ANALYSIS OF THE REMOTELY SENSED WATER QUALITY PARAMETERS OF THE INSUBRIC LAKES: METHODS AND RESULTS OF THE INTERREG SIMILE PROJECT

J.F Amieva (1), A. Austoni (1), M. Bresciani (2), M.A. Brovelli (1,2)

(1) Politecnico di Milano, Dept. of Civil and Environmental Engineering, Lecco Campus, Via Gaetano Previati 1/c, 23900, Lecco, Italy maria.brovelli@polimi.it, juanfrancisco.amieva@mail.polimi.it, alessandro.austoni@mail.polimi.it

(2) Istituto per il Rilevamento Elettromagnetico dell'Ambiente, CNR-IREA, via Bassini 15, 20133 Milano – bresciani.m@irea.cnr.it , maria.brovelli@polimi.it

KEY WORDS: Water Quality Monitoring, Landsat-8 TIRS, Sentinel-3 OLCI, Chlorophyll-a, Total Suspended Matter, Lake Water Surface Temperature, Spatial and Temporal Series.

ABSTRACT:

Lakes are a fundamental component of the environment and the territory and represent a precious source of fresh water for various uses. The area of the Prealps north of the Po valley in Italy is characterized by the presence of lakes which represent almost 80% of the total volume of fresh water in Italy (Rogora et al., 2018). The Insubric lakes (Lugano, Maggiore and Como) have their shared basins between Italy and Switzerland, and they are the objective of the SIMILE project, a cross-border Italian-Swiss project that aims to improve their coordinated management and strengthen stakeholder participation in the processes of knowledge and monitoring of water resources (Brovelli et al., 2019) by analyzing data acquired from in-situ to satellite sensors. The present work refers to data collected by remote sensing methods which offer the possibility to obtain synoptic views of water bodies to monitor water quality parameters (WQPs) such as the chlorophyll-a (Chl-a), the total suspended matter (TSM) and the lake surface water temperature (LSWT) (Giardino et al., 2013). This work presents an extensive evaluation of the space-time trends of the parameters based on the SIMILE remote sensing database.

Dataset	Number of maps	Initial acquisition	Final acquisition
Chl-a	389	15/01/2019	5/11/2022
TSM	392	15/01/2019	5/11/2022
LWST	69	01/01/2019	11/09/2022

Table 1. Summary of available maps per dataset.

1. INTRODUCTION

The paper presents the different analyses that were performed in order to understand the main spatial and temporal trends of some parameters in three insubric lakes, which are part of the SIMILE project: Lake Maggiore, Lake Como and Lake Lugano. The analyses were based on a previously produced set of maps which comprises 3 time-series datasets: chlorophyll-a (Chl-a), Total Suspended Matter (TSM) and Lake Surface Water Temperature (LSWT). Table 1 details the number of maps available in each dataset with the initial and final acquisition dates.

The main objective of the work is to determine if in the analyzed datasets there are trends which can be related to particular moments of the year but also determine if there are present trends in specific areas of the lakes. For the temporal analyses, it was necessary to adopt the seasons defined in the Water Framework Directive (WFD) (Bresciani et al., 2011), according to which are specified the following six time intervals (called, for brevity, "seasons"). The initial and final dates of each of the WFD seasons are specified in the table 2.

WFD Season	Initial date	Final date
Winter	01-Jan	20-March
Spring	01-Apr	15-May
Transit. spring-summer	15-May	15-Jun
Summer	01-Jul	31-Aug
Transit. summer-autumn	01-Sep	01-Oct
Autumn	01-Oct	30-Nov

Table 2. Water Framework Directive seasons' initial and final dates. (Bresciani et al., 2011)

2. MATERIALS AND METHODS

2.1 Analyzed parameters

Remote sensing is one of the possible tools that can help researchers and decision-makers to determine which are the conditions of a lake and, according to them, define the actions to be taken to properly manage its state in the challenging context that nowadays every lake exhibits due to climate change. Among the parameters that could be evaluated according to (Remote sensing for lake research and monitoring – Recent advances, 2016) are the following ones:

• **Transparency:** in (Remote sensing for lake research and monitoring – Recent advances, 2016) it is mentioned the importance of keeping good levels of transparency because it facilitates the penetration of the light in the lakes, allowing "biological, chemical and physical processes such as primary production and formation of macrophytes" (Remote sensing for lake research and monitoring – Recent advances, 2016). In this project, it was used the Total Suspended Matter (TSM) as an indicator for its measurement.

The suspended matter attenuates the lake transparency; it is typically present in lakes due to the contributions of lake tributaries and originates from soil and bedrock erosion or from internal resuspension (Remote sensing for lake research and monitoring – Recent advances, 2016).

- **Chlorophyll-a:** in order to understand the level of nutrient availability which can be a measure of the trophic level of a lake, it is possible to evaluate the amount of phytoplankton which is present in the studied waterbodies. For this purpose, it was adopted the Chloropyll-a indicator, which is a proxy of the level of phytoplankton. In order to determine the value of chlorophyll-a, it is necessary to analyze the absorption in the spectrum that ranges from 440 nm to 560 nm and also at 670nm (Remote sensing for lake research and monitoring Recent advances, 2016).
- **Temperature:** According to (Remote sensing for lake research and monitoring Recent advances, 2016), the water temperature is the product of several energy fluxes and it is a factor that affects different aspects of the ecosystem functioning, determining changes in the biodiversity of lake ecosystems and their oxygen concentrations. Specifically, thermal remote sensing instruments provide the capability to derive the lake's surface water temperature.

The analysis of the before-mentioned parameters in time is something that helps to understand which are their potential interactions and it contributes to the determination of the evolution of the global state of the lake's water quality. As was introduced before, this is one of the main objectives of this project.

2.2 Image processing

The data analyzed in the current work was produced in the first part of the SIMILE project and the main aspects of the processing are mentioned in (Toro Herrera et al., 2022). According to (Toro Herrera et al., 2022), in order to obtain the Chl-a and the TSM maps, it was necessary to use images provided by the ESA's Sentinel-3 A/B OLCI instrument which have a daily revisiting time with spatial resolution of 300 m. On the other hand, to produce the LSWT maps it was employed the Landsat-8 TIRS instrument of NASA whose termal infrared sensor has a spatial resolution of 100 m and a revisiting time of 16 days. Regarding the processing techniques, in (Toro Herrera et al., 2022) it is also mentioned that the Case 2 Regional Coast Colour (C2RCC), presented in (Brockmann et al., 2016), was employed to perform radiometric and atmospheric corrections as well as to compute Chl-a and TSM concentrations using Sentinel-3 data by means of a Neural Network model and the LSWT was determined using the Barsi method. The Barsi method is described in (Barsi et al., 2005) In (Toro Herrera et al., 2022) it is also explained that the C2RCC neural net flags were exploited to determine anomalies in the water spectra and also to determine the possible presence of clouds. The pixels with flags were excluded. Finally, (Toro Herrera et al., 2022) describes that in order to remove outliers, the 3 σ rule was applied.

2.3 Data availability

Considering that the different datasets were derived from multispectral imagery, it is possible to anticipate that there are some environmental conditions that may produce the absence of their values or that could also produce undesired values. For this reason, given the different datasets, it was quite important to estimate which are the levels of data availability in each pixel because, according to those levels, it was possible to select which pixels were the proper ones for the later analyses performed in this project. To determine the percentage of available data in each pixel of the different lakes for each of the three studied datasets, it was taken into consideration the total number of maps present for each dataset and then this number was compared with the total count of times that the same pixel did not get a null value.

Finally, were studied the spatial patterns of the data availability by means of a set of maps that analyzed the percentages of available data for all the pixels in each of the datasets by considering the full period of acquisitions but also for each season.

2.4 Statistical analyses

The project has also taken into account statistical analyses to understand the behaviour of the values of each dataset. For this purpose, one of the first considerations was to establish a threshold on the percentage of data availability per pixel: it was selected a minimum requirement of 50% of available data for each pixel to be considered in the analyses. Once defined the group of pixels that was taken into consideration, a histogram representation was prepared to analyze the distribution of the mean values of the pixels of each dataset in each of the different WFD seasons. In the same line, it was created a box-plot representation to complement the analysis of the distribution and to obtain further insights. These statistical analyses were performed for the group of the three lakes but also individually in order to determine if there exist differences in the distributions. In the section 3 section of this document, just the relevant results are described.

2.5 Spatial analysis

In order to understand the spatial distribution of the higher values of TSM, Chlorophylle-a and LSWT and to detect possible recurring patterns, it was chosen to implement a code automated procedure. First, the algorithm allows setting a group of initial parameters, depending on the desired output:

- Filter_Availability boolean type, if set to true, the analysis will be performed by also filtering data according to their availability
- **filter_season** boolean type, if set to true, the analysis will be performed individually for each season of the WFD periods, otherwise all periods will be considered together as one
- **lake** string type, to set according to the lake the study needs to focus on (Maggiore, Lugano and Como). If set to "full" the maps will be produced for all the three lakes, otherwise they will be cropped to a specific water basin
- **region** string type, it represents the region in which we want to focus our study (north,center or south). If set to "full" the maps will be fully produced, otherwise they will be cropped to a specific region

After that, the algorithm imports the correct datacube, from a local repository, according to the selected data type (TSM,

LSWT, Chl-a). To let the algorithm properly work, data has to be previously downloaded in a parent repository, in a sub-folder called "tsm", "lswt" and "chl" (depending on the type) inside a folder named "stacks". The datacube has three dimensions:

- dimension 1 and 2: these are the dimensions of a 2D matrix that corresponds to a 2D map of the Insubric lakes. By indexing dimension 1 and 2 it is possible to access a single data pixel from its pixel coordinates
- dimension 3: it corresponds to the time dimension, in particular each maps are stacked so one by one so that each map corresponds to a different day of the year, following a chronological order

For each day in the datacube it is computed the "90th percentile", meaning the pixel value in the 2D map under which 90% of the data falls. Therefore, a single value for each day of the year is extracted from each 2D matrix in the datacube. This "90th percentile" is a fundamental measure, since it will be used as a threshold to find higher pixel values relatively to the pixel values assumed in a specific day of the year. Due to this, greater values will be detected in all WFD periods, and not only in these periods where values are usually higher (for example temperatures will be higher in summer). As already anticipated, each 2D matrix is extracted from the datacube and its pixels are compared to the threshold relative to that day of the year (" 90^{th} percentile"). By a assigning a value of 1 if the pixel is greater than the threshold and 0 if not, a new matrix will be created in a form of a binary mask. Later all binary masks (corresponding to each day of the year) are adjusted according to the parameters mentioned at the top of this section and only after that, they are summed up to obtain a single 2D matrix (map), in which each pixel will count the number of days in which the 90^{th} percentile have been exceeded. Finally, the map is normalized with the number of days contained on the datacube, so that each pixel will then contain a percentage value. Thanks to this algorithm, it will be possible to produce a single 2D map (starting from a datacube), where regions in which it is more likely to find high values during the different WFD periods, are characterized by an elevated percentage value.

3. RESULTS AND DISCUSSION

3.1 Data Availability

In the maps that will be presented in the following section, data availability is characterized (as explained in Chapter 2.3) by a value that ranges between 0 (blue) and 1 (yellow), where 0 corresponds to complete unavailability and 1 to the maximum availability. Satellite data provides many benefits to lake monitoring, as it is possible to collect data from lakes without the need to be close to the field. Due to this, sometimes the distance between the sensor and the observed data, may cause some uncertainties in the observations, because of unwanted effects. As a matter of fact, satellite data suffers from the presence of "mixed pixels", shadows or adiacency effects. These effects bring uncertainty in reflectance values. This comes from the fact that these phenomena can generate some unusual responses, which may not be correctly detected from the sensor.

It can be evicted from Figure 1 and Figure 2, representing the data coming from the Sentinel-3 satellite (Chl-a and TSM) for the entire period, that pixels along the coast and in the Lugano



Figure 1. Data availability - Chlorophyll-a, full period.



Figure 2. Data availability - TSM, full period.

lake, assume a blue color, which means a level of availability lower than the 50%. Due to this, it is possible to state that these zones are characterized by bigger uncertainty. The reason for that can be attributed to the low resolution of the satellite, that causes a relatively big GSD, which implies that in one single pixel, many different classes of land may be included. Moreover, lake's border are characterized by the presence of both water and coast which may cause mixed pixels and coastal adjacency effects as well. Although this effects are quite evident in the maps, it is also true that borders represents, at least for the two largest lakes, a strong minority of the data. Moreover, TSM and Chl-a data cover almost completely all the zones with an high percentage of availability. For what it concerns the data availability analysis for TSM and Chl-a by period, results are not stating anything different from the full period already mentioned above and therefore the maps are not shown here.

Switching to the temperature, the first remark is that, as immediately evident from Figure 3, the difference in the spatial resolution of Landsat-8 Surface Water Temperature acquisitions, 3 times higher with respect to TSM and Chl-a maps (100m against 300m), makes maps definitely more detailed. Apart from that, data availability is clearly higher with respect to the other two data types, and also the Lugano lake and the coast are characterized by an optimal data coverage. Additionally, in Figure 4 it is possible to appreciate the difference in availability in the various periods, and it can be noticed how the situation is almost optimal in summer-autumn transition where most of the data has 100% availability, but worst in spring and



Figure 3. Data availability - LSWT, full period.



Figure 4. Data availability - LSWT, per season.

spring-summer transition. A further inspection on cloud coverage during this periods can possibly reveal the reason of this phenomenon. Despite that, it can be stated that the situation is almost excellent for all the periods, since generally the availability is never dropping under the 50%.

3.2 Statistical Analysis

This subsection describes the results obtained in the performed statistical analyses. The main objective was to understand the distributions of each of the datasets and to comprehend possible changes in the distributions between seasons and lakes.

TSM

In Figure 5 it is possible to observe the unified distribution of the TSM values for all the three lakes together. Differently from other parameters, the TSM distribution looks almost the same for all the WFD periods. Regarding the boxplots, instead, by

Winter	Spring	Spring-summer	Summer	Summer-aut.	Autumn
5	4	4	4	4	6-7

Table 3. Chl-a peak location per season $\mu g/L$.

comparing the unified plot (three lakes together) in Figure 6 with the one created just with data from the Lugano lake in Figure 7, it is possible to appreciate how the TSM variability for the Lugano lake is increasing during winter period. Despite that, values in winter are generally the lowest between the different periods (as can be seen from the mode in winter season, identified by the light-blue line inside each box in Figure 7). The highest values are, instead, reached during autumn period. As already known, the lake of Lugano is the smallest of the three so it is more likely that the variability of its indexes is higher.



Figure 5. Histograms per season- TSM (three lakes).



Figure 6. Boxplot per season- TSM (three lakes).

CHLOROPHYLL - A

Regarding Chl-a, the graph in Figure 8 shows how the distributions of the three lakes together, arrange themselves to a normal distribution, by distributing almost homogeneously around a single peak. In particular, for each period the peak is located around a specific value as summarized in table 3.

This trend finds a confirmation also in the boxplot graph (Figure 9), in which it is clearly visible how mean and variance of winter and autumn season are clearly higher with respect to all the other periods.

The reason of this behavior is probably related to cyanobac-



Figure 7. Boxplot per season- TSM (Lake Lugano).



Figure 8. Histograms per season- Chlorophyll-a (three lakes).



Figure 9. Boxplot per season- Chlorophyll-a (three lakes).

teria's activity: due to the lack of light in these periods, these organisms exploit their buoyancy to reach to the surface in order to get more light from the sun and warmer temperatures, then they duplicate, causing a massive increase in chl-a variability. This phenomenon usually occur in at the beginning and at the end of winter (Garneau Marie-Ève, 2015) (Yoshihiko Fujita, 2001). It is important to remember that satellites detect only the

superficial layers of the water and by doing so acquisitions are mainly detecting surface activity. Moving to single lake analysis, the lake of Lugano shows a very peculiar situation. As can be seen from Figure 10, the distribution of chl-a values for the lake of Lugano tends to assume a bimodal distribution with a double peak particularly evident in winter and summer.

The reason of this is related to the characteristic of the lake,



Figure 10. Histograms per season- Chlorophyll-a (Lake Lugano).

which is split in two different basins: the north region and the south one, which tend to have two different behaviours.

Lake Water Surface Temperature

As anticipated in Table 1, the amount of maps for surface temperature is clearly lower with respect to tsm and chl-a (69 maps against 392 and 389 respectively), because of the different satellite's revisit times (16 days for Landsat and 1 day for Sentinel). As a consequence, the distribution in Figure 11 is showing how data has very low variability around a well defined peak. This graph prove that surface temperatures are continuously increasing starting from winter (which has a peak around 7-8°C), up to summer (with a peak around 24°C).

It is fundamental to remember that these temperature are de-

lswt values [°C] of pixels by season (THRESHOLD=> at least 50.0% availability)



Figure 11. Histograms per season- LSWT (Three lakes).

tected only at the lake surface, due to the nature of the satellite sensor, hence it is usual that temperature may reach such high values. This phenomenon is even more explicit in the boxplot in Figure 12, where it can be appreciated how the temperature are slowly increasing from winter up to summer and then decreasing again going toward autumn, being the two periods of transition and autumn the ones with more variability (as expected, since usually in those periods temperature can vary consistently day by day).

Regarding the lake of Lugano, these normal distributions are



Figure 12. Boxplot per season- LSWT (three lakes).

even more defined, producing a clear and finite number of peaks (especially in autumn, where four bells can be spotted). This phenomenon can be attributed to the even reduced amount of data. Indeed, not only Landsat maps are fewer in number, but the Lugano lake is also the smallest of the three Insubric lakes, hence the amount of pixels contributing to the distribution is even lower. Due to this Lugano's distribution is sparser.



Figure 13. Histograms per season- LSWT (Lake Lugano).

3.3 Spatial patterns Analysis

This subsection presents the results related to the spatial analyses that were implemented in this project. Results associated with the three studied datasets are described in order to emphasize the main spatial patterns which are present for each of them.

TSM

From the point of view of the spatial distribution of the highest values, it is clear how a greater activity belongs to the northern zones of the lakes. Indeed, in figures 14, 15 and 16, especially in the north of the Como lake, the percentage of values exceeding the percentile rises up to almost the 60% in spring, spring-summer transition and summer respectively.



Figure 14. TSM spatial pattern map spring



Figure 15. TSM spatial pattern map spring-summer transition



Figure 16. TSM spatial pattern map summer

This spatial effect can be caused by the presence of the main rivers on the northern side of the lakes, which are known to collect suspended matter, which is later brought inside the water basin. In particular the rivers affecting the TSM concentration are the Mera and Adda ones, which are flowing inside the northern side of the Como lake, and the Ticino river which is flowing in the Maggiore lake.

CHLOROPHYLL - A

For chl-a the situation is very close to the TSM one. Indeed, it can be observed from Figure 17, 18, and 19, that in the northern region of Maggiore and Como lake the amount of chlorophyll is increasing up to the 60-70% in the periods of spring-summer transition, summer and summer-autumn transition. Actually, it's not new that chl-a and TMS data are close to one another, as the increase of suspended matter is related to the rivers that flow in the lakes, and knowing that most of that TSM is, indeed, organic matter, it happens that the phytoplankton has more nutrients to reproduce. Due to this, the spatial patterns of chlorophylla can be considered as a direct consequence of the TSM increase, hence of the rivers flowing in the lakes.

chl pixel frequency (%) exceeding 90% percentile in ss transition (50.0% availability)



Figure 17. CHLOROPHYLL-A spatial pattern map in spring - summer transition



Figure 18. CHLOROPHYLL-A spatial pattern map in summer

Lake Water Surface Temperature

The maps in Figure 20, 21 and 22, corresponding to the two transition periods and summer (which are usually the periods where temperatures are usually the highest), show higher temperatures in the south region of the Maggiore lake and Lugano one. Usually, these temperatures may be affected by light condition and water circulation, since they are detected at the surface. Given that wind may be considered as one of the main parameters that are influencing water circulation, closed environments may "suffer" from higher temperature. Indeed, the southern region of the Maggiore lake, is known to be surrounded by mountains, that may shelter this lake portion from winds.



Figure 19. CHLOROPHYLL-A spatial pattern map in summer autumn transition



Figure 20. LSWT - spatial pattern map in spring - summer transition



Figure 21. LSWT - spatial pattern map in summer

Regarding the Lugano lake, on the other side, its size may be the main reason of this effect, since by being the smallest and shallowest lake, it may be the most affected by environmental temperatures.



Figure 22. LSWT - spatial pattern map in summer - autumn transition

4. CONCLUSIONS

The lakes which were studied in this project constitute an important source of fresh water for several different activities and for that reason they are a relevant resource for the well-being of the communities and ecosystems. However, their proper management exhibits different challenges; some of which are intrinsic to any water body but also, in this particular case, some of them are related to the fact that Insubric lakes are water basins shared between two countries. Among these challenges, it could be highlighted the need to monitor in an efficient way the main water quality parameters and, according to the analysis of their state during time, define the best actions to be tackled. SIMILE's main objective is the coordination of that management and, in this sense, the current work exhibits a contribution to the understanding of the lake water quality: spatial and temporal patterns have been analyzed by means of statistical techniques that exploited three previously existing datasets (TSM, chl-a and LSWT), developed in the scope of the SIMILE project. This work has shown how satellite data can bring benefits to lake monitoring to help the decision-making processes. The benefits of remote sensing in this field are several, among which:

- sparing costs of expensive field interventions, by collecting data from a far distance
- allows frequent data collection thanks to the short revisit time of newer satellites
- data is freely available on most remote sensing satellite web portals

Of course satellite data analyses are also characterized by minor drawbacks as lower spatial distribution or revisit time for some satellites, and the detection of superficial data only, but their contribution is fundamental if used in synergy with other insitu technologies, for a complete and accurate lake monitoring campaign.

5. ACKNOWLEDGEMENTS

The work outlined in this paper is part of project SIMILE (Integrated monitoring system for knowledge, protection and valorisation of the subalpine lakes and their ecosystems) and as such it is financed by the European Commission in the frame of Interreg Italy-Switzerland 2014-2021 program.

REFERENCES

Barsi, J. A., Schott, J. R., Palluconi, F. D., Hook, S. J., 2005. Validation of a web-based atmospheric correction tool for single thermal band instruments. J. J. Butler (ed.), *Earth Observing Systems X*, 5882, International Society for Optics and Photonics, SPIE, 58820E.

Bresciani, M., Stroppiana, D., Odermatt, D., Morabito, G., Giardino, C., 2011. Assessing remotely sensed chlorophyll-a for the implementation of the Water Framework Directive in European perialpine lakes. *Science of The Total Environment*, 409(17), 3083-3091.

Brockmann, C., Doerffer, R., Peters, M., Kerstin, S., Embacher, S., Ruescas, A., 2016. 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. *Living Planet Symposium*, ESA Special Publication, 740, 54.

Brovelli, M. A., Cannata, M., Rogora, M., 2019. SIMILE, A GEOSPATIAL ENABLER OF THE MONITORING OF SUS-TAINABLE DEVELOPMENT GOAL 6 (ENSURE AVAIL-ABILITY AND SUSTAINABILITY OF WATER FOR ALL). *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLII-4/W20, 3–10.

Garneau Marie-Ève, Posch Thomas, P. J., 2015. Seasonal patterns of microcystin-producing and non-producing. *Harmful Algae*, 21-31.

Giardino, C., Bresciani, M., Stroppiana, D., Oggioni, A., Morabito, G., 2013. Optical remote sensing of lakes: an overview on Lake Maggiore. *Journal of Limnology*, 73(s1).

Remote sensing for lake research and monitoring – Recent advances, 2016. Remote sensing for lake research and monitoring – Recent advances. *Ecological Indicators*, 64, 105-122.

Rogora, M., Buzzi, F., Dresti, C., Leoni, B., Lepori, F., Mosello, R., Patelli, M., Salmaso, N., 2018. Climatic effects on vertical mixing and deep-water oxygen content in the subalpine lakes in Italy. *Hydrobiologia*, 824.

Toro Herrera, J. F., Carrion, D., Bresciani, M., Bratić, G., 2022. SEMI-AUTOMATED PRODUCTION AND FILTER-ING OF SATELLITE DERIVED WATER QUALITY PARA-METERS. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLIII-B3-2022, 1019–1026.

Yoshihiko Fujita, Kaori Ohki, A. M., 2001. Acclimation of Photosynthetic Light Energy Conversion to the Light Environments. *Algal Adaptation to Environmental Stresses*, 203-58.