ГИДРОЛОГИЯ

УДК [556.555:574.5]:51-7 doi: 10.33933/2074-2762-2021-62-27-51

Выделение озер, обладающих уникальными свойствами, статистическими методами

А.В. Измайлова, А.М. Расулова, В.Ю. Шмакова

Институт озероведения Российской академии наук — обособленное структурное подразделение Федерального государственного бюджетного учреждения науки «Санкт-Петербургский Федеральный исследовательский центр Российской академии наук», Санкт-Петербург, ianna64@mail.ru

В рамках настоящей статьи на примере озерных экосистем обсуждается возможность применения математико-статистических подходов при обоснованиях их уникальности. Рассмотрены математические методы, которые по различным признакам (параметрам), характеризующим озерные экосистемы и происходящие в них процессы, позволяют выделить уникальные, единственные в своем роде, озера. Сделан обзор методов, которые могут быть применены для обработки данных и выявления выбросов морфометрических характеристик водоемов, имеющих вулканическое происхождение. Выполнено сопоставление результатов, полученных при использовании различных критериев, и предложена экспертная оценка итогового набора озер на предмет уникальности их экосистем.

Ключевые слова: озерная экосистема, уникальные озера, идентификация аномалий, статистические выбросы, вулканические озера.

Identification of lakes with unique properties using statistical methods

A.V. Izmailova, A.M. Rasulova, V.Yu. Shmakova

Institute of Limnology of the Russian Academy of Sciences — a separate structural unit of the Federal State Budgetary Institution of Science “St. Petersburg Federal Research Center of the Russian Academy of Sciences”, Saint-Petersburg, Russia

The article considers the possibility of using mathematical and statistical approaches to substantiate the uniqueness of lake ecosystems. The study of lakes of a volcanic type for anomalous values of morphometric characteristics has been carried out. The following morphometric characteristics of lakes have been considered: length, width, water area, coastline length, average and maximum depths, catchment area, as well as dimensionless quantities calculated on their basis. Analysis of the data has shown the statistical distribution of the values to correspond to the gamma distribution, lognormal, chi-square or Weibull and to be easily normalized using the Box-Cox transformation. The identification of anomalous values has been carried out using graphical methods (construction of scatter plots, probability graphs, box-and-whiskers plots) and statistical criteria. In this work, in accordance with the type of distribution and the size of the sample, the following criteria for identifying outliers have been chosen: Smirnov-Grubbs, Hoaglin-Iglewicz, Tietjen-Moore, Chauvinet, Irwin and Thompson's method (Rosher's test). Comparison of the results obtained according to different criteria has been carried out, the box diagram and the Thompson method proving to be the best.

As a result of the analysis, 65 lakes with anomalous values for a number of characteristics have been identified. An expert assessment of the list obtained has shown that most of the lakes can be considered as unique, taking into account other parameters that characterize lake ecosystems — due to the peculiarities of

their hydrochemical regime and the originality of the biota. Along with the large and deepest water bodies, the list includes 18 small lakes (less than 1 km^2), the identification of which being difficult in a large sample without using the proposed apparatus.

Keywords: lake ecosystem, unique lakes, anomaly identification, statistical outliers, volcanic lakes.

For citation: *A.V. Izmailova, A.M. Rasulova, V.Yu. Shmakova.* Identification of lakes with unique properties using statistical methods. *Gidrometeorologiya i Ekologiya.* Journal of Hydrometeorology and Ecology (Proceedings of the Russian State Hydrometeorological University). 2021. 62: 27—51. [In Russian]. doi: 10.33933/2074-2762-2021-62-27-51

Введение

В связи с необходимостью совершенствования научных основ развития и расширения сети особо охраняемых природных территорий (ООПТ) часто стоит вопрос об отсутствии объективных критериев, позволяющих отнести ту или иную территорию или акваторию к категории «уникальных», руководствуясь при этом не только экспертными оценками. Поиск таких критериев представляет крайне актуальную задачу, решение которой позволит выделять объекты, особо нуждающиеся как в охране, так и в проведении мониторинговых исследований. В рамках настоящей статьи на примере озерных экосистем рассмотрена возможность применения статистических подходов для выявления выбросов из выборок, построенных по различным параметрам озерной экосистемы, которые могут свидетельствовать об уникальности водного объекта.

Как отмечено в работе [1], каждая озерная экосистема сформирована при определенном наборе внешних факторов и отличается совокупностью характерных лишь ей внутренних свойств и параметров. Однако при всем своем многообразии, сложившиеся в схожих условиях озерные экосистемы обычно обладают значительным подобием. На основе сходства элементов строятся их классификации [2]. На фоне большинства озер, схожих по целому ряду параметров, всегда находятся водоемы, выделяющиеся экстремальными, резко отличными от других характеристиками, благодаря которым они могут рассматриваться как уникальные. Поэтому, хотя каждая озерная экосистема по-своему своеобразна, для признания ее уникальности она должна отличаться неким специфическим набором свойств или характеристик, свидетельствующих об ее исключительности. Другими словами, если все озера мира представить, как статистическую совокупность, обладающую набором признаков (параметров), то большинство выборок, построенных по этим параметрам, может быть описано с помощью теоретического распределения. В то же время по некоторым параметрам будут наблюдаться выбросы, резко выделяющиеся из общей выборки.

Описание озера включает в себя большое количество характеристик — параметры, описывающие морфологию и морфометрию его котловины, химический состав воды, ее физические свойства, а также биологическое разнообразие сложившейся в процессе эволюции экосистемы. При этом сами характеристики являются многопараметрической функцией, зависящей в основном от таких факторов, как географическое местоположение, рельеф, климатические условия, антропогенное воздействие и др.

Следует заметить, что параметры, представляющие абиотическую среду, легче определяются. Это позволило к настоящему времени накопить большой банк данных по озерам мира, содержащий сведения о различных морфометрических характеристиках озерных котловин, их происхождении, ряде химических показателей. Оценка биоценозов по степени сложности их организации менее очевидна и, несмотря на наличие индексов, характеризующих разнообразие, пока недостаточно параметризована. Анализ имеющихся в ИНОЗ РАН баз данных по озерам [3—6] свидетельствует, что накопленных к настоящему времени данных по морфометрии водоемов и ряду химических показателей качества воды достаточно для проведения полноценного математического анализа не только на уровне генеральной совокупности всех озер мира, но и по отдельной статистической совокупности озер, выделенных, например, с учетом происхождения их котловин. В то же время необходимо признать недостаточность данных для проведения корректного анализа с использованием параметров, характеризующих биотическую среду водоемов, расположенных в различных ландшафтах.

В соответствии с гипотезой, предложенной в [1], вследствие взаимной обусловленности происходящих в озерах процессов, если водная экосистема отлична от остальных на определенной территории по одному признаку, то она может выделяться и по ряду других признаков. Поэтому при идентификации выбросов по разным параметрам они часто будут выявлены для одних и тех же объектов. То есть озеро имеет аномальные значения по целому ряду характеристик. При этом, как неоднократно указывалось [2, 7—8], генетическая принадлежность озерных котловин и их морфометрические особенности в значительной степени предопределяют протекающие в них процессы и трансформацию устойчивости лимносистем. Например, морфометрические показатели формы озерной котловины влияют на устойчивость водных масс, определяя термодинамическое перемешивание. Это позволяет на начальном этапе исследования сконцентрироваться именно на анализе морфометрических характеристик озер.

Цель данной статьи — показать возможность выделения уникальных озер на основе критериев математической статистики по морфометрическим характеристикам водных объектов. По генеральной совокупности озер мира была проведена классификация по генезису озерной котловины и выделен класс объектов, имеющих вулканическое происхождение. Хорошо известно, что многие из таких озер имеют экстремальные гидрохимические характеристики, что позволит в дальнейшем детально проанализировать полученные в ходе исследования результаты на предмет аномалий по иным признакам. Кроме того, необходимо отметить и относительно высокую лимнологическую изученность данного типа озер, что облегчает поставленные задачи.

Методика исследования

В простейшем случае статистический поиск аномалий в распределении сводится к задаче о выявлении выбросов. Под выбросом понимается наблюдение, резко выделяющееся из набора данных. Однако стоит учитывать отличия аномалий

от выбросов — последние могут включать в себя ошибки измерений (в том числе точности измерений и неправильную регистрацию данных). В рамках настоящего исследования была использована дополненная авторами база данных (БД) WORDLAKE [6]. Поскольку БД представляет собой уже обработанный массив натурных наблюдений, при работе на выявление выбросов априори считалось, что данные не содержат ошибок измерений.

Морфометрические характеристики, проанализированные в рамках исследования, представлены в табл. 1. БД WORLDLAKE включает в себя 786 озер, которые по происхождению озерной котловины можно отнести к вулканическим.

Результаты предварительного анализа данных представлены в табл. 2. Следует обратить внимание, что не для всех озер имеется полный набор морфометрических характеристик, поэтому количество объектов N для разных параметров разное.

Анализ коэффициента сжатия k_s и показателя удлиненности $1/k_s$ показывает, что форма озера в горизонтальной плоскости, основываясь на классификации [9], у большинства вулканических озер близка к округлой. Форму озерной котловины в горизонтальной плоскости можно характеризовать коэффициентом «емкости» c [10]. Среднее значение $c = 0{,}52$ свидетельствует о параболоидной форме озерной котловины. Это характерно для озер с данным типом происхождения, заполняющих водой вулканические кратеры, кальдеры, маары. Стоит заметить, что рассматривались также котловины, образованные за счет преграждения вод лавовыми потоками, а также смешанного — тектонико-вулканического происхождения.

Поскольку выборка состоит из 786 озер, и минимальное значение валидных объектов для определенной морфометрической характеристики — 114 значений, то ее можно отнести к выборкам большого объема [11] и применять соответствующие критерии идентификации выбросов.

Из предварительной статистической оценки морфометрических параметров вулканических озер видно, что распределение вероятности у таких характеристик, как L, W, S, ℓ, V, C, $1/k_s$, S^*, V^{**}, H^{**}, S/H_{avg}, имеет сильную левую (положительную) асимметрию, т. е. максимум смещен к началу координат. У характеристик k_s, c, λ, φ наблюдается правая асимметрия. Как правило, для нормального распределения показатель асимметрии находится в интервале от –1 до +1 (некоторые авторы применяют оценку от –2 до +2). Таким образом, для основных характеристик можно предположить, что распределение соответствует гамма-распределению, логнормальному, хи-квадрат или Вейбулла. Данный факт говорит о том, что наиболее часто встречаются небольшие вулканические озера с длиной и шириной озерного зеркала менее 5 км и 2,5 км соответственно, площадь таких озер не достигает 20 км^2, а средние глубины менее 40 м. Значение эксцесса (коэффициента вариации) показывает, в какой мере сплюснут пик распределения, т. е. насколько далеко от центра распределения отклоняются отдельные наблюдения. Для нормального распределения значение эксцесса изменяется от –3 до +3 (для мягкого критерия) или от –1 до +1 (для жесткого), когда вид распределения является решающим для дальнейшего анализа. При расчете эксцесса для нормального распределения использовалось центрирование на 0 (эксцесс Фишера). Если эксцесс больше нуля, то это указывает на «островершинное» распределение, если меньше, то пик явно

Таблица 1

Морфометрические параметры, содержащиеся в БД WORDLAKE

Morphometric parameters contained in the WORDLAKE database

Обозначение	Характеристика и единицы измерения
L	Длина озера (км)
W	Ширина озера (км)
S	Площадь зеркала озера (км^2)
ℓ	Длина береговой линии озера (км)
H_{avg}	Средняя глубина озера (м)
$H_{\max}$	Максимальная глубина озера (м)
V	Объем озера (км^3)
C	Площадь водосбора (км^2)
k_s	Коэффициент сжатия зеркала озера: $k_s = \frac{W}{L}$
$1/k_s$	Показатель удлиненности, обратно пропорциональный к коэффициенту сжатия k
c	Коэффициент формы озерной котловины (коэффициент емкости) $c = \frac{H_{avg}}{H_{\max}}$
H^*_{avg}	Безразмерная средняя глубина (по отношению к средней глубине озера Байкал H_{avg} = 170 м) $H^*_{avg} = \frac{H_{avg}}{H_{avgB}}$
$H^*_{\max}$	Безразмерная максимальная глубина (по отношению к максимальной глубине озера Байкал $H_{\max}$ = 1650 м) $H^*_{\max} = \frac{H_{\max}}{H_{\max B}}$
S^*	Безразмерная площадь озера (относительно площади Каспийского моря $S_{KS} = 3{,}86 \cdot 10^5$ км^2) $S^* = \frac{S}{S_{KS}}$
V^*	Безразмерный объем озера (относительно объема Каспийского моря $V_{KS} = 7{,}87 \cdot 10^5$ км^3) $V^* = \frac{V}{V_{KS}}$
H^{**}	Относительная глубина, равная отношению между максимальным вертикальным масштабом и площадью зеркала $H^{**} = 10^{-3} \frac{H_{\max}}{\sqrt{S}}$
$\frac{S}{H_{avg}}$	Показатель открытости, равный отношению площади озера к средней глубине
α	Коэффициент «глубинности» озера $\alpha = \frac{H_{avg}}{S^{1/3}}$
φ	Широта (в градусах)
λ	Долгота (в градусах)
Z_{asl}	Высота озера над уровнем моря (м)
N	Количество валидных объектов в выборке

не выражен. Как видно из табл. 2, все характеристики, кроме k_s, λ, φ, имеют ярко выраженный пик.

Таблица 2

Статистические оценки морфометрических характеристик вулканических озер

Statistical estimates of the morphometric characteristics of volcanic lakes

Характеристика	N	Мин.	Макс.	Среднее	Стандартное отклонение	Дисперсия	Асимметрия	Эксцесс
L	307	0,02	89	4,58	8,5	72,17	4,69	34,44
W	278	0,04	29,5	2,34	4,05	16,37	3,78	18,35
S	638	$1{,}2\cdot10^{-4}$	1130	17,59	78,89	6224	8,75	93,57
ℓ	232	0,200	363	18,42	42,2	1781	5,13	31,68
H_{avg}	269	$7{,}5\cdot10^{-3}$	328,6	33,78	49,54	2454	2,93	10,17
$H_{\max}$	510	0,200	589	55,01	77,5	6006	2,82	10,53
V	268	$2{,}0\cdot10^{-6}$	240	2,44	15,63	244,39	13,51	202,01
C	114	0,10	$5{,}24\cdot10^{4}$	1170	5420	$2{,}94\cdot10^{7}$	8	72,47
k_s	277	0,10	1	0,64	0,23	0,05	–0,26	–0,81
$1/k_s$	277	1	9,65	1,9	1,17	1,37	3,08	12,93
c	256	0,05	0,88	0,52	0,14	$2{,}03\cdot10^{-2}$	–0,56	0,37
H^*_{avg}	269	$1{,}01\cdot10^{-5}$	0,44	$4{,}57\cdot10^{-2}$	0,07	$4{,}48\cdot10^{-3}$	2,93	10,17
$H^*_{\max}$	510	$1{,}21\cdot10^{-4}$	0,36	$3{,}33\cdot10^{-2}$	0,05	$2{,}2\cdot10^{-3}$	2,82	10,53
S^*	640	$3{,}2\cdot10^{-10}$	$3\cdot10^{-3}$	$4{,}54\cdot10^{-5}$	$2{,}04\cdot10^{-4}$	$4{,}16\cdot10^{-8}$	8,76	93,87
V^*	268	$2{,}54\cdot10^{-11}$	$3\cdot10^{-3}$	$3{,}1\cdot10^{-5}$	$1{,}98\cdot10^{-4}$	$3{,}94\cdot10^{-8}$	13,51	202,01
H^{**}	472	$6{,}15\cdot10^{-5}$	4,22	$7{,}85\cdot10^{-2}$	0,23	0,05	13,49	232,53
S/H_{avg}	268	$1{,}47\cdot10^{-4}$	262	4,21	24,95	622,48	7,87	65,5
α	268	$5{,}34\cdot10^{-2}$	198,83	24,27	26,52	703,48	2,45	9,31
φ	616	–41,13	66,09	12,57	28,67	822,04	–0,19	–0,9
λ	613	–179	177,4	50,5	96,37	$9{,}28\cdot10^{3}$	–0,56	–0,8
Z_{asl}	607	0	4278	958,92	930,73	$8{,}66\cdot10^{5}$	1,1	0,62

Результаты исследования

Графические методы идентификации выбросов

Наиболее простым способом поиска выбросов в распределении являются графические методы. К ним относится построение точечных диаграмм, графиков вероятности, диаграмм «стебель—листья», диаграмм «ящика с усами» и т. п. В процессе анализа данных графические методы были применены для каждого морфометрического параметра. В данной статье приведены несколько примеров обнаружения выбросов различными графическими способами для морфометрических данных вулканических озер.

Рассмотрим точечные диаграммы (или графики рассеяния). На них каждый объект представляется точкой. На рис. 1 приведена точечная диаграмма, построенная по данным о средней и максимальной глубинах вулканических озер. На ней по оси X отложены максимальные глубины $H_{\max}$ объектов, по OY — средние

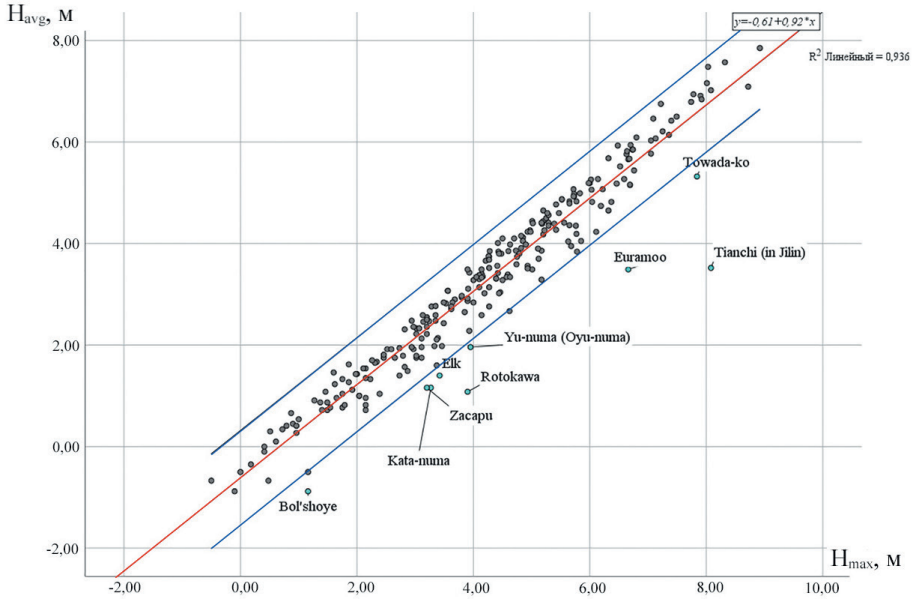

Рис. 1. Точечная диаграмма для средней и максимальной глубин вулканического озера.

Fig. 1. Scatter plot for the average and maximum depth of a volcanic lake.

глубины H_{avg}. На графике приведены аппроксимация линейной зависимостью (красная линия) и 95 % доверительный интервал (синие линии). Объекты, выходящие за границы доверительного интервала, можно рассматривать как кандидаты на выбросы (голубые точки). Стоит отметить, что точечные диаграммы в основном строят, чтобы показать корреляцию между величинами. В примере на рис. 1 между средней и максимальной глубиной есть почти функциональная взаимосвязь, и тангенс угла наклона аппроксимирующей прямой соответствует коэффициенту емкости.

Анализ графиков рассеяния для остальных морфометрических характеристик вулканических озер, характеризующихся сильной корреляцией между собой, показал, что меньше всего выбросов обнаружено по показателю открытости (S/H_{avg}). В результате анализа различными графическими методами были выявлены 12 озер, которые имеют выбросы по различным, аналитически не связанным, морфометрическим характеристикам (табл. 3).

Еще одним графическим методом для определения выбросов является построение графика вероятности. Он может быть представлен следующими видами: *P–P* (вероятностный график) или *Q–Q* (график квартилей). График *P–P* показывает кумулятивные пропорции переменной в сравнении с кумулятивными пропорциями теоретического распределения. Точки наблюдаемого распределения, лежащие наиболее удаленно от прямой теоретического распределения, следует рассматривать как возможные выбросы. На *Q–Q*-графике по оси X квартили стандартного теоретического распределения по OY — квартили данных. Так же, как

Таблица 3

Озера с аномалиями морфометрических характеристик, выявленные на основе графических методов

Lakes with anomalies of morphometric characteristics, identified on the basis of graphical methods

Код озера в БД	Название на русском	Альтернативное название	Страна
928	Корангамайт	Lake Corangamite	Австралия
5130	Лагуна-Колорада	Laguna Colorada	Боливия
10564	Тодос-лос-Сантос	Lago Esmeralda, Lago Todos los Santos	Чили
10988	Цзинпоху	Jingpo, Tzinpo Hu, Tzinpohu, Jingbo hu	Китай
14368	Цваи	Zway, Zwei, Ziway	Эфиопия
18422	Тоба	Toba, Danau Toba	Индонезия
26085	Таупо	Taupo	Новая Зеландия
27444	Тааль	Taal, Bombon	Филиппины
32205	Кроноцкое	Kronotskoye Ozero	Россия
45682	Моно	Mono	США
50391	Крейтер	Mazama, Crater Lake	США
52077	Йеллостоун	Yellowstone Lake	США

и для *P–P*-графика, принадлежность данных наблюдений к определенному виду распределения определяется близостью точек к прямой [12].

Интерес к этому методу заключается в том, что по графику вероятности можно также осуществить проверку принадлежности объектов наблюдений к тому или иному теоретическому виду распределения.

Для определения выбросов вероятностный и квартильный графики удобно представить в виде графиков с исключенными линейными трендами, что позволяет видеть несоответствие нормальному распределению более явным. Каждое значение на таком графике нормируется путем вычитания среднего значения и деления на соответствующее стандартное отклонение. Для каждого нормального вероятностного значения x_i значение с исключенным трендом z_i вычисляется следующим образом:

$$z_i = F^{-1}\frac{3i-1}{3n+1} - \frac{\left|x_i - \bar{x}\right|}{s}, \qquad (1)$$

где $i = 1, 2, \ldots, n$; F^{-1} — обратная функция нормального распределения; $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ — выборочное среднее значение; $s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$ — среднеквадратическое отклонение.

Экстремальные значения, отклоняющиеся более чем на +/– 0,04 стандартных отклонений, на таком графике являются кандидатами на выбросы. Если наблюдение превышает +/–1,96 стандартное отклонение, то оно является выбросом с 95%-ным доверительным интервалом.

Для примера рассмотрим *P–P* и *Q–Q* графики для некоторых морфометрических характеристик вулканических озер (рис. 2, 3). Проведенный анализ по данному графическому методу всех морфометрических характеристик показал, что большинство из них подчиняются законам логнормального, гамма распределения и Вейбулла. Такие морфометрические характеристики, как средняя глубина, форма озерной котловины и коэффициент сжатия близки по форме к нормальному распределению. На *P–P* и *Q–Q* графиках выбросы отображаются на наиболее

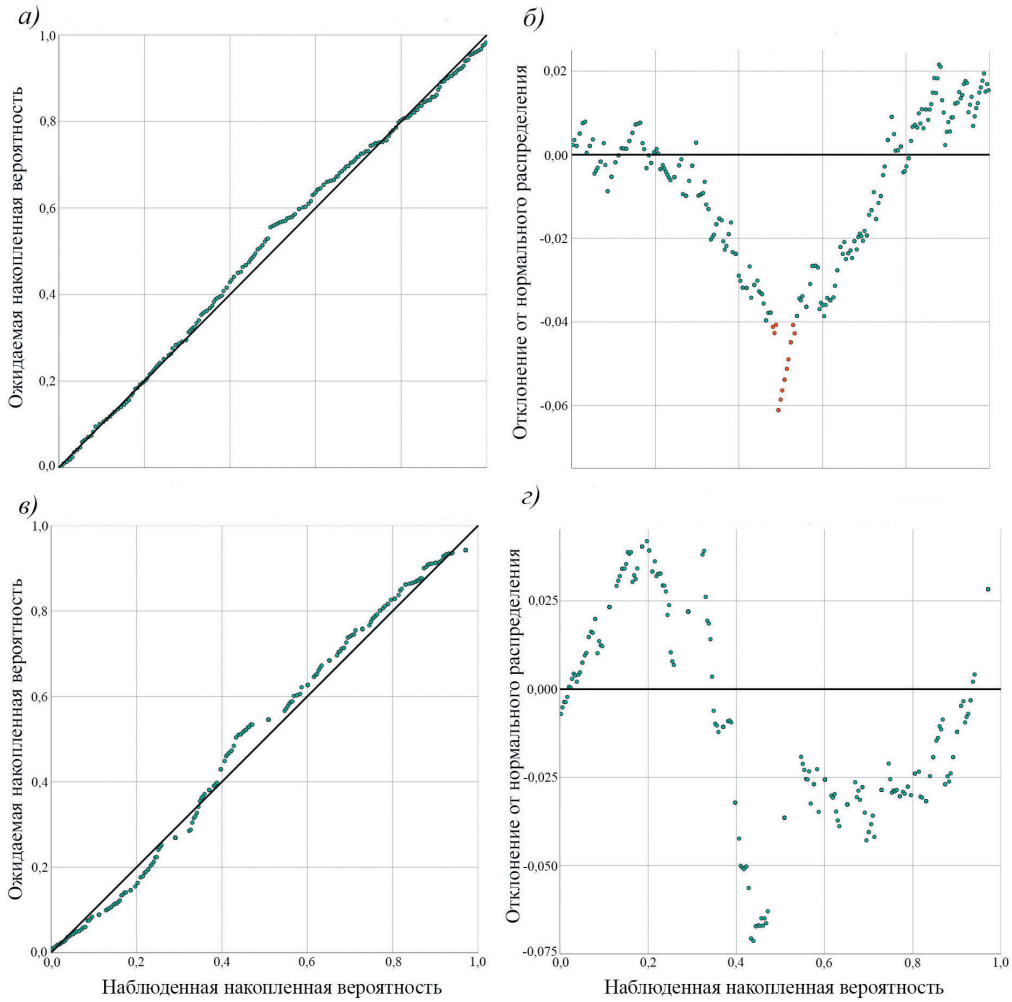

Рис. 2. Вероятностные графики (*а*, *в*) и вероятностные графики с удаленным трендом (*б*, *г*) для средней глубины (*а*, *б*) и коэффициента сжатия озера (*в*, *г*).

Fig. 2. Probabilistic plots (*а*, *в*) and probabilistic plots with a distant trend (*б*, *г*) for the average depth (*а*, *б*) and the compression ratio of the lake (*в*, *г*).

удаленные точки от прямой теоретического распределения. Они становятся наиболее заметны на графиках с удаленным трендом (оранжевые точки). Также было замечено, что по данному методу под потенциальные выбросы попадают те же озера, что и для метода точечной диаграммы (табл. 3).

Следующим графическим способом поиска выбросов в распределении является построение коробчатого графика (диаграмма «ящик с усами» или диаграмма размаха). Диаграмма «ящик с усами» была впервые предложена Джоном

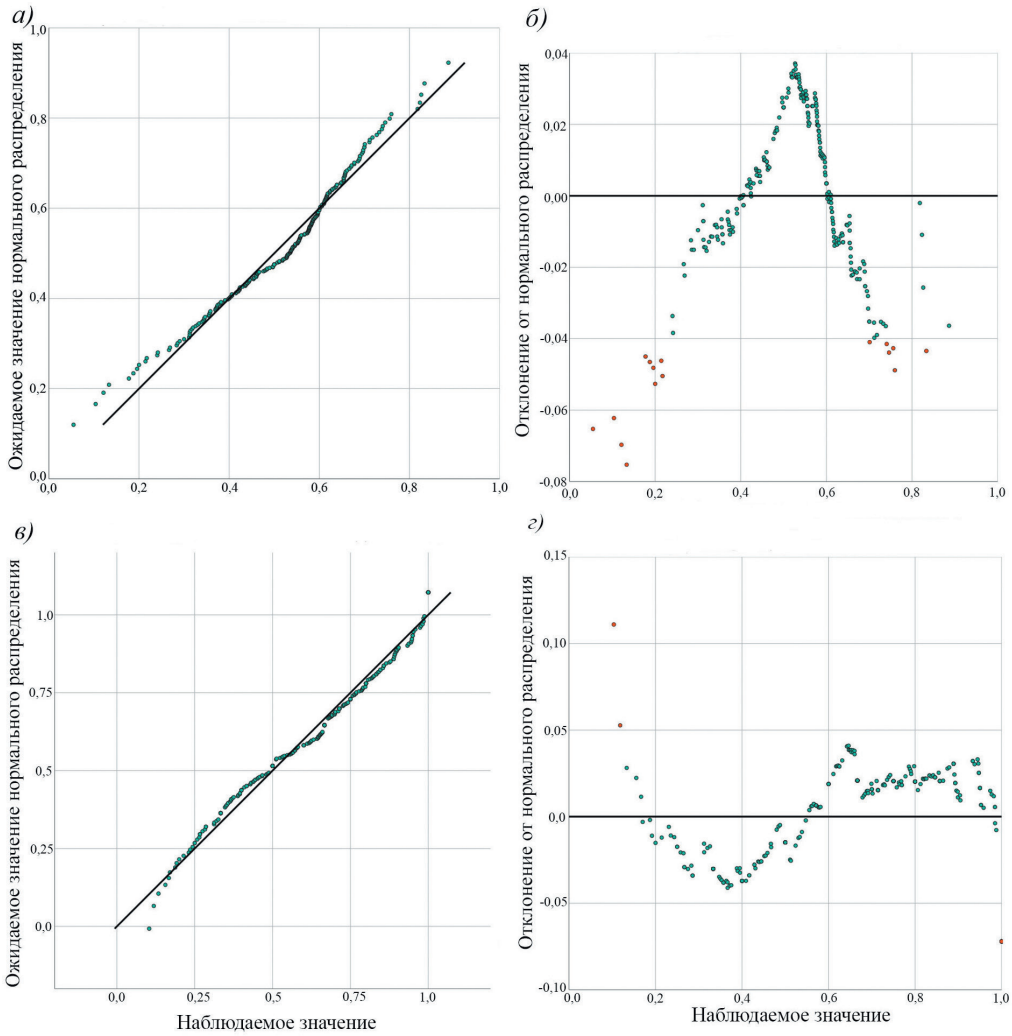

Рис. 3. Графики квартилей (*а*, *в*) и графики квартилей с удаленным трендом (*б*, *г*) для коэффициента формы озерной котловины (*а*, *б*) и коэффициента сжатия озера (*в*, *г*).

Fig. 3. Plots of quartiles (*а*, *в*) and plots of quartiles with a distant trend (*б*, *г*) for the shape factor of the lake hollow (*а*, *б*) and the compression factor of the lake (*в*, *г*).

Тьюки [13]. Для одной переменной такой график выглядит как диаграмма с переменной по OX и ее интервальной зависимости по OY. На рис. 4 приведены результаты квантильного анализа для некоторых морфометрических характеристик. Высота прямоугольника обозначает первый ($x_{0,25}$) и третий ($x_{0,75}$) квартили. К этим квартилям пристраиваются «усы», длина которых соответствует 25 % минимальным и 25 % максимальным значениям переменной. Горизонтальная линия внутри прямоугольника соответствует медиане ($x_{0,50}$), которая характеризует центр распределения. Высота ящика представляет интерквартильное расстояние $Q = x_{0,75} - x_{0,25}$. В случае нормального распределения данных выполняется следующее соотношение:

$$Q = 1{,}34s. \tag{2}$$

Для нормально распределенных данных медиана находится посередине и делит прямоугольник пополам. Смещение медианы относительно центра в верхнюю или нижнюю части ящика свидетельствует о наличии асимметрии.

В квантильном анализе точка, резко выделяющаяся от остальной совокупности, называется выбросом, если для нее выполняются следующие условия:

— для положительных аномалий: $x_{выб} > x_{0,75} + kQ$,

— для отрицательных аномалий: $x_{выб} < x_{0,25} - kQ$,

где k — подгоночный коэффициент, называемый коэффициентом выбросов. Обычно он принимается равным $k = 1{,}5$. Для случайной выборки, имеющей нормальное распределение, в диапазоне $\pm kQ$ содержится 99 % значений выборки. На рис. 4

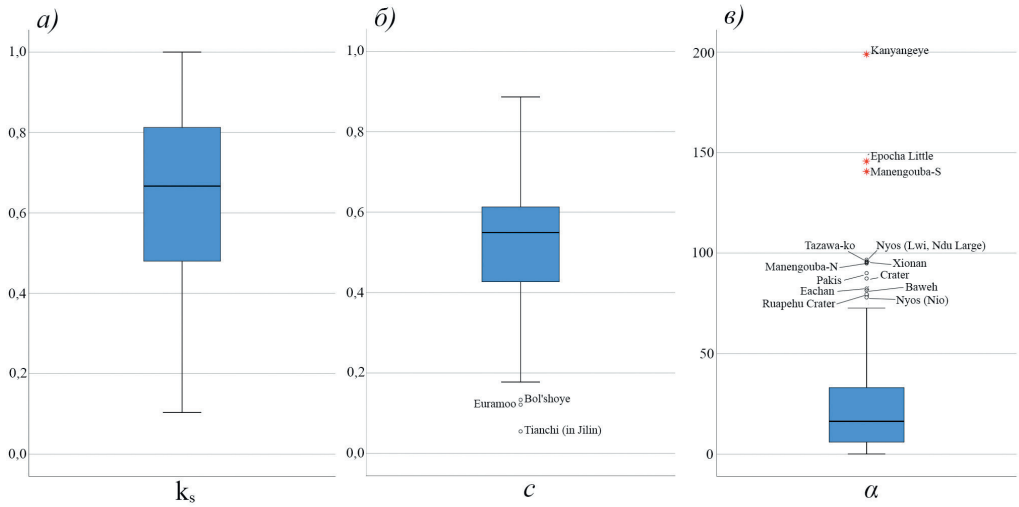

Рис. 4. Квантильный анализ для коэффициента сжатия (*а*),
коэффициента формы озерной котловины (*б*), коэффициента глубинности озера (*в*).

Fig. 4. Box diagrams for the compression ratio (*a*),
the shape factor of the lake hollow (*b*), and the lake depth factor (*c*).

выбросы обозначены оранжевыми звездочками. Как видно из рис. 4, значения коэффициента сжатия не имеют выбросов (рис. 4 *а*), а само распределение близко к нормальному. Для коэффициента формы озерной котловины (рис. 4 *б*) имеется несколько нехарактерных (несвойственных) значений, и само распределение имеет несимметричную форму. Для коэффициента глубины озера (рис. 4 *в*) имеются как нехарактерные значения, так и выбросы.

Анализ диаграмм размаха показал, что по некоторым морфометрическим характеристикам нет выбросов (k_s, c, φ, λ, Z_{als}), тогда как по остальным характеристикам найдено больше озер, чем в случае точечных диаграмм. Выбросы, найденные сразу по нескольким морфометрическим характеристикам для одного объекта БД, которые можно интерпретировать как «множественные», были обнаружены для 44 озер. Необходимо заметить, что все аномальные озера, ранее найденные по методу точечной диаграммы, вошли в список озер, найденных с помощью квантильного анализа. При этом предварительная экспертная оценка выделенных озер свидетельствует о необходимости признания их уникальными и по другим характеристикам, отражающим особенности их водной среды.

Статистические критерии идентификации выбросов

Наряду с графическими методами обнаружения выбросов было проведено выявление аномальных значений на основе классических статистических критериев. В качестве наиболее точных методов были выбраны критерии Смирнова—Граббса, Хоглина—Иглевича, Титьена—Мура, Шовене, Ирвина и метод Томпсона (критерий Рошера). Рассмотрим более подробно статистику каждого из критериев и возможности применения к выборке, рассматриваемой в данной работе. Все вышеприведенные критерии применимы только к гауссовому распределению, поэтому каждая из морфометрических характеристик приведена к нормальному виду с помощью преобразования Бокса—Кокса [14]. Также выбор данных критериев зависел от количества данных наблюдений.

Критерий Шовене предназначен для идентификации элемента выборки x_i, как аномального значения [15]. Идея статистики заключается в следующем: для любой случайной величины x_i, подчиняющейся нормальному закону распределения с центром рассеивания $x = a$, существует величина, называемая срединным (вероятностным) отклонением E, такая, что вероятность попадания случайной величины равна:

$$P\left(a - E < x_i < a + E\right) = 0,5. \tag{3}$$

Любые точки, вероятность которых попасть в указанный интервал меньше 0,5, считаются выбросами. В своей работе Шовене [16] выводит приблизительный критерий для исключения одного сомнительно наблюдения, основываясь на методе наименьших квадратов. Наблюдение x_i из выборки объема n будет выбросом, если соответствует следующему правилу:

$$n\left(1 - \Phi(t)\right) < 0,5, \tag{4}$$

или

$$\Phi(t) > \frac{2n-1}{2n}, \tag{5}$$

где $\Phi(x) = \frac{2}{\sqrt{2\pi}\sigma}\int_0^{k\sigma} e^{-x^2/2\sigma^2}dx$ — функция Лапласа. То есть можно сказать, что наблюдение является выбросом, если превышает $k\sigma$.

В современной записи критерий Шовене основан на статистике:

$$K = \frac{\left|x_i - \bar{x}\right|}{s}. \tag{6}$$

Критические значения K^* критерия Шовене приводятся в табличном виде. Если $K > K^*$, то значение x_i признается выбросом [17]. Также рядом автором [18, 19] описаны другие алгоритмы идентификации выбросов по критерию Шовене, дающие аналогичные результаты.

Критерий используется для поиска грубых промахов в выборках, распределенных по нормальному и экспоненциальному законам распределения. Его применение к исследуемой выборке выявило выбросы по W, ℓ, H_{avg}, c, H^*_{avg}, S/H_{avg}, α. При этом для всех данных характеристик были выявлены одиночные выбросы, совпадающие с другими критериями, за исключением выбросов для ℓ. В случае с длиной береговой линии при выявлении выбросов найдены критерием Шовене два озера — Китафутаго и Мидори в Японии *(Kitafutago*, *Midoriga-ike)*, которые по всем остальным методам выбросов не показали. В результате экспертной оценки в рамках данной работы было признано, что применение данного критерия к исследуемой выборке показывает неудовлетворительные результаты.

Критерий Ирвина предложен [20] для поиска выбросов в распределении с заранее известной дисперсией и хорошо работает на малых выборках. Статистика основана на критерии:

$$\tau = \frac{x_n - x_{n-1}}{s}, \tag{7}$$

и если подозрительным является первое значение:

$$\tau = \frac{x_2 - x_1}{s}, \tag{8}$$

то подозрительное значение сравнивают с табличным критическим значением τ^* [20]. В случае если критерий (7), (8) больше табличного значения $\tau > \tau^*$, то такое наблюдение является выбросом. В оригинале статьи Ирвина [20] критические значения находились численно, сейчас для них найдены аппроксимирующее уравнение вида [21]:

$$\tau^* = A\left(k - \frac{5}{n}\right)^B, \tag{9}$$

где k — номер значения в вариационном ряду; A, B — коэффициенты, зависящие от n и при уровне значимости $\alpha = 0{,}05$ имеющие вид:

$$A = -4,041n^{0,5} + 32,5148n^{0,4} - 103,5032n^{0,3} + 162,9495n^{0,2} - 127,32n^{0,1} + 40,7683, \quad (10)$$

$$B = -0,30595n^{0,5} + 6,7127n^{0,4} - 38,1211n^{0,3} + 93,2983n^{0,2} - \\ -106,1212n^{0,1} + 45,5395. \quad (11)$$

Максимальная и средняя ошибки аппроксимации равны 0,004 и 0,0011 соответственно, относительное отклонение уровней значимости от действительных значений при максимальной ошибки аппроксимации составляет 2,7 %.

Основным недостатком метода Ирвина является то, что он воспринимает как выбросы только экстремальные значения, что при поиске аномалий не всегда так.

Применение критерия Ирвина к исследуемой выборке показало лучшие результаты, чем применение критерия Шовене, и схожие с применением критерия Титьена—Мура и метода Томпсона, о которых будет указано далее. Исключением явились характеристики k_s, $1/k_s$, где выбросов найдено не было.

Критерий Смирнова—Граббса является модификацией критерия Граббса [22] для одного экстремального значения из нормального распределения. Данный критерий [23] основан на статистиках:

$$\tau = \frac{\max\limits_i(x) - \bar{x}}{s}, \quad (12)$$

или

$$\tau = \frac{\bar{x} - \min\limits_i(x)}{s}. \quad (13)$$

Критическое значение для статистики Смирнова—Граббса определяется формулой:

$$G_n = \frac{n-1}{\sqrt{n}} \sqrt{\frac{t^2_{\frac{\alpha}{n},n-2}}{t^2_{\frac{\alpha}{n},n-2} + n - 2}}, \quad (14)$$

где $t^2_{\frac{\alpha}{n},n-2}$ — квартиль распределения Стьюдента с n–2 степенями свободы, α — заданный уровень значимости (при расчетах было задано α = 0,05). Если величина статистики больше критического значения, то наблюдение признается выбросом.

При обработке исследуемой выборки на идентификацию выбросов с применением данного критерия не было обнаружено максимальных или минимальных значений, которые, согласно данной статистике, подпадают под выброс. Однако, его использование представлялось интересным, в силу некоторых отличий от критерия Шовене и в рамках работы выявило ряд дополнительных выбросов. При решении других задач, как показано в статье [11], результаты использования этих двух критериев в случае малых выборок также различаются.

Критерий Хоглина—Иглевича. Правило обнаружения выбросов согласно критерию Хоглина—Иглевича [24] формулируется следующим образом:

наблюдение признается выбросом, если его значение выходит за пределы интервала, границы которого определяются порядковыми статистиками:

$$(1+k)x_{[l]} - kx_{[n-l+1]} \tag{15}$$

и

$$(1+k)x_{[n-l+1]} - kx_{[l]}, \tag{16}$$

где $x_{[i]}$ — i-ый по величине член выборки, ранжированной по возрастанию; l представляет собой варианты:

$$l_1 = \frac{1}{2}\left[\frac{n+3}{2}\right],$$

$$l_2 = \frac{1}{2}\left[\frac{n}{4} + \frac{5}{12}\right], \tag{17}$$

$$l_3 = \frac{1}{2}\left[\frac{n}{4} + \frac{1}{4}\right].$$

где квадратными скобками обозначена целая часть числа. Коэффициент k выбирается в зависимости от значения l [25].

Применение критерия Хоглина—Иглевича к исследуемой выборке не показало эффективности обнаружения выбросов. Был обнаружен лишь один выброс по морфометрической характеристике H^{**} для канадского озера Каухако.

Критерий Титьена—Мура основан на выявлении выбросов к минимальным и максимальным значениям. Статистика критерия Титьена—Мура [26] идентификации k экстремальных значений основана на величине:

$$L_k = \frac{\sum_{i=1}^{n-k}\left(x_i - \overline{x_k}\right)^2}{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}, \tag{18}$$

в случае наибольших значений, где $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ — выборочное среднее для всей выборки, $\overline{x_k} = \frac{1}{n-k}\sum_{i=1}^{n-k} x_i$ — среднее значение после отбрасывания k наибольших значений.

И в случае наименьших значений:

$$L_k^* = \frac{\sum_{i=k+1}^{n-k}\left(x_i - \overline{x_k^*}\right)^2}{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}, \tag{19}$$

где $\overline{x_k^*} = \frac{1}{n-k}\sum_{i=k+1}^{n-k} x_i$ — среднее после отбрасывания k наименьших значений.

При наличии как наименьших, так и наибольших подозрительных значений используется следующее правило: находятся абсолютные отклонения $d_i = \left| x_i - \overline{x} \right|$, ранжируются по возрастанию. Пусть z_i — выборочное значение x_i, для которого d_i — i-е по величине. Тогда статистика критерия имеет вид [27]:

$$E_k = \frac{\sum_{i=1}^{n-k} \left(z_i - \overline{z_k} \right)^2}{\sum_{i=1}^{n} \left(z_i - \overline{z} \right)^2}, \tag{20}$$

где $\overline{z_k} = \frac{1}{n-k} \sum_{i=1}^{n-k} z_i$.

Если величина критерия больше некого критического табличного значения, то наблюдение признается выбросом.

При применении критерия Титьена—Мура к исследуемой выборке необходимо было изначально знать, сколько выбросов содержит та или иная характеристика. Предположение о количестве выбросов было взято на основе ее обработки графическими методами. Результаты применения теста показали, что для L, H_{avg}, $H_{\max}$, V, k_s, $1/k_s$, c, H^*_{avg}, $H^*_{\max}$, S^*, V^*, H^{**}, S/H_{avg}, α выбросы совпали с выбросами, полученными по квантильному анализу (для данных без трансформации). Для всех остальных характеристик данный метод не обнаружил выбросов. Таким образом, данный критерий является очень жестким для обнаружения выбросов.

Критерий Рошера (метод Томпсона) [28] для идентификации выбросов основан на статистике:

$$T_i = \frac{\left| x_i - \overline{x} \right|}{s}, \tag{21}$$

где s — выборочное стандартное отклонение.

Критическое значение для метода Томпсона находится согласно формуле:

$$T_n = \sqrt{\frac{(n-1)\, t^2_{\frac{1-\alpha}{n}, n-2}}{t^2_{\frac{1-\alpha}{n}, n-2} + n - 2}}. \tag{22}$$

Если величина критерия больше критического значения, то наблюдение признается выбросом.

Применение метода Томпсона для исследуемой выборки показало, что для $1/k_s$ значения выбросов полностью совпадают со значениями, полученными по критерию Титьена—Мура. Для L, V, S^* метод Томпсона идентифицировал намного больше выбросов, чем в случае применения критерия Титьена—Мура. Таким образом, метод Томпсона показал дополнительные выбросы, которые не были выявлены в результате применения остальных критериев. Несомненным плюсом данного метода является отсутствие предположения об изначальном количестве выбросов.

В табл. 4 приведен итоговый список озер, имеющих выбросы, найденные с помощью указанных выше критериев. Указаны водные объекты, у которых имеются некоторые аномалии по явно не связанным между собой морфометрическим характеристикам.

Таблица 4

Список озер с множественными выбросами,
полученными с использованием нескольких критериев

List of lakes with multiple emissions obtained with using multiple criteria

Код озера	Название	Альтернативное название	Происхождение	Страна
253	Уэчулафкен	Lago Huechulaufquen	Ледниковое / лавово-подпрудное	Аргентина
921	Колак	Lake Colac	Вулканическое (коллапс)	Австралия
928	Корангамайт	Lake Corangamite	Вулканическое (коллапс)	Австралия
5130	Лагуна-Колорада	Laguna Colorada	Озеро вулканического плато	Боливия
5703	Эпоча	Epocha Little	Вулканическое	Камерун
10564	Тодос Лос Сантос	Lago Esmeralda, Lago Todos los Santos	Ледниково-вулканическое	Чили
10988	Цзинпоху	Jingpo, Tzinpo Hu, Tzinpohu, Jingbo hu	Лавово-подпрудное	Китай
11015	Удальянчи		Вулканическое	Китай
11459	Небесное озеро	Tien-chih, Tianchi	В кратере влк. Пэктусан	Китай
13620	Сяонань	Xionan, Lesson South Lake	Вулканическое	Китай
13688	Аренль	Embalse Arenal	Вулканическое, преобразовано в водохранилище	Коста-Рика
14000	Илопанго	Lago de Ilopango	Вулканическое, (кратерное)	Эль Сальвадор
14368	Цваи	Zway, Zwei, Ziway Hayk’	Вулкано-тектоническое	Эфиопия
14390	Ауаса	Awasa Hayk; Awusa	Вулкано-тектоническое	Эфиопия
14412	Чоу-Бахр	Chew Bahir, Ch’ew Bahir, Lake Stefanie,	Вулкано-тектоническое	Эфиопия/ Кения
15637	Паравани	Paravani, Taparavani	Вулкано-тектоническое	Грузия
17627	Атитлан	Lago de Atitlan, Atitla	Вулканическое (кальдера)	Гватемала
17717	Миватн	Myvatn	Вулканическое (лавово-подпрудное)	Исландия
18281	Батур	Danau Barur	Вулканическое (кратерное)	Индонезия
18388	Клабат	Klabat	Вулканическое (кратерное)	Индонезия
18395	Бавех	Baweh, Danau di Baweh, Dibawah	Вулканическое	Индонезия
18411	Сорикмарапи	Sorikmarapi-2	Вулканическое (кратерное)	Индонезия
18412	Тандикат	Tandikat	Вулканическое (кратерное)	Индонезия
18415	Маниньжау	Maninjau, Maningam, Danau Manindjau	Вулканическое (кратерное)	Индонезия
18416	Сингкарак	Danau Singkarak	Вулкано-тектоническое	Индонезия

Продолжение табл. 4

Код озера	Название	Альтернативное название	Происхождение	Страна
18421	Ранау	Danau Ranau	Вулканическое (кратерное)	Индонезия
18422	Тоба	Danau Toba	Вулканическое (кратерное)	Индонезия
18424	Сурубель	Surubel	Вулканическое (кратерное)	Индонезия
18775	Больсена	Vulsino, Lago di Bolsena	Вулканическое (кратерное)	Италия
18776	Брачано	Lago di Bracciano	Вулканическое	Италия
19056	Товада	Towada-Ko	Вулканическое (двойная кальдера)	Япония
19082	Тадзава	Tazawa-ko	Вулканическое (кратерное)	Япония
19184	Аканума	Aka-numa	Вулканическое (запрудное в кальдерной депрессии)	Япония
19192	Инавасиро	Inawashiro-Ko	Вулканическое (запрудное)	Япония
19297	Тоя	Toya-ko	Вулканическое (кратерное)	Япония
19307	Шикотсю	Shikotsu-ko	Вулканическое (кратерное)	Япония
19379	Куттара	Kussharo-ko	Вулканическое (кратерное)	Япония
19380	Машу	Masyuko, Mashu-Ko	Вулканическое (кратерное)	Япония
19415	Окама	Okama	Вулканическое (кратерное)	Япония
19471	Тагаваура	Tagawaura	Вулканическое (лавово-подпрудное)	Япония
19613	Наканума	Nakanuma (Iwaonupuri)	Вулканическое (лавово-подпрудное)	Япония
19679	Каригоме	Karigome-ike	Вулканическое (маар)	Япония
19681	Мидори	Midoriga-ike	Вулканическое (маар)	Япония
23116	Магади	Magadi, Ngporonporo	Вулканическое (кратерное)	Кения
24469	Пацкуаро	Lago de Patzcuaro	Вулканическое (лавово-подпрудное)	Мексика
26062	Ротоковау	Rotokawau	Вулканическое (маар)	Новая Зеландия
26073	Роторуа	Lake Rotorua	Вулканическое (кратерное)	Новая Зеландия
26085	Таупо	Lake Taupo	Вулканическое (кратерное) / тектоническое	Новая Зеландия
27313	Кутубу	Lake Kutubu	Вулканическое (лавово-подпрудное)	Новая Гвинея\ Папуа
27344	Уисдом	Lake Wisdom	Вулканическое (кратерное)	Новая Гвинея\ Папуа
27444	Тааль	Taal, Bombon	Вулканическое (кратерное)	Филиппины
27456	Ланао	Lanao, Sultan Alonto	Тектонико-вулканическое (лавово-подпрудное)	Филиппины
32194	Банное		Вулканическое (кратерное)	Россия
32205	Кроноцкое		Вулканическое (лавово-подпрудное)	Россия
32207	Курильскоое		Тектонико-вулканическое (кратерное)	Россия

Окончание табл. 4

Код озера	Название	Альтернативное название	Происхождение	Страна
44142	Каньянгее	Kanyangeye	Вулканическое (кратерное)	Уганда
44190	Тарн 1	Tarn 1	Вулканическое (кратерное)	Уганда
44191	Тарн 2	Tarn 2	Вулканическое (кратерное)	Уганда
45665	Иньо	Inyo Crater Lake (Upper)	Вулканическое (кратерное)	США
45682	Моно	Mono	Вулкническое	США
46919	Каухако	Kauhako Lake	Вулканическое (кратерное)	США
46924	Уайеле	Wai'Ele'ele	Вулканическое (кратерное)	США
48651	Уашо	Big Washoe Lake; Washoe Lake	Вулканическое (лавово-подпрудное)	США
50391	Крейтер	Mazama, Crater Lake	Вулканическое (кратерное)	США
52077	Йеллостоун	Yellowstone Lake	Вулканическое (кратерное)	США

Необходимо отметить, что поиск аномальных значений в рамках данного исследования проводился классическими статистическими методами. Тем не менее, следует учитывать существование и других методов поиска аномалий, таких как метрические методы, метод 3σ, кластеризация, робастный подход, многокритериальный анализ и др. [2, 29].

Экспертная оценка полученного списка (табл. 4), состоящего из 65 водоемов, свидетельствует, что большинство из них могут рассматриваться как уникальные с учетом и других признаков, характеризующих озерные экосистемы. Обращает на себя внимание тот факт, что аномалии были выявлены приблизительно в равной пропорции среди всех рассмотренных подвидов озерных котловин (кратерные, лавово-подпрудные, смешанного происхождения). И если в случае достаточно древних котловин, расположенных в вулканических регионах (Лагуна Колорада на вулканическом плато Альтиплано, оз. Корангамайт, окруженное обнажениями вулкана, сформировавшего его чашу, оз. Пацкуаро и др.), выбросы по ряду параметров могли обуславливаться значительным преобразованием котловин с течением времени, то среди кратерных озер были выделены водоемы, занимающие наиболее интересные кальдеры, в том числе с геологической точки зрения. Так, в список попали озера, расположенные в кальдерах вулканов, извержения которых считаются наиболее разрушительными в истории Земли — Тобо (по оценкам во время извержения, произошедшего приблизительно 75 000 лет назад, было выброшено приблизительно 1500—2000 км^3 материала) и Таупо (извержение вулкана Таупо произошло 26 500 лет назад и сопровождалось выбросом 430 км^3 пирокластического материала, 320 км^3 игнимбрита и 420 км^3 первичного внутрикальдерного материала) [30]. Озера Тобо и Таупо — два самых крупных по площади, и имеющие котловины вулканического происхождения.

В списке выделенных озер есть и те, которые находятся в кратерах действующих вулканов, как оз. Батур. Последнее сильное извержение вулкана произошло в 1963 г., однако небольшие выбросы наблюдаются почти каждый год. Среди озер, расположенных в районах современной вулканической активности — озера

Курильское, Банное, Тоя, Йеллоустон и др. Как в самих озерах, так и в их окрестностях имеются многочисленные гейзеры, фумаролы и горячие источники. Оз. Банное занимает взрывную воронку, заполненную нагретой сернистой водой, его особенностью является наличие второго дна — верхнее, ложное дно состоит из застывшей серной корки, под которой находится резервуар расплавленной серы, мощностью в несколько метров.

Крайне интересны и попавшие в итоговый список некратерные озера. Так оз. Брачано обязано своим происхождением интенсивной вулканической и тектонической активности, происходившей на протяжении 600 000 лет и приведшей к образованию множества небольших вулканов. Главный магматический очаг находился под нынешним озером Брачано, обязанным своей депрессией его обрушению. Оз. Миватн сформировано около 2300 лет назад выплеснувшимся из большой трещины лавовым потоком. Лава текла по долине Лаксардалур в низменную равнину Адалдалур, где она впадала в Северный Ледовитый океан примерно в 50 км от Миватна. В то время здесь было большое озеро, предшественник современного. Когда светящаяся лава столкнулась с озером, часть его заболоченных отложений оказалась под ним. Последовавшие за этим паровые взрывы разорвали лаву на мелкие кусочки, которые были подброшены в воздух вместе с частью озера. В результате неоднократных взрывов в нескольких местах образовались группы кратеров, которые теперь доминируют над ландшафтом на берегу озера, а также образуют некоторые острова [31].

Среди озер, котловины которых имеют смешанное происхождение, особо примечательно оз. Тодос Лос Сантос. Это одно из немногих в мире озер ледниково-вулканического происхождения, в то время как большинство озер смешанного происхождения обязаны появлением своих котловин наряду с вулканическими также тектоническим процессам. В течение ледникового периода два рукава ледника, стекавших с гор в противоположных направлениях, сошлись в центральной долине и, расширяясь на запад, выпахали ее до депрессии, которую теперь занимают озера Тодос-Лос-Сантос и Льянкиуэ. После отступления ледника взрывы вулканов Осорно и Кальбуко разделили два озерных бассейна вулканическим материалом [32].

Анализ полученного списка озер свидетельствует, что в него попали не только крупные, наиболее глубокие водоемы, выделение которых обычно не вызывает особой сложности, но и целый ряд небольших озер, примечательных своими необычными морфометрическими характеристиками. Так, из 65 озер, выделенных благодаря поиску выбросов, 18 имеют площадь зеркала, не превышающую 1 км^2. Без использования, предложенного в рамках данной статьи аппарата, выделение таких озер в большой выборке представляет огромную сложность.

Как уже указывалось выше, морфометрические характеристики водоема во многом определяют происходящие в нем термические и динамические процессы. Так озера Тоя и Шикотсю практически никогда не замерзают и являются самыми северными незамерзающими озерами Японии. Целый ряд озер характеризуются значительными колебаниями уровня воды, как оз. Арениаль, сезонные колебания уровня которого составляют около 30 м [32].

Среди выделенных озер интересными по своим гидрохимическим характеристикам можно назвать озера Батур, Сингкарак, Ранау, Магади, Уисдом и др. Оз. Батур отличается необычным для большинства озер Индонезии химическим составом воды и повышенной минерализацией (электропроводность 1750—1800 мкСм/см) [33]. Для озер Сингкарак и Ранау характерно наличие значительного по объему аноксичного гиполимниона, на глубинах обнаруживается гидроген сульфид [34]. Для озер Уисдом (глубина 360 м) и Курильского (316 м), напротив, характерна относительно высокая концентрация кислорода в глубинных частях.

С биологических позиций обращают на себя внимание озера Аканума, Моно, Тааль и др. Оз. Аканума известно, как аборигенное обиталище «Marimo» — шаровидных колоний водоросли *Aegagropila sauteri*, 2—15 см в диаметре. Предполагается, что шаровидная форма сформирована за счет скопления волокнистой водоросли и вращательного движения, вызываемого турбулентностью воды [35]. Сходный вид сферических масс водорослей наблюдается еще в нескольких озерах Японии, однако они менее регулярны по форме. Вся природная цепочка питания животного мира оз. Моно основана на одноклеточных водорослях, используемых в пищу рачками *Artemia monica*. В 2009 г. в озере найдены уникальные бактерии GFAJ-1, примечательные своей способностью выживать при очень высоких концентрациях мышьяка [36]. В оз. Тааль найдено четыре эндемичных вида рыб.

Все вышесказанное свидетельствует, что полученный список требует дальнейшего более детального рассмотрения на предмет его уникальности. А также проведения статистического анализа других характеристик, описывающих экосистему озера.

Следует также отметить, что около трети озер, содержащихся в списке, имеют охранный статус и признаны уникальным природными объектами.

Выводы

В рамках данного исследования проведен анализ возможностей использования различных статистических методов для поиска выбросов, которые могут характеризовать водный объект как претендента на звание «уникального». Изначально, с использованием графических методов, был получен список озер, большинство из которых, согласно экспертной оценке, могут рассматриваться как уникальные. Далее выявление аномальных значений было проведено на основе целого ряда классических статистических критериев.

Среди методов, протестированных на примере исследуемых данных по морфометрическим характеристикам котловин озер вулканического происхождения, наибольшее соответствие с экспертной оценкой проявили диаграммы размаха (из визуальных критериев) и метод Томпсона. Критерий Смирнова—Граббса даже для идентификации экстремальных значений исследуемой выборки не показал хорошего результата. В то же время следует отметить — аналитические методы показали большее количество аномальных значений по сравнению с визуальными, что требует дальнейшего осмысления.

Экспертный анализ полученного списка озер подтвердил, что большинство из них могут быть признаны уникальными, поскольку обладают экстремальными характеристиками не только по своим морфометрическим параметрам.

Представляется необходимым с учетом полученного тестирования статистических методов, провести в дальнейшем анализ выборок гидрохимических характеристик, используя при этом наиболее подтвердившие себя для поставленных целей критерии. Другой задачей является переход к системной оценке уникальности на основе применения многокритериального анализа.

Благодарности

Работа выполнена в рамках проекта РФФИ 20-05-00303\20.

Acknowledgments

The work was carried out within the project of the RFBR 20-05-00303\20.

Список литературы

1. *Поздняков Ш.Р., Измайлова А.В., Расулова А.М.* Уникальные озера как объект научного интереса // Изв. РГО. 2020. Т. 152. № 3. С. 17—31.
2. *Мякишева Н.В.* Многокритериальная классификация озер. СПб.: РГГМУ, 2009. 160 с.
3. *Рянжин С.В., Ульянова Т.Ю.* Геоинформационная система для озер мира — ГИС «Озера Мира» (GIS WORLDLAKE) // Доклады РАН. 2000. Т. 370. № 4. С. 542—545.
4. *Румянцев В.А., Измайлова А.В., Ульянова Т.Ю.* Информационная система «Озера Земли» // Изв. РГО. 2009. Т. 141. Вып. 3. С. 8—14.
5. *Измайлова А.В., Ульянова Т.Ю.* Информационная система «Озера России» // Водное хозяйство России. 2014. № 6. С. 21—28.
6. *Кочков Н.В., Рянжин С.В.* Озера мира WORLDLAKE. Свидетельство о государственной регистрации базы данных № 2015621549.
7. *Драбкова В.Г., Сорокин И.Н.* Озеро и его водосбор — единая природная система. Л.: Наука. Ленингр. отд-ние, 1979. 196 с.
8. *Якушко А.Ф., Новик А.А.* Проблемы экологической устойчивости ледниковых ложбинных озер Белорусского поозерья // Вестник БГУ. 2005. № 1. С. 55—59.
9. *Григорьев С.В.* О некоторых определениях и показателях в озероведении // Труды Карельск. фил. АН СССР. 1959. Вып. 18. С. 29—45.
10. *Верещагин Г.Ю.* Методы морфологической характеристики озер // Тр. Олонецкой науч. экспедиции. Ч. 2. Вып. 1. Л.: ГГИ, 1930. 116 с.
11. *Попукайло В.С.* Обнаружение аномальных измерений при обработке данных малого объема // Технология и конструирование в электронной аппаратуре. 2016. № 4—5. С. 42—46.
12. *Gibbons J.D., Chakraborti S.* Nonparametric Statistical Inference, Fifth Edition // CRC Press. Boca Raton, London, New-York. 2014. 614 p.
13. *Tukey J.W.* Exploratory Data Analysis // Reading, Mass: Addison-Wesley, 1977.
14. *Box G.E.P., Cox, D.R.* An analysis of transformations // Journal of the Royal Statistical Society. Series B. 1964. No. 26. P. 211—252.
15. *Шенк Х.* Теория инженерного эксперимента / Пер. с англ. Е.Г. Коваленко; под ред. чл.-кор. АН СССР Н.П. Бусленко. М.: Мир, 1972. 381 с.
16. *Chauvenet W.* A manual or Spherical and practical astronomy. Vol. 2. Theory and use of astronomical instruments. Method of least squares. Philadelphia: J.B. Lippincott Company, London: Trubner. 1876. 632 p.
17. *Кобзарь А.И.* Прикладная математическая статистика. Для инженеров и научных работников. М.: ФИЗМАТЛИТ, 2006. 816 с.

18. *Сергеев А.Г., Тергеря В.В.* Метрология, стандартизация и сертификация: учебник. М.: Юрайт, 2011. 820 с.
19. *Радкевич Я.М.* Метрология, стандартизация и сертификация: учебник для вузов. М.: Высшая школа, 2004. 767 с.
20. *Irwin J.O.* On a Criterion for the Rejection of Outlying Observations // Biometrika. 1925. V. 17. No. 3/4. P. 238—250. https://doi.org/10.2307/2332079.
21. *Заляжных В.В.* Расширение области применения критерия Ирвина при обнаружении аномальных измерений // Вестник СибГУТИ. 2020. № 2. С. 95—104.
22. *Grubbs F.E.* Sample Criteria for Testing Outlying Observations // Ann. Math. Statist. 1950. V. 21. No. 1. P. 27—58. doi:10.1214/aoms/1177729885.
23. *Лемешко Б.Ю., Лемешко С.Б.* Расширение области применения критериев типа Граббса, используемых при отбраковке аномальных измерений // Измерительная техника. 2005. № 6. С. 13—19.
24. *Hoaglin D.C., Iglewicz B.* Fine-tuning some resistant rules for outlier labeling // JASA. 1987. V. 82, No. 400. P. 1147—1149.
25. *Большов Л.Н., Смирнов Н.В.* Таблицы математической статистики. М.: Наука, 1965. 464 с.
26. *Tietjen G., Moore H.* Some Grubb's type statistics for the detection of several outliers // Technometrics. 1972. V. 14. P. 583—597.
27. *Caliskan Karagöz, Derya & Aktaş Altunay, Serpil.* Generalized Tietjen–Moore test to detect outliers // Mathematical Sciences. 2018. V. 12. P. 7—15. https://doi.org/10.1007/s40096-017-0239-8.
28. *Thompson, R.* A Note on Restricted Maximum Likelihood Estimation with an Alternative Outlier Model // Journal of the Royal Statistical Society. Series B (Methodological). 1985. V. 47. No. 1. P. 53—55.
29. *Chandola V., Banerjee A., Kumar V.* Anomaly detection: A survey // ACM Computing Surveys. 2009. V. 41. No. 3. Article 15. 58 p.
30. *Wilson C.J.N., Ambraseys N.N., Bradley J., Walker G.P.L.* A new date for the Taupo eruption, New Zealand // Nature. 1980. V. 288. № 5788. P. 252—253.
31. *Einarsson Á., Stefánsdóttir G., Jóhannesson H., Ólafsson J.S., Gíslason G.M., Wakana I., Gudbergsson G., Gardarsson, A.* The ecology of Lake Mývatn and the River Laxá: variation in space and time // Aquatic Ecology. 2004. V. 38. P. 317—348.
32. Реки и озера мира. Энциклопедия. М.: ООО «Изд-во «Энциклопедия», 2012. 928 с.
33. *Lehmusluoto P., Machbub B.* Three tropical crater lakes in Bali (Indonesia). A re-examination of some lakes visited by the German Sunda-Expedition in 1929 // Archiv fur Hydrobiologie. 1989. V. 114. P. 537—553.
34. Limnology in Developing Countries / ed. B. Gopal, R.G. Wetzel. Intern. sci. publ. for the Intern. assoc. for limnology, 1995. V. 1. 230 p.
35. Data book of world lakes // Published by National Institute for Research Advancement. Tokyo, 1984. 512 p.
36. *Wolfe-Simon F., Blum J.S., Kulp T.R., Gordon G.W., Hoeft Sh.E., Pett-Ridge J., Stolz J.F., Webb S.M., Weber P.K., Davies P.C.W., Anbar1 A.D., Oremland R.S.* A bacterium that can grow by using arsenic instead of phosphorus // Science. 2011. V. 332. P. 1163—1166.

References

1. *Pozdnyakov Sh.R., Izmailova A.V., Rasulova A.M.* Unique lakes as an object of scientific interest. *Izvestiya Russkogo Geograficheskogo Obshchestva.* News Russian Geographical Society. 2020, 152 (3): 17—31. [In Russian].
2. *Myakisheva N.V. Mnogokriterialnaya klassifikacija ozer*. Multi-criteria classification of lakes. St. Petersburg: RSHU publishing, 2009. 160 p. [In Russian]
3. *Ryanzhin S.V., Ul'yanova T. Yu.* Geographic information system for lakes of the world — GIS «Lakes of the World» (GIS WORLDLAKE). *Doklady RAN.* Reports RAS. 2000, 370 (4): 542—545. [In Russian].
4. *Rumyantsev V.A., Izmailova A.V., Ul'yanova T.Yu.* Information system «Lakes of the Earth». *Izvestiya Russkogo Geograficheskogo Obshchestva.* News Russian Geographical Society. 2009, 141 (3): 8—14. [In Russian].
5. *Izmailova A.V., Ul'yanova T.Yu.* Information system «Lakes of Russia». *Vodnoe khozyaistvo Rossii.* Water industry of Russia. 2014, 6: 21—28. [In Russian].

6. *Kochkov N.V., Ryanzhin S.V.* Lakes of the world WORLDLAKE. Certificate of state registration of the database No. 2015621549.
7. *Drabkova V. G., Sorokin I. N. Ozero i ego vodosbor — edinaya prirodnaya sistema.* The lake and its catchment are a single natural system. Leningrad: Science. Leningrad department, 1979: 196 p. [In Russian].
8. *Yakushko A.F., Novik A.A.* Problems of ecological sustainability of glacial hollow lakes in the Belarusian lake area. *Vestnik Belorusskogo Gosudarstvennogo Universiteta*. Bulletin of the Belarusian State University. 2005, 1: 55—59. [In Russian].
9. *Grigor'ev S. V.* On some definitions and indicators in lake research. *Trudy Karel'skogo filiala Akademii nauk SSSR.* Proceedings of the Karelian department of the USSR Academy of Sciences. 1959, 18: 29—45. [In Russian].
10. *Vereshchagin G.Yu. Metody morfologicheskoi kharakteristiki ozer.* Methods for the morphological characteristics of lakes. Leningrad: State Hydrological Institute, 1930: 116 p. [In Russian].
11. *Popukailo V.S.* Detection of abnormal measurements when processing small data. *Tekhnologiya i konstruirovanie v ehlektronnoi apparature.* Technology and design in electronic equipment. 2016, 4—5: 42—46. [In Russian].
12. *Gibbons J.D., Chakraborti S.* Nonparametric Statistical Inference, Fifth Edition. CRC Press. Boca Raton, London, New-York. 2014: 614.
13. *Tukey John W.* Exploratory Data Analysis. Reading, Mass: Addison-Wesley, 1977.
14. *Box G.E.P., Cox D.R.* An analysis of transformations. Journal of the Royal Statistical Society. Series B. 1964, 26: 211—252.
15. *Shenk Kh. Teoriya inzhenernogo ehksperimenta.* Engineering experiment theory. Moscow: Mir, 1972: 381 p. [In Russian].
16. *Chauvenet W.* A manual or Spherical and practical astronomy Vol. 2. Theory and use of astronomical instruments. Method of least squares. Philadelphia: J.B. Lippincott Company, London: Trubner, 1876: 632 p.
17. *Kobzar A.I. Prikladnaja matematicheskaja statistika. Dlja ingenerov i nauchnyh rabotikov* Applied Mathematical Statistics. Moscow: FIZMATLIT, 2006: 816 p. [In Russian].
18. *Sergeev A.G., Teregerja V.V. Metrologija, standartizacija I sertifikacija, uchebnik.* Metrology, standardization and certification: textbook Moscow: Urigt. 2011. 820 p. [In Russian].
19. *Radkevich Ja.M. Metrologija, standartizacija I sertifikacija, uchebnik dlja vuzov.* Metrology, standardization and certification: a textbook for universities. Moscow: High school, 2004: 767 p. [In Russian].
20. *Irwin J.O.* On a Criterion for the Rejection of Outlying Observations. Biometrika. 1925, 17 (3/4): 238—250. https://doi.org/10.2307/2332079.
21. *Zaljazhnyh V.V. Rasdhirenie oblasti primenenija kriterija Irvina pri obnaruzhenii anomalnyh izerenij.* Expanding the scope of Irwin's criterion when detecting abnormal measurements // Vestnik SibGUTI. 2020, № 2: 95—104. [In Russian].
22. *Grubbs F.E.* Sample Criteria for Testing Outlying Observations. Ann. Math. Statist. 1950, 21 (1): 27—58. doi:10.1214/aoms/1177729885.
23. *Lemeshko B.Yu., Lemeshko S.B.* Expansion of the scope of the Grubbs-type criteria used to reject anomalous measurements. *Izmeritel'naya tekhnika.* Measuring technology. 2005, 6: 13—19. [In Russian].
24. *Hoaglin D.C., Iglewicz B.* Fine-tuning some resistant rules for outlier labeling. JASA. 1987, 82 (400): 1147—1149.
25. *Bol'shov L.N., Smirnov N.V. Tablitsy matematicheskoi statistiki.* Mathematical statistics tables. Moscow: Nauka, 1965: 464 p. [In Russian].
26. *Tietjen G., Moore H.* Some Grubb's type statistics for the detection of several outliers. Technometrics. 1972, 14: 583—597.
27. *Caliskan Karagöz, Derya & Aktaş Altunay, Serpil.* Generalized Tietjen–Moore test to detect outliers // Mathematical Sciences. 2018, 12: 7—15. https://doi.org/10.1007/s40096-017-0239-8.
28. *Thompson R.* A Note on Restricted Maximum Likelihood Estimation with an Alternative Outlier Model. Journal of the Royal Statistical Society. Series B (Methodological). 1985, 47 (1): 53—55.
29. *Chandola V., Banerjee A., Kumar V.* Anomaly detection: A survey. ACM Computing Surveys. Article 15. 2009, 41 (3): 58 p.

30. *Wilson C.J.N., Ambraseys N.N., Bradley J., Walker G.P L.* A new date for the Taupo eruption, New Zealand. Nature. 1980, 288 (5788): 252—253.
31. *Einarsson Á., Stefánsdóttir G., Jóhannesson H., Ólafsson J.S., Gíslason G.M., Wakana I., Gudbergsson G., Gardarsson A.* The ecology of Lake Mývatn and the River Laxá: variation in space and time. Aquatic Ecology. 2004, 38: 317—348.
32. *Reki i ozera mira. Ehntsiklopediya.* Rivers and lakes of the world. Encyclopedia. Moscow: Publishing house Encyclopedia. 2012: 928 p. [In Russian].
33. *Lehmusluoto P., Machbub B.* Three tropical crater lakes in Bali (Indonesia). A re-examination of some lakes visited by the German Sunda-Expedition in 1929. Archiv fur Hydrobiologie. 1989, 114: 537—553.
34. Limnology in Developing Countries. Ed. B. Gopal, R.G. Wetzel. Intern. sci. publ. for the Intern. assoc. for limnology, 1995, 1: 230 p.
35. Data book of world lakes. Published by National Institute for Research Advancement. Tokyo, 1984: 512 p.
36. *Wolfe-Simon F., Blum J.S., Kulp T.R., Gordon G.W., Hoeft Sh.E., Pett-Ridge J., Stolz J.F., Webb S.M., Weber P.K., Davies P.C.W., Anbar1 A.D., Oremland R.S.* A bacterium that can grow by using arsenic instead of phosphorus. Science. 2011, 332: 1163—1166.

Конфликт интересов: конфликт интересов отсутствует.

Статья поступила 25.11.2020
Принята к публикации после доработки 29.01.2021

Сведения об авторах

Измайлова Анна Владиленовна, д-р геогр. наук, ведущий научный сотрудник лаборатории географии и гидрологии Института озероведения РАН — СПБФИЦ РАН, ianna64@mail.ru.

Расулова Анна Мурадовна, канд. физ.-мат. наук, младший научный сотрудник лаборатории математических методов моделирования Института озероведения РАН — СПБФИЦ РАН, arasulova@gmail.com.

Шмакова Виктория Юрьевна, младший научный сотрудник лаборатории гидрохимии Института озероведения РАН — СПБФИЦ РАН, kulechka08@mail.ru.

Information about authors

Izmailova Anna Vladilenovna, Grand PhD (Geog. Sci.), Leading researcher, laboratory of geography and hydrology of the Institute of Limnology of RAS — SPIIRAS.

Rasulova Anna Muradovna, PhD (Phys. and Math. Sci.), Junior researcher, laboratory of mathematical modeling methods of the Institute of Limnology of RAS — SPIIRAS.

Shmakova Viktoriya Yurevna, Junior researcher, laboratory of hydrochemistry of the Institute of Limnology of RAS — SPIIRAS.