Analysis of Phytoplankton in the Volga River Using Satellite Monitoring with Sentinel-2 Data

Ivanova Yu.S.[1], Gorbunov M.Yu.[2], Tarasova N.G.[2,3] , Umanskaya M.V[2]. and Fedoseev V.A.[1*]

[1]1 Samara National Research University, Moskovskoe shosse 34, Samara, Russia

[2] Institute of Ecology of Volga Basin RAS, Samara Federal Research Center RAS, ul. Komzina 10, Togliatti, Russia

[3]Papanin Institute for Biology of Inland Waters RAS, Borok, Russia

*vicanfed@gmail.com

Abstract *Phytoplankton is a common type of plankton in the Kuibyshev Reservoir, located in the middle course of the Volga River. Observations of phytoplankton concentration are conducted as part of regular field expeditions. However, due to the large surface area of the water, these measurements do not give us clear understanding of changes in chlorophyll a concentration and phytoplankton biomass over time and space. Meanwhile, during the active reproduction period of phytoplankton, it can be observed visually. Therefore, our study aims to analyze the possibility of using remote sensing data to assess phytoplankton development. For this purpose, data from Sentinel 2 satellites (4 channels with a resolution of 10 meters and 6 channels with a resolution of 20 meters) were used. We used Sentinel-2 data corresponding to the coordinates of the points and the dates of field studies. The total number of points with laboratory-measured indicators and corresponding 10-channel remote sensing data exceeded 80. These data were used to train some regression models including linear regression, Ridge regression, Lasso regression, and other models to estimate chlorophyll a concentration. The results showed that the estimation error decreases as the number of samples increases, and overall, Sentinel-2 data can be used for rough concentration estimates. Due to the insufficiency of laboratory data for training high-precision models, we also conducted one more experiment with manually labeled images to train binary classification models for detecting areas with high plankton concentration. More than 20 satellite images containing the water area of the Zhiguli Sea (the Priplotinnoye Reach of the Kuybyshev Reservoir) were used for training and testing the classifiers. As a result, XGBoost and CatBoost classifiers were trained on the labeled data. Both showed an accuracy of around 0.89 on the test set. The experimental results demonstrated that this approach is an appropriate way to train high-quality classification models, enabling a global analysis of phytoplankton distribution over time and space.*

Keywords: *phytoplankton, remote sensing, Sentinel-2, machine learning, regression*

Introduction

"Blooming" of water refers to any excessive growth of planktonic unicellular, colonial, and filamentous algae, in which they become visually noticeable, affect water management activities, or cause the death of aquatic organisms [1-2]. In freshwater bodies of the temperate climate zone, cyanobacterial blooms occur most frequently [3]. In recent decades, the intensity, frequency, and duration of cyanobacterial blooms in freshwater continental bodies of water have been increasing [4] (see Fig. 1).

Figure 1: Examples of "blooming water".

The Kuybyshev Reservoir, located in the middle course of the Volga River, is the largest in Eurasia. Long-term monitoring of phytoplankton and the assessment of chlorophyll "a" concentration in the reservoir is carried out through a network of permanent stations [5] with varying frequencies, ranging from monthly to one-time or three-time observations between May and October in different years. Recent observations show that the timing of the onset of cyanobacterial blooms and their duration have changed significantly compared to long-term data from 1960-2000 [5-8]. Therefore, to ensure more reliable monitoring, an increase in the number of observations throughout the year is necessary. Another problem with the current situation is that the relatively small number of stations over the vast surface area of the Kuybyshev Reservoir does not allow for a comprehensive and complete understanding of the changes in chlorophyll "a" concentration and phytoplankton biomass over time and space.

The intensity of algal blooms is typically assessed by measuring phytoplankton concentration or the concentration of its main photosynthetic pigment, chlorophyll-a (Chl a). Due to the high spatial and temporal heterogeneity of phytoplankton distribution across reservoirs and large inland water bodies—caused, in part, by wind-driven currents and convectional phenomena in the water column—bloom intensity indicators at individual stations vary greatly. As a result, relying solely on data from a limited network of stations on specific dates makes it difficult to evaluate the integral characteristics of phytoplankton in the reservoir.

To address these challenges, remote monitoring methods can be applied. According to several studies [9-13], using remote sensing methods through satellite imagery can be effective in detecting algal bloom patches in large water bodies, tracking seasonal dynamics, and identifying other features of cyanobacterial mass development in various freshwater bodies [14]. In particular, satellite remote sensing methods can, with a certain degree of error, estimate Chl a concentration, biomass, and, with some limitations, the composition of phytoplankton across the entire water body. Thus, remote sensing methods can serve as an effective tool for early warning of algal blooms and rapid monitoring of bloom conditions in different parts of large water bodies.

The objective of this study was to analyze the potential of using Earth remote sensing data to assess cyanobacterial blooms in the Kuybyshev Reservoir.

Monitoring of water blooms in the Kuibyshev reservoir

Since the creation of the Kuybyshev Reservoir, periods of summer cyanobacterial blooms have been characteristic, but they were relatively short-lived and mostly recorded in July-August [15]. In recent years, the duration of these blooms has increased, and from 2020 to 2023, they were recorded at various stations from mid-June to mid-September. The dominant species during the cyanobacterial bloom in the Priplotinnoye stretch of the reservoir (Zhiguli Sea) and the Usinsky and Cheremshansky bays was, and remains, Aphanizomenon flos-aquae Ralfs ex Bornet & Flahault (order Nostocales), accompanied by Dolichospermum flos-aquae (Bornet & Flahault) P. Wacklin, L. Hoffmann & Komárek (order Nostocales), Microcystis aeruginosa (Kütz.) Kütz. (Chroococcales), as well as some other cyanobacterial species (see Fig. 2).

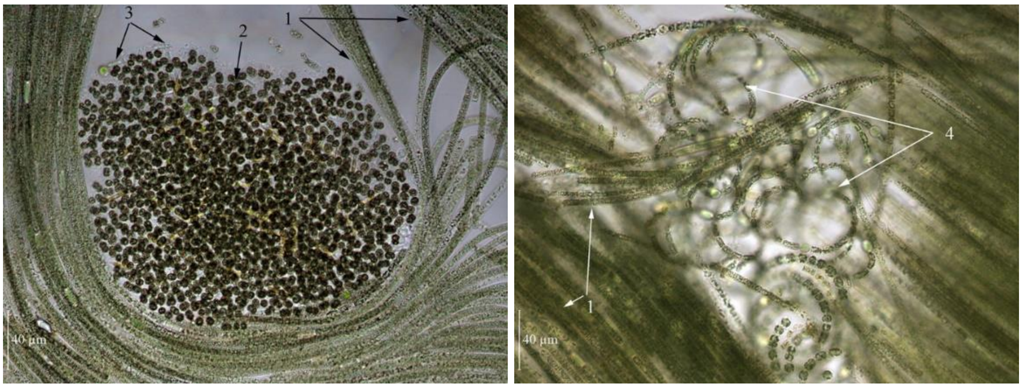

Figure 2: The dominant species of cyanobacteria in the Kuibyshev reservoir in June 2023: 1 – Aphanizomenon flos-aquae Ralfs ex Bornet & Flahault (Nostocales); 2 – Microcystis aeruginosa (Kütz.) Kütz. (Chroococcales); 3 – Pseudanabaena mucicola (Naumann & Huber-Pestalozzi) Schwabe (Pseudanabaenales); 4 – Dolichospermum flos-aquae (Bornet & Flahault) P.Wacklin, L.Hoffmann & Komárek (Nostocales).

Phytoplankton development in the Kuybyshev Reservoir and nearby water bodies is measured at stationary stations, as well as by sampling at individual points. Integrated samples are collected from the surface to the bottom. Simultaneously with sampling, measurements are taken of water depth, water transparency using a Secchi disk (m), specific electrical conductivity (S/cm), and water temperature (°C).

To determine the concentration of photosynthetic pigments, water samples with volumes of 0.2–1 L are filtered through FPSV glass filters (Vladisart, Russia) with a nominal retention threshold of 1.2 µm. The seston collected on the filters is extracted with 90% acetone in the dark at 4°C for 24 hours. Pigment concentrations in acetone extracts are determined according to [16-17] and the spectral reconstruction method.

For morphological identification and quantitative accounting of phytoplankton cells, water samples are fixed with formalin and processed according to standard methodology [18]. Species identification is carried out in accordance with [19].

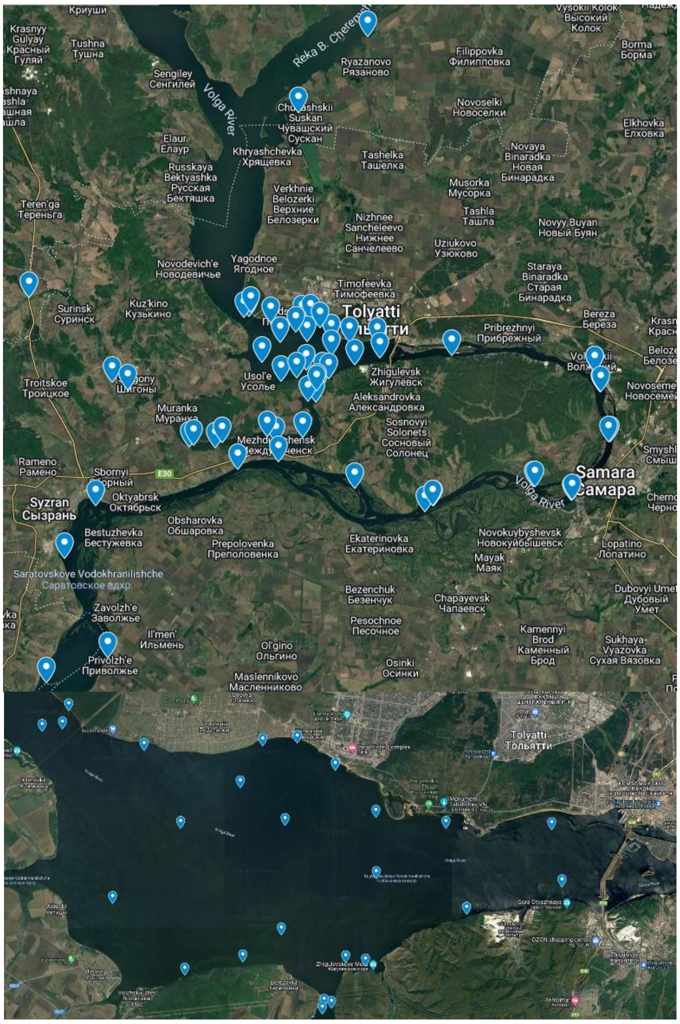

Figure 3: Location of Sampling Points.

To train and assess the quality of methods for detecting and estimating phytoplankton concentrations using Earth remote sensing data, 81 laboratory samples were selected. Samples were collected in September and October 2020, June and July 2021, as well as May, June, July, and September 2023. Figure 3 on the map shows the locations of the sampling points. The points in the Priplotinnoye stretch (Zhiguli Sea) are shown in more detail.

Satellite monitoring for phytoplankton assessment

a. Overview of Existing Studies on Remote Sensing Data for Phytoplankton Analysis:

The potential of remote sensing for assessing phytoplankton development has been actively researched by scientists [9-13]. In particular, several studies have documented the successful use of Sentinel-2 satellite data for monitoring aquatic ecosystems using machine learning methods.

For instance, in the study [9], the authors used Sentinel-2 satellite data to estimate chlorophyll-a concentrations in Chinese water bodies using regression models. They applied a linear regression model, support vector machines (SVM), and Catboost. The models were trained on 273 laboratory measurements. The best result was achieved using SVM, with an r^2 score of 0.91 (mean squared error). Study [10] emphasized the importance of combining satellite data and machine learning-based estimation methods with field measurements to improve prediction accuracy. It was shown that Sentinel-2 multispectral images and regression methods can effectively assess chlorophyll concentrations and phytoplankton biomass in the Barents Sea. In research [11], the authors proposed an automatic correction system to improve the accuracy of phytoplankton concentration estimates based on satellite images. Enhanced algorithms demonstrated increased prediction accuracy in water bodies under various atmospheric conditions. In article [12], neural network methods were used to detect phytoplankton. Studies [13, 20-26] also explore the use of machine learning methods for phytoplankton analysis.

b. Sentinel-2 Satellites and the MSI Sensor:

The Sentinel-2 satellite, launched as part of the European Commission's "Copernicus" program, is equipped with a multispectral imager (MSI), which provides imaging with resolutions ranging from 10 to 60 meters across 13 spectral bands. These data, collected in the visible, near-infrared, and shortwave infrared ranges, help detect changes on the Earth's surface, including vegetation, land use, and water resources.

The first satellite, Sentinel-2A, was launched in 2015, followed by its twin, Sentinel-2B, in 2017. This allows for imaging every 2-3 days in mid-latitude regions, including areas like the Volga River and the Kuybyshev Reservoir. Sentinel-2 data enable monitoring the ecological state of water bodies, assessing water quality, tracking vegetation changes, and analyzing human impacts.

Table 1: MSI bands and their parameters.

Band	Resolution	Central wavelength	Description
B2	10 m	490 nm	Blue
B3	10 m	560 nm	Green
B4	10 m	665 nm	Red
B5	20 m	705 nm	Vegetation red edge
B6	20 m	740 nm	Vegetation red edge
B7	20 m	783 nm	Vegetation red edge
B8	10 m	842 nm	Near Infrared (NIR)
B8a	20 m	865 nm	Vegetation red edge
B11	20 m	1610 nm	Shortwave Infrared (SWIR)
B12	20 m	2190 nm	Shortwave Infrared (SWIR)

c. Data Preparation for the Study:

For the analysis of water bloom conditions, we used 10-meter and 20-meter MSI channels, whose characteristics are listed in Table I. Sentinel-2 data was downloaded for the date of water sample collection. If no images were available for the specific date, the closest cloud-free image to the sampling date was selected. Figure 4 shows statistics reflecting the number of cases where the time interval between the image date and the sample collection date was 0, 1, 2, and so on, days. As seen from the graph, only for 19 out of 81 samples were we able to find an image taken on the same day. Due to the influence of wind and currents, even a few hours of deviation can alter the actual situation.

Thus, the amount of data collected, along with temporal deviations, does not allow us to train a highly reliable model for the studied area. However, at this stage of the research, our primary interest was to obtain a rough estimate of chlorophyll-a or at least detect the presence of phytoplankton, so we can later expand the collection of laboratory observations and improve machine learning models.

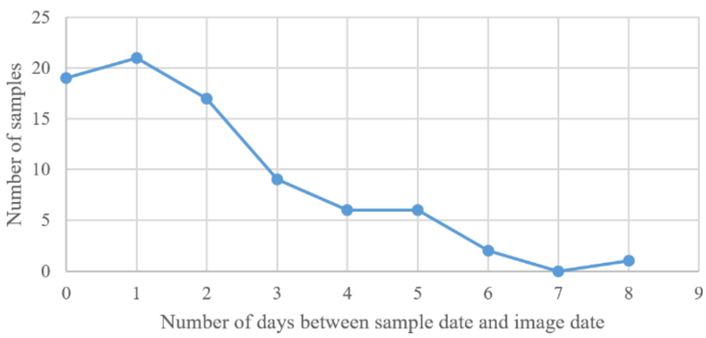

Figure 4: Statistics on the number of days between the laboratory measurement and the image.

Experimental Investigation

a. Regression Models and quality measures:

During the experiments, four regression models were examined: linear regression, Ridge regression, Lasso, and ElasticNet.

Linear regression is one of the most common machine learning methods used to predict the values of a dependent variable (or target variable) based on one or more independent variables (or features). It is based on the assumption that there is a linear relationship between the dependent and independent variables.

Linear regression can be expressed in a compact matrix form as follows:

$$Y = X\beta + \varepsilon,$$

where Y is the dependent variable (target variable), $\beta_1, \beta_2, \ldots, \beta_k$ are the coefficients (weights) of the linear regression, $x_{i1}, x_{i2}, \ldots, x_{ik}$ are the feature values for the i-th observation, and ε_i is the error or residual for the i-th observation.

Ridge regression is an extension of classical linear regression used to address problems of multicollinearity among features and improve the model's generalization ability. The main difference in Ridge regression is the introduction of regularization, which prevents model overfitting by adding a penalty for the magnitude of the regression coefficients. The equation for Ridge regression can be written as follows:

$$\hat{\beta} = arg\min_{\beta}\left(\sum_{i=1}^{n} \|Y_i - X_i\beta\|^2 + \lambda\|\beta\|^2\right),$$

where Y_i is the vector of dependent variables for the i-th observation (with dimensions $m \times 1$), X_i is the feature vector for the i-th observation (with dimensions $k \times 1$), β is the matrix of regression coefficients (with dimensions $k \times m$), λ is the regularization parameter.

Lasso regression is a type of linear regression that uses L1 regularization to improve prediction quality and address the problem of overfitting. One of the key features of Lasso regression is its ability to perform feature selection, making it particularly useful in tasks with a large number of features. Lasso can shrink the coefficients of irrelevant features to zero, effectively removing them from the model, which enhances interpretability and reduces model complexity.

The equation for Lasso regression can be written as follows:

$$\hat{\beta} = arg\min_{\beta}\left(\sum_{i=1}^{n} \|Y_i - X_i\beta\|^2 + \lambda\sum_{j=1}^{k}\left\|\beta_j\right\|_1\right),$$

where X_i is the feature vector for the i-th observation (with dimensions $k \times 1$), β is the matrix of regression coefficients (with dimensions $k \times m$, where m is the number of dependent variables), Y_i is the vector of dependent variables for the i-th observation (with dimensions $m \times 1$), $\|\beta_j\|_1$ is the L1-norm of the coefficients for feature j, λ is the regularization parameter that controls the strength of the penalty.

ElasticNet regression is a generalization of the Ridge and Lasso regression methods, combining their strengths to solve linear regression problems with regularization. This method is particularly useful for analyzing high-dimensional data where strong correlations between features exist. The main goal of ElasticNet is to improve the model's robustness and interpretability, as well as to prevent overfitting.

The loss function for ElasticNet in the multivariate case is as follows:

$$\hat{\beta} = arg \min_{\beta} \left(\sum_{i=1}^{m} \|Y_i - X\beta_i\|_2^2 + \lambda_1 \sum_{i=1}^{m} \|\beta_i\|_1 + \lambda_2 \sum_{i=1}^{m} \|\beta_i\|_2^2 \right),$$

where Y_i- is the vector of values for the i-th dependent variable, β_i is the vector of coefficients for the i-th dependent variable, λ_1 и λ_2 are regularization parameters for each target column, $\|\beta_i\|_1$ is the L1-norm for feature selection, and $\|\beta_i\|_2^2$ is the L2-norm for coefficient regularization and reducing the impact of multicollinearity..

The quality metrics used were MSE and r^2. The Mean Squared Error (MSE) is a measure that represents the average of the squared prediction errors. It is calculated using the following formula:

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y_i})^2,$$

where N is the number of observations, y_i is the actual value, and $\hat{y_i}$ is the predicted value. MSE is used to assess the accuracy of the model, and its value is always non-negative. The smaller the MSE, the better the model.

The coefficient of determination (r^2) measures the proportion of the variance in the dependent variable that is explained by the model. It is defined as:

$$r^2 = 1 - \frac{SS_{res}}{SS_{tot}},$$

where $SS_{res} = \sum_{i=1}^{N} (y_i - \hat{y_i})^2$ is the sum of squared residuals, and $SS_{tot} = \sum_{i=1}^{N} (y_i - \bar{y})^2$ is the total sum of squares (with $\bar{y}$ being the mean of the actual values).

The maximum value of r^2 is 1, indicating a perfect model. Values of r^2 less than or equal to 0 indicate no explanatory power.

a. Training Regression Models:

The sample of 81 instances was divided into training and test sets in a 4:1 ratio. Chlorophyll-a was predicted using 10 MSI bands. Additionally, chlorophyll-a was assessed separately using channel B3 and channel B8 for comparison. The results presented in Table II indicate that high accuracy has not yet been achieved due to the small size of the training dataset and differences in observation dates (as shown in Figure 4). Nevertheless, the best model managed to estimate the desired indicator with an r^2 of 0.5696, suggesting that the regression model can explain more than half of the error variance. It is also noteworthy that when estimating chlorophyll concentration using a single indicator, r^2 is less than 0, meaning that relying on just one channel does not yield a reasonably accurate estimate.

Table 1: MSI bands and their parameters.

Model	**10 bands: MSE**	**10 bands: r^2**	**B3: r^2**	**B8: r^2**
Linear	326.41	0.570	-0.483	-0.504
Ridge	760.57	-0.003	-0.469	-0.489
Lasso	513.36	0.323	-0.352	-0.152
ElasticNet	678.78	0.105	-0.301	-0.235

a. Classification of the Aquatic Area Using a Larger Dataset

Due to the lack of accurate chlorophyll-a concentration data from laboratory samples, we attempted to tackle the classification task—detecting phytoplankton. This task is simpler to solve than estimating chlorophyll-a and does not require laboratory observations, as the presence of blooming phytoplankton in the water column can be visually detected through the analysis of satellite images in the RGB or RGNir range.

For data labeling, training classification models, and evaluating their performance, we obtained images of the Zhiguli Sea from August, September, and October 2020, June, July, and August 2021, as well as August and September 2022, and July, August, and September 2023. These images were aligned for resolution and offset, focusing on the same area, which includes the Priplotinnoye Bay (Zhiguli Sea). Using the Cvat program, the obtained images were visually annotated into three classes: water with clear phytoplankton presence, clear water, and other areas. Pixels in the first two classes formed the dataset used for training, validating, and testing the classification models. Figures 5-6 show examples of images and the corresponding two masks: the phytoplankton mask and the water mask.

In the study, the XGBoost and CatBoost models were tested.

XGBoost (Extreme Gradient Boosting) is a powerful machine learning algorithm based on the gradient boosting method. It was developed to enhance the performance and accuracy of models, as well as to handle large datasets. XGBoost includes the following key stages:

- Tree construction: The algorithm sequentially builds decision trees, where each new tree is trained on the errors of the previous trees. This helps improve predictions and minimize error.
- Regularization optimization: XGBoost incorporates L1 and L2 regularization, which helps prevent overfitting and improves the model's generalization ability.
- Parallel computations: XGBoost implements parallel tree construction, significantly speeding up the training process compared to traditional gradient boosting methods.

CatBoost is a machine learning algorithm developed by Yandex, specifically optimized for working with categorical data. CatBoost stands out for its high performance, accuracy, and ease of use, making it a popular choice for data analysis in various fields. CatBoost operates on the principle of gradient boosting, but with a focus on handling categorical features.

As a result, both models showed comparable performance based on the Accuracy metric: 89.97% for XGBoost and 89.95% for CatBoost. The classification errors were not significant. These results suggest the promising potential of using MSI data and machine learning methods for analyzing phytoplankton distribution in the conditions of the Middle Volga.

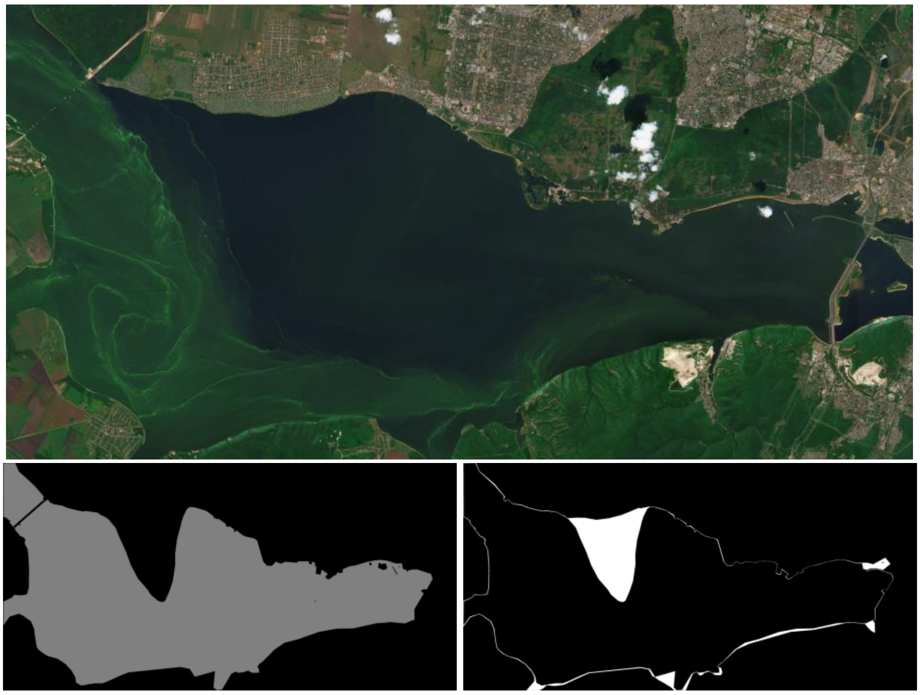

Figure 5: Figure 6. Example of an image from July 18, 2021 (top) with masks for phytoplankton classes (center) and water (bottom).

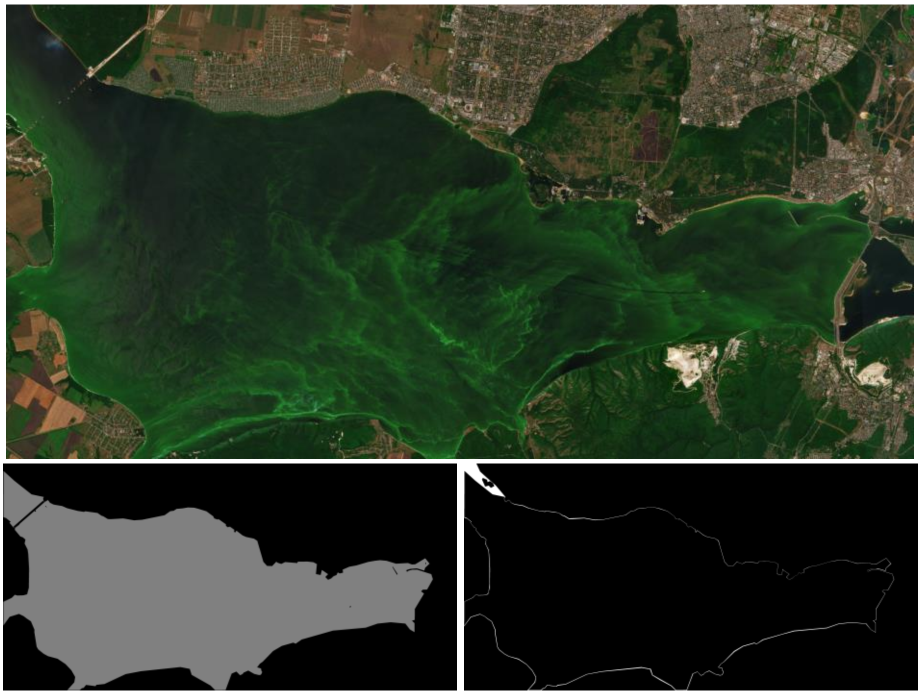

Figure 5: Example of an image from June 25, 2020 (top) with masks for phytoplankton classes (center) and water (bottom).

Conclusion

This study demonstrates the potential of using Sentinel-2 satellite data for assessing phytoplankton development in the Kuibyshev Reservoir. While traditional field measurements provide important localized data, the large surface area of the reservoir limits the ability to accurately track changes in chlorophyll-a concentration and phytoplankton biomass over time and space. Our experiments with remote sensing data, combined with machine learning models, showed promising results in estimating chlorophyll-a concentration, particularly as the sample size increased. Additionally, classification models successfully predicted whether chlorophyll-a concentrations exceeded specific thresholds, achieving high accuracy (over 90%) in test areas.

Despite the limitations posed by the relatively small number of field samples, our manual labeling experiment demonstrated that binary classification models can effectively detect areas with high phytoplankton concentrations. This approach paves the way for more comprehensive monitoring and analysis of phytoplankton distribution across time and space, offering valuable insights for future ecological monitoring and water management efforts. Expanding the dataset and refining machine learning models will further improve the precision of these estimations.

Acknowledgment

The study was supported by the Russian Science Foundation grant 23-14-20005, https://rscf.ru/project/23-14-20005/.

References

1. Lopez, C. B., Jewett, E. B., Dortch, Q., Walton, B. T., & Hudnell, H. K. (2008). Scientific assessment of freshwater harmful algal blooms. Interagency Working Group on Harmful Algal Blooms, Hypoxia, and Human Health of the Joint Subcommittee on Ocean Science and Technology. http://hdl.handle.net/1834/30787
2. Wurtsbaugh, W. A., Paerl, H. W., & Dodds, W. K. (2019). Nutrients, eutrophication and harmful algal blooms along the freshwater to marine continuum. WIREs Water, 6(5), e1373. https://doi.org/10.1002/wat2.1373
3. Huisman, J., Codd, G. A., Paerl, H. W., Ibelings, B. W., Verspagen, J. M. H., & Visser, P. M. (2018). Cyanobacterial blooms. Nature Reviews Microbiology, 16(8), 471–483. https://doi.org/10.1038/s41579-018-0040-1
4. Paerl, H. W., & Paul, V. J. (2012). Climate change: Links to global expansion of harmful cyanobacteria. Water Research, 46(5), 1349–1363. https://doi.org/10.1016/j.watres.2011.08.002
5. Rozenberg, G. S., & Vikchristyuk, L. A. (2008). Kuibyshev Reservoir (Scientific-Information Reference Book). Institute of Ecology of the Volga River Basin RAS, Tolyatti Russia (in Russian).
6. Tarasova N.G., Bykova S.V., Umanskaya M.V., Gorbunov M.Yu., Mukhortova O.V., Krasnova E.S. Formation of a community of planktonic organisms in the Dam Reach of the Kuibyshev Reservoir in the autumn of 2020 // In the collection: Modern problems of reservoirs and their catchments. Proceedings of the IX All-Russian scientific and practical conference with international participation. Perm, 2023. Pp. 228-231.

7. Tarasova, N. G. (2005). Composition, seasonal dynamics and invasive species of phytoplankton of the Kuibyshev reservoir The author's abstract dis. Cand. Sc.
8. M. V. Umanskaya, S. V. Bykova, M. Yu. Gorbunov, E. S. Krasnova, N. G. Tarasova "Transformation of unicellular plankton in the river-bay-plain reservoir system in the initial phase of cyanobacterial bloom". Bulletin of the Samara Scientific Center of the Russian Academy of Sciences. 2021. Vol. 23. No. 5 (103). P. 144-151.
9. Li, S., Song, K., Wang, S., Liu, G., Wen, Z., Shang, Y., Lyu, L., Chen, F., Xu, S., Tao, H., Du, Y., Fang, C., & Mu, G. (2021). Quantification of chlorophyll-a in typical lakes across China using Sentinel-2 MSI imagery with machine learning algorithm. Science of The Total Environment, 778, 146271. https://doi.org/10.1016/j.scitotenv.2021.146271
10. Asim, M., Brekke, C., Mahmood, A., Eltoft, T., & Reigstad, M. (2021). Improving Chlorophyll-A Estimation From Sentinel-2 (MSI) in the Barents Sea Using Machine Learning. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 14, 5529–5549. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. https://doi.org/10.1109/JSTARS.2021.3074975
11. Gómez, D., Salvador, P., Sanz, J., & Casanova, J. L. (2021). A new approach to monitor water quality in the Menor sea (Spain) using satellite data and machine learning methods. Environmental Pollution, 286, 117489. https://doi.org/10.1016/j.envpol.2021.117489
12. Watanabe, F. S. Y., Miyoshi, G. T., Rodrigues, T. W. P., Bernardo, N. M. R., Rotta, L. H. S., Alcântara, E., & Imai, N. N. (2020). Inland water's trophic status classification based on machine learning and remote sensing data. Remote Sensing Applications: Society and Environment, 19, 100326. https://doi.org/10.1016/j.rsase.2020.100326
13. Pahlevan, N., Smith, B., Schalles, J., Binding, C., Cao, Z., Ma, R., Alikas, K., Kangro, K., Gurlin, D., Hà, N., Matsushita, B., Moses, W., Greb, S., Lehmann, M. K., Ondrusek, M., Oppelt, N., & Stumpf, R. (2020). Seamless retrievals of chlorophyll-a from Sentinel-2 (MSI) and Sentinel-3 (OLCI) in inland and coastal waters: A machine-learning approach. Remote Sensing of Environment, 240, 111604. https://doi.org/10.1016/j.rse.2019.111604
14. Rolim, S. B. A., Veettil, B. K., Vieiro, A. P., Kessler, A. B., & Gonzatti, C. (2023). Remote sensing for mapping algal blooms in freshwater lakes: A review. Environmental Science and Pollution Research, 30(8), 19602–19616. https://doi.org/10.1007/s11356-023-25230-2
15. Ecology of Phytoplankton in the Kuibyshev Reservoir. L.: Nauka, 1989. 304 p.
16. Jeffrey, S. W., & Humphrey, G. F. (1975). New spectrophotometric equations for determining chlorophylls a, b, c1 and c2 in higher plants, algae and natural phytoplankton. Biochemie Und Physiologie Der Pflanzen, 167(2), 191–194. https://doi.org/10.1016/S0015-3796(17)30778-3
17. Lorenzen, C. J. (1967). Determination of Chlorophyll and Pheo-Pigments: Spectrophotometric Equations. Limnology and Oceanography, 12(2), 343–346. https://doi.org/10.4319/lo.1967.12.2.0343

18. Methodology for the Study of Biogeocenoses of Inland Water Bodies. Mordukhai-Boltovskoi F. D., ed. M.: Nauka, 1975. 240 p.
19. Guide to Freshwater Algae of the USSR. Vol. 1-14. M., L.: Soviet Science - Nauka, 1951-1983.
20. Grendaitė, D., & Petkevičius, L. (2024). Identification of Algal Blooms in Lakes in the Baltic States Using Sentinel-2 Data and Artificial Neural Networks. IEEE Access, 12, 27973–27988. https://doi.org/10.1109/ACCESS.2024.3366491
21. Woo Kim, Y., Kim, T., Shin, J., Lee, D.-S., Park, Y.-S., Kim, Y., & Cha, Y. (2022). Validity evaluation of a machine-learning model for chlorophyll a retrieval using Sentinel-2 from inland and coastal waters. Ecological Indicators, 137, 108737. https://doi.org/10.1016/j.ecolind.2022.108737
22. Bresciani, M., Cazzaniga, I., Austoni, M., Sforzi, T., Buzzi, F., Morabito, G., & Giardino, C. (2018). Mapping phytoplankton blooms in deep subalpine lakes from Sentinel-2A and Landsat-8. Hydrobiologia, 824(1), 197–214. https://doi.org/10.1007/s10750-017-3462-2
23. Viso-Vázquez, M., Acuña-Alonso, C., Rodríguez, J. L., & Álvarez, X. (2021). Remote Detection of Cyanobacterial Blooms and Chlorophyll-a Analysis in a Eutrophic Reservoir Using Sentinel-2. Sustainability, 13(15), 8570. https://doi.org/10.3390/su13158570
24. Caballero, I., Fernández, R., Escalante, O. M., Mamán, L., & Navarro, G. (2020). New capabilities of Sentinel-2A/B satellites combined with in situ data for monitoring small harmful algal blooms in complex coastal waters. Scientific Reports, 10(1), 8743. https://doi.org/10.1038/s41598-020-65600-1
25. Colkesen, I., Ozturk, M. Y., & Altuntas, O. Y. (2024). Comparative evaluation of performances of algae indices, pixel- and object-based machine learning algorithms in mapping floating algal blooms using Sentinel-2 imagery. Stochastic Environmental Research and Risk Assessment, 38(4), 1613–1634. https://doi.org/10.1007/s00477-023-02648-1
26. Chusnah, W. N., Chu, H.-J., Tatas, & Jaelani, L. M. (2023). Machine-learning-estimation of high-spatiotemporal-resolution chlorophyll-a concentration using multi-satellite imagery. Sustainable Environment Research, 33(1), 11. https://doi.org/10.1186/s42834-023-00170-1