда эмпирический закон распределения совпадает с нормальным, и аппроксимировать целесообразно «хвосты» распределений. Такой подход отличается от подхода, рассмотренного в [20], где показано, что результат измерения объемной активности радона следует характеризовать смесью двух распределений, например, смещенного распределения Вейбулла–Гнеденко и распределения минимального значения или смещенного распределения Вейбулла–Гнеденко и двойного показательного распределения.

Таким образом, состыковка нормального распределения с одним из известных распределений принята в настоящее время при проведении оценки соответствия и представляется достаточно перспективной для описания эмпирических распределений изменения контролируемых показателей. Подобный подход позволяет достаточно адекватно описать эмпирическое распределение. С другой стороны, он дает возможность использовать преимущества нормального закона распределения, в частности, развитый математический аппарат, основанный на нормальном законе распределения.

Многие водохозяйственные решения были бы более надежными при статистическом анализе результатов измерений контролируемых показате-

лей качества воды с использованием эмпирических функций распределения полученных лабораториями массивов данных. При этом целесообразен метод унификации эмпирических функций распределения путем рассмотрения их формы как результата влияния факторов случайной (штатной) и неслучайной (нештатной) природы. Под воздействием первых формируется нормальный закон распределения, возможно, усеченный, как на рис. 2, вследствие соизмеримости \bar{C} и СКО. Под влиянием вторых – преимущественно «тяжелый» правый хвост, позволяющий контролировать аномально высокие уровни загрязнения воды – особые события, образно обозначенные в литературе по качеству управления «черными лебедями» [21]. При условии устойчивого водопользования «черные лебеди» встречаются нечасто. И если причины их появления выясняются, то это позволяет принимать необходимые технологические или экологические решения. Унификация законов распределения может быть полезна для выработки единообразных правил водопользования, в частности, путем системного наблюдения за «черными лебедями» мира водных отношений.

СПИСОК ЛИТЕРАТУРЫ

1. Walker D. Accuracy and precision in sampling water// ISO Focus. 2006. № 6. P. 21–24.
2. Сибгатуллина А.М., Мазуркин П.М. Измерение загрязненности речной воды (на примере малой реки Малая Кокшага). М.: Изд-во «Академия естествознания», 2009. 72 с.
3. Розенталь О.М., Авербух А.И. Введение в квалиметрию воды // Водные ресурсы. 2013. Т. 40. № 4. С. 1–16.
4. Авербух А.И., Дмитрук В.И. Опыт организации контроля качества воды на КОО «АЗОТ» // Методы оценки соответствия. 2012. № 7. С. 24–29.
5. Водные ресурсы и водное хозяйство России: стат. сборник. М.: НИИ-Природа, 2010. 380 с.
6. Лепихин А.П., Возняк А.А. Статистические функции распределения гидрохимических показателей качества воды поверхностных водных объектов // Водное хозяйство России. 2012. № 4. С. 21–32.
7. Долгоносов Б.М., Власов Д.Ю., Дятлов Д.В., Сураева Н.О., Корчагин К.А. Статистические характеристики изменчивости качества воды, поступающей на водопроводную станцию // Инженерная экология. 2004. № 3. С. 2–20.
8. Лепихин А.П., Мирошниченко С.А. Особенности задания «фоновой» концентрации в естественных водотоках // Водное хозяйство России. 2002. Т. 4. № 3. С. 247–262.
9. Банникова О.А., Бычкова Е.Н. К вопросу об установлении региональных нормативов качества воды // Водное хозяйство России. 2011. № 6. С. 54–68.
10. Clemen R.T. Making Hard Decisions: an Introduction to Decision Analysis. Brooks. Cole Publishers. 1996.
11. ГОСТ Р 51292–2000. Вода. Общие требования к отбору проб. Введ. 21.04.2000. № 117-ст.
12. НВН 33-5.3.01-85 «Нормативы водного надзора. Инструкция по отбору проб для анализа сточных вод». Утв. 13.06.1985. № 223.
13. ГОСТ 25150–82. Канализация. Термины и определения. Введ. 24.02.1982. № 805.
14. Долгоносов Б.М., Корчагин К.А. Экстремальные гидрохимические явления в речной системе в период снеготаяния // Водное хозяйство России. 2013. № 4. С. 20–31.

15. ГОСТ Р 8.613–2006. Государственная система обеспечения единства измерений. Методики количественного химического анализа проб вод. Общие требования к разработке. Введ. 26.10.2005. № 264-ст.
16. ПМГ 96-2009. Государственная система обеспечения единства измерений. Результаты и характеристики качества измерений. Формы представления. Введ. 05.04.2010. № 59-ст.
17. *Guerrero V.* Power Law Distribution: Method of Multi-scale Inferential Statistics // *Journal of Modern Mathematics Frontier*. 2012. V. 1 No. 1 P. 21–28.
18. *Кузнецов А.Г., Александровская Л.Н.* Непараметрические методы «измерения» малых рисков в задачах оценки соответствия требований к безопасности автоматической посадки самолетов нормам летной годности // *Труды Московского института электромеханики и автоматики (МИЭА)*. 2011. № 3. С. 2–11.
19. *Кузнецов А.Г., Александровская Л.Н.* Методы «измерения» малых рисков в задачах оценки соответствия требованиям к безопасности // *Методы менеджмента качества*. 2011. № 3. С. 48–53.
20. *Полосин В.Г., Тертычная С.В.* Методика разделения статистических данных для смеси двух распределений на примере результатов измерения объемной активности радона // *Изв. вузов. Поволжский регион. Физико-математические науки*. 2008. № 4. С. 79–86.
21. *Taleb N.N.* *The Black Swan: The Impact of the Highly Improbable*. New York: Random House, 2007. 370 p.

Сведения об авторах:

Авербух Алла Израйлевна, главный специалист, АНО «Институт независимых экспертиз», 620010, г. Екатеринбург, а/я 267; e-mail: allaaver@mail.ru

Розенталь Олег Моисеевич, д-р техн. наук, профессор, главный научный сотрудник, Институт водных проблем Российской академии наук, 117971, Москва, ул. Губкина, 3; e-mail: orosental@rambler.ru