

УДК [556.555+574.5]:51-7

## ВЫЯВЛЕНИЕ УНИКАЛЬНЫХ ОЗЕР ОСОБО ОХРАНЯЕМЫХ ПРИРОДНЫХ ТЕРРИТОРИЙ ИШИМСКОЙ РАВНИНЫ С ИСПОЛЬЗОВАНИЕМ ГЕОИНФОРМАЦИОННЫХ СИСТЕМ

Корнеенкова Н.Ю., Расулова А.М.

*Институт озероведения Российской академии наук – Санкт-Петербургский федеральный исследовательский центр Российской академии наук, Санкт-Петербург,  
e-mail: ntkorn87@gmail.com*

В настоящее время, когда существующая сеть наблюдений за природными объектами, в том числе водными, крайне разрежена и не отвечает степени современных экологических угроз, как никогда актуальным является вопрос выделения водоемов с особыми характеристиками на заповедных территориях, основная задача которых состоит в сохранении природных комплексов во всем их многообразии. На примере водоемов особо охраняемых природных территорий (ООПТ) Тоболо-Ишимского междуречья продемонстрирована возможность использования методов машинного обучения для идентификации аномальных значений морфометрических характеристик озерных котловин. Морфометрические параметры озерных котловин получены с помощью дешифрирования водных поверхностей по спутниковым снимкам Sentinel-2. Среди методов машинного обучения для идентификации аномалий был выбран одноклассовый метод опорных векторов (One Class Support Vector Machine (OC-SVM)), основанный на классификации объектов в многомерном пространстве признаков. Идентификация аномальных объектов происходит путем построения в пространстве признаков гиперплоскости, разделяющей объекты. Гиперплоскость строится таким образом, что все нормальные данные находятся с одной ее стороны. Аномальным объектом считается точка в n-мерном пространстве, которая не проходит за гиперплоскость. Применение алгоритма одноклассового метода опорных векторов к выборке морфометрических параметров озер, расположенных в пределах ООПТ юга Западной Сибири, показало, что выбранная группа водных объектов является достаточно однородной. В ней были идентифицированы 10 объектов, признанных аномалиями, однако по своим морфометрическим характеристикам они не сильно отклоняются от статистических средних оценок всей выборки. Однако аномальные озера требуют к себе особого внимания в силу взаимозависимости морфометрических параметров и гидрологических процессов, происходящих в водной экосистеме.

**Ключевые слова:** уникальные озера, Ишимская равнина, поиск аномалий, выбросы, методы машинного обучения, одноклассовый метод опорных векторов

## IDENTIFICATION OF UNIQUE LAKES OF PROTECTED AREAS OF THE ISHIM PLAIN USING GEOINFORMATION SYSTEMS

Korneenkova N.Yu., Rasulova A.M.

*Institute of Limnology RAS – Saint Petersburg Federal Research Center RAS, Saint Petersburg,  
e-mail: ntkorn87@gmail.com*

At present, when the existing network of observations of natural objects, including water ones, is extremely sparse and does not correspond to the degree of modern environmental threats, the issue of identifying water bodies with special characteristics in protected areas, the main task of which is to preserve natural complexes in all their diversity. On the example of reservoirs of specially protected natural territories (protected areas) the possibility of using machine learning methods to identify abnormal values of morphometric characteristics of lake basins is demonstrated in the Tobolo-Ishim interfluvium. Morphometric parameters of lake basins were obtained by decoding water surfaces from Sentinel-2 satellite images. One Class Support Vector Machine (OC-SVM), based on the classification of objects in a multidimensional feature space, was chosen among the machine learning methods for identifying anomalies. Identification of anomalous objects occurs by constructing in the space of signs of a hyperplane separating the objects. The hyperplane is constructed in such a way that all normal data is on one side of it. An anomalous object is a point in n-dimensional space that does not pass through a hyperplane. The application of the algorithm of the single-class method of reference vectors to the selection of morphometric parameters of lakes located within the protected areas of the south of Western Siberia showed that the selected group of water bodies is fairly homogeneous. 10 objects recognized as anomalies were identified in it, however, according to their morphometric characteristics, they do not deviate much from the statistical average estimates of the entire sample. However, abnormal lakes require special attention due to the interdependence of morphometric parameters and hydrological processes occurring in the aquatic ecosystem.

**Keywords:** unique lakes, Ishim Plain, anomaly search, outliers, machine learning methods, one-class support vector machine

Особенности климатических и почвенных условий, а также географического положения обуславливают интенсивное сельскохозяйственное освоение лесостепной зоны юга Западной Сибири.

Большая часть Ишимской равнины относится к бессточным районам, в пределах

которых поверхностные воды характеризуются низким потенциалом самоочищения. Установлено, что в районах интенсивного животноводства техногенная миграция веществ сопровождается изменением химического состава и общей минерализации природных вод [1–3]. При этом существуют

пая сеть наблюдений за природными объектами, в том числе водными, крайне разрежена и не отвечает степени современных экологических угроз.

При значительном ухудшении экологического состояния озер на фоне активной антропогенной деятельности требуется принятие специальных консервативных мер для охраны наиболее ценных объектов природы. Выделение водоемов с особыми характеристиками как никогда актуально на заповедных территориях, основная задача которых состоит в сохранении природных комплексов во всем их многообразии.

Задача настоящего исследования – на примере водоемов особо охраняемых природных территорий (ООПТ) Тоболо-Ишимского междуречья продемонстрировать возможность использования ГИС (с целью определения морфометрических характеристик неизученных водоемов) и кластерных моделей идентификации аномальных значений с дальнейшей экспертной оценкой их уникальных характеристик.

Район исследования расположен в средней части Тоболо-Ишимского междуречья в пределах плоской Ишимской равнины, слабо наклонной в северном направлении, со средними высотами 100 м.

Ишимская равнина сложена лессовидными суглинками различного механического состава, глинами, аллювиальными супесями и песками, которые развиты на континентальных отложениях неогенового возраста. В южной части равнины, соответствующей региону исследования, преобладает денудационный рельеф. Гидрографическая сеть представлена озерами и болотами, речная сеть выражена слабо, преимущественно на севере равнины.

Для поверхности равнины характерны суффозионные котловины, вытянутые пологие гряды с широкими плоскими вершинами и пологими склонами. Широко распространены межгрядные ложбины, а также фрагменты днищ древних русел и речных долин, в которых размещаются крупные – площадью несколько десятков квадратных километров (Сиверга, Медвежье и др.), а также средние по площади озера, нередко соленые или горько-соленые.

Тоболо-Ишимское междуречье относится к регионам водного дефицита, тем не менее характерной чертой гидрографической сети рассматриваемой территории является значительная озерность. Помимо крупных и средних по величине остаточных водо-

емов, распространены также многочисленные бессточные озера, занимающие микрозападины площадью, как правило, менее 1 км<sup>2</sup>, которые приурочены к древним поверхностям выравнивания, развитым на дочетвертичных отложениях и перекрытых маломощным лессовидным покровом – пылеватыми известковистыми суглинками и супесями лессовидного облика [4]. Берега таких озер в основном слабо изрезанные, низменные, зачастую заболоченные. Значительная часть пресных мелководных озер подвержена зарастанию высшей водной растительностью. Наличие большого числа водных объектов, расположенных в пределах генетически однородной территории – поверхности выравнивания дочетвертичного возраста, дает хорошую перспективу для исследования морфометрических характеристик озер математическими методами идентификации аномалий.

#### Материалы и методы исследования

В настоящем исследовании рассмотрено 161 озеро в границах Частоозерского государственного природного зоологического заказника, а также в пределах западной части водно-болотных угодий «Озера Тоболо-Ишимской лесостепи», объединяющих несколько заказников.

Оценка количества водоемов и площадей их водной поверхности базировалась на автоматизированном дешифрировании данных Sentinel-2 с использованием многоканального спектрального индекса AWEI (Automated Water Extraction Index) [5]. Обработка данных Sentinel-2 и извлечение площадных характеристик водных объектов осуществлялись в программной среде QGIS.

Для получения морфометрических параметров озер проанализирован снимок Sentinel-2 с пространственным разрешением 10 м за 11 октября 2017 г., полученный посредством сервиса Earth-Explorer.

В ходе автоматизированной обработки спутникового изображения производилась векторизация растровых данных. Методом визуального дешифрирования выполнялось исправление поврежденной или неправильной геометрии полигонов, удаление островов высшей водной растительности и уточнение площадей, после чего производилось вычисление морфометрических характеристик озер. Из итоговых расчетов были исключены водоемы менее 0,005 км<sup>2</sup> в связи с погрешностями определения их площадных и линейных характеристик.

Таблица 1

Статистические оценки морфометрических характеристик  
161 озера Ишимской равнины

Параметр	Мин.	Макс.	Среднее	Стандартное отклонение	Дисперсия	Асимметрия	Экссесс
S	0,02	1,78	0,44	0,33	0,11	1,38	2,12
L	0,2	2,67	0,81	0,37	0,13	1,58	4,55
$W_{max}$	0,16	1,33	0,64	0,24	0,06	0,40	-0,23
$W_{avg}$	0,12	1,01	0,48	0,18	0,03	0,34	-0,18
P	0,72	8,84	3,11	1,39	1,93	1,09	1,63
$k_s$	0,22	0,84	0,62	0,11	0,01	-1,03	1,47
$1/k_s$	1,19	4,46	1,69	0,46	0,21	3,29	14,19
$S^*$	$6,42 \cdot 10^{-8}$	$4,62 \cdot 10^{-6}$	$1,15 \cdot 10^{-6}$	$8,62 \cdot 10^{-7}$	$7,42 \cdot 10^{-13}$	1,38	2,12
K	0,6	0,31	0,36	0,04	$1,90 \cdot 10^{-3}$	2,58	9,17

Районирование в целях выделения статистических совокупностей озер с учетом рельефа территории проводилось на основе данных геоморфологической схемы масштаба 1:2 500 000 [6].

В рамках работ были получены площадные ( $S$ , км<sup>2</sup>) и линейные характеристики 161 водоема в пределах ООПТ Тоболо-Ишимского междуречья. Линейные характеристики включали: длину ( $L$ , км), среднюю ( $W_{avg}$ , км) и максимальную ( $W_{max}$ , км) ширину и периметр ( $P$ , км). Дополнительно рассчитывались: коэффициент сжатия зеркала озера ( $k_s = W_{avg} / L$ ), показатель удлиненности ( $1/k_s$ ), безразмерная площадь озера  $S^* = S / S_{KS}$  (относительно площади Каспийского моря), а также коэффициент изрезанности береговой линии  $\left( K = \frac{L}{2\pi R} = \frac{L}{2\sqrt{\pi S}} \right)$ .

Из числа озер, по которым были получены площадные и линейные характеристики, 150 озер не превышают по площади 1 км<sup>2</sup>.

Статистические оценки морфометрических параметров изучаемых озер Тоболо-Ишимского междуречья приведены в табл. 1. Оценки показывают, что распределение большинства параметров близко к гауссовому, а для  $W_{avg}$  является таковым. Несмотря на то, что распределение морфометрических параметров легко приводится стандартными преобразованиями к нормальному виду и можно применить статистические критерии идентификации выбросов, как это было сделано в [7], в данном исследовании для того, чтобы не трансформировать исходную выборку, воспользуемся нестатистическим методом поиска аномалий, не требующим нормального распределения.

Для анализа выборки использовался одноклассовый метод опорных векторов (One Class Support Vector Machine (OC-SVM)). Реализация алгоритма OC-SVM осуществлялась с помощью библиотеки Python Scikit-learn. Группа методов опорных векторов (Support Vector Machine, SVM) основывается на классификации данных [8, 9]. В этой группе методов выделение отдельных классов происходит с помощью разделения  $n$ -мерного пространства признаков объектов гиперплоскостью, которая располагается так, чтобы расстояние между классами было максимальным.

Пусть  $\{(x_1, y_1), \dots, (x_i, y_i)\} \in X$ , – входные данные, представляющие собой векторы размерности  $n$ , а  $X \in R^n$ , где  $R^n$  – пространство признаков. Каждый  $x_i$  – вектор данных, а  $y_i$  – отвечает за класс, к которому принадлежит  $x_i$ . Основной задачей всех алгоритмов, основывающихся на методе опорных векторов, является нахождение функции  $f(x)$ , такой что отклонение от  $y_i$  меньше  $\varepsilon$  и  $f(x)$  – максимально плоская. В простейшем случае  $f(x)$  может иметь линейный вид [10], т.е.

$$f(x) = \langle w, x \rangle + b, \quad (1)$$

где  $\langle , \rangle$  – скалярное произведение в  $X$ ,  $w$  – вектор нормали к гиперплоскости, параметр  $b$  отвечает за смещение гиперплоскости относительно начала координат. Поскольку было принято допущение, что гиперплоскость описывается линейным уравнением (1), то для этого норма  $\|w^2\| = \langle w, w \rangle$  должна быть минимальна. Тогда задачу поиска такой  $f(x)$ , которая аппроксимирует все  $\{(x_1, y_1), \dots, (x_i, y_i)\}$  с точностью до  $\varepsilon$  можно сформулировать как

$$\min \left\{ \frac{1}{2} \|w\|^2 \right\}, \text{ при } \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon \end{cases} \quad (2)$$

То есть решается задача выпуклой оптимизации.

Особенностью алгоритма одноклассового метода опорных векторов является применение так называемой функции ядра, которая преобразует исходное пространство признаков размерности  $n$  в пространство размерности  $m$ , причем  $m > n$ . Гиперплоскость, отделяющая нормальные данные от аномальных, строится в пространстве  $R^m$ .

При поиске аномалий морфометрических параметров озер Тоболо-Ишимского междуречья использовалось гауссово ядро с радиальной базисной функцией, имеющее вид [11]:

$$K(x_i, x_j) = \exp\left(\gamma \|x_i - x_j\|^2\right), \quad (3)$$

где  $\gamma$  – параметр ядра, который в данной работе выбран пропорционально матрице данных.

Как и в простейшем случае с линейной функцией (1) для группы методов опорных векторов, решается задача выпуклой оптимизации, аналогичная (2), которую можно сформулировать следующим образом:

$$\min \left\{ \frac{1}{2} \|w\|^2 \right\} + \frac{1}{\eta n} \sum_{i=1}^n \xi_i - \rho, \quad (4)$$

где  $\xi_i$  – дополнительные переменные, характеризующие значения ошибки для точек в  $X$ ,  $\eta$  – параметр регуляризации, который отвечает за верхнюю границу доли выбросов,  $\rho$  – параметр, характеризующий максимальное расстояние гиперплоскости

от начала координат. В этом случае функция решения  $f(x)$  приобретает более сложный вид, в отличие от линейной функции (1), через кусочно-постоянную функцию  $\text{sgn}(x)$  и функцию ядра, отвечающую за форму гиперплоскости:

$$f(x) = \text{sgn} \left\{ \sum_{i=1}^n \alpha_i K(x, x_i) - \rho \right\}, \quad (5)$$

где  $\alpha_i$  – множители Лагранжа.

В алгоритме ОС-SVM аномальным объектом считается точка в  $n$ -мерном пространстве, которая не проходит за гиперплоскость.

### Результаты исследования и их обсуждение

Применение алгоритма одноклассового метода опорных векторов к выборке морфометрических параметров озер, расположенных в пределах ООПТ юга Западной Сибири, позволило идентифицировать 10 объектов, которые были признаны аномалиями (рисунок, табл. 2).

Многие из вошедших в итоговый набор водоемов ранее практически не изучались и могут рассматриваться как потенциальные объекты при проведении полевых исследований.

Перечень озер с аномальными характеристиками включает озера с площадями зеркала от 0,02 до 1,78 км<sup>2</sup>. Рассчитанная площадь семи водоемов более 0,5 км<sup>2</sup>, четырех – более 1 км<sup>2</sup>. Наиболее крупные озера – Большое (№ 7), Большое Бердюжье (№ 3), Богородское (№ 5), Чебачье (№ 6).

Таблица 2

Перечень выявленных озер с аномальными параметрами

№ п/п*	Координаты		L	W <sub>ав</sub>	W <sub>max</sub>	P	S	k <sub>s</sub>	1/k <sub>s</sub>	K
	X	Y								
1	68,0466	55,9280	1,5	0,91	1,25	6,96	1,37	0,61	1,64	0,36
2	68,0937	55,9216	0,87	0,2	0,3	2,36	0,17	0,22	4,46	0,6
3	68,3029	55,7981	1,61	1,01	1,3	6,2	1,62	0,63	1,59	0,36
4	68,0692	55,7328	1,4	0,46	0,77	4,52	0,64	0,32	3,08	0,49
5	68,4325	55,6302	2,08	0,54	0,96	6,92	1,12	0,25	3,87	0,55
6	68,9289	55,6128	1,07	0,9	0,98	4,84	0,96	0,84	1,19	0,31
7	68,9017	55,4765	2,67	0,67	1,05	8,84	1,78	0,25	4	0,56
8	68,8229	55,9003	0,98	0,74	0,93	3,92	0,73	0,76	1,32	0,32
9	68,7944	55,8980	0,31	0,25	0,31	1,28	0,08	0,8	1,24	0,31
10	68,3442	55,6860	0,2	0,12	0,16	0,72	0,02	0,62	1,61	0,36



*Озера с аномальными характеристиками, выявленные в пределах ООПТ Тоболо-Ишимского междуречья*

*Условные обозначения: 1 – исследованные озера; 2 – озера с аномальными характеристиками; 3 – границы ООПТ: 1 – государственный природный заказник федерального значения «Белоозерский» (водно-болотные угодья международного значения), II – Частозерский государственный природный зоологический заказник, III – «Уктузский» (перспективный), IV – государственный природный заказник регионального значения «Южный» (водно-болотные угодья международного значения), V – водно-болотные угодья международного значения «Тоболо-Ишимская лесостепь»*

Коэффициент сжатия  $k_s$  у шести водоемов превышает 0,6, в том числе у двух озер  $k_s$  более 0,8 (№ 6 и 9). Для четырех водоемов выявленный  $k_s$  находится в диапазоне от 0,2 до 0,3. Коэффициент изрезанности береговой линии  $K$  изменяется от 0,31 до 0,59. Наибольшие значения  $K$ , превышающие 0,5, рассчитаны для трех озер с минимальным в группе коэффициентом сжатия  $k_s$ .

Наибольший показатель удлиненности  $1/k_s$  получен для озера № 2 (4,46). Наименьшие значения  $1/k_s$  и  $K$  отмечены для озера Чебачье (№ 6 в табл. 2).

Анализ показателя удлиненности  $1/k_s$  показывает, что у шести выявленных озер с аномальными характеристиками форма котловины округлая или близка к округлой, у четырех водоемов близкая к овальной.

#### Заключение

Исследование морфометрических характеристик озер Тоболо-Ишимского меж-

дуречья показало, что выбранная группа водных объектов является достаточно морфометрически однородной. Водоемы с аномальными морфометрическими характеристиками не отклоняются сильно от статистических средних оценок выбранной группы озер. По происхождению озера суффозионные, преимущественно округлой формы, ряд выявленных водоемов отмечается изменением морфометрии в связи с зарастанием высшей водной растительностью (№ 2, 4, 5). Тем не менее идентификация данных водных экосистем одноклассовым методом опорных векторов, как аномальных, указывает на особенности совокупности их морфометрических параметров по сравнению со всей выборкой. Как было показано авторами [7], аномалии по морфометрическим параметрам, в силу взаимообусловленности происходящих в озерах процессов, могут быть связаны с уникальными значениями гидрохимиче-

ских, гидрологических и гидробиологических характеристик водной экосистемы. Вместе с тем говорить об их исключительности в настоящем случае не приходится, как это было сделано в предыдущих исследованиях на глобальных выборках озер [7], в силу недостаточной изученности данных водоемов. Тем не менее выделенные озера, несмотря на достаточно скромные размеры, следует учитывать при изучении возможностей выявления аномальных характеристик водной среды в данном регионе.

*Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта 20-05-00303\22 «Выявление критериев уникальности озерных экосистем».*

#### Список литературы

1. Новохатин В.В., Шепелева Н.А. Картографический метод в анализе динамики состояния озерных экосистем территории водно-болотных угодий Тоболо-Ишимской лесостепи // Вестник ТюмГУ. Науки о Земле. 2013. № 4. С. 154–160.
2. Кошечева Г.С. Особенности геохимического режима и качества вод ландшафтов Ишимской равнины. Ишим: Издательство ИГПИ им. П.П. Ершова, 2014. 172 с.
3. Токарь О.Е., Николаенко С.А. Оценка экологического состояния водных объектов Ишимского района по данным фитоиндикации и прямого гидрохимического анализа // Вестник Тамбовского университета. Сер: естественные и технические науки. 2014. Т. 19. № 5. С. 1573–1576.
4. Зылева Л.И., Коркунов К.В., Козырев В.Е., Пестова Л.Е., Калашникова О.С., Макарова А.Н., Новикова Л.П. Цифровая модель листа Государственной геологической карты Российской Федерации масштаба 1:1 000 000. Третье поколение. Серия Западно-Сибирская. Лист N-42 – р. Ишим. Объяснительная записка. СПб.: Картографическая фабрика ВСЕГЕИ, 2020. 157 с. [Электронный ресурс]. URL: [https://webftp.vsegei.ru/GGK1000/2020\\_n-42/N-42\\_ObZap.pdf](https://webftp.vsegei.ru/GGK1000/2020_n-42/N-42_ObZap.pdf) (дата обращения: 28.06.2022).
5. Feyisa G.L., Meilby H., Fensholt R., Proud S.R. Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery. Remote Sensing of Environment. 2014. Vol. 140. P. 23–35.
6. Пестова Л.Е. Цифровая модель листа Государственной геологической карты Российской Федерации масштаба 1:1 000 000. Третье поколение. Серия Западно-Сибирская. Лист N-42 – р. Ишим. Картографическая фабрика ВСЕГЕИ, 2020. [Электронный ресурс]. URL: [https://vsegei.ru/ru/info/pub\\_ggk1000-3/Zapadno-Sibirskaya/n-42.php](https://vsegei.ru/ru/info/pub_ggk1000-3/Zapadno-Sibirskaya/n-42.php) (дата обращения: 27.06.2022).
7. Измайлова А.В., Расулова А.М., Шмакова В.Ю. Выделение озер, обладающих уникальными свойствами, статистическими методами // Гидрометеорология и экология. 2021. № 62. С. 27–51.
8. Cortes C., Vapnik V. Support-vector networks. Machine learning. 1995. Vol. 20. P. 273–297. DOI: 10.1007/BF00994018.
9. Tax D.M.J., Duijn R. Support vector data description. Machine learning. 2004. Vol. 54 (1). P. 45–66. DOI: 10.1023/B:MACH.0000008084.60811.49.
10. Smola A.J., Schölkopf B. A tutorial on support vector regression. Statistics and Computing Archive. 2004. Vol. 14. No. 3. P. 199–222. DOI: 10.1023/B:STCO.0000035301.49549.88.
11. Lamrini B., Gjini A., Daudin S., Pratumarty P., Armando F. and Travé-Massuyès L. Anomaly Detection using Similarity-based One-Class SVM for Network Traffic Characterization. Proceedings of the 29th International Workshop on Principles of Diagnosis. Warsaw, Poland, 27–30 August, 2018. [Электронный ресурс]. URL: <https://ceur-ws.org/Vol-2289/> (дата обращения: 04.06.2022).