

**PREDICTIVE PERFORMANCE OF LINEAR REGRESSION MODELS  
IN ESTIMATION OF ARTEMIA SALINA ABUNDANCE USING FIELD  
AND REMOTE SENSING DATA**

**D. Krivoguz, R. Borovskaya**

Azov-Black Sea Branch of “VNIRO” (“AzNIIRKH”),  
RF, Rostov-on-Don, Beregovaya St., 21v.  
E-mail: [krivoguz\\_d\\_o@azniirkh.ru](mailto:krivoguz_d_o@azniirkh.ru)

The main goal of this research involved finding the possibility of using the regression models, as a part of the machine learning methods, for the purposes of identification the spatial patterns of *Artemia salina* distribution in the Southern Sivash. Development of such models facilitates the assessment of *A. salina* abundance in water bodies and makes it possible to achieve a higher level of accuracy. In order to identify the maximum absorption levels in different parts of the light spectrum, spectral signatures at each of the monitoring stations have been taken and compared with the satellite data. Astaxanthin, hemoglobin and their absorption spectrums have been also analyzed, with the use of a spectrophotometer. It has been concluded that the spatial data provided by Sentinel-2 satellite show a good potential to become a crucial data source, which will help increase the frequency of monitoring of *A. salina* in the Southern Sivash. It should also be noted that small-sized datasets do not pose an issue, and good accuracy results, appropriate for future investigations in this field, can be obtained by means of simple machine learning algorithms.

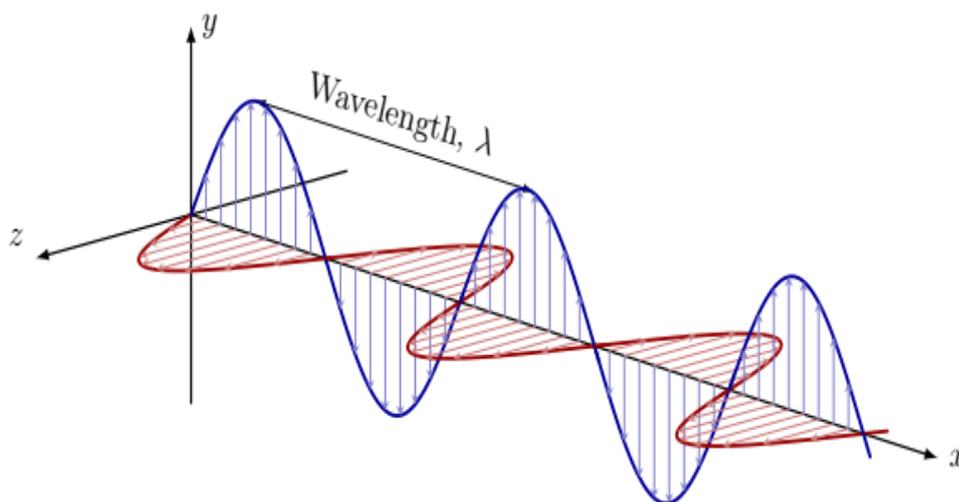
**Keywords:** Remote sensing, spectral analysis, *Artemia salina*, Sivash, machine learning.

Received in the editor: 24.03.2021. After revision: 01.06.2021.

**Introduction.** One of the most valuable methods of studying the Earth surface and objects located on it is remote sensing [1, 2]. It is based on the investigation of the spectral reflecting properties of certain areas of the Earth surface. The measurement of solar radiation in each of them makes it possible to determine the properties and characteristics of the investigated objects [3].

Remote methods used for the assessment of the objects on the Earth surface are based on the analysis of the results of measuring the electromagnetic radiation, reflected from the Earth surface and registered at a remote point [4].

The electromagnetic radiation is characterized by two parameters: wavelength ( $\lambda$ ) and frequency ( $\nu$ ) (Fig. 1).



**Fig. 1.** Schematic presentation of electromagnetic waves  
**Рис. 1.** Схематическое представление электромагнитных волн

The connection between these two values can be estimated as

$$\lambda = \frac{C_0}{v} ,$$

which is equal to the light speed.

Properties of the reflected, absorbed and transmitted radiation differ for different objects and depend on the nature of these objects and conditions in which they exist [5]. At the same time, the curve of the spectral reflection represents the plot of the reflectance of the object in different parts of the spectrum. Pigments, contained in living objects, play the major role due to the fact that the maximum of absorption of different pigments lays in different parts of the spectrum [6].

The flow of the solar energy consisting of solar radiation and rays scatters in the atmosphere and, upon hitting the surface of nature objects, undergoes various transformations, which include absorption, transmission and reflection. Depending on the characteristics of a surface, the reflection can be direct or indirect; however, in the real world, we can usually see a combined type of reflection.

The reflection properties of the surface can be measured by evaluating the coefficients of brightness, which are calculated as

$$r = \frac{B_2}{B_1} ,$$

where  $B_1$  is the brightness of a perfect diffuser,  $B_2$  is the brightness of the investigated surface. Spectral coefficients of brightness depend on physicochemical properties of these objects, such as their temperatures, wetness, etc.

The main goal of this research is to find the possible use of regression models, in the context of application of machine learning methods, in the assessment of the spatial patterns of *Artemia salina* distribution in the Southern Sivash. Development of such type of models facilitates the assessment of the abundance of *A. salina* in water bodies and increases its accuracy, which, along with the field research of the Azov-Black Sea Branch of VNIRO, will help increase

the quantity and quality of the collected data, which is instrumental in the matter of fisheries assessment.

In the course of this research, we have used some assumptions:

- due to the fact that Sivash Bay is a shallow water body [7], its water column is fully mixed. Based on the research conducted by Vesnina [8], the averaged concentration of *Artemia* at the depth of 0-2 m is rather uniform, mainly due to the temperature distribution. It means that vertical distribution of *A. salina* in the water of the Sivash Bay is even or has the same dependence;
- in term of mathematical description of the top water layer, the area of *A. salina* distribution exceeds the area of water

$$\begin{cases} S_{total} = S_{Artemia} + S_{water} \\ S_{Artemia} > S_{water} \end{cases} ,$$

which suggests the idea that, during the spatial analysis of remote sensing data, the reflectance of *Artemia* will get to the pixel value instead of water.

The key feature of this study is also the use of a small-sized dataset, resulting from the low number of samples, with simple machine learning algorithm — linear regression, which, in our opinion, will contribute to dealing with a problem of data scarcity that can lead to underfitting or overfitting of the model.

**Materials and Methods.** *Artemia salina*. *Artemia salina* is an organism targeted by the larvae of 70 species of hydrobionts feeding on it [9].

Species of the genus *Artemia* are related to arthropods. At present, there are 7 recorded species of *Artemia* genus, 3 of which are mainly spread in Eastern Europe:

- *Artemia sinica* (Yaneng, 1983)
- *Artemia salina* (Linnaeus, 1758)
- *Artemia urmiana* (Gunther, 1900)

Life cycle of *Artemia* (Fig. 2) is characterized by a high level of adaptation and changeability, ensuring the survival of the species in harsh conditions of hypersaline water bodies [10]. Environmental conditions, unfavorable for *Artemia*, include temperature below zero in winter season, the level of salinity

exceeding 300‰, drying of rivers, lack of oxygen, etc. To survive under these circumstances, *Artemia* transforms into cysts — eggs, covered with thick brown

shell, containing an embryo, which development is stopped at the gastrula stage [11, 12].

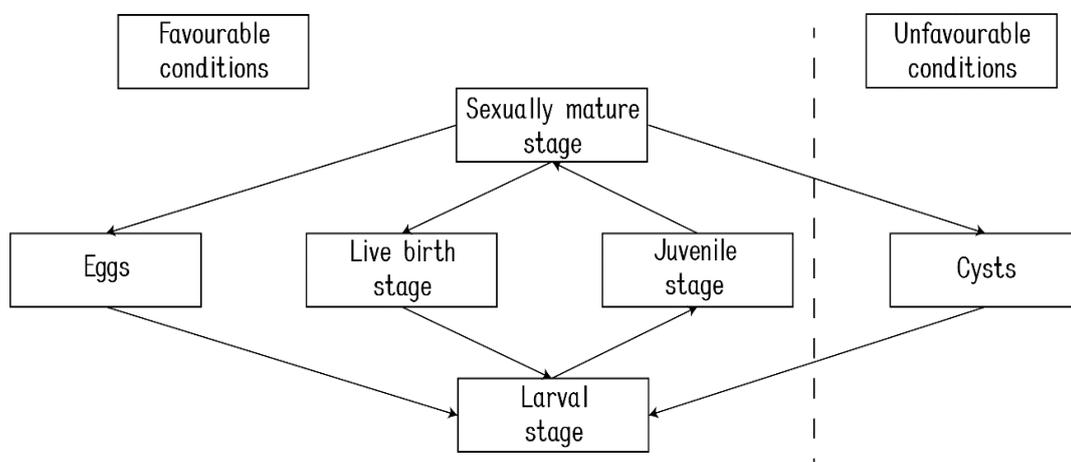


Fig. 2. Life cycle scheme for *Artemia salina* [13]

Рис. 2. Жизненный цикл *Artemia salina* [13]

Dried cysts can withstand extreme freezing and heating (over 60°C). Dehydrated cysts have a higher level of temperature tolerance, so they become viable in the temperature range from -18 to +40°C [14].

**Pigments of *Artemia salina*.** Eggs of *Artemia* vary by their color from cream to dark brown [15]. Generally, it depends on hematin content in the shell of an egg. Nauplii, unlike eggs, have bright orange color, primarily due to esterified astaxanthin. This carotenoid is also present in *Artemia* eggs [16].

Adult *Artemia* specimens do not have carotenoids, unless a small part of astaxanthin ester from previous lifeforms [11, 14]. Pigmentation of adult individuals is contingent on their nutrition. Carotenoids found in such species as *Dunaliella terlioelecta* and *Phacodactylum tricorutum*, are absorbed in *Artemia* organism through their consumption [17]. For *Dunaliella*, these carotenoids are β-carotene, γ-carotene, lutein, violaxanthin and neoxanthin; for *Phacodactylum*, they are β-carotene, diadinoxanthin, fucoxanthin and neoxanthin [18]. Generally, the most widespread carotenoid in *Artemia* organism is astaxanthin [19, 20].

Another common cause of red color occurring in *Artemia* is the presence of hemoglobin [17]. Hemoglobin content in its blood basically depends on the amount of oxygen contained in the water where it lives [16]. Sometimes, the increase in red coloration is caused by the rise of salinity level [21, 22]. Some authors have also identified the presence of hemochromogen in the guts of *Artemia*, haematin in its shell and eggs, and ommochrome pigment in its eyes.

**Astaxanthin, hemoglobin, and their absorbing spectrums.** Astaxanthin or 3,3'-dihydroxy-β,β-carotene-4,4'-dione is a pigment belonging to oxygen-containing carotenoids xanthophylls [15]. The chemical formula for astaxanthin is C<sub>40</sub>H<sub>52</sub>O<sub>4</sub>. In aquatic environment, it can include liposome and polymeric nanoparticles, which consequently increases its stability and dissolubility [17, 21].

Spectral and photochemical properties of astaxanthin are based on positions of the lowest energetic levels. A molecule of astaxanthin absorbs the energy in visible part of the specter due to its transformation from one energetic state into another by transition of π-electrons into vacant antibonding π-orbitals. The maximum of

astaxanthin absorption lies further than in other carotenoids due to its longer system of conjugated connections. In the ultraviolet part of the light spectrum, the absorption of energy is equal to transitions into another, higher energetic levels. Nevertheless, the main feature of astaxanthin is a hemoglobin, a chromoprotein that has a prosthetic group (heme) connected to a protein globin. In the heme of hemoglobin, the first atom of *Fe* is connected by a coordinate covalent bond with histidine residue of globin, and the second one is connected to oxygen. The main function of hemoglobin is oxygen transfer. Usually, a molecule of hemoglobin consists of four polypeptide chains, each of them connected with one group of hemes [21].

**Artemia salina data.** The data on *Artemia salina* in the Southern Sivash have been collected by Semik and Zamyatina [10], who collected it at 10 stations. The location of these stations has also been presented in their research [10]. In the course of their investigation, samples of *Artemia* were collected 3 times at each station. Sampling was conducted with the use of Epstein net, and the volume of filtered water equaled 50 liters. In the field, samples were fixed in 4% formalin solution. After that, they were processed in the laboratory. Estimation of *Artemia* abundance in samples was conducted in Bogorov counting chamber using binocular microscope MBS-10.

Water samples, containing *Artemia*, were filtered using special filters under the pressure, then kept in a refrigerator for about 12 hours till full drying. Then they were placed in centrifuge tubes, where about 8 ml of fresh 90% solution of acetone was added. Their extraction was carried out for 2 hours. After the extraction, the obtained samples were centrifuged for 10 minutes with the speed of 5000 cycles per minute, poured into measuring tubes and analyzed in spectrophotometer SPECORDUV-VIS.

**Linear and polynomial regressions.** Regression analysis is basically a statistical approach used for investigation and modeling of the relationships between

variables [23, 24]. Its application can be found in various research fields, such as engineering, physical or chemical studies, economics and management, biological and Earth sciences, sociological sciences, etc. [25]. Thus, regression analysis is one of the most widespread methods of statistical analysis.

Simple linear regression can be defined as

$$y = \beta_0 + \beta_1 x + \varepsilon_i ,$$

where  $\beta_1$  is a slope, and  $\varepsilon_i$  is the sum of squared residuals.

In machine learning, creation of the regression model is an iterative process, directed at finding the effective independent variables to explain dependent variables, which we tried to model, describe and comprehend in order to define which of these analyzed variables describe a model with a higher accuracy.

Determination coefficient  $R^2$  is usually considered to be the most important value [23, 26] that describes the quality measure of the model and is defined as

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} ,$$

where  $y_i$  is a value of observed variable,  $\bar{y}_i$  is an averaged value of  $y$ , and  $\hat{y}_i$  is a modeled value, built by evaluation parameters.

The main advantages of regression models are:

- simplicity of evaluation algorithms;
- visibility and interpretability of results.

The main disadvantages are the following:

- low level of prediction accuracy due to frequent interpolation of data;
- subjective choice of model type due to fitting of the model to empirical data;
- difficulty in interpretation of causal relationship.

Polynomial regression is one of the regression analysis forms [24, 25], where dependence between the independent value  $x$  and the dependent value  $y$  is modeled like

$$y(x) = a + b \cdot x + c \cdot x^2 + \dots + h \cdot x^n .$$

The case, where  $n = 1$  and this polynomial is a straight line, is called a linear regression. The main difference between polynomial and simple linear regressions is an ability to model nonlinearly separated data, and, due to its higher flexibility, polynomial regression can model more complex interrelations.

The main goal of polynomial regression is to address the case, where  $p$  is a polynomial of the given power, then

$$y_i = \sum_{j=1}^p x_j x_i^{j-1} + \varepsilon_i .$$

**Results and discussion.** For investigation of maximum absorption levels in different parts of the light spectrum, the spectral signatures have been taken for 10 monitoring stations in the Southern Sivash in 2018 [10], which covers 13 spectral channels of Sentinel-2 satellite.

Generally, the maximum of absorption (Fig. 3) lies at 0.56 nm and 0.705 nm, which corresponds to green and red parts of the light spectrum. These data also comply with the maximum of astaxanthin and hemoglobin absorption.

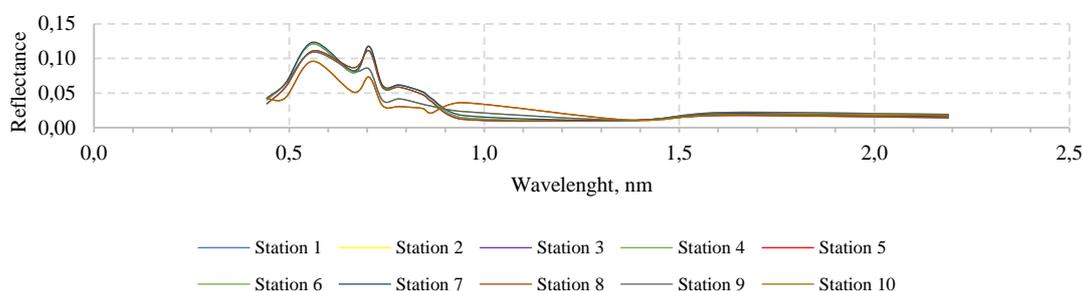


Fig. 3. Spectral curves of reflectance for 10 stations in the Southern Sivash

Рис. 3. Спектральные кривые по 10 станциям Южного Сиваша

The differences and similarities between the spectral signatures at various stations should be noted, in particular at stations 4 and 5, which, in our opinion, are explained by the absence of *Artemia* in the samples collected at these stations. Fig. 3 allows to see pairing of closely-spaced signatures, collected close to each other due to the same absorption properties of the object and environment.

By application of regression analysis, the equations describing the dependencies between spectral reflectance and abundance of *Artemia salina* for 2 bands of Sentinel-2 satellite — bands 3 and 5, — have been solved. To achieve this, the fitting of linear regression with polynomials of degree 3 and 4 (Fig. 4) was used, which made it possible to reach 79 and 87 percent of accuracy for each of these channels.

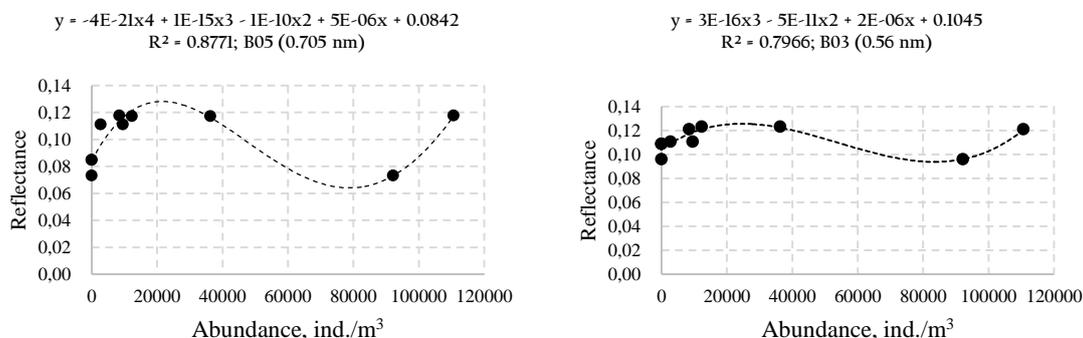
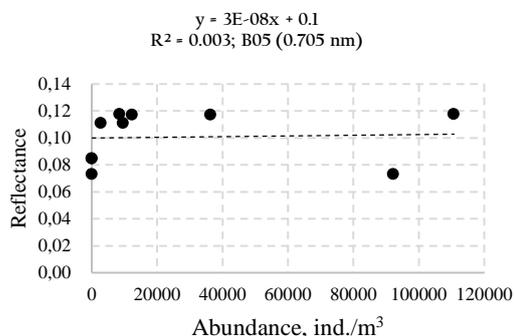


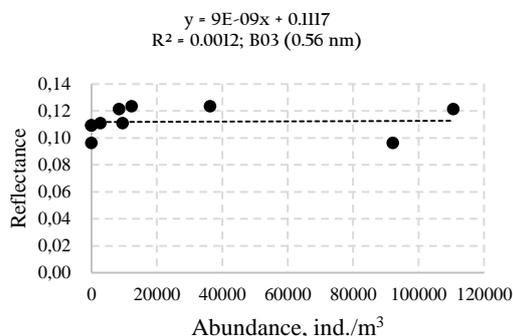
Fig. 4. Results of fitting the linear regression models using the polynomials of degree 4 and 3 for channels 3 and 5 of Sentinel-2 satellite

Рис. 4. Результаты подгонки линейно-регрессионной модели полиномами 4-й и 3-й степени для 3-го и 5-го каналов спутника Sentinel-2

Then, fitting by simple linear regression (by polynomial of degree 1) was used; however, the achieved results of the model accuracy were insufficient (Fig. 5). The



created models poorly describe dependencies between analyzed data, and their determination coefficients are close to zero (0.0012 and 0.003, correspondingly).



**Fig. 5.** Results of fitting the linear regression models using a polynomial of degree 1 for channels 3 and 5 of Sentinel-2 satellite

**Рис. 5.** Результаты подгонки линейно-регрессионной модели полиномом 1-й степени для 3-го и 5-го каналов спутника Sentinel-2

In future, these regression models, applied to corresponding spectral channels of remote sensing data, can provide a means of facilitating the spatial analysis of the data on *Artemia salina*, collected in the Southern Sivash, and allow to increase the frequency of environmental monitoring in this area.

Based on the results of this investigation, it seems to be the most optimal to use the spectral channel B05 of the Sentinel-2 satellite with wavelength of 0.705 nm for monitoring the abundance of *Artemia salina* in this area, due to the best level of accuracy achieved by the model for this channel.

**Conclusion.** As a result of this research, several regression models have been tested, which will facilitate environmental spatial monitoring of *Artemia salina* abundance in the Southern Sivash. Spatial data provided by Sentinel-2 satellite show a good potential to become crucial in the matter of increasing the frequency of monitoring *Artemia salina* in the Southern Sivash.

This research also gives evidence that small-sized datasets do not pose a serious problem, and simple machine learning algorithms can achieve good accuracy results, appropriate for future investigations in this field.

Thus, the results presented in this paper show sufficient potential for being applied

in the course of environmental monitoring of *Artemia salina* in the Southern Sivash. It should also be noted that the continuation of monitoring in this area will increase the total accuracy of the model and provide us with a better result.

## REFERENCES

1. Buongiorno Nardelli B. et al. A re-analysis of Black Sea surface temperature // *J. Mar. Syst.* 2010. Vol. 79, № 1–2. P. 50–64.
2. Krivoguz D., Bepalova L. Landslide susceptibility analysis for the Kerch Peninsula using weights of evidence approach and GIS // *Russ. J. Earth Sci.* 2020. Vol. 20, № 1. P. 1–12.
3. Krivoguz D. Methodology of physiography zoning using machine learning: A case study of the Black Sea // *Russ. J. Earth Sci.* 2020. Vol. 20, № 1. P. 1–10.
4. Demchev D. et al. Sea Ice Drift Tracking From Sequential SAR Images Using Accelerated-KAZE Features // *IEEE Trans. Geosci. Remote Sens.* 2017. Vol. 55, № 9. P. 5174–5184.
5. McFeeters S.K. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features // *Int. J. Remote Sens.* 1996. Vol. 17, № 7. P. 1425–1432.

6. Hung Trinh L., Tuyen Vu D. Application of remote sensing technique for drought assessment based on normalized difference drought index, a case study of Bac Binh district, Binh Thuan province (Vietnam) // *Russ. J. Earth Sci.* 2019. Vol. 19. P. ES2003.
7. Sovga E.E., Eryemina E.S., Khmara T. V. Water Balance in the Sivash Bay as a Result of Variability of the Natural-Climatic and Anthropogenic Factors // *Phys. Oceanogr.* 2018. Vol. 25, № 1. P. 67–76.
8. Vesnina L.V., Permyakova G.V. Dynamics of number and distribution of uneven-age individuals of *Artemia* in deep-water Bolshoe Yarove lake (Altaysky Kray) // *Tomsk State Univ. J. Biol.* 2013. Vol. 1. P. 89–102.
9. Shadrin N. V., Anufrieva E. V., Shadrina S.N. Brief review of phototrophs in the Crimean hypersaline lakes and lagoons: diversity, ecological role, the possibility of using // *Mar. Biol. J.* 2017. Vol. 2, № 2. P. 80–85.
10. Semik A.M., Saenko E.M., Zamyatina E.A. Current status of the Brine shrimp population *Artemia* Leach, 1819 in the Eastern Sivash Bay // *Aquat. Bioresour. Environ. Aquatic Bioresources & Environment, FSBSI VNIRO, Azov-Black Sea Branch of the FSBSI VNIRO (AzNIIRKH).* 2019. Vol. 2, № 2. P. 45–56.
11. Gajardo G.M., Beardmore J.A. The Brine Shrimp *Artemia*: Adapted to Critical Life Conditions // *Front. Physiol. Frontiers.* 2012. Vol. 3. P. 185.
12. Shaala N.M.A. et al. Lethal Concentration 50 (LC50) and Effects of Diuron on Morphology of Brine Shrimp *Artemia Salina* (Branchiopoda: Anostraca) Nauplii // *Procedia Environ. Sci. Elsevier BV.* 2015. Vol. 30. P. 279–284.
13. Litvinenko L.I. et al. Methodical recommendations for stock assessment and prediction of recommendational amount of catching *Artemia*. Moskow. 2019. 50 p.
14. Anufrieva E., Shadrin N. The long-term changes in plankton composition: Is Bay Sivash transforming back into one of the world's largest habitats of *Artemia* sp. (Crustacea, Anostraca)? // *Aquac. Res.* Blackwell Publishing Ltd. 2020. Vol. 51, № 1. P. 341–350.
15. Gilchrist B.M., Green J. The pigment of *Artemia*. // *Proc. R. Soc. Lond. B. Biol. Sci. The Royal Society London.* 1960. Vol. 152. P. 118–136.
16. Boonyaratpalin M. et al. Effects of  $\beta$ -carotene source, *Dunaliella salina*, and astaxanthin on pigmentation, growth, survival and health of *Penaeus monodon* // *Aquac. Res.* Blackwell Publishing Ltd. 2001. Vol. 32, № SUPPL. 1. P. 182–190.
17. Czygan F.C. On the Metabolism of Carotenoids in the Crustacean *Artemia salina* // *Zeitschrift für Naturforsch. - Sect. B J. Chem. Sci. Verlag der Zeitschrift für Naturforschung.* 1968. Vol. 23, № 10. P. 1367–1368.
18. Amarouayache M., Kara M.H. Aspects of life history of *Artemia salina* (Crustacea, Branchiopoda) from Algeria reared in different conditions of salinity // *Vie milieu - Life Environ.* 2017. Vol. 67, № 1. P. 15–20.
19. Shadrin N., Yakovenko V., Anufrieva E. Suppression of *Artemia* spp. (Crustacea, Anostraca) populations by predators in the Crimean hypersaline lakes: A review of the evidence // *Int. Rev. Hydrobiol. Wiley-VCH Verlag.* 2019. Vol. 104, № 1–2. P. 5–13.
20. Ha N.-T. et al. Detecting Multi-Decadal Changes in Seagrass Cover in Tauranga Harbour, New Zealand, Using Landsat Imagery and Boosting Ensemble Classification Techniques // *ISPRS Int. J. Geo-Information. Multidisciplinary Digital Publishing Institute.* 2021. Vol. 10, № 6. P. 371.
21. Vandenberg C.J., Matthews C.M., Trotman C.N.A. Variant Subunit Specificity in the Quaternary Structure of *Artemia* Hemoglobin // *Mol. Biol. Evol. Society for Molecular Biology and Evolution.* 2002. Vol. 19, № 8. P. 1288–1291.
22. Clegg J.S., Trotman C.N.A. Physiological and Biochemical Aspects of *Artemia* Ecology // *Artemia: Basic and Applied Biology.* Springer Netherlands. 2002. P. 129–170.
23. Uyanik G.K., Güler N. A Study on Multiple Linear Regression Analysis // *Procedia - Soc. Behav. Sci.* 2013. Vol. 106. P. 234–240.

24. Seber G.A., Lee A.J. Linear regression analysis. Wiley-Interscience. 2003. 557 p.

25. Weisberg S. Applied linear regression. Wiley-Interscience. 2005. 310 p.

26. Pesaran M.H., Smith R.J. A Generalized  $R^2$  Criterion for Regression Models Estimated by the Instrumental Variables Method // *Econometrica*. 1994. Vol. 62, № 3. P. 705.

## ОЦЕНКА ПРЕДСКАЗАТЕЛЬНЫХ ВОЗМОЖНОСТЕЙ ПРИМЕНЕНИЯ МОДЕЛЕЙ ЛИНЕЙНОЙ РЕГРЕССИИ ДЛЯ ОПРЕДЕЛЕНИЯ БИОМАССЫ *ARTEMIA SALINA*, ИСПОЛЬЗУЯ ПОЛЕВЫЕ И СПУТНИКОВЫЕ ДАННЫЕ

Д.О. Кривогуз, Р.В. Боровская

Азово-Черноморский филиал ФГБНУ «ВНИРО» («АзНИИРХ»),  
г. Ростов-на Дону, ул. Береговая, 21в

Основная цель данного исследования заключалась в нахождении возможности использования линейных регрессионных моделей как части методов машинного обучения, чтобы показать пространственные закономерности распространения *Artemia salina* в Южном Сиваше. Создание таких моделей позволяет с высокой точностью определять биомассу *A. salina* в водоемах. Для исследования максимальных уровней поглощения в различных частях светового спектра мы сопоставили спектральные сигнатуры на всех станциях мониторинга со спутниковыми данными, а также провели анализ спектров поглощения астаксантина и гемоглобина с помощью спектрофотометра. Мы пришли к выводу, что данные спутника Sentinel-2 показывают хороший потенциал для того, чтобы стать ключевым источником данных, который поможет увеличить частоту мониторинга *Artemia salina* в Южном Сиваше. Также следует отметить, что небольшие наборы данных не являются большой проблемой в данном исследовании, и простые алгоритмы машинного обучения могут дать хорошие результаты точности, подходящей для будущих исследований в этой области.

**Ключевые слова:** дистанционное зондирование, спектральный анализ, *Artemia salina*, Сиваш, машинное обучение.