UNIVERSITY OF TARTU
Faculty of Science and Technology
Institute of Computer Science
Data Science Curriculum

Muhammad Kumail Raza, Sushant Aryal, Anderson Isaac Guamán Viveros

Spatio-Temporal Changes in Optical Water Quality Parameters

Data Science in Remote Sensing (LTTO.00.027)

Supervisors:
Krista Alikas
Fjodor Ševtšenko

Tartu 2024

Abstract

Water bodies on land are critical ecosystems that play an important role in different domains. Changes in the characteristics of water bodies can influence internal processes, potentially causing long-term negative impacts on ecosystem features and biodiversity. A common indicator of degradation in lakes is brownification. This research focuses on Colored Dissolved Organic Matter (CDOM), which, at high concentrations, turns water brown and alters underwater light conditions. This phenomenon significantly affects ecosystem characteristics and diversity.

To monitor and understand these changes, it is crucial to represent the state of water bodies using a brownification scale. Previous studies have addressed water body classification through Optical Water Types (OWT).

This study examines the spatio-temporal dynamics of optical water quality in Estonia's Lake Võrtsjärv, emphasizing the key parameter CDOM. It builds upon CDOM-based Optical Water Type classification, expanding the analysis to a broader spatio-temporal range. Additionally, the study investigates how CDOM levels vary with factors such as wind speed, temperature, and solar radiation, using data recorded from stations near Lake Võrtsjärv.

Keywords: Water bodies, Colored Dissolved Organic Matter (CDOM), Optical Water Types (OWT), Spatio-temporal dynamics, brownification.

Table of Contents

Abstract **2**
1 Introduction **4**
2 Background **5**
3 Study Area **6**
4 Dataset **7**
4.1 Sentinel-2 Imagery 7
4.2 Meteorological Data 8
5 Methods **9**
5.1 Pixel-wise Classification into CDOM-OWT Classes 9
5.2 Temporal Analysis of CDOM Concentration and Meteorological Data 10
6 Results **11**
6.1 Temperature and CDOM 16
6.2 Solar Energy and CDOM 16
6.3 Precipitation and CDOM 16
6.4 Wind Speed and CDOM 16
6.5 Differences between Stations 16
7 Discussions **17**
8 Conclusions **17**
9 References **18**
10 Annexes **18**

1 Introduction

Water bodies on land are critical ecosystems that play an important role in different domains. These natural elements (i.e., rivers, lakes, and wetlands) provide drinking water, support biodiversity, and even help to mitigate natural disasters such as floods and droughts (Grigg, 2011). Despite their significance, these ecosystems are under threat globally due to shrinking and pollution caused by global warming and anthropogenic activities.

A common indicator of degradation in lakes is brownification. This process is caused by high concentration of both organic and non-organic particles. Among the principal agents that contribute to this problem are high concentrations of Total Suspended Matter (TSM), Chlorophyll-a (Ch-a) and Colored Dissolved Organic Matter (CDOM) (Dubois, 2011). Although this is a natural process, anthropogenic activities such as agriculture, mining, deforestation, increase in urbanization, and others, intensify this phenomenon.

This project focuses on CDOM, a key water quality component. CDOM consists of organic molecules derived from the remains of plants and animals altered through chemical reactions. Both human activities, such as agriculture and deforestation, and natural processes, like soil runoff and plant decay, can increase CDOM concentrations. High levels of CDOM lead to water brownification, a phenomenon that alters light conditions, affects aquatic habitats, and impacts ecosystem diversity. CDOM particles absorb incoming light, reducing the light available to aquatic organisms and negatively affecting their behavior and habitat conditions (Madhav et al., 2023).

To assess water quality, it is essential to represent the state of water bodies using a brownification scale. Traditional methods involve collecting in-situ measurements of chemical and optical properties like CDOM. While precise, this approach faces challenges in spatial and temporal coverage, cost, and accessibility. Remote sensing offers a practical alternative, allowing for large-scale monitoring of water bodies.

Remote sensing involves using satellite-based sensors, such as those on the Landsat and Sentinel missions, to gather data on water quality. By measuring light reflectance from water surfaces, the status of brownification can be determined. The reflectance values are influenced by optically active substances' absorption and scattering properties, including CDOM, chlorophyll-A, and total suspended solids (TSS) (Arst, 2003). Each water body exhibits unique combinations of these substances, and analyzing reflectance across multiple spectral bands allows for accurate classification of Optical Water Types (OWTs).

This project aims to understand water quality and its variability over time. A key objective is to explore the application of remote sensing and data science techniques in water quality studies. This includes leveraging machine learning (ML) methods, focusing on AutoML, to simplify and enhance classification processes. The project involves acquiring data from sources like ESTHUB, preprocessing and normalizing the data for ML analysis, and classifying satellite data to derive meaningful insights over a sizable temporal range. The project aims to develop a comprehensive approach for monitoring and analyzing water quality by integrating advanced remote sensing and ML techniques.

2 Background

The concept of water type classification was first introduced by Nils Jerlov in 1951, who categorized ocean waters on a clear-to-turbid scale using the downwelling diffuse attenuation coefficient (Jerlov, 2014). By 1976, ten water types were defined within this framework. Morel and Prieur (1977) further developed water classification based on reflectance measurements, identifying two cases (Das, 2023): Case 1, dominated by phytoplankton, and Case 2, characterized by inorganic particles, using reflectance data and optical coefficients.

Reinart et al. (2003) applied OWT classification to lakes and coastal waters in Estonia and Finland, focusing on Case 2 water types. Using K-Means clustering, they defined five optical classes: clear, moderate, turbid, very turbid, and brown. Similarly, Moore et al. (2009) expanded the classification to eight water types based directly on radiance measurements.

Spyrakos et al. (2018) identified 13 spectrally distinct clusters for inland waters and nine for marine environments, using clustering algorithms and functional analysis. Uudeberg et al. (2019) classified boreal lakes and coastal areas into five OWT classes using mathematical clustering rules. This work was extended in 2020 to map optical water quality parameters to empirical algorithms, addressing the inefficiencies of traditional water monitoring.

In Brazil, da Silva et al. (2020) used in-situ reflectance data and K-Means clustering for OWT assessment. Their findings were later applied to Sentinel-2 MSI data (2021), employing Support Vector Machines for classification with novelty detection techniques (da Silva et al., 2021).

Recent advances, such as those by Ševtšenko (2024), utilized a series of machine learning (ML) algorithms for supervised OWT classification. This approach leveraged AutoML techniques to streamline the classification process, linking OWT categories to CDOM-based classes through robust evaluation methods.

3 Study Area

Lake Võrtsjärv, located in southern Estonia, is the second-largest lake in the country by surface area and plays a vital role in the region's ecosystem and local livelihoods. It is a shallow, eutrophic freshwater lake with an average depth of approximately 2.8 meters and a maximum depth of about 6 meters. The lake covers an area of around 270 square kilometers, making it one of the largest inland water bodies in the Baltic region (Moora et al., 2002).

Lake Võrtsjärv is fed by multiple small rivers and drains into the Emajõgi River, which eventually flows into Lake Peipsi. The lake is a crucial habitat for various aquatic species, including fish, aquatic plants, and migratory birds. Its surrounding wetlands and reed beds provide nesting grounds and biodiversity support (Moora et al., 2002).

The lake experiences significant seasonal variability due to its shallow depth and geographic location. During the winter months, it typically freezes, while in warmer months (April to October), it becomes an active site for ecological processes such as photosynthesis and nutrient cycling (Pettersson et al., 2010). This period was chosen for the study as it allows for consistent monitoring and analysis of water quality.

Lake Võrtsjärv is characterized by its high concentration of dissolved organic matter, suspended sediments, and nutrient load, largely influenced by agricultural activities and natural processes in its catchment area (Moora et al., 2002). These factors make it a representative study site for exploring the dynamics of Optical Water Types (OWTs) and Colored Dissolved Organic Matter (CDOM) classification. Understanding the lake's water quality is critical for ecological conservation, sustainable management, and mitigating the impacts of anthropogenic activities.

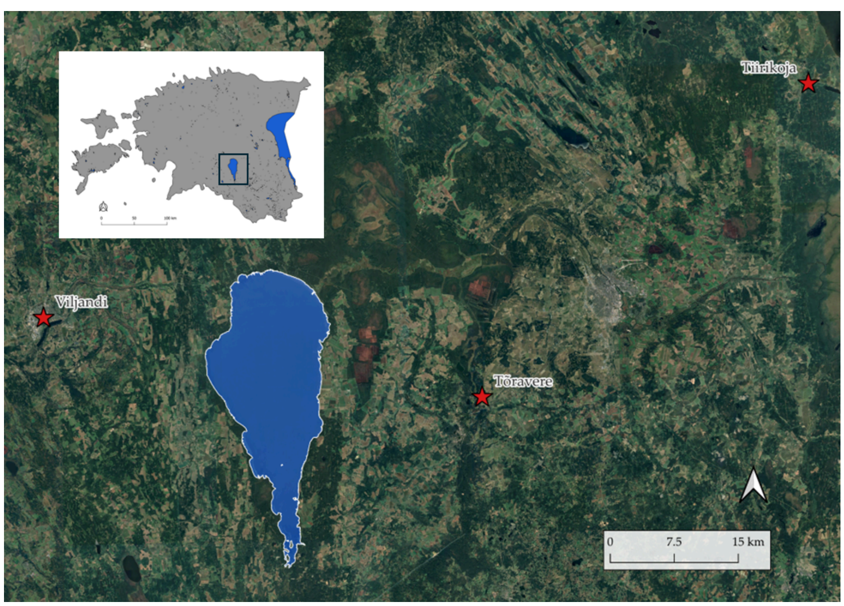

Figure 1 Study Area

4 Dataset

4.1 Sentinel-2 Imagery

The dataset utilized in this study comprised Sentinel-2 (S2) satellite imagery sourced from ESTHub (Estonian Land Board, n.d.). The images were processed at the Top-of-Atmosphere (TOA) level and corrected using the Polymer L2 MSI v4.16.1 processor, specifically designed for scientific use. To ensure high-quality reflectance data, ancillary data from NCEP (National Centers for Environmental Prediction) were integrated to support the atmospheric correction process.

The study's Area of Interest (AOI) was Lake Võrtsjärv in Estonia. Data were collected from 2016 to 2023, focusing on April to October. This time frame was selected as the lake remains unfrozen during these months, facilitating the observation of water quality parameters.

Spectral data from Sentinel-2 bands 1 through 8A were included in the analysis. These bands were normalized to ensure uniformity and compatibility with machine learning algorithms.

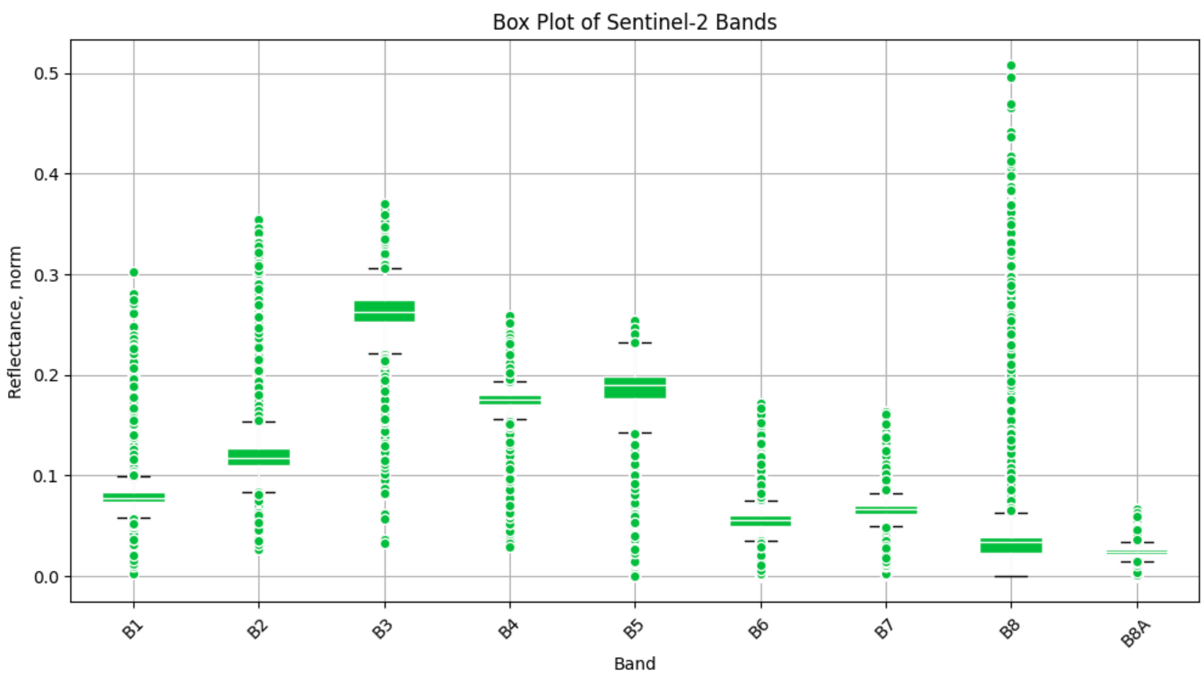

Figure 2. Normalized Features of Sentinel-2 Imagery

This dataset forms the foundation for studying the spatio-temporal dynamics of water quality in Lake Võrtsjärv, emphasizing analyzing Optical Water Types (OWTs) and Colored Dissolved Organic Matter (CDOM).

4.2 Meteorological Data

In this study, meteorological data from weather stations located at Tartu-Tõravere, Tiirikoja, and Viljandi was incorporated to further analyze the environmental factors influencing the optical water quality of

Lake Võrtsjärv. The dataset included hourly measurements of variables such as radiation, air pressure, precipitation, humidity, temperature, and wind parameters. These measurements were analyzed and processed to derive monthly averages, providing a comprehensive temporal perspective.

The integration of weather station data allowed for a deeper exploration of the relationships between meteorological conditions and the lake's optical properties. By combining this data with satellite observations, this study investigated how factors like temperature fluctuations, wind patterns, and radiation levels impact the dynamics of Colored Dissolved Organic Matter (CDOM) and other water quality parameters over time. This approach provides a more holistic understanding of the interactions between atmospheric conditions and the lake's ecosystem.

5 Methods

5.1 Pixel-wise Classification into CDOM-OWT Classes

After preprocessing the data, the study shifted to applying machine learning algorithms for classifying Optical Water Types (OWTs) and analyzing key water quality parameters, such as Colored Dissolved Organic Matter (CDOM). The classification process was influenced by the methodology outlined in Ševtšenko (2024) on CDOM-based Optical Water Type classification and adapted to suit the specific characteristics of Lake Võrtsjärv. The normalized reflectance values from Sentinel-2 bands 1 through 8A formed the input for this phase, enabling the identification of optical water clusters and their variations over time.

The classification was performed using a K-Means clustering algorithm, optimized through the AutoGluon framework, to divide the data into eight distinct clusters. These clusters corresponded to unique Optical Water Types, capturing variations in optical properties across Lake Võrtsjärv. The eight clusters were selected based on prior studies and refined through evaluation metrics to ensure meaningful classification. A Weighted Ensemble (L3) classifier further enhanced accuracy and robustness. This approach combined multiple models to improve the reliability of predictions, ensuring accurate separation of water types across diverse spatial and temporal datasets.

The performance of the clustering algorithm was evaluated using several metrics. The silhouette score, which measures the separation between clusters, was 0.308, indicating moderate differentiation among the identified clusters. Balanced accuracy was high, with values of 0.979 for the training split and 0.963 for the test split, demonstrating the model’s reliability across different data subsets. The Macro F1-Score and Micro F1-Score were also notable, at 0.970 and 0.971, respectively, underscoring the precision and recall of the clustering process.

To ensure the robustness of the classification, penalties were incorporated to assess errors in the predicted order of key water quality parameters, including CDOM, chlorophyll-a (Chla), total suspended solids (TSS), and Secchi depth. The CDOM order penalty was 0.000, indicating perfect alignment with expected classifications, while the penalties for Chla and TSS were 0.125 each, reflecting minimal deviations. Similarly, the Secchi depth penalty was 0.000, validating the model's accuracy.

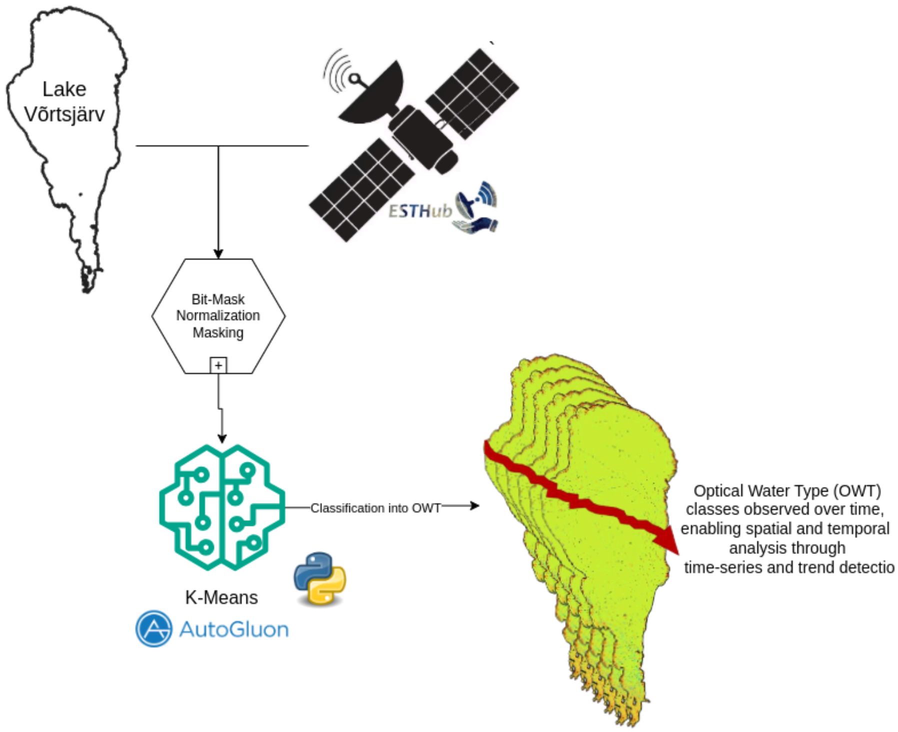

Figure 3. Methodology Diagram for Classification

The classification results were applied to generate time-series data and spatial maps of Optical Water Types and CDOM concentrations across Lake Võrtsjärv. This allowed for a detailed analysis of water quality trends and seasonal dynamics between 2016 and 2023. By leveraging unsupervised clustering with the K-Means algorithm and integrating AutoML tools, the methodology provided a robust framework for analyzing large-scale water quality datasets. The approach demonstrated the effectiveness of combining data-driven machine learning techniques with domain knowledge, offering insights into water quality parameters' spatial and temporal behavior in large freshwater ecosystems. By adapting Fyodor Ševtšenko's framework, the study successfully developed a scalable method for monitoring and understanding water quality in Lake Võrtsjärv.

5.2 Temporal Analysis of CDOM Concentration and Meteorological Data

The analysis of CDOM concentrations and meteorological data was conducted temporally, focusing on the percentage of CDOM concentration across the lake. To comprehensively understand, meteorological data was collected from three weather stations at Tartu-Tõravere, Tiirikoja, and Viljandi, strategically positioned around Lake Võrtsjärv. These stations were selected to capture the spatial variability of atmospheric conditions influencing different parts of the lake.

Using the data from these stations, comparisons were made between temporal trends in meteorological variables such as temperature, wind speed, radiation, and precipitation, and the average CDOM concentrations across the lake. The analysis was carried out in Jupyter Notebook, where visualizations, comparison matrices, and correlation matrices were created to explore and quantify these relationships.

6 Results

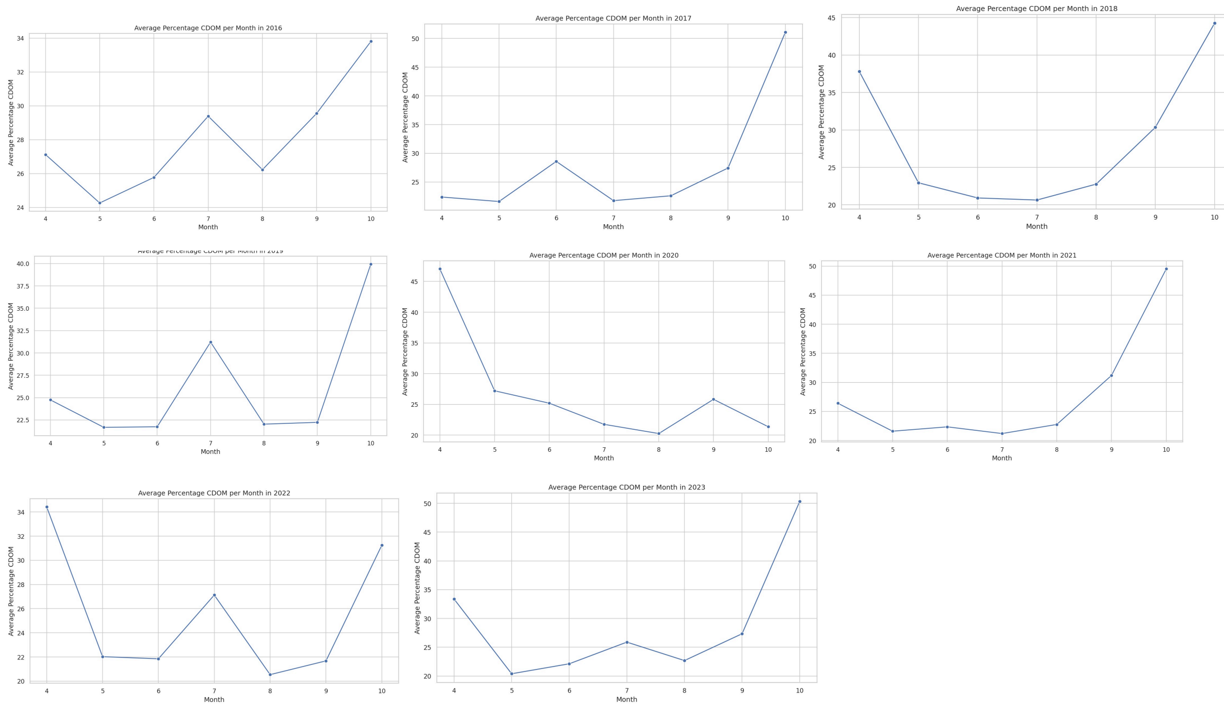

Figure 4. CDOM Concentrations Every Month 2016 - 2023

Figure 4 presents line graphs showing the average monthly percentage of CDOM (Chromophoric Dissolved Organic Matter) in Lake Võrtsjärv across multiple years. Each year’s curve generally exhibits a characteristic “U” shape, with relatively moderate CDOM values at the start of the year, followed by a notable decline mid-year, and then a pronounced rise toward the year’s end. While the specific magnitudes and timing of these fluctuations vary somewhat from year to year, the recurring pattern suggests that overarching seasonal processes, such as changes in precipitation regimes, temperature variations, and the seasonal input of organic matter, consistently influence the temporal distribution of CDOM in this lake system.

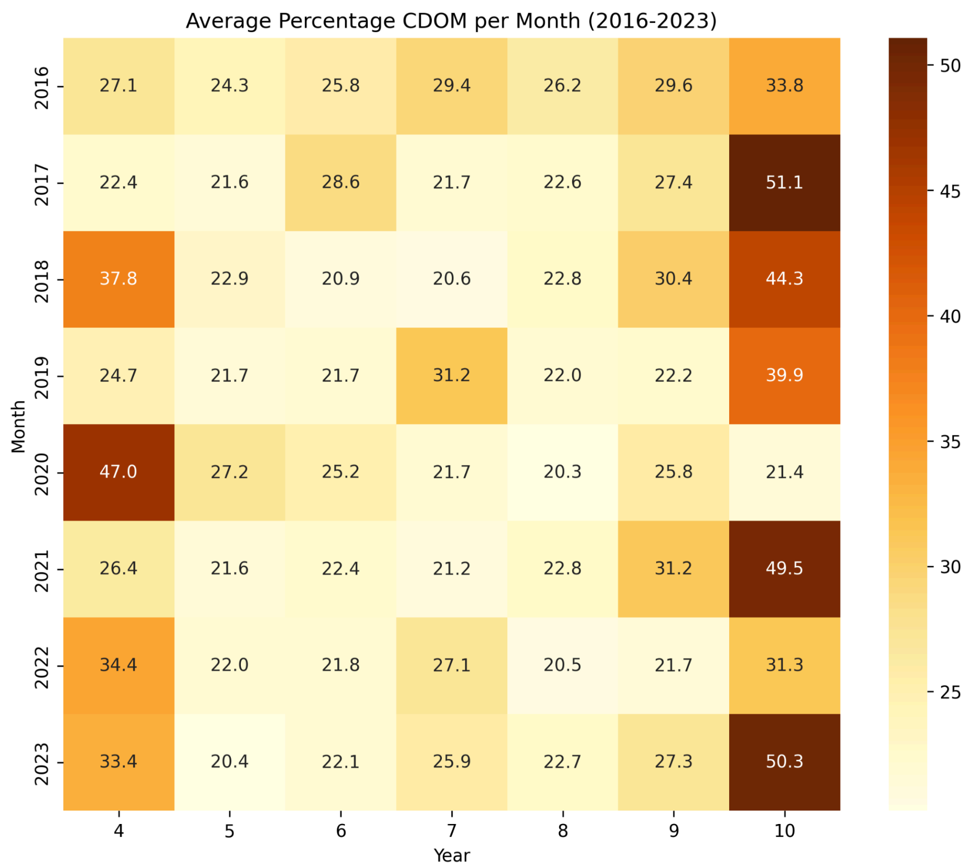

Figure 5. Average monthly percentage of CDOM Concentration 2016 - 2023

A noticeable feature in the data is that certain months and years show particularly high CDOM values. For instance, some years exhibit peaks in CDOM during the late summer or early autumn months, which may coincide with periods of increased runoff from surrounding lands, enhanced decomposition of organic materials in the water, or changes in precipitation and temperature that influence the input and breakdown of organic matter. In contrast, other months typically present lower CDOM levels, suggesting times of the year when environmental conditions are not as conducive to the buildup of dissolved organic matter or when photodegradation and microbial activity may reduce CDOM concentrations.

Over time, these monthly patterns can vary from year to year. The differences might be driven by changes in weather patterns, land use, or other environmental factors, such as shifts in local vegetation or hydrology. For example, a particularly wet year could increase runoff and lead to higher CDOM levels during certain months, while a dry or sunny period might promote more photodegradation and lower CDOM readings.

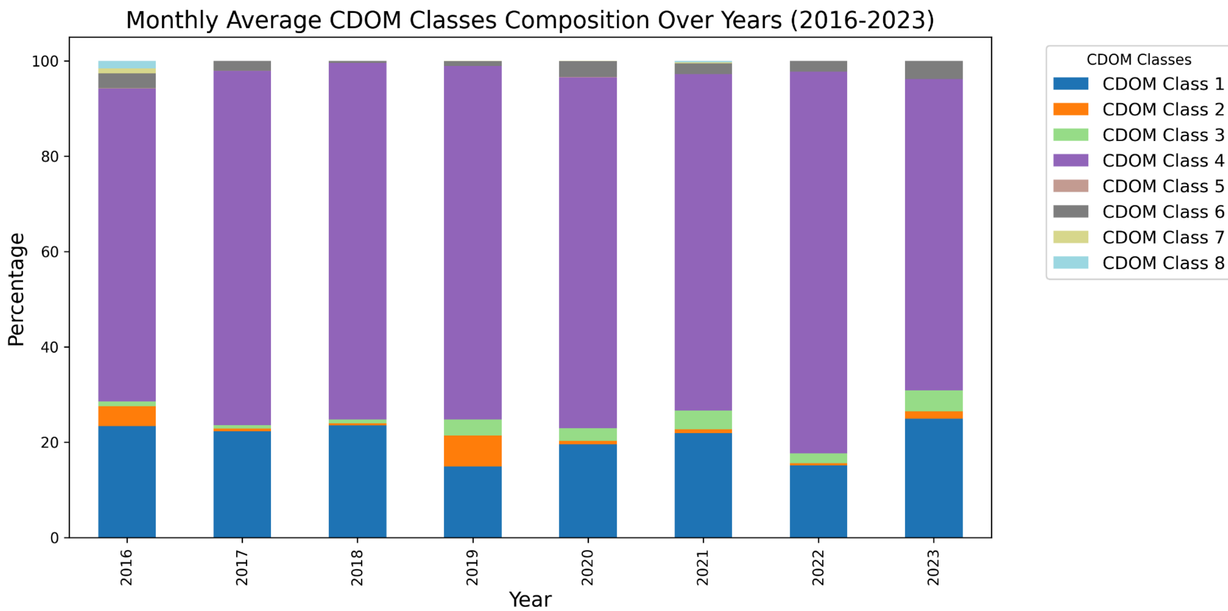

Figure 6. Average CDOM Composition Over Years

Figure 5 illustrates the relative composition of various CDOM classes over the period from 2016 to 2023, presenting the average monthly percentage contributions of each class. The stacked bar chart highlights that while multiple classes are present, the overall composition is dominated by two or three major CDOM classes. Over time, these dominant classes remain fairly stable, with only subtle variations in their proportional contributions. Minor classes, although present, persist at comparatively low percentages and do not exhibit significant long-term trends. This consistency in the proportional makeup of CDOM suggests that the underlying environmental conditions and biogeochemical processes influencing the lake's organic matter inputs and transformations have remained relatively steady throughout the observed timeframe.

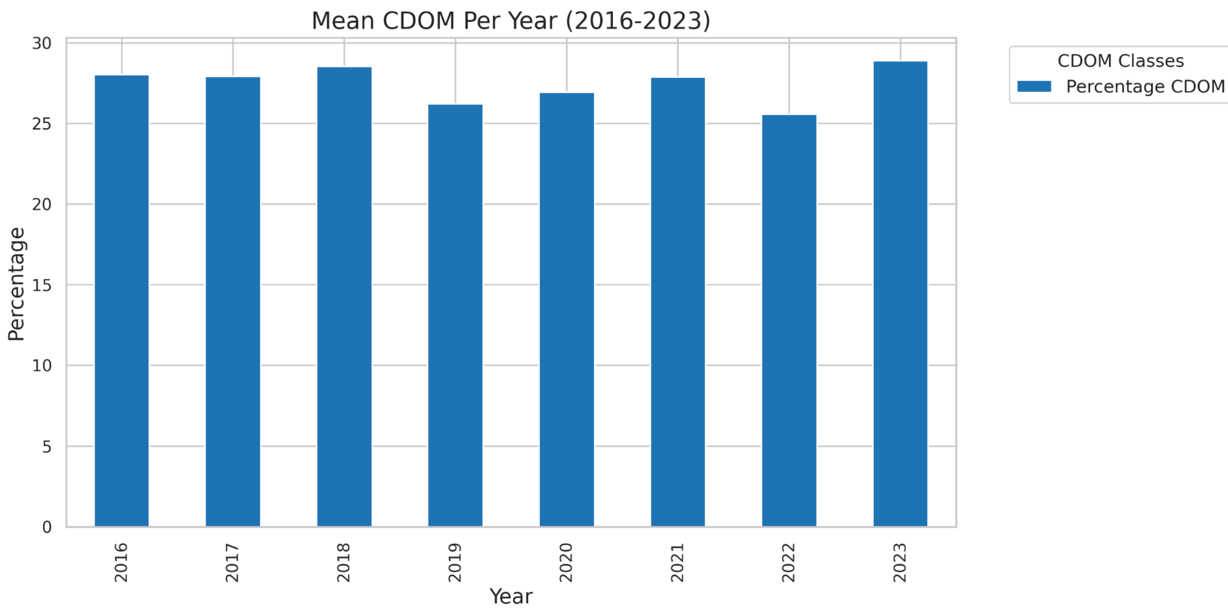

Figure 7. Mean CDOM Over Yearsimplies

The charts Figure 6 and 7 imply that, on an annual basis, CDOM levels have remained broadly consistent from 2016 through 2023. While the data do not point to dramatic long-term shifts, the slight annual differences may still offer insights into how local climate conditions, hydrology, or watershed management practices influence the presence and concentration of dissolved organic matter in the environment.

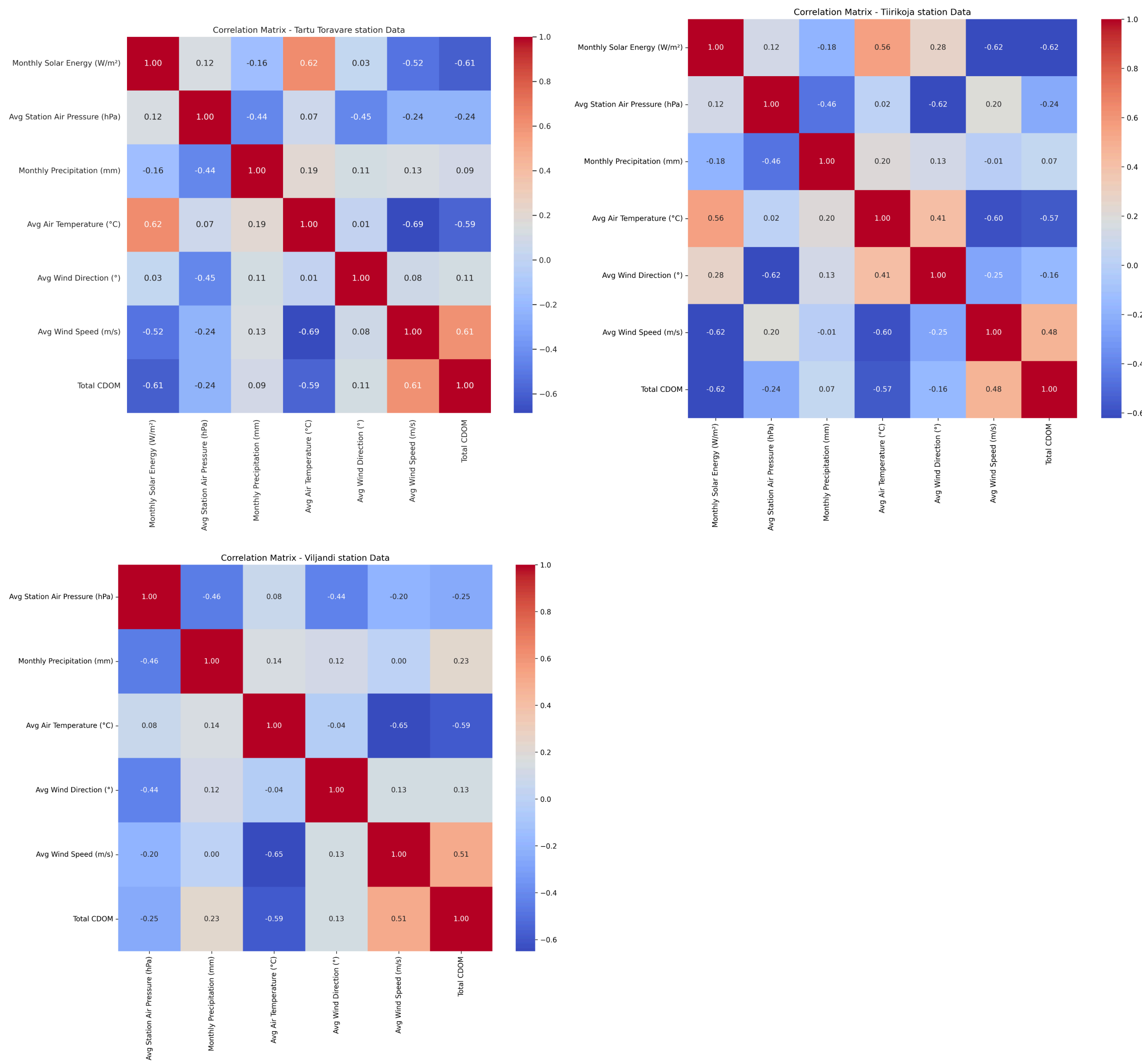

Figure 8. Correlation matrices show the statistical relationships between Total CDOM (Chromophoric Dissolved Organic Matter) concentrations and Meteorological Data from Different Stations

Monthly solar energy, measured in watts per square meter (W/m²), represents the average solar energy received during the month. The average station air pressure, expressed in hectopascals (hPa), indicates the local atmospheric pressure recorded at the station. Monthly precipitation, measured in millimeters (mm), reflects the monthly rainfall. The average air temperature, calculated in degrees Celsius (°C), provides the mean temperature for the same period. Similarly, the average wind direction, expressed in degrees (°), highlights the predominant direction of the wind, while the average wind speed, measured in meters per second (m/s), captures the mean wind velocity. Additionally, the total concentration of colored dissolved organic matter (CDOM) in the water offers insights into water quality. These metrics allow for analyzing general patterns and provide valuable interpretations of environmental conditions.

6.1 Temperature and CDOM:

At multiple stations, there tends to be a negative correlation between Total CDOM and Average Air Temperature. This might mean that during warmer months, CDOM levels are generally lower. One possible explanation is that higher solar irradiation and temperature can lead to increased photodegradation (CDOM breaks down under strong sunlight) or differences in seasonal runoff patterns that affect when organic matter enters the water.

6.2 Solar Energy and CDOM:

Often, Total CDOM shows a negative correlation with Monthly Solar Energy. Higher solar energy could accelerate the breakdown of organic compounds that make up CDOM or coincide with conditions when less terrestrial runoff (a primary CDOM source) occurs. In other words, sunnier months might lead to more photobleaching (light-driven degradation) of CDOM.

6.3 Precipitation and CDOM:

Some stations show positive correlations between Monthly Precipitation and Total CDOM. Rainfall can wash soil organic matter and leaf litter into streams, rivers, and lakes, increasing the concentration of dissolved organic matter. Thus, more rain often means more CDOM from terrestrial sources.

6.4 Wind Speed and CDOM:

The relationship with wind speed can vary by station. In some cases, higher wind speeds correlate positively with CDOM—potentially due to increased mixing of water layers and resuspension of sediment-bound organic matter. In other stations, higher wind speeds might coincide with periods or seasons when CDOM input is lower, resulting in negative correlations. Local geography, land use, and water body characteristics can cause these differences.

6.5 Differences between stations:

Each station has a unique pattern. For example, one station might have a strong negative correlation between temperature and CDOM, while another shows a weaker relationship. These differences highlight that the factors influencing CDOM are complex and site-specific.

The correlation matrices indicate that Total CDOM levels in a water body are not static—they respond to various environmental conditions. Seasonal changes in temperature, sunlight, rainfall, and wind patterns can influence the production and degradation of CDOM. Understanding these correlations helps researchers infer potential drivers of CDOM variability and can guide sampling strategies, environmental management, and water quality modeling in the regions.

7 Discussions

The stable annual patterns in CDOM percentages suggest that, at the scale of multiple years, Lake Võrtsjärv's DOM regime may be relatively resilient or in a state of quasi-equilibrium concerning major environmental shifts. The absence of a pronounced long-term trend could imply that climatic and land-use drivers have remained relatively consistent in this timeframe or that the lake's ecological and biogeochemical processes inherently buffer against substantial year-to-year changes.

Monthly patterns, however, highlight the dynamic interplay of hydrological and ecological processes. Elevated CDOM percentages in certain months may coincide with increased runoff events, as higher precipitation can mobilize terrestrial organic matter and deliver it into the lake. The timing of these pulses, along with seasonal variations in primary productivity, decomposition rates, and photodegradation, likely shapes the observed seasonal signals. Seasonal variation of hydrological parameters controlled the lake CDOM levels (Kutser, Li, Toming, & Noges, 2022). For instance, the correlation of warmer months and higher solar energy with lower CDOM could reflect enhanced photobleaching of organic matter and greater microbial and photochemical breakdown, thereby reducing CDOM concentrations.

The variability in correlations across different stations suggests that local factors mediate the relationship between environmental drivers and CDOM. This spatial heterogeneity is expected in an extensive, complex lake system like Võrtsjärv, where inflows, surrounding terrestrial ecosystems, and internal mixing processes differ around the lake's perimeter. Lake volume variation can potentially indicate variations in lake water color and DOC fluxes (Kutser, Li, Toming, & Noges, 2022). The interplay between climate-driven factors (e.g., precipitation and temperature) underscores the complexity of predicting future CDOM trends under changing environmental conditions.

8 Conclusions

This comprehensive examination of CDOM concentrations in Lake Võrtsjärv over an eight-year period reveals a stable annual baseline punctuated by more pronounced monthly and seasonal variability. The lack of a clear long-term trend suggests relative stability in the key processes governing CDOM inputs and breakdown within the lake. Nevertheless, the pronounced seasonal peaks and troughs, along with their year-to-year variability, highlight the importance of episodic events and local conditions that shape the lake's dissolved organic matter dynamics.

These findings have several implications for the management and study of Lake Võrtsjärv. Understanding that significant year-to-year changes may not be evident at the annual scale, stakeholders and researchers should pay closer attention to finer temporal resolutions and seasonal and monthly cycles to detect shifts in water quality and ecosystem function. Identifying the timing and drivers of seasonal peaks could inform more targeted sampling strategies, remediation efforts, and predictive modeling. Ultimately, ongoing monitoring and research will be crucial for anticipating how broader environmental changes, such as altering precipitation regimes or increasing temperatures, may influence Lake Võrtsjärv's CDOM dynamics and, by extension, its overall ecological health.

9 References

Arst, K. I. (2003). Optical properties and remote sensing of multicomponental water bodies. Springer Science & Business Media.

Das, S. (2023). Transparency of water column. In *An introduction to water quality science: Significance and measurement protocols* (pp. 71–78). Springer International Publishing. https://doi.org/10.1007/978-3-031-42137-2_4

Dubois, O., (2011). The state of the world's land and water resources for food and agriculture: managing systems at risk (pp. xxiii+-285).

Estonian Land Board. (n.d.). *National Satellite Data Centre (ESTHub)*. Retrieved December 12, 2024, from https://geoportaal.maaamet.ee/eng/spatial-data/national-satellite-data-centre-esthub-p654.html

Ševtšenko, F. (2024). CDOM-based Optical Water Types Classification [Master's thesis, Institute of Computer Science]. University of Tartu

Grigg, N. S. (2011). Total Water Management. United States: American Water Works Association.

Jerlov, N. G. (2014). Optical oceanography (Vol. 5). Elsevier.

Madhav, S., Singh, V. B., Kumar, M., & Singh, S. (Eds.). (2023). Hydrogeochemistry of Aquatic Ecosystems. John Wiley & Sons, Incorporated.

Moora, T., Raukas, A., & Tavast, E. (2002). Geological history of Lake Vortsjarv. Proc. Estonian Acad. Sci. Geol, 51, 157-179.

Moore, T. S., Campbell, J. W., & Dowell, M. D. (2009). A class-based approach to characterizing and mapping the uncertainty of the MODIS ocean chlorophyll product. *Remote Sensing of Environment, 113*(11), 2424–2430. https://doi.org/10.1016/j.rse.2009.07.016

Pettersson, K., George, G., Nõges, P., Nõges, T., & Blenckner, T. (2010). The impact of the changing climate on the supply and re-cycling of phosphorus. In G. George (Ed.), *The Impact of Climate Change on European Lakes* (pp. 121–137). Springer Netherlands. https://doi.org/10.1007/978-90-481-2945-4_8

Reinart, A., Herlevi, A., Arst, H., & Sipelgas, L. (2003). Preliminary optical classification of lakes and coastal waters in Estonia and south Finland. *Journal of Sea Research, 49*(4), 357–366. https://doi.org/10.1016/S1385-1101(03)00019-4

da Silva, E.F.F., Novo, E.M.L.M., Lobo, F.L. et al. Optical water types found in Brazilian waters. Limnology 22, 57–68 (2021). https://doi.org/10.1007/s10201-020-00633-z

Spyrakos, E., O'Donnell, R., Hunter, P.D., Miller, C., Scott, M., Simis, S.G.H., Neil, C., Barbosa, C.C.F., Binding, C.E., Bradt, S., Bresciani, M., Dall'Olmo, G., Giardino, C., Gitelson, A.A., Kutser, T., Li, L., Matsushita, B., Martinez-Vicente, V., Matthews, M.W., Ogashawara, I., Ruiz-Verdú, A., Schalles, J.F., Tebbs, E., Zhang, Y. and Tyler, A.N. (2018), Optical types of inland and coastal waters. Limnol. Oceanogr., 63: 846-870. https://doi.org/10.1002/lno.10674

Uudeberg, K., Ansko, I., Põru, G., Ansper, A., & Reinart, A. (2019). Using Optical Water Types to Monitor Changes in Optically Complex Inland and Coastal Waters. Remote Sensing, 11(19), 2297. https://doi.org/10.3390/rs11192297

10 Annexes

Weekly report

Week1

1. What I have done since last meeting	The topic of the group work was chosen. I have started to read about optical water types (OWT). The group members met and discussed about the organization of the work and further steps. The group has scheduled an appointment with the supervisors for the second week.	Explored the topic. Skimmed through research done in similar domains. Reviewed the published article, and understood terminology and general concept of project. Coordinated with group on the project.	WE discussed about the project with group mate and explored the topic. We reviewed the article on CDOM-based Optical Water Type classification and emailed the supervisor to schedule a meeting for next week.
2. Questions, issues	How to create an account on the ESThub platform, and how to use it? Which criteria are we going to consider to get the data? Start and end date? Temporal and spatial resolution? How are we going to store and access the data? Online drive? Hard drive?	Sourcing and Storing Data. Resolution of our Temporal Analyis. Target Area of Interest. How to run the model effectively?	How do we create and use the ESThub platform? What types of data will we be working with? How will we execute the model?
3. What I plan to do before the next meeting	Read about water classification methods based on transparency and active substances. Read about the basics of machine learning. Create an ESThub account and start learning about the platform.	Devise a method to run the model on a sample area of interest. View outcomes and understand the output of the model.	Get familiar with Machine Learning and remote sensing concepts. Figure out about optical water types (OWT). Investigate Sentinel imagery, focusing on its various bands and their specific purposes. Explore ESThub.

Week 2

1. What I have done since last meeting	Request an ESTHub account. Had a meeting with supervisors. Review the material shared by the supervisors. Explored sentinel images. Met with group members to understand the progress on Kumail's script.	Sourced a sample imagery of lake Vjotsroki using Google Earth Engine and Sentinel 2 for low cloud cover in 2024. Extract band 1-8a from imagery and store as csv through a python script. Written a script in python to select the best model based on sillouhette score. Ran model on sample dataset.	I explored Sentinel images and had a meeting with my supervisor. I visualized the Sentinel images in QGIS and emailed the ESThub coordinator for portal access. I met with my group members to discuss the progress on script and learned about the various steps Kumail took to make the model work. Additionally, I reviewed the materials shared by the coordinator.
2. Questions, issues	How can I get my digital sign to finish the ESTHub contract and get access to the platform. How to interpret clusters from the model's output and which of them could be usefull to our project	1. What kind of meaningful information we can extract from model classification relevant to our project. 2. What are the next steps required?	How can I gain access to the ESTHub processing platform using a digital signature? What was the optimal model used in the reference article? How can we interpret the image classification results from the model?
3. What I plan to do before the next meeting	Get access to ESTHub platform, learn how to used it, and get the needed data. Run the models, fix possible bugs, identify the best model, and get some preliminary/initial results	Devise method to automate the process of data acquiring and classification. Coordinate with team and supervisors on the direction of our project.	Continue emailing the ESThub coordinator to obtain the credentials. Run the model independently with Kumail's assistance, explore the results, and learn the process.

Week 3

1. What I have done since last meeting	Run the classification model with simple satelite image. Ask for ESTHub account. Download satellite images from Copernicus	Interpreted Model Results and Compared with Previous Studies.	1. Copernicus Browser Exploration and Sentinel Image Download. Familiarize myself with the Copernicus Browser platform and its capabilities for accessing and downloading Sentinel satellite imagery. 2. Model Execution on my Device. Ran the model in my local machine.
2. Questions, issues	Bugs in the classification model execution Digital sign for the ESTHub account	Metrices used for getting the correct model. Missing files in model directory.	-
3. What I plan to do before the next meeting	Get access to ESTHub platform and learn how to use it features.	Go through research to determine correct models which can predict correct classes accurately	1 . Obtain Access to ESTHub Account Access. 2. Familizer myself with ESTHub Portal: Review the Instruction PDF provided by the supervisor.

Week 4

1. What I have done since last meeting	Run the classification model. Started to write the report.	Shortlisted Correct Model and Map OWT Classes to CDOM Concentrations.	1. We had a meeting with the supervisor to discuss the progress. 2. Explored the different models from the provided file and determined the best model for the classification. 3. I looked at the report format at the University of Tartu.
2. Questions, issues	I have never got the reply from ESTHub	Data is not normalized	I had no idea about the bitmask and other band information, which were presented in different naming formats instead of standard naming band 1, band 2, and band 3. The processing method to use within ESThub.
3. What I plan to do before the next meeting	Get some preliminar results - classified images Discuss and organize the work for disertation.	Normalize Data	Explore continuously the metadata information from the ESThub image. Download and Normalize the data. Until then, Utilize the Google Earth engine for the image to run the model until we figure out the ESThub image.

Week 5

1. What I have done since last meeting	Generate preliminar classification images	Obtained Normalized Surface Reflectance for short time period using Polymer Correction on EstHUB.	I explored the ESThub and asked the supervisor about my issue with the ESThub image. I got a response from the supervisor regarding the correct way to download the image. I downloaded the image from the study time frame using polymer correction.
2. Questions, issues	Does the code classification is correct?	Is the normalization algorithm correct?	Confusion regarding whether our normalization method is the same as the technique used by the model while training the model.
3. What I plan to do before the next meeting	Undertand the basis of the code and the result	Understand the Normalization and Collect Data for whole time period.	Normalize the data. Make some output using the Google Earth engine image until we figure out everything with normalization. Make presentation slides and prepare myself for the presentation.

Week 6

1. What I have done since last meeting	Prepare presentation about work progress	Focused on Presentation and making sensible output.	I was able to normalize the image. Discuss the model's output and the appropriate way to represent those outputs from the model. We choose gifs to represent a change in CDOM level for different months within a year.
2. Questions, issues		Alot of issues regarding model output and results.	The issue was with the model output and different parameters that were present in the model.
3. What I plan to do before the next meeting	Follow recomendations after presentation	Address to issues and recommendations raised in Presentation.	Workout with the issue commented out by the supervisor during the presentation. Continue working on the ESThub image, replacing the GEE image, since we have been using it till now.

Week 7

1. What I have done since last meeting	Read and understand the classification code	Obtained Correct OWT CDOM Classes	Begin normalization with the ESThub downloaded data in the correct format based on the suggestion made by the supervisor. Before normalization, I removed the cloud and cloud shadow along with other noise using the bitmask value based on the instruction from the supervisor. Made charts to visualize the normalized data and see whether it was similar to the reference article or not.
2. Questions, issues	Understand normalization process	Running the model on large dataset	Running the model for the whole dataset for each image from 2016 to 2023.
3. What I plan to do before the next meeting	Try different parameters and models to better understand the code and result	Run the model and generate classes for all time period.	Understand the model parameters and how to choose the optimal model for the classification. Understand the classification code and how it corresponds to the OWT code.

Week 8

1. What I have done since last meeting	Make different test and executions of the code	Obtained CDOM Classes for all time period.	We had a discussion on the model output and how different classes can be interpreted since the model generates the CSV file with a CDOM class for each point. So, we discussed the output and learned about the model from Kumail.
2. Questions, issues	The comprehension of the code have been more difficult than expected	How to interpolate to raster and validate results.	How should we represent this point in the appropriate format, in raster? Which interpolation algorithm should we use?
3. What I plan to do before the next meeting	Schedule metting with supervisor for guide and advices	Generate Maps and Validate Results.	Figure out how to represent the output. Generate different raster files from it, write code to automate this process for each image and convert them to the image for visualization as well.

Week 9

1. What I have done since last meeting	Meet and talk with supervisor	Generate Maps for Time Period and Calculated Statistics (Mean CDOM etc)	We had a meeting with the supervisor and outlined the different methods and work that we had previously completed. Discussed the findings from the output from the year 2016 to 2023. What pattern did we see, and how can we interpret the output? Discussed how it might have happened and what the trend was annually and monthly.
2. Questions, issues	What is the correct procedure for normalization	What Metreological Variables should be included and where to get data.	Should we include the metrological variable and try to show the relationship between the CDOM value and environmental factors like solar radiation, temperature, and precipitation? If so, where will we find this data? Is it freely available, and what factor should we include?
3. What I plan to do before the next meeting	Try suggestions: normalization and models	Compare with metreological data.	Prepare for the presentation in the coming week. Understand the relationship between the metrological data and CDOM concentration. Download Metrologial data. Generate different images (covariance matrix image, bar chart) that show the relationship between meteorological data and CDOM value.

Week 10

1. What I have done since last meeting	Run the code correctly and generate optimal results	Obtained dataset from 3 stations near lake and preprocessed data and generate monthly means.	Discussed the metrological data findings from the Kumail research. Explored the data and how it should be used to interpret the relationship between environmental factors and CDOM value. We had a meeting with a supervisor as well. We discussed our progress until now.
2. Questions, issues	Interpretation of images	Which graphs and metrices would help compare data best	Interpretation of the data.
3. What I plan to do before the next meeting	Prepare presentation	Generate Visualizations for metreological data and CDOM concentrations.	Prepare for the presentation in the coming week.

Week 11

1. What I have done since last meeting	Prepare and interprete results	Visalized Metreological data and CDOM Concentration using different graphs in Jupyter Notebook.	Interpreted the results and finalized the presentation, including different results that we should include in the presentation. Explored code to visualize the CDOM data and discussed the best way to represent this information in charts and graphs.
2. Questions, issues	Recall guideline for writen report	How to properly present these results and justify them.	----------
3. What I plan to do before the next meeting	Gather generated results	Interpret results.	Start including this finding in the report.

Week 12

1. What I have done since last meeting	Sintesize information	Interpreted results with their explanations.	I looked at the guidelines for report writing and continued with the literature review to further understand our results' findings.
2. Questions, issues		None	None
3. What I plan to do before the next meeting	Write report	Write Report for this.	Write report

Week 13

1. What I have done since last meeting	Write final report	Written Report	Written Final Report.
2. Questions, issues	Expected issues with deadline due to exams week	None	_
3. What I plan to do before the next meeting	Write final report	Finalize and Submit Report.	Submit the Report after finalizing it.

Github repo: https://github.com/mohnark/spatio-temporal-changes-owq-params