**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, and Fedoseev V.A.1\*

1 Geoinformatics and Information Security Department, Samara National Research University, Russia

2 Laboratory of hydrobiology, Institute of Ecology of Volga Вasin RAS, Samara Federal Research Center RAS, Russia

3 Laboratory of Algology, Papanin Institute for Biology of Inland Waters RAS, Russia

\*vicanfed@gmail.com

***ABSTRACT***

Phytoplankton is an essential component of aquatic ecosystems. The Kuibyshev reservoir, located in the middle reaches of the Volga River, is the largest in Eurasia. Phytoplankton monitoring in it is carried out using a grid of permanent stations. However, due to its large water area, this does not allow obtaining the most complete and comprehensive picture of the spatio-temporal changes in the chlorophyll "a" concentration and phytoplankton biomass. 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 60. These data were used to train 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. Additionally, some classification models (Random Forest, SVM, Decision Tree, etc.) were similarly trained to predict whether chlorophyll a concentration exceeds a certain threshold. The experiments showed that the classification accuracy on test areas exceeds 0.9. 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. The experiment 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