

Estimation of water quality parameters from space images

A K Popova

Matrosov Institute for System Dynamics and Control Theory SB RAS, 134
Lermontova Street, Irkutsk, 664033, Russia

chudnenko@icc.ru

Abstract. Water quality affects many human activities. Remote sensing is efficient and economical instrument for water monitoring. The paper investigates the problem of choosing an algorithm for Chl-a concentration determination. In this study, we made calculations for Multispectral Instrument (MSI) on Sentinel-2 for Lake Baikal by different empirical algorithms and C2RCC tool. We choose 3 band combination that have high correlation with in situ data of Chl-a. Resultant distribution map display spatial dynamics of Chl-a in the lake. Our research is intended to help environmental scientist to assess pollution level of the Lake Baikal and interpret the ecological meaning of results.

1. Introduction

In the traditional approach, water quality is estimated using water samples from different depths. The samples are analyzed in a laboratory to determine the physical and chemical properties of the water. There are also automatic stations and autonomous underwater vehicles capable of determining water parameters. Such methods are accurate but require investment of time, money, trained personnel, and laboratory equipment. In some areas, estimation of water quality is complicated by low transport accessibility.

These problems can be eliminated by using remote sensing, which provide spatial measurements in near real time in almost any area. With remote sensing we can detect, track and map water pollutants - oil and chemical spills, algal blooms and high concentrations of suspended solids. Satellite data has a wide temporal coverage and allows to estimate the quality over an extended water area. Traditional measurements are carried out from once a month to once every five years, and remote sensing have a frequency of survey from one day. Spatial coverage plays an important role in determining pollution in a reservoir as a whole, since one or more sampling points do not always accurately describe the state of the entire object [1-2].

2. Materials and Methods

Water quality can be determined by its optical properties, which depend on many factors, for example, on the amount of suspended organic and inorganic particles and dissolved substances. The most commonly used indicators for quality assessment are the concentration of chlorophyll-a (Chl-a) and the total suspended matter (TSM), the absorption coefficient of colored dissolved organic matter (CDOM), and the transparency of water measured by Secchi disk (SD). Lakes are optically more complex and diverse than sea or ocean waters, therefore these indicators can have large temporal and

spatial variations, depending on changes in weather, biological composition and physical properties of the lake [3-4].

Phytoplankton forms the basis of the aquatic food web and it is the main driver of biogeochemical processes in the ocean. In inland waters, phytoplankton biomass is an important parameter of water quality, since the abundance of algae can potentially indicate the degree of eutrophication in a particular water body. The main phytoplankton pigment is Chl-a, one of the most well-known parameters of water quality, therefore it is used as a simple substitute for phytoplankton biomass all over the world [5].

Lakes are increasingly subject to phytoplankton blooms due to eutrophication caused by cyanobacteria, which produce potentially lethal toxins that pose health risks. Also, phytoplankton bloom is one of the main factors affecting water transparency, and there is a negative correlation between these two parameters. Consequently, Chl-a is an important parameter for tracking eutrophication, lake ecological status and health risks from cyanobacterial blooms [6].

TSM determines the organic and mineral suspended solid particles that can come from rivers or rise from the bottom: living and dead phytoplankton, humic substances, clay minerals, detritus, etc. CDOM is a mixture of organic molecules produced by the decomposition of terrestrial vegetation, higher aquatic plants, phytoplankton, or bacteria. SD is an estimation of the water transparency with a Secchi disk, which is defined as the reflection of light from disk surface. All these parameters can be determined with satellite images.

Phytoplankton is an effective light absorber in blue and red bands due to the presence of intracellular pigments. Therefore, remote sensing of phytoplankton is usually based on the detection of its absorption of light. Currently, there are no generally accepted algorithms for accurately obtaining Chl-a in inland waters based on the reflection coefficient of remote sensing. Unlike the open oceans, inland waters are influenced by terrestrial materials such as humus and iron-rich minerals, which contribute significantly to the absorption and scattering of light in water. As a result, the standard blue-to-green bands ratio algorithms that are acceptable for open ocean are not suitable for inland waters.

The near infrared (NIR) bands are also used to determine Chl-a. The advantage of longer wavelengths is to reduce the effect of absorption of light by non-algae particulate matter and CDOM, as their absorption coefficients decrease exponentially with wavelength [7-8].

The choice of the Chl-a estimation algorithm depends on the optical properties of the water. Most remote sensing algorithms for assessing Chl-a concentration are based on the principles of changing light absorption depending on the amount of algae pigment. The calculations use the following properties of the reflection coefficients:

- the first peak of strong absorption in the blue region with a wavelength between 440 and 510 nm;
- minimum absorption in the green region between 550 and 555 nm;
- a second absorption peak in the red region between 670 and 675 nm;
- minimum absorption in the NIR region between 685 and 710 nm [9].

Algorithms based on two or three spectral band ratios are more preferable, since they help to reduce the influence of light, atmosphere and water surface [10].

Currently, water color data is provided by a number of satellite sensors: the geostationary ocean color imager (GOCI), the moderate resolution spectroradiometer (MODIS), the color ocean and land imaging sensor (OLCI), the Landsat operational land imager (OLI), and the Sentinel-2 multispectral instrument (MSI). All these sensors have suitable spectral and spatial resolution, capable of detecting pollutants and assessing water quality parameters. MSI and OLCI sensors are more often used in research now.

In the work we used Sentinel-2 MSI images to determine the amount of chlorophyll in Lake Baikal. Sentinel-2 was launched under the European Space Agency Copernicus program in 2015 (Sentinel-2A) and 2017 (Sentinel-2B). Both carry a MultiSpectral Instrument (MSI) with a multispectral imager with 13 spectral bands (Table 1), covering channels from blue to short wave infrared (SWIR) with a resolution of 10 to 60 m. Provides global coverage every 5 days.

Table 1. Spectral bands for the Sentinel-2 sensors.

| Band | Central wavelength (nm) |
|------------------------------|--------------------------------|
| Band 1 – Coastal aerosol | 442 |
| Band 2 – Blue | 492 |
| Band 3 – Green | 559 |
| Band 4 – Red | 665 |
| Band 5 – Vegetation red edge | 705 |
| Band 6 – Vegetation red edge | 740 |
| Band 7 – Vegetation red edge | 782 |
| Band 8 – NIR | 832 |
| Band 8A – Narrow NIR | 864 |
| Band 9 – Water vapour | 945 |
| Band 10 – SWIR – Cirrus | 1373 |
| Band 11 – SWIR | 1613 |
| Band 12 – SWIR | 2202 |

S2 MSI has high potential for monitoring Chl-a in coastal and inland waters due to the red channel at around 705 nm (band 5) and near the second absorption peak of Chl-a in the red band (band 4, 665 nm).

3. Results and Discussion

We took data of the Chl-a concentration near the village Bolshie Koty (105°04' E, 51°53' N) from open sources as in situ data [11]. This is data from an automatic station for measuring water quality parameters; it collects data obtained from sensors of dissolved oxygen, hydrogen index (pH), redox potential, chlorophyll-a and other indicators of the lake Baikal. Measurement results are available on the server [14].

We downloaded Sentinel-2 MSI images for the southern end of Lake Baikal for July 2017 from Copernicus Open Access Hub. The calculations were performed in the SNAP. SNAP is an open-source software with toolboxes for processing of Earth Observation data from Sentinel satellites.

Satellite images were previously preprocessed: all MSI bands were previously resampled to 20 m and passed atmospheric correction by the algorithm Case-2 Regional CoastColour (C2RCC). Atmospheric correction is an essential procedure as for 90% of the signal that reaches the image sensor is affected by the absorption and scattering by aerosols and different particles in the atmosphere (water vapor, carbon dioxide, ozone). The retrieval of water constituents, or its optical properties, is achieved by inversion of the water leaving reflectance spectrum, measured at top of atmosphere and thus requiring a correction for atmospheric effects [12].

The C2RCC processor relies on a large database of simulated water leaving reflectances and related top-of-atmosphere radiances. Neural networks are trained in order to perform the inversion of spectrum for the atmospheric correction, i.e. the determination of the water leaving radiance from the top of atmosphere radiances, as well as the retrieval of inherent optical properties of the water body. Also, C2RCC gives outputs results of chl-a and TSM and provides the possibility to add additional background information such as salinity, elevation, ozone, temperature, and air pressure [13].

Based on the literature overview, 13 empirical algorithms were selected. There are the various combinations of the ratios of two and three S2 MSI bands. In table 2 we present the results of calculations by empirical algorithms and by built-in function of C2RCC. Column “Range of values” presents min and max values obtained in calculations for the whole image, column “Value at Bolshie Koty” presents the value at the point of in situ data.

The mean absolute percentage error (MAPE) is a measure of prediction accuracy of a forecasting method in statistics, for example in trend estimation, also used as a loss function for regression problems in machine learning. It usually expresses the accuracy as a ratio defined by the formula:

$$M = \frac{1}{n} \sum_{i=1}^n \left| \frac{x - y_i}{x} \right|$$

where x is the actual value and y is the forecast value. Last column in table 2, “M” is MAPE for each algorithm. The larger M value mean the larger error of algorithm.

Table 2. Comparison of results.

| Empirical algorithms | Range of values | Value at Bolshie Koty | M |
|---|-----------------|-----------------------|--------|
| C2RCC MSI | 0,009...28,87 | 0,54 | 0,5 |
| R_{665}/R_{705} | 0,4...1,258 | 1,11 | 0,0277 |
| R_{705}/R_{740} | 0,26...1,38 | 0,94 | 0,129 |
| R_{665}/R_{740} | 0,12...1,582 | 0,98 | 0,0925 |
| R_{740}/R_{665} | 0,63...8,2 | 1,04 | 0,037 |
| $\log(R_{705}/R_{665})$ | -0,23...0,885 | -0,11 | 1,101 |
| $R_{740}/R_{705} - R_{705}/R_{665}$ | -0,32...2,157 | 0,22 | 0,796 |
| R_{560}/R_{665} | 0,67...1,983 | 1,59 | 0,472 |
| R_{783}/R_{665} | 0,45...10,97 | 0,80 | 0,259 |
| $(R_{490} - R_{665})/R_{560}$ | -0,548...1,064 | 0,84 | 0,22 |
| R_{705}/R_{665} | 0,596...10,2 | 0,94 | 0,129 |
| $(R_{490} - R_{665})/(R_{560} - R_{665})$ | -142,9...195,03 | 2,19 | 1,027 |
| $(R_{665}^{-1} - R_{705}^{-1}) * R_{740}$ | -0,17...4,82 | -0,10 | 1,09 |

The best result with smallest error was obtained for the ratios of bands R_{665}/R_{705} , in second place – R_{740}/R_{665} , and third is R_{665}/R_{740} . All these algorithms use red and vegetation red edge bands with 665, 705, 740 nm wavelength. At the same time C2RCC showed values far from in situ as well as some other algorithms, although they were good in research on European lakes [3, 6].

Maps of chlorophyll distribution in the South Baikal and Angara river were constructed in SNAP. Figure 1 shows result of the calculations for some algorithms. Blue color matches the low values and dark red – high values for calculated chl-a.

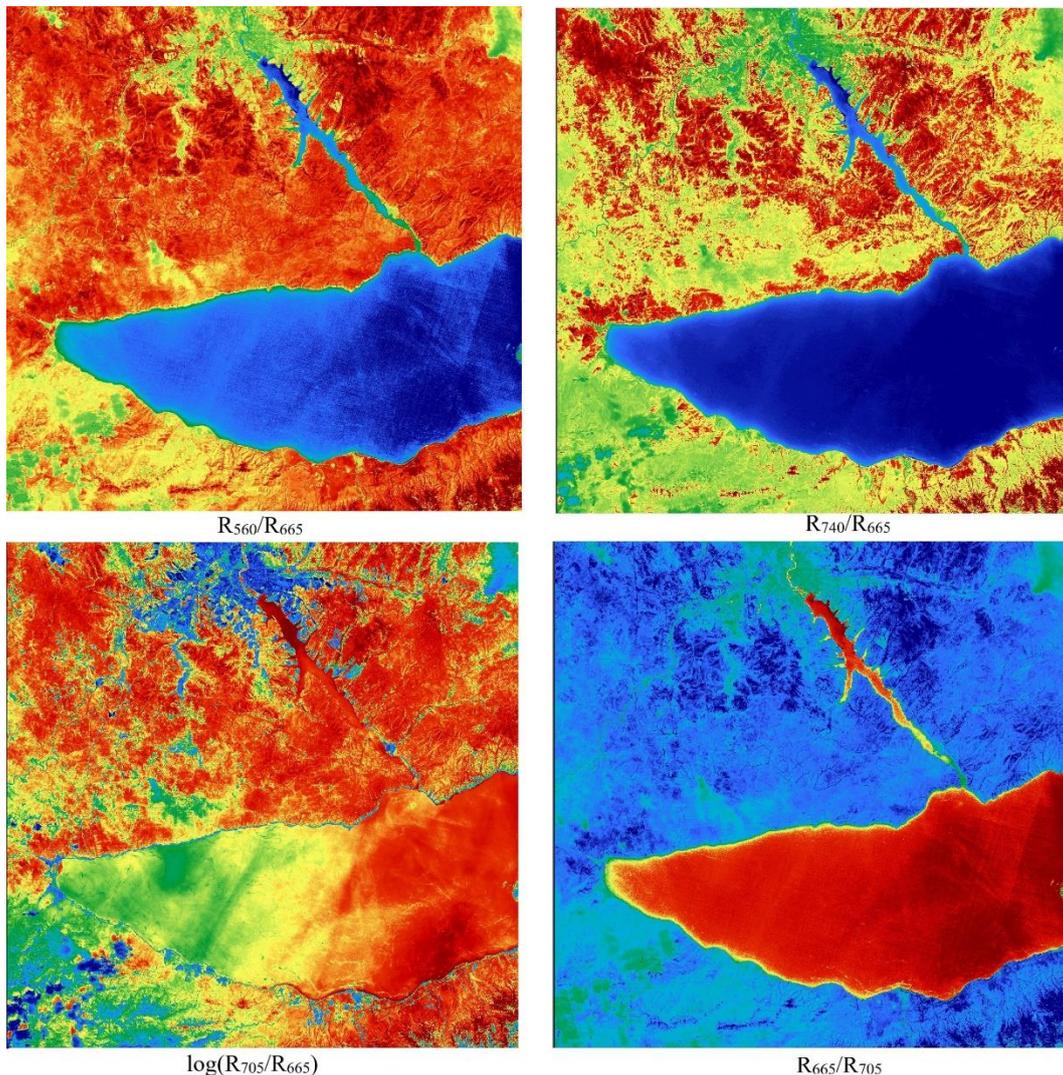


Figure 1. Maps of chlorophyll distribution in South Baikal calculated by empirical algorithms.

4. Conclusion

Remote sensing is an effective method for water bodies monitoring. It helps to manage rivers, lakes and marine coastal area and support decision making by quickly measurement of various water quality parameters. Real-time or near-real-time measurements of water pollutants and toxins at various spatial scales are necessary to monitor and manage environmental impacts and understand the processes governing their spatial distribution.

The purpose of the study is to explore algorithms for chlorophyll-a concentration calculations for Lake Baikal. It did not include an assessment of the pollution level of the lake and the ecological meaning of the results. We hope that the chosen algorithms will help environmental scientists to estimate these parameters.

In the future, we plan to expand the studies with calculations on the images from Sentinel-3 with the OLCI sensor, which has more bands in the red and NIR spectrum, and to calculate other water quality parameters for Lake Baikal.

Acknowledgments

The results were obtained within the framework of the State Assignment of the Ministry of Education and Science of the Russian Federation for the project "Methods and technologies of cloud-based service-oriented platform for collecting, storing and processing large volumes of multi-format interdisciplinary data and knowledge based upon the use of artificial intelligence, model-guided approach and machine learning" (state registration number 121030500071-2). Some results were obtained using the facilities of the Centre of collective usage "Integrated information network of Irkutsk scientific educational complex".

References

- [1] Hafeez S, Sing Wong M, Abbas S, Y. T. Kwok C, Nichol J, Ho Lee K, Tang D and Pun L 2019 Detection and Monitoring of Marine Pollution Using Remote Sensing Technologies *Monit. Mar. Pollut.*
- [2] Topp S N, Pavelsky T M, Jensen D, Simard M and Ross M R V 2020 Research trends in the use of remote sensing for inland water quality science: Moving towards multidisciplinary applications *Water (Switzerland)* **12** 1–40
- [3] Soomets T, Uudeberg K, Jakovels D, Brauns A, Zagars M and Kutser T 2020 Validation and Comparison of Water Quality Products in Baltic Lakes Using Sentinel-2 MSI and Sentinel-3 OLCI Data *Sensors* **20** 742
- [4] Toming K, Kutser T, Uiboupin R, Arikas A, Vahter K and Paavel B 2017 Mapping water quality parameters with Sentinel-3 Ocean and Land Colour Instrument imagery in the Baltic Sea *Remote Sens.* **9**, 10
- [5] Alikas K, Kangro K and Reinart A 2010 Detecting cyanobacterial blooms in large North European lakes using the maximum chlorophyll index *Oceanologia* **52** 237–57
- [6] Anspér A and Alikas K 2019 Retrieval of chlorophyll a from Sentinel-2 MSI data for the European Union water framework directive reporting purposes *Remote Sens.* 2019 **11**, 64
- [7] Zheng G and DiGiacomo P M 2017 Remote sensing of chlorophyll-a in coastal waters based on the light absorption coefficient of phytoplankton *Remote Sens. Environ.* **201** 331–41
- [8] Filipponi F 2018 River Color Monitoring Using Optical Satellite Data *Proceedings* **2** 569
- [9] Ha N T T, Thao N T P, Koike K and Nhuan M T 2017 Selecting the best band ratio to estimate chlorophyll-a concentration in a tropical freshwater lake using sentinel 2A images from a case study of Lake Ba Be (Northern Vietnam) *ISPRS Int. J. Geo-Information* **6** 290
- [10] Conopio M, Japor R K, Blanco A C and Tamondong A M 2019 Estimation of chlorophyll-a concentration in Laguna de Bay using Sentinel-3 satellite data *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. - ISPRS Arch.* **42** 125–32
- [11] Shimaraeva, S.V., Pislegina, E.V., Krashchuk, L.S. et al. Dynamics of chlorophyll a concentration in the South Baikal pelagic during the direct temperature stratification period. *Inland Water Biol* 10, 59–63 (2017). <https://doi.org/10.1134/S1995082917010163>
- [12] Pereira-Sandoval M, Ruescas A, Urrego P, Ruiz-Verdú A, Delegido J, Tenjo C, Soria-Perpinyà X, Vicente E, Soria J and Moreno J 2019 Evaluation of atmospheric correction algorithms over Spanish inland waters for sentinel-2 multi spectral imagery data *Remote Sens.* **11** 1–23
- [13] Brockmann C, Doerffer R, Peters M, Stelzer K, Embacher S and Ruescas A 2018 Evolution of the C2RCC neural network for Sentinel 2 and 3 for the retrieval of ocean colour products in normal and extreme optically complex waters *J. Mater. Process. Technol.* **1** 1–8
- [14] Data server of Baikal water parametrs. Income accessed online on 22th December 2020 via <https://hlserver.lin.irk.ru/shs/rinko/>