ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

UDC 528.854.04:004.85:581.132.1 doi: 10.32620/reks.2024.2.05

Bohdan YAILYMOV[1], Nataliia KUSSUL[1,2], Pavlo HENITSOI[2], Andrii SHELESTOV[1,2]

[1] *Space research institute of National Academy of Sciences of Ukraine and State Space Agency of Ukraine, Kyiv, Ukraine*
[2] *National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine*

IMPROVING SPATIAL RESOLUTION OF CHLOROPHYLL-A IN THE MEDITERRANEAN SEA BASED ON MACHINE LEARNING

*The **objective of this study** is to increase the spatial resolution of data on the level of chlorophyll-a in the Mediterranean Sea using satellite images and ground measurements. The **goal** of this study is to develop an information technology based on machine learning to create chlorophyll-a concentration maps with high spatial resolution for the pilot areas of the Mediterranean Sea. Traditional ground-based methods for measuring chlorophyll-a are time-consuming, expensive, and have limited spatial and temporal coverage. Therefore, satellite observations have become an effective tool for monitoring chlorophyll-a over large areas. Low spatial resolution satellite data such as GCOM-C/SGLI and Sentinel-3 OLCI allow measurements of chlorophyll-a concentration at the sea surface. However, these data have limited accuracy and spatial resolution, which creates challenges for monitoring local changes in coastal zones and small water areas. **Tasks**: to analyze available satellite data and ground-based measurements of chlorophyll-a for the Mediterranean Sea; to investigate the correlation between satellite data of different spatial resolutions and ground measurements; to select informative features from satellite data for building machine learning models; and to develop models for increasing the spatial resolution of chlorophyll-a based on **regression and machine learning algorithms**. **Obtained results**: information technology combining satellite data with ground measurements in the Google Earth Engine cloud platform is proposed; correlations between satellite measurements of chlorophyll-a and ground data are investigated; models based on Random Forest and Multilayer Perceptron with coefficients of determination up to 0.36 and correlation of 0.6 with test data are built; chlorophyll-a maps with a spatial resolution of 10 m are created for the pilot area near Cyprus. **Conclusions**. The developed information technology allows the effective combination of satellite data of different spatial resolutions and ground measurements to increase the accuracy and detail of chlorophyll-a maps in the Mediterranean Sea. Further research involves improving the preprocessing of satellite data, using more features, involving data from other regions, and applying more sophisticated machine learning models.*

Keywords: *machine learning; satellite data; chlorophyll-a; cloud technologies, information technology; iMERMAID.*

1. Introduction

1.1. Motivation

Measuring chlorophyll-a is an important way to assess water quality because it reflects the photosynthetic activity of phytoplankton, which is the main product of organic matter in aquatic ecosystems [1]. Chlorophyll-a concentration is also linked to biological productivity, biodiversity, the carbon cycle, and global climate change. However, measuring chlorophyll-a using traditional methods, such as water sample collection and laboratory analysis, is time-consuming, expensive, and limited in spatial and temporal coverage. Therefore, satellite observations have become an alternative and effective tool for monitoring chlorophyll-a in large areas.

Satellite monitoring of chlorophyll-a is based on the measurement of sunlight reflected from the water surface in various spectral ranges [2]. Chlorophyll-a absorbs light in the blue and red ranges and reflects light in the green range. Thus, the concentration of chlorophyll-a can be determined using special algorithms that use signal ratios in different ranges. However, using satellite data to estimate chlorophyll-a concentration also presents challenges and limitations. In particular, the influence of the atmosphere, which scatters and absorbs light, leads to errors in the estimation of chlorophyll-a. To correct this effect, atmospheric correction is needed, the algorithms of which can be complex and unreliable, especially for coastal waters with high aerosol concentrations. Another problem is the influence of the sea surface, which reflects light and creates noise in the signals picked up by satellites. To reduce this effect, data corrections for the sea surface are necessary, which may not be of sufficient quality or unsuitable for various conditions of the sea surface, such as waves, foam, and reflection. The higher the spatial resolution of the data, the stronger the effect of internal scattering, which depends on the optical properties of water, such as transparency, color, turbidity, and concentration of dissolved and suspended substances. These factors affect the spectral shape of the signal coming from the water and can mask or alter the chlorophyll-a signal. To take these factors into account, complex optical water models are required, which may be unavailable or inaccurate for different types of water. Another important factor is the effect of seasonal and spatial changes in the concentration of chlorophyll-a, which require frequent and regular validation of satellite data using ground measurements. However, ground-based measurements of chlorophyll-a may be scarce or absent in many regions, especially in remote or inaccessible locations. In addition, ground-based measurements of chlorophyll-a may be inconsistent with satellite data due to differences in time, space, and measurement depth.

Given these challenges and limitations, many researchers have attempted to improve the accuracy and reliability of chlorophyll-a measurements from satellite and ground-based water data, particularly for the Mediterranean Sea.

1.2. State of the Art

There are many satellites that do not measure the value of chlorophyll-a directly, but have a wide set of spectral bands that make it possible to calculate it using various algorithms and approaches. In particular, in [3], the authors investigated the effectiveness of algorithms for calculating chlorophyll-a indicators based on Sentinel-3 data (Ocean Color 4 for MERIS [OC4Me]) in comparison with neural network approaches and concluded that the OC4Me algorithm showed better results in comparison with in situ measurements than neural networks. The authors of the article [4] compared three more algorithms for calculating the concentration of chlorophyll-a: OC4, OC5 and OC6, using data from stations to validate their accuracy and reliability.

In [5], a new methodology is proposed for automatically combining surface reflectance values using multi-sensor satellite observations (Landsat, Sentinel-2, Terra ASTER) with ground samples of water quality in time and space in the Google Earth Engine cloud platform. In [6], the authors developed a model for remote estimation of chlorophyll-a concentration in Lake Dianshan, China, based on Landsat-8 satellite data. Based on the data of the MODIS device, an algorithm was developed [7], which returns the concentration of chlorophyll-a near the water surface in mg/m^3, calculated using the empirical relationship obtained from ground measurements of chlorophyll-a and the ratios of the blue-green bands of the ground remote reflection.

Another promising direction is the application of the latest technologies and methods of machine learning to improve the accuracy and spatial resolution of satellite data. In article [8], the authors propose a new algorithm for restoring gaps in data on the concentration of chlorophyll-a on the sea surface based on data from the Chinese HY-1C satellite, which considers the influence of the atmosphere, sea surface, and internal scattering. In the study [9], the authors used multi-time OLCI data and the Light Gradient Boosting Machine (LightGBM) machine-learning model together with four spectral indices based on the characteristic ranges of OLCI as additional input characteristics to estimate the concentration of chlorophyll-a in Fujian coastal waters. In the article [10], the authors used Sentinel-2 data to determine chlorophyll-a levels in Marmaris Bay, Turkey, integrating satellite data with a geographic information system (GIS) to improve the spatial resolution and accuracy of chlorophyll-a measurements.

The comparison and validation of different chlorophyll-a recovery algorithm for different regions and conditions is also an important stage of research. For example, work [11] analyzed the relationship between chlorophyll-a and sea surface temperature in the Mediterranean Sea using Landsat-8, Sentinel-2 data, and temperature data from the sea and land surface temperature radiometer (Sea and Land Surface Temperature Radiometer - SLSTR.

In [12], the authors analyzed the financial component of land-based measurements of water quality in seas and lakes. They emphasize the importance of supporting both ground-based and satellite monitoring of marine pollution indicators, which provide complementary information for water quality management.

The Mediterranean Sea is one of the most biodiverses and ecologically valuable regions in the world but it is also one of the most polluted. The level of chlorophyll-a is an important indicator of the productivity of marine ecosystems and the impact of pollution. To increase the spatial resolution of chlorophyll-a, this paper proposes information technology in the Google Earth Engine cloud platform [13]. The use of cloud platforms makes it possible to increase the efficiency of training models and their application to data [14, 15]. To implement technology for increasing the spatial resolution of chlorophyll-a data, the possibilities of using satellite data to analyze the level of chlorophyll-a in the Mediterranean Sea were investigated. For this purpose, free satellite data measuring chlorophyll-a in the Mediterranean Sea and their characteristics, such as spatial, temporal, and spectral resolution, availability, etc., were investigated. A review of available ground-based data measuring chlorophyll-a in the Mediterranean Sea, such as buoys, ships, etc., and their characteristics, such as spatial and temporal coverage, availability, etc., has also been reviewed.

To obtain information about the concentration of chlorophyll-a at each point on the sea surface, a comparative correlation analysis of the values of different spectral bands of satellite data (Sentinel-2,3) with indicators from marine stations measuring chlorophyll-a was performed. Spectral bands with the highest level of correlation with data from marine stations were used as informative features in building a model to obtain chlorophyll-a maps with a spatial resolution higher than that of existing products.

1.3. Objective and Approach

Within the scope of this work, it is proposed to achieve an increase in the spatial resolution of chlorophyll-a concentration by applying regression methods of machine learning [16]. The use of regression algorithms of machine learning allows us to consider the complex nonlinear relationships between the spectral characteristics of water surfaces and the level of chlorophyll-a, as well as to effectively generalize the acquired knowledge to new satellite data, ensuring an increase in the spatial detail of the resulting maps.

1.4. Content of the Paper

The structure of the remaining sections in this paper is as follows: Data used and Materials are elaborated in Sections 2 and 3. Section 3.1 introduces the satellite data used, and Section 3.2 describes the in situ dataset and preprocessing steps. The proposed information technology and the process of training and testing the models are described in Section 4. Section 5 describes the metrics used in the article, and Sections 6 and 7 describe some preliminary data analysis results. Section 8 outlines the results of the conducted experiments, followed by the conclusions and discussions.

2. Case Study

In this study, the Mediterranean Sea, which is of great importance for the ecology, economy, and culture of many countries, was chosen as the research area. This sea is also heavily influenced by anthropogenic activities, which lead to water pollution with chemicals that pose a threat to marine ecosystems and biodiversity. To solve this problem, the Horizon Europe iMERMAID project was launched, which aims to integrate innovative solutions for the prevention, monitoring, and restoration of chemical pollution in the Mediterranean Sea.

Within the framework of this project, five pilot areas were selected, which represent different types of marine environments and require special attention and protection. On Fig. 1 shows the location of these territories on the map of the Mediterranean Sea.

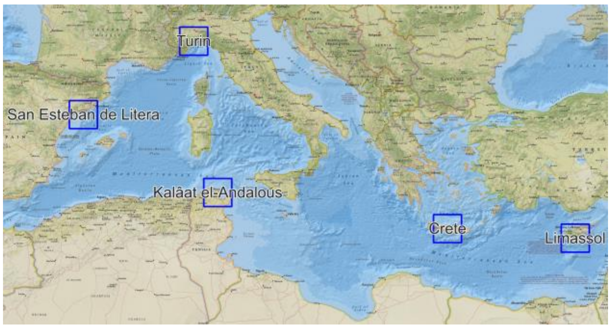


Fig. 1. Pilot territories of the iMERMAID project

Available data from chlorophyll-a measurements were used to analyze the state and dynamics of these territories. Data on chlorophyll-a were obtained from two sources: measurements made at various stations and depths and satellite data obtained from remote monitoring of the Earth. All available chlorophyll-a data were used to train the model. One of the test pilot territories located near Cyprus was chosen for the construction of a high spatial resolution map.

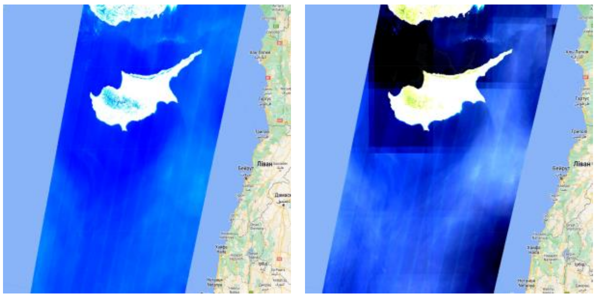
3. Data and Preprocessing

3.1. Satellite Data

In this study, satellite data with different spatial resolutions were used to analyze chlorophyll-a in the Mediterranean Sea. Among the most popular low spatial resolution satellite data used to measure chlorophyll-a is the GCOM-C/SGLI with a spatial resolution of 4638.3 m. The GCOM-C/SGLI L3 Concentration of Chlorophyll-a is a product with a delay of 3-4 days. They make it possible to measure the concentration of photosynthetic pigment (chlorophyll-a) in phytoplankton in the surface layer of the sea. Another satellite used in this study, capable of measuring Chlorophyll-a, is Sentinel-3 OLCI EFR (Ocean and Land Color Instrument Earth Observation Full Resolution) data with a spatial resolution of 300 m. Shooting occurs in 21 spectral bands from visible to near-infrared (400 to 1029 nm). Of these, 7 spectral bands contain a mention of chlorophyll measurements; accordingly, these bands were further investigated for comparison with ground-based measurement data. These data provide the ability to measure large areas, but they have limited accuracy due to their low spatial resolution.

In addition, and considering that Chlorophyll-a absorbs light in the blue (about 458-523 nm) and red (about 650-680 nm) regions of the spectrum and reflects light in the green region of the spectrum (about 543-578 nm), the high spatial resolution data from Sentinel-2 satellites were used to investigate the relationship between satellite and ground data, as well as to improve spatial resolution. Although these data do not directly measure chlorophyll-a, they allow us to explore relationships with other parameters and use them to improve the accuracy and spatial resolution of the product. Table 1 presents satellite data and spectral bands that were studied within the scope of this study.

One of the challenges in using Sentinel-2 data for water quality monitoring is the preprocessing of artifacts. In particular, it was found that collections of harmonized data in the Google Earth Engine (GEE) cloud platform after atmospheric correction using the Sen2Cor algorithm have certain inconsistencies between granules for one date. An example of a satellite image without atmospheric correction and with correction for October 25, 2023 is shown in Fig. 2.



Sentinel-2 L1C (RGB) – without atmospheric correction

Sentinel-2 L2A (RGB, Sen2Cor) – with atmospheric correction

Fig. 2. Example of the Sentinel-2 data near Cyprus on 2023-10-25

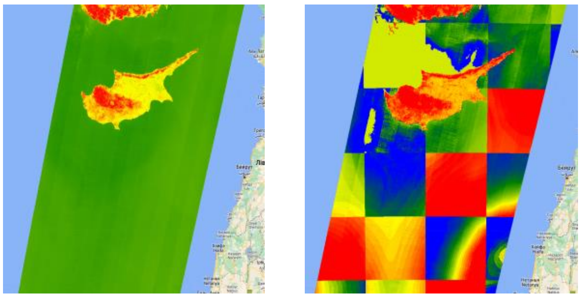
A particularly outlined problem is noticeable after calculating the Normalized Difference Chlorophyll Index (NDCI) [17] (Fig. 3), which is determined by the following formula:

$$\mathrm{NDCI} = \frac{\mathrm{B5} - \mathrm{B4}}{\mathrm{B5} + \mathrm{B4}}, \qquad (1)$$

where B5 is the Visible and Near Infrared (VNIR) band of the Sentinel-2 satellite and B4 is the Red band.

This means that for further use of the data, we should use another approach for atmospheric data correction, or change the settings of the standard Sen2Cor algorithm [18]. In our study, we used data from the cloud-based GEE platform without L1C atmospheric correction and with L2A correction to compare the results.

Other problems that can be seen from Fig. 2, Fig. 3 are bands on the data and solar glare in the sea [19]. The banding problem is associated with odd and even Sentinel-2 sensors, which shoot at different angles [20]. The first problem is not completely solved today, but it is possible to improve the quality of the picture by removing the sun glare, but not for all cases.



Normalized Difference Chlorophyll Index (L1C)

Normalized Difference Chlorophyll Index (L2A)

Fig. 3. NDCI index based on Sentinel-2 data near Cyprus for 2023-10-25

One of the works, where the authors also encountered such problems for the estimation of chlorophyll-a from Sentinel-2 data, is the work [21], where different methods of atmospheric correction were used, and an algorithm based on obtaining BRDF values was used to eliminate the effect of solar glare (Bidirectional Reflectance Distribution Function) of the image in the SWIR (Short Wavelength InfraRed) ranges. There are also other approaches based on the estimation of the brightness of the glow from the signal in the near-infrared range [22]. This is the approach used in our study.

Table 1

Satellite data were used in this study

Satellite	Revisit time (days)	Spatial resolution (m)	Used spectral bands
GCOM-C/SGLI L3 Chlorophyll-a Concentration	2	4638.3	CHLA_AVE
Sentinel-3 OLCI EFR	2	300	Oa03_radiance, Oa04_radiance, Oa05_radiance, Oa06_radiance, Oa08_radiance, Oa10_radiance, Oa11_radiance.
Sentinel-2	5	10 - 60	B1 (aerosol), B2 (blue), B3 (green), B4 (red), B5 - B8a (Visible and Near Infrared), B9 - B12 (Short Wave Infrared)

3.2. In situ Data

Coriolis data are used for ground-based monitoring of ocean parameters [22]. These data typically include measurements of chlorophyll-a, water temperature, salinity, and other parameters using instruments located on buoys, drifting platforms, or ships. Coriolis data play an important role in the validation and calibration of satellite data because they provide an opportunity to obtain accurate sea surface measurements that can be compared with satellite observations.

In addition, Coriolis data were used to create models and analyze the dynamics of marine ecosystems. They allow the study of changes in chlorophyll-a and other important parameters in time and space, which is key to understanding the impact of climate change and human activity on the marine environment.

Coriolis data are widely used in scientific research and practical applications, such as weather forecasting, natural resource management, and environmental protection. For our study, all available chlorophyll-a data until February 2024 were used (Fig. 4).

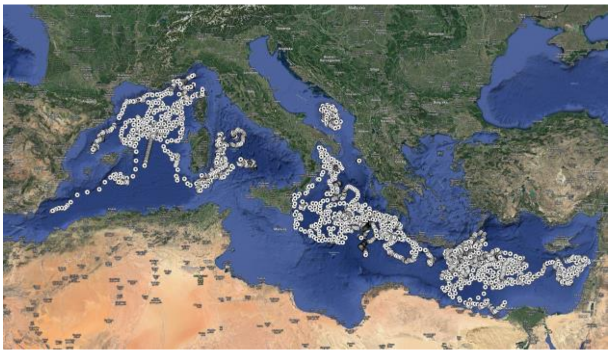


Fig. 4. Geospatial distribution of chlorophyll-a measurement data for the Mediterranean Sea

Before using the data, they were preprocessed. All data with the highest quality according to measurement quality control (identifier in attributes) were downloaded in csv format and converted to vector format (according to coordinates in each file). All values of chlorophyll-a that had negative values were excluded because in this case they have no physical meaning. Measurements at each point were performed several times, each of which corresponds to the measurement depth. Instead of the measurement depth, the attributive information contains the pressure measured in decibars. To determine the depth of chlorophyll-a measurement, the approach described in the Copernicus Marine Service product quality technical documentation [23] was used. Research was also conducted based on the method proposed in [24] and Archimedes’ law, considering the correction for the acceleration of free fall. Results obtained by different methods showed similar results. After obtaining the depth of measurement of chlorophyll-a, for each point, the values closest to the water surface are selected.

For comparison with satellite data, the data were screened by the depth value; in particular, only those points with a measurement depth of no more than 20 m were left for consideration. This depth was chosen considering the Sentinel-2 spatial resolution for SWIR bands (B11 and B12) and experimental comparison with satellite spectral bands. For the period 2015-2024, after data pre-processing, a total of 4547 points were obtained (Fig. 4), and after comparing these measurements with the corresponding satellite data, 1569 of them remained (satellites may not have taken the data on the day when the measurement took place, or the data could get into the cloud mask).

4. Information Technology for Increasing the Spatial Resolution of Chlorophyll-a

The goal of this study was to create a map of Chlorophyll-a concentration with high spatial resolution for a pilot area in the Mediterranean Sea. To achieve this goal, an information technology was developed that integrates satellite data of different spatial resolutions and ground-based measurements of Chlorophyll-a in the Google Earth Engine (GEE) cloud platform. The proposed methodology is based on the use of machine learning to identify relationships between satellite data and ground measurements, and to improve the spatial resolution of chlorophyll-a maps. The general scheme of information technology is presented in Fig. 5 and consists of the following stages.

Preliminary processing of ground measurement data. At this stage, an analysis of ground-based measurements of chlorophyll-a obtained from the Coriolis service is performed. Data are downloaded and filtered before being imported into the GEE cloud platform for further analysis.

Preprocessing of satellite data. Data selection was carried out according to the date of ground measurements of chlorophyll-a. In particular, low spatial resolution data to analyze their dependence with ground data and high spatial resolution data (10 meters) Sentinel-2, which has 13 spectral bands for visible and near-infrared spectra. Preprocessing of satellite data is carried out, such as cloud masking and atmospheric correction.

Data analysis. GCOM-C/SGLI and Sentinel-3 satellite data are compared (chlorophyll concentration is calculated using the OC4ME algorithm [3]) with ground-based chlorophyll-a data from the Coriolis service. The purpose of this comparison is to determine the degree of connection between satellite and ground measurements and to identify informative features and the optimal depth of ground measurements for model training. For comparison, correlation analysis is used, which allows the

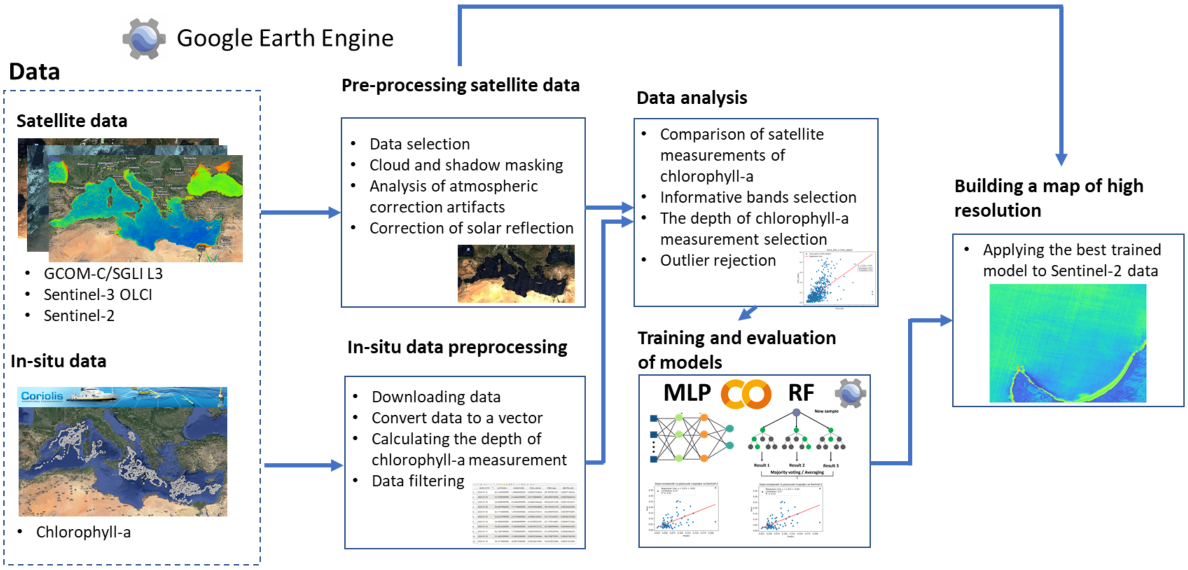

Google Earth Engine
Data
Satellite data
GCOM-C/SGLI L3
Sentinel-3 OLCI
Sentinel-2
In-situ data
Coriolis
Chlorophyll-a
Pre-processing satellite data
Data selection
Cloud and shadow masking
Analysis of atmospheric correction artifacts
Correction of solar reflection
In-situ data preprocessing
Downloading data
Convert data to a vector
Calculating the depth of chlorophyll-a measurement
Data filtering
Data analysis
Comparison of satellite measurements of chlorophyll-a
Informative bands selection
The depth of chlorophyll-a measurement selection
Outlier rejection
Training and evaluation of models
MLP
RF
Building a map of high resolution
Applying the best trained model to Sentinel-2 data

Fig. 5. General scheme of chlorophyll-a mapping based on satellite data with a spatial resolution of 10 m

assessment the dependence between variables. This made it possible to remove anomalous values of chlorophyll-a, and to determine the maximum depth with chlorophyll-a measurements, where there is a dependence on satellite data. After data preparation, they are divided into training (80%) and test (20%) samples to build the model and check its quality.

Development of a model to calculate chlorophyll-a. Two machine-learning algorithms – Random Forest (RF) and Multilayer Perceptron (MLP) – were chosen as models. Both algorithms were trained on 80% of the ground data that were pre-processed and tested on the 20% of ground data that were left for validation. For each model, the best parameters that minimize the prediction error are selected. Optimization of RF parameters included selection of the number of trees, depth, and number of samples for branching. The following parameters are defined as optimal: number of trees: 100; maximum depth: 4; maximum percentage of data usage for the bootstrap sample: 60%; minimum number of samples per sheet: 4; minimum number of samples for branching: 12. MLP optimization includes tuning the number of layers, the number of neurons in each layer, and the learning rate. The following parameters are defined as optimal: activation function: logistic; regularization parameter (alpha): 0.1; learning rate: 0.001; size of hidden layers: (7, 7); maximum number of iterations: 1000. Simulation results were evaluated using the coefficient of determination (R2), mean squared error (MSE) and correlation with the test data set.

Construction of a high-resolution chlorophyll-a map. After training and testing the model, it was applied to Sentinel-2 images to construct a high spatial resolution chlorophyll-a map. For this, spectral bands with resolutions of 10 and 20 m were used.

The developed information technology makes it possible to create maps of water pollution (chlorophyll-a) in the pilot area of the Mediterranean Sea based on Sentinel-2 satellite data in the free GEE cloud platform. Considering the speed of the cloud platform, it is possible to easily expand the construction area of the chlorophyll-a concentration map and input satellite data.

5. Metrics for evaluating the models and data

The following metrics were used to assess the dependence between satellite data and ground measurement data, as well as the quality of the model and its ability to reproduce the dependence between variables: the Pearson correlation coefficient, the coefficient of determination, and the root mean square error [3]. The Pearson correlation coefficient is a measure of the linear relationship between two variables, which ranges from 1 to 1. The closer the coefficient is to 1 or 1, the stronger the relationship between the variables. The correlation coefficient is calculated by the following formula:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}, \qquad (2)$$

where x_i and y_i – values of variables x and y in the i^{th} observation, $\bar{x}$ and $\bar{y}$ – average values of the variables x and y.

The coefficient of determination (R^2) is a measure of how well the model explains the variation of the dependent variable. It shows the proportion of the dependent variable that can be explained by the model. The coefficient of determination takes values from 0 to 1. The closer the coefficient is to 1, the better the model reproduces the data. The coefficient of determination is calculated by the following formula:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y})^2}, \tag{3}$$

where y_i – the actual value of the dependent variable in the i^{th} observation, $\hat{y}_i$ – the predicted value of the dependent variable in the i^{th} observation, $\overline{y}$ – the mean value of the dependent variable.

The mean squared error (MSE) is a measure of how accurately the model predicts the value of the dependent variable. It shows how the actual and predicted values differ on average. The smaller the root mean square error, the better the model. The root mean square error is calculated according to the following formula:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2, \tag{4}$$

where y_i and $\hat{y}_i$ – the actual and predicted values of the dependent variable in the i^{th} observation and n is the number of observations.

6. Analysis of chlorophyll-a measurements based on satellite data

Chlorophyll-a is an important indicator of marine ecosystem pollution that can be measured using satellite data. However, different satellites have different characteristics, such as spatial, temporal and spectral resolution, which can affect the accuracy and comparability of measurements. This study analyzes the chlorophyll-a measurement between GCOM-C/SGLI satellites and Sentinel-3 data within one pixel for the period 01/01/2023 - 12/31/2023.

To compare chlorophyll-a measurements between GCOM-C/SGLI satellites and the Sentinel-3 data, spectral bands related to chlorophyll-a measurements were selected for Sentinel-3 and GCOM-C/SGLI (Table 1). Since the GCOM pixel is significantly larger than Sentinel-3, an average of 4638.3 m was performed using the arithmetic mean of the pixels belonging to one GCOM-C/SGLI pixel (Fig. 6).

In addition, a study was conducted on the calculation of chlorophyll-a according to Sentinel-3 data. As you know, there are different methods: OC4ME, OC4, OC5, OC6, and methods based on neural networks [4]. In this study, the OC4ME method was used to calculate the chlorophyll-a index, which is described in more detail in [3], where the authors made calculations and selected coefficients specifically for the Mediterranean Sea.

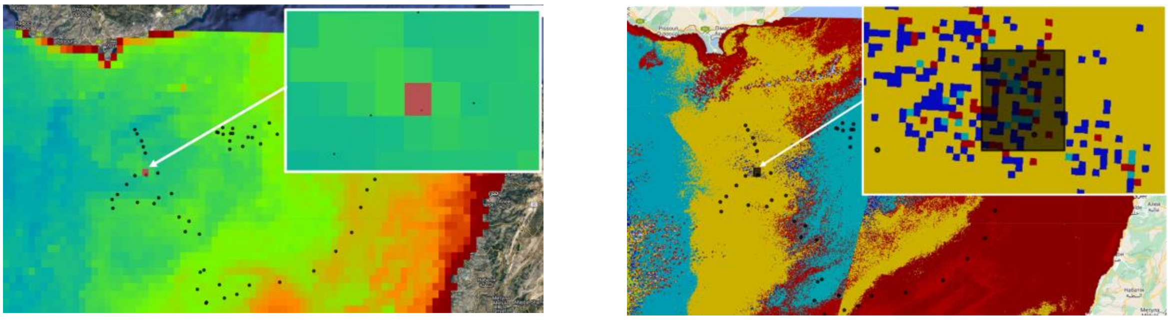
A comparison of the correlation coefficients between the measurements of chlorophyll-a by the GCOM-C/SGLI satellite and the Sentinel-3 data is given in Table 2.

Table 2

Correlations between Sentinel-3 data
with GCOM-C/SGLI

Band	Correlation
Oa011	-0.082851
Oa03	-0.272541
Oa04	-0.255240
Oa05	-0.214820
Oa06	-0.175918
Oa08	-0.114631
s3-Index (method C4ME)	0.379838

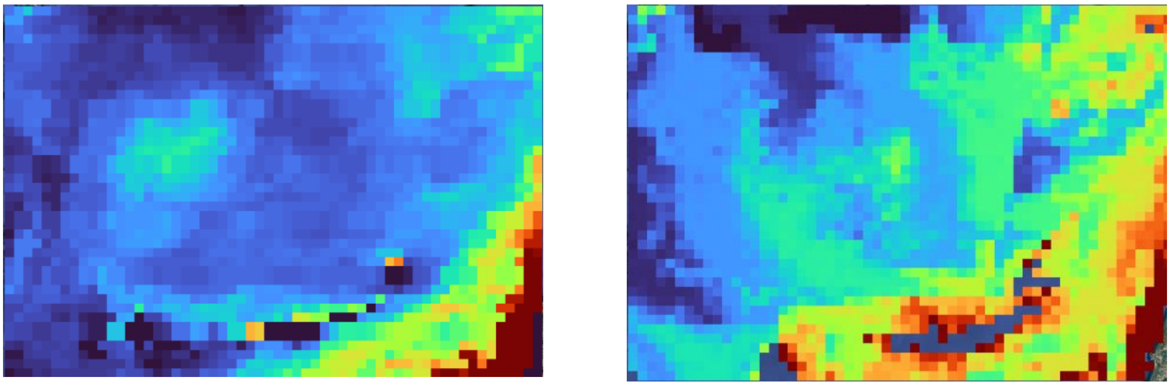
Based on the obtained results between the GCOM-C/SGLI and Sentinel-3 satellites, we can draw conclusions that the correlations with all bands are negative, but the best (the largest by module) is the correlation with the Oa03 band, which indicates their similarities. Measurement of chlorophyll-a concentration using the OC4ME method has the highest correlation with GCOM-C/SGLI, indicating its effectiveness.



GCOM-C/SGLI

Sentinel-3 (OC4ME)

Fig. 6. Example of the comparison of chlorophyll-a for the Mediterranean Sea within one pixel near Cyprus



GCOM-C/SGLI (24.11.23)

Sentinel-3 (OC4ME) (24.11.23)

Fig. 7. An example of the comparison of chlorophyll-a for the Mediterranean Sea near Cyprus

To confirm the results of the correlation, the analysis was performed not within the limits of one pixel, but of the chlorophyll-a maps for the same dates, which had an area of approximately 3,400 thousand hectares. The comparison was also carried out for the period 01.01.2023 - 12.31.2023 (Fig. 7).

For the comparison of satellite data for this period, only those images obtained on the same dates were selected, after which a correlation analysis of individual maps was carried out for each common date. On Fig. 8 shows the frequency distribution of correlations for the period 01.01.2023 - 12.31.2023. Because of cloudiness, the number of pixels for each date could be different when calculating correlations.

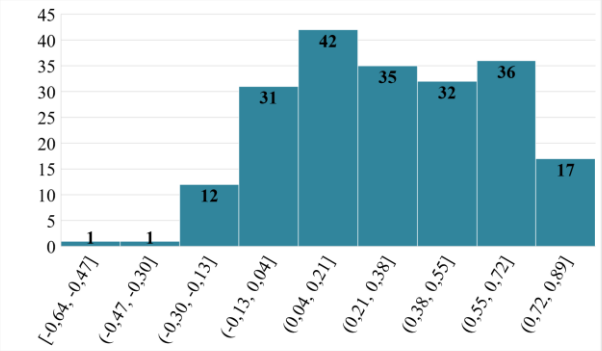


Fig. 8. Frequency distributions of chlorophyll-a correlations between GCOM-C/SGLI and Sentinel-3 (OC4ME) during 2023 for common dates

As can be seen from Fig. 8 emissions are present (in particular, negative values of chlorophyll-a concentrations), but they have been excluded for the sake of transparency of the results. The average correlation for all dates was 0.30. The obtained result is close to the comparison of data within one pixel.

7. Analysis of chlorophyll-a data from ground measurements

To check the reliability and calibration of the satellite data, they were compared with ground measurements obtained with the help of the Coriolis service. These ground data are more accurate and reliable although they are limited in spatial coverage and measurement frequency for one territory. For correct calculations, it is necessary to consider that the measurements are made at different depths, and the satellite, in turn, can only shoot on the surface of the reservoir. To do this, research was conducted to determine at what minimum depth of measurement of chlorophyll-a should be trained in the future model. Table 3 presents the results of the ground data correlation with the GCOM-C/SGLI satellite.

Table 3

Correlations between ground data with the GCOM-C/SGLI satellite

Samples number	Depths (m)	Coefficient of determination	Correlation
377	5	0.3558	0.596
466	10	0.3013	0.5489
600	20	0.3266	0.5715
886	50	0.1883	0.4340
1569	All depths	0.0126	0.1121

To have more data for training the model and considering the spatial resolution of the Sentinel-2 data, data with chlorophyll-a measurements of at least 20 m were selected. Analysis of chlorophyll-a measurement data showed that the dataset contains data with negative values (which is physically impossible for such an indicator) and data with very high values. Accordingly, there was a need to remove outliers, for which we used the three-sigma rule. The result obtained by comparing the filtered ground measurements of chlorophyll-a with the GCOM-C/SGLI satellite data (CHLA_AVE) is presented in Fig. 9. The calculated coefficient of determination is R^2=0.3266, and the correlation is 0.5715.

When comparing ground-based measurements of chlorophyll-a (CPHL_ADJUS) with data from the Sentinel-3 satellite, correlations with all bands were negative (Fig. 10). The ground data correlates best with the Oa03_radiance band (correlation r=-0.56 and R^2=0.32). Therefore, to further use the Sentinel-3 data, it is worth improving the methods of counting chlorophyll-a based on different combinations of bands. Within the scope of this study, Sentinel-3 data were used to check the reliability of the GCOM-C/SGLI satellite measurement and to optimally choose the minimum depth of chlorophyll-a measurement from ground stations.

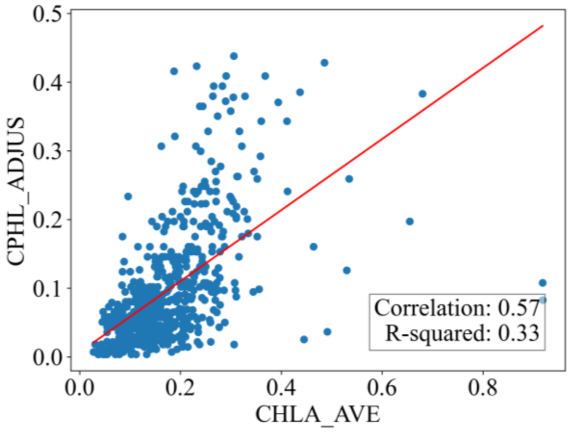


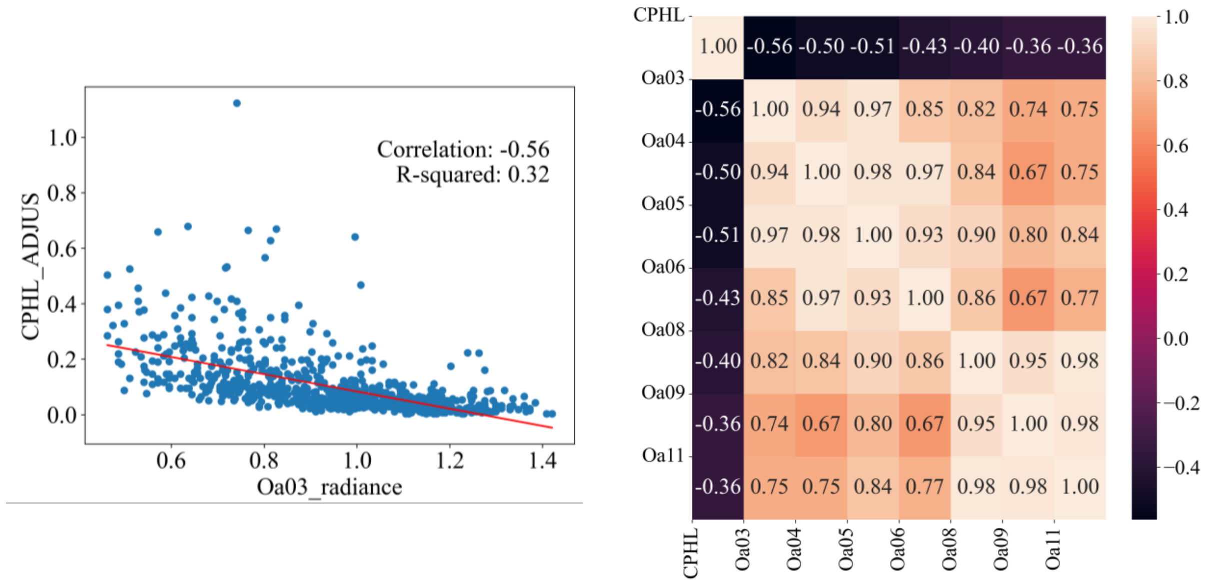
Fig. 9. Comparison of ground-based chlorophyll-a measurements (CPHL_ADJUS) and GCOM-C/SGLI satellite data (CHLA_AVE)

Having confirmed the reliability of ground data, the dependence of chlorophyll-a data on Sentinel-2 satellite data was investigated. For this, too, were the data obtained for each of the 10 bands accurate for the corresponding date. The Sentinel-2 data also lack a band that measures chlorophyll-a directly. Correlations with ground data are generally higher for atmospherically adjusted L2A data (Fig. 11). The highest correlation was obtained with the B1 band used for aerosols, but this may be a consequence of the existing artifacts observed in the data and discussed above.

8. Results

This study developed an information technology using satellite data and ground measurements to improve the spatial resolution of chlorophyll-a in the GEE cloud platform. The trained model was adapted for application to a pilot area in the Mediterranean Sea, with the possibility of use for any date for which Sentinel-2 optical satellite data is available. Pre-processing of the data was limited by the capabilities of the Google Earth Engine cloud platform, which affected the quantity and quality of data in the two pre-prepared datasets, with and without Sentinel-2 correction. The different cloud masks used could also partially affect the results.

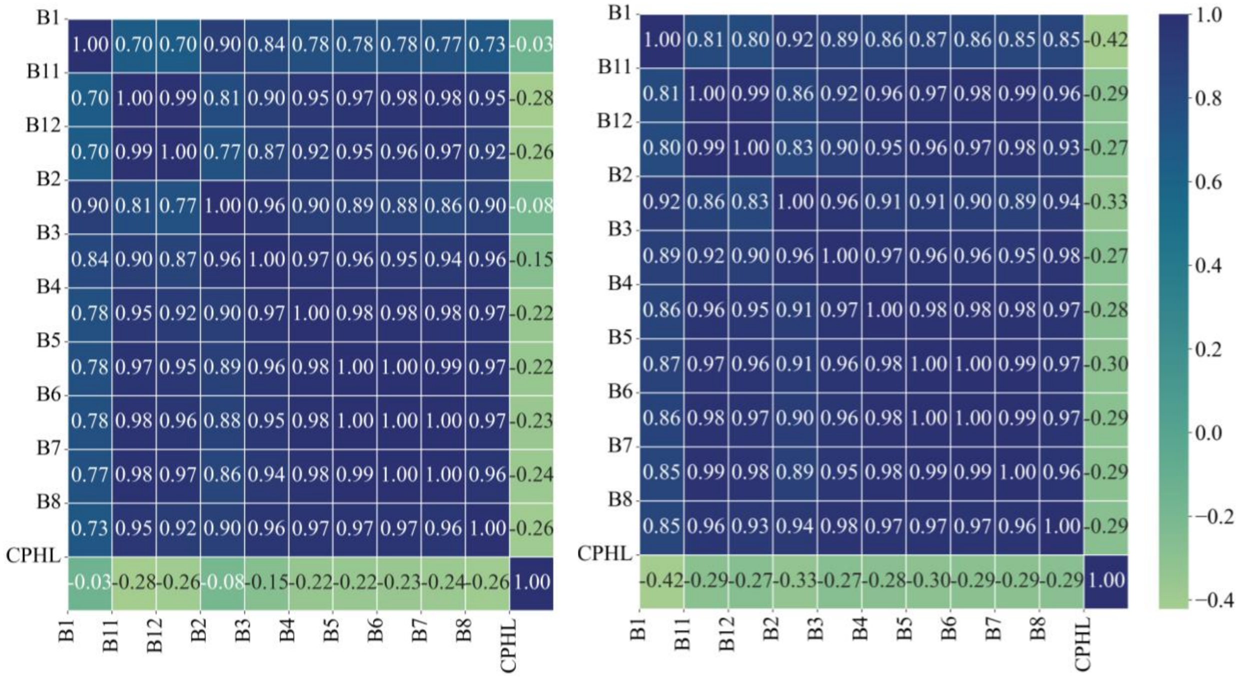
For model training, 80% of the data was used, followed by cross-validation, where the training data was divided into 10 subsets. To train the models, the optimal parameters were searched for the Random Forest (RF) and Multilayer Perceptron (MLP) models to minimize the risk of overtraining. Random Forest was chosen for its ability to efficiently process large datasets while ensuring accuracy. In addition, it is less sensitive to different ranges of values and is available for use in the GEE cloud platform. The Multilayer Perceptron model was used to identify implicit relationships in the data and was performed using Google Collab, which is not available in the GEE cloud platform. For RF, R^2 was 0.33 for uncorrected data and 0.36 for corrected data, and the correlation was 0.56 and 0.6, respectively. For MLP, worse R^2 results were obtained, but MSE performed better on uncorrected data for the two models. The results of the obtained model on the test dataset are presented in Fig. 12.



Dependence of chlorophyll-a Oa03_radiance band of Sentinel-3 with ground measurements (CPHL_ADJUS)

Correlation matrix of Sentinel-3 bands with ground data (CPHL_ADJUS)

Fig. 10. Dependence of in-situ chlorophyll-a measurements on Sentinel-3 bands



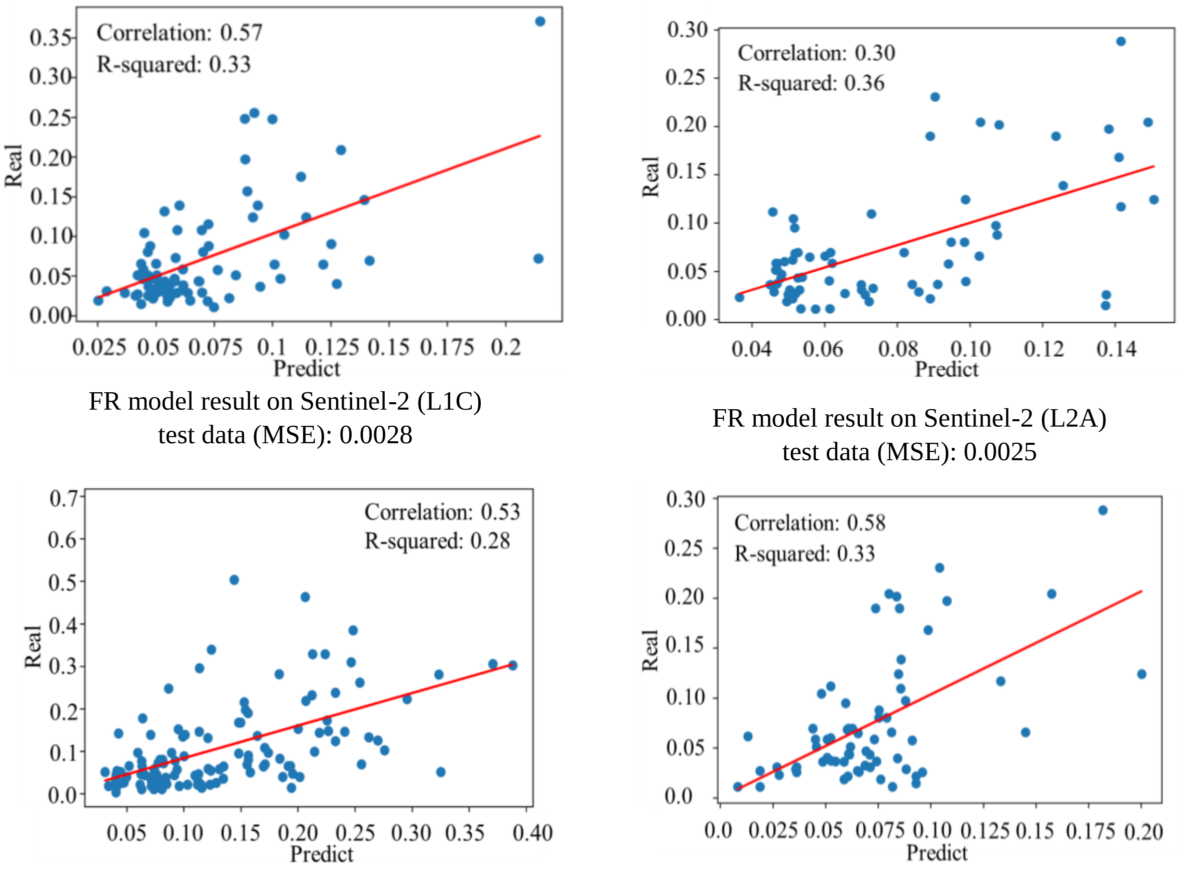
	B1	B11	B12	B2	B3	B4	B5	B6	B7	B8	CPHL
B1	1.00	0.70	0.70	0.90	0.84	0.78	0.78	0.78	0.77	0.73	-0.03
B11	0.70	1.00	0.99	0.81	0.90	0.95	0.97	0.98	0.98	0.95	-0.28
B12	0.70	0.99	1.00	0.77	0.87	0.92	0.95	0.96	0.97	0.92	-0.26
B2	0.90	0.81	0.77	1.00	0.96	0.90	0.89	0.88	0.86	0.90	-0.08
B3	0.84	0.90	0.87	0.96	1.00	0.97	0.96	0.95	0.94	0.96	-0.15
B4	0.78	0.95	0.92	0.90	0.97	1.00	0.98	0.98	0.98	0.97	-0.22
B5	0.78	0.97	0.95	0.89	0.96	0.98	1.00	1.00	0.99	0.97	-0.22
B6	0.78	0.98	0.96	0.88	0.95	0.98	1.00	1.00	1.00	0.97	-0.23
B7	0.77	0.98	0.97	0.86	0.94	0.98	0.99	1.00	1.00	0.96	-0.24
B8	0.73	0.95	0.92	0.90	0.96	0.97	0.97	0.97	0.96	1.00	-0.26
CPHL	-0.03	-0.28	-0.26	-0.08	-0.15	-0.22	-0.22	-0.23	-0.24	-0.26	1.00

Correlation matrix of Sentinel-2 (L1C) with ground data (CPHL_ADJUS)

	B1	B11	B12	B2	B3	B4	B5	B6	B7	B8	CPHL
B1	1.00	0.81	0.80	0.92	0.89	0.86	0.87	0.86	0.85	0.85	-0.42
B11	0.81	1.00	0.99	0.86	0.92	0.96	0.97	0.98	0.99	0.96	-0.29
B12	0.80	0.99	1.00	0.83	0.90	0.95	0.96	0.97	0.98	0.93	-0.27
B2	0.92	0.86	0.83	1.00	0.96	0.91	0.91	0.90	0.89	0.94	-0.33
B3	0.89	0.92	0.90	0.96	1.00	0.97	0.96	0.96	0.95	0.98	-0.27
B4	0.86	0.96	0.95	0.91	0.97	1.00	0.98	0.98	0.98	0.97	-0.28
B5	0.87	0.97	0.96	0.91	0.96	0.98	1.00	1.00	0.99	0.97	-0.30
B6	0.86	0.98	0.97	0.90	0.96	0.98	1.00	1.00	0.99	0.97	-0.29
B7	0.85	0.99	0.98	0.89	0.95	0.98	0.99	0.99	1.00	0.96	-0.29
B8	0.85	0.96	0.93	0.94	0.98	0.97	0.97	0.97	0.96	1.00	-0.29
CPHL	-0.42	-0.29	-0.27	-0.33	-0.27	-0.28	-0.30	-0.29	-0.29	-0.29	1.00

Correlation matrix of Sentinel-2 (L2A) with ground data (CPHL_ADJUS)

Fig. 11. Dependence of ground chlorophyll-a measurements on Sentinel-2 bands (without and with atmospheric correction)



FR model result on Sentinel-2 (L1C) test data (MSE): 0.0028

FR model result on Sentinel-2 (L2A) test data (MSE): 0.0025

MLP model result on Sentinel-2 (L1C) test data (MSE): 0.0027

MLP model result on Sentinel-2 (L2A) test data (MSE): 0.0025

Fig. 12. Results of training on the test dataset

After training the models, they were used to construct a chlorophyll-a map for a pilot area near the northern part of Cyprus. On Fig. 13 presents the input Sentinel-2 images fed to the model. In particular, images without correction (left column) for October 25, 2023 and with correction (right column). As can be seen from the figures, the corrected data have bands on the territory of the sea. Accordingly, the results of building the map also have them. On data without atmospheric correction, such artifacts are much less noticeable, especially based on the FR algorithm, which is not as sensitive to a different range of spectral values. Although the results of the MLP model also have a band due to the measurement sensor on the satellite, their results may be more promising in the future after this problem is eliminated. To obtain uniform results, a solution may be the correction of data based on existing methods or the development of your own. The obtained results confirm the effectiveness of the Random Forest method, which can be used in real time in the GEE cloud platform.

9. Discussions

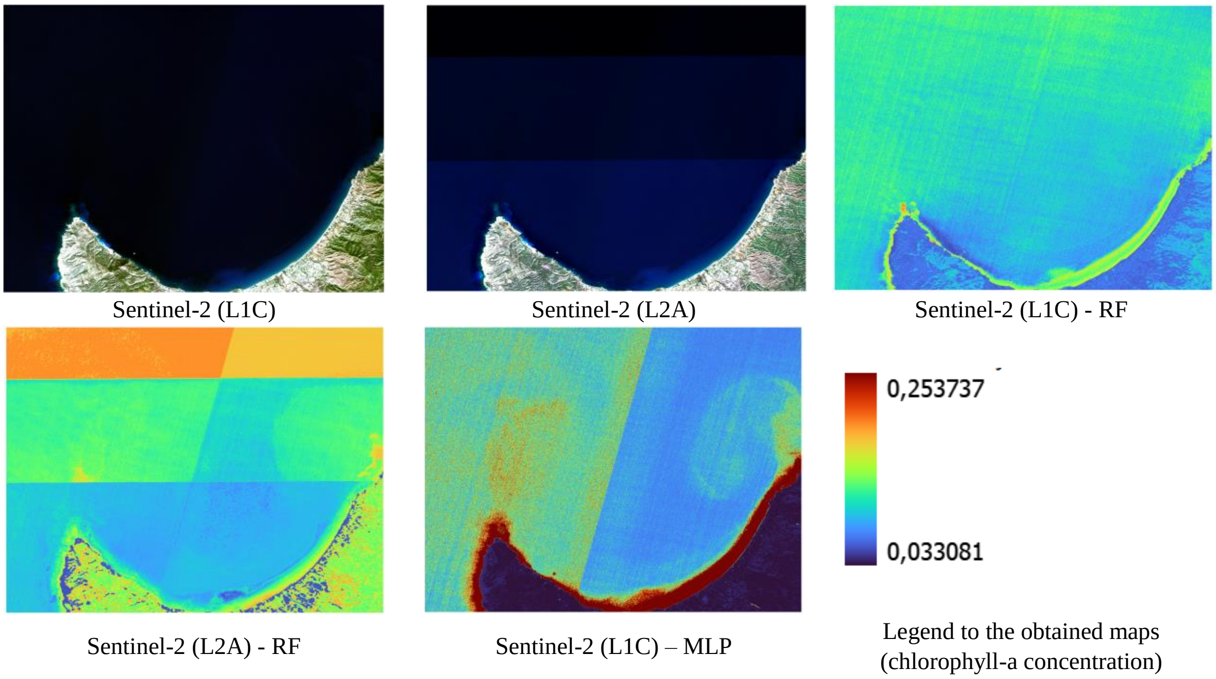
The proposed information technology for increasing the spatial resolution of chlorophyll-a concentration maps demonstrates the potential of using satellite data, ground measurements, and machine learning methods in the GEE cloud platform. Although the obtained coefficients of determination have not yet shown high results, the correlation with test ground data is more than 0.6, which indicates the effectiveness of the developed approach.

However, there are opportunities to improve the accuracy and reliability of the developed models. One of the main challenges faced in this study was the inconsistencies and artifacts of the Sentinel-2 data after atmospheric correction using the Sen2Cor algorithm. In addition, solar glare on the water surface can significantly affect the accuracy of the obtained chlorophyll-a maps. Investigating alternative methods of atmospheric correction or improving existing algorithms could mitigate these problems.

Another aspect that can be further explored is the inclusion of additional features such as spectral indices, bands combinations or data from other satellites to train the models. By expanding the feature space, models can identify complex relationships between satellite observations and chlorophyll-a concentrations, potentially increasing their accuracy.

Alternatively, the use of more advanced machine learning techniques, such as deep learning architectures or ensemble models, can be explored.

It is worth noting that the current study focused on the Mediterranean region. Expanding the training and testing datasets to include data from other geographic areas can increase the robustness of the models and their adaptability to different environmental conditions and water characteristics. If more data are used, it will be possible to select chlorophyll-a measurements taken at a depth of no more than 1 m, which should have the greatest impact on the accuracy of the results.



Sentinel-2 (L1C)

Sentinel-2 (L2A)

Sentinel-2 (L1C) - RF

Sentinel-2 (L2A) - RF

Sentinel-2 (L1C) – MLP

Legend to the obtained maps (chlorophyll-a concentration)

Fig. 13. Chlorophyll-a maps obtained using Random Forest and Multilayer Perceptron

In addition, the integration of additional environmental variables such as sea surface temperature, salinity, or other water quality parameters can provide valuable contextual information and potentially improve the accuracy of chlorophyll-a concentration estimation models.

Although the proposed information technology is a promising step toward high-resolution chlorophyll-a mapping, further research is needed to address the existing challenges and limitations.

10. Conclusions

This paper proposes an information technology for increasing the spatial resolution of chlorophyll-a. For this purpose, an analysis was conducted regarding the possibility of satellite monitoring of the quality of water bodies (chlorophyll-a concentration), especially in the Mediterranean Sea. The available satellite data, which make it possible to measure the concentration of chlorophyll-a on the surface of water bodies, were studied and compared during 2023 for the pilot area. It was found that the average correlation between GCOM-C/SGLI and Sentinel-3 data is equal to 0.3 both at the level of one pixel and when comparing chlorophyll-a maps in the pilot area. To improve the correlations, further research is needed to improve the calculation of chlorophyll-a from Sentinel-3 data.

The available ground-based data of chlorophyll-a measurements for the Mediterranean Sea were also investigated. One of the best and most effective sources is the Coriolis service. To use these data with satellite data, we conducted a study on the dependence of chlorophyll-a between different satellite data and measurement depths. It was important to choose the minimum measurement depth at which it is possible to obtain the largest amount of data that is used for training and testing models. Following these studies, a dataset was created to train models to map chlorophyll-a using high-resolution satellite data.

The results of model training demonstrate a correlation with ground test data at the level of 0.6 and the coefficient of determination at the level of 0.36. To improve the results of the model in the future, it is planned to investigate more features (indices, band combinations, data from other satellites) that can be used, as well as to improve the preprocessing of Sentinel-2 satellite data. In particular, to improve the methods of atmospheric correction, the methods of removing sunlight from water, and the coordination of different satellite sensors to eliminate artifacts. It is also planned to apply more complex machine learning models and use data not only from the Mediterranean Sea but also from other regions to increase the amount and variability of training and test data.

Contributions of authors: conceptualization, methodology, development of model, analysis, writing – original draft preparation – **Bohdan Yailymov;** formulation of tasks, analysis of results, writing – review and editing **– Andrii Shelestov, Nataliia Kussul;** software, verification, visualization **– Pavlo Henitsoi.**

Conflict of Interest

The authors declare that they have no conflict of interest in relation to this research, whether financial, personal, authorship or otherwise, that could affect the research and its results presented in this paper.

Financing

The work was founded within the HORIZON Europe project iMERMAID "Innovative solutions for Mediterranean Ecosystem Remediation via Monitoring and Decontamination from Chemical Pollution", (contract number 101112824).

Data Availability

The data will be made available upon reasonable request.

Use of Artificial Intelligence

During the preparation of this work, the authors used ChatGPT to improve language and readability. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

All the authors have read and agreed to the published version of this manuscript.

References

1. Roussillon, J., Fablet, R., Gorgues, T., Drumetz, L., Littaye, J., & Martinez, E. A Multi-Mode Convolutional Neural Network to reconstruct satellite-derived chlorophyll-a time series in the global ocean from physical drivers. *Frontiers in Marine Science*, 2023, vol. 10, article no. 1077623. DOI: 10.3389/fmars.2023.1077623.

2. Zennaro, F., Furlan, E., Canu, D., Alcazar, L. A., Rosati, G., Solidoro, C., Aslan, S., & Critto, A. Venice lagoon chlorophyll-a evaluation under climate change conditions: A hybrid water quality machine learning and biogeochemical-based framework. *Ecological Indicators*, 2023, vol. 157, article no. 111245. DOI: 10.1016/j.ecolind.2023.111245.

3. Moutzouris-Sidiris, I., & Topouzelis, K. Assessment of Chlorophyll-a concentration from Sentinel-3 satellite images at the Mediterranean Sea using CMEMS open source in situ data. *Open Geosciences*, 2021, vol. 13, no. 1, pp. 85-97. DOI: 10.1515/geo-2020-0204.

4. Binh, N., Hoa, P., Thao, G., Duan, H., & Thu, P. Evaluation of Chlorophyll-a estimation using Sentinel 3 based on various algorithms in southern coastal Vietnam. *International Journal of Applied Earth Observation and Geoinformation*, 2022, vol. 112, article no. 102951. DOI: 10.1016/j.jag.2022.102951.

5. Wang, L., Xu, M., Liu, Y., Liu, H., Beck, R., Reif, M., Emery, E., Young, J., & Wu, Q. Mapping freshwater chlorophyll-a concentrations at a regional scale integrating multi-sensor satellite observations with Google earth engine. *Remote Sensing,* 2020, vol. 12, no. 20, article no. 3278. DOI: 10.3390/rs12203278.

6. Theologou, I., Kagalou, I., Papadopoulou, M., & Karantzalos, K. Multitemporal mapping of chlorophyll-α in Lake Karla from high resolution multispectral satellite data. *Environmental Processes*, 2016, vol. 3, pp. 681-691. DOI: 10.1007/s40710-016-0163-1.

7. *MODIS Chlorophyll-a Concentration*. Available at: https://modis.gsfc.nasa.gov/data/dataprod/chlor_a.php. (accessed 29.01.2024).

8. Yang, G., Ye, X., Xu, Q., Yin, X., & Xu, S. Sea surface chlorophyll-a concentration retrieval from hy-1c satellite data based on residual network. *Remote Sensing*, 2023, vol. 15, no. 14, article no. 3696. DOI: 10.3390/rs15143696.

9. Su, H., Lu, X., Chen, Z., Zhang, H., Lu, W., & Wu, W. Estimating coastal chlorophyll-a concentration from time-series OLCI data based on machine learning. *Remote Sensing,* 2021, vol. 13, no. 4, article no. 576. DOI: 10.3390/rs13040576.

10. Kaymaz, Ş., & Ates, E. Estimating chlorophyll-a concentration using remote sensing techniques. *Annals of Reviews and Research*, 2018, vol. 4, no. 2, pp. 555-633. Available at: https://juniperpublishers.com/arr/pdf/ARR.MS.ID.555633.pdf (accessed 04.01.2024).

11. Tuzcu Kokal, A., & Musaoğlu, N. Monitoring chlorophyll-a and sea surface temperature with satellite data derived from multiple sensors. *Proceedings of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2021, vol. 43, pp. 515-520. DOI: 10.5194/isprs-archives-XLIII-B3-2021-515-2021.

12. Papenfus, M., Schaeffer, B., Pollard, A. I., & Loftin, K. Exploring the potential value of satellite remote sensing to monitor chlorophyll-a for US lakes and reservoirs. *Environmental Monitoring and Assessment*, 2020, vol. 192, no. 12, article no. 808. DOI: 10.1007/s10661-020-08631-5.

13. Yailymov, B., Shelestov, A., Yailymova, H., & Shumilo, L. Google Earth Engine framework for satellite data-driven wildfire monitoring in Ukraine. *Fire*, 2023, vol. 6, no. 11, article no. 411. DOI: 10.3390/fire6110411.

14. Shelestov, A., Lavreniuk, M., Vasiliev, V., Shumilo, L., Kolotii, A., Yailymov, B., Kussul. N., & Yailymova, H. Cloud approach to automated crop classification using Sentinel-1 imagery. *IEEE Transactions on Big Data*, 2019, vol. 6, no. 3, pp. 572-582. DOI: 10.1109/TBDATA.2019.2940237.

15. Shelestov, A., & Kussul, N. Using the fuzzy-ellipsoid method for robust estimation of the state of a grid system node. *Cybernetics and Systems Analysis*, 2008, vol. 44, no. 6, pp. 847-854. DOI: 10.1007/s10559-008-9057-1.

16. Kussul, N., Kravchenko, A., Skakun, S., Adamenko, T., Shelestov, A., Kolotii, A., & Gripich, Y. Crop Yield Forecasting Regression Models based on MODIS Data. *Current problems in remote sensing of the Earth from space*, 2012, vol. 9, no. 1, pp. 95-107.

17. Mishra, S., & Mishra, D. Normalized difference chlorophyll index: A novel model for remote estimation of chlorophyll-a concentration in turbid productive waters. *Remote Sensing of Environment*, 2012, vol. 117, pp. 394-406. DOI: 10.1016/j.rse.2011.10.016.

18. Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., & Gascon, F. Sen2Cor for Sentinel-2. *Proceedings of Image and signal processing for remote sensing XXIII. SPIE,* 2017, vol. 10427, pp. 37-48. DOI: 10.1117/12.2278218.

19. Harmel, T., Chami, M., Tormos, T., Reynaud, N., & Danis, P. Sunglint correction of the Multi-Spectral Instrument (MSI)-SENTINEL-2 imagery over inland and sea waters from SWIR bands. *Remote Sensing of Environment*, 2018, vol. 204, pp. 308-321. DOI: 10.1016/j.rse.2017.10.022.

20. Gascon, F., Bouzinac, C., Thépaut, O., Jung, M., Francesconi, B., Louis, J., Lonjou, V., Lafrance, B., Massera, S., Gaudel-Vacaresse, A., Languille, Florie, F., Alhammoud, B., Viallefont, F., Pflug, B., Bieniarz, J., Clerc, S., Pessiot, L., Trémas, T., Cadau, E., De Bonis, R., Isola, C., Martimort, P., & Fernandez, V. Copernicus Sentinel-2A calibration and products validation status. *Remote Sensing,* 2017, vol. 9, no. 6, article no. 584. DOI: 10.3390/rs9060584.

21. Tavares, M. H., Lins, R. C., Harmel, T., Fragoso Jr, C. R., Martínez, J. M., & Motta-Marques, D. Atmospheric and sunglint correction for retrieving chlorophyll-a in a productive tropical estuarine-lagoon system using Sentinel-2 MSI imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2021, vol. 174, p. 215-236. DOI: 10.1016/j.isprsjprs.2021.01.021.

22. *Coriolis: In situ data for operational oceanography*. Available at: https://dataselection.coriolis.eu.org/ (accessed 29.01.2024).

23. IN SITU TAC INSITU_GLO_BGC_DISCRETE_MY_013_046 Available at: https://catalogue.marine.copernicus.eu/documents/QUID/CMEMS-INS-QUID-013-046.pdf (accessed 29.01.2024).

24. Saunders, P. Practical conversion of pressure to depth. *Journal of Physical Oceanography*, 1981, vol. 11, no. 4, pp. 573-574. DOI: 10.1175/1520-0485(1981)011<0573:PCOPTD>2.0.CO;2.

Received 04.02.2024, Accepted 15.04.2024

ПІДВИЩЕННЯ ПРОСТОРОВОГО РОЗРІЗНЕННЯ ХЛОРОФІЛУ-А В СЕРЕДЗЕМНОМУ МОРІ НА ОСНОВІ МАШИННОГО НАВЧАННЯ

Б. Я. Яйлимов, Н. М. Куссуль,
П. О. Геніцой, А. Ю. Шелестов

Предметом вивчення в статті є підвищення просторового розрізнення даних про рівень хлорофілу-а в Середземному морі за допомогою супутникових знімків та наземних вимірювань. **Метою** статті є розробка інформаційної технології на основі машинного навчання для створення карт концентрації хлорофілу-а з високим просторовим розрізненням для пілотних територій Середземного моря. Традиційні методи наземного вимірювання хлорофілу-а є трудомісткими, дорогими та мають обмежене просторове й часове покриття. Тому супутникові спостереження стали ефективним інструментом для моніторингу хлорофілу-а на великих територіях. Супутникові дані низького просторового розрізнення, такі як GCOM-C/SGLI та Sentinel-3 OLCI, дозволяють вимірювати концентрацію хлорофілу-а на поверхні моря. Однак, ці дані мають обмежену точність та просторову роздільну здатність, що створює виклики для моніторингу локальних змін у прибережних зонах та невеликих акваторіях. **Завдання:** проаналізувати наявні супутникові дані та наземні вимірювання хлорофілу-а для Середземного моря; дослідити залежності між супутниковими даними різного просторового розрізнення та наземними вимірюваннями; обрати інформативні ознаки з супутникових даних для побудови моделей машинного навчання; розробити моделі для підвищення просторового розрізнення хлорофілу-а на основі **регресійних алгоритмів** та алгоритми **машинного навчання**. **Отримані результати**: запропоновано інформаційну технологію, що поєднує супутникові дані з наземними вимірюваннями в хмарній платформі Google Earth Engine; досліджено кореляції між супутниковими вимірюваннями хлорофілу-а та наземними даними; побудовано моделі на основі Random Forest та Multilayer Perceptron з коефіцієнтами детермінації до 0,36 та кореляцією 0,6 з тестовими даними; створено карти хлорофілу-а з просторовим розрізненням 10 м для пілотної території біля Кіпру. **Розроблена інформаційна технологія дозволяє** ефективно поєднувати супутникові дані різного просторового розрізнення та наземні вимірювання для підвищення точності та деталізації карт хлорофілу-а в Середземному морі. Подальші дослідження передбачають вдосконалення попередньої обробки супутникових даних, використання більшої кількості ознак, залучення даних з інших регіонів та застосування складніших моделей машинного навчання.

Ключові слова: машинне навчання; супутникові дані; хлорофіл-а; хмарні технології, інформаційна технологія; iMERMAID.

Яйлимов Богдан Ялкапович – канд. техн. наук, зав. відділу космічних інформаційних систем та технологій, Інститут космічних досліджень НАН України та ДКА України, Київ, Україна.

Куссуль Наталія Миколаївна – д-р техн. наук, проф., зав. каф. математичного моделювання та аналізу даних, Навчально-науковий Фізико-технічний інститут НТУУ "КПІ ім. Ігоря Сікорського", Київ, Україна.

Геніцой Павло Олексійович – магістр каф. математичного моделювання та аналізу даних Фізико-технічний інститут НТУУ "КПІ ім. Ігоря Сікорського", Київ, Україна.

Шелестов Андрій Юрійович – д-р техн. наук, проф., проф. каф. математичного моделювання та аналізу даних, Навчально-науковий Фізико-технічний інститут НТУУ "КПІ ім. Ігоря Сікорського", Київ, Україна.

Bohdan Yailymov – Candidate of Technical Sciences, Head of the Department of Space Information Systems and Technologies, Space Research Institute NAS of Ukraine and SSA of Ukraine, Kyiv, Ukraine,
e-mail: yailymov@gmail.com, ORCID: 0000-0002-2635-9842, Scopus Author ID: 56335387700.

Nataliia Kussul – Doctor of Technical Sciences, Professor, Head of the Department of Mathematical Modeling and Data Analysis, Educational and Research Institute of Physics and Technology NTUU "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine,
e-mail: nataliia.kussul@gmail.com, ORCID: 0000-0002-9704-9702, Scopus Author ID: 6602485938.

Pavlo Henitsoi – Master's Student at the Department of Mathematical Modeling and Data Analysis, Educational and Research Institute of Physics and Technology NTUU "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine,
e-mail: pavlogenizoy@gmail.com, Scopus Author ID: 58079724400.

Andrii Shelestov – Doctor of Technical Sciences, Professor, Professor at the Department of Mathematical Modeling and Data Analysis, Educational and Research Institute of Physics and Technology NTUU "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine,
e-mail: andrii.shelestov@gmail.com, ORCID: 0000-0001-9256-4097, Scopus Author ID: 6507365226.