

Л.А. Ковальчук

ПОГРЕШНОСТИ ИНТЕГРАЛА ВЕРОЯТНОСТЕЙ СТАТИСТИЧЕСКОГО РАСПРЕДЕЛЕНИЯ ОТЛИЧНОГО ОТ НОРМАЛЬНОГО, НО ВЫЧИСЛЕННОГО ПО МОДЕЛИ ГАУССА

Представлено аналитическую модель, описывающую симметричные логарифмически экспоненциальные распределения концентраций гидрохимических компонент. Рассчитаны погрешности оценивания вероятностей превышения ПДК по модели Гаусса, построены диаграммы погрешностей. Установлен диапазон значений эксцесса распределений, при которых вероятности превышения ПДК, оценённые по модели Гаусса с учётом погрешностей оценивания, приемлемы для решения практических задач.

Состояние проблемы

Достоверность оценивания вероятностей превышения предельно допустимых концентраций (ПДК) гидрохимических компонент посредством интеграла вероятностей зависит от законов статистических распределений концентраций. При гидрохимических исследованиях часто принимается гипотеза о нормальном или логарифмически нормальном распределении концентраций, что не всегда подтверждается результатами статистического анализа. Процедуры идентификации законов статистических распределений требуют привлечения специального математического аппарата и связаны со значительными объёмами вычислений. В случаях, когда анализируемые распределения близки к нормальным распределениям, оценки статистических параметров гидрохимических компонент, в частности вероятности превышения ПДК, выполняются по правилам нормального распределения, но при этом не указываются возникающие погрешности оценивания.

Цель настоящего исследования – установить допустимые пределы отклонений эксцесса от нормального значения, при которых погрешности интеграла вероятностей и вероятностей превышения ПДК, оценённых по нормальному закону с учётом погрешностей оценивания, приемлемы для решения практических задач.

Материалы и методика

В основу исследования положены материалы сезонных наблюдений Государственной гидрометеорологической службы Украины за последние три десятилетия на пятидесяти стационарных пунктах в основных речных бассейнах Днепра, Дуная, Днестра, Южного Буга, Западного Буга, Сиверского Донца, Приазовья и Крыма. Анализу были подвергнуты временные ряды сульфатов (SO_4^{2-}), хлоридов (Cl^-), кальция (Ca^{2+}), магния (Mg^{2+}), натрия (Na^+), железа ($\text{Fe}^{\text{общ.}}$), меди (Cu), цинка (Zn), хрома (Cr^{6+}), аммония (NH_4^+), нитратов (NO_2^-), нитритов (NO_3^-), фосфатных ионов, кремния, фенолов, нефтепродуктов, синтетических поверхностно активных веществ (СПАВ). Установлено, что гистограммы статистических распределений логарифмированных значений концентраций зачастую имеют куполообразный вид.

Методическая часть включала четыре этапа. На первом этапе, для идентификации законов статистических распределений концентраций, была привлечена аналитическая модель, описывающая широкий класс симметричных экспоненциальных распределений [1]:

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

где $\lambda = \sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}}$, $\Gamma(z)$ – гамма-функция: ($\Gamma(z) = \int_0^\infty e^{-t} t^{z-1} dt$, при $\text{Re} > 0$),

σ – стандартное отклонение, X_y – координата центра распределения, α – некоторая характеристика постоянная для анализируемого статистического распределения, которая однозначно определяет параметр формы экспоненциальных распределений – эксцесс (ε):

$$\varepsilon = \Gamma(1/\alpha)\Gamma(5/\alpha) / [\Gamma(3/\alpha)]^2, \quad (2)$$

При $\alpha < 1$ модель описывает распределения, близкие по своим свойствам распределению Коши; $\alpha = 1$ модель (1) 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 трапецеидальным распределениям; $\alpha \rightarrow \infty$ модель выражает равномерное распределение.

Приняв условие $\alpha = 2$, $X_{ii} = 0$, $\sigma = 1$, модель (1) была преобразована в модель нормального распределения:

$$p(x) = \frac{1}{\sqrt{2\pi}} \exp \left[-\left(\frac{x^2}{2} \right) \right], \quad (3)$$

посредством которой были рассчитаны значения плотности вероятностей распределения от 0 до 6. Для этого распределение центрировалось: $X_{ii} = 0$ и представлялось к вычислениям в виде отклонений от центра через 0,1 от -3 до +3, как аналог ранжированной последовательности при построении гистограмм распределений. Значения плотности вероятностей интегрировались методом трапеций, в результате чего были получены значения функции распределения (интеграл вероятностей).

На втором этапе был установлен диапазон значений эксцесса временных рядов гидрохимических компонент. Значения эксцесса были пересчитаны посредством выражения (2) в значения параметра α (табл. 1). Сохраняя условия: $X_{ii} = 0$, $\sigma = 1$ и меняя значения α согласно табл. 1, посредством модели (1) были вычислены с дискретностью 0,1 значения кривых распределения, плотности вероятностей и интегралы вероятностей различных законов статистических распределений от -3 до +3.

Таблица 1

Эксцессы распределений и соответствующие им значения α ,
использованные при численных экспериментах

Эксцесс	Параметр α	Эксцесс	Параметр α
1,900000	15,326920	3,761722	1,500000
2,000000	8,114994	4,000000	1,383831
2,631091	2,500000	5,000000	1,130523
2,787330	2,250000	6,000000	0,987010
3,000000	2,000030	7,000000	0,892513
3,302482	1,750000	8,000000	0,824588

На третьем этапе были вычислены разницы между значениями функции нормального распределения и значениями функций распределений, отличных от распределения Гаусса, когда $\alpha \neq 2$, но вычисленных по модели Гаусса. Полученные разницы были пересчитаны в относительные погрешности функции нормального распределения и представлены на диаграммах в процентах.

На четвёртом этапе для характерных временных рядов концентраций гидрохимических компонент посредством модели (1) были идентифицированы законы статистических распределений концентраций и вычислены вероятности превышения ПДК. Затем по модели Гаусса для тех же временных рядов были вычислены вероятности превышения ПДК, которые корректировались при помощи диаграмм и сопоставлялись с истинными значениями вероятностей превышения ПДК.

Обсуждение результатов исследования

При расчётах по модели Гаусса интегралов вероятностей пологих распределений со значениями эксцессов от 2,78 до 3,0, возникают как положительные, так и отрицательные погрешности: положительные – когда значения функции распределения Гаусса оказываются больше значений функций распределений идентифицированных статистических законов; отрицательные – когда функция распределения Гаусса меньше фактической функции распределения (рис. 1). В диапазоне вероятностей от 0 до 5% относительные погрешности расчётов стремительно уменьшаются с 25% до нуля; при вероятностях от 5 до 50% погрешности становятся отрицательными, не опускаясь ниже 5%, и остаются близкими к нулю при вероятностях более 50%.

В случае распределений со значениями эксцессов от 3 до 3,3, более островершинных чем кривая Гаусса, погрешности интегралов вероятностей, рассчитанные по модели Гаусса, в диапазоне вероятностей 0-5% имеют отрицательный знак и стремительно уменьшаются от 30% до 0%. Затем, при вероятностях от 5 до 50%, погрешности приобретают положительные значения и по абсолютной величине не превышают 5%, а при вероятностях более 50% – колеблются около 0% (рис. 1).

В диапазоне значений эксцесса 2,63-2,79 и 3,30-4,53 относительные погрешности интегралов вероятностей, вычисленных посредством модели Гаусса, достигают 50-80% при минимальных вероятностях. Тем не менее, сохраняются общие черты погрешностей: они максимальные при вероятностях от 0 до 5-6%, изменяют свой знак в диапазоне вероятностей 6-60%, и незначительны при вероятностях более 60% (рис. 2).

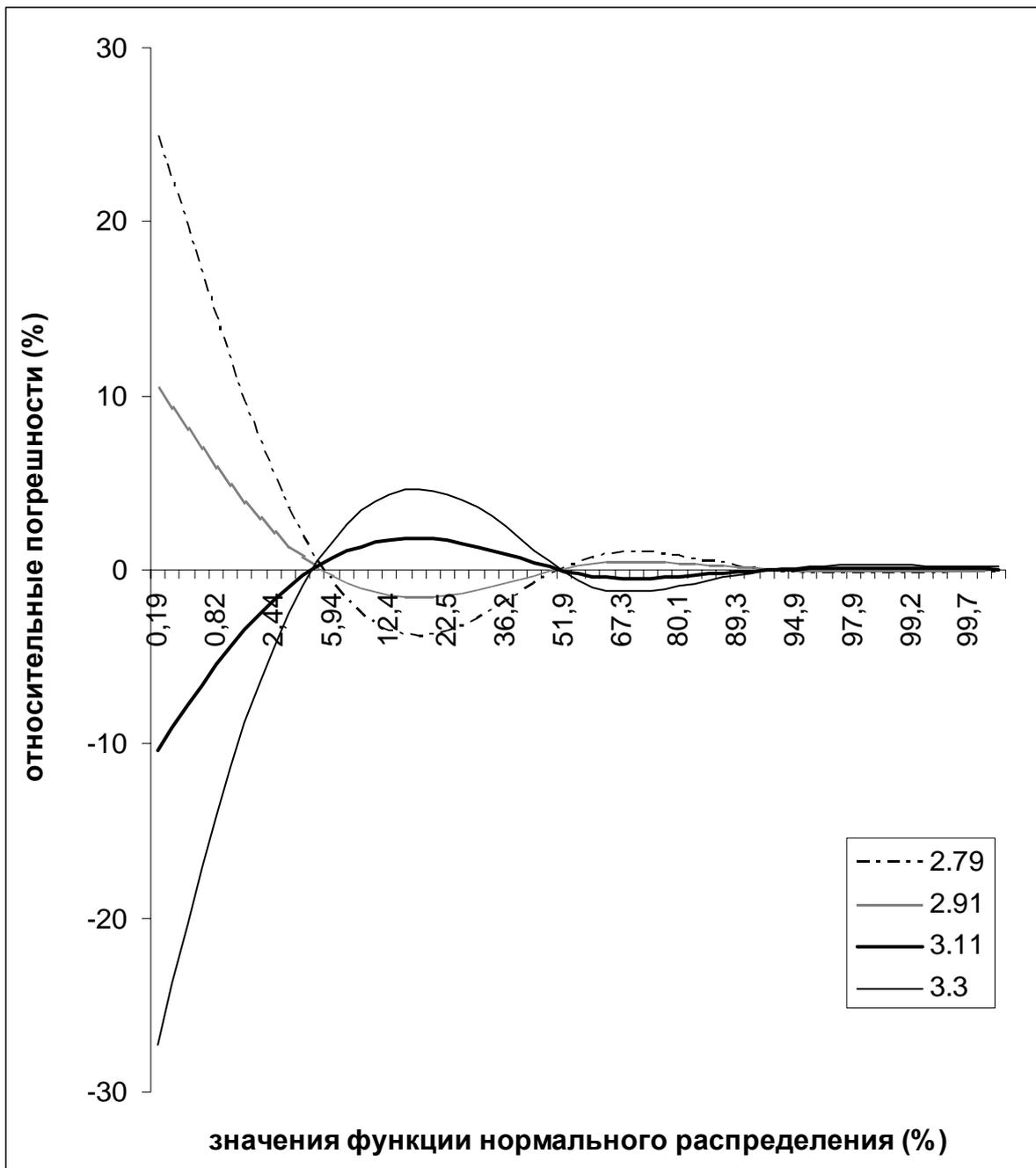


Рис. 1. Погрешности интегралов вероятностей распределений со значениями эксцесса 2,79, 2,91, 3,11, 3,30, оценённых по модели Гаусса

В диапазонах значений эксцесса 1,90-2,63 и 4,53-8,00 сохраняются общие черты относительных погрешностей интегралов вероятностей, вычисленных по модели Гаусса (рис. 3). Однако их значения существенно увеличиваются. При трапецеидальных распределениях с эксцессом около 2 и значениями интеграла вероятностей менее 4% относительные погрешности достигают +100%, опускаются до -40% в диапазоне вероятностей 10-50%, снова поднимаются до +10% при вероятностях

около 70-80%, после чего плавно уменьшаются до незначительных величин.

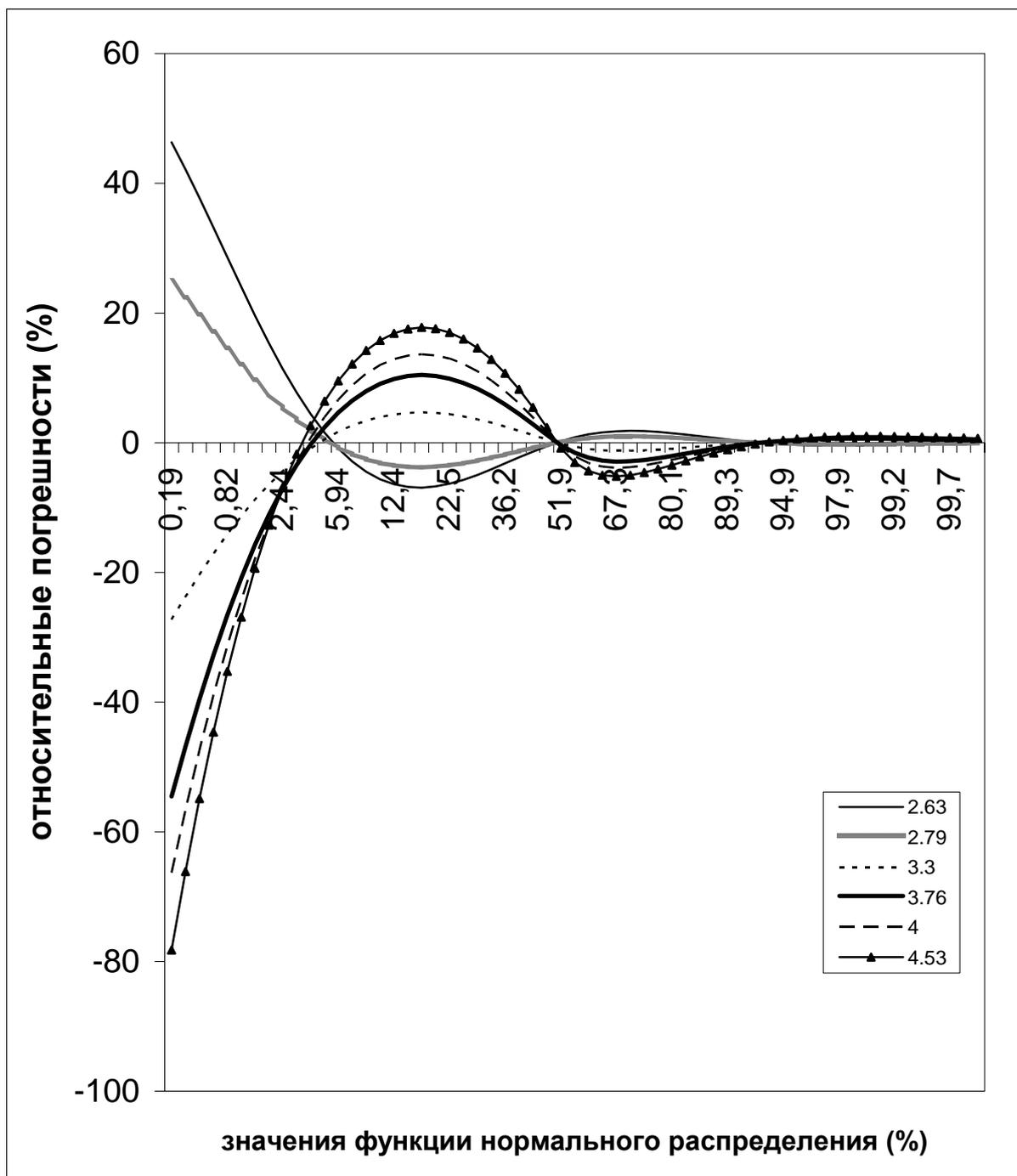


Рис. 2. Погрешности интегралов вероятностей распределений со значениями эксцесса 2,63, 2,79, 3,30, 3,76, 4,0 и 4,53, оценённых по модели Гаусса

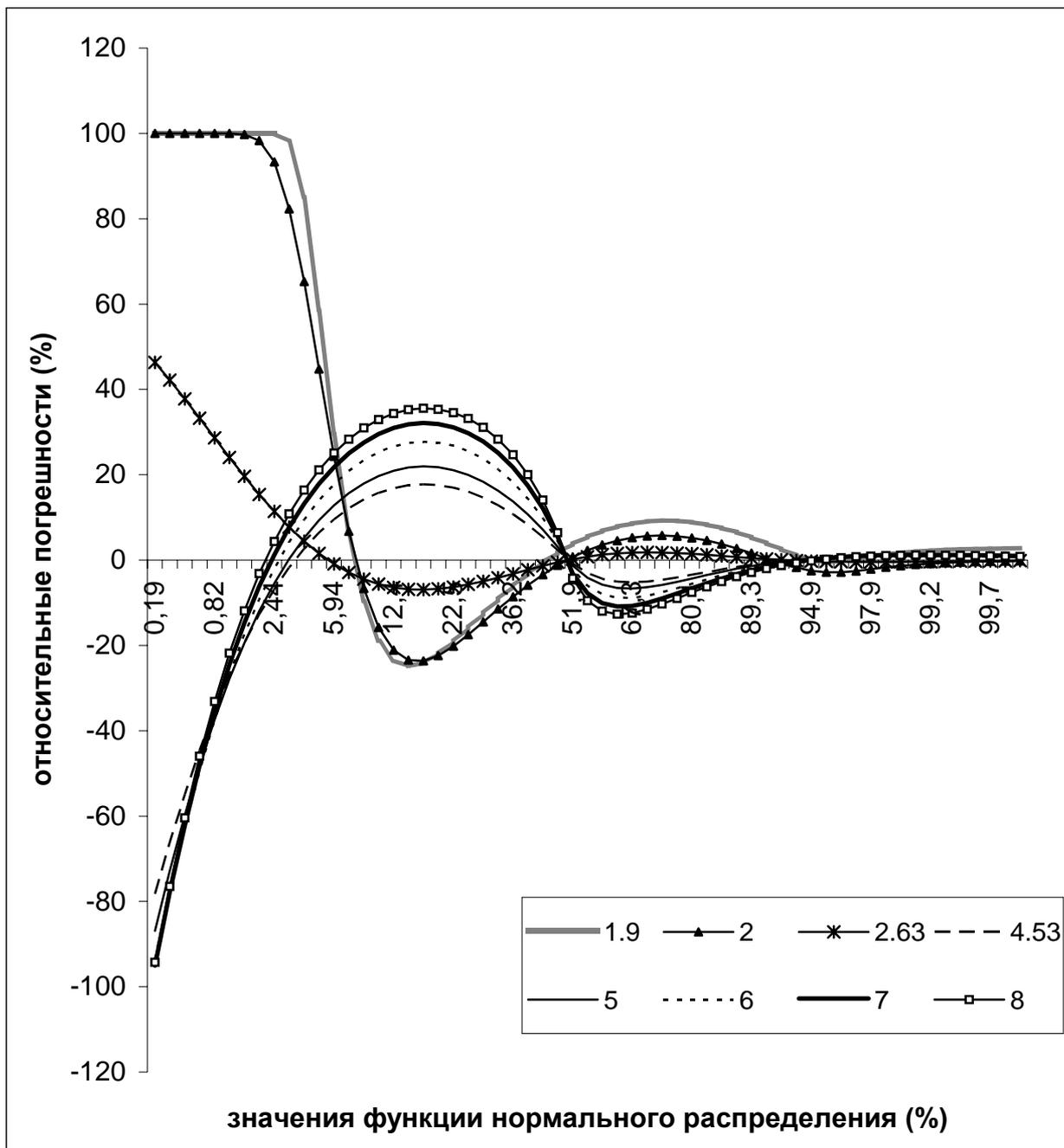


Рис. 3. Погрешности интегралов вероятностей распределений со значениями эксцесса 1,90, 2,00, 2,63, 4,53, 5,00, 6,00, 7,00, 8,00, оценённых по модели Гаусса

В диапазоне значений эксцесса 4,53-8,00 распределение погрешностей однородно: при малых вероятностях от 0 до 5% отрицательные значения погрешностей уменьшаются от 80-100% до 0%, в диапазоне вероятностей 5-50% погрешности увеличиваются до +20-+40%, при вероятностях более 50% погрешности сначала снижаются до -15%, а затем уменьшаются до 2-5%.

Таблица 2

Сопоставление вероятностей превышения ПДК фосфатных ионов, вычисленных по фактическим логарифмически экспоненциальным моделям и по моделям Гаусса с последующим учётом погрешностей оценивания

Местоположение пункта наблюдений			Эксцесс	Вероятности превышения ПДК (%), вычислен. по		Корректировка вероятн. превыш. ПДК, вычисленной по модели Гаусса (%)		
Бас-сейн	Река	Пункт		фактической модели	модели Гаусса	поправка	исправленное значение	Погрешн. коррек-тики
1	2	3	4	5	6	7	8	9
Дунай	Прут	Яремча	3,86	10,81	11,14	-1,31	9,83	0,98
Западн. Буг	Луга	Вл.-Волынский	4,21	19,78	24,87	-3,93	20,94	-1,16
Закарпатье	Латорица	Подполье	2,19	33,22	26,82	3,48	30,30	-2,92
Днестр	Зол. Липа	Бережаны	4,68	83,74	81,03	4,86	85,89	-2,15
Припять	Турья	Ковель	5,26	5,91	6,89	1,3091	5,58	-0,33
Днепр	Хорол	Миргород	3,22	91,15	90,09	0,09	90,18	-0,97
Сив. Донец	Уды	Харьков	4,97	98,61	99,35	-0,49	98,85	0,24
Южн. Буг	Ингул	Кировоград	3,18	93,13	92,93	0,09	92,84	-0,29
Приазовье	Кальчик	Кременёвка	3,05	29,37	28,19	0,06	28,13	-1,24
Крым	Салгир	Пионерское	1,96	30,25	23,82	3,79	27,61	-2,64

Сопоставление вероятностей превышения ПДК фосфатных ионов в поверхностных водах Украины, вычисленных по фактическим логарифмически экспоненциальным моделям и по моделям Гаусса, с последующим учётом погрешностей оценивания показало, что наибольшие различия, достигающие 2,64% и 2,92% характерны для трапецеидальных распределений со значениями эксцессов 1,96 и 2,19 (табл. 2, рис. 4 а, б).

а)



б)



в)



Рис. 4. Плотность вероятностей логарифмов концентрации фосфатных ионов: а) р. Латорица, с. Подполозье, эксцесс 2,19; б) р. Салгир, с. Пионерское, эксцесс 1,96; в) значения $p(x)$ кривой распределения логарифмов концентрации фосфатных ионов р. Турья, г. Ковель, эксцесс 5,26

В тоже время обострённость кривой распределения со значением эксцесса 5,26 не всегда влечёт за собой значительные различия в оценках вероятностей превышения ПДК (табл. 2, рис. 4 в).

Выводы

1. Статистические распределения логарифмов концентраций гидрохимических компонент поверхностных вод Украины характеризуются значениями эксцесса от 1,90 до 8,00, что охватывает диапазон симметричных логарифмически экспоненциальных распределений.

2. Если не требуется высокая точность расчёта, то оценки вероятностей превышения ПДК гидрохимических компонент допустимо проводить согласно логарифмически нормальной модели с учётом погрешностей оценивания посредством предложенных диаграмм.

3. При необходимости высокоточного оценивания вероятностей превышения ПДК гидрохимических компонент рекомендуется использовать аналитическую модель, описывающую весь диапазон симметричных экспоненциальных распределений.

* *

Представлено аналітичну модель, що описує симетричні логарифмічно експонентні розподіли концентрацій гідрохімічних компонент. Розраховані похибки оцінювання ймовірностей перевищення ГДК по моделі Гауса, побудовані діаграми похибок. Встановлено діапазон значень ексцесу розподілів, при яких імовірності перевищення ГДК, оцінені по моделі Гауса з урахуванням похибок оцінювання, прийнятні для рішення практичних задач.

* *

1. Новицкий П.В., Зограф И.А. Оценка погрешностей результатов измерений. – Л.: Энергоатомиздат, Ленингр. отд-ние, 1985. – 248 с.

*Український науково-дослідницький
гідрометеорологічний інститут, Київ*