

Automated high frequency monitoring of Lake Maggiore through *in situ* sensors: system design, field test and data quality control

Rocco Tiberti,¹ Rossana Caroni,¹ Massimiliano Cannata,² Andrea Lami,¹ Dario Manca,¹ Daniele Strigaro,^{2,3} Michela Rogora^{1*}

¹National Research Council of Italy, Water Research Institute (CNR-IRSA), Largo Tonolli 50, 28922 Verbania, Italy; ²Institute of Earth Sciences, Department of Environment, Construction and Design, University of Applied Sciences of Southern Switzerland (SUPSI), Campus Mendrisio, Via Francesco Catenazzi 23, 6850 Mendrisio, Switzerland; ³Department of Earth and Environmental Sciences (DSTA), University of Pavia, Via Ferrata 9, 27100 Pavia, Italy

ABSTRACT

A high frequency monitoring (HFM) system for the deep subalpine lakes Maggiore, Lugano and Como is under development within the EU INTERREG project SIMILE. The HFM system is designed to i) describe often neglected but potentially relevant processes occurring on short time scale; ii) become a cost-effective source of environmental data; and iii) strengthen the coordinated management of water resources in the subalpine lake district. In this project framework, a first HFM station (LM1) consisting of a monitoring buoy was placed in Lake Maggiore. LM1 represents a pilot experience within the project, 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. LM1 is presently equipped with solar panels, a weather station, and sensors for water temperature, pH, dissolved oxygen, conductivity, and chlorophyll-a. In this study, we describe the main features of LM1 (hardware and software) and the adopted Quality Assurance/Quality Control (QA/QC) procedures. To this end, we provide examples from a test period, *i.e.*, the first 9-months of functioning of LM1. A description of the software selected as data management software for the HFM system (IstSOS) is also provided. Data gathered during the study period provided clear evidence that coupling HFM and discrete sampling for QA/QC controls is necessary to produce accurate data and to detect and correct errors, mainly because of sensor fouling and calibration drift. These results also provide essential information to develop further the HFM system and shared protocols adapted to the local environmental (*i.e.*, large subalpine lakes) and technical (expertise availability) context. The next challenge would be making HFM not only a source of previously unaffordable information, but also a cost-effective tool for environmental monitoring.

INTRODUCTION

Monitoring the quality of surface waters is compulsory under current environmental policies, in particular in Europe, under the EU Water Framework Directive (WFD; European Commission, 2000). The WFD requires the classification of the ecological status of surface waters with an integrated approach based on several biological

indicators and physico-chemical and hydro-morphological variables. The principal aim of the WFD is to protect and enhance the status of aquatic ecosystems in the European Union. Monitoring is a basic effort to assess the present-day quality of European waters, describe their evolution through time, and, ultimately, assess the effects of the WFD and design mitigation options (Skeffington *et al.*, 2015).

Traditionally, monitoring freshwaters involves field sampling and subsequent laboratory work. Samples are collected manually from selected areas of the waterbody and times of the year. This conventional approach provides essential information on the water quality and the main ecological processes. However, this approach often fails to capture the fast-paced dynamics of many biotic and abiotic processes because of low sampling frequency (Marcé *et al.*, 2016). As an example, the conventional monitoring of the trophic state of lakes based on microscopic characterization of phytoplankton communities and chlorophyll-a (Chl-a) extractions (Lorenzen, 1967; ISO, 1992), due to its demanding field work and laboratory effort, is necessarily conducted at a low frequency resolution. This kind of monitoring may be insufficient to detect the short-lived, but significant changes in the phytoplankton communities, *e.g.*, rapid turnover of phytoplankton species, cyanobacterial increases, and bloom

Corresponding author: michela.rogora@cnr.it

Key words: Water quality; chlorophyll; data management; sensors; SIMILE project.

Edited by: Diego Fontaneto, CNR-IRSA Water Research Institute, Verbania, Italy.

Received: 3 March 2021.

Accepted: 19 April 2021.

This work is licensed under a Creative Commons Attribution Non-Commercial 4.0 License (CC BY-NC 4.0).

©Copyright: the Author(s), 2021

Licensee PAGEPress, Italy

J. Limnol., 2021; 80(2):2011

DOI: 10.4081/jlimnol.2021.2011

formation (Hunter *et al.*, 2009; Stumpf *et al.*, 2012; Song *et al.*, 2012; Tran Khac *et al.*, 2018).

In general, all the processes occurring over a temporal scale shorter than the sampling frequency may go unnoticed in traditional monitoring programs (*e.g.*, short-lived, extreme episodic, or unpredictable events; Dubelaar *et al.*, 2004; Banas *et al.*, 2005; Itvánovics *et al.*, 2005; Stockwell *et al.*, 2020). Both trophic state indicators, and single limnological variables (*e.g.*, temperature, pH, dissolved oxygen) are indicative of lake water quality and may contribute to identify conditions favourable to -or indicative of- algal blooms development. For instance, summer heat waves may promote blooms of harmful cyanobacteria and other worrisome taxonomic shifts (Wagner and Adrian, 2009). Sudden changes in water temperature, conductivity and dissolved oxygen can support the identification of upwelling events due to internal seiches, mixing episodes, intrusion of river waters and the calculation of lake mixing and stratification indices (Read *et al.*, 2011). Short term changes in pH and dissolved oxygen may also help investigating inorganic carbon dynamics in lakes (Khan *et al.*, 2020). Further, high frequency limnological data, possibly gathered at different locations within a lake, represent a valuable resource for lake modelling (Laborde *et al.*, 2010; Pilotti *et al.*, 2013).

Recent advances in sensor technology can overcome the drawbacks of conventional monitoring, giving the opportunity to measure an increasing number of limnological and ecological parameters at unprecedented short temporal intervals and also during adverse meteorological events (Banas *et al.*, 2005; Johnson *et al.*, 2007; Le Vu *et al.*, 2011; Jennings *et al.*, 2012; Klug *et al.*, 2012; Hamilton *et al.*, 2014; Meinson *et al.*, 2016). Over the past decade, there has been a drastic increase in the use of automated collection of high-frequency data in scientific research and environmental monitoring (Horsburgh *et al.*, 2015). In addition, the installation of automated high frequency monitoring (HFM) systems is often expected to become a cost-effective monitoring tool, to cope with the scarcity of specialized personnel for field and laboratory work and to reduce the analytical costs (Le Vu *et al.*, 2011). A great impulse to the development of HFM systems for European lakes came from the COST Action NETLAKE (ES1201 - Networking Lake Observatories in Europe) which prompted the establishment of a network of scientists, technologists, managers, and stakeholders focused on the application of sensor technology for lake monitoring (Marcé *et al.*, 2016). The cooperative network GLEON (Global Lake Ecological Observatory Network) also supports sharing and interpreting high resolution sensor data from lakes worldwide (Weathers *et al.*, 2013). An increasing number of environmental authorities from all around the world are integrating some HFM systems in their monitoring programmes (Le Vu *et al.*, 2011; Bertone

et al., 2018), also to improve provision of high-quality freshwater, provision of food (aquaculture) and recreation (Marcé *et al.*, 2016). A few examples from Italy include the HFM network deployed in 18 Sardinian reservoirs used to identify the optimal depth for water withdrawal and to early detect the development of algal blooms (Marcé *et al.*, 2016); a monitoring system which has been developed in Lake Iseo, measuring the main thermal, radiative, and mechanical fluxes on the lake surface (Pilotti *et al.*, 2013); and a high-frequency monitoring of meteorological parameters and water temperature at different locations in Lake Como as input to a three-dimensional hydrodynamic model (Laborde *et al.*, 2010).

The implementation of a HFM system is also projected for the lakes of the so-called Insubric region (lakes Maggiore, Como, and Lugano) and will be developed within the cross-border cooperation project SIMILE (Italian acronym for “Integrated monitoring system for knowledge, protection and valorisation of the subalpine lakes and their ecosystems”). While the general aim of SIMILE is to improve and optimize lake monitoring integrating conventional monitoring, satellite data, *in situ* HFM data, and user-contributed georeferenced data (Brovelli *et al.*, 2020), specific aim of the HFM system is to make available high frequency data of basic limnological and meteorological parameters in order to provide information on lake processes occurring over a short time scale, such as algal blooms, or weather-related episodic events.

The HFM station LM1 was the first station developed within SIMILE (Fig. 1). In this project framework, LM1 represents a pilot experience, aimed at providing the practical know-how needed for the creation of the larger HFM system projected in SIMILE. To enable the replicability and transferability of this first experience, LM1 was conceived as a low-cost modular system. Indeed, although full-equipped systems for HFM of water bodies are commercially available, their long-term maintenance may be difficult because the understanding of the hardware and software components is often not an easy process. Therefore, an increasing number of research groups are developing their own modular system independently (Albaladejo *et al.*, 2012; Tran Khac *et al.*, 2018; Vitale *et al.*, 2018): in-house development provides the flexibility needed to recover the system when malfunctioning or damages occur (Vitale *et al.*, 2018) and to tweak its configuration to monitoring needs.

In this paper, we describe LM1 (software and hardware components, *i.e.*, sensors, electronic devices, power supply and transmission systems), providing an example of a cost-effective and modular limnological buoy for HFM. In addition, we used the first nine months of HFM data to provide examples of: i) field validation of sensor data, comparing sensor readings with *in situ* (multiparameter probe) and laboratory measurements; ii) the most

common source of errors in sensor data; and iii) the adopted Quality Assurance and Quality Control procedures. This study mainly aimed at developing a sound protocol for HFM data collection, quality check, data transmission, and storage within the SIMILE project. For this reason, we have focused on the approach adopted to get affordable HFM data, while the data analysis will be the subject of a later study within the SIMILE project once the whole data flow procedure will be established. The information provided by our study can be used in HFM monitoring of lakes to get affordable data, to be integrated into institutional monitoring programs of environmental agencies and research institutes.

LAKE MAGGIORE, ITS MONITORING HISTORY AND LONG-TERM DYNAMICS

Lake Maggiore is a large (surface area 213 km²; volume 38.1 km³) and deep (370 m) subalpine lake located between Northern Italy and Southern Switzerland (Fig. 1). Its watershed (6,600 km²) is shared almost equally between the two countries while most of lake surface (~80%) lies in Italy. Lake Maggiore has 33 inflows - main

inflows are the Ticino and Toce Rivers - and one outflow (Ticino River). Its theoretical water renewal time is 4.20 years. Lake level is controlled by the Miorina Dam to ordinarily allow for a 0.5 m difference between summer and winter periods (Fenocchi *et al.*, 2017).

Lake Maggiore has been regularly monitored on a monthly base since 1978 with the support of the International Commission for the Protection of Swiss-Italian Waters (CIPAIS; CNR-IRSA, 2019). Samples and measures were taken at the deepest point of the lake (Ghiffa station; 45°58'30" N; 8°39'09" E). Since 2019, a second monitoring site in Pallanza (100 m depth) has been included in the monitoring programme (Fig. 1). Monitoring at both stations include 1) sampling for chemical and biological analyses, and 2) a vertical profile of temperature through a multiparameter probe (Idronaut CTD304). Sampling for chemical analyses involves the collection of 12 water samples along the water column at the Ghiffa station and 3 samples at the Pallanza station. Samples are routinely analysed for the main chemical variables (pH, electrical conductivity, alkalinity, major anions and cations, reactive and total phosphorus, ammonium, total nitrogen, reactive silica) at the water chemistry laboratory of the CNR Water Research Institute (CNR-IRSA) in Verbania using stan-



Fig. 1. LM1. a) Description of the main components of LM1, parts in red are projected, not yet present components. b) LM1 buoy. c) Location of LM1 and of the long-term sampling stations (white dots) in Lake Maggiore.

dard methods for freshwater samples (APHA, AWWA, WEF 2012; APAT IRSA-CNR 2003). In addition, Chlorophyll-a concentrations are determined fluorometrically (Fluoroprobe BBE Moldaenke) in an integrated sample through the 0-20 m layer. Integrated phytoplankton samples for biovolume and density through the 0-20 m layer are also collected (CNR- IRSA, 2019).

Because of its long monitoring history, Lake Maggiore was included in the Long-Term Ecological Research Network - LTER site “Southern Alpine Lakes” (LTER_EU_IT_008), also comprising lakes Orta, Como, Iseo, and Garda. These lakes are located south of the Alps, in one of the most densely populated and urbanized areas of Europe. They represent a strategic water supply for agriculture, industry, fishing, and civil use and are an important resource for recreation and tourism (Salmaso and Mosello, 2010). Long-term trends in trophic state and other chemical and biological variables of the deep sub-alpine lakes have been assessed in a series of synoptic studies (Salmaso and Mosello, 2010; Rogora *et al.*, 2018; Salmaso *et al.*, 2020).

According to long-term monitoring data, the most relevant changes in Lake Maggiore include oligotrophication and increased thermal stability. Lake Maggiore has undergone oligotrophication since the 1980s due to reduced external nutrient loads. The lake is presently oligomesotrophic, with annual average total phosphorus concentration of 12-13 $\mu\text{g P L}^{-1}$ (Rogora *et al.*, 2018). Trends of phytoplankton biomass and chlorophyll concentrations confirmed the oligotrophication of the lake (Morabito *et al.*, 2012). After the last episode of complete mixing in 2005-2006 due to exceptionally cold and windy winter conditions (Ambrosetti and Barbanti, 1999), a tendency towards increasing stability of the water column and decreasing mixing depth has been reported as an effect of climate warming (Rogora *et al.*, 2018). The effects of climate change on Lake Maggiore have been extensively described and involve both the hydrodynamic and thermal

features of the lake (Fenocchi *et al.*, 2018) as well as its trophic and oxygen status (Rogora *et al.*, 2018) and various biological compartments (Morabito *et al.*, 2012, 2018; Nöges *et al.*, 2017; Tanentzap *et al.*, 2020). As regards phytoplankton, while nutrient inputs were the main drivers of its dynamics until the early 1990s, climatic factors started to play an important role during the oligotrophication phase (Morabito *et al.*, 2012).

Until present, monitoring of Lake Maggiore has been based on the conventional sampling methods described above (CNR-IRSA 2019; Rogora *et al.*, 2018) and HFM has been limited to meteorological and hydrological variables. However, recent evidence of the relevance of rapid ecological changes and extreme meteorological events prompted the idea that more frequent limnological data are needed to describe and timely predict fast-paced processes (Morabito *et al.*, 2018), such as algal blooms in spring and summer (Callieri *et al.*, 2014; Tapolczai *et al.*, 2015).

THE HFM SYSTEM

Hardware: the buoy, power supply, sensors, and data transmission

LM1 buoy is located in the Pallanza basin, at about 50 m from the shoreline (latitude N 45.92440°; longitude E 8.54774°; Fig. 1), anchored to three concrete blocks by steel chains at a depth of about 40 m. Two 50-Watt solar panels give electricity to a lead battery (50 A/h), which supplies all the power needed by the HFM system including the electronic control unit, the submerged sensors, the meteorological station (model gmx501, Gill Instruments, placed on the top of the buoy), the system for data transmission and a security signal light.

At present, LM1 is equipped with sensors for water temperature, pH, dissolved oxygen, conductivity, chlorophyll-a (Tab. 1). Sensors are held inside a stainless-steel guard placed 2.5 meters beneath the lake surface; an ad-

Tab. 1. Characteristics of the sensors presently installed on the LM1 buoy in Lake Maggiore.

Variable	Sensor (manufacturer)	Measuring principle/technology	Range	Resolution/accuracy
Temperature*	PHEHT, C4E, OPTOD (Ponsel/Aqualabo)	NTC (Negative Temperature Coefficient) thermistors	0-50°C	0.01±0.5°C
pH/Redox	PHEHT (Ponsel/Aqualabo)	Combined electrode (pH/reference): special glass, Ag/AgCl ref., gelled electrolyte (KCl)	0-14	0.01±0.10
Electrical conductivity	C4E (Ponsel/Aqualabo)	Conductivity sensor with 4 electrodes (2 graphic, 2 platinum)	0-200 $\mu\text{S cm}^{-1}$ at 25°C	0.01±2.00 $\mu\text{S cm}^{-1}$
Dissolved oxygen	OPTOD (Ponsel/Aqualabo)	Optical measure with luminescence	0-20 mg L^{-1} (0-200%)	0.01±0.10 mg L^{-1}
Chlorophyll-a	Cyclops-7 (Turner design)	Fluorescence	0-50 $\mu\text{g L}^{-1}$	<0.008±0.030 $\mu\text{g L}^{-1}$

*A NTC thermistor is included in each of the Ponsel sensor (for pH, conductivity, and dissolved oxygen). In our evaluation period we refer to these temperature sensors. We plan to add a thermistor chain with PT100 (Platinum resistance thermometers) to the buoy (Fig. 1).

divitive sensor for chlorophyll-a concentration is placed at a depth of about 8 m. Each sensor is wire connected to the electronic control unit and waterproofed connectors. Further sensors (for turbidity/TSS, phycocyanin and phycoerythrin, at 2.5 m depth; dissolved oxygen, at 30 m depth) and a webcam have been added to LM1 in March 2021 (*i.e.*, after the study period); a thermistor chain is planned to be added during the project.

The electronic control unit is protected by a waterproof box inside the buoy and is accessible from outside the buoy by lifting a lid. The electronic control unit has been specifically designed for signal acquisition from the sensors, data storage, basic data elaboration and wireless transfer. The electronic motherboard includes a power supply board, a Beagle Back CPU board (Processor: AM335x 1GHz ARM® Cortex-A8), serial boards for sensor data acquisition, high-resolution input boards ADC, and boards for wireless communication and solar panel management.

Software: two tiers' data storage using IstSOS

The CPU board support a Linux system allowing the following tasks:

- Local data storage (on USB support);
- Real/near real time data transfer through Ethernet, 3G/GPRS modem or radio;
- Elaboration/aggregation of local data and transfer of, *e.g.*, daily report;
- Remote reconfiguration and software update.

During the test period (December 2019-September 2020), raw data were stored in daily csv files at 1 min frequency and transmitted via wireless connection to the

CNR-IRSA server. These data were used to assess different aspects of the system functioning (*e.g.*, sensor calibration, remote and field controls, problems related to sensor drift and fouling), and how to manage them. Further, data gathered during the test period were used to test Quality Assurance/Quality Control (QA/QC) procedures to be successively adopted in the whole HFM system. In particular, data quality issues will be implemented in IstSOS, the open-source data management software selected for the HFM system (Fig. 2). IstSOS was selected due to its specific features specialized in hydro-climate data management (Pozzoni *et al.*, 2020; Strigaro *et al.*, 2019). IstSOS is a python implementation of the Sensor Observation Service standard (SOS) from the Open Geospatial Consortium (OGC), which defines a web service interface to register and retrieve metadata and observations of sensors using a standard protocols and formats, in order to increase the data consistency and interoperability. Among the numerous features of IstSOS (Cannata *et al.*, 2015), native support of data quality management is one of the most important for this application since it allows to develop data quality control procedures that associates observations with quality index (QI) that can be later used to filter or properly weight observations in subsequent elaborations. In IstSOS each registered observation is strictly bonded to a code (QI) that identifies the latest quality check test performed and successfully passed. The tests will be divided into 4 main categories:

- Tests performed on raw data coming from the sensor;
- Checks performed on aggregated data;
- Statistical tests on the time-series;
- Human driven tests and the final 'correct' flag.

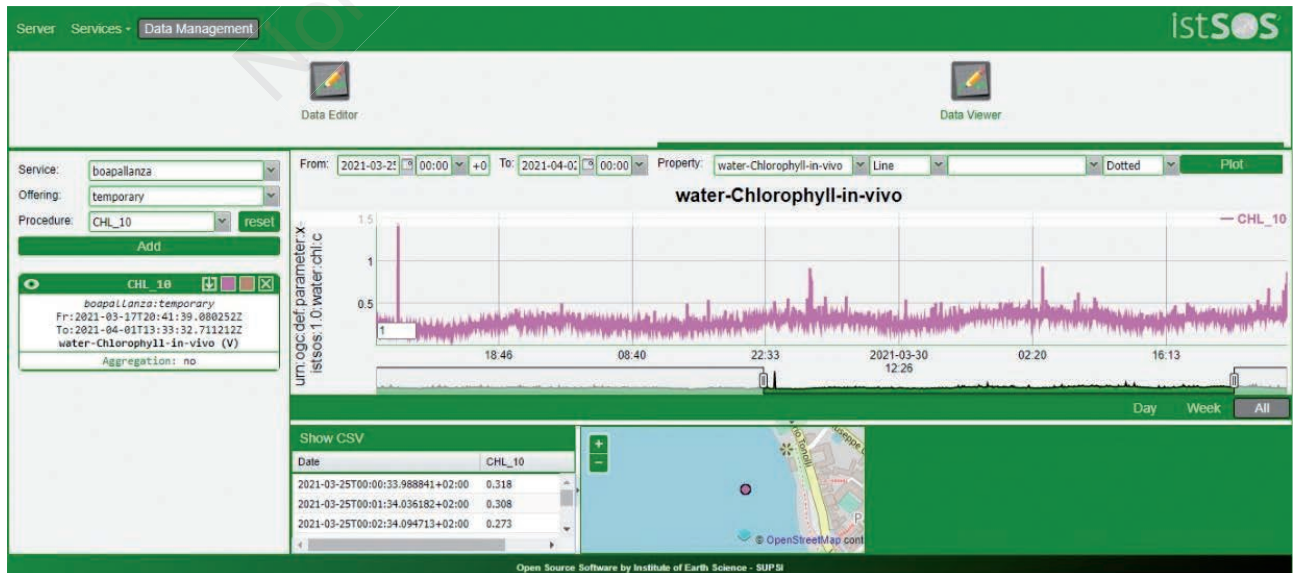


Fig. 2. Web interface of the data viewer of IstSOS showing data flowing from LM1.

The quality check tests are performed consequentially based on an increasing level of quality evaluation. The IstSOS's web interface allows to register new QC tests, such as those proposed in this work. Thanks to the lightweight of IstSOS and to the availability of a CPU support and a Linux system, it will be possible to install IstSOS directly on the monitoring buoys/platforms. It will act as a local data storage system dedicated to collect and serve raw data using standards and perform preliminary quality checks directly at the node before data are transmitted to the data warehouse. To this end, a standard communication between two IstSOS instances is enabled. The described two tiers' data storage approach is depicted in Fig. 3. The general data flow is divided into a tier at sensor side (buoy) and a tier at server side (data centre). At the buoy, data are collected from sensors and then checked for range and step tests. After that, raw observations and related QC codes are inserted into the IstSOS instance installed at the buoy. At a specific configurable time interval (suggested between 10 and 30 mins), an algorithm retrieves raw data from IstSOS, performs quality checks, aggregates and store resulting values with related quality code back to IstSOS. At user-defined interval data are then transmitted from the buoy to the data centre using selected protocol (WiFi, LoRa, NBIoT, 4G) and SOS standard requests to directly ingest them into the second tier IstSOS instance. Instantly, ingested observations are accessible for intensive processing such as for example the creation of reports, additional data validation tests, the calculation of indicators, scientific analysis, alert notifications, forecast analysis and client interfaces creation. This approach permits to locally store high frequency data and commu-

nicate aggregated data at an adequate frequency for the purpose of the monitoring. However, high-frequency data are also transmitted at fixed intervals (daily in the case of LM1 buoy) for raw data backup on a local server. This solution reduces battery consumption, bandwidth usage and transmission costs.

QUALITY ASSURANCE AND QUALITY CONTROL PROCEDURES FOR SENSOR DATA

Background and definitions

Although HFM by sensor networks can provide many benefits, sensors are susceptible to some inevitable levels of failure, which can result in lost or poor-quality data (Campbell *et al.*, 2013). It may be therefore inappropriate to make HFM data available to the public/users in their raw form before checking them, or without having taken care of the correct functioning of the entire HFM system. Establishing proper QA/QC routines guarantees that used/released data meet the expected quality standards and, at the same time, provides data usage protection. QA/QC procedures are adopted to minimize errors in data streaming from the HFM system and to document the overall quality of the data. QA and QC are closely related, but they have distinct meanings: in our specific context: we define 1) QA as the procedures taken to ensure that the HFM system and its maintenance protocols are developed in a way to minimize data loss, sensor errors and the need of subsequent corrective measures on data, and 2) QC as the procedures to test whether streaming data meet the expected quality standards including passing a series

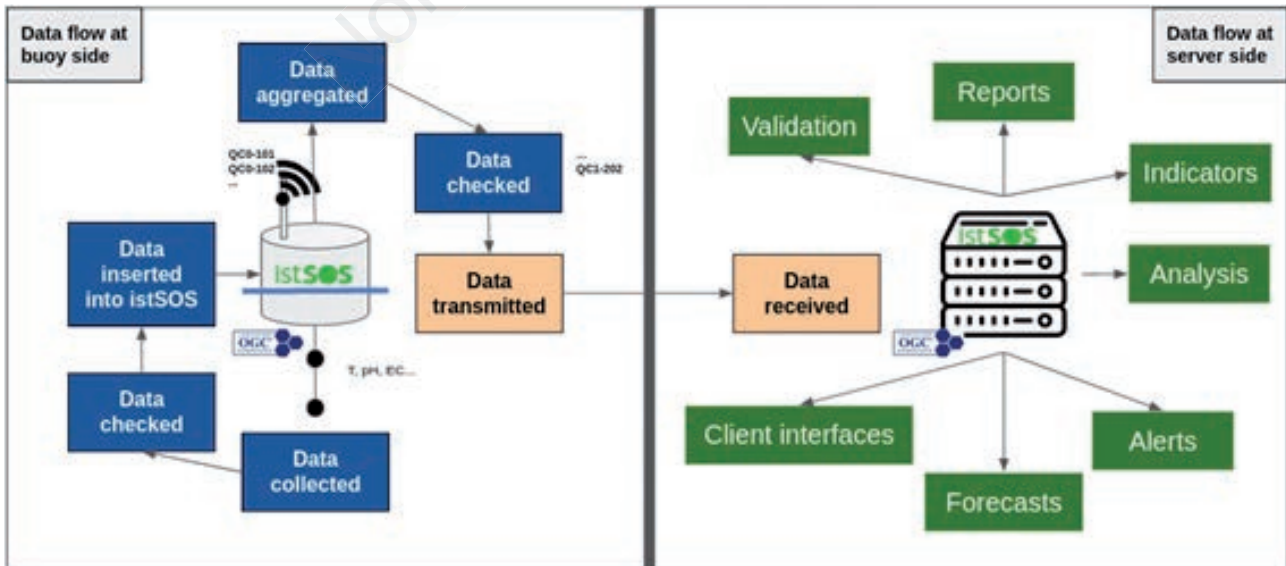


Fig. 3. Data flow from the monitoring system to server-side architecture.

of semi-automated QC tests (see paragraph “QC procedures”) and an expert-based inspection. QC procedures are necessary to make sure that field data measurements are right on (*i.e.*, accurate), reproducible (precise; consistent), and have a good estimate of their uncertainty. Associate to proper QA protocols, QC is essential to produce reliable data.

Errors in high frequency long-term datasets primarily occur from sensor biofouling and calibration shift; but errors may also occur for many other reasons (Wagner *et al.*, 2006). In Fig. 4 we provide some examples from LM1 of the most common types of error in hydrological and water quality sensor data:

- Skipped or no-data values when there are gaps in the data (Fig. 4a).
- Sensor drift, occurring because electronic drift in sensors reading away from the instrument’s calibration during the period between calibrations (Fig. 4b). This may also occur due to the aging of parts of the sensor, *e.g.*, membrane aging or dye degradation in oxygen sensors.

- Out of range values, when data values are beyond the range of plausible values for the specific phenomenon being measured (Fig. 4c).
- Sensor fouling due to biological growth (biofouling; Fig. 4d) or other residue build-up on the sensor between maintenance visits (*i.e.*, sensor cleaning), which can produce erroneous values and data truncation, *i.e.*, when data values are recorded at the reporting limit for a sensor because its maximum or minimum recording level has been exceeded.
- Failed sensor, that is an error occurring when sensor fails (*e.g.*, Chl-a sensor fails, and the system records a value equal to 0; Fig. 4e).
- Incorrect offset or calibration, when data values are in error by a constant value.
- Data value persistence, when data show constant values that are recorded when a sensor becomes stuck in a single position or when a sensor fails and the data-logger repeatedly records the last measured value.
- Power failure, when the power supply fails to provide

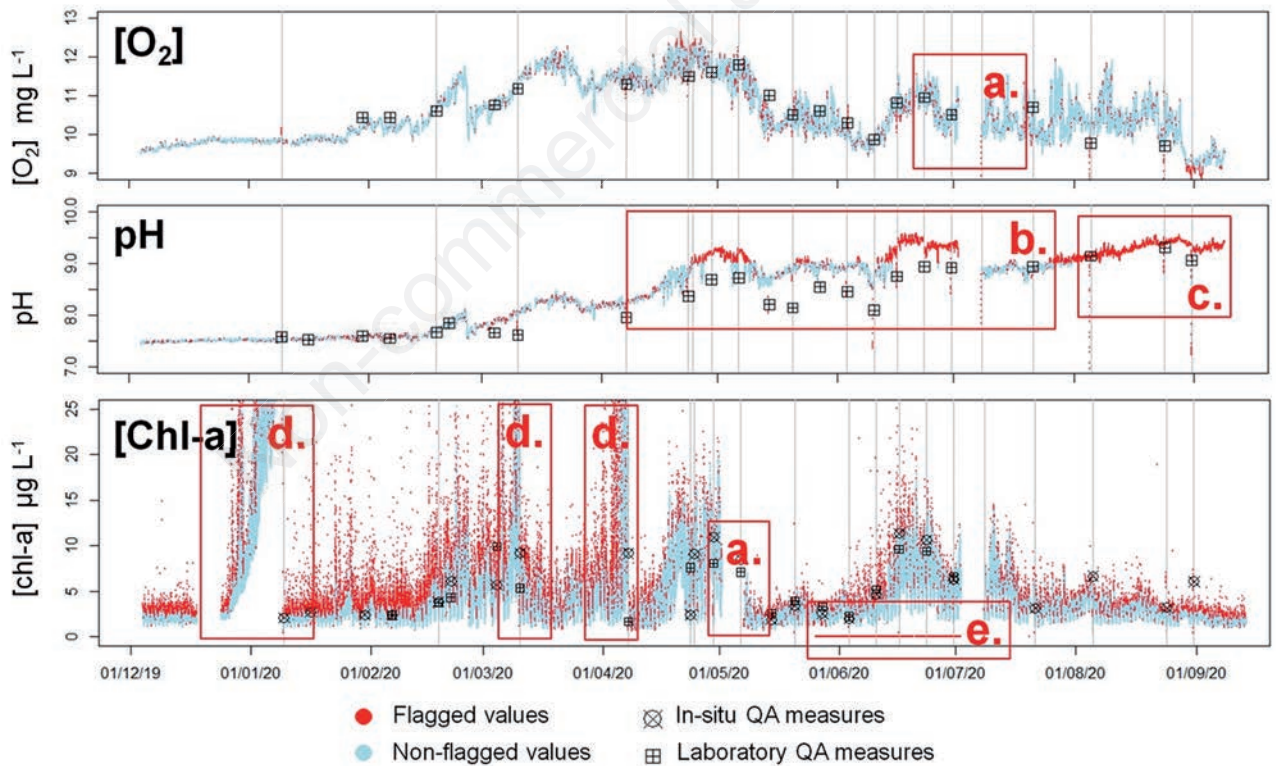


Fig. 4. Example of pre-processed data series from LM1 with i) data flags based on semi-automated quality control (see paragraph *Semi-automated QC controls*), ii) results of *in situ* and laboratory quality assurance measures (see paragraph *Is LM1 working well? Field controls and periodic reporting*), and iii) sensor cleaning dates (vertical grey lines). Red boxes highlight some common sources of error: a. missing data can occur for electric supply and memory saturation issues; b. pH sensor underwent a drift, but it was recovered by re-calibration in July 2020; c. pH values not complying with the local range criteria were flagged, but, in this case, they represent extreme but accurate values, as shown by their good match with QA measures; d. fouling is a common source of error for Chl-a sensor and it becomes clear if sensor values decrease abruptly after sensor cleaning; if data reach the reporting limits of the sensor, data truncation can be observed; e. Chl-a sensor fails produced anomalous zero values.

the required levels of power to a sensor resulting in suspect measurements.

- Adverse site conditions, when conditions immediately surrounding the sensor are not representative of the site (*e.g.*, sediment build-up in a sensor cup, ice build-up around a sensor).

Since error accumulation can rapidly disrupt data series, producing reliable HFM data requires that such errors are: i) avoided, through appropriate QA procedures; ii) identified, by correct QC procedures; and iii) eventually corrected through interactive editing to mitigate for error occurrence (Mourad and Bertrand-Krajewski, 2002). In particular, QA and QC procedures are strictly interconnected, because the identification of errors and functioning anomalies by QC procedures provide up to date information to ensure the proper maintenance and functioning of the system. Our QA/QC procedures include all the calibration, performance verification, and maintenance of LM1.

QA procedures

Sensor calibration in laboratory

Before installation, all the sensors underwent a laboratory calibration. Calibration of pH, conductivity and oxygen sensors were done according to the provision of the manufacturers, using pH buffers and standard solutions. The calibration of the chlorophyll sensors was performed using dilution series of single-species cultures of *Chlorella* sp. (Chlorophyta). We utilized two different chlorophyll sensors, by the same manufacturer (Turner Design), to assess whether their performance was similar and hold common calibration coefficients. The phytoplankton culture was added to filtered lake water (Whatman glass fibre filters, GF/C with a nominal porosity of ca. 1 μm) using 12 litres black buckets, with non-reflective linear surface. Five algal dilutions were used, covering a Chl-a range of about 0-30 $\mu\text{g L}^{-1}$. During measurements, Chl-a sensors were held 1 cm beneath the surface of the sample and an automatic mixer used. Readings were taken once the instruments had stabilized (after ca. 10 s). The sensors were rinsed and dried thoroughly between control (MilliQ water and filtered lake water) and test samples measurements. Sensor fluorescence (F; in Volt) were plotted against Chl-a concentrations ([Chl-a]; in $\mu\text{g L}^{-1}$) obtained by spectrophotometric reading after acetone extraction (APAT IRSA-CNR, 2003). The sensors yielded a linear response up to the highest chlorophyll concentration tested. However, performance of the two chlorophyll sensors differed slightly, resulting in different regression equations: $[\text{Chl-a } 1] = 7.01 F - 0.3053$, $R^2 = 0.998$; $[\text{Chl-a } 2] = 11.421 F - 0.2078$, $R^2 = 0.999$. It is thus evident that, when dealing with field fluorescent sensors, each chlorophyll sensor needs to be calibrated individu-

ally. These regression equations were provisionally set to convert sensor readings into Chl-a concentrations, subject to confirmation by field validation.

Is LM1 working? Remote control

An operator oversees controlling that LM1 is working and sending data properly on a daily basis. This procedure greatly limits the possibility of losing data but does not allow it to detect malfunctioning in real time. To overcome this problem, we created an automated procedure to query LM1 every 10 minutes and to send an email alarm message in case HFM data is not received by the server associated with LM1. Such alarm indicates that wireless transmission failed and does not necessarily imply the loss of sensor data. If the problem is limited to the transmission system, data can be recovered from a local memory placed on LM1 exactly to fix this eventual problem. However, if transmission fails because power supply problems or local memory saturation, data cannot be recorded, and a timely intervention may save a lot of data. This system allows us to intervene promptly, minimize the loss of data, and detect functioning problems.

Is LM1 working well? Field controls and periodic reporting

The correct functioning of LM1 was assessed by periodic field validations, to test the consistency between sensor data and data from standard monitoring methods. Tab. 2 provides the activity-log of the field surveys to the LM1 buoy. Data from field and laboratory measurements were compared with sensor data in order to i) check their accuracy; ii) recalibrate sensors, when necessary; iii) correct sensor data when possible.

Field surveys were planned approximately on a bi-monthly basis, unless urgent maintenance was required (*e.g.*, based on remote controls). However, to preserve one of the main advantages of HFM (*i.e.*, keeping monitoring costs low), we tried to reduce the number of field surveys when possible, *e.g.*, in periods of low lake productivity (winter), when biofouling was likely to be less severe and frequent sensor cleaning unnecessary.

To check sensor data accuracy and get information for subsequent field surveys planning, HFM data underwent periodic reporting. A mechanism for periodic reporting is an essential part of QA procedures (von Lehmden and Nelson, 1977). A weekly report was produced to provide an easy visualization of the comparison between sensor data and conventional monitoring data. Each weekly report is a description of the dynamics of the HFM sensor data collected over the previous two weeks (reporting period). The core part of the reports is a comparison between sensor data and QA data collected by conventional monitoring methods (infrequent *in situ* observations and

sampling, and laboratory analyses). Reports contain a graphical representation of the recent dynamics of sensor data and an assessment of sensor data quality based on semi-automated QC procedures (see paragraph “Quality controls”), their deviation from QA monitoring results, and an expert opinion on the data quality and desirable treatment for low-quality data (*e.g.*, data correction or deletion).

Data validation in the field

Once placed in the field, sensors should be checked for calibration and, when necessary, re-calibrated. According to Wagner *et al.* (2006), field calibration is done by

using calibration standards (*i.e.*, solutions of known quality) during field visits. This procedure ensures that sensor readings are correct and provides necessary data for re-calibration. We periodically performed such calibration checks to recalibrate the pH and conductivity sensors (Tab. 1). For all the sensors, instead -or in addition to- calibration standards, we used field data obtained from *in situ* and laboratory measurements (QA). This was done because i) we were particularly interested in assessing the performance of the sensors in the field and in comparing them to the methods conventionally used in Lake Maggiore monitoring; ii) appropriate laboratory instrumentation was available at the CNR IRSA, such as Fluoroprobe for Chl-a determination.

Tab. 2. QA activity log for LM1 station in Lake Maggiore.

Date	Time	QA Procedures		LM1 maintenance and functioning
		Cleaning	Measures	
03/12/2019	12:00			Starting LM1
18/12/2019	12:00			Electric interruption
24/12/2019	12:00			Electric problem fixed
09/01/2020	12:20	✓	✓	
16/01/2020	09:52		✓	
20/01/2020	12:00		✓	
27/01/2019	12:00	✓		
30/01/2020	08:37		✓	
06/02/2020	10:51		✓	
18/02/2020	10:18	✓	✓	
21/02/2020	14:37		✓	
04/03/2020	10:03		✓	
10/03/2020	10:22	✓	✓	
07/04/2020	11:30	✓	✓	
23/04/2020	11:47	✓	✓	
24/04/2020	12:53	✓	✓	
29/04/2020	11:44	✓	✓	Installation meteorological station
06/05/2020	11:40	✓	✓	
14/05/2020	09:27		✓	
20/05/2020	10:13	✓	✓	
27/05/2020	11:28		✓	
03/06/2020	10:37	✓	✓	
10/06/2020	09:50	✓	✓	
16/06/2020	08:53	✓	✓	
23/06/2020	09:19	✓	✓	
30/06/2020	10:17	✓	✓	
02/07/2020	07:50			Sensors removed for laboratory calibration
08/07/2020	08:07	✓		Sensors remitted
21/07/2020	11:39	✓	✓	
05/08/2020	09:15	✓	✓	
24/08/2020	10:50	✓	✓	<i>In situ</i> recalibration of chlorophyll sensor
31/08/2020	11:25	✓	✓	<i>In situ</i> test for chlorophyll sensor functioning

Field measurements and laboratory analyses from standard methods were used to check sensor accuracy and, if needed, for sensor recalibration in the field. Water temperature was measured *in situ* by taking a vertical temperature profile with a multiparameter probe (In-Situ SmarTroll MP). A 0.5 L water sample was taken with a Ruttner bottle at the sensor depth (2.5 m) for laboratory determination of pH, conductivity and dissolved oxygen (Winkler's method). Additional 2.0 L water samples were taken at the sensor depths (2.5 or 8 m; Fig. 1) for chlorophyll-a determination using two methods:

- fluorometric determination by FluoroProbe (BBE Moldaenke, GmbH, Germany). Measures were done with a Workstation benchtop unit, using a 25 ml cuvette of lake water samples;
- acetone extraction followed by spectrophotometric reading (method APAT IRSA 9020; APAT IRSA-CNR 2003).

The accuracy of the temperature, pH, conductivity, and dissolved oxygen sensors was assessed comparing results from *in situ* and laboratory QA measures with sensor data, *i.e.*, average of the last 30 measures after QA measure/sampling. For these sensors, we followed the calibration criteria of Wagner *et al.* (2006) to decide when sensor recalibration was necessary. In such cases, calibration QA measures were used as a reference to recalibrate the sensors. However, we usually observed a good fit between sensor and QA measures and limited calibration problems and periodic recalibrations using calibration standards were sufficient to maintain sensor data within the acceptable accuracy limits (*e.g.*, $\pm 0.3 \text{ mg L}^{-1}$ for oxygen concentration); a clear sensor drift was however observed for the pH sensor (Fig. 4b), needing sensor recalibration and drift correction after about six months of sensor operation. A possible problem related to the optical oxygen sensor is dye degradation; however, during the nine -months operation period, we did not observe any deviation in the oxygen sensor readings.

Because Chl-a sensors were calibrated and tested in laboratory conditions, their performance may differ in the field with lake phytoplankton communities. In addition, there are two very common factors which may affect the linear relationship between fluorescence and Chl-a concentrations, producing significant deviations between field and laboratory measures:

- the so-called fluorescence quenching (*i.e.*, the depression of the fluorescence signal in surface waters during daylight and especially at noon), which is a ubiquitous phenomenon in lakes and oceans (Marra, 1997);
- phytoplankton community composition (Richardson *et al.*, 2010).

Both factors should be taken into account to recalibrate the sensor for field measurements, *i.e.*, find an ac-

curate conversion factor for raw sensor data (from Volts to Chl-a concentrations).

Concerning quenching, this mechanism produces a daily cycle in fluorescence data which is regarded as a consequence of solar light photoinhibition on algal photosynthetic activity (McBride and Rose, 2018). As commonly reported in other HFM studies, we have observed in our data the depressing effects of quenching on the fluorescence of Chl-a during midday hours. This pattern was much attenuated at a greater depth (8 m; Fig. 5). In Fig. 5, it can be observed that around the time of sunrise (at hour 6:00) chlorophyll-a values from night values of about 1.0-1.5 Volt sharply declined to daylight values of 0.3-0.4 Volt, remaining at low levels until crepuscular time (sunset at hour 20:15), when values gradually increased again.

Neglecting quenching may have disruptive consequences on the interpretation of chlorophyll-a data series; we therefore considered the night surface measurements as the most indicative of chlorophyll-a concentration, as suggested in McBride and Rose (2018). To illustrate this issue, we used the Chl-a data (in Volt) from the surface sensor (2.5 m) and we calculated the average Chl-a concentrations in the 24 hours, or in the night immediately following water sampling (using data from one hour around midnight); we used only values recorded after sampling because of possible biofouling events (Fig. 4d). Whether the field calibration was performed using the daily or night averages, this significantly changed the calibration parameters of the relationship between fluorescence (F; in Volt) and laboratory concentration ([Chl-a]; in $\mu\text{g L}^{-1}$), *i.e.*, $[\text{Chl-a}] = 11.52 F + 0.42$, $R^2 = 0.79$ for the daily averages; $[\text{Chl-a}] = 9.02 F + 0.22$, $R^2 = 0.83$ for the night-time averages. The former equation does not account for quenching, having a higher slope and worse fit of the latter. Interestingly, accounting for quenching also provided a better concordance between the parameters of the field calibration and the laboratory calibration, run under controlled experimental conditions *i.e.*, $[\text{Chl-a}] = 7.01 F - 0.31$, $R^2 = 0.998$ (see paragraph "Sensor calibration in laboratory").

Concerning phytoplankton community composition, sensor measures of chlorophyll-a by fluorometric determination may produce an underestimation of chlorophyll-a concentrations when cyanobacteria are abundant (Bowling *et al.*, 2016), because chlorophyll-a fluorescence induction works differently for cyanobacteria and eukaryotic algae (Stirbet *et al.*, 2019). We found a good agreement between sensor data and chlorophyll-a concentrations measured by laboratory extractions, likely because cyanobacteria have usually a low biomass in Lake Maggiore (Fastner *et al.*, 2016), as confirmed by Fluoroprobe data (Fig. 6). Patterns observed in the phytoplankton community over the study period (December 2019 and September 2020; Fig. 6) have

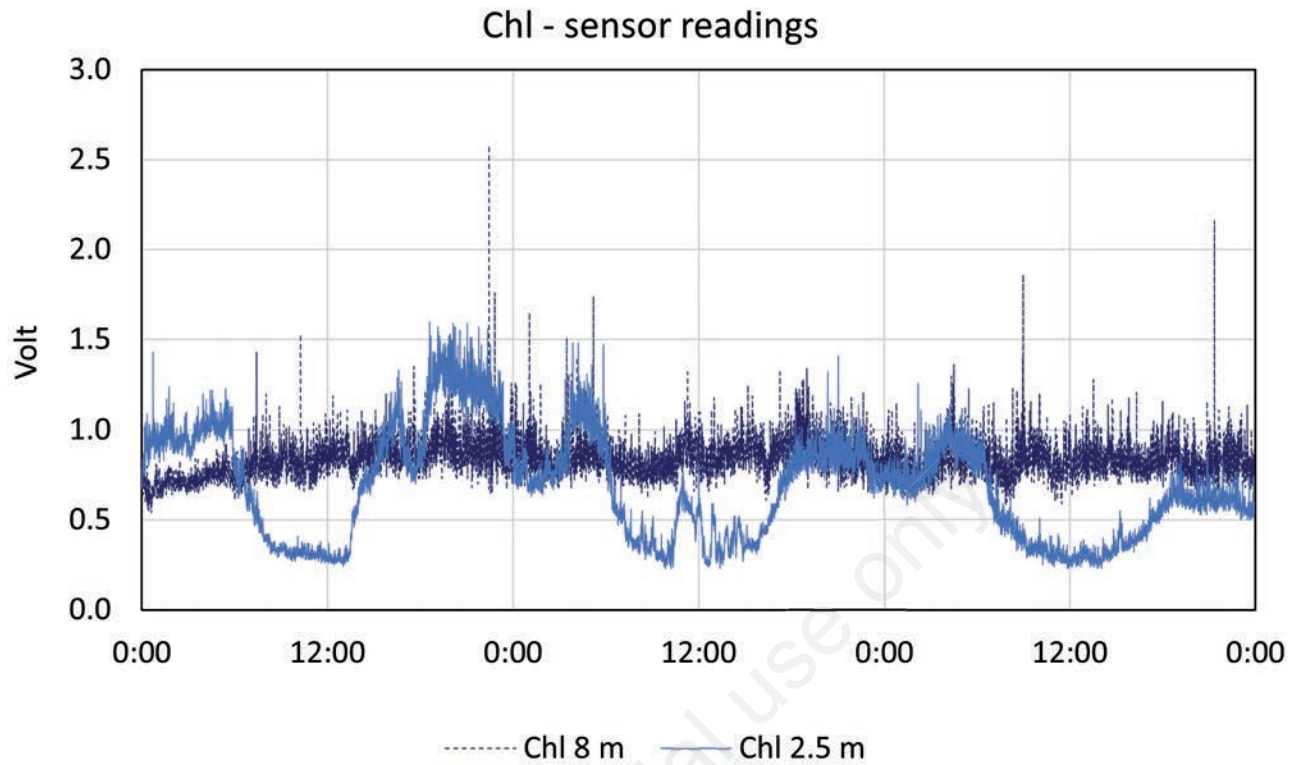


Fig. 5. Daily cycle of chlorophyll-a value recorded by the in-field CYCLOPS-7 sensors (surface sensor, at 2.5 m depth, and deep sensor, at about 8.0 m depth) on sunny days in summer 2020. The readings of the deep sensor are steadily around 0.8 Volt.

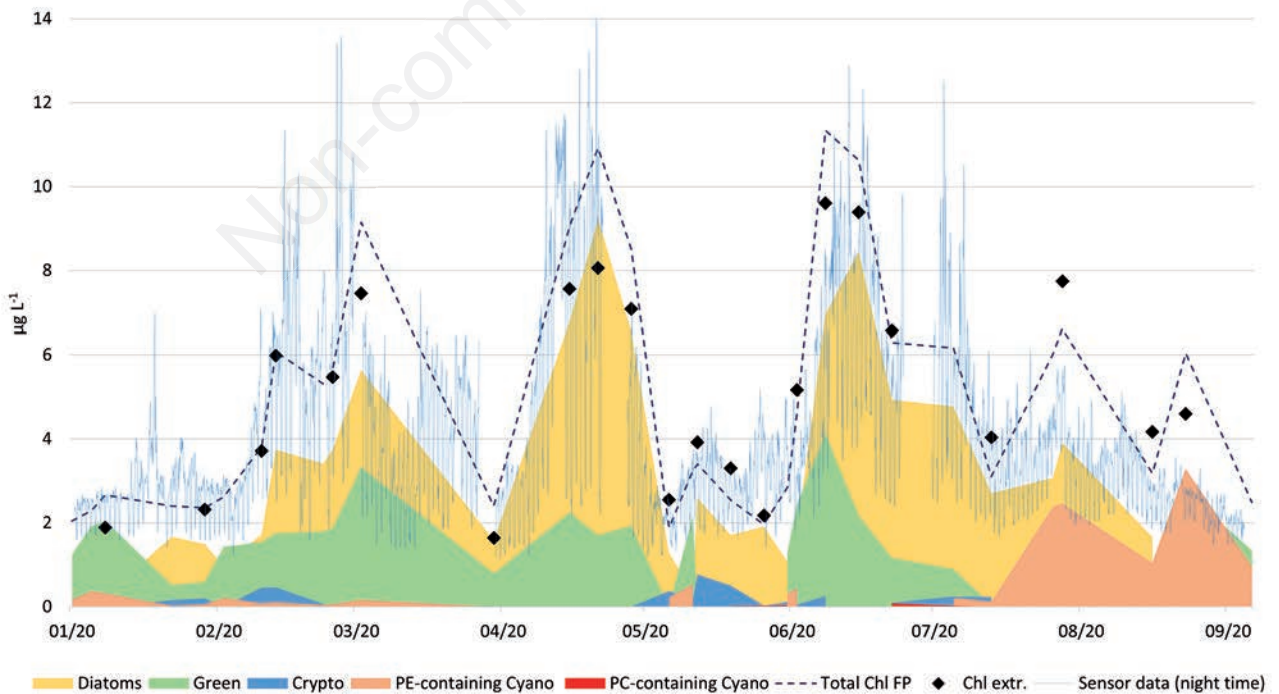


Fig. 6. Chlorophyll-a trend in Lake Maggiore as recorded by the CYCLOPS-7 sensor (night-time mean hourly values) and total chlorophyll-a concentration provided by Fluoroprobe and by laboratory extraction. The contributions of the main phytoplankton groups according to Fluoroprobe are also shown.

followed a dynamic typically observed in Lake Maggiore, with a dominance of Bacillariophyta during spring and early summer, and a more diverse community during summer including Chlorophyta and Cyanobacteria (Morabito *et al.*, 2012). In August 2020 a deviation between sensor and laboratory chlorophyll data was observed and associated to an increase of phycoerythrin containing Cyanobacteria, detected and measured by the Fluoroprobe. The increase of phycoerythrin might be due to picocyanobacteria (*i.e.*, *Synechococcus*), since they usually became abundant in late summer in Lake Maggiore (Callieri and Piscia, 2002). This condition may suggest that sensor field calibration and chlorophyll data interpretation in Lake Maggiore require careful consideration because of its phytoplankton community composition during summer periods. Performing frequent field sampling and analysis of the phytoplankton community composition could improve the comparison with sensor chlorophyll data, at least during the first phase of the HFM system running. Further, the adoption of field sensors for the detection of phyco-cyanin and phycoerythrin, as foreseen for the LM1 buoy, may be crucial to disentangle the contribute of different algal groups on the overall signal of the chlorophyll registered by the field sensor.

Sensor cleaning

Sensor fouling, in particular biofouling, is a major source of errors in limnological HFM systems. Fouling occurs because particulate matter, sediments, bacteria, microalgae or larger organisms adhere to the surface of the sensors, of their supports and cables, with possible interferences with data quality. At some places and occasions, sensors may face actual short-term biofouling effects, disrupting the quality of the measurements, sometimes in less than a week. Unless autonomous and effective antifouling techniques are adopted, a regular manual cleaning of the sensors is needed to assure the quality of the data. However, this is likely to become the most laborious task related to system maintenance (Laas *et al.*, 2016).

For LM1, biofouling was a major issue for the sensor for chlorophyll-a, requiring frequent field visits to the buoy for sensor cleaning (Tab. 2). Even if fouling was observed only for the chlorophyll sensor, all sensors were cleaned with the same frequency required by the chlorophyll-a sensor, which was probably sufficient to avoid any fouling problems for the other sensors. The minimum time elapsed between the last chlorophyll-a sensor cleaning and the recurrence of the problem was less than two weeks, but longer recurrence times were usually observed. Evidence of fouling arises from the observation of the data series, which show an abrupt decrease, just after sensor cleaning (Fig. 4d). Until now, biofouling was addressed through regular manual cleaning of the sensor surface and support and by data correction. However, finding autonomous and

effective anti-fouling systems is probably a priority to reduce the maintenance efforts and preserve one of the most attractive advantages of automated monitoring systems, *i.e.*, saving time and personnel costs.

Some techniques and technologies are available to reduce sensor fouling (Pires, 2010); they include using automated mechanical cleaning systems, antifouling paints (with active biocides such as copper), and active systems based on electro-mechanical vibration removing fouling material from the surface. However, antifouling paints are not adapted to protect sensors' sensitive parts (the interface between the measurement medium and the sensor sensitive area), which should remain as much as possible unmodified, and the power requirement of electro-mechanical methods may be too high for HFM systems. Up to now, standard methods to avoid and manage sensor fouling are not available. It is likely that biofouling may become an even more pressing problem for meso-eutrophic lakes, such as lakes Como, Varese and Lugano. An important step for the creation of an efficient HFM system for the subalpine lakes is the creation of standard antifouling protocols, which should be considered a priority for our project.

QC procedures

Semi-automated QC controls

Automated QC controls do not replace manual data inspection by experts which is ultimately needed to take the final decision on deletion or correction of erroneous or low-quality data. However, automated QC represents a necessary improvement. It ensures consistency, reduces human bias, and enables to cope with the huge flow of data from HFM systems, which is incompatible with manual QC methods historically used by ecologists (Campbell *et al.*, 2013). Automated QC control classifies data according to mathematical or logical criteria. When they do not meet such criteria, data are not removed, but labelled (or flagged) as suspicious, preserving raw data unmanipulated (*i.e.*, pre-processed data). The final decision on retaining, correcting, or deleting flagged data is taken at the final data editing. For example, extreme events may produce extreme but accurate values which may be flagged for falling outside a certain range criterion (Fig. 4c).

Most automated quality control procedures are focused on flagging raw data values that do not meet one of several plausibility tests (Sheldon, 2008; Lerner *et al.*, 2011; Taylor and Loescher, 2013). Flagging erroneous or suspicious values can require simple algorithms (*e.g.*, simple range checks) or more complicated techniques (Moatar *et al.*, 2001; Hill *et al.*, 2009; Fiebrich *et al.*, 2010). We provide an example of semi-automated QC procedures based on the package sensorQC (Read *et al.*,

2015) of the R statistical environment version 4.3.1 (R Development Core Team, 2019). This package was developed by the United States Geological Survey and specifically designed for automated QC controls for sensor data (Gries *et al.*, 2014). Our semi-automated QC procedures provide for four plausibility tests. A flag is generated and associated to the data which do not meet the criteria (Tab. 3) of the following tests:

- global range test: data are flagged unless they fall within valid regional lake ranges or instrumental limits (whichever is more restrictive) (*i.e.*, $>MAX_{global}$ and $<MIN_{global}$, respectively representing the maximum and minimum plausible values at a global level); we set MAX_{global} and MIN_{global} based on instrumental limits and literature data;
- local range test: data are flagged unless they fall within locally valid site-specific ranges (*i.e.*, $>MAX_{local}$ and $<MIN_{local}$, respectively representing the maximum and minimum plausible values at a site-specific level); we set MAX_{local} and MIN_{local} respectively equal to the upper and lower 95% Confidence Intervals of the surface values recorded during the long-term monitoring campaign of Lake Maggiore at the Ghiffa sampling station;
- stuck value test (or value persistence test): data are flagged if neighbouring values differ by less than the resolution of the sensor for more than N_{rep} repetitions, where N_{rep} is an arbitrary number of repeated data that is considered indicative of sensor stuck;
- spike test or absolute deviation around median test: data are flagged when their absolute deviation around the median value calculated over the last w previous values is larger than 3 times their median (*i.e.*, median absolute deviation - $MAD > 3$).

Visual inspection of data and manual QC editing

Many sensor errors, such as sensor drift and fouling may swell the automated procedures but become evident when sensors undergo periodic maintenance operations (*i.e.*, cleaning) or sensor accuracy is tested, *i.e.*, using calibration standards or field measures (see paragraph “Qual-

ity assurance”). Automated QC can identify and flag potentially erroneous values. However, there is also a need of choose and apply the appropriate data corrections (Horsburgh *et al.*, 2015). As stated above, to this aim, the attention and expertise of field or data technicians are required (Fiebrich *et al.*, 2010). At this stage, flags can be conveniently used to sort the data according to the criteria for automated QC, which enables saving considerable time. During the final editing, data undergo visual inspection. A readable graphic representation of HFM data associated to QA monitoring data (*e.g.*, sensor cleaning events, laboratory measurements, *etc.*; Fig. 4) is therefore necessary.

In Tab. 4 we provide the logical scheme (*i.e.*, articulated in a series of dichotomous questions) that we used for the final editing. It should be noted that at first data is evaluated based on the results of automated QC controls (flagged data), but then also non-flagged data are inspected. Final editing provides for one of three possible actions: retain data (for data that are considered accurate enough), discard data (when data quality is considered too low and correction methods inapplicable), or correct (*i.e.*, subtract/add a quantity to the raw sensor value). The final decision is often based on data accuracy as compared to QA measurements, so that QA measurements should be frequent enough to enable such comparisons and provide guarantees about the quality of HFM data. Sensor data are retained without any manipulations when they match the calibration standards or QA measures used to evaluate the sensor accuracy. A certain deviation between sensor data and QA measures is however expected, but data can be retained without corrections if they meet fixed quality standard (Tab.4).

More sophisticated methods are needed to correct for i) calibration offset, ii) sensor drift, and iii) sensor fouling (Wagner *et al.*, 2006). Hereafter we describe some methods to correct data. Being aware that these methods are not representative of the entire range of possible correction methods. However, a full description of these methods is out of the scopes of the present study. We just provide some examples of data correction from our study that can be applied in similar cases.

Tab. 3. Parameters of the semi-automated QC tests applied to the sensor data from the LM1 buoy.

Sensor	Unit	Parameters of the QC tests						
		MIN_{GLOBAL}	MAX_{GLOBAL}	MIN_{LOCAL}	MAX_{LOCAL}	N_{REP}	w	MAD
pH		3	11	7.37	9.01	5	60	3
Temperature	°C	0	50	6.60	23.00	5	60	3
Electrical conductivity at 25 °C	$\mu S\ cm^{-1}$	0	2000	135.84	166.68	5	60	3
Oxygen concentration	$mg\ L^{-1}$	0	20	9.17	12.28	5	60	3
Chlorophyll-a concentration	$\mu g\ L^{-1}$	0	34.43 ^a	0.03	34.43 ^a	3	60	3

^aUpper instrument limit.

Calibration offset occurs when sensor data deviate from QA measurements by a constant; sensor data can be corrected adding this constant and further errors avoided through recalibration. To compensate for sensor drift, we considered the difference, or deviation, between QA measures and sensor data. Sensor data were estimated as the mean value of the 60 measures (1 hour) taken starting

from 10 mins after QA measures, to avoid values temporarily affected by QA procedures. Even if sensor drift may follow different patterns, we assumed it followed a straight linear course with time. We therefore run a linear regression between sensor deviation and time to correct for sensor drift (Fig. 7a and a').

Sensor fouling was observed only for chlorophyll-a

Tab. 4. Line of reasoning for the final HFM data editing.

1.	Are the values under inspection flagged according to at least one of the semi-automated Quality Control criteria?	
	Yes	go to 2
	No	go to 12
2.	Are the values flagged according to the stuck value test?	
	Yes	go to 3
	No	go to 4
3.	Does data inspection confirm sensor stuck?	
	Yes	DISCARD
	No	go to 4
4.	Are data values flagged according to the spike test?	
	Yes	go to 5
	No	go to 6
5.	Is there a clear mechanism (ecological or physical) explaining the spikes?	
	Yes	go to 6
	No	DISCARD
6.	Are data values flagged according to the global range test?	
	Yes	go to 7
	No	go to 8
7.	Are values at or outside instrumental limits?	
	Yes	DISCARD
	No	go to 8
8.	Are values flagged according to the local range test?	
	Yes	go to 9
	No	go to 11
9.	Are values outside local range because of sensor drift or fouling?	
	Yes	go to 10
	No	go to 11
10.	Do sensor drift or fouling caused data truncation at the instrumental limits?	
	Yes (remove all data which can be affected by drift or fouling, also those data recorded before the instrumental limits were reached. Then, recalibrate or clean the sensors)	DISCARD
	No (recalibrate or clean the sensors)	CORRECT
11.	The match between sensor data and calibration standards/available Quality Assurance (QA) measures is good enough?	
	Yes	RETAIN
	No	go to 12
12.	Do values deviate from QA measures by a constant?	
	Yes	CORRECT
	No	go to 13
13.	Do values deviate from QA measures according to a recognizable pattern attributable to sensor drift or fouling?	
	Yes	go to 10
	No (sensor should be repaired or changed)	DISCARD

*Temperature error less than $\pm 0.2^{\circ}\text{C}$; EC error less than $\pm 5 \mu\text{S cm}^{-1}$; $[\text{O}_2]$ error less than 0.3 mg L^{-1} ; pH error less than 0.3 pH units; [Chl-a] error less than 15% the values recorded the night after the QA measures. Temperature, EC, pH, and $[\text{O}_2]$ criteria are based on Wagner et al., 2006; the criterion for [Chl-a] is based on personal observations of data accuracy.

sensors, becoming evident after sensor cleaning (Fig. 2d). The removal of fouling by sensor cleaning (fouling end point) results in a shift of chlorophyll-a values, which provides a measure of maximum fouling, *i.e.*, estimated as the difference between the chlorophyll-a average values 12 hours before (ca. 720 points) and 12 hours (ca. 720 points) after cleaning. The onset of fouling was estimated visually (*i.e.*, the time Chl-a data appeared to begin drift-

ing upward). We assumed fouling to be linear, so the correction for fouling (the quantity to be subtracted from the raw Chl-a data) is the linear interpolation between 0 at the estimated onset of fouling and the maximum fouling at the point of sensor cleaning (Fig. 7b).

Releasing raw or automatically corrected data from HFM systems may provide inaccurate baseline data for environmental management, scientific, sanitary, and com-

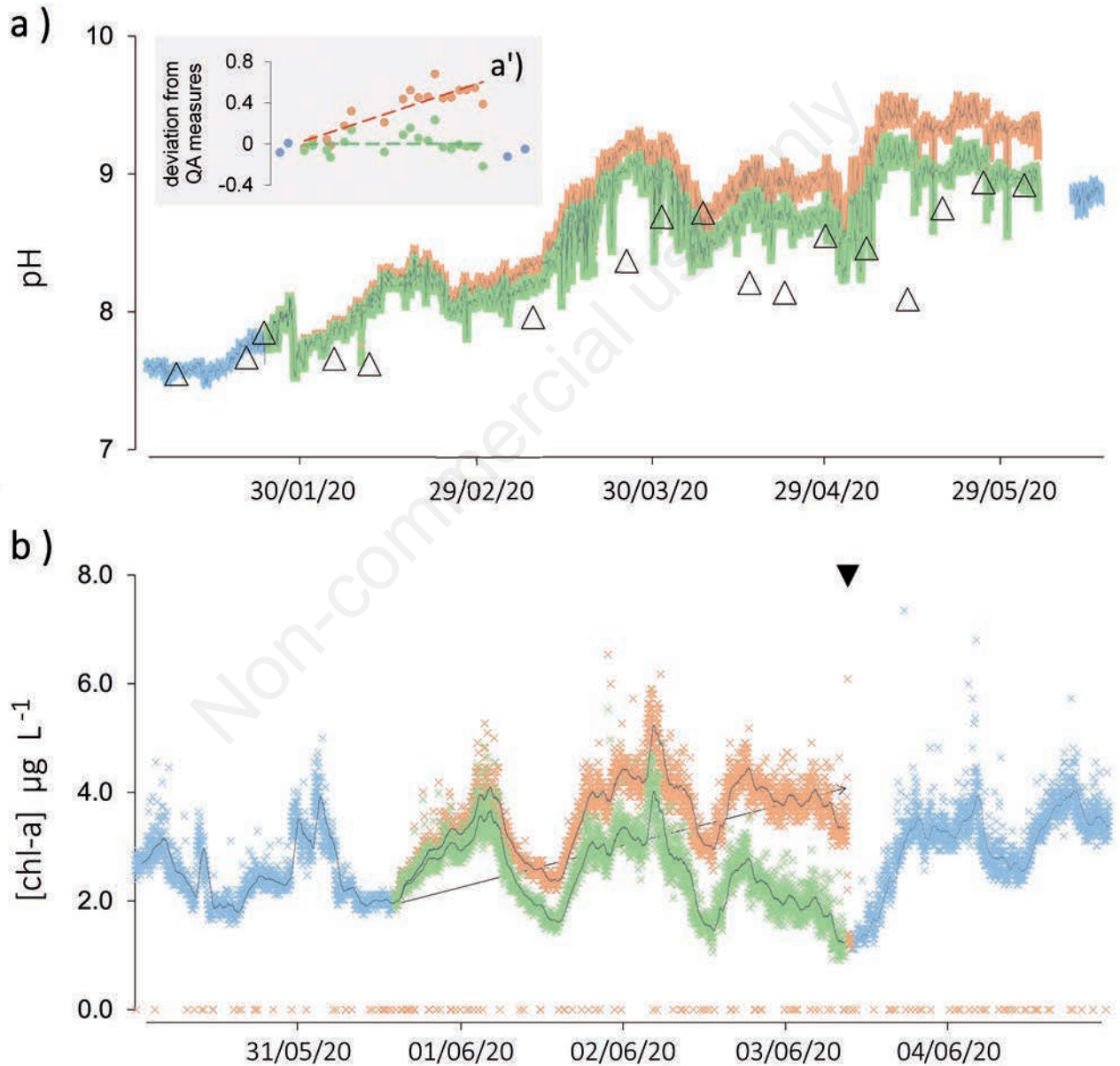


Fig. 7. a) Sensor drift correction using the parameters of the linear regression between sensor deviation and time (a'). b) Sensor fouling correction. Blue dots, retained values; red dots, discarded values affected by drift, fouling or sensor cleaning operations (*i.e.*, 10 mins before and after cleaning hour); green dots, corrected values; moving averages of the values recorded in the previous 1 hour are provided; empty triangles, laboratory pH measures; solid reversed triangle, sensor cleaning date; red and green dashed lines (a') regression lines between sensor a data and time before and after correction.

munication purposes. This should be avoided by checking the data before release or, when data are released in real- or near-to-real-time, all the possible source of errors and their effects on the flow of data should be clearly described.

TOWARDS A HFM NETWORK FOR LAKES IN THE INSUBRIC REGION

Newly available sensor technologies give the opportunity of monitoring lakes at unprecedented frequencies. However, HFM systems entails some difficulties related to the complex technology of sensors and to the management of a huge amount of data. Depending on the local context (*i.e.*, expertise availability), creating a functional HFM network may present several difficulties, essentially related to the gap in the technical know-how. LM1 should be considered as a pilot experience to start filling this knowledge gap. The first nine months of LM1 functioning enabled us to test the advantages and shortcomings of HFM and will provide the needed background to implement the rest of the HFM network in the Insubric lakes Maggiore, Como and Lugano, none of which was previously subject to *in situ* real-time monitoring.

Based on this pilot experience, we highlight four major issues for the development of an efficient HFM system at a regional level:

Sensor data accuracy: positioning sensors and data acquisition must be accompanied by shared QA/QC procedures, to produce accurate data and provide reliable and consistent information on lake water quality.

Value for money: frequent field surveys required for sensor maintenance (*e.g.*, cleaning) may undermine the cost-effectiveness of HFM (Garel *et al.*, 2009; Le Vu *et al.*, 2011). Reducing survey frequency is a major issue for the implementation of a long-term HFM of Insubric lakes, for example, developing shared methods to reduce fouling.

Long-term sustainability: LM1 hardware and software underwent several changes (*e.g.*, changing/adding electronic components, sensors, and data transmission protocols) which have been possible because LM1 was developed in-house, having full access to all the hardware and software components. This also increased replicability and transferability of our experience within and outside the SIMILE project.

Data access: making raw, real time HFM data available to the public/users before validation may be inappropriate, even when users are provided with an alerting about possible error sources. Sound quality check and validation protocols are essential to provide useful data to the users. Within SIMILE the following strategy will be adopted, in order to manage that data flow: data are automatically checked in real-time for soundness and gross-

error detection and are stored in the local istSOS (on-buoy) with an associated quality index. At specific time intervals, data are aggregated and associated with a new quality index resulting from a number of data checks, including those tested in the present study (*e.g.*, step test, local and global range tests). Successively, aggregated and quality checked data are transmitted using the selected transmission protocol to the data-centre tier consisting of several Web services dedicated to data collection, protection and serving.

The present study provides essential indication for the installation of a HFM system and for a shared QA/QC procedures to guarantee the good functioning of the system itself and a wise management of HFM data. A shared data validation protocol is still under development, based on the results we gathered during this study. The final protocol should aim at reducing the validation time to facilitate public access to the data with controlled quality. Currently, taking advantage of the experience gained with LM1, a mooring platform equipped with limnological sensors has been installed in Lake Lugano, and three buoys will be placed in different areas of Lake Como.

ACKNOWLEDGEMENTS

This work was supported by SIMILE (ID: 523544), an Interreg Italian-Swiss project funded by the European Regional Development Fund (ERDF). The *Chlorella* culture for the Chl sensor calibration experiment was kindly provided by Roberta Piscia (CNR – IRSA, Verbania). We are grateful to Cristiana Callieri for her support during the laboratory work for sensor calibration. We are also indebted with Paola Giacomotti, Arianna Orrù and Gabriele Tartari for their assistance during field sampling and chemical analysis. A great input to the development of HFM in Italian lakes (Maggiore and Orta) came from the research community of the NETLAKE and GLEON networks. We thank two anonymous reviewers for their helpful comments and suggestions on the manuscript.

REFERENCES

- Albaladejo C, Soto F, Torres R, Sánchez P, López JA, 2012. A low-cost sensor buoy system for monitoring shallow marine environments. *Sensors* 12:9613-9634.
- Ambrosetti W, Barbanti L, 1999. Deep water warming in lakes: an indicator of climatic change. *J. Limnol.* 58: 1-9.
- APAT, IRSA-CNR, 2003. [Metodi analitici per le acque. Manuali e linee guida 29/2003]. [Handbook in Italian]. APAT, IRSA-CNR.
- APHA, AWWA, WEF, 2012. Standard Methods for the examination of water and wastewater. American Public Health Association, Washington DC.

- Banas DP, Grillas I, Auby F, Lescuyer E, Coulet JC, Moreteau J, 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 of weather data to predict irradiance in the water column. *Hydrobiologia* 551:3-16.
- Bertone E, Burford M, Hamilton D, 2018. Fluorescence probes for real-time remote cyanobacteria monitoring: A review of challenges and opportunities. *Water Res.* 141:152-162.
- Bowling LC, Zamyadi A, Henderson RK, 2016. Assessment of *in situ* fluorometry to measure cyanobacterial presence in water bodies with diverse cyanobacterial populations. *Water Res.* 105:22-33.
- Brovelli MA, Cannata M, Rogora M, 2020. SIMILE, a geospatial enabler of the monitoring of Sustainable Development Goal 6 (Ensure availability and sustainability of water for all). *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* 42:4/W20.
- Callieri C, Piscia R, 2002. Photosynthetic efficiency and seasonality of autotrophic picoplankton in Lago Maggiore after its recovery. *Freshwater Biol.* 47:941-956.
- Callieri C, Bertoni R, Contesini M, Bertoni F, 2014. Lake level fluctuations boost toxic cyanobacterial oligotrophic blooms. *PLoS ONE* 9:e109526.
- Campbell JL, Rustad LE, Porter JH, Taylor JR, Dereszynski EW, Shanley JB, Gries C, Henshaw DL, Martin ME, Sheldon WE, Boose ER, 2013. Quantity is nothing without quality: automated QA/QC for streaming environmental sensor data. *BioScience* 63:574-585.
- Cannata M, Antonovic M, Molinari M, Pozzoni M, 2015. Ist-SOS, a new sensor observation management system: software architecture and a real-case application for flood protection. *Geomat. Nat. Haz. Risk* 6:635-650.
- CNR-IRSA, 2019. [Ricerche sull'evoluzione del Lago Maggiore. Aspetti limnologici. Programma triennale 2016 – 2018. Campagna 2018 e rapporto triennale 2016-18]. [Report in Italian]. Commissione Internazionale per la protezione delle acque italo-svizzere. 160 pp.
- Dubelaar GBJ, Geerders PJF, Jonker RR, 2004. High frequency monitoring reveals phytoplankton dynamics. *J. Environ. Monit.* 6:946–952.
- European Commission, 2000. Council Directive 2000/60/EC of 23 October 2000 establishing a framework for Community action in the field of water policy. *Off. J. Eur. Commun.* L327. p. 73.
- Fastner J, Abella S, Litt A, Morabito G, Voros L, Palffy K, Straile D, Kummerlin R, Matthews D, Phillips G, Chorus I, 2016. Combating cyanobacterial proliferation by avoiding or treating inflows with high P load - experiences from eight case studies. *Aquat. Ecol.* 50:367-383.
- Fenocchi A, Rogora M, Sibilla S, Dresti C, 2017. Relevance of inflows on the thermodynamic structure and on the modeling of a deep subalpine lake (Lake Maggiore, Northern Italy/Southern Switzerland). *Limnologica* 63:42-56.
- Fenocchi A, Rogora M, Sibilla S, Ciampittello M, 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 Dynam.* 51:3521–3536.
- Fiebrich CA, Morgan CR, McCombs AG, Hall PKJ, McPherson RA, 2010. Quality assurance procedures for mesoscale meteorological data. *J. Atmos. Ocean. Technol.* 27:1565-1582.
- Garel E, Nunes S, Neto J, Fernandes R, Neves R, Marques J, Ferreira, 2009. The autonomous Simpatico system for real-time continuous water-quality and current velocity monitoring: examples of application in three Portuguese estuaries. *Geo-Mar. Lett.* 29:331–341.
- Gries C, Read JS, Winslow LA, Hanson PC, Weathers KC, 2014. Enabling innovative research by supporting the life cycle of high frequency streaming sensor data in the Global Lake Ecological Observatory Network (GLEON). In: AGU Fall Meeting Abstracts 2014.
- Hamilton DP, Carey CC, Arvola L, Arzberger P, Brewer C, Cole JJ, Gaiser E, Hanson PC, Ibelings BW, Jennings E, Kratz TK, Lin FP, McBride CG, de Motta MD, Muraoka K, Nishri A, Qin B, Read JS, Rose KC, Ryder E, Weathers KC, Zhu G, Trolle D, Brookes JD, 2014. A global lake ecological observatory network (GLEON) for synthesising high-frequency sensor data for validation of deterministic ecological models. *Inland Waters* 5:49-56.
- Hill DJ, Minsker BS, Amir E, 2009. Real-time Bayesian anomaly detection in streaming environmental data. *Water Resour. Res.* 45:10.1029/2008WR006956.
- Horsburgh JS, Reeder SL, Spackman Jones A, Meline J, 2015. Open source software for visualization and quality control of continuous hydrologic and water quality sensor data. *Environ. Modell. Softw.* 70:32-44.
- Hunter PD, Tyler AN, Gilvear DJ, Willby NJ, 2009. Using remote sensing to aid the assessment of human health risks from blooms of potentially- toxic cyanobacteria. *Environ. Sci. Technol.* 43:2627-2633.
- ISO, 1992. Water quality - Measurement of biochemical parameters - Spectrometric determination of the chlorophyll-a concentration. Norm ISO 10260:1992. International Organization for Standardization Publ., Geneva.
- Iványovics V, Honti M, Osztóics A, Shafik HM, Padisák J, Yacobi Y, Eckert W, 2005. Continuous monitoring of phytoplankton dynamics in Lake Balaton (Hungary) using on-line delayed fluorescence excitation spectroscopy. *Freshwater Biol.* 50:1950-1970.
- Jennings E, Jones S, Arvola L, Staehr PA, Gaiser E, Jones ID, Weathers KC, Weyhenmeyer GA, Chiu CY, de Eyto E, 2012. Effects of weather-related episodic events in lakes: an analysis based on high-frequency data. *Freshwater Biol.* 57:589-601.
- Johnson KS, Needoba JA, Riser SC, Showers WJ, 2007. Chemical sensor networks for the aquatic environment. *Chem. Rev.* 107:623-640.
- Khan H, Laas A, Marcé R, Obrador B, 2020. Major effects of alkalinity on the relationship between metabolism and dissolved inorganic carbon dynamics in lakes. *Ecosystems* 23: 1566–1580.
- Klug JL, Richardson DC, Ewing HA, Hargreaves BR, Samal NR, Vachon D, Pierson DC, Lindsey AM, O'Donnell DM, Effler SW, Weathers KC, 2012. Ecosystem effects of a tropical cyclone on a network of lakes in Northeastern North America. *Environ. Sci. Technol.* 46:11693-11701.
- Laas A, de Eyto E, Pierson D, Jennings E, 2016. NETLAKE Guidelines for automatic monitoring station development. Technical report. NETLAKE COST Action ES1201. 58 pp.

- Laborde S, Antenucci J P, Copetti D, Imberger J, 2010. Inflow intrusions at multiple scales in a large temperate lake. *Limnol. Oceanogr.* 55:1301-1312.
- Lerner B, Boose E, Osterweil L, Ellison A, Clarke L, 2011. Provenance and quality control in sensor networks, p. 98-103. In: M. Jones and C. Gries (eds.), *Proceedings of the Environmental Information Management Conference*, Santa Barbara, University of California.
- Le Vu B, Vincon-Leite B, Lemaire BJ, Bensoussan N, Calzas M, Drezen C, Deroubaix JF, Escoffier N, Degres Y, Freissinet C, Groleau A, Humbert JF, Paolini G, Prévot F, Quiblier C, Rioust E, Tassin B, 2011. High-frequency monitoring of phytoplankton dynamics within the European water framework directive: application to metalimnetic cyanobacteria. *Biogeochemistry* 106:229-242.
- Lorenzen CJ, 1967. Determination of chlorophyll and phaeophytin Spectrophotometric equations. *Limnol. Oceanogr.* 12:343-346.
- Marra J, 1997. Analysis of diel variability in chlorophyll fluorescence. *J. Mar. Res.* 55:767-784.
- Marcé R, George G, Buscarinu P, Deidda M, Dunalska J, de Eyto E, Flaim G, Grossart H, Istvanovics V, Lenhardt M, Moreno-Ostos E, Obrador B, Ostrovsky I, Pierson DC, Potužák J, Poikane S, Rinke K, Rodríguez-Mozaz S, Staehr PA, Šumberová K, Waajen G, Weyhenmeyer GA, Weathers KC, Zion M, Ibelings BW, Jennings E, 2016. Automatic high frequency monitoring for improved lake and reservoir management. *Environ. Sci. Technol.* 50:10780-10794.
- McBride C, Rose KC, 2018. Automated high-frequency monitoring and research, p. 419-461. In: D.P. Hamilton, K.J. Collier, J.M. Quinn and C. Howard-Williams (eds.), *Lake restoration handbook: A New Zealand perspective*. Cham: Springer.
- Meinson P, Idrizaj A, Nöges P, Nöges T, Laas A, 2016. Continuous and high-frequency measurements in limnology: history, applications, and future challenges. *Environ. Rev.* 24:1-11.
- Morabito G, Oggioni A, Austoni M, 2012. Resource ratio and human impact: How diatom assemblages in Lake Maggiore responded to oligotrophication and climatic variability. *Hydrobiologia* 698:47-60.
- Morabito G, Rogora M, Austoni M, Ciampittiello M, 2018. Could the extreme meteorological events in Lake Maggiore watershed determine a climate-driven eutrophication process? *Hydrobiologia* 824:163-175.
- Mourad M, Bertrand-Krajewski JL, 2002. A method for automatic validation of long time series of data in urban hydrology. *Water Sci. Technol.* 45:263-270.
- Moatar F, Miquel J, Poirel A, 2001. A quality-control method for physical and chemical monitoring data. Application to dissolved oxygen levels in the river Loire (France). *J. Hydrol.* 252:25-36.
- Nöges T, Anneville O, Guillard J, Haberman J, Järvalt A, Manca M, Morabito G, Rogora M, Thackeray SJ, Volta P, Winfield IJ, Nöges P, 2017. Fisheries impacts on lake ecosystem structure in the context of a changing climate and trophic state. *J. Limnol.* 77:1640.
- Pilotti M, Valerio G, Leoni B, 2013. Data set for hydrodynamic lake model calibration: A deep pre-alpine case. *Water Resour. Res.* 49:1-5.
- Pires MD, 2010. Evaluation fluorometers for the *in situ* monitoring of chlorophyll and/or cyanobacteria. *Deltares Report*. Project 1203593-000. 17 pp.
- Pozzoni M, Salvetti A, Cannata M, 2020. Retrospective and prospective of hydro-met monitoring system in the Canton Ticino, Switzerland. *Hydrol. Sci. J. Ahead-of-print* 1-15.
- R Development Core Team, 2019. R: A Language and Environment for Statistical Computing. Version 3.5.2. R Foundation for Statistical Computing, Vienna, Austria.
- Read JS, Garner B, Pellerin B, Loken L, 2015. sensorQC. USGS-R.
- Read JS, Hamilton DP, Jones ID, Muraoka K, Winslow LA, Kroiss R, Wu CH, Gaiser E, 2011. Derivation of lake mixing and stratification indices from high-resolution lake buoy data. *Environ. Modell. Softw.* 26:1325-1336.
- Richardson TL, Lawrenz E, Pinckney JL, Guajardo RC, Walker EA, Paerl HW, MacIntyre HL, 2010. Spectral fluorometric characterization of phytoplankton community composition using the Algae Online Analyser®. *Water Res.* 44:2461-2472.
- 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* 824:33-50.
- Salmaso N, Mosello R, 2010. Limnological research in the deep southern subalpine lakes: synthesis, directions and perspectives. *Adv. Oceanogr. Limnol.* 1:29-66.
- Salmaso N, Buzzi F, Capelli C, Cerasino L, Leoni B, Lepori F, Rogora M, 2020. Responses to local and global stressors in the large southern perialpine lakes: Present status and challenges for research and management. *J. Great Lakes Res.* 46:752-766.
- Sheldon WMJ, 2008. Dynamic, rule-based quality control framework for real-time sensor data, p. 145-150. In: C. Gries and M.B. Jones (eds.), *Proceedings of the Environmental Information Management Conference: Sensor Networks*.
- Skeffington RA, Halliday SJ, Wade AJ, Bowes MJ, Loewenthal M, 2015. Using high-frequency water quality data to assess sampling strategies for the EU Water Framework Directive. *Hydrol. Earth Syst. Sci.* 19:2491-2504.
- Song KS, Li L, Li S, Tedesco L, Li LH, Hall B, 2012. Hyperspectral remote sensing of total phosphorus (TP) in three central Indiana water supply reservoirs. *Water Air Soil Poll.* 223:1481-1502.
- Stirbet A, Lázár D, Papageorgiou GC, 2019. Chlorophyll a fluorescence in cyanobacteria: relation to photosynthesis, p. 79-130. In: A.K. Mishra, D.N. Tiwari and A.N. Rai (eds.) *Cyanobacteria*. Academic Press.
- Stockwell JD, Doubek JP, Adrian R, et al., 2020. Storm impacts on phytoplankton community dynamics in lakes. *Glob. Change Biol.* 26:2756-2784.
- Strigaro D, Cannata M, Antonovic M, 2019. Boosting a weather monitoring system in low income economies using open and non-conventional systems: Data quality analysis. *Sensors* 19:1185.
- Stumpf RP, Wynne TT, Baker DB, Fahnenstiel GL, 2012. Inter-annual variability of cyanobacterial blooms in Lake Erie. *PLoS One* 7:e2444.
- Tanentzap AJ, Morabito G, Volta P, Rogora M, Yan ND, Manca M, 2020. Climate warming restructures an aquatic food web over 28 years. *Glob. Change Biol.* 26:6852-6866.

- Tapolczai K, Anneville O, Padisák J, Salmaso N, Morabito G, Zohary T, Tadonlécé RD, Rimet F, 2015. Occurrence and mass development of *Mougeotia* spp. (Zygnemataceae) in large, deep lakes. *Hydrobiologia* 745:17-29.
- Taylor JR, Loescher HL, 2013. Automated quality control methods for sensor data: a novel observatory approach. *Biogeosciences* 10:4957-4971.
- Tran Khac V, Hong Y, Plec D, Lemaire BJ, Dubois P, Saad M, Vinçon-Leite B, 2018. An automatic monitoring system for high-frequency measuring and real-time management of cyanobacterial blooms in urban water bodies. *Processes* 6:11.
- Vitale AJ, Perillo GE, Genchi SA, Arias AH, Piccolo M, 2018. Low-cost monitoring buoys network tracking biogeochemical changes in lakes and marine environments – a regional case study. *Pure Appl. Chem.* 90:1631-1646.
- von Lehmden DJ, Nelson C, 1977. Quality assurance handbook for air pollution measurement systems. Volume II. Ambient air specific methods (No. PB-273518; EPA-600/4/77/027a). Environmental Protection Agency, Research Triangle Park, NC: 346 pp.
- Wagner C, Adrian R, 2009. Cyanobacteria dominance: Quantifying the effects of climate change. *Limnol. Oceanogr.* 54:2460-2468.
- Wagner RJ, Boulger RWJ, Oblinger CJ, Smith BA, 2006. Guidelines and standard procedures for continuous water-quality monitors: station operation, record computation, and data reporting. U.S. Geological Survey Techniques and Methods 1–D3, Reston: 51 pp.
- Weathers KC, Hanson PC, Arzberger P, Brentrup J, Brookes J, Carey CC, Gaiser E, Gaiser E, Hamilton DP, Hong GS, Ibelings B, Istvánovics V, Jennings E, Kim B, Kratz T, Lin F-P, Muraoka K, O'Reilly C, Rose KC, Ryder E, Zhu G, 2013. The Global Lake Ecological Observatory Network (GLEON): the evolution of grassroots network science. *Limnol. Oceanogr. Bull.* 22:71-73.