

**ПРОБЛЕМЫ ЯДЕРНОЙ, РАДИАЦИОННОЙ  
И ЭКОЛОГИЧЕСКОЙ БЕЗОПАСНОСТИ**

УДК 504.4.054

**ОЦЕНКА СОСТОЯНИЯ ВОДНОГО ОБЪЕКТА В ЗОНЕ  
РАЗМЕЩЕНИЯ РОАЭС С ИСПОЛЬЗОВАНИЕМ МЕТОДА  
ГЛАВНЫХ КОМПОНЕНТ**

© 2018 Е.А. Абидова, О.И. Бейсуг, А.Е. Дембицкий, О.Е. Драка, О.Ю. Пугачёва

*НИИ АЭМ, Волгодонский инженерно-технический институт – филиал Национального исследовательского ядерного университета «МИФИ», Волгодонск, Ростовская обл., Россия*

Метод главных компонент был применен для анализа данных экологического мониторинга р. Дон. Метод позволил произвести оценку большого числа разнородных и быстро меняющихся параметров. Результаты указывают на возможность автоматической классификации состояния водного объекта на основании наблюдения его химических, биологических и физических параметров. Предложенная модель классификации может быть применена для обработки данных мониторинга водных экосистем, в том числе Цимлянского водохранилища и водоема-охладителя Ростовской АЭС. Работа выполнена по материалам, полученным в комплексных совместных экспедициях с Донской устьевой станцией Росгидромета в нижнем течении р. Дон на участке от г. Аксай до г. Ростов-на-Дону. Пробы воды отбирались на стационарных пунктах и створах гидрохимической сети мониторинга поверхностных вод суши.

*Ключевые слова:* метод главных компонент, спектр собственных значений, матрица счетов, матрица нагрузок, активность внеклеточных эстераз, щелочная фосфатаза сестона.

Поступила в редакцию: 24.07.2018

Загрязнение водных экосистем в настоящее время приобрело глобальные масштабы. Для поверхностных вод оно имеет особенное значение, так как запасы пресных вод на земле ограничены. Дон является основным источником питьевого водоснабжения и является объектом непосредственного воздействия Ростовской АЭС. При определении экологического состояния дельты р. Дон учитывается достаточно большое количество гидрохимических характеристик, полученных в результате отбора и обработки проб. В настоящем исследовании рассматриваются химические параметры, такие как рН, концентрации кислорода, соединений азота, биологические, среди которых БПК, показатели активности внеклеточных эстераз (АЭ) и активности щелочной фосфатазы (АЩФ), температура.

Для повышения качества обработки данных экологического мониторинга необходим подход, позволяющий автоматически производить анализ данных экологического мониторинга и классифицировать состояние объекта наблюдения. Для решения аналогичных задач в различных областях деятельности (техническая диагностика, финансы) [1, 2, 3, 4] используются алгоритмы классификации. Существующие алгоритмы можно разделить на две группы: вероятностные (статистические) и признаковые. Статистический подход позволяет получить достоверные результаты, если исходные данные получены в результате многолетних, всесезонных, ежесуточных наблюдений, что возможно при работе стационарных экологических постов. В рамках настоящей работы данные получены путем мобильного мониторинга в отдельные моменты времени. Для извлечения информации из этих данных целесообразна реализация одного из признаковых подходов – метода главных компонент [5, 6].

Метод главных компонент позволяет произвести классификацию состояния объекта на определенные группы на основе признаков, полученных по измеренным данным, за счет того, что объекты в одной и той же группе похожи друг на друга с определенной точки зрения. Если классовые метки групп изначально неизвестны, то классификация становится задачей кластеризации. Кластеризация разбивает множество объектов на кластеры, так что объекты в одном кластере более похожи друг на друга, чем на другой кластер. Классификатор позволяет получить абсолютно достоверные результаты только для ситуаций, описанных обучающей выборкой. В нашем случае доступны только неразмеченные данные измерений. Для таких ситуаций были разработаны модели частичной классификации [7, 8].

В модели частичной классификации дополнительная информация по непомеченным данным извлекается из распределения данных и структуры множеств. Частичная классификация основывается на предположении, что соседние точки данных, скорее всего, принадлежат к одному и тому же классу, а также на предположении, что данные с одинаковой структурой, вероятно, будут в одном классе. Корреляции обычно существуют между различными параметрами состояния окружающей среды в связи с их химическим или биологическим взаимодействием. Таким образом, данные, соответствующие одному и тому же состоянию, как правило, соответствуют одному и тому же статистическому распределению или структуре множества.

Очевидно, все рассматриваемые в настоящей работе показатели экологического состояния водотока зависят друг от друга. Значения параметров, соответствующих состоянию объекта, определяются также географией (в отборах, взятых на расстоянии нескольких километров, результаты могут отличаться на порядки), зависят и от сезонной и суточной изменчивости.

Следует отметить, что высокие концентрации оксидов азота и аммиака сами по себе являются показателями экологического неблагополучия. А такие биологические показатели как АЩФ и АЭ показывают реакцию экосистемы на неблагоприятные факторы. Показано [9, 10], что АЩФ и АЭ возрастают в соответствии с ростом экологической нагрузки, их высокие значения сохраняются даже после снижения нагрузки, однако, под действием особо токсичных загрязнителей оба параметра снижаются. То есть нормативные значения АЩФ и АЭ лежат в определенном диапазоне.

Створ	Дата	Время	АЩФ	АЭ	t	С	рН	O <sub>2</sub>	БПК <sub>5</sub>	NH <sub>4</sub>	NO <sub>2</sub>	NO <sub>3</sub>	N общ	N вал	N орг	N взв
			мкмоль/лч	мкмоль/лч				мг/л	мг/л	мг/л	мг/л	мг/л	мг/л	мг/л	мг/л	мг/л
Выше г. Аксай	24.07.2002	12:20	0,562	4,25	25	8,1	7,9	1,67	0,003	0,003	0,012	0,562	1,03	0,544	0,463	
Водозабор Ростов	24.07.2002	13:20	0,314	2,68	25	8,3	6,6	1,63	0,038	0,007	0,168	0,531	0,96	0,318	0,425	
Ниже р.Темерник	24.07.2002	14:28	1,136	4,6	25	8,2	7,3	2,26	0,003	0,009	0,207	0,643	1,16	0,424	0,514	
***																

13

Рисунок 1 – Вид исходных данных [Type of source data]

Данные для анализа представлены в удобной форме для применения метода главных компонент: в виде прямоугольной матрицы  $X$ , строки которой (их принято называть образцами) соответствуют времени и месту отбора проб, а столбцы (переменные) значениям параметров. Размерность матрицы  $I \times J = 30, J = 13$ . Возможно, часть данных является избыточной, не содержит нужной для классификации информации. Метод главных компонент позволяет понизить размерность, т.е. заменить исходную матрицу  $X$  двумя новыми  $T$  и  $P$ , размерность которой  $A$ , меньше числа столбцов исходной матрицы (число строк сохраняется):

$$X = TP^t + E = \sum_{a=1}^A t_a P_a^t + E \tag{1}$$

где  $T$  – матрица счетов, размерностью  $I \times A$ ;  
 $P$  – матрица нагрузок, размерностью  $J \times A$ ;  
 $E$  – матрица остатков, размерностью  $I \times J$ .

Если декомпозиция выполнена правильно, то матрица  $T$  несет в себе столько же информации, сколько ее было в начале, в матрице  $X$ . При этом матрица  $T$  меньше, и, стало быть, проще, чем  $X$ . Матрица счетов  $T$  дает проекции исходных образцов на подпространство главных компонент. Строки  $t_1, \dots, t_I$  матрицы  $T$  – это координаты образцов в новой системе координат. Счета несут информацию, полезную для понимания того, как устроены данные. На графике счетов каждый образец изображается в координатах  $(t_i, t_j)$ , чаще всего –  $(t_1, t_2)$ , обозначаемых PC1 и PC2. Близость двух точек означает их схожесть, т.е. положительную корреляцию. Точки, расположенные под прямым углом, являются некоррелированными, а расположенные диаметрально противоположно – имеют отрицательную корреляцию.

Успех применения метода главных компонент зависит от выбора координат и их числа. Полезным инструментом для решения этой проблемы является график, на котором изображаются собственные значения  $\lambda_i$  в зависимости от числа PC<sub>i</sub> [5]. Собственные значения получаются в результате сингулярного разложения, преобразования близкого к разложению на главные компоненты. В результате сингулярного разложения исходная матрица  $X$  разлагается в произведение трех матриц

$$X=USV^t \tag{2}$$

где  $U$  – матрица, образованная ортонормированными собственными векторами матрицы  $XX^t$ ;  
 $V$  – матрица, образованная ортонормированными собственными векторами матрицы  $X^tX$ ;  
 $S$  – диагональная матрица, элементами которой являются сингулярные значения равные квадратным корням из собственных значений  $\lambda_i$ .

Связь между двумя разложениями определяется следующим простым соотношением:

$$T = US$$

В настоящей работе выбираются главные компоненты, которым соответствуют наибольшие собственные значения.

Прежде чем подвергнуть анализируемые данные разложению на главные компоненты и сингулярному разложению к ним была применена процедура шкалирования с тем, чтобы привести матожидание всех параметров к нулю, а дисперсию к единице.

Створ	АЩФ	АЭ	t C	pH	O2	БПК5	NH4	NO2	NO3	N общ	N вал	N орг	N взв
	мкмоль	мкмоль/лч			мг/л	мг/л	мг/л	мг/л	мг/л	мг/л	мг/л	мг/л	мг/л
Выше г. Аксай	-0.16	1.04	1.37	-2.06	-0.53	-1.18	-1.08	-0.99	-1.22	-0.67	-0.35	0.74	-0.02
Водозабор Ростов	-0.90	-0.54	1.37	-0.24	-1.62	-1.24	-0.44	-0.71	-0.05	-0.85	-0.55	-0.72	-0.19
Ниже р.Темерник	1.56	1.39	1.37	-1.20	-1.03	-0.32	-1.08	-0.57	0.24	-0.23	0.02	-0.03	0.22
...													

Рисунок 2 – Вид данных после шкалирования [Type of data after scaling]

К подготовленным данным применялось сингулярное разложение (стандартная

процедура обработки сигналов, предусмотренная в пакете LabView). В результате разложения получился спектр собственных значений.

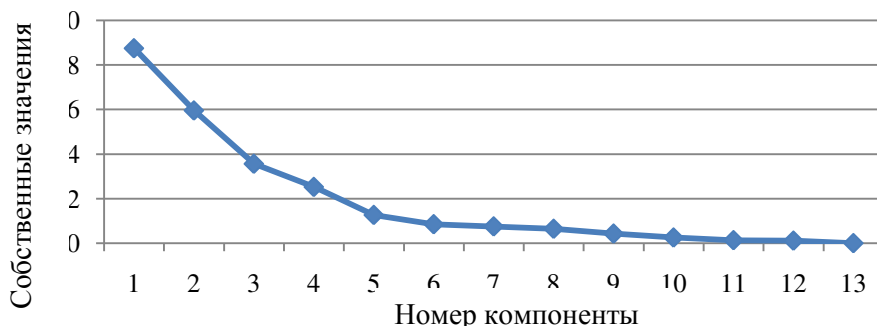


Рисунок 3 – Спектр собственных значений [Eigenvalue spectrum]

График показывает, что главными являются компоненты с номерами 1-4.

После разложения шкалированной матрицы на  $T$  и  $P$  (тоже стандартная процедура обработки сигналов предусмотренная в пакете LabView) была получена матрица счетов. Первые два столбца этой матрицы (из тринадцати) соответствуют двум первым главным компонентам. Исходная матрица была спроецирована на первые главные компоненты. Результаты были представлены в виде тридцати точек на плоскости.

Для интерпретации полученных результатов замеры, соответствующие относительно высокой концентрации кислорода (более 8,5 мг/л) обозначены круглой формой маркера. Если замер соответствует относительно высокой концентрации аммиака (более 0,08 мг/л), маркер имеет форму ромба. Большая концентрация вещества обозначается большим размером.

Характерно, что концентрация аммиака возрастает, хоть и не вполне монотонно, по горизонтали вдоль первой главной компоненты PC1 вплоть до максимального значения данного параметра 0,293 мг/л, а концентрация кислорода в целом возрастает по вертикали вдоль второй главной компоненты PC2.

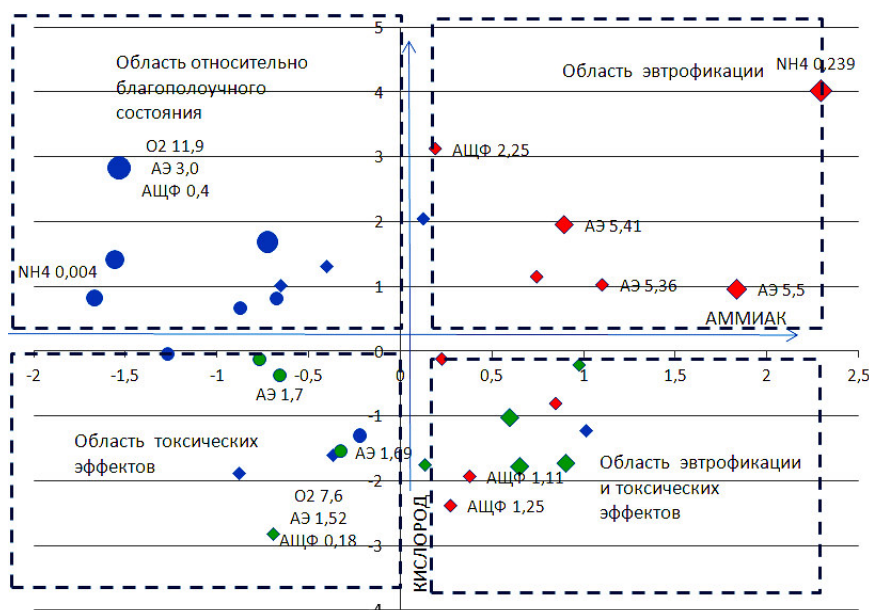


Рисунок 4 – График счетов (PC1 – PC2) [Chart of Accounts (RS1 - RS2)]

Легко видеть, что все круглые маркеры (одиннадцать из тридцати) находятся в

левой полуплоскости, а большинство ромбов в правой.

Можно отметить, что экстремальные значения АЩФ и АЭ наблюдаются везде, кроме левой верхней четверти плоскости, в которой сосредоточены замеры с высоким содержанием кислорода и низким содержанием аммиака. Можно считать, что замеры в этой части плоскости соответствуют относительному (по сравнению с другими замерами) экологическому благополучию. С другой стороны в правой верхней четверти сосредоточились замеры, соответствующие, судя по значениям АЩФ и АЭ, эвтрофикации объекта. В нижней полуплоскости преобладают замеры, которые на основании значений АЩФ и АЭ соответствуют токсическим эффектам.

Таким образом, в результате применения удалось классифицировать результаты замеров по четырем параметрам – концентрация кислорода, аммиака АЩФ и АЭ.

Теперь рассмотрим проекции на компоненты с номерами 3 и 4, которые также отражают существенную информацию о структуре данных. Исследовав новый график счетов совместно с таблицей исходных данных, можно сделать вывод о том, что РС3 отделяет относительно высокую концентрацию оксида азота  $\text{NO}_3$  (более 0,17 мг/л) от более низкой. Высокая концентрация оксида азота обозначена здесь круглой формой маркера, низкая ромбами.

РС4 отделяет относительно низкий рН (менее 8,2) от высокого. Соответственно преимущественно слева оказываются пробы с щелочной реакцией, а справа с кислой. Отметим, что группы в данном случае менее отчетливы, чем на графике счетов первых двух компонент.

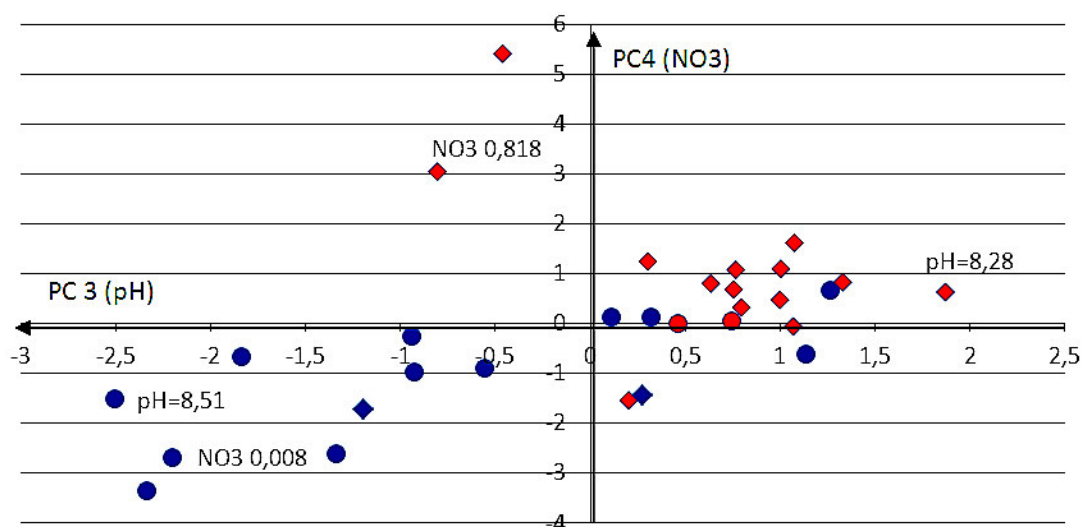


Рисунок 5 – График счетов (PC3 – PC4) [Chart of Accounts (RS3 – RS4)]

Таким образом, метод главных компонент был применен для анализа данных экологического мониторинга р. Дон. Метод позволил произвести оценку большого числа разнородных и быстро меняющихся параметров. Результаты указывают на возможность автоматической классификации состояния водного объекта на основании наблюдения его химических, биологических и физических параметров. Предложенная модель классификации может быть применена для обработки данных мониторинга различных экосистем, в том числе водоема-охладителя Ростовской АЭС. В качестве развития настоящего исследования планируется изучение нетривиальных зависимостей между параметрами состояния экосистем для их комплексной оценки и прогнозирования.

## СПИСОК ЛИТЕРАТУРЫ

1. *Галиаскаров, М.Р.* Сравнительный анализ характеристик мониторинга технологического процесса с использованием линейного и нелинейного метода главных компонент [Текст] / М.Р. Галиаскаров, И.В. Рудакова, Л.А. Русинов // Известия Санкт-Петербургского государственного технологического института (технического университета). – 2013. – №22 (48). – С. 106-110.
2. *Галиаскаров, М.Р.* Система диагностики на базе нелинейного метода главных компонент и дискриминантного анализа Фишера [Текст] / М. Р. Галиаскаров, Л. А. Русинов // Известия Санкт-Петербургского государственного технологического института (технического университета). – 2016. – №33 (59). – С. 91-96.
3. *Захарова, А.А.* Информационная система управления риском банкротства предприятия. Юргинский технологический институт [Текст] / А.А. Захарова. – Томск : Изд-во Томского политехнического университета, 2013.
4. *Zakharova, A.A.* (2013). Fuzzy swot analysis for selection of bankruptcy risk factors. *Applied Mechanics and Materials*. 379 (2013). P. 207-213.
5. *Померанцев, А.* Метод главных компонент (PCA) [Текст] / А. Померанцев. – Российское хемометрическое общество [Электронный ресурс]. – URL : <http://rcs.chemometrics.ru/Tutorials/pca.htm> (дата обращения: 23.12. 2013).
6. *CUAHSI WaterOneFlow Workbook (version 1.1).* A guide to using CUAHSI's WaterOneFlow web services to retrieve hydrologic time series data. 2010 г. [Электронный ресурс]. – URL : [http://his.cuahsi.org/documents/HISDoc5\\_UseWebServices11](http://his.cuahsi.org/documents/HISDoc5_UseWebServices11)
7. *Pfahring, B.A.* semi-supervised spam mail detector. ECML-PKDD Discovery Challenge Workshop. 2006. – URL : <http://www.ecmlpkdd2006.org/pfahring>
8. *Anshul Bhargav, Munish Bhargav.* Pattern Discovery and Users Classification Through Web Usage Mining. – International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT). 2014. С. 12.
9. *Бейсуг, О.И.* Биоиндикация экологического состояния водных экосистем территории воздействия Ростовской АЭС с использованием ферментативных показателей / О.И. Бейсуг // Глобальная ядерная безопасность. – 2013. – №1 (6). – С. 29-32.
10. *Предеина, Л.М.* Влияние повышенных концентраций цинка и железа на активность внеклеточных эстераз и щелочной фосфатазы в природных и модельных пресноводных экосистемах [Текст] / Л.М. Предеина, О.И. Бейсуг, М.Н. Предеин // Изв. вузов. Сев.-Кавк. регион. Естественные науки. – 2006 – №7. – С. 69-81.

## REFERENCES

- [1] *Galiaskarov M.R., Rudakov I.V., Rusinov L.A.* Sravnitel'ny`j analiz karakteristik monitoringa tehnologicheskogo processa s ispol'zovaniem linejnogo i nelinejnogo metoda glavny`x komponent [Comparative Analysis of Process Monitoring Performance Using Linear and Non-Linear Method of Principal Components]. *Izvestiya Sankt-Peterburgskogo gosudarstvennogo tehnologicheskogo instituta (texnicheskogo universiteta)* [Proceedings of the St.Petersburg State Technological Institute (technical university)]. 2013. № 22 (48). P. 106-110 (in Russian).
- [2] *Galiaskarov M.R., Rusinov L.A.* Sistema diagnostiki na baze nelinejnogo metoda glavny`x komponent i diskriminantnogo analiza Fishera [Diagnostic System Based on Nonlinear Principal Components Method and Discriminant Analysis of Fisher]. *Proceedings of the St. Petersburg state technological Institute (technical University)*. 2016. № 33 (59). P. 91-96 (in Russian).
- [3] *Zaharova A.A.* Informacionnaya sistema upravleniya riskom bankrotstva predpriyatiya. Yurginskij tehnologicheskij institut [Information System of Enterprise Bankruptcy Risk Management. Yurga technological Institute]. Tomsk. Publishing House of Tomsk Polytechnic University. 2013 (in Russian).
- [4] *Zakharova, A.A.* (2013). Fuzzy Swot Analysis for Selection of Bankruptcy Risk Factors. *Applied Mechanics and Materials*. 379 (2013). P. 207-213.
- [5] *Pomerantsev A.* Metod glavny`x komponent (PCA) [Principal Component Method (PCA)]. *Rossijskoe xемометрическое obshhestvo* [Russian Chemometric Society]. URL: <http://rcs.chemometrics.ru/Tutorials/pca.htm> (in Russian).
- [6] *CUAHSI WaterOneFlow Workbook (version 1.1).* A guide to using CUAHSI's WaterOneFlow web services to retrieve hydrologic time series data. 2010 г. URL: [http://his.cuahsi.org/documents/HISDoc5\\_UseWebServices11](http://his.cuahsi.org/documents/HISDoc5_UseWebServices11).

- [7] Pfahringer B.A. semi-supervised spam mail detector. ECML-PKDD Discovery Challenge Workshop. 2006. Available at: <http://www.ecmlpkdd2006.org/pfahringer>
- [8] Anshul Bhargav, Munish Bhargav. Pattern Discovery and Users Classification Through Web Usage Mining. – International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT). 2014. P. 12.
- [9] Beysug O.I. Bioindikaciya e`kologicheskogo sostoyaniya vodny`x e`kosistem territorii vozdejstviya Rostovskoj AE`S s ispol`zovaniem fermentativny`x pokazatelej [Bioindication of Ecological State of Aquatic Ecosystems of Rostov NPP Area with the Use of Enzymatic Indicators]. Global'naya yadernaya bezopasnost` [Global Nuclear Safety]. 2013. № 1 (6). P. 29-32 (in Russian).
- [10] Predeina, L.M., Beysug O.I., Predein M.N. Vliyanie povu`shenny`x koncentracij cinka i zheleza na aktivnost` vnekletochny`x e`steraz i shhelochnoj fosfatazy` v prirodny`x i model'ny`x presnovodny`x e`kosistemax [Effect of Elevated Concentrations of Zinc and Iron on the Activity of Extracellular Esterases and Alkaline Phosphatase in Natural and Model Freshwater Ecosystems]. Izvestiya vuzov. Severo-Kavkazskij region. Estestvenny`e nauki [News of Universities. North Caucasus region. Natural Sciences]. 2006. № 7. P. 69-81 (in Russian).

### Assessment of Water Body State in Rostov NPP Area Using the Principal Components Method

E.A. Abidova<sup>1</sup>, O.I. Bejsug<sup>2</sup>, O.E. Draka<sup>3</sup>, A.E. Dembickij<sup>4</sup>, O.Yu. Pugachyova<sup>5</sup>

*Volgodonsk Engineering Technical Institute the branch of National Research Nuclear University "MEPhI",  
Lenin St., 73/94, Volgodonsk, Rostov region, Russia 347360*

<sup>1</sup>ORCID iD: 0000-0003-0258-5543

Wos Researher ID: O-1870-2018

e-mail: [nii\\_energomash@mail.ru](mailto:nii_energomash@mail.ru)

<sup>2</sup>ORCID iD: 0000-0001-9218-9354

Wos Researher ID: O-6513-2018

e-mail: [nii\\_energomash@mail.ru](mailto:nii_energomash@mail.ru)

<sup>3</sup>ORCID iD: 0000-0002-3397-6830

Wos Researher ID: O-2222-2018

e-mail: [nii\\_energomash@mail.ru](mailto:nii_energomash@mail.ru)

<sup>4</sup>ORCID iD: 0000-0003-4616-621X

Wos Researher ID: O-3133-2018

e-mail: [nii\\_energomash@mail.ru](mailto:nii_energomash@mail.ru)

<sup>5</sup>ORCID iD: 0000-0002-9168-3008

Wos Researher ID: O-2103-2018

e-mail: [nii\\_energomash@mail.ru](mailto:nii_energomash@mail.ru)

**Abstract** – The principal component method is used to analyze the environmental monitoring data of the river Don. The method made it possible to estimate a large number of heterogeneous and rapidly changing parameters. The results indicate the possibility of automatic classification of the water body state on the basis of observation of its chemical, biological and physical parameters. The proposed classification model can be applied to the processing of monitoring data of various ecosystems, including Tsimlyansk reservoir and water-cooler of Rostov NPP. Water samples are taken at stationary points and sections of the hydrochemical network for monitoring surface land waters.

**Keywords:** principal components method, eigenvalue spectrum, a matrix of accounts, the matrix of loadings, extracellular esterases activity, alkaline phosphatase of zeston.