

УДК [556.555:574.5]:51-7
AGRIS M40

<https://doi.org/10.33619/2414-2948/72/08>

ПРИМЕНЕНИЕ АЛГОРИТМА ISOLATION FOREST ДЛЯ ОБОСНОВАНИЯ УНИКАЛЬНОСТИ ВОДОЕМОВ В ГРУППЕ КАРСТОВЫХ ОЗЕР

©*Расулова А. М.*, ORCID: 000-0003-4400-2000, SPIN-код: 3888-6462, канд. ф.-м. наук, Санкт-Петербургский Федеральный исследовательский центр РАН, г. Санкт-Петербург, Россия, ARasulova@gmail.com

©*Измайлова А. В.*, ORCID: 0000-0003-1705-6863, SPIN-код: 9660-9805, д-р геогр. наук, Санкт-Петербургский Федеральный исследовательский центр РАН, Государственный гидрологический институт, г. Санкт-Петербург, Россия, ianna64@mail.ru

APPLICATION OF THE ISOLATION FOREST ALGORITHM TO SUBSTANTIATE THE UNIQUENESS OF WATER BODIES IN THE GROUP OF KARST LAKES

©*Rasulova A.*, ORCID: 000-0003-4400-2000, SPIN-code: 3888-6462, Ph.D., St. Petersburg Federal Research Center of the Russian Academy of Sciences, St. Petersburg, Russia, ARasulova@gmail.com

©*Izmailova A.*, ORCID: 0000-0003-1705-6863, SPIN-code: 9660-9805, Dr. habil., St. Petersburg Federal Research Center of the Russian Academy of Sciences, State Hydrological Institute, St. Petersburg, Russia, ianna64@mail.ru

Аннотация. Совершенствование научных основ развития и расширения сети особо охраняемых природных территорий требует поиска алгоритмов, которые можно было бы использовать для выявления уникальных экосистем. Алгоритмизация процесса идентификации аномалий предоставляет возможность не только обрабатывать большие массивы данных, но и приводит к получению объективных и сопоставимых между собой оценок. Целью настоящего исследования является выявление наиболее оптимальных механизмов для идентификации аномальных значений среди морфометрических характеристик карстовых озер, которые могут свидетельствовать об уникальности всей озерной экосистемы. В рамках данной статьи исследование проводилось на основе математического анализа выборок, построенных по различным признакам на основе базы данных WORLDLAKE. В качестве методов анализа были использованы статистические методы и алгоритм машинного обучения «Изолированный лес» (Isolation Forest (*iForest*)). В результате применения алгоритма *iForest* к выборке морфометрических параметров карстовых озер, состоящей из 738 объектов, были выявлены 43 аномальных водоема. Экспертная оценка итогового набора озер на предмет уникальности их экосистем показала, что выбранный метод идентификации аномальных значений хорошо подходит для поставленной задачи. Многие озера с индексом аномальности выше 60% могут быть признаны уникальными благодаря необычности своих абиотических характеристик, ряд обладает и своеобразной биотой. В аномальные объекты попали такие хорошо известные озера, как Церик-Кель, Чрвено, Салда, Трихонис, Вегоритис, Петрон и др. При этом для большинства из них аномалии были выявлены сразу для нескольких параметров. Таким образом примененный алгоритм для выявления аномальных морфометрических

характеристик озерных котловин, позволил получить интересные выборки для дальнейшего экспертного анализа всей озерной экосистемы на предмет ее уникальности.

Abstract. Improving the scientific foundations for the development and expansion of the network of specially protected natural areas requires the search for algorithms that could be used to identify unique ecosystems. Algorithmization of the anomaly identification process provides an opportunity not only to process large amounts of data but also leads to obtaining objective and comparable estimates. The purpose of this research is to identify the most optimal mechanisms for identifying anomalous values for the morphometric characteristics of karst lakes, which may indicate the uniqueness of the entire lake ecosystem. Within the framework of this article, the study was carried out based on a mathematical analysis of samples built for various characteristics based on the WORLDLAKE database. Statistical methods and the Isolation Forest (*iForest*) machine learning algorithm were used as methods of analysis. As a result of applying the *iForest* algorithm to a sample of morphometric parameters of karst lakes, consisting of 738 objects, 43 anomalous water bodies were identified. An expert assessment of the final set of lakes for the uniqueness of their ecosystems showed that the chosen method for identifying anomalous values is well suited for the task at hand. Many lakes with an anomaly index above 60% can be recognized as unique due to the unusualness of their abiotic characteristics; a number of them also have a peculiar biota. The anomalous objects included such well-known lakes as Tserik-Kol', Crveno, Salda Lake, Trihonida, Vegoritida, Petron, etc. Moreover, for most of them, anomalies were detected for several parameters at once. Thus, the applied algorithm for identifying anomalous morphometric characteristics of lakes made it possible to obtain interesting samples for further expert analysis of the entire lake ecosystem for its uniqueness.

Ключевые слова: озерная экосистема, уникальные озера, идентификация аномалий, карстовые озера, изолированный лес, дерево решений.

Keywords: lake ecosystem, unique lakes, identification of anomalies, karst lakes, Isolated Forest, decision tree.

Введение

В настоящее время крайне актуален вопрос поиска объективных критериев обоснования уникальности какого-либо объекта (в нашем случае – озера) с целью рассмотрения возможности придания ему охранного статуса. Антропогенное загрязнение водных объектов на значительной территории страны достигло столь значимых масштабов, что сохранность наиболее ценных из них может обеспечиваться только в случае закрепления за ними режима охраняемых территорий различного уровня. При этом уровень охраны должен определяться не только эстетической, культурной и рекреационной значимостью водоема, но и сложностью и уникальностью его экосистемы. Оценка последней требует научного подхода, тщательного изучения особенностей функционирования конкретного озера. Согласно последней оценке [1], в России дешифрируется около 3,8 млн. водоемов естественного происхождения. Лимнологическая изученность страны остается достаточно низкой. В этой связи использование математических подходов для первичного выделения водоемов – претендентов на охранный статус — на основе аномальности их характеристик представляется крайне перспективным.

Как указывалось [2], среди всех параметров, характеризующих озерную экосистему, наиболее легко определяются абиотические, и, прежде всего, – морфометрические.

Определение последних возможно не только благодаря полевым исследованиям, но и с помощью геоинформационных систем (ГИС) и спутниковых снимков. К настоящему времени нами накоплен большой объем данных по озерам мира, содержащий сведения о различных морфометрических характеристиках озерных котловин, их происхождении. В этой связи первичная оценка озер на предмет уникальности их экосистем возможна именно с применением выборок различных морфометрических характеристик. Генетическая принадлежность озерных котловин и их морфометрические особенности в значительной степени определяют протекающие в них процессы и трансформацию устойчивости лимносистем [3, 4]. В то же время, как отмечалось в работе [5], в силу 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. То есть при идентификации выбросов по разным параметрам, они часто будут выявлены для одних и тех же объектов, что позволяет изначально ориентироваться на наиболее изученные характеристики.

В статье [2] на примере анализа морфометрических параметров озер вулканического происхождения нами была удачно продемонстрирована возможность применения различных статистических критериев идентификации выбросов, которые могут характеризовать водный объект как претендент на звание «уникального». В рамках настоящей статьи аналогичное исследование проводится на примере выборки озер карстового происхождения.

Необходимо отметить, что в силу геологических особенностей, для значительной части России характерно широкое распространение озер карстового происхождения, в том числе в наиболее освоенных регионах. Чаще всего это небольшие по площади водоемы, крайне уязвимые к антропогенному вмешательству. Значительная часть из них уже имеет охранный статус. Среди 1113 ООПТ России, в которых озера являются объектом охраны, на долю карстовых озер приходится 8%.

В ходе исследования на основе классических статистических методов проведено описание исходной выборки морфометрических параметров карстовых озер. Показано что, применение статистических критериев выявления выбросов, как было сделано нами для озер вулканического происхождения, оказалось неоптимальным для карстовых озер. Для идентификации аномальных озер в рамках данной работы применялся алгоритм машинного обучения «Изолированный лес» (Isolation Forest, далее *iForest*). Выявленные в результате работы алгоритма аномальные значения показали хорошее согласие с экспертной оценкой.

Карстовые водоемы и их распространение

К карстовым водоемам обычно относят озера, образованные подземными и поверхностными водами путем растворения веществ, выноса тонких частиц и появления в результате этой деятельности подземных пустот, впоследствии заполненных водой. Такие пустоты могут располагаться как непосредственно в растворимых породах, так и в покрывающие их (иногда – прилегающих к ним) толщах нерастворимых пород. Чаще всего карстовые озера имеют небольшой размер, округлую, овально-вытянутую или лопастную форму. Для многих из них характерна значительная глубина. Формы озерных котловин изменяются от конусообразной до полусферической [6] (показатель емкости конуса – 0,33, а полусферы – 0,67).

К карстовым котловинам, в силу схожести происходящих процессов, ряд лимнологов относят и заполненные водой углубления в земле, образованные в зонах вечной мерзлоты в результате таяния подземных пластов льда, так называемые термокарстовые озера. Однако, в

силу заведомых различий в формировании экосистем в водоемах зоны вечной мерзлоты, термокарстовые озера нами в выборку включены не были. Кроме того, как было показано в [7], характер поведения морфометрических параметров озерных котловин карстовых и термокарстовых озер несколько отличен. Б.Б. Богословский [8] считал, что по действующему фактору образования термокарстовые водоемы наиболее близки к ледниковым. В то же время нами рассматривались некоторые водоемы смешанного происхождения, расположенные в карстовых регионах, так как предполагалось, что их экосистемы будут характеризоваться схожим функционированием.

В Российской Федерации озера карстового происхождения имеют обширное распространение на Онежско-Двинском водоразделе, в центральной части Русской равнины, особенно в Поволжье, на юге Западной Сибири; их наибольшая концентрация характерна для плоских равнинных участков, где достаточно увлажненно и распространены карбонатные породы. Наиболее глубокие карстовые водоемы встречаются в горных областях – на Северном Кавказе, Южном Урале и Приуралье, в Приангарье. В зарубежной Европе карстовые озера наиболее распространены на Балканском полуострове, однако небольшие группы встречаются и в других регионах Европы. Согласно недавним картографическим оценкам, 21,6% поверхности суши в Европе характеризуется карбонатными породами [9]. Карстовые озера распространены и еще в целом ряде регионов мира. В зарубежной Азии они прежде всего встречаются на полуострове Малая Азия, на Большом и Малом Кавказе, а также, единичные водоемы – в различных частях Южной, Юго-Восточной, Восточной и Средней Азии. В Северной Америке обилие карстовых озер обнаруживается во Флориде, Юкатане и в Кордильерах.

К сожалению, в рамках настоящего исследования мы были ограничены имеющимися у нас базой данных (далее БД). Доступные данные охватывают далеко не все карстовые озера мира, однако по России они включают практически все водоемы, происхождение которых в литературе указывается как карстовое. Изначально БД включала только батиметрически измеренные водоемы, по всем карстовым озерам мы сочли необходимым дополнить ее морфометрическими характеристиками, полученными с использованием спутниковых снимков.

Методика исследования

Проведенный в ходе исследования математический анализ исходной выборки, опираясь на морфометрические характеристики озер (Таблица 1). Для выбора методов идентификации аномалий была проведена предварительная статистическая проверка на нормальность распределения по указанным морфометрическим параметрам, результаты которой приведены в табл. 2. Выборка карстовых озер состояла из 738 объектов, однако не все из них имели полный набор данных (некоторые морфометрические характеристики не были определены). Минимальное количество валидных объектов в выборке содержится по параметру S (площади водосбора озера) и составляет 248 озеро. Максимальное — для длины L , ширины W , коэффициента сжатия k_s , показателя удлиненности l/k_s — 738 объектов. Таким образом для любого из параметров по размеру выборки можно говорить о статистически значимых результатах.

Таблица 1

МОРФОМЕТРИЧЕСКИЕ ПАРАМЕТРЫ, СОДЕРЖАЩИЕСЯ В БД WORDLAKE

| 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} | Максимальна 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 $H_{avg} = 740$ м) $H_{avg}^* = \frac{H_{avg}}{H_{avgB}}$ |
| H_{max}^* | Безразмерная максимальная глубина (по отношению к максимальной глубине озера Байкал $H_{max} = 1650$ м) $H_{max}^* = \frac{H_{max}}{H_{maxB}}$ |
| 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 |

Таблица 2

СТАТИСТИЧЕСКИЕ ОЦЕНКИ МОРФОМЕТРИЧЕСКИХ ХАРАКТЕРИСТИК КАРСТОВЫХ ОЗЕР

| Хар-ка | N | Диапазон | Минимум | Максимум | Среднее | Стандартная ошибка отклонения среднего | Дисперсия, D | Асимметрия, А | Стандартная ошибка асимметрии | Экцесс, Е | Стандартная ошибка эксцесса |
|--------|-----|-----------------------|-----------------------|----------------------|-----------------------|--|-----------------------|-----------------------|-------------------------------|-----------------------|-----------------------------|
| L | 738 | 54,996 | $4,20 \cdot 10^{-3}$ | 55 | 1,734 | 0,130 | 3,523 | 12,41 | $8,358 \cdot 10^{-1}$ | 104,046 | 0,18 |
| W | 738 | 19,996 | $4,00 \cdot 10^{-3}$ | 20 | 0,794 | $5,27 \cdot 10^{-2}$ | 1,432 | 2,05 | $8,988 \cdot 10^{-1}$ | 69,625 | 0,18 |
| S | 737 | 651 | $1,10 \cdot 10^{-3}$ | 651 | 3,980 | 1,126 | 30,556 | 933,679 | $9,005 \cdot 10^{-1}$ | 340,926 | 0,18 |
| ℓ | 555 | 225,93 | $7,00 \cdot 10^{-2}$ | 226 | 6,123 | 0,563 | 13,269 | 176,079 | 0,104 | 147,089 | 0,207 |
| Havg | 453 | 244,75 | 0,25 | 245 | 6,202 | 0,651 | 13,864 | 192,223 | 0,115 | 197,84 | 0,229 |
| Hmax | 590 | 286,5 | 0,50 | 287 | 15,440 | 1,005 | 24,401 | 595,41 | 0,101 | 58,297 | 0,201 |
| V | 446 | 2,751 | $4,63 \cdot 10^{-7}$ | 2,75 | $3,29 \cdot 10^{-2}$ | $8,43 \cdot 10^{-3}$ | 0,178 | $3,17 \cdot 10^{-2}$ | 0,116 | 135,712 | 0,231 |
| C | 248 | 5019,9 | 0,10 | $5,02 \cdot 10^{-1}$ | 112,159 | 25,056 | 394,589 | $1,557 \cdot 10^{-1}$ | 0,155 | 101,175 | 0,308 |
| ks | 738 | 2,913 | $8,72 \cdot 10^{-2}$ | 3 | 0,565 | $9,136 \cdot 10^{-1}$ | 0,248 | $6,16 \cdot 10^{-2}$ | 1,253 | $8,988 \cdot 10^{-1}$ | 11,31 |
| I/ks | 738 | 11,137 | 0,333 | 11,47 | 2,306 | $5,808 \cdot 10^{-1}$ | 1,578 | 2,489 | 2,581 | $8,988 \cdot 10^{-1}$ | 8,031 |
| c | 408 | 3,895 | $7,81 \cdot 10^{-2}$ | 3,97 | 0,481 | $1,187 \cdot 10^{-1}$ | 0,240 | $5,747 \cdot 10^{-1}$ | 0,121 | 109,897 | 0,241 |
| H*avg | 453 | 0,331 | $3,38 \cdot 10^{-4}$ | 0,331 | $8,382 \cdot 10^{-3}$ | $8,803 \cdot 10^{-1}$ | $1,874 \cdot 10^{-1}$ | $3,51 \cdot 10^{-4}$ | 12,267 | 0,115 | 197,840 |
| H*ma | 590 | 0,174 | $3,03 \cdot 10^{-4}$ | 0,174 | $9,358 \cdot 10^{-3}$ | $6,088 \cdot 10^{-1}$ | $1,479 \cdot 10^{-1}$ | $2,187 \cdot 10^{-1}$ | 6,558 | 0,101 | 58,297 |
| S* | 737 | $1,687 \cdot 10^{-3}$ | $2,85 \cdot 10^{-11}$ | 1,687 | $1,031 \cdot 10^{-3}$ | $2,917 \cdot 10^{-1}$ | $7,918 \cdot 10^{-1}$ | $6,27 \cdot 10^{-9}$ | 17,650 | $9,005 \cdot 10^{-1}$ | 340,962 |
| V* | 446 | $3,496 \cdot 10^{-5}$ | $5,8810^{-11}$ | $3,50 \cdot 10^{-1}$ | $4,181 \cdot 10^{-7}$ | $1,071 \cdot 10^{-1}$ | $2,262 \cdot 10^{-1}$ | $5,119 \cdot 10^{-1}$ | 10,502 | 0,116 | 135,734 |
| H** | 590 | 5,665 | $7,89 \cdot 10^{-5}$ | 5,66 | $9,816 \cdot 10^{-2}$ | $1,615 \cdot 10^{-1}$ | 0,392 | 0,154 | 10,876 | 0,101 | 141,752 |
| S/Hav | 453 | 460 | $1,10 \cdot 10^{-5}$ | 460 | 2,116 | 1,032 | 21,965 | 482,458 | 20,187 | 0,115 | 420,355 |
| α | 453 | 816,537 | 0,130 | 816,67 | 11,72 | 2,050 | 43,635 | $1,904 \cdot 10^{-1}$ | 14,9 | 0,115 | 261,398 |
| φ | 542 | 113,235 | -44,995 | 68,240 | 48,395 | 0,577 | 13,434 | 180,473 | -3,204 | 0,105 | 14,818 |
| λ | 544 | 333,222 | 164,547 | 168,676 | 26,606 | 1,744 | 40,673 | 1654,29 | -1,509 | 0,105 | 4,73 |
| Zasl | 503 | 2970,30 | 0,70 | 2971 | 484,667 | 26,794 | 600,933 | $3,611 \cdot 10^{-1}$ | 2,082 | 0,109 | 3,812 |

Анализ Таблицы 2 и сравнение результатов с полученными для вулканических озер [2], показывает, что для карстовых водоемов средние характеристики площади и глубины существенно меньше. Среднее значение коэффициента формы озерной котловины, равное 0,481, говорит, что такие озера по большей части имеют параболическую форму котловины. Для описания формы зеркала озера обычно используют показатель удлиненности. Согласно [10], для водоемов с формой зеркала, близкой к округлой, показатель удлиненности составляет 1,5-3,0. Увеличение показателя удлиненности соответствует вытягиванию формы зеркала озера. В горизонтальной плоскости карстовые озера имеют округлую форму, так как среднее значение их показателя удлиненности $1/k_s=2,306$, однако при этом могут встречаться и озера с сильно вытянутой эллипсоидальной формой зеркала (максимальное $1/k_s=11,47$). То есть, большинство озер карстового типа имеют небольшие по площади, неглубокие озерные котловины параболической или конусообразной формы.

При идентификации аномалий статистическими методами и выборе критериев обработки выбросов необходимо знать вид распределения. Приблизительная форма распределения видна из Таблицы 2. Поскольку статистическими методами наиболее просто идентифицировать выбросы для нормального распределения, то охарактеризуем распределение параметров с точки зрения их приближенности к нормальному распределению. Как видно из значений эксцесса (E), приведенного в Таблице 2, для всех параметров распределение имеет «островершинный» вид, по краям которого находится больше значений, чем вокруг среднего. Для нормального распределения значения $-3 \leq E \leq 3$ (для мягкого критерия) или $-1 \leq E \leq 1$ (для жесткого). Значения асимметрии (A) для морфометрических параметров карстовых озер показывает, что распределение сильно смещено в сторону меньших значений, т.е. имеет сильную левую асимметрию, за исключением коэффициента емкости, значение которого подпадает под нормальное распределение. Более точная проверка на нормальность с помощью критерия Колмагорова-Смирнова ($p=0,05$), показала, что нулевая гипотеза не принимается. Привести все остальные распределения к нормальному виду с помощью преобразования Бокса-Кокса [11], что было возможно для вулканических озер [2], для карстовых водоемов не удалось. Таким образом, при работе с выборкой по карстовым озерам применение статистических критериев идентификации выбросов в случае нормального распределения, оказалось невозможным. Подбор других форм распределения (экспоненциального и Вейбула) к морфометрическим параметрам карстовых озер не показал эффективности. Следовательно, применение классических критериев поиска выбросов в данном случае оказалось неоптимальным.

В качестве альтернативы статистическим методам поиска выбросов, в рамках данной работы, была выбрана идентификация аномальных значений с помощью алгоритма машинного обучения *iForest*, основанного на теории графов. Алгоритм *iForest* предложен в 2008 г. Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou [12–14]. Данный подход к обнаружению аномалий, базируется на изоляции экземпляров (в нашем случае озер) исследуемой выборки без привлечения меры расстояний или плотности (таких как расстояние Махаланобиса или метод k -ближайших соседей). Метод *iForest* основывается на свойствах аномалий:

- 1) аномальных значений всегда мало по сравнению с размером выборки;
- 2) значения атрибутов аномальных экземпляров всегда разнообразно и резко отличаются от нормальных.

Метод *iForest* основан на построении множества нескольких непересекающихся неориентированных бинарных деревьев решений (или деревьев изоляции, *Isolation Tree*, далее *iTree*), так называемый ансамбль деревьев решений. В свою очередь дерево решений

представляет собой граф такой, что: между любой парой вершин имеется только один маршрут, и он не содержит ни одного цикла. Т.е. дерево решений является связным ациклическим графом.

Схематическое представление ансамбля деревьев решений представлено на Рисунке. Дерево состоит из «листьев» (концевые узлы), «веток» (ребра графа), узлов ветвления (вероятностных узлов) и иногда «корня». В машинном обучении «листья» представляют собой целевую функцию, а признаки, от которых она зависит, - ребра, «корень» описывает все пространство признаков. Движение при принятии решения происходит от «корня» к «листьям».

Дадим определение дерева принятия решений (*i*Tree). Пусть T — узел дерева, такой что T — либо концевой узел без дочерних узлов, либо внутренний узел принятия решения с двумя дочерними узлами (T_l , T_r). Деревом принятия решений называется такое дерево, что принятие решения на узле T состоит из атрибута q и значения разделения p , так что принятие решения $q < p$ определяет переход либо к T_l либо к T_r .

Изоляционное дерево строится следующим образом: пусть $X = \{x_1, \dots, x_n\}$ — набор данных из n точек с количеством вариационных атрибутов d . Для построения *i*Tree используется подвыборка из ψ -экземпляров, таких что $X' \subset X$. X' делится рекурсивно, случайным образом выбирается атрибут q и значение разделения p пока: 1) узел имеет только один экземпляр или 2) все данные на узле не принимают одинаковые значения. *i*Tree считается полностью выросшим если все экземпляры разные и каждый из них изолирован от «листа». Полное дерево решений имеет ψ — внешних, $\psi - 1$ — внутренних и $2\psi - 1$ — общего количества узлов. Аномалией признается экземпляр, который наиболее просто изолировать.

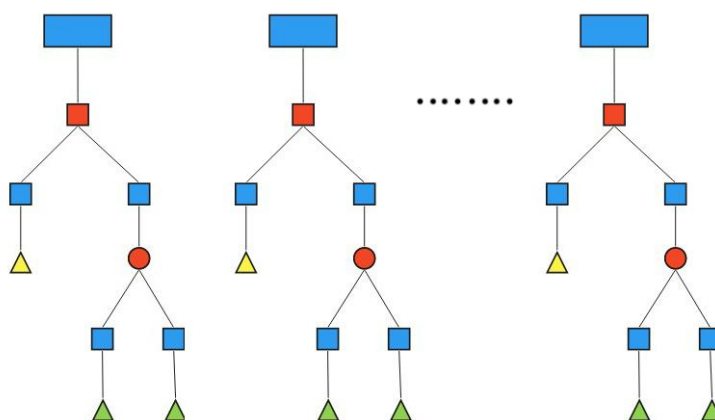


Рисунок. Структура ансамбля решающих деревьев. Обозначения: квадраты - узлы принятия решений, кружки - вероятностные узлы, треугольники – концевые узлы

Мерой восприимчивости к изоляции экземпляра считают длину пути от «корня» до «листа». Или более в строгом математическом определении: длиной пути $h(x)$ точки данных x называется количество ребер, используемых в пути по дереву решений от корневого узла до внешнего узла. Если длина пути $h(x)$ короткая, то восприимчивость к изоляции высокая, т.е. экземпляр с минимальным $h(x)$ является аномальным. Следовательно, конечный узел (лист) желтого цвета на Рисунке является аномальным.

Оценка аномалий, как уже было сказано выше, основана на $h(x)$. Однако, расчет длины пути затруднительны из-за разницы в скорости роста средней высоты *i*Tree, которая имеет порядок $\log \psi$, и максимально возможной высоты, которая увеличивается в линейном порядке ψ . Поэтому невозможно напрямую сравнить $h(x)$ для подвыборок различного

значения. Для нормализации используется оценка среднего $h(x)$ взятая аналогично неудачному поиску внешних узлов в двоичном дереве поиска [15]. Согласно нему средняя длина неудачных поисков в двоичном дереве поиска имеет вид:

$$c(\psi) = \begin{cases} 2H(\psi - 1) - 2(\psi - 1)/n, & \text{при } \psi > 2 \\ 1, & \text{при } \psi = 2 \\ 0, & \text{при } \psi < 2 \end{cases}$$

где $H(i)$ — гармоническое число, равное $\ln(i) + \gamma$, где γ — постоянная Эйлера. Тогда оценка аномальности экземпляра x имеет вид (так называемый «индекс аномальности»):

$$s(x, \psi) = 2 \frac{E(h(x))}{c(\psi)} \quad (1)$$

где $E(h(x))$ — среднее значение $h(x)$ из ансамбля i Tree. Рассмотрим предельные случаи оценки аномальности экземпляра согласно (1):

1. Если $E(h(x)) \rightarrow 0$, то $s \rightarrow 1$
2. Если $E(h(x)) \rightarrow \psi - 1$, то $s \rightarrow 0$
3. Если $E(h(x)) \rightarrow c(\psi)$, то $s \rightarrow 0,5$

Отсюда следует следующие критерии признания экземпляра аномальным:

1. Если $s \rightarrow 1$, то такой экземпляр можно с уверенностью признать аномальным,
2. Если $s \ll 0,5$, то такой экземпляр можно считать нормальным,
3. Если оценка аномальности дает для всех значений $s \approx 0,5$, то выборка не имеет аномалий.

Liu F. T., Ting K. M., Zhou Z. H. эмпирическим путем нашли, что размера подвыборки $\psi = 2^8 = 256$ обычно достаточно для обнаружения аномалий. В этой же работе было показано как размер подвыборки влияет на производительность идентификации аномалий и достигает своего оптимального значения при размере 2^8 и нечувствителен к дальнейшему увеличению размера [14].

Результаты исследования

При поиске аномалий с помощью алгоритма i Forest в выборке морфометрических характеристик карстовых озер размер подвыбоки ψ равнялся 256. Предполагаемое количество аномалий было ограничено только размером выборки, т.е. 738 экземплярами. Поскольку выборка имеет пропуски данных и не обладает высокоточными значениями, аномальными признавались значения с $s > 0,6$. Результаты работы алгоритма по идентификации аномалий показаны в Таблице 3.

В Таблице 3 приведены озера с аномальными значениями более чем у двух морфометрических параметров.

Таблица 3

ОЗЕРА С АНОМАЛИЯМИ МОРФОМЕТРИЧЕСКИХ ХАРАКТЕРИСТИК, ВЫЯВЛЕННЫХ НА
 ОСНОВЕ МЕТОДА IFOREST

| Код | Название английское (латинское) | Название | Страна | Макс. индекс аномальности | Кол-во выявленных аномалий по разным параметрам |
|-------|---------------------------------------|--------------------------------|-------------------------|------------------------------|---|
| 44012 | Beuşehir (Kiriligol) | Бейшехир | Турция | 82,94% | 8 |
| 21369 | Kush-Murun | Кушмурун | Казахстан | 81,39% | 9 |
| 13772 | Tsrveno (Crveno) | Чрвено (Красное) | Хорватия | 81,14% | 7 |
| 53083 | Salda Lake | Салда | Турция | 79,20% | 11 |
| 31493 | Dukhovoye | Духовое | Россия | 78,83% | 3 |
| 30650 | Bisericii | Бисеричи | Румыния | 78,71% | 5 |
| 53064 | Tserik-Kol' | Церик-Кель (Нижнее Голубое) | Россия | 78,60% | 6 |
| 16 | Band-i-Haibat | Хайбат | Афганистан | 78,60% | 6 |
| 5174 | Livanjsko polje | Бушко Блато | Босния и Герцеговина | 78,31% | 9 |
| 24905 | Sumiyn nuur | Сумийнуур | Монголия | 77,75% | 6 |
| 46205 | George | Джордж | США | 74,96% | 5 |
| 18443 | Parishan | Паришан | Иран | 74,95% | 8 |
| 17594 | Vegoritida | Вегоритис | Греция | 72,80% | 10 |
| 46049 | Crescent | Кресент | США | 72,60% | 9 |
| 17352 | Spiekeroog | Спикеруг | Германия | 72,41% | 4 |
| 15463 | Girotte | Жиротт | Франция | 72,36% | 5 |
| 46831 | Tsala Aporka | Тсала Апока | США | 71,52% | 3 |
| 17589 | Trihonida | Трихонис | Греция | 71,41% | 7 |
| 30647 | Baia Verde I | Бая Верде I | Румыния | 71,36% | 2 |
| 39734 | Uguy | Угуй | Россия | 71,24% | 6 |
| 1311 | Luner See | Лунерзее | Австрия | 70,87% | 5 |
| 30649 | Baia Verde III | Бая Верде III | Румыния | 70,51% | 2 |
| 39744 | Yarkul' | Яркуль | Россия | 70,37% | 6 |
| 18294 | Danau Anggi Giti | Данау Ангги Гити | Индонезия | 69,77% | 9 |
| 18293 | Danau Anggi Gigi | Данау Ангги Гиги | Индонезия | 69,59% | 7 |
| 13731 | Vransko | Вранско | Хорватия | 69,38% | 6 |
| 17585 | Limni Tsivlou | Цивлос | Греция | 69,22% | 5 |
| 43458 | Murtensee | Мютензее | Швейцария | 68,35% | 7 |
| 15532 | Mont Cenis | Мон-Сени | Франция/Италия | 67,56% | 2 |
| 46331 | Jackson | Джэксон | США | 67,55% | 5 |
| 17588 | Ozeros | Озерос | Греция | 67,40% | 3 |
| 44007 | Karaviran | Каравиран | Турция | 67,21% | 7 |
| 39922 | Chusovskoye | Чусовское | Россия | 66,88% | 4 |
| 45345 | Chandler | Чандлер | США | 65,45% | 2 |
| 53150 | Svitiaz | Свитязь | Украина | 65,36% | 8 |
| 46305 | Iamonia | Чмония | США | 65,25% | 5 |
| 46864 | Weir | Вейр | США | 65,20% | 5 |

| Код | Название английское (латинское) | Название | Страна | Макс. индекс аномальности | Кол-во выявленных аномалий по разным параметрам |
|--------|---------------------------------------|--------------|------------|------------------------------|---|
| 31697 | Kaldu | Калды | Россия | 65,17% | 3 |
| 175941 | Petron | Петрон | Греция | 64,97% | 4 |
| 30648 | Baia Verde II | Бая Верде II | Румыния | 64,73% | 2 |
| 17402 | Arendsee | Арендзее | Германия | 64,57% | 4 |
| 40790 | Serebryanka | Серебрянка | Россия | 62,19% | 2 |
| 19 | Band-i-Paner | Панир | Афганистан | 61,17% | 4 |

Как видно из Таблицы 3 аномалиями было признано 43 озера. Для всех водоемов выбросы были выявлены одновременно по ряду параметров, характеризующих глубину (средняя, максимальные и относительная глубины, коэффициент глубинности и коэффициент формы). Также было выявлено от 1 до 23 озера (в зависимости от конкретного морфометрического параметра) с индексами аномальности от 60,01% до 67,57% только по одному морфометрическому параметру, которые не вошли в выборку Таблицы 3.

Обсуждение

Экспертная оценка полученного набора озер свидетельствует, что в него попали многие водоемы, интересные не только своими аномальными морфометрическими характеристиками, но и обладающие уникальными экосистемами. Последнее, отчасти может быть обусловлено своеобразием строения их котловин, а также особенностью питания озера. Среди водоемов с наиболее высокими, множественными аномалиями – оз. Чрвено, Церик-Кель, Салда, Хайбат, Жирот и др.

Известно, что самым глубоким карстовым озером считается Чрвено (Красное), одно из Имотских озер Хорватии, глубина которого оценивается в 287 м [16]. Озеро лежит в глубокой карстовой воронке, имеющей глубину 530 м, и сформированной в результате разрушения подземной пещеры. Соседнее с ним оз. Модро занимает воронку поменьше и имеет глубину 91 м. Вода оз. Чрвено темно-синего цвета. Размеры составляют около 150 x 180 м, в зависимости от уровня воды [17]. В озере обитает эндемичная рыба *Delminichthys adspersus*, в сухое время года встречающаяся также и в окрестных реках и озерах, что предполагает наличие между ними подземных протоков. Согласно проведенному исследованию, индекс аномальности s для оз. Чрвено по параметрам, характеризующим глубину и форму котловины, превысил 81%, для оз. Модра — около 60%.

Чуть меньшая глубина зафиксирована для оз. Церик-Кель, входящего в группу Голубых озер, расположенных в Черекском районе Кабардино-Балкарии на северном склоне Скалистого хребта. Площадь озера 0,0216 км², глубина — 279 м [18]. Озеро представляет собой одновременно карстовую шахту (пропасть) и карстовый источник. Оно относится к коррозионным водоемам, образованным напорными водами зоны вертикальной восходящей циркуляции [19]. Поверхностный приток отсутствует. Вода озера зеленовато-голубого цвета, имеет преимущественно сульфатно-кальциевый состав, содержит растворенный сероводород, наличие которого, согласно [20], обусловлено развитием в водоносном горизонте анаэробных бактериальных процессов. В отличие от других озер, входящих в группу Голубых, оз. Церик-Кель характеризуется отсутствием ихтиофауны. Согласно [18], в

нем обнаружен рачок гаммарус, небольшое количество пиявок вблизи берегов, бокоплавов и водяных жуков около текущего из озера ручья. Индекс аномальности для оз. Церик-Кель по параметрам, характеризующим глубину, превысил 78%.

Среди карстовых водоемов аномальностью не только по глубине ($s \sim 75\%$), но и по форме чаши ($s > 68\%$) характеризуется оз. Салда в Турции (максимальный индекс аномальности выявлен по объему воды – 79,2%). Это солонатоводное карстовое озеро площадью около 44 км² глубиной 184 м с бирюзовой водой. Вода жесткая и высокощелочная [21]. Поверхностный сток отсутствует. Оз. Салда — одно из наиболее глубоких и чистых озер Турции, объявленное особой природоохранной зоной, согласно постановлению Президента от 14.03.2019 №824. Среди эндемиков, обитающих в озере — яровой голянь Бурдур (*Pseudophoxinus burduristicus*) семейства карповых. В зимние месяцы на озере селятся редкие представители семейства утиных — белоглазый нырок (*Aythya nyroca*) и савка (*Oxyura leucosephala*), благодаря чему оно отнесено к водно-болотным угодьям мирового значения.

Среди озер со значительной площадью водного зеркала аномальные характеристики выявлены и для оз. Трихонис (Трихонида), расположенного в Западной Греции, в районе с широким развитием карста. Наибольший выявленный индекс аномальности $s > 71\%$. Площадь озера Трихонис составляет 98,6 км², глубина 58 м. Происхождение водоема смешанное. Возраст древний. Озеро имеет охранный статус (входит в сеть охранных участков «Натура 2000»). Его биота характеризуется высоким разнообразием. В озере обитает 25 видов рыб, 11 из которых являются эндемичными для Греции. Карликовый бычок трихонис (*Economidichthys trichonis*) встречается только в данном озере [22].

Среди озер Греции аномалии выявлены и для таких известных озер как Вегоритис и Петрон, расположенных в 2 км друг от друга на севере Греции. Наибольший индекс аномалии для оз. Вегоритис составляет $s > 72\%$, а для оз. Петрон $s > 65\%$. Оз. Вегоритис имеет площадь зеркала 54 км², оз. Петрон — 12 км², глубина обоих составляет около 70 м. На берегах озер встречаются 162 вида птиц, из которых гнездятся 87. На озерах зимуют утки и гуси, воспроизводятся белоглазый нырок и малый баклан [23]. Благодаря богатству орнитофауны западное побережье озера Вегоритис и область озера Петрон также внесены в сеть охранных участков «Натура 2000».

Известным и достаточно глубоким карстовым озером является оз. Свитязь (площадь зеркала 26,2 км², макс. глубина 58,4 м), расположенное на Украине и, наряду с другими водоемами из группы Шацких озер, входящее в состав Шацкого природного национального парка. Вода в озере гидрокарбонатно-кальциевая, минерализация около 200 мг/дм³ [24]. Наибольший выявленный индекс аномальности составляет $s > 65\%$.

С максимальным индексом аномальности $s > 78\%$ в список уникальных водоемов попало и искусственное озеро Бушко Блато, расположенное на территории Боснии и Герцеговины в карстовом Ливанском поле, характеризующимся множеством уникальных природных явлений и карстовых особенностей.

Большое место в итоговом списке аномалий занимают озера, расположенные на территории нашей страны. Среди них, наряду с уже упомянутом оз. Церик-Кель, такие достаточно крупные озера, в образовании котловин которых принимали участие карстовые процессы, как оз. Чусовское, Духовое, Калды, Угуй. Оз. Чусовское (площадь зеркала 27,8 км²) находится в Пермском крае и включено в ООПТ «Озеро Чусовское». Наибольший выявленный индекс аномальности $s > 66\%$. Озеро расположено на дне заболоченной эрозионно-карстовой котловины. Вода гидрокарбонатно-хлоридно-натриевая, минерализация от 33 до 347 мг/л, богата органикой, имеет желто-коричневый цвет [25]. Озеро является

местом скоплений водоплавающих и околоводных птиц во время весенних и осенних миграций. Гнездятся виды, занесенные в Красную книгу, такие как скопа и чернозобая гагара. Оз. Духовое (6,16 км²) расположено в Бурятии и включено в ООПТ местного значения «Баргузинское побережье Байкала». Наибольший выявленный индекс аномальности составляет $s > 78\%$. Озеро Угуй (24,3 км²) находится в Новосибирской области. Наибольший индекс аномальности $s > 71\%$. У южного берега озера утвержден памятник природы «Мирнинский рям», площадью 476 га, представляющий собой изолированный островной участок таежно-болотной растительности, расположенный среди лесостепного ландшафта на берегу озера, сохранившийся в малоизмененном виде на протяжении тысяч лет. Оз. Калды (16,7 км²) находится в Челябинской области и считается одним из самых ее теплых водоемов. Наибольший выявленный индекс аномалии $> 65\%$.

Исходя из полученной нами выборки, можно сделать вывод, что при индексе аномальности морфометрических характеристик, превышающем 60%, озеро с высокой вероятностью может быть признано уникальным и еще по целому ряду параметров. При этом, при поиске уникальных карстовых озер, характеристики формы чаши (прежде всего показатели глубины), более интересны, нежели характеристики, описывающие форму зеркала.

Проведенное исследование также показало, что аномальные значения морфометрических характеристик нередко были выявлены для озер смешанного происхождения. Это объясняется наличием у них нетипичной чаши и свидетельствует о том, что предложенный аппарат достаточно хорошо определяет выбросы, в нашем случае – котловины иной формы.

Необходимо отметить, что в полученной нами итоговой выборке большинство озер имеют охранный статус, некоторые (например оз. Салда) включены в состав Рамсарских угодий. В то же время, многие водоемы, признанные памятниками природы регионального значения, не были идентифицированы. Это объясняется тем, что присуждение таким озерам охранный статус связано не с их аномальными морфометрическими характеристиками, а с обитанием в них редких краснокнижных видов растений и животных, или с их эстетической, культурной и рекреационной значимостью. Последнее представляется крайне важным, однако может обосновываться только экспертными оценками. С другой стороны, включение в итоговый список озер, по которым была выявлена только одна аномалия по какому-либо параметру, позволяет получить более широкую выборку для ее дальнейшего экспертного анализа. В такой выборке могут быть представлены озера – претенденты на присуждение им охранный статус. Однако, при их поиске необходимо наряду с морфометрическими параметрами рассматривать и другие характеристики среды обитания. Для конкретной территории в этом случае могут быть учтены параметры, описывающие геохимию ландшафта. При этом идентификация аномалий с использованием проверенного в данном исследовании с помощью алгоритма машинного обучения Isolation Forest представляется оправданным.

Выводы

Как показывают, проводимые исследования, использование алгоритма машинного обучения *iForest* для выделения водоемов с уникальными свойствами дает хорошие результаты. Многие озера из полученного в ходе анализа итогового списка отличаются не только аномалиями своих морфометрических параметров, но и обладают своеобразием других абиотических характеристик, а подчас – характеризуются уникальной биотой.

Существенным достоинством примененного в настоящем исследовании алгоритма *iForest* является скорость его выполнения и меньшая трудоемкость в сравнении с анализом, построенным с учетом статистических критериев.

В то же время, необходимо признать, что при анализе различных типов озер мы не смогли применить единый математический аппарат поиска выбросов. В случае с озерами вулканического происхождения предварительный анализ исходных выборок свидетельствовал о возможности применения классических статистических критериев выявления выбросов. На примере озер карстового происхождения нами была продемонстрирована допустимость применения алгоритма машинного обучения *iForest*, в случае, когда распределение не относится к распределениям Гаусса, Вейбула или экспоненциальному (или его нельзя привести к нормальному). В этой связи надо отметить, что в каждом конкретном случае возможность применения различных математических методов при поиске аномалий должна быть надежно обоснована на этапе предварительного анализа выборки.

Таким образом, выбор оптимального метода идентификации аномалий основывается на предварительном анализе выборки и ее принадлежности к той или иной области исследований. Надежность выбранного математического метода достигается путем верификации полученных результатов экспертом.

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

Список литературы:

1. Измайлова А. В. Водные ресурсы озер России // География и природные ресурсы. 2016. №4. С. 5-14.
2. Измайлова А. В. Расулова А. М., Шмакова В. Ю. Выделение озер, обладающих уникальными свойствами, статистическими методами // Гидрометеорология и экология. 2021. № 62. С. 27-51.
3. Драбкова В. Г., Сорокин И. Н. Озеро и его водосбор — единая природная система. Л.: Наука. Ленингр. отд-ние, 1979. 196 с.
4. Мякишева Н. В. Многокритериальная классификация озер. СПб.: Изд-во РГГМУ, 2009. 160 с.
5. Поздняков Ш. Р., Измайлова А. В., Расулова А. М. Уникальные озера как объект научного интереса // Известия РГО. 2020. Т. 152. №3. С. 17-31.
6. Wajkiewicz-Grabowska E., Magnuszewski A., Mikulski Z. Hydrometria. PWN, Warszawa, 1993. 313 P.
7. Ахмедова Н. С. Особенности морфометрического строения различных типов карстовых озер // География: проблемы науки и образования. LXIII Герценовские чтения. Материалы ежегодной Всероссийской научно-методической конференции. СПб.: Полиграф-Ресурс, 2010. С. 360–362.
8. Богословский Б. Б. Озероведение. Изд-во МГУ. 1960. 336 с.
9. Chen Z., Auler A. S., Bakalowicz M., Drew D., Griger F., Hartmann J., Jiang G., et al. The World Karst Aquifer Mapping project: concept, mapping procedure and map of Europe // Hydrogeology J., 2017. No 25 (3). P. 771–785. doi: 10.1007/s10040-016-1519-3
10. Григорьев С. В. О гидрологических типах водоемах // Труды V научной конференции по изучению внутренних водоемов Прибалтики. Минск, 1959. С.46-55.

11. Box G. E. P., Cox D. R. An analysis of transformations // Journal of the Royal Statistical Society: Series B (Methodological). 1964. V. 26. №2. P. 211-243. <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>
12. Liu F. T., Ting K. M., Zhou Z. H. Isolation forest //2008 eighth iee international conference on data mining. IEEE, 2008. P. 413-422. <https://doi.org/10.1109/ICDM.2008.17>
13. Tan S. C., Ting K. M., Liu T. F. Fast anomaly detection for streaming data //Twenty-Second International Joint Conference on Artificial Intelligence. – 2011.
14. Liu F. T., Ting K. M., Zhou Z. H. Isolation-based anomaly detection // ACM Transactions on Knowledge Discovery from Data (TKDD). 2012. V. 6. №1. P. 1-39. <https://doi.org/10.1145/2133360.2133363>
15. Preiss B. R. Data structures and algorithms with object-oriented design patterns in C++. John Wiley & Sons, 2008.
16. Ozimec R., Jalžić B., Jelić D. Preliminarni izvještaj prirodoslovnih istraživanja u okviru Speleoronilačke ekspedicije Crveno jezero 2017 // Subterranea Croatica. 2017. V. 15. №2. P. 34-41.
17. Garasic M. Crveno jezero-the biggest sinkhole in Dinaric Karst (Croatia) // Abbasi, A. et Giesen, N., éditeurs: EGU General Assembly Conference Abstracts. 2012. V. 14. P. 7132.
18. Ефремов Ю. В. Голубое ожерелье Кавказа. Л.: Гидрометеиздат, 1988. 160 с.
19. Максимович Н. Г., Мещерякова О. Ю., Деменев А. Д. Уникальность карстового озера Церик-Кель (Голубое озеро) в Кабардино-Балкарии // Озера Евразии: проблемы и пути решения. Петрозаводск, 2017. С. 92-98.
20. Хмурчик В. Т. Микробиологические исследования оз. Черик-Кель (Кабардино-Балкарская Республика) // Известия Высших учебных заведений. Северо-Кавказский регион. Сер. Естественные Науки, 2013. №3 (175). С. 60-62.
21. Kazanci N., Girgin S., Dügel M. On the limnology of Salda Lake, a large and deep soda lake in southwestern Turkey: future management proposals // Aquatic Conservation: Marine and Freshwater Ecosystems. 2004. V. 14. №2. P. 151-162. <https://doi.org/10.1002/aqc.609>
22. LIMNES TRICHONIDA KAI LYSIMACHEIA GR2310009 Natura 2000. <https://natura2000.eea.europa.eu/Natura2000/SDF.aspx?site=GR2310009>
23. LIMNI PETRON N2K GR1340007 dataforms. Natura 2000. <https://natura2000.eea.europa.eu/Natura2000/SDF.aspx?site=GR1340007>
24. Хільчевський В. К. Перші комплексні гідрохімічні дослідження Шацьких озер на Волині у 1975 р. // Гідрологія, гідрохімія і гідроекологія. 2015. Т. 4(39). С. 64-71.
25. Кадастровый отчет по ООПТ охраняемый природный ландшафт регионального значения «Чусовское озеро» // ООПТ России. <http://oopt.info/>

References:

1. Izmailova, A. V. (2016). Water resources of lakes in Russia. *Geography and natural resources*, (4), 5-14. (in Russian)
2. Izmailova, A. V. Rasulova, A. M., & Shmakova, V. Yu. (2021). Identification of lakes with unique properties using statistical methods. *Hydrometeorology and ecology*, (62), 27-51. (in Russian).
3. Drabkova, V. G., & Sorokin, I. N. (1979). Ozero i yego vodosbor — yedinaya prirodnyaya sistema. Leningrad. (in Russian)
4. Myakisheva, N. V. (2009). *Mnogokriterial'naya klassifikatsiya ozer*. St. Petersburg. (in Russian)

5. Pozdnyakov, Sh. R., Izmailova, A. V., & Rasulova, A. M. (2020). Unique lakes as an object of scientific interest. *Izv. Russian Geographical Society*, 152(3), 17-31. (in Russian).
6. Bajkiewicz-Grabowska, E., Magnuszewski, A., & Mikulski, Z. (1993). *Hydrometria*. PWN, Warszawa.
7. Akhmedova, N. S. (2010). Features of the morphometric structure of various types of karst lakes. *Geography: problems of science and education. LXIII Herzen's Readings. Materials of the annual All-Russian scientific and methodological conference, St. Petersburg, April 22-24, 2010*. St. Petersburg. 360-362. (in Russian)
8. Bogoslovsky, B. B. (1960). *Ozerovedeniye*. Moscow. (in Russian)
9. Chen, Z., Auler, A. S., Bakalowicz, M., Drew, D., Griger, F., Hartmann, J., & Jiang, G., et al. (2017). The World Karst Aquifer Mapping project: concept, mapping procedure and map of Europe. *Hydrogeol. J*, 25 (3), 771–785. <https://doi.org/10.1007/s10040-016-1519-3>
10. Grigoriev, S. V. (1959). About hydrological types of reservoirs. *Proceedings of the V scientific conference on the study of the Baltic inland water bodies*. Minsk. 46-55. (in Russian).
11. Box, G. E., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2), 211-243. <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>
12. Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008, December). Isolation forest. In *2008 eighth IEEE international conference on data mining* (pp. 413-422). IEEE. <https://doi.org/10.1109/ICDM.2008.17>
13. Tan, S. C., Ting, K. M., & Liu, T. F. (2011, June). Fast anomaly detection for streaming data. In *Twenty-Second International Joint Conference on Artificial Intelligence*.
14. Liu, F. T., Ting, K. M., & Zhou, Z. H. (2012). Isolation-based anomaly detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 6(1), 1-39. <https://doi.org/10.1145/2133360.2133363>
15. Preiss, B. R. (2008). *Data structures and algorithms with object-oriented design patterns in C++*. John Wiley & Sons.
16. Ozimec, R., Jalžić, B., & Jelić, D. (2017). Preliminarni izvještaj prirodoslovnih istraživanja u okviru Speleoronilačke ekspedicije Crveno jezero 2017. *Subterranea Croatica*, 15(2), 34-41.
17. Garasic, M. (2012, April). Crveno jezero-the biggest sinkhole in Dinaric Karst (Croatia). In *Abbasi, A. et Giesen, N., éditeurs: EGU General Assembly Conference Abstracts* (Vol. 14, p. 7132).
18. Efremov, Yu.V. (1988). *Goluboye ozherel'ye Kavkaza*. Л.: Гидрометеоздат, 160 p. (in Russian)
19. Maksimovich, N.G., Meshcheryakova O. Yu., Demenev A.D. (2017) The uniqueness of the karst lake Tserik-Kelt (Blue Lake) in Kabardino-Balkaria. *Lakes of Eurasia: problems and solutions*. Petrozavodsk. 92-98. (in Russian)
20. Khmurchik, V.T. (2013). Microbiological studies of the lake. Cherik-Kel (Kabardino-Balkar Republic). *Izv. Institutions of higher education. North Caucasian region. Ser. Naturally. Science*. No 3 (175), 60-62. (in Russian)
21. Kazanci, N., Girgin, S., Dügel, M. (2004). On the limnology of Salda Lake, a large and deep soda lake in southwestern Turkey: Future management proposals. *Aquatic Conservation: Marine and Freshwater Ecosystems*. No. 14, 151-162. <https://doi.org/10.1002/aqc.609>
22. LIMNES TRICHONIDA KAI LYSIMACHEIA GR2310009 Natura 2000. <https://natura2000.eea.europa.eu/Natura2000/SDF.aspx?site=GR2310009>

23. LIMNI PETRON N2K GR1340007 dataforms. Natura 2000.
<https://natura2000.eea.europa.eu/Natura2000/SDF.aspx?site=GR1340007>

24. Khilchevsky, V. K. (2015) The first complex hydro-chemical preliminaries of the Shatsk lakes on Volin at 1975 p. *Hydrology, hydrochemistry and hydroecology*, 4(39), 64-71. (in Ukrainian).

25. Kadastryy otchot po OOPT okhranyayemyy prirodnyy landshaft regional'nogo znacheniya "Chusovskoye ozero". OOPT Rossii. <http://oopt.info/> (in Russian).

*Работа поступила
в редакцию 19.10.2021 г.*

*Принята к публикации
22.10.2021 г.*

Ссылка для цитирования:

Расулова А. М., Измайлова А. В. Применение алгоритма Isolation Forest для обоснования уникальности водоемов в группе карстовых озер // Бюллетень науки и практики. 2021. Т. 7. №11. С. 63-79. <https://doi.org/10.33619/2414-2948/72/08>

Cite as (APA):

Rasulova, A., & Izmailova, A. (2021). Application of the Isolation Forest Algorithm to Substantiate the Uniqueness of Water Bodies in the Group of Karst Lakes. *Bulletin of Science and Practice*, 7(11), 63-79. (in Russian). <https://doi.org/10.33619/2414-2948/72/08>