

# ПРОГНОЗИРОВАНИЕ РАСПРЕДЕЛЕНИЯ ЗООПЛАНКТОНА В ЮЖНОМ ОКЕАНЕ С ПОМОЩЬЮ МАШИННОГО ОБУЧЕНИЯ

## PREDICTION OF ZOOPLANKTON DISTRIBUTION IN SOUTHERN OCEAN USING MACHINE LEARNING

**D. Krivoguz  
S. Malko  
A. Semenova**

*Summary.* The article discusses applying possibilities of machine learning to predict zooplankton concentrations in the Southern Ocean, depending on various environmental factors. Since zooplankton is an important part of the trophic chains of aquatic ecosystems, the role of predicting its quantity and distribution becomes more significant. Zooplankton of the Southern Ocean is represented by four large groups — copepods, euphausiids, salps and pteropods. They role in ecosystem of the Southern Ocean is to transfer energy from phytoplankton to the fishes. To reach our goals we analyzed modern approaches in zooplanktons modeling and identified 3 main groups of them — ecosystem, biogeochemical and size-based models. On opposite side, machine learning methods have a number of significant advantages, allowing you to make a prediction with a rather limited variety of ecosystem data. In this research we used data obtained from long-term monitoring mission of zooplankton in the Southern Ocean that was divided into test and training sets by 7 to 3 ratio. The spatial distribution of samples is mainly related to the Pacific sector of the Southern Ocean. Basic descriptive statistical analysis showed that high concentrations of zooplankton are common only in some areas of the Southern Ocean. Correlation analysis revealed no connection between analyzed environmental factors and the concentration of zooplankton, which indicates either the dependence of these concentrations on other factors not used in the study, or on their combined effect. To compare the possibilities of predicting zooplankton concentrations by machine learning, we chose 4 algorithms (k-nearest neighbors, random forest, AdaBoost, and artificial neural network) that were sequentially trained by the training data and then applied on test data. As a result, Random forest and AdaBoost algorithms showed the highest rating with 100% accuracy results. The worst result of prediction was shown by artificial neural networks with an accuracy of 86%.

*Keywords:* zooplankton, machine learning, Southern Ocean, modeling, statistical analysis.

**Кривогуз Денис Олегович**

*К.г.н., г.с., Азово-Черноморский филиал ФГБНУ  
«Всероссийский научно-исследовательский институт  
рыбного хозяйства и океанографии» («АзНИИРХ»),  
г. Керчь*

*krivoguz\_d\_o@azniirkh.ru*

**Малько Сергей Владимирович**

*К.б.н., доцент, ФГБОУ ВО «Керченский  
государственный морской технологический  
университет», г. Керчь*

**Семенова Анна Юрьевна**

*К.э.н., доцент, ФГБОУ ВО «Керченский  
государственный морской технологический  
университет», г. Керчь*

*Аннотация.* В статье рассматриваются возможности применения методов машинного обучения для прогнозирования концентраций зоопланктона в Южном океане в зависимости от различных факторов среды. Поскольку зоопланктон является важным звеном в трофических цепях водных экосистем, роль прогнозирования его количества и распространения обретает большую значимость. Методы машинного обучения обладают рядом существенных преимуществ, позволяя осуществлять прогноз при довольно ограниченном разнообразии данных об исследуемой экосистеме. В работе были использованы данные, полученные в результате многолетнего мониторинга зоопланктона в Южном океане, и были разделены на тестовую и обучающую выборки в пропорции 7:3. Первичный статистический анализ показал, что высокие концентрации распространены лишь в отдельных местах Южного океана. Проведение корреляционного анализа выявило отсутствие взаимосвязи между анализируемыми факторами окружающей среды и концентрацией зоопланктона, что говорит или о зависимости этих концентраций от других факторов, не используемых в исследовании, либо о комбинированном их воздействии. Для сравнения возможностей прогнозирования концентраций зоопланктона с помощью методов машинного обучения были выбраны 4 алгоритма, которые последовательно обучались на тренировочной выборке, а затем проходили проверку на тестовых данных. В результате наибольшую точность показали алгоритмы Random forest и AdaBoost с результатами точности в 100%. Худший результат прогнозирования показали искусственные нейронные сети с точностью в 86%.

*Ключевые слова:* зоопланктон, машинное обучение, Южный океан, моделирование, статистический анализ.

## Введение

**К** зоопланктону относится довольно широкая группа морских организмов, которая в трофических цепях занимает промежуточное место между фитопланктоном и рыбами. Он является важной составляющей как питательного цикла, так и круговорота углерода, передавая огромное количество энергии находящимся выше него в трофических цепях рыбам. Отсюда вытекает необходимость в разработке адекватных и современных методов прогнозирования его пространственно-временной динамики. К сожалению, на данный момент, используемые модели далеки от совершенства, что обуславливает необходимость рассмотрения современных средств моделирования для изучения динамики численности популяций зоопланктона. На данный момент в исследовательской практике распространено использование имитационного моделирования. Примерами таких работ могут служить исследования В.Ф. Зайцева [8], В.И. Лазаревой [9] и Н.В. Шадрина [11]. Также существует ряд исследований, использующих статистический анализ, к примеру работы С.Э. Болотова [7] и А.В. Русакова [10]. Несмотря на огромную работу, проведенную вышеупомянутыми авторами, представленные ими модели или требовательны к входящим параметрам, описывающим биогеоценозы или применимы касательно только одной экосистемы и при переносе самой модели в другие условия велика вероятность увеличения уровня прогностических ошибок.

В целом, все существующие способы и модели, описывающих жизненный цикл зоопланктона в различных экосистемах, можно разделить на 3 группы (Рис. 1) — биогеохимические, размерные и экосистемные модели [1].

На текущем этапе развития научной мысли применение алгоритмов машинного обучения занимает все большую значимость в решении прогностических задач разного рода направленности. Их неоспоримым преимуществом перед остальными способами прогнозирования и моделирования сложных и динамических систем является применение метода «черного ящика», что заключается в отсутствии необходимости использования каких-либо знаний о внутреннем устройстве и аспектах функционирования исследуемого объекта, получение которых в достаточном объеме бывает достаточно сложно. Но, в противовес своим неоспоримым достоинствам, у данного подхода есть сильная зависимость к объему используемых для тренировки и обучения модели данных. Поэтому цель данной работы заключалась в исследовании возможностей прогнозирования численности популяции зоопланктона в Южном океане на основе данных многолетнего мониторинга.

## Характеристика района исследований

Южный океан — признанный официально, но фактически не ратифицированный, 4-й по площади океан, содержащий около 20 млн. км<sup>2</sup> воды. Шельфовая зона Южного океана погружена достаточно глубоко — приблизительная глубина составляет от 400 до 800 м. Самое глубокое место — Южно-Сандвичев желоб. Его условными границами являются зона антарктической конвергенции, широта, южнее мыса Горн, граница плавучих льдов, зона конвергенции южнее 60° Ю.Ш. Расположение донных отложений, мощность которых здесь составляет от 150 до 2000 м — концентрическое, вокруг Антарктиды. В литологическом отношении они представлены айсберговыми глинистыми и глинисто-песчаными, кремнистыми диатомовыми илами, фораминиферовыми илами, смешанными кремнисто-известковыми и вулканогенными породами [5].

Средняя температура воды Южного океана колеблется в районе от -2 до + 10 °С. Температура воздуха в январе у побережья не превышает 0 °С, в то время как в районе морей Уэддела и Росса в среднем её значения составляют около -6 °С. По мере удаления от материковой зоны, в индийском и атлантическом секторах, температура воздуха возрастает до 7–12 °С.

Несмотря на суровые климатические условия и изолированность экосистемы, биота Южного океана представлена довольно широко. Основную биомассу составляют фито- и зоопланктон, криль, губки и различные представители иглокожих [6]. Также из важных представителей фауны здесь присутствует несколько семейств рыб, в особенности нототении. Также широко распространены китовые (синий кит, финвал, сейвал, горбач и др.) и тюлени (тюлень Уэдделла, тюлень-крабод, морской леопард, морской котик). Фитобентос представлен красными и бурыми водорослями.

Зоопланктон Южного океана обладает большим разнообразием планктонных беспозвоночных, отличающихся по форме тела и образу жизни. Весь зоопланктон в Южном океане можно разделить на четыре больших группы в зависимости от размера тела — от крошечных (<100 мкм) веслоногих ракообразных до гигантских (более 1 м) цепей медуз или сальпы. Их жизненный цикл может варьироваться от нескольких дней до нескольких лет. Как правило, желатиновые или желеобразные организмы, которые фильтруют или устраивают засаду добыче, имеют более энергоэффективный образ жизни, чем быстро движущиеся ракообразные, и, следовательно, имеют тенденцию иметь более высокие темпы популяционного роста для данного размера [7].

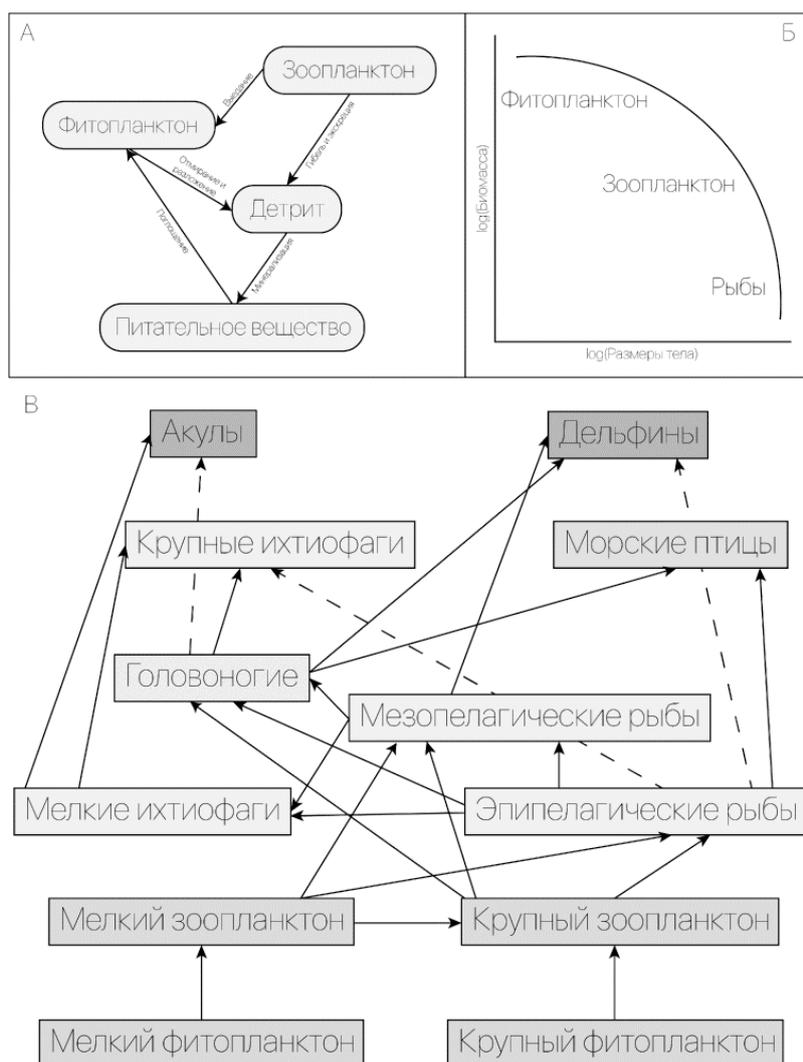


Рис. 1. Схематическое представление основных групп моделей для исследования зоопланктона (А – Биогеохимические, Б — Размерные, В — Экосистемные)

## Материалы и методы

### Данные наблюдений

Используемый в работе набор данных является частью исследовательских материалов, полученных в результате исследований в рамках проекта Continuous Plankton Recorder в Южном океане. Временной период, охватываемый данными мониторинга используемых данных — 2008–2010 гг. Места отбора проб представлены на карте (Рис. 2).

### Используемые алгоритмы

Для сравнения точности прогнозирования распределения запасов криля в зависимости от различных факторов окружающей среды были выбраны 4 метода: Метод

k-ближайших соседей, алгоритм случайного леса, искусственная нейронная сеть и алгоритм AdaBoost [3, 4].

## Результаты и обсуждение

Статистическая вариативность анализируемых количественных факторов довольно различна (Табл. 1). Наименьшую вариативность имеет показатель солёности воды, который находится в пределах от 32.32 до 35.63‰. Температура воды, также, как и показатель флуоресцентности, более вариативны, меняя свои показатели в диапазоне от 0 до 20 единиц.

Для анализа взаимосвязи исследуемых факторов с концентрацией зоопланктона был проведен корреляционный анализ (Табл. 2). Его сущность заключалась в последовательном попарном сравнении взаимосвязи

Таблица 1. Статистические характеристики исследуемых количественных факторов

Статистический показатель	Флуоресцентность, пр.е.	Соленость,‰	Температура, °С	Фотосинтетически активная радиация, мкЭ·м <sup>-2</sup> ·с <sup>-2</sup>
Минимум	0.000	32.32	0.000	0.00
1-й квартиль	0.520	33.74	1.930	2.37
Медиана	1.020	33.83	3.660	18.63
Среднее	2.449	33.85	4.445	232.85
3-й квартиль	3.390	33.91	6.320	325.03
Максимум	21.430	35.63	19.100	2247.93

Таблица 2. Результаты корреляционного анализа взаимосвязи исследуемых факторов с концентрациями зоопланктона

Фактор	Показатель коэффициента корреляции
Флуоресцентность	0.04881911
Соленость	-0.07733507
Температура	-0.18477536
Фотосинтетически активная радиация	-0.17130239

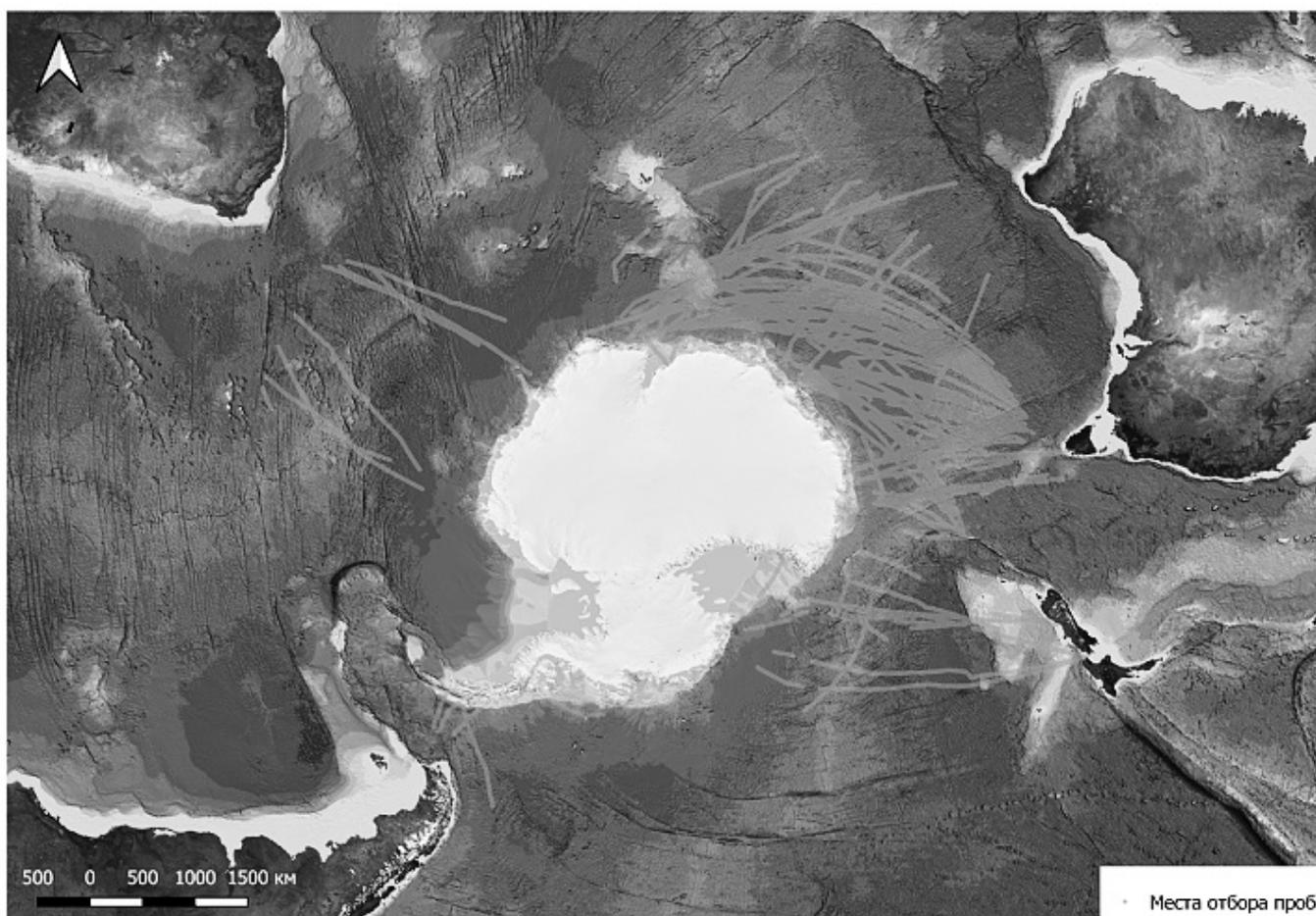


Рис. 2. Места отбора проб, используемые в анализе

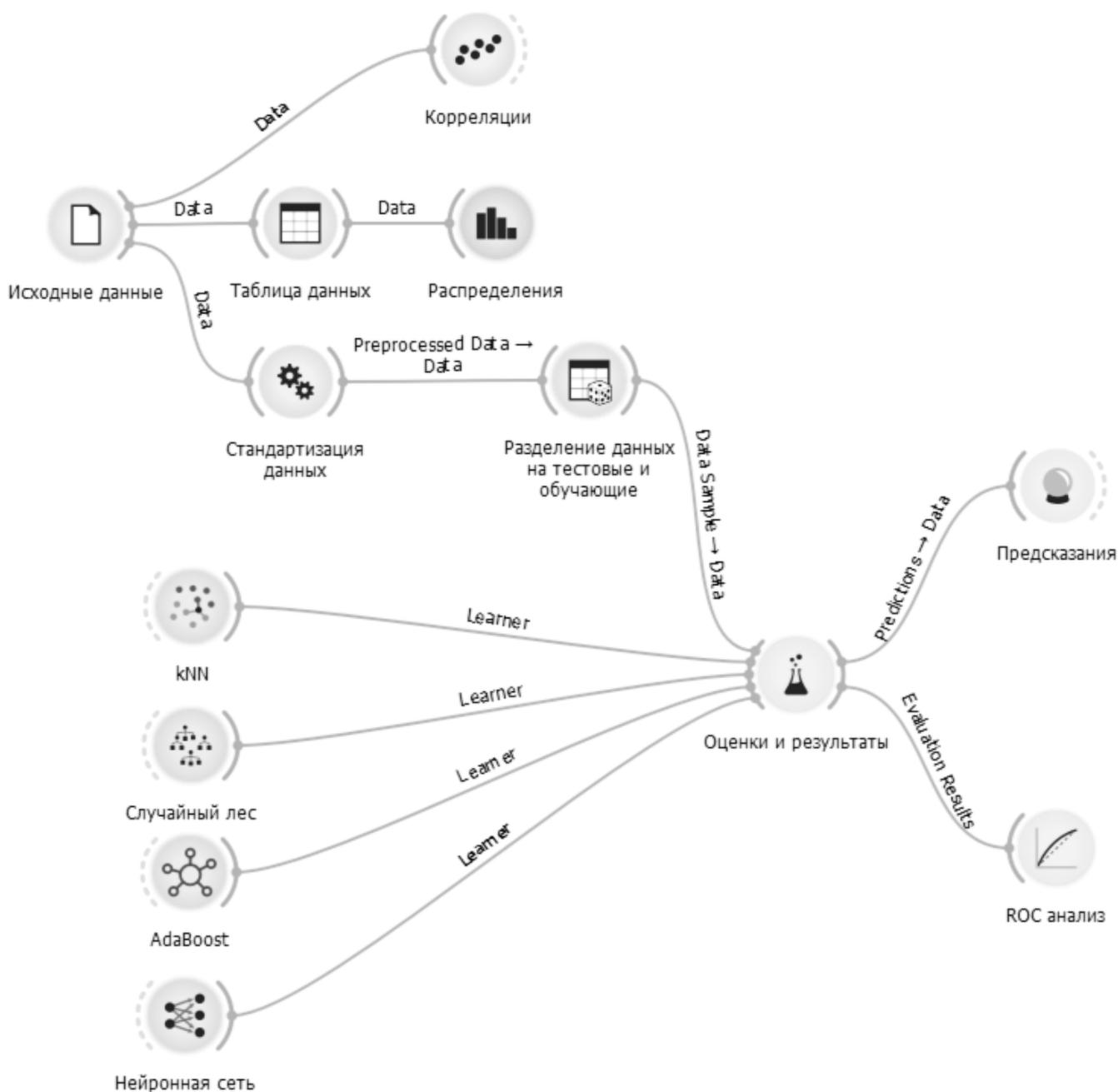


Рис. 3. Схема проведения анализа в Data mining среде Orange

показателя отдельного взятого фактора с количеством зоопланктона в отдельно взятой пробе.

Согласно шкале Чеддока, которая показывает качественную характеристику силы связи между анализируемыми факторами, все рассчитанные коэффициенты корреляции показывают слабую взаимосвязь. Это говорит об отсутствии какого-либо заметного влияния исследуемых факторов на количество зоопланктона. С другой стороны, это может говорить о существовании или более

глубокой и сложной факторной связи, когда проявляется комбинационное воздействие факторов на исследуемое явление или о существовании иного фактора, который находится в более тесном взаимодействии с явлением и количественное или качественное изменение которого приведет к изменениям в количестве зоопланктона.

Таким образом, с одной стороны, все это говорит об отсутствии какого-либо одного фактора, который бы мог объяснить высокую концентрации зоопланктона

Таблица 3. Результаты точности предсказания исследуемых алгоритмов

Метод	AUC	Accuracy	F <sub>1</sub>	Precision	Recall
kNN	1.000	0.998	0.998	0.998	0.998
Случайный лес	1.000	1.000	1.000	1.000	1.000
Искусственная нейронная сеть	0.861	0.855	0.833	0.832	0.855
AdaBoost	1.000	1.000	1.000	1.000	1.000

в отдельных местах. С другой стороны, проведенный анализ позволяет выделить оптимальные условия, в которых экосистема Южного океана имеет наибольшую продуктивность [2].

Для анализа возможностей прогнозирования распределения зоопланктона в Южном океане были выбраны 4 наиболее популярных и часто используемых алгоритма машинного обучения со следующими настройками:

- ◆ kNN: количество соседей — 5, метрика — «евклидова», вес — «унифицированный»;
- ◆ Random forest: количество деревьев — 10, максимальное количество подмножеств при которых происходит разделение — 5, лимит глубины отдельных деревьев — 3;
- ◆ ANN: количество нейронов в скрытых слоях — «100–100–100», активация — «логистическая», оптимизационный алгоритм решателя — «Adam», регуляризация — 0,0001, максимальное количество итераций — 100.
- ◆ AdaBoost: основной классификатор — «дерево», количество классификаторов — 50, скорость обучения — 1, алгоритм классификации — «SAMME.R», функция потерь — «линейная».

Общая схема проведения анализа представлена на рисунке 3.

Так как порядки анализируемых данных существенно отличались, были проведены стандартизация и нормализация, для приведения их к одной размерности. Для стандартизации и нормализации были использованы следующие параметры:  $\mu = 0$ ,  $\sigma^2 = 1$ .

Затем данные делились на две выборки — обучающую и тестовую (70% и 30% соответственно). Таким образом в обучающую выборку попало 27452 проб, а в тестовую — 11764.

Результаты анализа приведены в таблице 3.

Наибольший уровень обучаемости и предсказания показали алгоритмы Random forest и AdaBoost (табл. 4).

Уровень их прогноза концентрации зоопланктона в зависимости от анализируемых факторов окружающей среды чрезвычайно высок и достигает 100%. Чуть меньший уровень прогноза показывает алгоритм kNN.

Наименьший уровень из всех алгоритмов прогнозирования показали искусственные нейронные сети. Процент правильно предсказанных концентраций у них составил около 85%, что тоже является довольно хорошим показателем.

### Заключение

Таким образом, можно сделать следующие выводы:

1. Современные методы прогнозирования и анализа водных сообществ хоть и широко распространены, но имеют множество значительных недостатков, усложняя процесс изучения биоценозов. В тоже время современные подходы, основанные на методах машинного обучения, хорошо проявляют себя даже в условиях ограниченности исходных данных и не требуют построения сложных математических моделей.
2. Распределение зоопланктона по территории Южного океана довольно широко, но зоны его высокой концентрации локализованы лишь в отдельных местах, учитывая не однородность и изменчивость параметров окружающей среды, это говорит о существовании между ними зависимости.
3. Корреляционный анализ исследуемых факторов показал отсутствие прямой зависимости, а следовательно, и их влияния на концентрации зоопланктона по отдельности. Это может говорить как о комбинированном воздействии этих факторов на биомассу зоопланктона, так и о существовании неких факторов, не попавших в исследуемые, но в более значительной мере имеющие с ней прямую зависимость.
4. Анализируемые алгоритмы машинного обучения в целом показали хорошие результаты в предсказании концентраций биомассы зоопланктона по показателям факторов окружающей среды.

Наибольшими прогностическими возможностями обладают алгоритмы Random forest и AdaBoost, предсказывающие концентрации зоопланктона

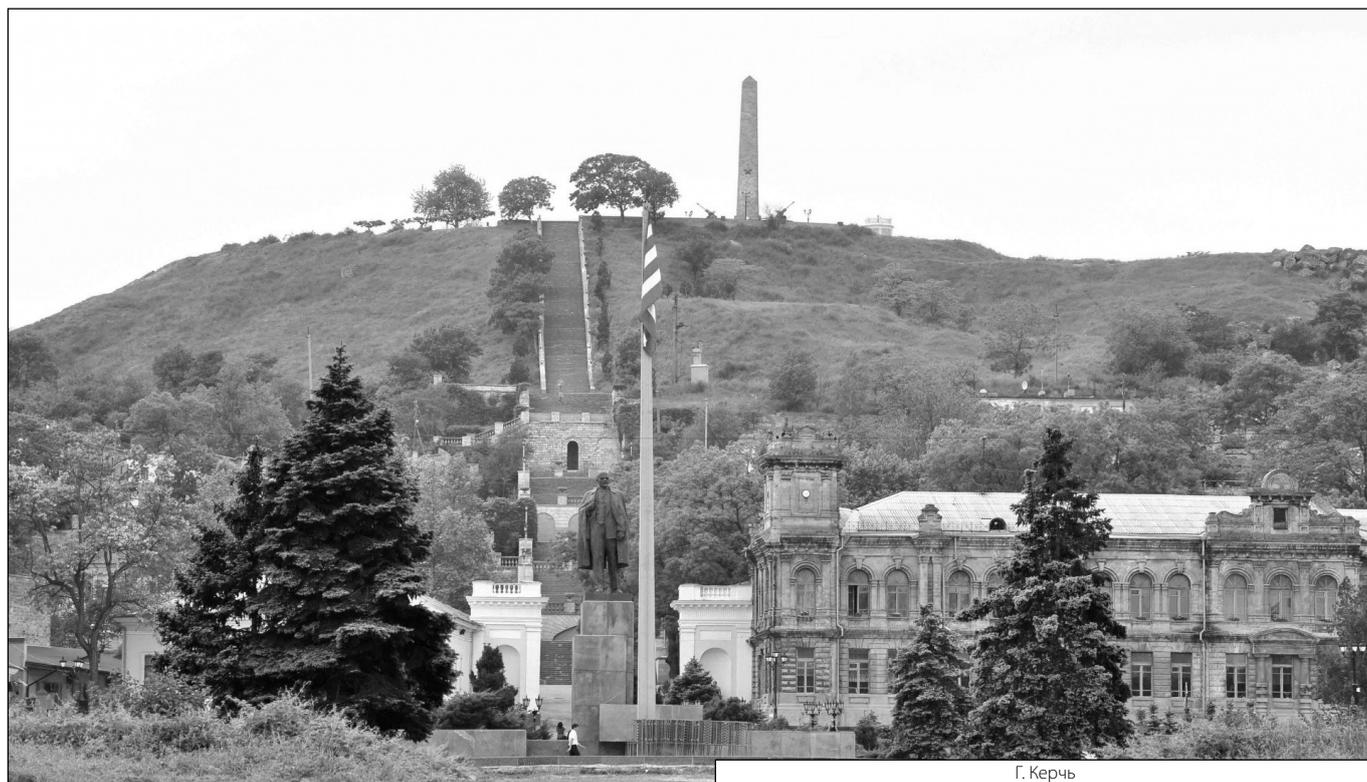
со 100% точностью. Худший результат показали искусственные нейронные сети, точность прогноза которых составила чуть более 86%.

## ЛИТЕРАТУРА

1. Everett J.D. [и др.]. Modeling What We Sample and Sampling What We Model: Challenges for Zooplankton Model Assessment // *Frontiers in Marine Science*. 2017. (4).
2. Krivoguz D. Methodology of physiography zoning using machine learning: A case study of the Black Sea // *Russian Journal of Earth Sciences*. 2020. № 1 (20). С. 1–10.
3. Krivoguz D., Bespalova L. Landslide susceptibility analysis for the Kerch Peninsula using weights of evidence approach and GIS // *Russian Journal of Earth Sciences*. 2020. № 1 (20). С. 1–12.
4. Krivoguz D.O., Borovskaya R.V. Methodological aspects of natural territories zoning using machine learning // *Monitoring systems of environment*. 2020. № 1. С. 13–20.
5. Pollard R.T. [и др.]. Zooplankton distribution and behaviour in the Southern Ocean from surveys with a towed Optical Plankton Counter. 2002.
6. Schnack-Schiel S.B., Isla E., Isla E. The role of zooplankton in the pelagic-benthic coupling of the Southern Ocean // *Scientia Marina*. 2005. № S2 (69). С. 39–55.
7. Болотов С.Э. Оценка и моделирование динамики экологической структуры пойменных биоценозов планктонных животных // *Динамика окружающей среды и глобальные изменения климата*. 2008. № 15 (1). С. 167–175.
8. Зайцев В.Ф. [и др.]. Моделирование динамики популяции некоторых видов гидробионтов Каспийского моря в связи с предстоящим освоением нефтегазовых месторождений // *Экология, рыбное хозяйство*. 2006. № 35 (6). С. 223–229.
9. Лазарева В.И., Минеева Н.М., Жданова С.М. Пространственное распределение планктона в водохранилищах верхней и средней Волги в годы с различными термическими условиями // *Поволжский экологический журнал*. 2012. (4). С. 394–407.
10. Русаков А.В. [и др.]. Математическая модель озерного сообщества с учетом целочисленности размера популяции: хаотические и долгопериодные колебания. 2016.
11. Шадрин Н.В., Ануфриева Е.В. Интегральные показатели изменчивости *Arctodiaptomus salinus* (Daday, 1885) (Copepoda, Diaptomidae) и возможности их использования для оценки состояния популяций // *Биология внутренних вод*. 2018. (4). С. 54–63.

© Кривогуз Денис Олегович (krivoguz\_d\_o@azniir.kh.ru), Малько Сергей Владимирович, Семенова Анна Юрьевна.

Журнал «Современная наука: актуальные проблемы теории и практики»



Г. Керчь