

УДК 502.35

КЛАСТЕРНЫЙ АНАЛИЗ ДЛЯ ОБЪЕКТИВНОГО ЗОНИРОВАНИЯ ВОДНЫХ ОБЪЕКТОВ ПО СТЕПЕНИ ЗАГРЯЗНЕНИЯ

© 2025 Н.А. Жильникова

Санкт-Петербургский государственный университет аэрокосмического приборостроения,
г. Санкт-Петербург, Россия

Статья поступила в редакцию 25.09.2025

В данной статье представлена методика классификации зон загрязненности водных объектов, основанная на алгоритмах кластерного анализа, которая позволяет на основе многомерной оценки гидрохимических и гидробиологических показателей выделять участки водных объектов со схожими характеристиками качества воды и классифицировать их по уровню антропогенной нагрузки. Описан математический аппарат методики, включающий четыре последовательных этапа: подготовку данных, определение мер сходства, иерархическую классификацию и определение оптимального количества кластеров. Предложенный подход обеспечивает объективное разграничение зон различной степени загрязненности водных объектов, что имеет важное значение для обоснованного выбора водоохранных мероприятий и мониторинга их эффективности. Показана возможность интеграции предложенного подхода в существующую систему мониторинга водных объектов. Методика предоставляет объективные данные для оптимизации природоохранных мероприятий и повышения эффективности управления водными ресурсами.

Ключевые слова: водные объекты, кластерный анализ, экологическое зонирование, качество воды, многомерный анализ, иерархическая классификация, антропогенная нагрузка, водоохранные мероприятия.

DOI: 10.37313/1990-5378-2025-27-4(2)-318-325

EDN: QFZGMK

ВВЕДЕНИЕ

Эффективное управление водными ресурсами и обеспечение их экологической безопасности предполагают наличие комплексного подхода к оценке качества водной среды и идентификации источников антропогенного воздействия. Однако сложившаяся практика экологического зонирования водных объектов преимущественно опирается на субъективные экспертные заключения либо на анализ ограниченной выборки показателей, что не позволяет достичь необходимой степени объективности и полноты исследования [1].

В качестве решения данной проблемы предлагается методология, направленная на объективную классификацию зон загрязненности. В ее основе лежит применение методов многомерного статистического анализа, в частности кластеризации данных. Использование данного инструментария обеспечивает возможность интегральной оценки состояния водных участков по всей совокупности анализируемых гидрохимических и гидробиологических показателей и их последующую типологизацию по характеру и интенсивности антропогенной нагрузки

МЕТОДИКА КЛАССИФИКАЦИИ ЗОН ЗАГРЯЗНЕННОСТИ ВОДНЫХ ОБЪЕКТОВ

Задача классификации зон загрязненности водного бассейна формулируется следующим образом. Пусть множество $I = \{I_1, I_2, \dots, I_w\}$ обозначает w объектов (участков или створов водных бассейнов, а также источников загрязнения), принадлежащих некоторой совокупности n . Каждый объект характеризуется набором показателей качества воды $C = \{C_1, C_2, \dots, C_n\}$. Результат измерения k -го показателя для i -го объекта обозначается x_{ik} , а вектор $X_i = [x_{ik}]$ соответствует набору всех показателей для i -го объекта. Таким образом, множество объектов I описывается множеством векторов измерений $X = \{X_1, X_2, \dots, X_w\}$.

Постановка задачи кластеризации заключается в разбиении исходного множества объектов I на t непересекающихся подмножеств (кластеров) $\{C_1, C_2, \dots, C_m\}$ при соблюдении следующих условий:

1. Каждый объект из множества I принадлежит одному и только одному кластеру.

2. Объекты, принадлежащие одному кластеру, должны быть максимально сходны между собой по заданному набору признаков.

Жильникова Наталья Александровна, доктор технических наук, доцент, профессор кафедры инновации и интегрированных систем качества Института фундаментальной подготовки и технологических инноваций.

E-mail: n.zhilnkova@guap.ru

3. Объекты, принадлежащие разным кластерам, должны существенно различаться.
Предлагаемая методика включает четыре последовательных этапа (рис.1).

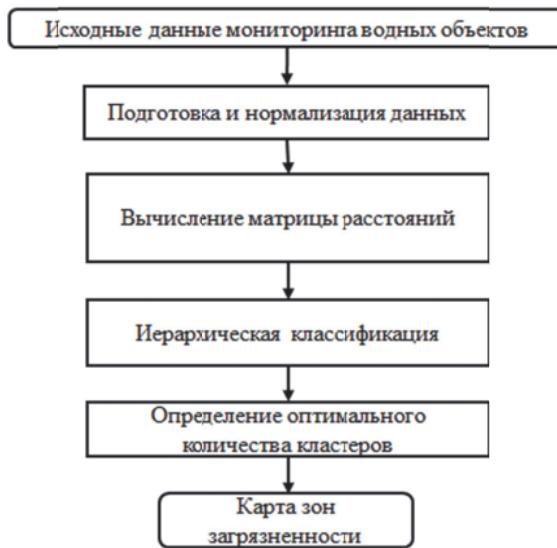


Рис. 1. Алгоритм методики классификации зон загрязненности водных объектов

На начальном этапе исследования выполняется формирование исходной матрицы данных. Со-вокупность эмпирических данных, включающая гидрохимические и гидробиологические показатели качества воды на контролируемых участках водного бассейна, агрегируется в матрицу признаков X следующего вида:

$$\begin{matrix} x_{11} & x_{12} & \dots & x_{1k} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2k} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{i1} & x_{i2} & \dots & x_{ik} & \dots & x_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{w1} & x_{w2} & \dots & x_{wk} & \dots & x_{wn} \end{matrix}, \quad (1)$$

где w – число объектов (участков речного бассейна, створов);

n – число показателей качества воды;

x_{ik} – значение показателя k для i -го объекта.

С целью устранения влияния разномасштабности и различных единиц измерения используемых показателей, что является необходимым условием для корректного вычисления метрик расстояния, применяется процедура стандартизации. В ходе данной процедуры каждое значение x_{ik} преобразуется в стандартизированное значение z_{ik} согласно выражению:

$$Z_{ik} = \frac{x_{ik} - \bar{x}_k}{S_k}, \quad (2)$$

причем

$$\bar{x}_k = \frac{1}{w} \cdot \sum_{i=1}^w x_{ik}, \quad (3)$$

$$S_k = \left[\frac{1}{w} \cdot \sum_{i=1}^w (x_{ik} - \bar{x}_k)^2 \right]^{\frac{1}{2}}, \quad (4)$$

где $k=1,2,\dots,n$;

\bar{x}_k – среднее арифметическое значение k -го показателя;

S_k – стандартное (среднеквадратическое) отклонение k -го показателя;

Z_{ik} – стандартизированное значение k -го показателя для i -го створа.

Второй этап методики заключается в количественной оценке степени сходства между всеми парами объектов наблюдения. В качестве метрики расстояния в многомерном признаковом пространстве используется евклидово расстояние [2].

Результатом данного этапа является симметричная матрица попарных расстояний. Значение $d(X_i, X_j)$ для заданных векторов измерений X_i и X_j , называется расстоянием между X_i и X_j и эквивалентно расстоянию между I_i и I_j соответственно выбранным показателям $\{C_1, C_2, \dots, C_n\}$. Для выполнения процедуры классификации необходимо вычисление элементов матрицы расстояний C между каждой парой объектов i и j . Данная матрица служит основой для последующей процедуры кластеризации. Матрица расстояний между всеми объектами должна иметь следующий вид:

$$C = \begin{bmatrix} 0 & C_{12} & \dots & C_{1i} & \dots & C_{1p} & \dots & C_{1w} \\ C_{21} & 0 & \dots & C_{2i} & \dots & C_{2p} & \dots & C_{2w} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots & \ddots & \vdots \\ C_{i1} & C_{i2} & \dots & 0 & \dots & C_{ip} & \dots & C_{iw} \\ \dots & \dots & \dots & \dots & 0 & \dots & \dots & \dots \\ C_{p1} & C_{p2} & \dots & C_{pi} & \dots & 0 & \dots & C_{pw} \\ \dots & \dots & \dots & \dots & \dots & \dots & 0 & \dots \\ C_{w1} & C_{w2} & \dots & C_{wi} & \dots & C_{wp} & \dots & 0 \end{bmatrix}, \quad (5)$$

где C_{ij} обозначает расстояние между объектами i и j .

Третий этап предполагает проведение иерархической кластеризации на основе полученной матрицы расстояний. Алгоритм начинается с инициализации, при которой каждый из w объектов рассматривается как отдельный кластер. Далее на каждом итеративном шаге два наиболее близких (наименее удаленных) кластера объединяются в один, уменьшая общее число кластеров на единицу. Этот процесс последовательно повторяется до тех пор, пока все объекты не будут объединены в единый глобальный кластер. Визуальное представление результатов иерархического объединения обеспечивается построением дендрограммы, которая графически иллюстрирует последовательность слияния кластеров и расстояния, на которых эти слияния происходят [3].

Элементы этой матрицы служат основой для проведения на третьем этапе иерархической классификации зон загрязненности водных объектов. Для решения рассматриваемой задачи с помощью автоматизации можно рекомендовать способ построения иерархической дендрограммы [3]. Процесс классификации начинается с рассмотрения w объектов множества $I = \{I_1, I_2, \dots, I_w\}$. Множество I представляется как множество кластеров $\{I_1\}, \{I_2\}, \dots, \{I_w\}$. Затем два наиболее сходных по своему множеству признака (показателя) объекта I_i и I_j объединяются в один кластер. Новое множество кластеров, состоящее уже $w-1$ кластеров, будет иметь вид:

$$\{I_1\}, \{I_2\}, \dots, \{I_i, I_j\}, \dots, \{I_w\}. \quad (6)$$

Повторяя процесс, получаются последовательные множества кластеров, состоящие из $w-2$, $w-3$ т.д. кластеров.

Заключительным этапом процедуры иерархической кластеризации является визуализация ее результатов посредством построения дендрограммы, которая представляет собой графическое отображение иерархической структуры данных, иллюстрируя последовательность объединения кластеров и соответствующие им межкластерные расстояния. Процесс формально завершается, когда все исходные объекты объединяются в один глобальный кластер.

Основной задачей на данном этапе является выбор оптимального уровня среза дендрограммы, что эквивалентно определению наиболее устойчивого и содержательного разбиения объектов на кластеры. Критерий оптимальности основывается на фундаментальном принципе кластерного анализа: максимизации межкластерной гетерогенности (различий между кластерами) при одновременной максимизации внутрикластерной гомогенности (сходства объектов внутри одного кластера).

Для формализации этого принципа применяется следующий алгоритм. Величины расстояний $\{d_1, d_2, \dots, d_{w-1}\}$, на которых происходили объединения, ранжируются в убывающем порядке. Далее для каждого уровня слияния вычисляется отношение двух соседних уровней расстояний согласно выражению:

$$i_2 = \frac{d_1}{d_2}, i_3 = \frac{d_2}{d_3}, \dots, i_{w-1} = \frac{d_{w-2}}{d_{w-1}}, \quad (7)$$

где d_1, d_2, \dots, d_{w-1} – упорядоченные по убыванию величины межкластерных расстояний;

i_2, i_3, \dots, i_{w-1} – последовательность вычисленных отношений расстояний.

Выбор оптимального количества кластеров k производится на основе анализа данной последовательности с использованием следующего неравенства:

$$i_k < i_{k+1} \text{ (для } k=2,3,\dots,w-1\text{),} \quad (8)$$

Выполнение данного неравенства для некоторого k свидетельствует о резком скачке расстояния при переходе от k к $k-1$ кластерам, что указывает на объединение двух существенно различных групп объектов и, следовательно, на потенциальную оптимальность разбиения на k кластеров.

В случае, если неравенство (8) выполняется для нескольких значений k , что указывает на наличие нескольких устойчивых иерархических уровней в структуре данных, применяется дополнительное условие для выбора наилучшего разбиения. Данное условие позволяет сравнить два конкурирующих разбиения на k и l кластеров и определяется соотношением:

$$i_k < i_l . \quad (9)$$

Если данное условие выполняется, то можно утверждать, что лучшим является разбиение на k кластеров.

Полученное в результате анализа оптимальное разбиение участков речного бассейна на кластеры является эффективным инструментом для оперативной оценки и прогнозирования качества воды. Методика предполагает применение принципа типизации: из каждого сформированного кластера выбирается один репрезентативный (типовой) участок для углубленного мониторинга. Результаты исследований и прогностические модели, разработанные для данного репрезентативного участка, экстраполируются на все остальные объекты (створы), входящие в тот же кластер. Такой подход позволяет существенно сократить объем полевых и лабораторных работ, оптимизируя затраты на систему мониторинга. На основе типизации также проводится ранжирование объектов по уровню антропогенной нагрузки, социально-экономической значимости и другим критериям [4, 5].

Предложенный подход может быть также использован для сравнительного анализа и классификации объектов из различных водных бассейнов. Для этого формируется объединенная матрица данных вида:

$$\begin{array}{ccccccc} x_{11} & x_{12} & \dots & x_{1k} & \dots & x_{1n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \underline{x_{w1}} & \underline{x_{w2}} & \dots & \underline{x_{wk}} & \dots & \underline{x_{wn}} \\ x_{11} & x_{12} & \dots & x_{1k} & \dots & x_{1n} , \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \underline{x_{j1}} & \underline{x_{j2}} & \dots & \underline{x_{jk}} & \dots & \underline{x_{jn}} \\ \underline{x_{N1}} & \underline{x_{N2}} & \dots & \underline{x_{Nk}} & \dots & \underline{x_{Nn}} \end{array} \quad (10)$$

где w – число объектов исследуемого бассейна;

N – число объектов нового (сравниваемого) бассейна;

x_{ik} – значение k -го показателя для i -го объекта исходного бассейна;

x_{ik} – значение k -го показателя для j -го створа другого (нового) бассейна.

После формирования объединенной матрицы ко всему набору данных применяется процедура стандартизации и последующего кластерного анализа, описанная выше. В результате будет получено новое, единое для двух бассейнов разбиение объектов на классы, что позволит провести их корректный сопоставительный анализ и выявить общие закономерности в формировании качества воды.

Применение описанной выше методики к объединенной выборке объектов из двух речных бассейнов W и N позволяет получить их итоговое распределение по классам (зонам), характеризующимся сходными гидрохимическими и гидробиологическими параметрами.

Результаты кластеризации, демонстрирующие итоговое разбиение, целесообразно представлять в форме сводной таблицы. В таблице 1 представлен пример итоговой классификации, где объекты из исходного W_i бассейна и нового N_j распределены по трем гомогенным классам. Такое распределение наглядно показывает, что участки, географически принадлежащие разным водным системам, могут быть отнесены к одному классу на основании общности их состояния и уровня антропогенной нагрузки. Например, объекты W_2, W_7 (из первого бассейна) и N_1, N_8 (из второго бассейна) объединены в кластер 2, что свидетельствует о высокой степени сходства их показателей качества воды.

Для каждого класса могут быть рассчитаны и представлены средние значения по основным показателям качества воды, что позволяет дать содержательную интерпретацию каждому классу (например, «условно чистые воды», «зоны умеренного загрязнения», «участки с высоким уровнем антропогенной нагрузки»).

Таблица 1 – Результаты кластеризации объектов двух речных бассейнов по показателям качества воды (пример интерпретации)

Номер кластера	Состав кластера (объекты)	Показатели качества воды	Интерпретация (характеристика класса)
1	W_1 W_i N_j	$x_{11} \ x_{12} \dots x_{1k} \dots x_{1n}$ $\underline{x_{i1}} \ \underline{x_{i2}} \dots \underline{x_{ik}} \dots \underline{x_{in}}$ $\underline{x_{j1}} \ \underline{x_{j2}} \dots \underline{x_{jk}} \dots \underline{x_{jn}}$	Условно чистые воды с низкими концентрациями загрязнителей
2	W_2 W_7 N_1 N_8	$x_{11} \ x_{12} \dots x_{1k} \dots x_{1n}$ $\underline{x_{71}} \ \underline{x_{72}} \dots \underline{x_{9k}} \dots \underline{x_{9n}}$ $\underline{x_{11}} \ \underline{x_{12}} \dots \underline{x_{1k}} \dots \underline{x_{1n}}$ $x_{81} \ x_{82} \dots x_{8k} \dots x_{8n}$	Зоны умеренного загрязнения органическими веществами
...	
k	W_3 W_9 N_2 N_5	$x_{31} \ x_{32} \dots x_{3k} \dots x_{3n}$ $\underline{x_{91}} \ \underline{x_{92}} \dots \underline{x_{9k}} \dots \underline{x_{9n}}$ $\underline{x_{21}} \ \underline{x_{22}} \dots \underline{x_{2k}} \dots \underline{x_{2n}}$ $x_{51} \ x_{52} \dots x_{5k} \dots x_{5n}$	Участки с высоким уровнем антропогенной нагрузки

Такая форма представления результатов не только демонстрирует итог работы алгоритма методики классификации зон загрязненности водных объектов, но и служит основой для дальнейшего анализа, включая разработку адресных природоохранных мероприятий и оптимизацию сети мониторинга для обоих водных бассейнов.

СРАВНИТЕЛЬНЫЙ АНАЛИЗ И ВЕРИФИКАЦИЯ РЕЗУЛЬТАТОВ КЛАССИФИКАЦИИ ЗОН ЗАГРЯЗНЕННОСТИ ВОДНЫХ ОБЪЕКТОВ

Для верификации результатов, полученных методом иерархической кластеризации, и для их сопоставления с действующими нормативными подходами проведен сравнительный анализ предложенной методики с подходом к классификации качества воды водного объекта на основе удельного комбинаторного индекса загрязненности воды (УКИЗВ), который официально используется для оценки степени загрязненности поверхностных вод в Российской Федерации и обобщает информацию по ограниченному набору лимитирующих гидрохимических показателей [6, 7].

Проведена оценка данных, полученных в результате анализа проб воды из 6 створов реки Луга (Ленинградская область), по основным гидрохимическим показателям. В таблице 2 приведены исходные данные и значения предельно допустимой концентрации (ПДК) для водоемов рыбохозяйственного значения [8].

Для каждого створа рассчитан УКИЗВ, который учитывает сколько показателей превышают ПДК и насколько. По результатам расчетов створы относятся к разным классам качества воды (таблица 3).

Вывод по УКИЗВ: есть один условно чистый участок (W_1), один слабо загрязненный (W_2), два очень загрязненных (W_3, W_4) и два грязных участка (W_5, W_6). По данной классификации участки четко ранжируются по интегральной степени загрязненности.

Результаты применения кластерного анализа ко всему массиву данных из таблицы 2 (без учета ПДК) в соответствии с предложенной методикой указали на необходимость выделить 3 кластера (таблица 4).

Вывод по кластеризации: алгоритм сгруппировал створы по типу их состояния. Кластер 1 – условно чистые; кластер 2 – участки с органическим/биогенным загрязнением (возможно сточные воды очистных сооружений); кластер 3 – участки с явным промышленным загрязнением (нефтепродукты), которое усугубляется и органикой.

Таблица 2 – Фактические и нормативные значения концентраций гидрохимических показателей в р. Луга

Створ	БПК ₅ , мгО ₂ /л (ПДК=3,0)	Азот аммонийный, мг/л (ПДК=0,4)	Азот нитритный, мг/л (ПДК=0,02)	Нефтепродукты, мг/л (ПДК=0,05)	Растворенный кислород, мг/л (ПДК>6,0)
W ₁	1,5	0,10	0,010	0,02	8,5
W ₂	2,0	0,20	0,015	0,06	8,0
W ₃	7,0	1,50	0,100	0,02	4,5
W ₄	6,5	1,20	0,090	0,04	5,0
W ₅	3,5	0,40	0,050	0,40	5,5
W ₆	8,0	1,00	0,150	0,35	4,0

Таблица 3 – Результаты классификации участков р. Луга по УКИЗВ

Створ	Расчетное значение УКИЗВ	Класс качества воды
W ₁	0,8	1 («условно чистая»)
W ₂	1,1	2 («слабо загрязненная»)
W ₃	3,8	3 («очень загрязненная»)
W ₄	3,5	3 («очень загрязненная»)
W ₅	4,9	4а («грязная»)
W ₆	7,2	4б («грязная»)

Таблица 4 – Результаты иерархической кластеризации участков р. Луга

Номер кластера	Состав кластера (створы)	Характеристика кластера (на основе средних значений)
1 (W ₁ , W ₂)	Фоновые участки	Низкие значения всех загрязнителей, высокое содержание растворенного кислорода
2 (W ₃ , W ₄)	Биогенное загрязнение	Высокие концентрации органики и азота, дефицит кислорода
3 (W ₅ , W ₆)	Промышленное загрязнение	Высокие концентрации нефтепродуктов, органики и азота

Для наглядного сопоставления двух подходов и выявления их взаимосвязи, результаты классификации по УКИЗВ и кластерного анализа для исследуемых створов сведены в единую таблицу (таблица 5), анализ которой позволяет выявить как области согласованности, так и принципиальные различия в логике двух методик.

Результаты сравнительного анализа демонстрируют фундаментальные различия и взаимодополняющий характер двух подходов.

Классификация по УКИЗВ представляет собой нормативно-иерархический подход, основная задача которого – ранжирование водных объектов по интегральной степени загрязненности в соответствии с установленной нормативной шкалой. Данный подход отвечает на вопрос о степени соответствия качества воды нормативным требованиям, осуществляя, таким образом, «вертикальную» оценку состояния объекта.

Кластерный анализ, в свою очередь, реализует исследовательско-структурный подход, цель которого заключается не в ранжировании, а в объективной группировке (типовизации) объектов на основе сходства их многомерных характеристик. Этот подход позволяет выявить доминирующие типы и факторы загрязнения, отвечая на вопрос о внутренней структуре и природе антропогенного воздействия, то есть выполняя «горизонтальную» типизацию.

Проведенный анализ на примере створов W₅ и W₆ наглядно иллюстрирует это принципиальное различие: УКИЗВ дифференцирует их по степени общего загрязнения (классы 4а и 4б), тогда как кластерный анализ объединяет их в одну группу (кластер 3) по общему признаку – наличию промышленного загрязнителя. Одновременно с этим, данный подход четко отделяет их от объектов кластера 2 (W₃, W₄), которые, несмотря на отнесение к близкому классу по УКИЗВ (класс 3), характеризуются принципиально иной природой загрязнения – биогенной.

Таблица 5 – Сводная таблица сопоставления результатов классификации зон загрязненности разными подходами

Створ	Класс качества воды по УКИЗВ	Номер кластера	Интерпретация соответствия
W ₁	1 («условно чистая»)	1	Высокая степень соответствия. Объекты, отнесенные к кластеру 1, однозначно идентифицируются обоими подходами как условно чистые.
W ₂	2 («слабо загрязненная»)	1	
W ₃	3 («очень загрязненная»)	2	Соответствие с уточнением типа загрязнения. Объекты кластера 2 соответствуют классу 3 по УКИЗВ. Однако кластерный анализ позволяет дифференцировать их как участки с преобладающим биогенным типом загрязнения.
W ₄	3 («очень загрязненная»)	2	
W ₅	4а («грязная»)	3	Различие в принципах группировки. УКИЗВ ранжирует объекты W ₅ и W ₆ по разным классам загрязненности (4а и 4б), в то время как кластерный анализ объединяет их в единый кластер 3 на основании общего доминирующего фактора — промышленного загрязнения нефтепродуктами.
W ₆	4б («грязная»)	3	

Таким образом, совместное применение двух подходов обеспечивает комплексную и многоаспектную оценку состояния водных объектов. Нормативная классификация по УКИЗВ дает интегральную оценку уровня загрязненности, в то время как кластерный анализ раскрывает его внутреннюю структуру, что является критически важным для разработки эффективных и целенаправленных водоохраных мероприятий. Например, для решения проблем кластера 2 нужны меры по борьбе со сбросом органики, а для кластера 3 — строгий контроль промышленных источников загрязнения нефтепродуктами.

ЗАКЛЮЧЕНИЕ

В настоящей работе разработана и апробирована методика классификации зон загрязненности водных объектов, основанная на применении методов кластерного анализа. Предложенный математический аппарат, включающий последовательное выполнение этапов подготовки данных, расчета мер сходства, проведения иерархической классификации и определения оптимального числа кластеров, продемонстрировал свою эффективность для обработки многомерных массивов гидрохимических и гидробиологических данных.

Практическая значимость работы заключается в том, что результаты классификации служат надежной основой для разработки целенаправленных водоохраных мероприятий, а также для количественной оценки эффективности их реализации.

Предложенная методика не заменяет, а эффективно дополняет существующую нормативную базу, предоставляя ценную информацию для принятия управленческих решений в области охраны водных ресурсов, и является современным аналитическим инструментом, способствующим повышению эффективности управления водными ресурсами и обеспечению экологической безопасности.

СПИСОК ЛИТЕРАТУРЫ

- Жильникова, Н. А. Методика оценки эффективности функционирования систем водоотведения в жизненном цикле производства продукции в условиях неопределенности / Н.А. Жильникова, А.А. Березина И.А., Шишкун // Датчики и системы. – 2024. – № 4(276). – С. 31-36.
- Кондратьев, С.А. Метод оценки биогенной нагрузки на водные объекты Северо-Запада России / С.А. Кондратьев, С.Л. Басова // Изв. Русского географического общества. – 2009. – Т. 141 (2). – С. 42-52.
- Бова, В. В. Методы и алгоритмы кластеризации текстовых данных (обзор) / В. В. Бова, Ю. А. Кравченко, С. И. Родзин // Известия ЮФУ. Технические науки. – 2022. – № 4(228). – С. 122-143. – DOI 10.18522/2311-3103-2022-4-122-143.
- Корженевский, Б. И. Принципы мониторинга водных объектов на основании изучения деградационных

- процессов / Б. И. Корженевский, А. В. Матвеев // Сельский механизатор. – 2020. – № 2. – С. 10-11.
5. Цибudeева, Д. Ц. Ранжирование водных объектов Республики Бурятия по интенсивности антропогенной нагрузки / Д. Ц. Цибudeева, И. Д. Рыбкина // Водные и экологические проблемы Сибири и Центральной Азии : труды II Всероссийской научной конференции с международным участием, Барнаул, 25–29 августа 2014 года. Том 2. – Барнаул: Типография «Графикс», 2014. – С. 299-306.
 6. Адам, А. М. Применение удельного комбинаторного индекса загрязненности воды как агрегированного показателя качества поверхностных вод на примере р. Томи / А. М. Адам, В. И. Борисенко, Г. И. Мершина // Экология и управление природопользованием. Экологическая безопасность территорий (проблемы и пути решения) : сборник научных трудов IV Всероссийской научно-практической конференции с международным участием, Томск, 27 ноября 2020 года. Том Выпуск 4. – Томск: Общество с ограниченной ответственностью «Литературное бюро», 2021. – С. 4-7.
 7. Фатхутдинова, И. Ш. Оценка изменения мутности и удельного комбинаторного индекса загрязненности (УКИЗВ) воды реки Дема / И. Ш. Фатхутдинова // География, экология, туризм: научный поиск студентов и аспирантов : Материалы XI Всероссийской научно-практической конференции, Тверь, 07 апреля 2023 года. – Тверь: Тверской государственный университет, 2023. – С. 51-54.
 8. Федюченко, Н. Р. Управление качеством воды в бассейне реки Луга / Н. Р. Федюченко // World of science : сборник статей XI Международной научно-практической конференции, Пенза, 30 января 2025 года. – Пенза: Наука и Просвещение (ИП Гуляев Г.Ю.), 2025. – С. 56-62.

CLUSTER ANALYSIS FOR OBJECTIVE ZONING OF WATER BODIES BY POLLUTION DEGREE

© 2025 N.A. Zhilnikova

Saint-Petersburg State University of Aerospace Instrumentation, St. Petersburg, Russia

This article presents a methodology for classifying water body pollution zones based on cluster analysis algorithms. This methodology, based on a multidimensional assessment of hydrochemical and hydrobiological indicators, allows for identifying water body sections with similar water quality characteristics and classifying them by anthropogenic load. The mathematical framework of the methodology is described, including four sequential stages: data preparation, determination of similarity measures, hierarchical classification, and determination of the optimal number of clusters. The proposed approach ensures objective delineation of zones of varying degrees of water body pollution, which is essential for the informed selection of water protection measures and monitoring their effectiveness. The feasibility of integrating the proposed approach into an existing water body monitoring system is demonstrated. The methodology provides objective data for optimizing environmental protection measures and improving the efficiency of water resource management.

Keywords: water bodies, cluster analysis, ecological zoning, water quality, multivariate analysis, hierarchical classification, anthropogenic load, water protection measures.

DOI: 10.37313/1990-5378-2025-27-4(2)-318-325

EDN: QFZGMK

REFERENCES

1. Zhil'nikova, N. A. Metodika ocenki jeffektivnosti funkcionirovaniya sistem vodootvedenija v zhiznennom cikle proizvodstva produkciy v uslovijah neopredelennosti / N.A. Zhil'nikova, A.A. Berezina I.A., Shishkin // Datchiki i sistemy. – 2024. – № 4(276). – S. 31-36.
2. Kondrat'ev, S.A. Metod ocenki biogennoj nagruzki na vodnye ob'ekty Cevero-Zapada Rossii / S.A. Kondrat'ev, S.L. Basova // Izv. Russkogo geograficheskogo obshhestva. – 2009. – Т. 141 (2). – S. 42-52.
3. Bova, V. V. Metody i algoritmy klasterizacii tekstovyyh dannyyh (obzor) / V. V. Bova, Ju. A. Kravchenko, S. I. Rodzin // Izvestija JuFU. Tehnicheskie nauki. – 2022. – № 4(228). – S. 122-143. DOI 10.18522/2311-3103-2022-4-122-143.
4. Korzhenevskij, B. I. Principy monitoringa vodnyh ob'ektov na osnovanii izuchenija degradacionnyh processov / B. I. Korzhenevskij, A. V. Matveev // Sel'skij mehanizator. – 2020. – № 2. – S. 10-11.
5. Cibudeeva, D. C. Ranzhirovaniye vodnyh ob'ektov Respubliki Burjatija po intensivnosti antropogennoj nagruzki / D. C. Cibudeeva, I. D. Rybkina // Vodnye i jekologicheskie problemy Sibiri i Central'noj Azii : trudy II Vserossijskoy nauchnoj konferencii s mezhdunarodnym uchastiem, Barnaul, 25–29 avgusta 2014 goda. – Tom 2. – Barnaul: Tipografia “Grafiks”, 2014. – S. 299-306.
6. Adam, A. M. Primenenie udel'nogo kombinatornogo indeksa zagrjaznennosti vody kak agregirovannogo pokazatelia kachestva poverhnostnyh vod na primere r. Tomi / A. M. Adam, V. I. Borisenko, G. I. Mershina // Jekologija i upravlenie prirodopol'zovaniem. Jekologicheskaja bezopasnost' territorij (problemy i puti reshenija) : sbornik nauchnyh trudov IV Vserossijskoy nauchno-prakticheskoy konferencii s mezhdunarodnym uchastiem, Tomsk, 27 nojabrja 2020 goda. Tom Vypusk 4. – Tomsk: Obshhestvo s ogranicennoj otvetstvennost'ju “Literaturnoe bjuro”, 2021. – S. 4-7.
7. Fathutdinova, I. Sh. Ocenka izmenenija mutnosti i udel'nogo kombinatornogo indeksa zagrjaznennosti (UKIZV) vody reki Dema / I. Sh. Fathutdinova // Geografija, jekologija, turizm: nauchnyj poisk studentov i aspirantov : Materialy XI Vserossijskoy nauchno-prakticheskoy konferencii, Tver', 07 aprelja 2023 goda. – Tver': Tverskoy gosudarstvennyj universitet, 2023. – S. 51-54.
8. Fedjuchenko, N. R. Upravlenie kachestvom vody v bassejne reki Luga / N. R. Fedjuchenko // World of science : sbornik statej XI Mezhdunarodnoj nauchno-prakticheskoy konferencii, Penza, 30 janvarja 2025 goda. – Penza: Nauka i Prosveshhenie (IP Guljaev G.Ju.), 2025. – S. 56-62.

Natalya Zhilnikova, Doctor of Technical Sciences, Professor of the Department of Innovation and Integrated Quality Systems.
E-mail: n.zhilnikova@guap.ru