LAKE WATER QUALITY MONITORING TOOLS

G. Bratic[1], D. Carrion[1], M. Cannata[2], M. Rogora[3], D. Strigaro [2], M.A. Brovelli[1,4]*

[1] Politecnico di Milano, Dept. of Civil and Environmental Engineering, Via Gaetano Previati 1/c, 23900 Lecco, Italy - (gorica.bratic; daniela.carrion; maria.brovelli)@polimi.it
[2] Institute of Earth Sciences, University of Applied Sciences and Arts of Southern Switzerland (SUPSI), Campus Trevano, 6952 Canobbio, Switzerland - (massimiliano.cannata; daniele.strigaro)@supsi.ch
[3] CNR Water Research Institute, L.go Tonolli 50, 28922 Verbania, Italy - michela.rogora@irsa.cnr.it
[4] CNR-IREA, Institute for Electromagnetic Sensing of the Environment, Via Edoardo Bassini 15, 20133 Milano, Italy

Commission IV, ICWG IV/III

KEY WORDS: Water Quality, Sensors, Earth Observation, Satellite Imagery, Citizen Science, Integrated Tools

ABSTRACT:

Lakes as ecosystems provide many goods and services. To benefit from them in long term we must assure sustainable management. SIMILE (informative System for the Integrated Monitoring of Insubric Lakes and their Ecosystems) project is focused on developing efficient monitoring of lake water quality since it gives the critical input for adequate management. The lakes of interest for SIMILE are the Insubric lakes Como, Lugano, and Maggiore. The paper is focused on describing which tools are used in the SIMILE project to exploit different sources of lake water quality data: *in-situ* high-frequency monitoring (HFM) through sensors, satellite observations, and data collected by citizens. Even though the paper is focused on the SIMILE project, and thus on tools and procedures for the Insubric lakes, it can serve as an example for other lakes too, especially because the tools developed in the project, such as a collaborative platform for sharing satellite-derived water quality parameters, and mobile application and web administrator interface for citizen science, are free and open-source, they can be easily adapted if needed. Moreover, the procedures for the processing of data coming from different sources are based on free (and often also open source) software and are well documented. The tools and procedures described in this paper might be a foundation for similar practice for lakes worldwide, and thus a step forward the 6th Sustainable Development Goal (SDG) of the United Nations ("Ensure availability and sustainable management of water and sanitation for all").

1. INTRODUCTION

Lakes are essential resources for satisfying the freshwater needs of households, industry, and agriculture. Moreover, lakes are attractive locations for many touristic and recreational activities that bring revenue to the local and regional communities. However, water quality may be a limiting factor for the above-listed purposes. In this view, SIMILE (Informative System for the Integrated Monitoring of Insubric Lakes and their Ecosystems) is an important project since it is focused on the lake quality monitoring that gives critical input for successful lake management.

The objective of the paper is to describe how and which tools were implemented to support decision-makers to better monitor the lake water quality involving, at the same time, a wide community of interested parties which is important, as the lakes are a common good. The tools were developed in the context of SIMILE project and they represent distinct, yet complementary tools for monitoring the deep subalpine lakes Maggiore, Como and Lugano. These lakes and their catchments are partially shared between Italy and Switzerland, so that the governance of the lakes should be performed in cooperation by the two countries Interested parties for the SIMILE project include administrations, research facilities, citizens, local associations, and schools from both Swiss and Italian sides. In addition, the development of the tools is followed by the creation of guidelines, and also by related lessons, webinars, and workshops. This will extend the project effects beyond the time span of the project.

SIMILE relies on *in-situ* sensor data, satellite observations, and a mobile application for citizen science to collect data about lake water quality. Such data will be eventually integrated in a Business Intelligence (BI) platform and presented in form of indicators (Brovelli et al., 2019).

The details about the tools used for lake water monitoring in SIMILE are organized in 5 remaining sections. In section 2, we share information about types of high-frequency monitoring (HFM) sensors installed, the parameters they measure, quality check (QC) and quality assurance (QA) procedures, and software employed for sensor data elaboration. Section 3 is dedicated to lake water quality monitoring with help of satellite observations. More in detail, section 3 describes how three water quality parameters (WQPs) can be extracted from satellite imagery, which imagery and which procedures are used to do so, how the derived products are shared by producers with users through a collaborative platform, and the zonation analyses made with satellite-imagery-based products. Section 4 includes a description of the tools for citizen science - mobile application and web administrator interface, and the experiment with machine learning for automatic validation of data collected by citizens. Section 5 briefly presents the concept of the BI platform, while the section 6 contains conclusions.

2. SENSOR DATA

Within SIMILE, high-frequency data are collected through HFM systems - *in-situ* sensors placed on buoys or platforms. The use of sensors may be convenient for integrating the traditional discrete monitoring of lakes, because of their high temporal resolution, consistency in time, low cost, and wireless sensor network. Data acquired by sensors are useful for detecting rapid changes

*Corresponding author

in physical, biological, and chemical parameters. They are particularly useful for detecting short-duration phenomena (e.g. algal bloom) (Banas et al., 2005; Le Vu et al., 2011) . Besides, they can have the role of "early warning" systems for algal bloom, anoxia, significant meteorological events that can completely change the state of the quality parameters or potentially cause floods, etc. The *in-situ* sensors have some drawbacks as well. They require maintenance (e.g. calibration, cleaning) and collected data should be regularly checked trough comparison with filed and laboratory methods. Sensors are prone to fouling that prevents accurate measurements. Furthermore, they can be damaged in case of extreme weather events or vandalism. Depending on the occurrence and frequency of fouling and damages, the costs for maintenance can become very high (Garel et al., 2009; Le Vu et al., 2011), in addition to the cost of data processing. It is clear that *in-situ* sensors cannot replace traditional water quality monitoring. If nothing else, the traditional quality measurements are needed for *in-situ* sensor calibration, but they are also irreplaceable for measurement of the parameters for which reliable sensors do not exist yet (e.g. algal nutrients) (Rinke et al., 2013). However, *in-situ* sensors can complement traditional measurements, and even serve to optimize the frequency of samples withdrawal and reduce costs in this way.

Figure 1: LM1 buoy in Maggiore lake

2.1 Sensors

A monitoring buoy (LM1) was placed in Lake Maggiore, Pallanza basin, in 2020 as a pilot experience, aimed at providing the practical know-how needed for the development of the whole HFM system. To increase replicability and transferability, LM1 was developed in-house, and conceived as a low-cost modular system. Several *in-situ* sensors were placed on the LM1 (Figure 1). The details of the sensors installed are provided in Tiberti et al. (2021). They include sensors for water temperature, pH/redox, dissolved oxygen, conductivity, turbidity, algal pigments (chlorophyll-a, phycocyanin and phycoerythrin). The HFM system is powered by a solar panel. It also has an electronic control unit for signal acquisition from the sensors, data storage, basic data elaboration, and wireless transfer. The sensor measurements are stored on a memory card and sent to CNR IRSA's servers over a wireless connection on a daily basis. The same sensors (excluding turbidity) have been also recently installed on two buoys in Lake Como. These sensors are still in a test phase and data will be subject to the same validation procedure of those in Lake Maggiore.

On the basis of a similar philosophy used to build LM1, an automatic HFM system was developed and deployed on Lake Lugano. This monitoring system implements, using an open approach, a customized solution where some elaboration on the collected data is moved on the edge. This transboundary lake is still recovering from the eutrophication process which had a peak during the 80s (Lepori et al., 2018). To this end, the monitoring system was equipped with six OPTOD by Ponsel sensors to measure the dissolved oxygen in water at six different depths (0.4, 2.5, 5.0, 8.0, 12.5, 20 m). The aim is to use such timeseries to study new approaches for estimating the Primary Production of the lake taking advantage of quasi real-time measurements (Cannata et al., 2021; Staehr et al., 2010). Such system was deployed in November 2020 and collected more than one year of data which are currently under validation and elaboration. In Figure 2, an extrapolation of the data collected is shown. Starting from the top, data about the dissolved oxygen and temperature are available. Such data collected by the HFM system are compared with wind speed and global radiation timeseries collected by the nearest MeteoSwiss Lugano station. Looking at this graph, it is possible to highlight how this data can, for example, help in better understanding the water dynamics of the lake such as possible mixing events and vertical movements of water oxygen.

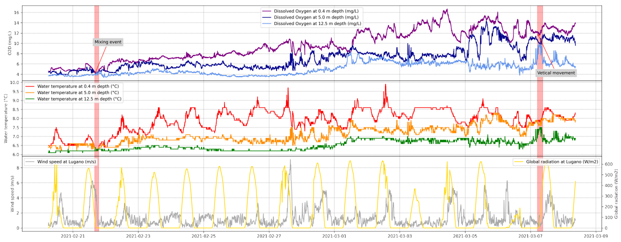

Figure 2: Data comparison between data collected by the Lugano HFM system and wind speed and global radiation data collected by the MeteoSwiss Lugano weather station.

2.2 Quality assurance and quality control procedures

Once the data is properly delivered, system functions can be monitored remotely (Tiberti et al., 2021). However, data flow interruptions can occur due to data collection or delivery errors. Moreover incorrect data can be generated due to sensor fouling/calibration drift or other error sources. It was necessary to establish QA procedures to minimize data loss, sensor errors, and the need for data correction. Cleaning of the HFM system, as well as *in-situ* measurements coupled with laboratory analyses, were performed. Particular attention was paid to the validation of chlorophyll-a measurements trough comparison with laboratory analysis based on standard methods. As an example, Figure 3 shows the evaluation of sensor readings in 2021 through the comparison with chlorophyll-a data gathered from grab samples analysed in laboratory.

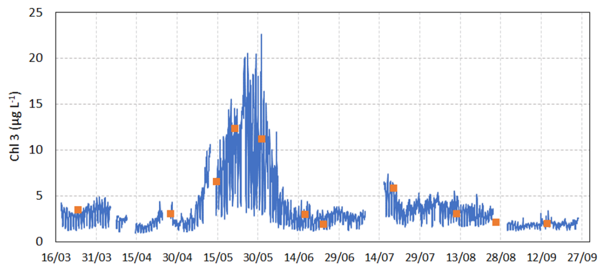

Figure 3: Sensor data (hourly average) of chlorophyll-a (2.5 m depth) in 2021. Orange dots: chlorophyll-a measured in the laboratory on grab samples.

Data collected during the first period of functioning of the LM1 buoy provided clear evidence that coupling HFM and discrete sampling for QA/QC controls is necessary to produce accurate data and detect and correct errors, mainly because of sensor fouling and calibration drift (Tiberti et al., 2021).

2.3 Software for sensor data - istSOS

Data collected by different sensors will be processed with istSOS software. It will serve as two-tier data storage - one tier on a buoy CPU board, and the other in the data center (Tiberti et al., 2021). The istSOS is selected for some of its many advantages (Cannata et al., 2015). First of all, it makes sensor observations compliant with the Sensor Observation Service (SOS) standard of the Open Geospatial Consortium. Thanks to this feature, data consumers (persons, machines, or processes) can easily request an overview of the available SOS capabilities as well as obtain details about offered sensors, service providers, observed properties, and areas. istSOS feature to associate quality index (QI) to the sensor records simplifies subsequent elaborations by allowing filtering or weighting the observations. Light-weight is another favorable characteristic of istSOS that is very significant for having the software on the buoy; however, it is important to mention that the software installation is enabled by the CPU that supports the Linux system. The istSOS on a buoy side performs a series of tests in order to assign a QI to observation and aggregates them at specific time intervals. Afterward, the observations are transferred to the data-center tier. WiFi, LoRa, NBIoT, and 4G are the protocols that can be used for the transfer of data between two tiers. Once the data are available on the data center tier they can be used for further processing in the scope of analyses, validation, calculation of indicators, etc. Data-user will have the possibility to select the specific sensor and deepen into details related to the selected sensor such as data types, responsible party, sensor specification, observing period, and observed phenomena.

Besides high-frequency monitoring (HFM), sensor data are intended to be used at later steps for modeling of lacustrine dynamics, with the aim to forecast future evolution considering nutrients and climate change scenarios (Fenocchi et al., 2018).

3. SATELLITE OBSERVATIONS

Satellite images in service of lake quality monitoring enrich the traditional approach by providing high-frequency data covering full areas of lakes of interest. Various free satellite images are downloaded from the spatial agencies' web portals and processed with open source algorithms in order to observe WQPs - suspended matter (TSM), chlorophyll-a (CHL-a), and lake surface water temperature (LSWT) (Luciani et al., 2021). The first two parameters were extracted from Sentinel-3 A/B OLCI (Ocean and Land Colour Instrument), while LSWT was extracted from Landsat 8 TIRS (Thermal Infrared Sensor) (Table 1). Anomalies in CHL-a or TSM were further investigated by using Sentinel-2 A/B MSI (MultiSpectral Instrument).

Water Quality Parameters	Spatial resolution	Source
Chlorophyll	300m	ESA - Sentinel-3 A/B OLCI
Total Suspended Matter	300m	ESA - Sentinel-3 A/B OLCI
Lake Surface Water Temperature	30m*	NASA - Landsat 8 TIRS

*the resolution of original Landsat8 TIRS is 100m, but the LSWT products are resampled to 30m

Table 1: Overview of the source data for WQP maps

3.1 Computation of water quality parameters

Even if satellite observations are coming from different sensors, they were all processed with SNAP (SentiNel Application Platform) software with different procedures (Luciani et al., 2021). For processing imagery from OLCI sensors, the C2RCC (Case 2 Regional Coast Colour) processor plugin was used. The plugin takes care of radiometric and atmospheric correction, and derivation of CHL and TSM. Moreover, the plugin produces flags for clouds and anomalies of water spectra, that are applied as a mask to CHL and TSM layers. In situations when flags are not successfully created by the C2RCC (e.g. thin clouds) NIR band is used for creating an additional mask. The masks from NIR are derived by setting a specific a threshold value. Since the NIR band is at 30m resolution, the images of the TIRS sensor are resampled also to 30m resolution, because the masking would not be feasible if the masked layer and the mask are at different resolutions. The C2RCC can be parametrized with fixed site-specific characteristics of the ecosystem to account for the lake's inherent optical properties (IOPs) on the WQPs estimates. To account for the temperature input required, C2RCC processor considered the data retrieved from the website of regional environmental protection agencies.

The LSWT was determined using the Barsi method (Barsi et al., 2005), which accounts for the radiometric correction of the images retrieved by the TIRS sensor (Luciani et al., 2021). It relies on atmospheric correction parameters that are obtained from USGS (United States Geological Survey) sources. Subsequently, the BOA (Bottom of atmosphere) radiances and LSWT were computed. Prevalently, for masking clouds the Landsat cloud masks were used. However, when they were not sufficiently efficient, the NIR band with specific thresholds was used.

Images from MSI were subject to atmospheric correction by Second Simulation of a Satellite Signal in the Solar Spectrum - Vector (6Sv) code. The 6Sv does not have cloud masking integrated, so the clouds were masked based on the thresholds on NIR bands. MSI observations were useful for the detection of the presence of cyanobacteria in the surface layer, as well as for foam detection. For the former one, the ratio between the red and green band with threshold 1 was useful, while for the latter the BOMBER model (Bio-Optical Model-Based tool for Estimating water quality and bottom properties from Remote sensing images) (Giardino et al., 2012) was utilized.

The output of the processing, the WQP datasets, were named following a specific naming convention to facilitate data upload and access. The name of each WQP product contains information about sensor name, quality parameter abbreviation, projection code, acquisition date and time of the original image used for WQP computation, and processing level of the original image. For example, for the WQP product with the name "S3A_CHL_IT_20200105T100533_L1" is acquired by the S3A sensor, it refers to chlorophyll-a concentrations, the reference map projection is for Italy (WGS 84 / UTM zone 32N), it is acquired on 05 January 2020 at 10:05:33, and the processing level of the image is Level 1.

Since the lakes are located between Italy and Switzerland, all products were reprojected to the two most common map projections in these two countries - CH1903+ / LV95 - Swiss CH1903+ / LV95, and WGS 84 / UTM zone 32N.

In the period from the beginning of January 2019 until the end of December 2021 675 products were derived. This includes 283 CHL products, 283 TSM products, and 109 LSWT products. The revisit time of Sentinel-3 is shorter than the revisit time of

Landsat 8, and thus there are many more observations of CHL and TMS. This results in weekly temporal resolution of CHL and TSM, and monthly temporal resolution of LSWT on average, however, the temporal resolution can be higher, depending on the image quality and on the cloud coverage.

Every WQP product is analyzed for detecting ”out- of-range” values, through statistical analysis, evaluation of the neighboring conditions and visual inspection of the original image.

3.2 Collaborative platform for producers and users

The WQP products are accessible for users on a collaborative platform (Toro Herrera et al., 2021). The platform is based on two separate applications: one for WQP producers (providers), and one for WQP users.

The application for WQP producers is realized as a geospatial data-sharing platform through which WQP products can be uploaded, together with their metadata and style. The application for producers is built upon a customized GeoNode application that runs on a docker container. WQP products providers may or may not have the permission to modify the already uploaded products, depending on their privileges in the application. An example of the interface of the application for producers is displayed in Figure 4. The figure shows a section of the application in which single WQP product and its metadata can be visualized and/or downloaded. The data sharing platform is available at https://www.geonode.eo.simile.polimi.it/.

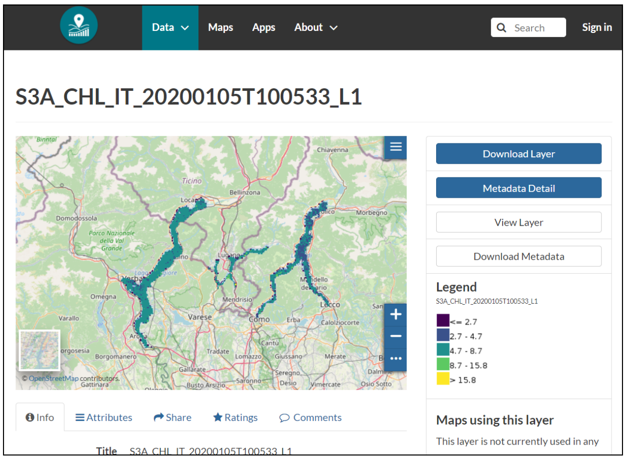

Figure 4: An example of an uploaded WQP product on producers’ application of collaborative platform

The application for users is a WebGIS through which WQP products are accessible as interactive maps. Some of the WebGIS base functionalities that are included in the user application are a table of contents, overview map, base map selection, zoom in/zoom out buttons, etc. The appearance of the WebGIS is displayed in Figure 5. The figure shows the WebGIS with several panels active: *Layer-Panel*, *Metadata-Panel*, and *Time-Panel*. *Layer-Panel* enables users to select the WQP of interest and to display it with the preferred opacity or to download it. The *Metadata-Panel* provides information about selected WQP in terms of typology, measurement units, legend, time stamp, etc. The *Time-Panel* allows users to navigate through time series of the selected WQP. Besides the panels displayed in the figure, the WebGIS integrates also *Basemap-Panel* from which the background map can be selected, and *Slide-Panel* with user guides for the WebGIS. The WebGIS functionalities can be further explored at https://www.webgis.eo.simile.polimi.it/. The user’s application is developed with a node.js JavaScript runtime environment that runs on a docker container different from the GeoNode one, however, the two containers/applications communicate by using open standards.

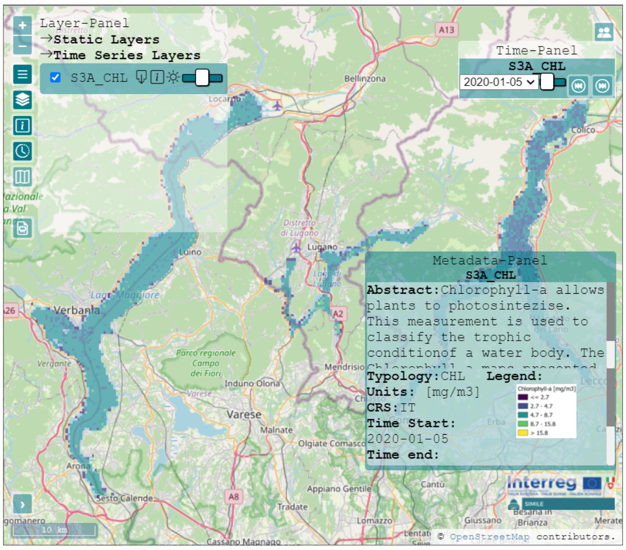

Figure 5: An example of appearance of the WebGIS (users’ application of the collaborative platform)

3.3 Analyses of water quality parameters

Some experiments were made in order to make zonation of the lakes according to water quality parameters (Gerosa et al., 2021). The zonation can be useful to detect the areas with the most critical water quality that require urgent interventions. The distinction among the lake zones was based on the combination on the clustering of three WQPs - CHL, TSM, and LSWT.

When the zonation experiments were carried out there were 198 CHL maps, 198 TSM maps, and 61 LSWT maps available. At that point, the detection of the ”out-of-range” values had not been introduced in the production of WQPs, yet. Therefore, prior to performing the cluster analyses the preprocessing of WQPs was done in order to detect, eliminate, or fix, when possible, the outliers and anomalies. After the outliers/anomalies analysis over the collection of WQPs maps were done, and suspicious WQPs were eliminated, around 60 combinations of the WQPs (30 for year 2019, and 30 for 2020) were created for approximately the same dates. The maximum number of combinations was primarily determined by the low temporal resolution of LSWT. The WQPs products were considered separately for the three lakes under study. This was necessary since the values of quality parameters vary depending on the lake. Then, values of the WQPs were normalized according to their maximum values reported in literature to ensure that they have the same importance in the clustering. The normalized WQPs were assembled into a virtual raster in which each WQP represents one band: for each date, there were 3 virtual rasters for 3 different lakes. The rasters of clusters for each virtual raster were computed with the K-means algorithm, in which the number of clusters was 5. The clusters were visually interpreted, with an additional processing step that facilitates interpretation. Namely, starting from the rasters of clusters new 3 band rasters were computed so that each band represented the average value of a WQP per cluster. In such a way, it was possible to display 3 WQPs with RGB colors, where LSWT was in the red color channel, CHL in green, and TSM in blue. The map legend has been designed to indicate the presence of high or low values for the WQPs in each cluster.

Apart from the zonation of lakes at specific dates, this was done also on a yearly basis. The first step to do so was aggregating the previous output (time series of average WQP per cluster) over time by computing the average values of pixels. Then, 5 new clusters were computed starting from the standard deviation of the WQPs normalized by the average value of the WQP for each year (2019 and 2020).

The analyses showed high variability of WQPs depending on season and years. Even so, it was possible to notice systematic behavior in some parts of some lakes. For example, Como Lake has three branches, and the quality parameters behave differently in these branches during the year. Similarly, the behavior of WQPs in Lake Lugano is different for the two distinct basins. For Lake Maggiore, a gradient with higher WQPs in the north and lower WQPs in the south was observed.

4. CITIZEN SCIENCE

The citizen science aspect of SIMILE is important for at least two reasons. Firstly, it helps to raise awareness of citizens about the significance of lakes ecosystems, and secondly, the data collected by citizens are supporting the analyses and monitoring of the lake water quality. The activities related to citizen science in SIMILE project were dedicated to the development of the free and open-source mobile application that allows citizens to report various issues observed in a lake by submitting pictures, filling out a simple survey, or inserting values of measurement performed by themselves (Carrion et al., 2020; Vavassori et al., 2021). Since the data collected by citizens might contain irrelevant or inappropriate contributions, the use of Artificial Intelligence for pre-filtering of data was studied in order to avoid/reduce manual checks (Biraghi et al., 2021).

4.1 SIMILE - Lake Monitoring mobile application

The application for citizen science, SIMILE - Lake Monitoring, was based upon lessons learned after detailed research of existing tools with similar use. The application was designed as a free and open source application that is user-friendly, functions offline, and provides user support (Jovanovic et al., 2019). The architecture of the mobile application is high-level three-tier system (Figure 6) (Carrion et al., 2020). The client is a mobile application developed on AngularJS and Ionic frameworks of JavaScript. The data is stored on MongoDB, a NoSQL and object-oriented database. Client requests are made via a RESTful API written in JavaScript, using Node.js.

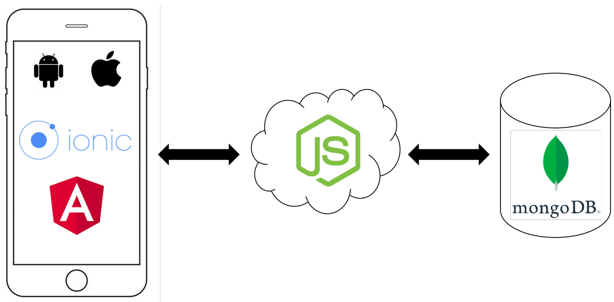

Figure 6: High-level three-tiear system

The application is available for Android and iOS via corresponding digital distribution services, while the source code is available on GitHub (https://github.com/interreg-simile/mobile-application). A registration is not mandatory in order to use the application. The design of the application was decided after ontological investigation by Biraghi et al. (2020) of the most suitable markers, feature attributes and symbols, and maps for representing dynamic environmental phenomena with aim to encourage participation by citizens.

The main aim of the application is to allow various users that may or may not be experienced to collect information relevant for lake quality monitoring. All observations collected by users are georeferenced and must be accompanied by a photo. User input is divided into 4 sections: *Weather*, *Details*, *Measures*, and *Other details*. The *Weather* section collects information about weather conditions at the moment when the new observation is inserted. The details on the weather are retrieved from the OpenWeatherMap service automatically and they can be refined by a user if needed. Some weather parameters are necessary to support the interpretation of collected data, such as water color as it depends on the illumination that can be affected by clouds. In the *Details* section, the users can insert parameters that can be identified visually i.e. algae, foams, oil stains, litters, drains, suspicious fauna, as well as odors. For each parameter, there is a list of questions specific to the characteristics of that parameter that helps a user to describe the observed issue. An example of questions for algae parameters is displayed in Figure 7. The *Measures* section

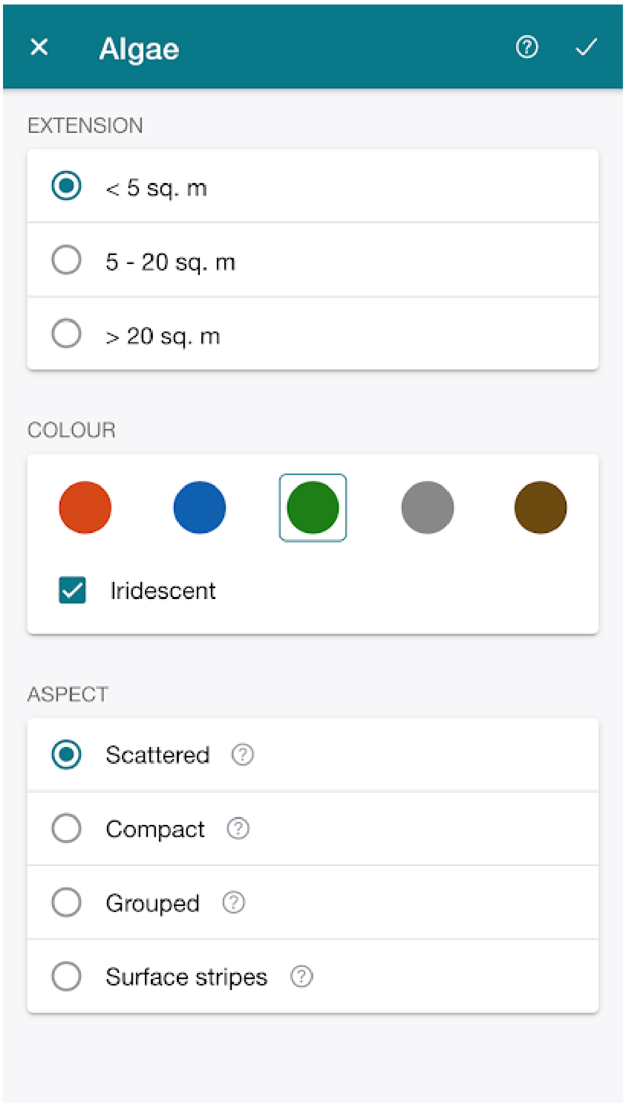

Figure 7: Example of parameters related to the observation of algae in SIMILE Lake monitoring application

is dedicated to the observations for which measurement tools are

https://doi.org/10.5194/isprs-archives-XLIII-B4-2022-599-2022

required such as Secchi disk, thermometer, or other limnological equipment. The parameters that can be inserted in this section are transparency, temperature, pH, oxygen, and bacteria. In the section *Other details*, users have 500 characters at their disposal to add further details if deemed necessary.

The *Help* button, available for each section and almost for every parameter, allows consultation of the detailed indication on how to use different sections of the application, or how to accurately insert observations. Figure 7 contains several examples of the *Help* button (question mark in a circle).

Contributions of users are stored in the database, and they can be visualized in the Map section of the application with the OpenStreetMap basemap in the background. The application contains also *Glossary*, *News*, *Project*, and *Useful links* sections. The *Glossary* contains definitions of the terms relevant to the lake quality parameters, *News* promotes educational or participatory events for the preservation of lakes, *Project* contains the description of the SIMILE project, and *Useful links* contains URLs leading to further details.

Except for the mobile application, there is the administrator web interface (Vavassori et al., 2021). It is primarily dedicated to the SIMILE project partners (or environmental agencies) for managing the SIMILE application and observations collected with it, however, it can be accessed also by guest users. Administrators and guests have the possibility to visualize observations in two ways, as tables or as points on the map. Unlike the mobile application, the web administrator interface contains the section *Analysis* where one can get statistics (e.g. pie charts of the count of observations) and trends (e.g. scatter plots of the count of observations with respect to date) of the collected observations. It is possible to filter the observations by date, weather conditions, water quality parameter, administrative region, or lake name. The filters can be used both for visualization and analyses purposes. There are some functionalities reserved only for the system administrators. The functionalities for administrators are insertion and modification of the events (visible by users in the *News* section of the mobile application), and modification and elimination of the observations. They can be used only if the web administrator interface is accessed with specific administrator credentials.

4.2 Validation of citizen science data with Machine Learning

The contributions of users through SIMILE applications might occasionally be irrelevant or inappropriate (Biraghi et al., 2021). Such a scenario is possible in the case of inexperienced individuals that may commit an error unintentionally, due to limited knowledge about lake ecosystems. However, given that the registration is not mandatory, unsuitable content might be submitted by malevolent users too. A validation of contributions by machine learning (ML) was explored as an automatic way to detect inadequate contributions, that otherwise would be managed by the system administrators manually which is time-consuming.

The automatic validation experiment was focused on algae and foams parameters, as they are the most frequently reported issues. Three different ML tools were involved in the experiment: Clarify platform, Convolutional neural network (CNN), and faster Region-based CNN (R-CNN), an object-detection algorithm. The observations collected by SIMILE mobile applications solely were not sufficient input datasets for the ML models. The initial dataset was enlarged by keyword and image searches on web engines (Google, Bing, etc) and by crawling Flickr data. Out of the three ML selected, the faster R-CNN object detection algorithm showed the best performance, however further improvements are required in order for ML to autonomously perform effective validation without human interaction.

5. BUSINESS-INTELLIGENCE PLATFORM

The final objective of the project is to consolidate information from traditional measurements, sensors, satellite observations, citizen science, and models of lake dynamics into a BI platform. Having the information from different sources on a single platform will promote the use of such information for decision and policymaking for the public administration of the Insubric lakes (Brovelli et al., 2019). The business intelligence platform will be a web-based data-driven system.

6. CONCLUSIONS

The paper outlined various tools for the monitoring of lake water quality, which is essential for efficient management of these precious resources. The Insubric lakes extend over Italian-Swiss border. They are utilized on both sides of the border for agriculture, domestic use, energy generation, tourism, and recreation, and other activities. For this reason, a cross-border water quality management is pivotal for ensuring integrity of the waters of Insubric lakes. The tools for the monitoring are developed within the SIMILE project thanks to the expertise and efforts of the Swiss and Italian project partners (universities, research institutes and public administrations), but also thanks to the participation and feedback of citizens, local associations, schools, etc. The SIMILE project focuses on three main data sources for monitoring of lake water quality: *in-situ* sensors for HFM, satellite imagery, and citizen science. For all three data sources it was necessary to develop procedures and tools for extraction of relevant information from raw data, as well as the platforms for distribution of the extracted information.

Sensors installed on the buoys/platform within the project provided useful information on the short-term variability of limnological variables and allowed a first assessment of short-term processes, such as those related to meteorological events. QA/QC procedures were adopted to get affordable data. The software selected for processing HFM sensor data is istSOS.

The data coming from satellite imagery do not have as high frequency as those from HFM sensors, however, they provide data for the whole surfaces of the lakes, and not just for immediate surrounding as it is case with *in-situ* sensors. The information from satellite imagery are extracted with SNAP software and procedures specific to satellite sensor with which imagery is acquired. Three WQPs — LSWT, TSM, and CHL — are extracted from satellite observations. The WQPs are examined for "out-of-range" values with statistical methods, and they are published on collaborative platform if they pass quality checks. The collaborative platform is designed within SIMILE project to facilitate WQPs publication, and their further exploitation in the project or by other users.

Finally, citizen science data is collected by citizens with the SIMILE - Lake monitoring mobile application developed specifically for SIMILE project. In addition to the mobile application, the web administrator interface was developed too to allow visualization and management of the observations from the mobile application. The observations by citizens are particularly useful for phenomena that cannot be detected by HFM sensors or satellite imagery, such as foams, litter, odors, etc. Moreover, the observations of citizens are not fixed to a single location, which is an advantage with respect to the HFM sensors used in SIMILE project. Nevertheless, the observations of citizens are prone to intentional or unintentional errors, and therefore their reliability must be verified. Some experiments were made in order to exploit possibility of integration of AI for validation of citizen science data with aim to minimize manual checks.

Information related to water quality coming from different sources are complementing each other. Finally, information coming from sensors, satellite imagery, citizen science, traditional measurements, and models of lake dynamics will be integrated in the business intelligence platform. This will encourage utilization of such information for different purposes including policy and decision making.

Even thought the SIMILE project is dedicated to Insubric lakes, the approach adopted in the project can be extended on the lakes worldwide, especially because the tools that are developed within the project are free and open source, and therefore can be customized if needed. Wider use of SIMILE outputs and principles would significantly contribute to the 6th Sustainable Development Goal (SDG) of the United Nations ("Ensure availability and sustainable management of water and sanitation for all").

ACKNOWLEDGEMENTS

The paper presents the research carried out within SIMILE project (informative System for the Integrated Monitoring of Insubric Lakes and their Ecosystems, ID: 523544), which is funded by the European Commission in the frame of Interreg Italy-Switzerland 2014-2021 program.

REFERENCES

Banas, D., Grillas, P., Auby, I., Lescuyer, F., Coulet, E., Moreteau, J.-C. and Millet, B., 2005. Short time scale changes in underwater irradiance in a wind-exposed lagoon (Vaccarès Lagoon, France): Efficiency of infrequent field measurements of water turbidity or weather data to predict irradiance in the water column. Hydrobiologia 551(1), pp. 3–16.

Barsi, J. A., Schott, J. R., Palluconi, F. D. and Hook, S. J., 2005. Validation of a web-based atmospheric correction tool for single thermal band instruments. San Diego, California, USA, p. 58820E.

Biraghi, C. A., Lotfian, M., Carrion, D. and Brovelli, M. A., 2021. AI in support to water quality monitoring. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLIII-B4-2021, pp. 167–174.

Biraghi, C. A., Pessina, E., Carrion, D. and Brovelli, M. A., 2020. VGI visualisation to support participatory lake monitoring: the case study of SIMILE project. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLIII-B4-2020, pp. 237–244.

Brovelli, M. A., Cannata, M. and Rogora, M., 2019. SIMILE, a geospatial enabler of the monitoring of sustainable development goal 6 (Ensure availability and sustainability of water for all). ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-4/W20, pp. 3–10.

Cannata, M., Antonovic, M., Molinari, M. and Pozzoni, M., 2015. istSOS, a new sensor observation management system: Software architecture and a real-case application for flood protection. Geomatics, Natural Hazards and Risk 6(8), pp. 635–650.

Cannata, M., Strigaro, D., Lepori, F., Capelli, C., Veronesi, M., Rogora, M., Lami, A. and Brovelli, M., 2021. SIMILE: An integrated monitoring system to understand, protect and manage sub-alpine lakes and their ecosystem. EGU General Assembly 2021, pp. EGU21–12404.

Carrion, D., Pessina, E., Biraghi, C. A. and Bratic, G., 2020. Crowdsourcing water quality with the SIMILE app. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLIII-B4-2020, pp. 245–251.

Fenocchi, A., Rogora, M., Sibilla, S., Ciampittiello, M. and Dresti, C., 2018. Forecasting the evolution in the mixing regime of a deep subalpine lake under climate change scenarios through numerical modelling (Lake Maggiore, Northern Italy/Southern Switzerland). Climate Dynamics 51(9-10), pp. 3521–3536.

Garel, E., Nunes, S., Neto, J. M., Fernandes, R., Neves, R., Marques, J. C. and Ferreira, , 2009. The autonomous Simpatico system for real-time continuous water-quality and current velocity monitoring: examples of application in three Portuguese estuaries. Geo-Marine Letters 29(5), pp. 331–341.

Gerosa, C., Bresciani, M., Luciani, G., Biraghi, C. A., Carrion, D., Rogora, M. and Brovelli, M. A., 2021. Zonation of subalpine lakes based on remotely sensed water quality parameters. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLIII-B3-2021, pp. 551–558.

Giardino, C., Candiani, G., Bresciani, M., Lee, Z., Gagliano, S. and Pepe, M., 2012. BOMBER: A tool for estimating water quality and bottom properties from remote sensing images. Computers & Geosciences 45, pp. 313–318.

Jovanovic, S., Carrion, D. and Brovelli, M. A., 2019. Citizen science for water quality monitoring applying FOSS. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-4/W14, pp. 119–126.

Le Vu, B., Vinçon-Leite, B., Lemaire, B. J., Bensoussan, N., Calzas, M., Drezen, C., Deroubaix, J. F., Escoffier, N., Dégrés, Y., Freissinet, C., Groleau, A., Humbert, J. F., Paolini, G., Prévot, F., Quiblier, C., Rioust, E. and Tassin, B., 2011. High-frequency monitoring of phytoplankton dynamics within the European water framework directive: application to metalimnetic cyanobacteria. Biogeochemistry 106(2), pp. 229–242.

Lepori, F., Roberts, J. J. and Schmidt, T. S., 2018. A paradox of warming in a deep peri-Alpine lake (Lake Lugano, Switzerland and Italy). Hydrobiologia 824(1), pp. 215–228.

Luciani, G., Bresciani, M., Biraghi, C. A., Ghirardi, N., Carrion, D., Rogora, M. and Brovelli, M. A., 2021. Satellite monitoring system of subalpine lakes with open source software: The case of SIMILE Project. Baltic Journal of Modern Computing.

Rinke, K., Kuehn, B., Bocaniov, S., Wendt-Potthoff, K., Büttner, O., Tittel, J., Schultze, M., Herzsprung, P., Rönicke, H., Rink, K., Rinke, K., Dietze, M., Matthes, M., Paul, L. and Friese, K., 2013. Reservoirs as sentinels of catchments: the Rappbode Reservoir Observatory (Harz Mountains, Germany). Environmental Earth Sciences 69(2), pp. 523–536.

Staehr, P. A., Bade, D., Van de Bogert, M. C., Koch, G. R., Williamson, C., Hanson, P., Cole, J. J. and Kratz, T., 2010. Lake metabolism and the diel oxygen technique: State of the science: Guideline for lake metabolism studies. Limnology and Oceanography: Methods 8(11), pp. 628–644.

Tiberti, R., Caroni, R., Cannata, M., Lami, A., Manca, D., Strigaro, D. and Rogora, M., 2021. Automated high frequency monitoring of Lake Maggiore through in situ sensors: System design, field test and data quality control. Journal of Limnology.

Toro Herrera, J. F., Carrion, D. and Brovelli, M. A., 2021. A collaborative platform for water quality monitoring: SIMILE webGIS. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLIII-B4-2021, pp. 201–207.

Vavassori, A., Biraghi, C. A., Bratic, G., Carrion, D., Zamboni, G. and Brovelli, M. A., 2021. Citizen science tools for lake monitoring in the framework of the United Nations Sustainable Development Goals: The project SIMILE. GI_Forum 1, pp. 179–186.