UNIVERSITY OF PAVIA

DEPARTMENT OF EARTH AND ENVIRONMENTAL SCIENCES

Ph.D. in Earth and Environmental Sciences

Management and protection of lake ecosystems by enabling data analytics based on a cost-effective, open and high-frequency monitoring system

---

Author:
Daniele Strigaro

Supervisor:
Prof. Silvio Seno

Co-Supervisor:
Prof. Massimiliano Cannata

Anno Accademico 2022-2023
Ciclo XXXV

A Giovanni e Paola

Al Prof. Silvio Seno per aver reso possibile il concretizzarsi di questo progetto ed aver sempre sapientemente guidato la direzione con placidità e fermezza.

Al Prof. Massimiliano Cannata per essere sempre una fonte di nuove sfide e spunti che stimolano e arricchiscono il mio percorso professionale.

Un semplice grazie alle colonne portante della mia vita Giovanni, Paola, la mia famiglia, i miei amici, la Fede.

Contents

1 Introduction **1**
1.1 Organization of the thesis 6

2 The regulatory framework **7**
2.1 Italy 7
2.2 Switzerland 9
2.3 The CIPAIS 11

3 The software architecture **13**
3.1 The authentication 18
3.2 The orchestrator 19
3.3 istSOS 19
3.3.1 The Sensor Observation Service (SOS) standard 21
3.4 Grafana 26
3.5 Web interface 27

4 Open and cost effective digital ecosystem for lake water quality monitoring **29**
4.1 Abstract 30
4.2 Introduction 30
4.3 Materials and Methods 35
4.3.1 Area of Study 35
4.3.2 The System Architecture 37
4.3.3 The Automatic High Frequency Monitoring (AHFM) 38
4.3.4 The Data Management 44
4.3.5 Data Quality Control Process 47
4.4 Results 48
4.4.1 Data Completeness 48
4.4.2 Data Quality 49
4.4.3 Maintenance and Calibration 51
4.4.4 The Powering System Data 52
4.4.5 Comparison between AHFM System and IDRONAUT's Probes . 54

4.5 Discussion and Conclusions 61

5 Automatic assessment of lake status using an open source approach: Lake Lugano's case study 65

5.1 Abstract 66
5.2 Introduction 66
5.3 Materials and methods 68
5.3.1 Study area 68
5.3.2 Data sources 68
5.3.3 The information standardization 69
5.3.4 Automatic Data elaboration 71
5.4 Results 74
5.5 Conclusions 77

6 The Primary Production 81

6.1 Introduction 81
6.2 Materials and Methods 83
6.2.1 The area of study 83
6.2.2 The Automatic High Frequency Monitoring System 84
6.2.3 Data Management System 85
6.2.4 The Primary Production estimation 87
6.3 Results 93
6.3.1 Results comparison between methods 98
6.4 Discussion 99
6.5 Conclusions 102
6.6 Acknowledment 103

7 Conclusions 105

Bibliography 109

List of Figures 122

List of Tables 122

1 Introduction

Water is a crucial resource to humans, life in general and to a sustainable process. However it is stressed more and more by the effects of the climate changes and the human activities [1, 2]. Water or more properly the hydrosphere includes all the liquid of frozen surface waters, the groundwater and the atmospheric vapour available on the Earth. In fact, water is the most abundant substance and cover more than the 70% of the Earth surface. Even though lakes are only a minimum part of all the water in the Hydrosphere, they are important storages for fresh water as well as source of recreation and represent an irreplaceable element for household, agricultural, and industrial usages. The consequences of climate change are affecting such resources with direct and indirect impacts [3]. The progressive increment of water temperature both on the upper layer, where it is more tangible, and on the deeper water layers is one of the most evident effects. Water temperature is an important physical property since many plants, animals, and other organisms flourish only within a specific range. In addition, it affects some aspects of water quality. For example, the growth of algae and bacteria is promoted by higher temperatures which can negatively affect the growth and productivity of other aquatic lives due to a reduction of dissolved oxygen in the water [4]. Higher temperatures bring also a change in the mixing patterns that are essential to redistribute nutrients over the water column [5]. These are some of the physical effects and dynamics that are involved when the temperature parameter is considered. However, other aspects must be taken into account such as the uses associated to lakes as, even if warmer water might seem more inviting for recreational purposes, it also contributes to blooms of harmful algae and bacteria which produce toxins that can cause illnesses and contaminate drinking water supplies [6]. While these are just some examples of direct effects of climate change, indirect effects can stress this water resource with same or even higher magnitude. For example, while glaciers shrinking can bring a surplus of water in short-term, in the middle and long-term a loss of snow and ice-masses alter spatial and temporal dynamics with important changes in the relative contributions to stream flow and consequently to water reservoirs [7]. The concomitance of climate changes and human activities effects is a new mix of co-factors that never happened in the history of environmental monitoring. This situation can induce unknown dynamics and feedbacks that need to be studied [8]. The monitoring of such ecosystems is of primary importance in order to analyse the processes and think of potential mitigation actions that can increase the

quality of water and prevent possible negative effects in time. Traditionally, the water lake monitoring is conducted through monthly or fortnightly campaigns where technicians and limnologists use a boat to reach the monitoring points. The monitoring locations are chosen as representative points of an area of the lake. Then, the activities concern in person data collection using sampling bottles at different depths, manual driven probes that are lowered to the bottom of the lake using a wrinch, and the adoption of other tools to collect information about transparency (Secchi Disk), etc. In general, the instruments, in particular the sensor probes, are bought from companies that provide their own software and datalogger. Hence, the management of the data is stuck with specific tools by vendors which do not have common protocols to retrieve data sensed with probes. Such kind of monitoring is fundamental for many reasons: it provides an high level of data quality; it is fully manually controlled, so that the human role is important to minimize the errors; it has potentially long time-series, since the monitoring campaign in some cases started very early in the 19th century. Another important aspect of lake monitoring is the management of data and time series. This is an aspect that is usually and very often underestimated or even ignored. Currently, data from monitoring campaigns including those from both external and internal laboratory analyses are generally saved in heterogeneous formats. Specifically, data can be saved as MS Excel files, text files, in database files such as MS Access or in several types of formats at the same time. This leads to:

- a loss of uniformity of the data format;
- the use of terms that do not follow a precise and unique ontology;
- the lack of interoperability between data, different time series and even different databases;
- a prone-to-errors process, since the replication of data in more than one format can be done incorrectly and introduce errors in the values or otherwise in the information copied;
- a lack of data integrity, i.e. the protection of data against accidental or intentional changes of content by third parties;
- a high data latency, i.e. the time taken to process the possible requests of the dataset.

Based on the described state of the art concerning the monitoring of lakes, the PhD work came about to address two main problems:

1. the lack of adequate temporal resolution for studying the stress factors to which the lake resource is subject;

2. the lack of a data management strategy that can utilise the advantages of the digital revolution we are currently experiencing.

Starting from the first problem, the traditional monitoring is something that can't be substituted and that represents a fundamental source of information, but it has some limitations, in particular in spatial and temporal resolutions which are not enough to answer the highlighted current challenges that are affecting lakes. The Nyquist-Shannon sampling theorem established conditions to provide a sample rate that permits to capture all information of a signal [9]. Considering a sinusoidal signal of 1 Hz, a minimum of two samples along the period must be captured to fully explore it. Thus, the frequency of traditional campaigns is not enough to fully exploit and study unknown dynamics 1.1.

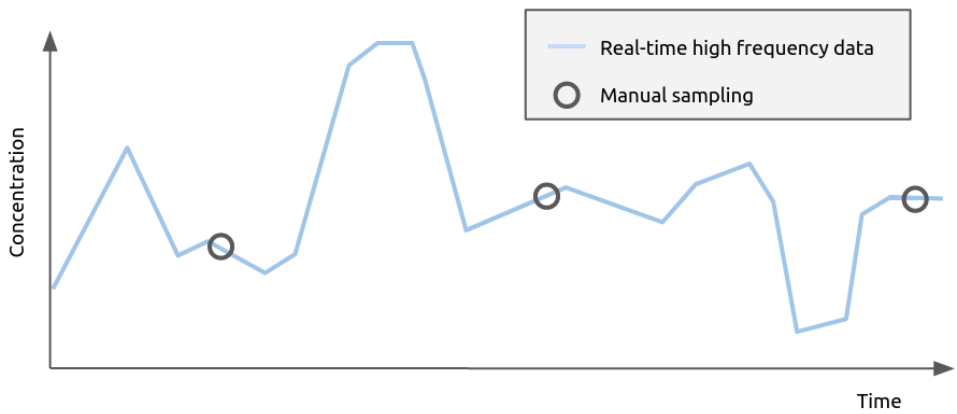

Figure 1.1: Comparison manual sampling and high frequency monitoring.

In addition to the frequency of the campaigns which could potentially further increase, there are other limits in terms of costs and of logistics. In lake monitoring, new solutions exist that could potentially solve these issues. Since February 2019 on Lake Geneva a monitoring platform, called LéXPLORE, with a size of 10 x 10 m has been installed [10]. The floating platform offers advanced sensors with high sensitivity for high-throughput analyses. However, the system is strongly oriented to scientific research, with very high level of

technical systems and very high costs of implementation, including purchase of fully proprietary solutions. Another example is the buoy built on Lake Maggiore which is a much smaller custom solution oriented toward specific needs of monitoring [11]. Unfortunately, this solution lacks in using standards and it is very difficult to replicate due to the strong custom components. While the described system could provide a solution to the temporal resolution, the satellite images can offer a way to an information spatially distributed along the surface of the lake. Nevertheless, it is not a robust source of information since strongly susceptible to many factors [12] (clouds, calibration, light scattering, etc.). The second problematic regards in particular the management of data. In the field of lake monitoring, there are many data sources that are characterised by systems that often do not talk to each other. The lack of interoperability is due to the inability to have systems and communication standards that can facilitate the exchange and sharing of information. The digital revolution and in particular the digitalisation process we have been witnessing over the last decade is increasingly raising awareness on the importance of data interoperability, which is essential for a full exploitation of data thanks to te creation of new series that can foster the understanding of a monitored object [13, 14]. So-called Big Data are not useful if it is not possible to carry out analyses between different types of data and find correlations and relationships that can create real value. Taking into account the data which are available on the lake resource, solving the problem of integration and analysis is key, in order to be able to have new information and systems that can also react in real time to certain events (e.g. algal blooms). The possible reasons that have led to a situation in which the digital revolution is struggling to take hold and bring about technological and analytical changes can be summarised as follows [15]:

- lack of digital expertise in a wide variety of fields of interest. This issue does not only relates to monitoring, but all areas of work. There is an increasing need for transversal professionals who master IT tools to facilitate data management, analysis and security [16];
- resistance to change [17]. In many situations it is difficult for workers to adapt to new technologies and solutions;
- lack of an easy-to-play solution. In environmental monitoring, as in other fields, it is normal to deal with a varied selection of software to manage data tools and analyses;

- lack of data interoperability between systems. The availability of different data sources (and software to manage them) leads to a lack of interaction caused by the fact that proprietary solutions often follow marketing strategies which offer 'closed' products that can only be solved using proprietary suites.

- costs. Proposing changes and new solutions certainly has the disadvantage of higher initial investment costs. However, in the long run, the benefits are often greater, both in terms of costs and in terms of new opportunities that can be created.

Based on the state of the art described and the problems highlighted, the doctoral research was based on two main research questions:

1. Can an integrated and replicable solution answer to these needs and be used in an applied operative monitoring at scale?

2. How might this system help in tackling the local effects of climate changes?

The project proposal aims at answering to the identified research questions through the creation of an integrated system which uses as much open source licensed technology as possible and is open source itself. The idea is to create a complete pipeline that can integrate different data sources by means of processes that can make the time series organised and accessible and then be served via standard services (Figure 1.2). Data integration allows further analysis of the data to produce new time series either by manual or automatic processes. This proposition also includes the creation of a Automatic High-Frequency Monitoring (AHFM) system built using cost-effective principles and meeting open design requirements.

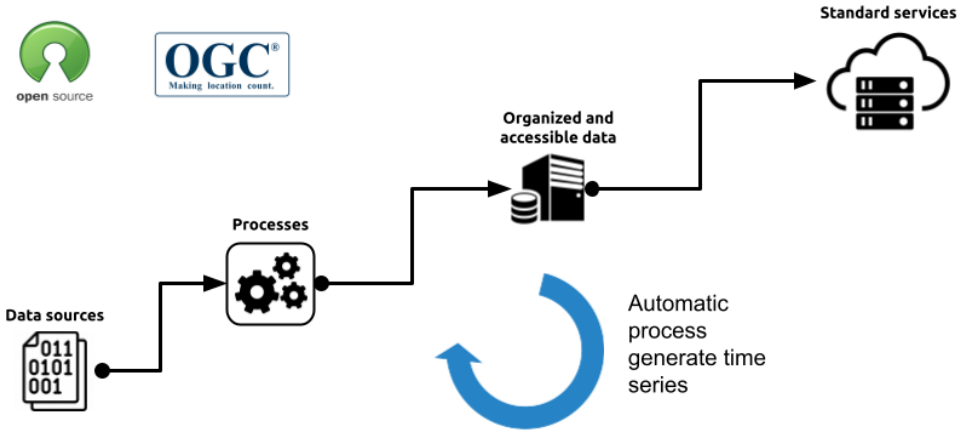

Figure 1.2: Scheme of the pipeline of the proposed solution.

1.1 Organization of the thesis

The PhD project is focused on the development of this architecture using Lake Lugano as the main testing, development and validation area for the entire system. Based on the above, this doctoral thesis is divided into five main chapters. The first presents an overview of regulations and laws to which the study area is subject, being Lake Lugano's waters extended into both Italy and Switzerland. The second chapter describes the software structure and architecture underlying the developed data management system. The third chapter presents the proposed hardware solution for the development of an AHFM system and the results of its validation process, which took over more than a year of operation. This chapter is an open access paper published into the peer-reviewed *Sensors* journal by MDPI. The fourth chapter details the automatic processes that were developed once the different information sources were imported into the software, including the algorithms and methodology used during the implementation. As the previous chapter it has been published as open access paper into *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* . Finally, the fifth chapter aims at assessing the impact the created system has towards environmental analysis and in particular the estimation of the metabolism of Lake Lugano using real-time data collected by the AHFM system. A comparison between the traditional method and the new one is proposed and evaluated. This last chapter is an original part which will be submitted for publication during the year 2023.

2 The regulatory framework

This research has been conducted on Lake Lugano which is a transboundary lake between Italy and Switzerland. In the following sections a breif overview on the regulatory framework of the two countries is presented.

2.1 Italy

In the water matter, Italy, as a member of the European Union (EU), refers to the Water Framework Directive 2000/60/CE which defines the quality status, the management and protection of the water resources and delineates in which directions the countries' actions . The Directive establishes a framework for the protection of inland surface waters, groundwater, transitional waters and coastal waters. The main goals are the prevention and reduction of pollution, the promotion of sustainable water use, the protection of the environment, improving the condition of aquatic ecosystems and mitigating the effects of floods and droughts. The main aim is to achieve an ecological and chemical 'good status' of all european waters. The Water Framework Directive pursues a systemic approach and stipulates that Member States, after having identified river basins and assigned them to river basin districts, are obliged to carry out analyses concerning the characteristics of each district, the impact of human activities on water, the economic aspect of water use, the drawing up of a register of areas to which special protection has been attributed and a census of all bodies of water used for the exploitation of water intended for human consumption with specific characteristics. The European legislation requires that each river basin district must have a management plan which can be adapted on the basis of the analyses results. Italy, in implementation of Delegated Law 308 of 15 December 2004, issued Legislative Decree no. 152 of 3 April 2006, Environmental Regulations, with which it brought together and coordinated the regulatory discipline of the different sectors of environmental law; specifically, in implementation of Directive 2000/60/EC, Title II of Part Three regulates the subject of water resources, defining the quality objectives to be achieved for surface and underground water bodies. Subsequent regulations contain technical criteria for the characterisation, monitoring and classification of water bodies for the amendment of the technical regulations of Legislative Decree 152/2006, namely Decree 131/2008 (typification, identification of water bodies, pressure analysis), Decree 56/2009 (monitoring and identification of reference conditions) and finally Decree 260/2010 (classification). In considera-

tion of what is regulated on the subject of hazardous substances by Directive 2000/60/EC, which provides for the reduction and gradual elimination of pollution and that the environmental quality standards is defined on the basis of the criteria of toxicity and ecotoxicity, Italy adopted Legislative Decree No. 219 of 10 December 2010 implementing Directive 2008/105/EC on environmental quality standards in the field of water policy. Finally, Italy transposed the Directive 2013/39/EU with Legislative Decree 172/2015, amending the list of priority substances set out in the Legislative Decree 152/2006.

Legislative Decree 152/06 provides for two levels of planning at different territorial scales:

- the river basin district management plan: the Po River Basin District Management Plan 2015, published on 22 December 2015 and approved by resolution of the Institutional Committee of the Po River Basin on 3 March 2016 by and with the DPCM of 27 October 2016, constitutes the latest update of the 2010 Management Plan;

- Water Protection Plan (PTA), which is a specific sector plan drawn up by the Regions: with D.C.R. 117-10731 of 13 March 2007, the Piedmont Regional Council approved the Water Protection Plan (PTA), an instrument aimed at achieving the quality objectives of water bodies and more generally to protect the entire surface and underground water system of Piedmont.

With regard to the use of water resources, there is another order of reference for the quality of water for human consumption, bathing and boating in lake waters.

With reference to legislation for human consumption of fresh surface water, Directive 2000/60/EC Article 7 states that, within each river basin district, Member States shall identify all water bodies used for the abstraction of water intended for human consumption, which provide more than 10 m^3 per day on average or serve more than 50 people and those water bodies intended for such future use. The reference legislation is Directive 98/83/EC (as amended by Directive 2015/1787/EU), transposed in Italy by Legislative Decree 31/2001.

As far as bathing is concerned, the Italian regulatory references are Legislative Decree No 116 of 30 May 2008, transposition of Directive 2006/7/EC on the management of bathing water quality, and the Ministerial Decree of 30 May 2010, definition of the criteria for determining the prohibition of bathing,

as well as the modalities and technical specifications for the implementation of Legislative Decree 116/2008.

Navigation on lakes is regulated at regional level by the Consolidated Text of Regional Laws on transport L.R. no. 11 of 14 July 2009 for Lombardy and the Provisions on inland navigation, inland waterways and conferral of functions to local authorities L.R. of 17 January 2008 no. 2 for Piedmont.

Legislative Decree No. 230 of 15 December 2017 sets out measures for the adaptation of national legislation to the provisions of Regulation (EU) No. 1143 of the European Parliament and of the Council of 22 October 2014, with particular reference to:

- official controls necessary to prevent the deliberate introduction of invasive alien species of EU relevance;
- the granting of authorisations provided for in Articles 8 and 9 of the Regulation;
- the establishment of the national surveillance system provided for in Article 14 of the Regulation;
- management measures aimed at the eradication, population control or containment of populations of invasive alien species of EU, transnational or national relevance;
- the discipline of sanctions for infringement of the provisions of the regulation and the decree.

2.2 Switzerland

On the basis of the powers assigned to the Confederation by the Federal Constitution (Art. 74 Protection of the Environment and Art. 76 Water), laws and ordinances have been enacted to ensure the sparing use of water resources, their protection and defence against the harmful effects of water. At the institutional level of the cantons, they are responsible for the implementation of federal laws and at the same time have their own hierarchical legislative competence (at cantonal level, the principles of federal law may be taken over or applied more restrictively). According to the Federal Law (LPAc) of 24 January 1991 on the Protection of Waters, the objective to be pursued is comprehensive protection; indeed, the law states that waters must be protected

from detrimental effects so that their multiple functions are not compromised. Within the scope of application of the Act include:

- water quality controls of lakes, rivers and groundwater;
- sewage treatment;
- controls of industrial and artisanal waste water;
- minimal outflows;
- the preservation of water as landscape elements and as ecosystems.

Subsequently, the Water Protection Ordinance (OPAc) of 28 October 1998 laid down ecological objectives for surface water and requirements concerning the quality of surface water and groundwater.

With the aim of publicising the development objectives for Swiss watercourses, in 2003 the responsible federal offices drew up the Guidelines for Swiss Watercourse Management, which outline measures that cantons, regions and municipalities can take and provide good examples for specialists and the interested public. In 2011, the UFAM then published the Guidelines for Integrated Water Management in Switzerland. In addition, since 1998, the Federal Office for the Environment, in cooperation with Eawag, the Institute for Water Research at the Swiss Federal Institutes of Technology and the cantons, has defined standardised methods for the analysis and assessment of watercourses using a system based on modules and levels, which includes the analysis of the following aspects: hydrology, water structure (ecomorphology), water chemistry and ecotoxicology, biocenosis of animals, plants and microorganisms (biology).

Existing regulations governing the different types of water use include:

- Confederation
  - Federal Fisheries Act of 21 June 1991 (LFSP);
  - Federal Act of 21 June 1991 on the Development of Watercourses.
- Canton Ticino
  - the Cantonal Law implementing the Federal Act against Water Pollution of 8 October 1971 (of 2 April 1975), which regulates the measures within the Canton's jurisdiction against water pollution;

- the Cantonal Groundwater Law (of 12 September 1978) regulating the assessment and usage of groundwater;
- Cantonal Water Supply Act (of 22 June 1994) regulating the competences and tasks of the municipalities and the canton to ensure a normal supply and sparing use of water, in particular drinking water, as well as the funding of measures necessary to achieve this goal. measures necessary to achieve this end;
- the Cantonal Regulation on the hygiene of lake and river bathing waters of 12 July 2011, which establishes the hygiene requirements for bathing water and the measures to be taken to protect the health of bathers;
- DFI Ordinance of 16 December 2016 on drinking water and water for swimming pools and showers accessible to the public and Cantonal Water Supply Act of 22 June 1994 which regulate competences and tasks in the area of the supply and sparing use of drinking water;
- the Federal Law of 3 October 1975 on Inland Navigation (LNG) and the corresponding Cantonal Law application law of 22 November 1982, which lays down the regulations relating to inland navigation within the scope of the powers delegated by federal legislation, taking environmental and ecological requirements and the welfare of the population. population.

Finally, we would like to point out that a draft of a new Ticino cantonal law on water management (LGA) is currently under consultation. This law brings together all the cantonal provisions that currently regulate the matter and at the same time repeals six existing cantonal laws, implementing the principle, now consolidated in practice, of integrated water management.

2.3 The CIPAIS

Since 40s, the scientific community and, in particular, members of the Italo-Swiss Fisheries Commission (CISPP) showed new interest in the topic of pollution. From 1945 and 1965 transbouandary water bodies were particularly stressed bacause of massive industrial discharges. Such events caused serious episodes of fish mortality. At the end of 50s, the CISPP commission decided to form an International Commission for the Protection of the Italy-Swiss

waters (CIPAIS) during the meeting helded on 4 November 1960 in Lugano. Such commission has operatioal autonomy at technical and scientific level, with the task of preparing regular reports. During the next years, the CIPAIS became more and more important so that with the agreements signed in Rome on 20 April 1972 and, after the planned Exchange of Notes, the Convention between Switzerland and Italy on the Protection of Swiss-Italian Waters from Pollution was ratified in 1973 and came into force on 7 August 1973. In order to coordinate their efforts, the Italian and Swiss Governments agreed to cooperate closely to protect the folowing lakes together with other watercourses:

- Lake Maggiore or Verbano
- Lake Lugano or Ceresio

Thanks to this cooperation, on Lake Lugano, the case study of this research, limnological research was conducted almost identical to that carried out for Lake Maggiore, focusing on the restoration prospects, in line with the objectives of achieving the mesotrophic condition provided by both the Federal Ordinance on the Protection of Water Protection Ordinance (OPAc 1998, annex 2, cf. 13) and the CIPAIS objectives. In particular, the three are the goals aimed at:

- reduction of external loads of total phosphorus and nitrogen;
- monitoring the evolution of the internal nutrient load;
- monitoring of trophic evolution.

3 The software architecture

The software architecture has been designed to host historical time series collected from the traditional surveys and from the monitoring system designed, developed and deployed within this doctoral project. From the elaboration of such data, it is possible to derive indicators which can offer an overview on the biological, chemical and physical status of the lake. The system design, showed in Figure 3.1, is composed by integrated services which different roles which communicate to each other using internal protocols. Docker is the core technology selected to permits the orchestration of all these services. In fact, each service is a docker container with the characteristic to be isolated and easily integrated with other containers to create composite software architectures. The docker containerization is a technology that allow and facilitate the replicability of the system to answer to one of the project need which was to enahnce and create and easy to go solution for the isntallation of a data management system. The application relies on some open source software:

- istSOS (www.istsos.org) to handle data according to the Sensor Observation Service (SOS) standard of the Open Geospatial Consortium (OGC);
- Grafana (www.grafana.org) to generate graphs and dashboards;
- PostgreSQL (www.postgresql.org) as database technologies where data are archived;
- Keaycloak (www.keycloak.org) to handle authentication, users privileges and single sign-on access;
- an Orchestrator developed in Python to handle permissions and shunt requests to the correct service. In addition, the orchestrator is responsible to enable the development of plugins to add features to the software suite, as for the indicators calculations.
- a Graphical User Interface to get access to all these feature to users as per their predefined role (viewer, editor and administrator). Th GUI has been developed using frameworks such as ReactJS, NextJS and Node.

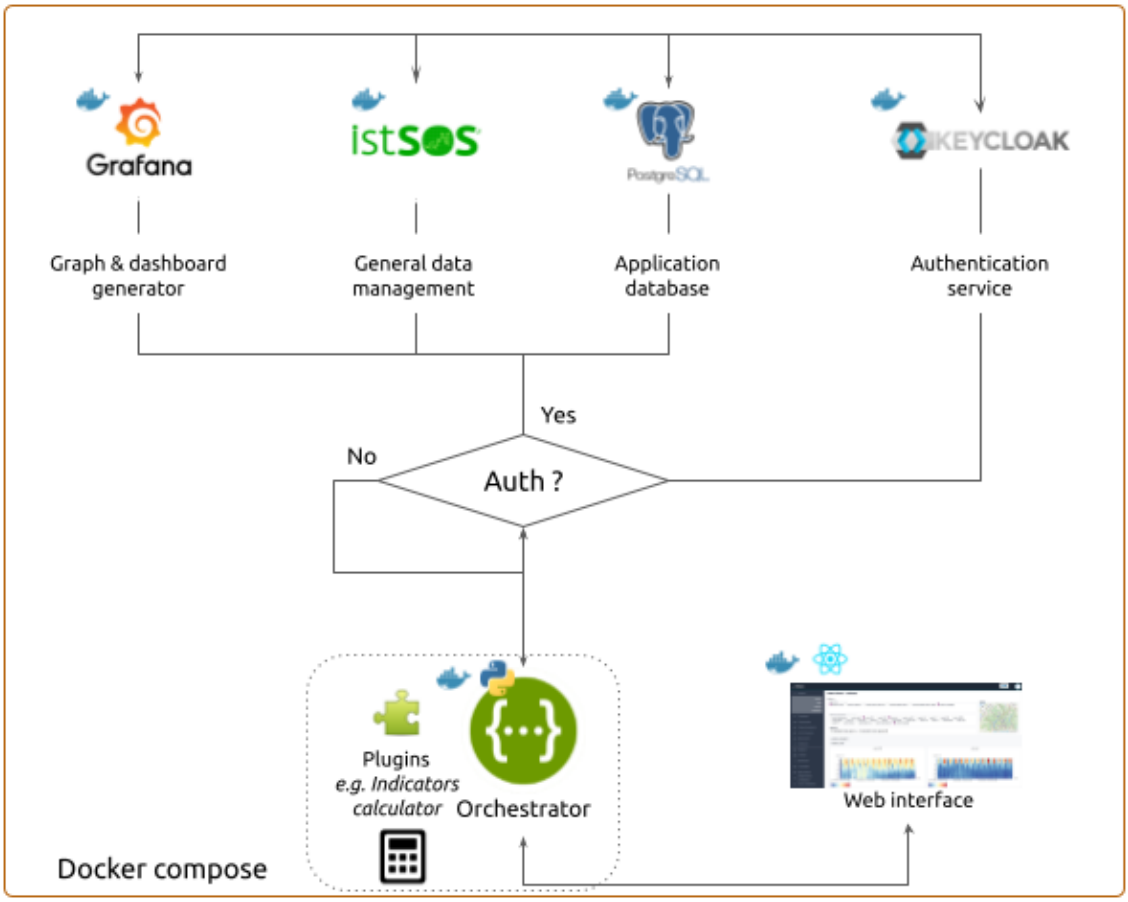

Figure 3.1: Design of the software architecture developed within this project.

The Compose technology has been used for the automatic generation of the services whose have been implemented as individual docker container. Compose is a tool for defining and starting multi-container docker applications which are defined in a YAML file where the various configurations can be set. The configuration file of the software in production mode is showed in the following code block. Various parameters have been defined and configured to allow the application, which is otherwise very complex to manage and launch, to be simplified, while maintaining the possibility of customising ports, domains and access credentials. These parameters are present in an environmental file containing environment variables that the user can modify before performing the first initialisation phase of the application. Once this phase has been started, the application will be available and accessible with all its functionalities. The following paragraphs summarise and better describe each service.

```
version: '3'
```

```
services:

  #################
  # VERNEMQ
  #################
  vernemq:
    image: vernemq/vernemq:1.10.4.1-alpine
    restart: always
    environment:
      - DOCKER_VERNEMQ_USER_SIMILE=simile
      - DOCKER_VERNEMQ_ACCEPT_EULA=yes
      - DOCKER_VERNEMQ_LOG__CONSOLE__LEVEL=debug
      - DOCKER_VERNEMQ_VMQ_PASSWD__PASSWORD_RELOAD_INTERVAL=10
      - DOCKER_VERNEMQ_PLUGINS__VMQ_PASSWD=on

  ##################
  # ISTSOS
  ##################
  istsos:
    image: registry.gitlab.com/ist-supsi/simile/istsos-suite/
       istsos:1.0.4
    restart: always
    depends_on:
      - postgresql
      - keycloak
    volumes:
      - ./services/istsos/services:/usr/share/istsos/services

  ##################
  # KEYCLOAK
  ##################
  keycloak:
    image: registry.gitlab.com/ist-supsi/simile/istsos-suite/
       keycloak:1.0.3
    restart: always
    depends_on:
      - postgresql
    environment:
```

```
    - KEYCLOAK_DATABASE_HOST=postgresql
    - KEYCLOAK_DATABASE_PORT=5432
    - KEYCLOAK_DATABASE_NAME=keycloak
    - KEYCLOAK_DATABASE_USER=${POSTGRES_USER}
    - KEYCLOAK_DATABASE_PASSWORD=${POSTGRES_PASSWORD}
    - KC_HOSTNAME_PATH=${GN_BASE_PATH}/auth
    - KC_PROXY=edge
    - KC_HOSTNAME_STRICT=false
    - KC_HTTP_RELATIVE_PATH=${GN_BASE_PATH}/auth

  ##################
  # POSTGRESQL
  ##################
  postgresql:
    image: registry.gitlab.com/ist-supsi/simile/istsos-suite/
       postgresql:1.0.2
    shm_size: '1gb'
    restart: always
    ports:
      - ${GN_PG_PORT}:5432
    environment:
      - POSTGRES_USER=${POSTGRES_USER}
      - POSTGRES_PASSWORD=${POSTGRES_PASSWORD}
      - TZ=Europe/Zurich
    volumes:
      - ./services/postgresql/data:/var/lib/postgresql/data

  ##################
  # GRAFANA
  ##################
  grafana:
    image: registry.gitlab.com/ist-supsi/simile/istsos-suite/
       grafana:1.0.6
    restart: always
    depends_on:
      - keycloak
    environment:
      - GF_DEFAULT_APP_MODE=production
      - GF_SERVER_HTTP_PORT=3000
```

```
    - GF_SERVER_ROOT_URL=${GN_BASE_PATH}/grafana/
    - GF_SERVER_SERVE_FROM_SUB_PATH=true
    - GF_SECURITY_ALLOW_EMBEDDING=true
    - GF_SECURITY_COOKIE_SAMESITE=lax
    - GF_PATHS_PROVISIONING=/var/lib/grafana/provisioning
    - GF_AUTH_GENERIC_OAUTH_AUTH_URL=${GN_BASE_PATH}/auth/realms
      /istsos/protocol/openid-connect/auth
    - GF_AUTH_GENERIC_OAUTH_TOKEN_URL=http://keycloak:8080${
      GN_BASE_PATH}/auth/realms/istsos/protocol/openid-connect/
      token
    - GF_AUTH_GENERIC_OAUTH_API_URL=http://keycloak:8080${
      GN_BASE_PATH}/auth/realms/istsos/protocol/openid-connect/
      userinfo
    - GF_RENDERING_CALLBACK_URL=/grafana
    - GF_AUTH_GENERIC_OAUTH_TLS_SKIP_VERIFY_INSECURE=true
    - GF_AUTH_LOGIN_COOKIE_NAME=grafana_session
    - GF_AUTH_OAUTH_AUTO_LOGIN=true
    - GF_AUTH_DISABLE_LOGIN_FORM=true
    - GF_AUTH_GENERIC_OAUTH_ENABLED=true
    - GF_AUTH_GENERIC_OAUTH_ALLOW_SIGN_UP=true
    - GF_AUTH_GENERIC_OAUTH_CLIENT_ID=istsos-istsos
    - GF_AUTH_GENERIC_OAUTH_SCOPES=openid profile email
    - GF_AUTH_GENERIC_OAUTH_ROLE_ATTRIBUTE_PATH=contains(
      realm_access.roles[*], 'admin') && 'Admin' || contains(
      realm_access.roles[*], 'editor') && 'Editor' || 'Viewer'

##################
# GUI
##################
gui:
  build:
    context: ./services/gui
    dockerfile: Dockerfile
  restart: always
  environment:
    - NEXT_PUBLIC_BASE_PATH=${GN_BASE_PATH}

##################
# ORCHESTRATOR
```

```
##################
orchestrator:
  image: registry.gitlab.com/ist-supsi/simile/istsos-suite/
     orchestrator:1.0.9
  restart: always
  depends_on:
    - vernemq
    - postgresql
  environment:
    - GN_BASE_PATH=${GN_BASE_PATH}
    - ORC_PROTOCOL=8888
    - MQTT_BROKER_HOST=${MQTT_BROKER_HOST}
    - MQTT_BROKER_PORT=${MQTT_BROKER_PORT}
    - MQTT_BROKER_USER=${MQTT_BROKER_USER}
    - MQTT_BROKER_TOPIC=${MQTT_BROKER_TOPIC}
    - MQTT_BROKER_PASSWORD=${MQTT_BROKER_PASSWORD}
    - MQTT_BROKER_CLIENT_ID=${MQTT_BROKER_CLIENT_ID}

##################
# NGINX
##################
nginx:
  image: registry.gitlab.com/ist-supsi/simile/istsos-suite/nginx
     :1.0.3
  restart: always
  ports:
    - ${GN_NGINX_PORT}:2000
    - ${GN_MQTT_PORT}:1883
  environment:
    - NGINX_BASE_PATH=${GN_BASE_PATH}
    - NGINX_PORT=${GN_NGINX_PORT}
```

3.1 The authentication

The access to data is regulated by a dedicated authentication service (Keycloak) protecting the various requests and functions. To this end, roles and groups have been used to better defined the priviligies of users and on a specific group of data. The authentication is designed to offer three levels of access

by profiling each user in a role: *admin*, *editor* and *viewer*. The *admin* has full access to the application; the *editor* has write and read privilegies to a limited goups of data; the *viewer* has just read only access to a predefined group of data. If roles defined the priviligies of users, groups related to a service of the software istSOS (see sec. 3.3). In such a way, if a user belong to predefined group it will have access to the correspongin service on istSOS, with the priviligies defined by roles. The adoption of roles and groups allows a better user profiling to determine if a can edit or view data and what functions it can use. The OAuth 2.0 authentication protocol has been used to standardise access flows for web, mobile or sensor applications. The relation between istSOS services and Keycloak groups is managed by the orchestrator 3.2.

3.2 The orchestrator

The orchestrator has been developed as a core component of the application since it has the role to manage all the requests and make all the services working togheter. It has been implemented from scratch using the Python programming language thanks to the adoption of the Tornado web framework (www.tornadoweb.org). The orchestrator works as a single access point to manage: the authentication phase; the web request sorting phase; and the possibility of extending the application with plugins. web; and the possibility of extending the application with plugins. The possibility of extending the orchestrator with plugins will make it easy to add functions for calculating indicators of the chemical and physical state of lakes. physical state of lakes. The orchestrator is externally exposed to provide the istSOS API to the various clients that need to interface with the application (e.g. sensors sending data or the web interface used by the end user).

3.3 istSOS

The istSOS software is the service adopted to archive and store all the data involved into this project. istSOS is based on five key objects, which derive from the Sensor Observation standard specifications (see sec. 3.3.1:

The SOS is based on five key objects as represented in Figure-2 :

Observations: they are the center of the standard and represent the values mesured at given time instants (e.g.: value: 0.2, time: 08-11-2012 12:12) and represented according to the O&M standard data model. Procedure: indicates

who provide the observations, this is generally the sensor but it may also be a generic process that leads to some observations (e.g.: procedure: TREVANO) and is represented as SensorML standard data model. Observed Properties: they represent the phenomena that are observed (e.g.: phenomenon: air-temperature) and is represented with a URI (uniform resource identifier) composed by colon separated text according to the om:observedProperty of the O&M standard. Feature of interest: it is the fature that relates to the observations, so for an in-place instrument is the sensor location, while for remote device it the target location (e.g.: location: Trevano, coordinates: 718345,99224,389, reference system: CH1903/LV03) represented according to the om:featureOfInterest element of the O&M standard. Offering: it is a collection of sensor used to conveniently group them up (e.g.: offering: weather-sensor-SUPSI) and is represented as sos:ObservationOffering element of the SOS standard.

Key object of the SOS standard

Thanks to the openness of the source code it has been possible to modify and further improve the software adding new features. istSOS has initially been modified to permit an integration with the OAuth 2.0 protocol. In fact, istSOS originally offered only a basic authorization using the Apache web server (apache.org). However, this resulted in several issues that led to a complicated configuration. For this reason, istSOS has been used just for data storing and management. The authentication pocess delegated to the orchestrator as described in the previous paragraph. While offering several features including the implementation of the SOS standard of the OGC for data sharing, istSOS had no support for profile-type data. This type of data, commonly used in Limnology, has the characteristic of having for a given time instant a series of measured depth-value rows (see Figure 3.2). In order to extend istSOS to support this type of data, either the source code and the database structure has been updated. istSOS works with different type of procedures. A procedure can be defined as a sensor or script that produces data (observations). Each observation is composed by a list of observed properties (parameters) and if needed unit of measures. The type of procedures originally supported are:

- insitu-fixed-point: it is used when a monitoring system is located into a fixed position (e.g. weather stations);
- insitu-mobile-point: it is used when a monitoring system is mobile (e.g. a

sensor on a bike)

- virtual: it is a procedure based on a Python script which can get data from other procedures and elaborate thse data with a formula on the fly

However, to support the profile type of data two more procedures two more types of procedure has been added: profile; and the virtual-profile. The first support the data type already described. The second allows to add dynamically a profile type data by merging multi insitu-fixed-point procedure. During the creation of the virtual-profile procedure, it is very important to set the Feature Of Interest (FOI), i.e. the position of the sensor, as accurately as possible, particularly with regard to the altitude relative over the sea surface. During development, it was decided, as assumption, to set the altitude is the hydrometric zero of the lake. This value is then used to substract the depth at which the sensor is placed and derive the altitude relative to the sea of that specific sensor placed at a given depth in the sensor chain.

3.3.1 The Sensor Observation Service (SOS) standard

The Sensor Observation Service (SOS) of the Open Geospatial Consortium (OGC) is a standard to share data and metadata. It is part of the Sensor Web Enablement (SWE) initiative. There are available two approved version 1 and 2 and the istSOS software enables both.

The aim of the OGC Sensor Web Enablement (SWE) working group, established to provide a definition of the system, is to provide a unique environment for a multitude of heterogeneous sensor networks by identifying technologies, language syntax and architecture to enable the development of the Sensor Web. In such a way, data can be shared, accessed and processed easily by researchers, scientists, decision makers, and users in general.

In the following list a series of OpenGIS® standards proposed by the SWE working group is present:

- SWE Common: XML model for sensors data
- SensorML (Sensor model Language): XML model for processes and components
- O&M (Observation and Measurements): XML model for the observation and measurements

- SOS (Sensor Observation Service): interface for data access and distribution
- SPS (Sensor Planning Service): interface for the sensor operation activation
- SAS (Sensor Alert Service): interface for dispatching alerts using the XMPP (extensible messaging and presence protocol)
- WNS (Web Notification Service): interface for the notification of information with different protocols

As part of the SWE, the Sensor Observation Service standard defines the interface to interact with sensor observations: from sensors exploration to measures retrieval and data management throughout transactional operations. It worth noting that this description refers to SOS in version 1; the version 2 of the standard was recently approved at the time of writing and introduces some minor changes.

The SOS standard is based on request-response pattern to exchange messages between the service and the consumer using the HTTP(s) protocol, like most of the OGC standards. Requests can use a GET or POST method. In the first case a key-value-pairs (KVP) is used to pass variables and values, in the second case the request must be formatted using an XML format whci must follows the specifications and requirements of the standard. The responses are always formated as XML compliant to the specificaitons. In Table 3.1, the list of the supported requests with a short description are shown. A software must implemt at least the mandatory request to be compliant to the standard.

SOS request	Profile	Mandatory	Short description
GetCapabilities	Core	Yes	Allow to describe the service providing information on administrator, offered capabilities, observed property and features, etc..
DescribeSensor	Core	Yes	It provides a potentially detailed description of a given registered component, system or process in SensorML format
GetObservation	Core	Yes	It provides observations based on the setting of filters that includes timing, processes, phenomena, feature of interest, and other parameters in O&M model

RegisterSensor	Transactional	No	It provides capability to automatically register a new sensor to the existing service
InsertObservation	Transactional	No	It provides capability to dynamically insert new observation(s) related to a registered sensor
GetFeatureOfInterest	Enhanced	No	It provides requested feature of interest in GML format
GetResult	Enhanced	No	It provides a light way to request observation without provide full request every time
GetObservationByID	Enhanced	No	It provides a quick access to observation by identification number
GetFeatureOfInterestTime	Enhanced	No	It provides the time interval when a given feature of interest has been observed

DescribeFeatureType	Enhanced	No	It provides the schema used to represent the features of interest
DescribeObservationType	Enhanced	No	It provides the schema used to represent the Observations
DescribeResultModel	Enhanced	No	It provides the schema used to represent the result object within the sml:observation

Table 3.1: List of SOS requests with their short description.

```
FI210310
10/03/2021 09:01:59
Ship Name
Depth Pressure Temperature Conductivity Oxygen mg/L Oxygen % pH Turbidity AutoScale Conductivity @20°C Trx-chl(a)  Trx-Neph Phycocyanin
0.00 0.116 8.1255 172.25 10.74 91.2 8.612 7.432 232.64 32.87 -1.04 8.93
0.50 0.490 8.1194 171.80 10.78 91.5 8.605 7.378 232.06 35.68 -1.05 9.80
1.00 0.981 8.1184 172.03 10.89 92.4 8.598 7.264 232.38 46.79 -1.05 11.84
1.50 1.471 8.0911 172.40 11.02 93.5 8.588 7.425 233.06 54.32 -1.04 13.40
2.00 1.961 8.0606 172.40 11.12 94.2 8.578 7.441 233.26 61.23 -1.05 14.59
2.50 2.452 8.0406 172.57 11.23 95.1 8.572 7.360 233.62 63.79 -1.05 14.49
3.00 2.942 8.0119 172.56 11.30 95.7 8.558 7.301 233.79 66.48 -1.05 13.84
3.50 3.432 8.0064 173.19 11.36 96.2 8.548 7.349 234.69 66.12 -1.04 13.76
4.00 3.923 7.9907 172.83 11.40 96.5 8.542 7.248 234.30 61.46 -1.05 13.43
4.50 4.413 7.9664 173.12 11.45 96.8 8.534 7.297 234.85 68.86 -1.05 14.79
5.00 4.903 7.9326 173.30 11.49 97.1 8.522 7.042 235.32 63.18 -1.05 14.65
5.50 5.394 7.9014 173.45 11.53 97.4 8.510 7.156 235.73 67.61 -1.05 14.96
6.00 5.884 7.8601 173.00 11.55 97.5 8.467 6.943 235.40 63.92 -1.05 14.02
6.50 6.374 7.8221 173.49 11.52 97.0 8.427 7.325 236.32 60.02 -1.05 13.38
7.00 6.865 7.7534 172.97 11.43 96.2 8.400 6.954 236.07 60.89 -1.05 14.08
7.50 7.355 7.6735 172.91 11.36 95.4 8.375 6.880 236.51 63.94 -1.05 12.04
8.00 7.845 7.6258 173.29 11.27 94.5 8.334 6.764 237.36 63.35 -1.05 14.70
8.50 8.336 7.5557 172.93 11.14 93.2 8.238 6.669 237.34 48.00 -1.05 10.32
9.00 8.826 7.4620 172.75 10.87 90.8 8.143 6.485 237.72 41.56 -1.05 10.00
9.50 9.316 7.3681 172.46 10.47 87.3 8.058 6.306 237.95 38.93 -1.05 8.03
10.00 9.807 7.2942 172.62 10.02 83.4 8.006 6.344 238.67 31.92 -1.05 7.14
10.50 10.297 7.2730 172.32 9.58 79.6 7.969 6.234 238.40 29.37 -1.05 7.60
11.00 10.787 7.2559 172.09 9.25 76.9 7.943 6.226 238.20 28.39 -1.05 6.76
11.50 11.278 7.1868 171.92 9.02 74.8 7.921 6.270 238.43 25.76 -1.05 6.93
12.00 11.768 7.0172 171.53 8.80 72.7 7.888 5.946 239.04 24.66 -1.05 5.40
12.50 12.258 6.8921 170.39 8.54 70.4 7.842 5.879 238.31 15.66 -1.05 4.97
13.00 12.749 6.8851 170.93 8.19 67.5 7.810 5.881 239.10 13.09 -1.05 4.30
13.50 13.239 6.8646 170.52 7.75 63.8 7.788 5.830 238.67 12.75 -1.05 4.68
14.00 13.729 6.8505 170.85 7.48 61.6 7.770 5.815 239.23 13.88 -1.05 4.37
14.50 14.220 6.7960 170.39 7.25 59.6 7.755 5.778 238.97 13.67 -1.05 4.63
15.00 14.710 6.7577 170.19 7.04 57.8 7.736 5.775 238.94 12.21 -1.05 4.33
15.50 15.200 6.7454 170.38 6.81 55.9 7.720 5.719 239.30 10.16 -1.05 3.83
16.00 15.691 6.7137 169.91 6.60 54.1 7.708 5.737 238.85 12.07 -1.05 3.90
16.50 16.181 6.6889 170.03 6.43 52.7 7.697 5.737 239.20 9.01 -1.05 3.81
```

Figure 3.2: Example of a text file format for data profile type.

3.4 Grafana

The activities concerning the software Grafana focused on three main tasks:

1. configuration of the software to integrate it inside the architecture;

2. implementation of two plugins;

3. definition of YAML files to automatically generate predefined and parametrized dashboards during the installation phase.

With regard to point 1, some configuration variables were parameterised to facilitate the integration of the service into the application. These variables are used for a general configuration of the service, but also to allow Grafana to communicate directly with the Keycloak software. Grafana uses the roles returned by Keycloak during the user authentication phase to offer accesses of Admin, Editor or Viewer access. Thanks to the support of the OAuth2.0 standard, it was possible to link the authentication service selected in the project and inhibit the internal one. The flexibility of the configuration permits to include Grafana in the project application as dashboard and chart generation service. In addition, Grafana offers the possibility to notify and alert on the basis of some custom rules and annotate on charts. However, such interesting features were still not explored. Step 2 initially involved the development of two plugins. The first allows to connect istSOS to Grafana as a data source, since by default Grafana has support for a limited number of data sources, among which does not include istSOS. The data source is the space from which Grafana takes the data to generate graphs. The second implemented plugin permits to create a panel with a heatmap inside for the visualisation of the limnological parameter profiles of the lake (Figure 3.3). Depending on the type of the visualization that the user most like, it can change using a simple line graph.

Each dashboard was parameterised on the basis of the observed properties, the time range and the selected procedures and their type (insitu-fixed-point, profile, virtual-profile, etc.). According to the combination of this last parameter, different dashboards are generated (Figures 7 and 8) which were preconfigured in six YAML-type configuration files.

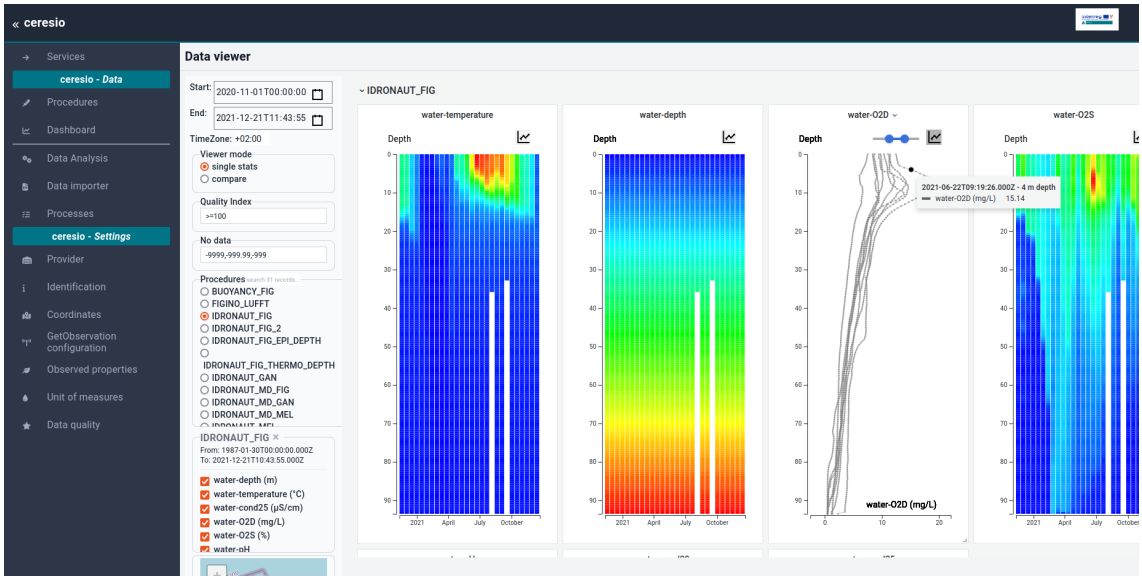

Figure 3.3: Example of the dashboard created using the developed profile plugin visualization.

3.5 Web interface

The last part of the software developed concerns the web Graphical User Interface (GUI), which is based on NextJS technologies (www.nextjs.org). It has been shaped on the user role adding more configuration pages and options when and admin or editor is logged. In Figures 3.5 and 3.4, it is possible to view part of the web interface created to facilitate the user in interacting with the features offered by the various services. Within this, a tool was implemented using WebAssembly technology for importing text files from the measurement campaigns carried out on the lakes. The monitoring campaigns produced very long time series and consequently a large amount of data. In order to facilitate the import, two modes are provided. The first is to be used when there is a need to import data and create the reference procedure within istSOS as it does not yet exist. In fact, in the istSOS, data always refer to a procedure that can be likened to a sensor sending data. In this case, only one reference data file can be selected. In addition, a series of metadata must be filled in to define the type of procedure, the data it provides and its geographical location (Figure 3.5). The tool automatically detects the number of columns to be imported. It only remains for the user to define for each column which is the correct property observed and its unit of measurement.

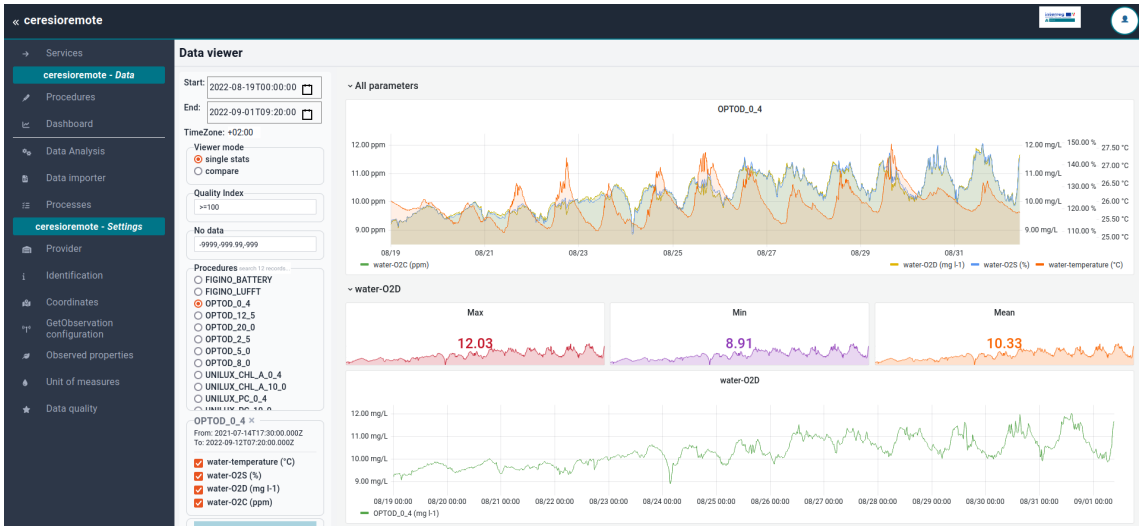

Figure 3.4: Dashboard for time series data.

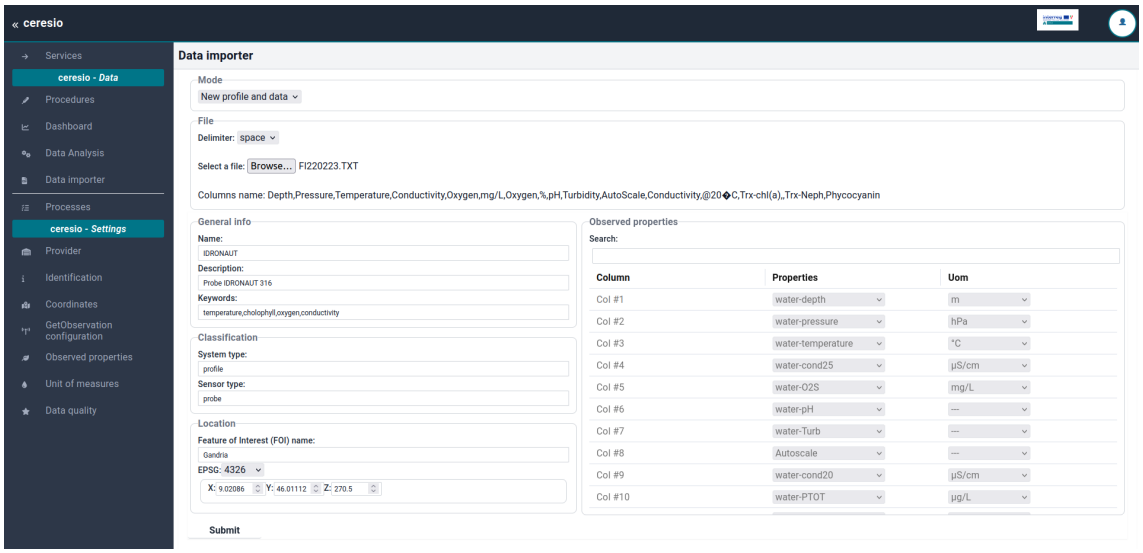

Figure 3.5: Example of the importer form to make easier for the user the importing of the profile data.

4 Open and cost effective digital ecosystem for lake water quality monitoring

Daniele Strigaro [1,2,*], Massimiliano Cannata [2], Fabio Lepori [2], Camilla Capelli [2], Andrea Lami [3], Dario Manca [3] and Silvio Seno [1,2]

[1] Department of Earth and Environmental Sciences (DSTA), University of Pavia, Via Ferrata 9, 27100 Pavia, Italy

[2] Institute 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;

[3] National Research Council of Italy, Water Research Institute (CNR-IRSA), Largo Tonolli 50, 28922 Verbania, Italy

Correspondence: daniele.strigaro01@universitadipavia.it

Strigaro, D., Cannata, M., Lepori, F., Capelli, C., Lami, A., Manca, D., Seno, S., 2022. Open and Cost-Effective Digital Ecosystem for Lake Water Quality Monitoring. Sensors 22, 6684. https://doi.org/10.3390/s22176684

4.1 Abstract

In some sectors of the water resources management, the digital revolution process is slowed by some blocking factors such as costs, lack of digital expertise, resistance to change, etc. In addition, in the era of Big Data, many are the sources of information available in this field, but they are often not fully integrated. The adoption of different proprietary solutions to sense, collect and manage data is one of the main problems that hampers the availability of a fully integrated system. In this context, the aim of the project is to verify if a fully open, cost-effective and replicable digital ecosystem for lake monitoring can fill this gap and help the digitalization process using cloud based technology and an Automatic High-Frequency Monitoring System (AHFM) built using open hardware and software components. Once developed, the system is tested and validated in a real case scenario by integrating the historical databases and by checking the performance of the AHFM system. The solution applied the edge computing paradigm in order to move some computational work from server to the edge and fully exploiting the potential offered by low power consuming devices.

monitoring; buoy; nb-iot; iot; open source; istSOS

4.2 Introduction

Water has always been a crucial but finite resource for life with strong connections to both food and energy sectors. According to the report of the Joint Research Center (JRC) of the European Commission (EC), in order to overcome the common view of resources as individual assets, the Water Energy Food and Ecosystem Nexus (WEFE Nexus) approach has been gaining growing interest. WEFE envisages several interactions between sectors, which were largely underappreciated [18]. Water bodies such as lakes are a perfect and concrete example of what the WEFE nexus approach aims at increasing the understanding on. In fact, lakes play a central role for stocking water and make it available for different usages, e.g., agriculture, industry, drinking water, etc. Therefore, a management and protection plan to preserve ecosystems that usually exist around these key resources is also strongly needed. Nevertheless, many touristic and economic activities are bounded to lakes, making preservation also a matter of economic interests, with different stakeholders involved. Unfortunately, climate change and

scarcity of water are affecting several places in the world. At the time of writing this paper, the North of Italy is facing one of the most severe drought of the last seventy years [19]. In particular, the Po watershed is dramatically under the mean level of accumulated precipitations. Such situation brings even more attention to the urgent need of finding solutions for a better water management and explore new frontiers to face the future challenges to which lakes and water resources are exposed. Furthermore, also the United Nations (UN), through their Sustainable Development Goals (SDGs), have set a dedicated goal to water quality: within Goal 6 "Ensure access to water and sanitation for all", the indicator 6.3.2 aims at monitoring the improvement of ambient water quality since water is essential for a large number of services and needs [20]. Although the depicted situation shows how lakes are strongly stressed by climate change effects and by human activities, over the last fifty years, developed countries have been implementing a number of actions in order to increase the quality of water. Focusing the attention on Lake Lugano as the area of study selected for this research, the mitigation actions imposed by administrative bodies since the 1970–1980s worked as counteraction of the galloping eutrophication of those years [21], a process which caused plants and algae to grow in excess due to the increased availability of nutrients [22]. According to the Organization for Economic Co-operation and Development (OECD) categorization [23], lakes' trophic status is usually classified in three ranges (oligotrophic, mesotrophic or eutrophic) with an increasing supply of nutrients and organic matter. Considering such classification, the reports of the CIPAIS commission (International Commission for the Protection of Italian-Swiss Water) highlight that, while over the last 25/30 years Lake Lugano recovered from an eutrophic to a mesotrophic status, it now appears at a standstill possibly due to unknown dynamics triggered by the changing environmental context [24]. Such scenario suggests the need for further actions in order to better understand the ecosystem dynamics and future trends and consequently react with calibrated new mitigation actions. There are two main lecture keys on which we would like to focus: the first concerns the lake monitoring while the second is related to the management of time series data. During the last few decades, environmental monitoring has been facing a deep revolution as a consequence of new technologies spreading across the market (e.g., the Internet of Things—IoT ; low cost sensors; and new data transmission methods [25, 26]). Such revolution is part of a wider process called digitalization [27] that is impacting all economic sectors as well as human

life. The new technologies, mentioned above, could potentially contribute to the densification of official monitoring systems, generally maintained by public administrations, which strongly need to improve the monitoring network for various scopes and with a different level of accuracy to support local needs. By way of example, the Swiss hydro-meteorological monitoring network includes a first level network, whose activity is focused on monitoring climatic variables with highly accurate sensors and tools and a second level network composed by small regional monitoring systems (one per canton), with more dense stations to guarantee effectiveness in supporting the local water management and in handling hazardous phenomena [28]. On the basis of such new innovations, it is common to find in literature several papers trying to solve research problems by building custom devices for monitoring environmental parameters [29–31]. These devices are mainly based on micro-controllers such as Arduino, Raspberry or Pycom solutions and make use of low cost sensors and modules to keep the costs of the system competitive [32, 33]. If, on one hand, such an approach demonstrates high potential and good results [34, 35], on the other hand, it could often lack in proved data quality replicability and in clear documentation to guide the user in building the system ("how to"). Furthermore, while high technical skills are required, in most cases, these solutions do not permit the deployment of a complete infrastructure chain that starts from data collection and ends with data management, comprising tools for data analysis, data quality control, notifications, etc. The availability of a tested and replicable open digital ecosystem composed of all these elements is crucial to increase the adoption of such solutions for the implementation of reliable real working monitoring systems. One of the reasons for the described situation could be related to the budget of research, which is usually limited considering their need to solve specific issues. In addition, focusing on water lake monitoring, periodic campaigns (i.e., weekly or monthly) are usually performed using on-board instruments (e.g., CTD probe measuring conductivity and water temperature or Secchi disk to measure transparency) and limnological vessels which are then further analysed in the laboratory [36], while a real-time automatic monitoring network is generally missing. In literature, there are examples of advanced systems that can improve the information gathered using traditional methods, but they are usually very expensive and strongly oriented to scientific experimental research [37, 38]. On Lake Geneva, a floating laboratory is installed, hosting more than 30 scientific projects with the aim of measuring and modelling key physical

and biogeochemical processes at high frequency [10]. Data management and integration is the second key issue that this paper would like to focus on. Unfortunately, such processes are facing a slow adaptation by the public and private sectors [39] in some of the water fields. For example, in the context of data management, a report on digital transformation in the U.S. water utilities sector by Dodge Data and Analytics highlighted that 90% of respondents to the survey confirmed that their data lie in disconnected systems such as spreadsheets, paper records, etc. Moreover, even digitized data often lack in uniformity of formats, interoperability and integrity. This situation is even more problematic in developing countries where the digitization and consequently the digitalization processes are still far from being in an advanced status [40, 41].

The possible reasons hampering the adoption of new technologies in lake monitoring and digitalization of the water sectors can be synthesized with the following points [42]:

- lack of digital expertise: the demand of hybrid professional profiles is growing in the market, especially in the environmental sector and more generally in data science [43, 44];
- resistance to change: human resources are often reluctant to change their mindset and way of working [15];
- lack of data interoperability between systems: this is a particularly delicate topic, as in the environmental system there are usually many proprietary solutions, each with its own standards for data collection, logging, transmission and sharing. This very wide topic concerns architecture, protocols and standards [45];
- lack of an easy-to-play solution: this aspect is strongly connected to the previous one, since the problem of having different kinds of proprietary solutions highlights the lack of a simple and smooth way of importing and integrating different data formats in a single solution;
- costs: it is a common belief that the benefit or value of having data in real time and with high frequency cannot outweigh the initial investment and maintenance costs [37].

In this context, this research would like to explore if an integrated and fully open software and hardware solution could address the limitations men-

tioned above and foster digitalization in the water sector. In fact, if successfully adopted, such a solution could help in understanding how water ecosystem management could benefit from the advantages of digitalization and in tackling the local effects of climate changes thanks to the support by high-frequency monitoring systems in discovering yet unknown dynamics. In particular, for testing and validating the system, we focused on the Primary Production (PP) estimation of the lake Lugano (the area of the study), a parameter which is generally calculated for the surveillance of the lake eutrophication process and that represents one of the most important indicators evaluating the lake metabolism. The PP can be defined as the process of assimilation of inorganic carbon into organic matter by autotroph organisms. In order to monitor the situation of Lake Lugano, twenty water samples are collected using little glass bottles at different depths, each at a distance of about 1 m from the previous one (from 0.5 to 20 m depth). Such samples are then analysed following the Nielsen and Schindler method [46, 47], whose results are processed using an algorithm that calculates the yearly amount of Carbon for squared meter (gC/m^2). Those values are finally used by administrative bodies to evaluate the status of the reservoir. Unfortunately, this methodology presents some critical aspects that can be synthesised with the following points:

- it is an expensive procedure since it requires costly external analyses and manpower/ working hours;
- highly qualified personnel is involved during the campaign, with very high cost per hour;
- results are based on a low temporal resolution time series (monthly samples);
- during the sampling process, a potentially dangerous radioactive component is manipulated by technicians.

For these reasons, an Automatic High Frequency Monitoring (AHFM) system was developed in order to apply, test and validate new technologies and recent methodology such as the PP estimation based on a number of parameters that can be measured in real time and high frequency using sensors (e.g., dissolved water oxygen, solar radiation, wind speed, water temperature) according to the Staher et al. method [48]. This approach calculates lake

metabolism rates on a daily time scale. Since it can potentially replace the traditional method but it is still unfortunately still underused, it must be highlighted that, in order to consider this new approach as a valid alternative, a wider in-depth investigation should be conducted in the future, in addition to the analysis of technological aspects on which this article is focused.

4.3 Materials and Methods

The existing limitations that block the digitalization process may be overcome by the development of a fully monitoring system (from data collection to data management), which follows some requirements, such as:

- use as much as possible an open source approach either for the software or for the hardware part;
- use open standards for data integration and sharability;
- maximise the possibility of easily extending and customising the software;
- use cloud technologies to deploy server side components.

Once the different parts of the system were developed, they were tested and validated in a real case scenario. In order to make things initially simpler, we focused on a specific problem that affects the selected study area and then verify if the proposed solution could effectively improve the ongoing situation. Since an all-in-one commercial solution does not exist according to the authors' knowledge, a fully open and replicable solution is proposed in this article.

Such activities were conducted in the framework of the SIMILE (Informative System for the Integrated Monitoring of Insubric Lakes and their Ecosystems) INTERREG Italy-Switzerland project, which aims at integrating different techniques (in-situ monitoring, remote sensing and citizen science) to monitor trans-boundary and sub-alpines lakes such as Maggiore, Como and Lugano.

4.3.1 Area of Study

The study area individuated to test and validate the monitoring system is Lake Lugano. It is composed of three sub-basins (see Figure 4.1). The North and South basins are divided by the Melide bridge, which was built on a

sublacual moraine. The third small basin is located in proximity to the Tresa river. From a morphometric point of view, the Lugano South basin has an area of 21.4 km^2 and a mean and maximum depth, respectively, of 85 m and 89 m (Table 4.1). This study is focused on the South basin because of the main interest due to its precarious water oxygenation, which, during the winter season, can reach the anoxide status in the deepest part of the lake. This is one of the negative effects of the euthrophication process started during the 50 s with a peak in the 80 s. Thanks to the European Directive (2000/60/CE and 2006/7/CE) and to the Switzerland law (Federal Act on the Protection of Waters and Water Protection Ordinance), the situation is improving, but it is still far from the fixed goal of 150 gC/m^{2}y (mean average of 300 gC/m^{2}y of last 5 years) [21, 49]. Hence, Lake Lugano has been constantly monitored since the 1970s through manual campaigns which occur one or two times per month depending on the type of monitoring (chemical, physical or biological).

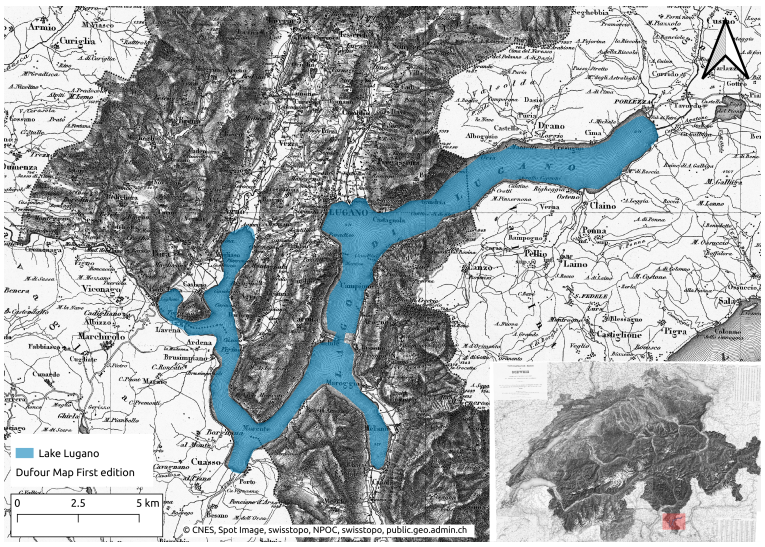

Figure 4.1: Map of Lake Lugano, southern part of Switzerland.

Table 4.1: Morphometric information about Lake Lugano.

	Total	**North Basin**	**South Basin**
Area of the basin (km^2)	565.6	269.7	295.9
Area (km^2)	48.9	27.5	21.4
Volume (km^3)	6.5	4.7	1.8
Mean depth (m)	134	171	85
Max depth (m)	288	288	89

4.3.2 The System Architecture

The AHFM system is part of a wider architecture that covers all the components engaged in the monitoring system. To this end, the first step concerned the research and collection of all the historical time series of Lake Lugano which are usually archived in different data formats (e.g., CSV or TXT file, Microsoft Access databases or Microsoft Excel files, etc.). One of the main objective was to integrate data sources into a single sink to obtain a unique entry point and simplify a situation dominated by a lack of integrity, interoperability, uniformity as well as being prone to the errors process because many pieces of data are replicated in different data formats. In Figure 4.2, the schema of the data flow is presented; different data sources were integrated into an organized and accessible system that can make them available to stakeholders using standard services. In addition to the reorganization of the existing datasets, which is continuously updated by the data collected during the monitoring campaigns, new observations were generated in quasi real-time by the AHFM system. Besides the development of components that can integrate and import already existing data, the system relies on the Sensor Observation Service (SOS) of the Open Geospatial Consortium (OGC) for data sharing and interoperability. It is a web service that allows for query environmental data encoded using the Observations and Measurements (O & M) model coming from sensors or more generally by a so-called "procedure". The metadata of each *procedure* are formatted following the Sensor Model Language (SensorML). In addition to the SOS OGC standard, data and metadata are available through a REST API interface in order to better link web applications. In summary, the approach proposed permits retrieving observations using a single point access and further elaborate on them to create new time series.

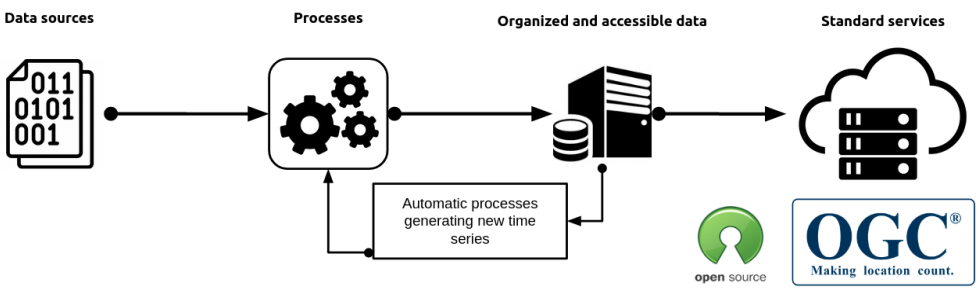

Figure 4.2: The data flow architecture.

During the development of the proposed architecture, of which components are described in detail in Section 4.3.4, three main approaches are used: the

Service-Oriented Architecture (SOA); Edge Computing (EC) and WebAssembly. The first is a development pattern where a service is a component application that communicates to the other through a communication protocol over a network [50, 51]. In SOA, services play a key role and become the basic computing unit to support development and composition of larger, more complex services, which in turn can be used to create flexible, ad-hoc and dynamic applications [52]. The second (EC) is a computing paradigm that brings data elaboration and storage closer to data sources such as in some IoT devices [53, 54]. Finally, WebAssembly provides a way to run low level code written in multiple languages (Rust, C++, C#, etc.) on the web at near native speed using a modern web browser.

4.3.3 The Automatic High Frequency Monitoring (AHFM)

The already cited AHFM is developed and then mounted on a platform of 2.5 × 2.7 m in the middle of Lake Lugano Southern Basin (45.95866° N, 8.89371° E, EPSG 4326). In Figure 4.3, the design of the platform is presented. It is composed by a metallic structure sustained by four floats and anchored to the bottom of the lake with four chains connected to concrete blocks. At the middle of the floor, there is a hole where a chain of sensors can be lowered. The purpose of this structure is to offer space and support to easily install and maintain the hardware components used to monitor the water environmental parameters.

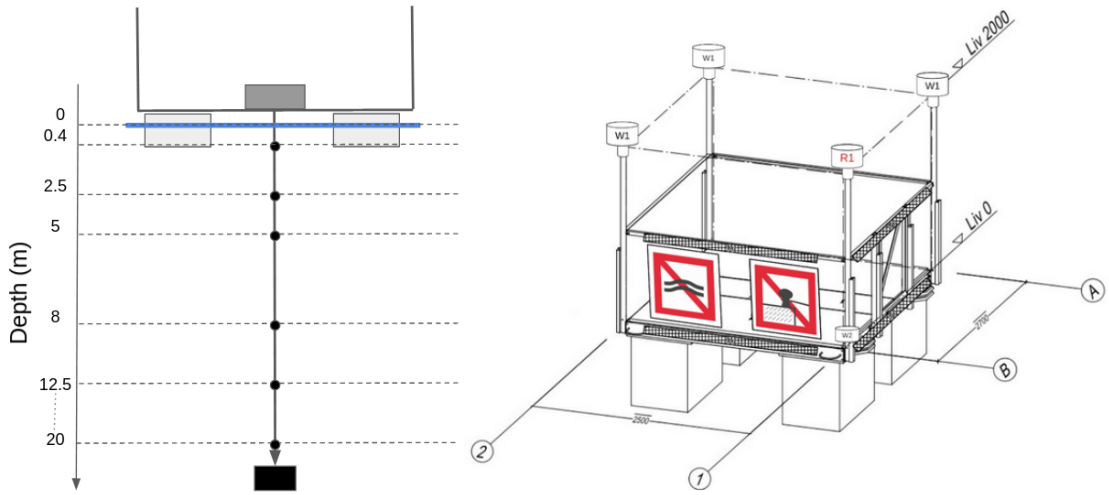

Figure 4.3: On the left, the design of the position of the chain sensors; on the right, the project of the platform designed and realized by Officine Ghidoni SA (Via al Pizzante 9, 6595 Riazzino, Switzerland).

The AHFM system, mounted on a customized structure fixed on the center of the platform, is composed of three main units (Figure 4.4): the power unit; the sensor unit; and the main unit.

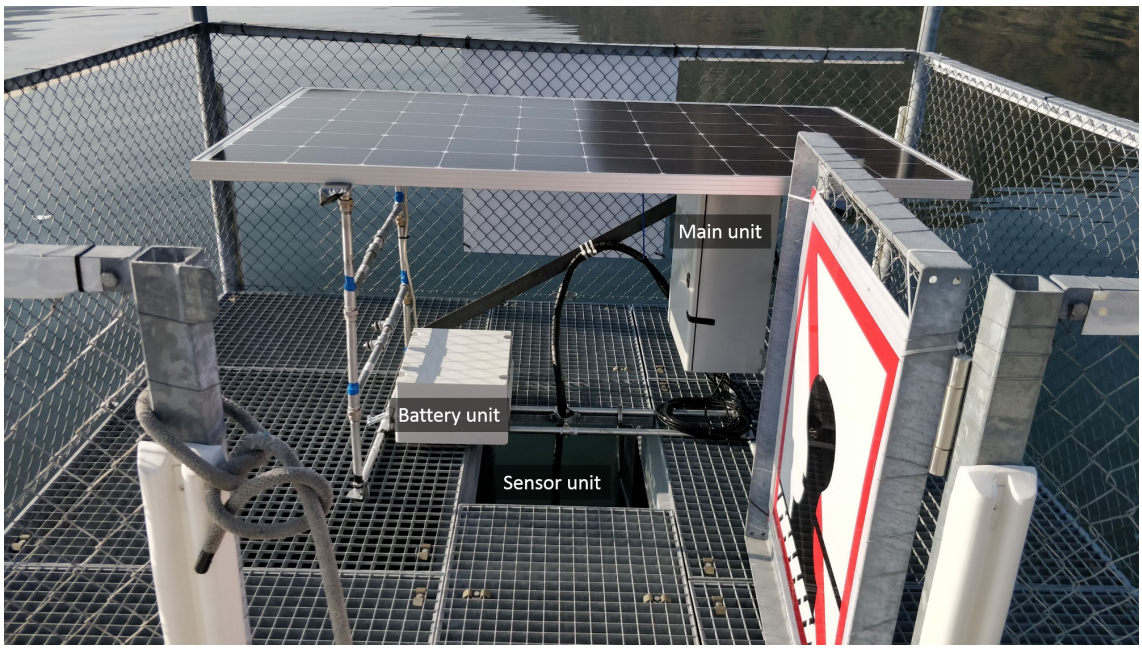

Figure 4.4: The AHFM system architecture.

The Power Unit

The power unit consists of a 12 V per 110 Ah battery (1320 Wh), a 20 A solar regulator and a photovoltaic panel with an estimated peak power of around 300 W. The solar regulator provides 12 V to the system voltage regulator, which then steps down the current to 5 V to power up the core unit, and to 12 V to power the sensor unit (Figure 4.5).

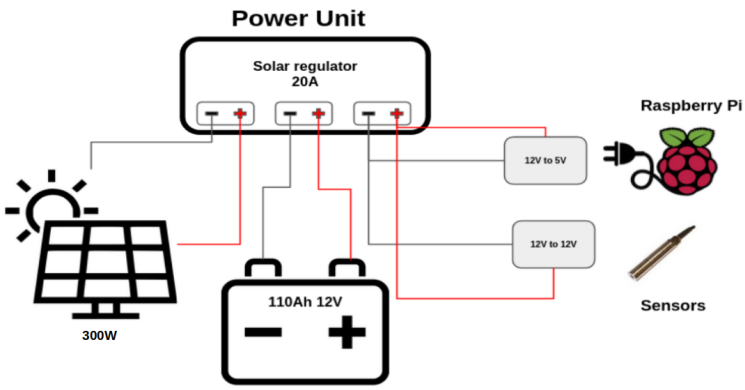

Figure 4.5: The power unit.

The Sensor Unit

The sensor unit relies on six Optical Sensors based on luminescent technology (OPTOD) by Ponsel, which measures the dissolved oxygen in terms of concentration (ppm and mg/L) and saturation (%) with an accuracy of, respectively, +/−0.1 ppm or mg/L and +/−1%. In addition, it provides the water temperature, which is used to estimate the saturation. The OPTOD sensor can be connected to any type of datalogger through the Modbus RS-485 connection since it has an open protocol for command and data transmission, and it requires 5–12 V current. The sensor is factory calibrated. The re-calibration process can also be performed, putting the sensor at open air checking that it measures 100% of oxygen saturation. Otherwise, when this process does not correct the measurements, a proper calibration should be performed in the laboratory or delivering the sensor to the main factory. The six sensors were assembled in order to form a chain in which each sensor is positioned at different depths (0.4, 2.5, 5.0, 8.0, 12.5 and 20.0 m). The depths correspond to the position where the traditional water samples are collected during the monitoring campaign performed by limnologists (Figure 4.6). They usually take one sample every 0.5 m, so the most significant point is chosen with a higher density in the first five meters since the productivity is concentrated in the photic area [55].

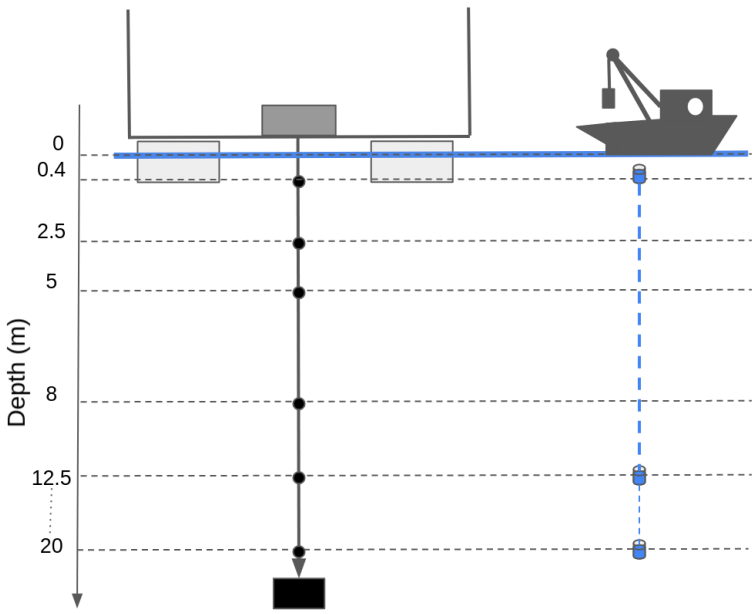

Figure 4.6: Comparison between AHFM and manual sampling depths.

The Main Unit

The main unit (Figure 4.7) includes solar and voltage regulators, a section with terminals for sensors and power inputs and finally a core box. The power inputs concern the connection with the power unit based on a miniature computing platform based on the Raspberry Pi 4 model. Raspberry 4 is chosen instead of other MicroController Units (MCU), like Arduino or PyCom devices because of its larger computational power, its flexibility and acceptable power consumption which therefore permit to perform some inbuilt analysis directly at the edge of data source as for the EC computational paradigm. The workflow implemented in the control unit is developed using the Python programming language, and it is available at the github repository under the GNU GPL v3 License (`https://gitlab.com/danistrigaro/station_configurator/`, accessed on 1st August 2022). The algorithm obtains data from the sensor unit every minute through a MODBUS RS-485 interface, performs some quality checks on the data and aggregates them every 10 min. After the aggregation, the data are sent to the server infrastructure. The control unit supports the SOS standard for data sharing thanks to the istSOS software, a lightweight data management system which is successfully used for various applications based on the hydro-meteo weather stations data of the Canton Ticino [56].

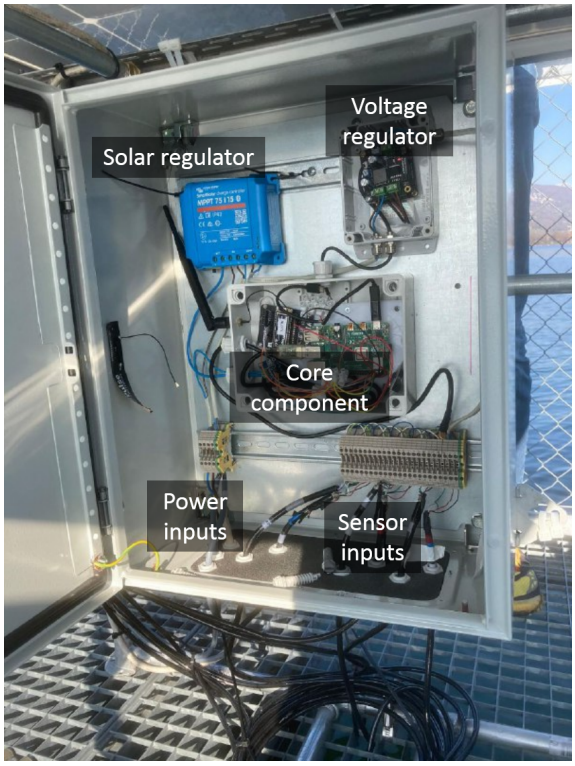

Figure 4.7: The main unit components.

The communication unit is based on a IoT shield which sends data through NB-IoT or LTE-CATM1. The data transmission takes advantage of the MQTT protocol which is very light and widely used in IoT systems. The NB-IoT or LTE-CATM1 are selected instead of the LoRa communication because they have more bandwidth for data transmission while maintaining very good performance in power consumption.

The Firmware Structure

The control unit, based on the Raspberry pi 4, hosts a Raspian Pi OS Lite distribution, which was customized in order to minimize the current consumption disabling the unused features. Therefore, istSOS, an implementation of the SOS standard, was installed to manage on the edge all the data collected using the SOS standard. The availability of a powerful device that manages data can help in performing some tasks that otherwise are usually performed on server side. In fact, the data flow (see Figure 4.8) concerns the collection of raw data from the chain of sensor every minute; then, these data are instantly checked to avoid null value and outliers by controlling that the value is in the range specified in the sensor datasheet. Data are inserted in the local istSOS

which will be assigned a flag to indicate the quality of the value. Every 10 min, a script is run to obtain the last 10 min data, to check the consistency of data in terms of variability between them, and finally to aggregate only the data which passed the tests. If the data flagged as correct are less than 60%, then the aggregate value is flagged with a quality index of 200, otherwise of 201. According to the configured sending time, the aggregated data are transmitted to the server side using NB-IoT through the MQTT protocol. If data are correctly sent to the broker, the flag of each observation is updated on local istSOS in order to be able to filter between sent and not sent data.

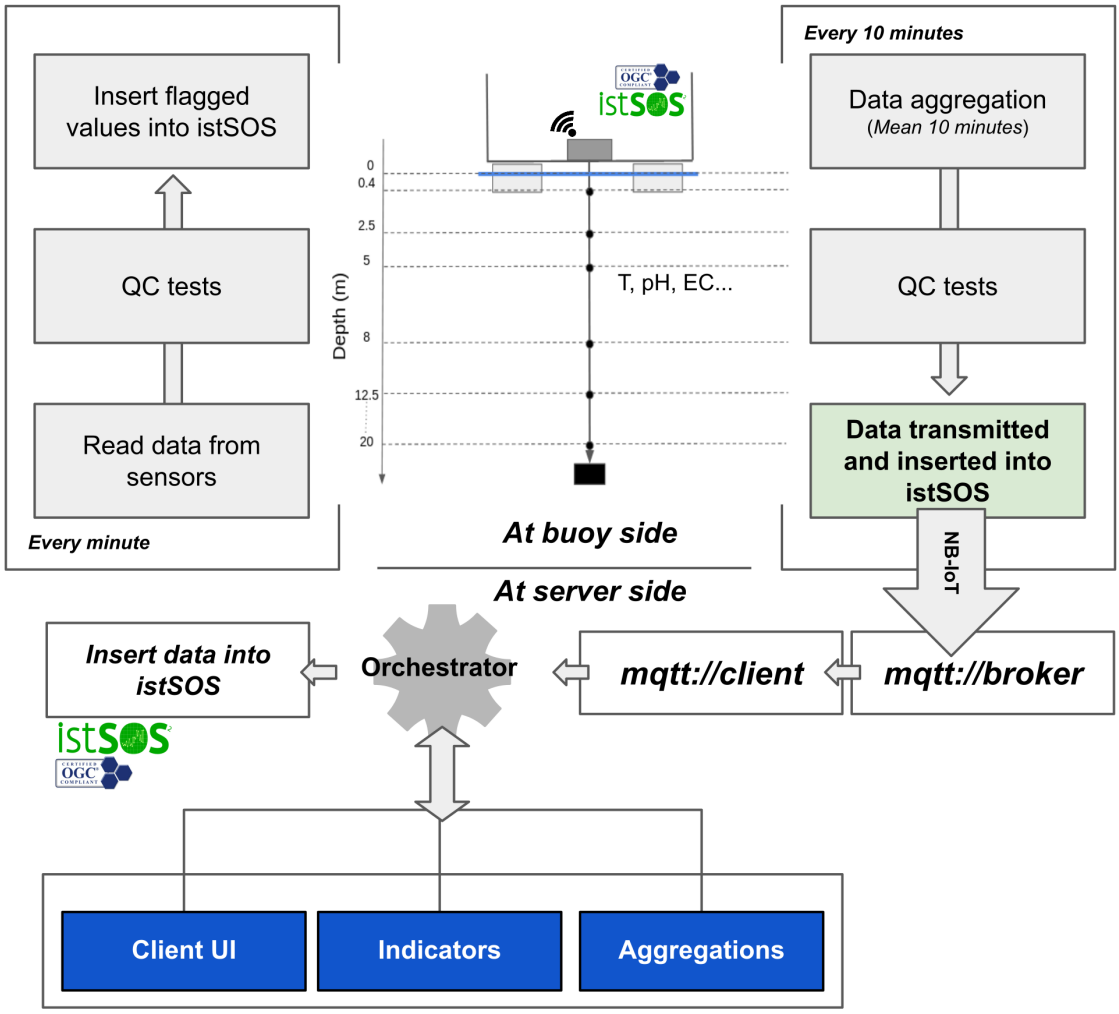

Figure 4.8: Data flow from buoy to server.

The data flow was implemented using custom Python scripts that are automatically generated running the initial configurator script that reads the YAML settings file and performs the following steps:

- check the availability of the configured sensors;
- register sensors in local and remote istSOS;
- create the read and write script that collects raw data from sensors based on the frequency set in the configuration file;
- create the aggregation script that collects raw data from istSOS and aggregate them using the frequency set in the configuration file;
- generate the Python script which obtains the aggregated data from local istSOS, activates the NB-IoT module and transmits each single observation to the remote istSOS.

The execution of the scripts is possible because, during the configuration process, scripts are added to the crontab schedule according to the specified frequency.

Once data are transmitted to the server, they are available for further validation processes, analysis, reports creation, etc.

4.3.4 The Data Management

The software architecture (Figure 4.9) is designed to store, share and elaborate the data coming from the AHFM system and from the historical time series collected by the traditional campaign since the 1970s. It is composed by seven services (Figure 4.5): a web interface; an orchestrator (Python); an authentication service (Keycloak); a database (`https://www.postgresql.org/` (PostgreSQL), accessed on 1st August 2022); a data management system (`http://istsos.org/` (istSOS), accessed on 1st August 2022); a graph and dashboard generator (`https://grafana.com/` (Grafana), accessed on 1st August 2022); an MQTT broker for AHFM system data (`www.vernemq.com` (VerneMQ), accessed on 1st August 2022); a processor for indicators' elaboration (Python); a user interface (`https://nextjs.org` (Next,.Js), accessed on 1st August 2022). The implementation of each single service was done using Docker technology. In this way, it was possible to create a Compose file to automatically generate the services. Compose is a tool to define and run multi-container Docker applications.

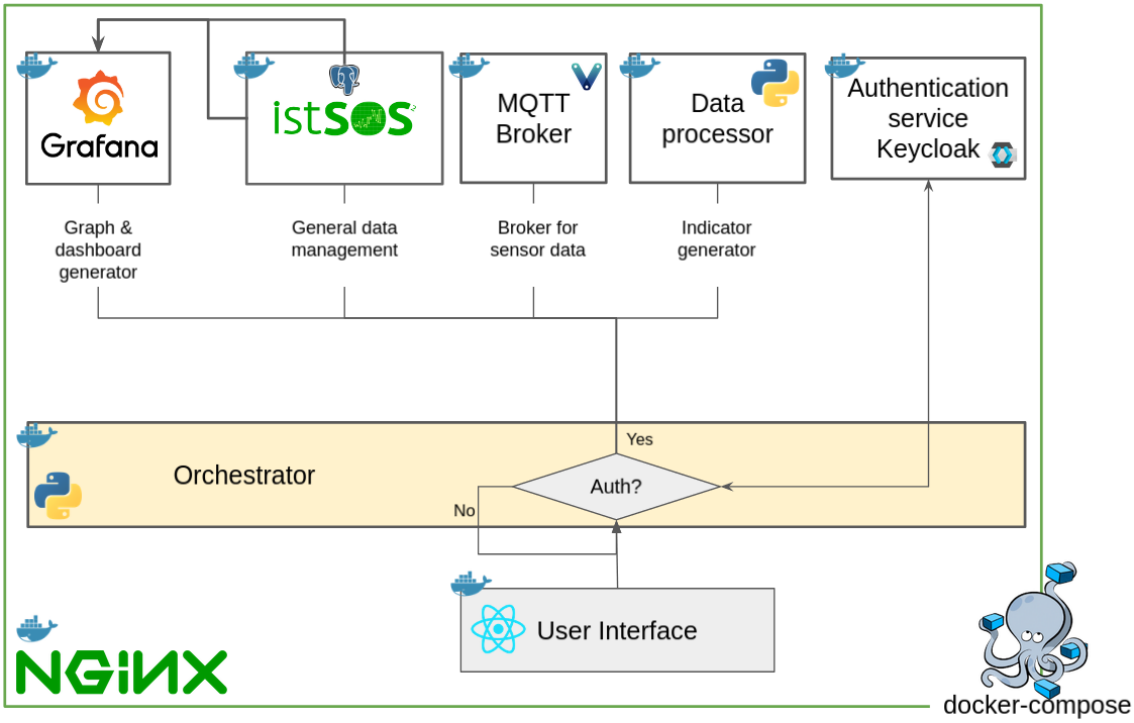

Figure 4.9: The system architecture.

Data Sources and Standards

In order to facilitate the decision makers through a synthetic interface based on dashboards and graphs, all information needs to be digitally accessible from a unique entry point. Therefore, all the data sources are initially grouped, then processed and digitized. Basically, four main sources exist:

1. data from manual sampling campaign based on sensor probes;
2. data from laboratory analysis performed on water samples;
3. data from satellites;
4. data coming from a real-time monitoring system.

To foster the data integration and exploitation, as well as data sharing, international standards are used as a powerful tool which provides a predefined language to increase the interoperability between data. In the environmental field, the Open Geospatial Consortium (OGC), founded in 1994, is a community which collaboratively develops open interface standards and associated encoding standards, and also best practices to exchange geospatial information.

To enable data integration, the Sensor Observation Service (SOS) [57] open standard is selected to handle all the data available.

Support for Profile Data Type

The profile data are a particular time series where, for each datetime, an array of data composed by depth and value is available. Instead of a bi-dimensional data representation (usually time and value), in profile data, there are three dimensions: the time, depth and value (Figure 4.10). For this reason, istSOS and Grafana, two of the selected applications which shape the server architecture, do not natively support the profile data type. In fact, istSOS, which was mainly used for time series data, it originally does not support the data profile format. To fill this gap, in this work, the software has been further improved by implementing a feature to support data profile type, which is then incorporated into the core of the software. The Grafana application has the function to generate dashboards and plots. In addition to istSOS, it is an open source software which allows for query different data sources and visualizes the data requested in many ways. It is easily extensible thanks to the possibility to develop plugins in order to add new data sources and new panels for data representation. Hence, a new data source plugin has been developed in order to enable the communication to the istSOS instance and a new panel to visualize the data profiles.

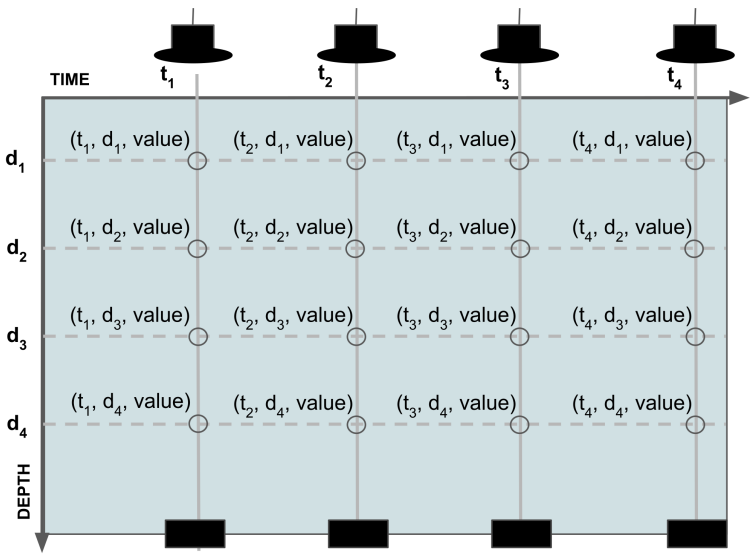

Figure 4.10: Profile data type.

4.3.5 Data Quality Control Process

The data Quality Control (QC) in a monitoring system aims at detecting erroneous or suspicious values which could be a sentinel of possible system malfunctions. Once detected, they should be automatically or manually corrected or deleted [58]. In Table 4.2, a list of Quality Index available on istSOS is used to flag values. Higher values correspond to a higher level of quality. The second digit of the QIs is used on the edge to understand if data are correctly transmitted to the MQTT broker (digit 1) or not (digit 0).

The QC tests implemented follow the WMO guidelines [59]. Tests are applied in real-time at the edge and in post-processing at the server side.

Table 4.2: Quality Indexes list.

QI	Description
0	Raw value erroneous
100	Raw value
101	The value is consistent with the sensors's range of measurement
102	The value is consistent with the previous one
103	The value is timely consistent
200	The value has been aggregated with less than 60% of valid values
201	The value has been correctly aggregated

QC Process at the Edge

The AHFM system collects new measures that undergoes two tests before being stored in istSOS. The first test consists of verifying that the measure is a valid number (is not a "Not a Number", NaN). The second applies a range test, to identify gross error and outliers (Table 4.3). The range test is based upon specifications and limits found in the sensor's datasheet.

Table 4.3: Range test limits used at a data sensing layer.

Observed Property	Range
Water temperature	$-10 < x < 60$ (°C)
Oxygen saturation	$0 < x \leq 200$ (%)
Oxygen concentration	$0 < x \leq 20$ (mg/L or ppm)

Once it is time to aggregate data, two other tests are performed. The step test (Equation (4.1)) controls the variation compared with the previous and the next observed values not being above a prefixed limit [59]:

$$|V_i - V_{i-1}| + |V_i - V_{i+1}| \leq \sigma_v \tag{4.1}$$

The second check performed (persistence test) controls the fact that between values there is a minimum variability. This test is performed to detect possible sensor malfunctioning that often results in a returned constant value, which is not corresponding to the current observed phenomena. While other methods are proposed in literature by [60–62], we implemented those proposed by [63] (Equation (4.2)):

$$AVG = n\frac{\sum_{i=1}^{n} V_i}{n} V_i \neq AVG \tag{4.2}$$

The average is then performed only on the values that passed such tests. If less than 60% of the values retrieved for the aggregation do not pass the tests, the value aggregated will be flagged with a lower QI flag.

4.4 Results

Hereinafter, the presented results refer to an entire year of data collected from 1 January to 31 December. The system was implemented and started the data collection in November 2020, but we considered the first two months as a testing and stabilization phase where small improvements and fixes were applied.

4.4.1 Data Completeness

The evaluation of data completeness is an important estimation of the reliability of a newly developed system. Basically, taking into account the whole period of work (from the 1 January to 31 December 2021), the amount of the data expected, calculated from the sampling frequency, is compared to the amount of data correctly archived. Since raw data are collected every minute for a total of 1440 observations per day, the daily archived observations were counted and then the daily percentage of completeness was calculated. These values are again grouped into five categories: bad if the percentage is less than 70%; poor if the percentage is less than 80% but more than 70%; medium

where the value is between 80% and 90%; good between 90% and 99%; and finally high, where the percentage is more than 99%. The graph in Figure 4.11 shows some results for all the sensors installed even if each sensor has its own connection and data are read from them one piece after another. Only 0.3% of the daily data completeness is categorized as bad, 2.2% in the medium class and finally 97.3% of the values has a daily percentage of completeness greater than 99%.

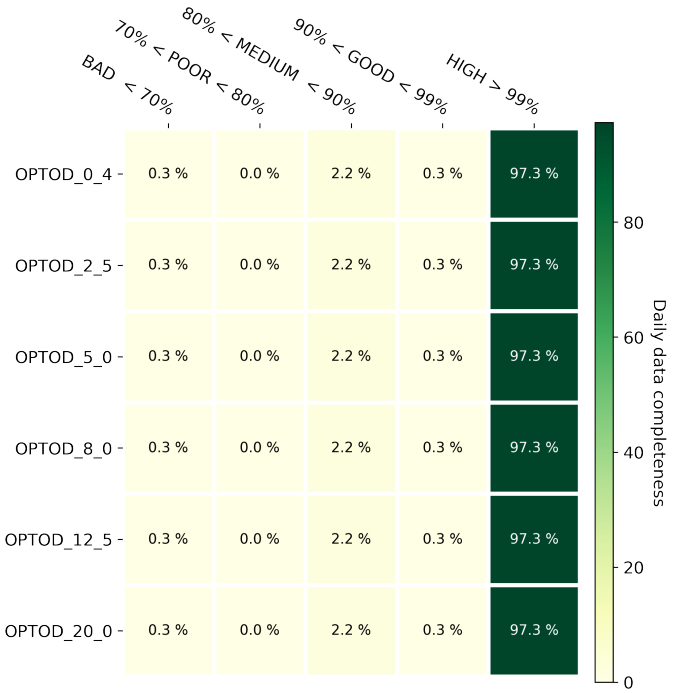

Figure 4.11: Daily data completeness heatmap based on the classification of the daily percent value of completeness in five FL classes: bad, poor, good, and high.

4.4.2 Data Quality

In this section, the results of the quality control process described in Section 4.3.5 are shown. In Figure 4.12, a heatmap per each observed property showing the result of such an analysis is proposed. In those graphs, data were aggregated per day by making a percent of data that passed a specific test. Then, each day is categorized according to five classes which can describe the percent of days that have bad, poor, medium, good, and high quality. Water oxygen concentrations expressed in ppm and mg/L have exactly the same results since they are basically the same measure but expressed with a different unit of measure. Almost 100% of the values passed the range and step test. Just

0.23% of the day were categorized with a good level of quality, which means that the remaining 99.7% have a high level of quality. A similar behavior is registered for the water oxygen saturation. In addition, 99.82% and 99.68% of the days have respectively more than 99% of data that passed the range test and the step test as well the persistence test. The last graph in Figure 4.12 is about the water temperature. The quality control process shows that 98.4% of day was categorized with a high level of quality since more than the 99% of the values of those days passed three tests.

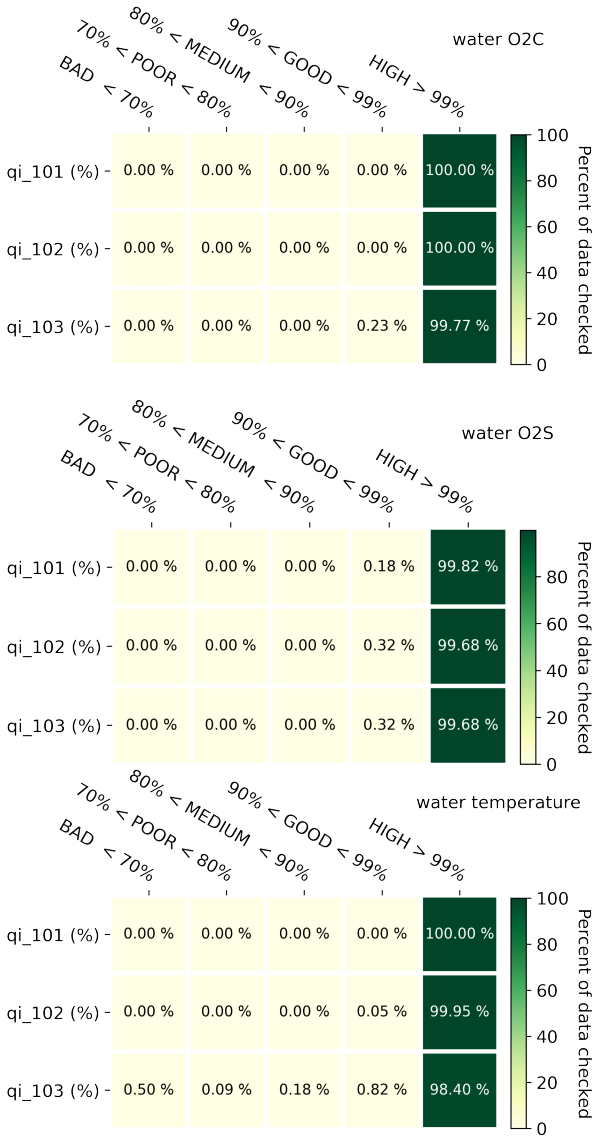

Figure 4.12: Daily aggregated data quality tests categorized into five FL discrete classes.

4.4.3 Maintenance and Calibration

The system deployment also coincided with the start of the maintenance activity, which is very important in particular when sensors are submerged in very productive lakes such as lake Lugano. During the winter, the maintenance does not highlight specific issues. Starting from the spring and especially in summer, sensors were covered by an algae film which could affect the measures of the dissolved oxygen. If on one hand the maintenance is focused on cleaning the sensors, during July, a communication issue blocked the data transmission, but not the data collection, which was regularly archived both on istSOS and on a text file on a USB stick. The flash drive was introduced after a few months for securing eventual SD card failures in order to have a data backup available.

In addition, during the maintenance activity, sensor calibration is verified by controlling the fact that measures reach a value near 100% of oxygen saturation in the open air. To this end, the chain of sensors is pulled out of the water waiting for them to be dry in order to prevent erroneous measures that could be affected by the presence of the water. If values deviate for more than ±5% from 100%, the sensor is re-calibrated applying a coefficient of correction which could be positive or negative. This process is performed by connecting the sensor with a laptop and by using the Calsens software. The coefficient is used in post-processing to correct the previous data collected by applying a linear regression between the last calibration and the new one. In Figure 4.13, an example of the data correction performed on sensors positioned at 2.5 and 12.5 m depths is shown by overlapping the raw data with the corrected one. In these cases, graphs highlight the fact that the sensors have generally a drift which causes an underestimation of the values. During the period considered, sensors generally suffer from an overestimation of the values. Considering that the platform was visited eight times to perform maintenance, the calibration coefficient was negative only two times, showing an overestimation of the values. From the experience gained during the working period of the AHFM system, in a highly productive lake such as Lake Lugano, which has a mesotrophic status, a maintenance program consisting of a visit every two weeks is enough to guarantee a good quality of measurements and a good status of the chain of sensors. Each visit lasted an average of three hours, but since this value comprises developing and testing actions, it is possible to consider one hour an half for a standard maintenance activity (cleaning sensors and solar panel, calibrating sensors if needed and downloading raw

data).

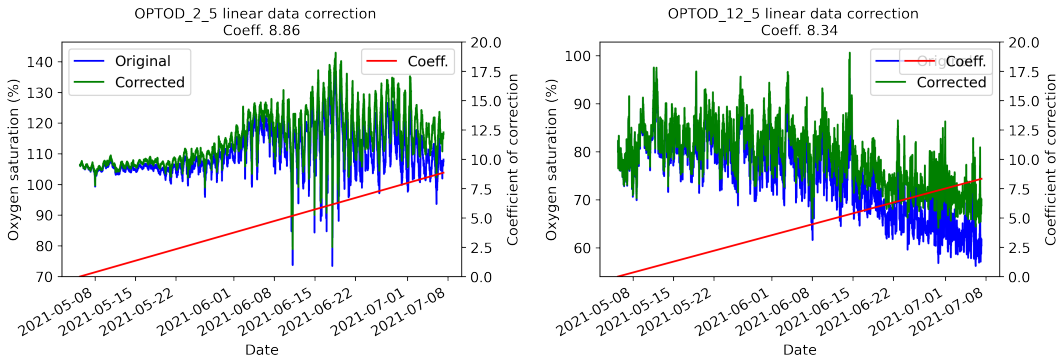

Figure 4.13: Data correction of time-series collected by the OPTOD sensors by Ponsel one positioned at 2.5 and another at 12.5 m depth.

4.4.4 The Powering System Data

Data collected to analyse the power management are retrieved from the Victron Energy solar regulator. Such data do not cover the entire year of analysis because data were not always downloaded from the device during the maintenance activity. The power system is based on three type of phases: bulk; absorption; and float. The bulk is the phase where the charger sends current and amperes to the battery so that the battery voltage rises. The bulk stops and absorption starts when the battery is full and reaches the specified absorption voltage. If the battery is full and the photovoltaic panel is in a situation to transform solar energy in current, the float phase begins with the simple job of keeping the battery full. In Figure 4.14, the trend of the three described phases is highlighted. The time in float has a peak during the summer of 800 min and decreases in cold seasons to a minimum of about 100. The time in absorption and in bulk are quite constant, respectively, around 160 and 40 min.

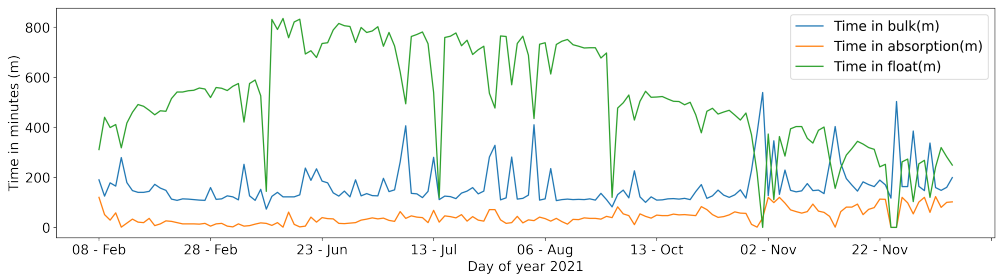

Figure 4.14: Time in bulk, absorption and in float.

In Figure 4.15, the comparison between the yield and the consumption energy expressed in watts per hour is shown. Even though the initial consumption is estimated to 110 Wh per day, the installation of a new Operating System, which is more optimized in power management, and the change in the communication hardware module, brings a drastic decreasing of the power consumption.

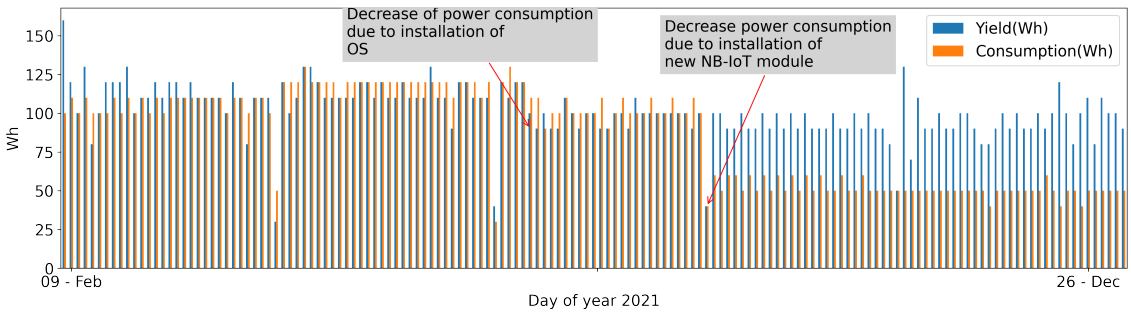

Figure 4.15: Energy yielded vs. energy consumed by the AHFM system.

A final consideration on the power management system concerns the trend of the voltage battery which rarely reaches values under 12.3 Volts (Figure 4.16).

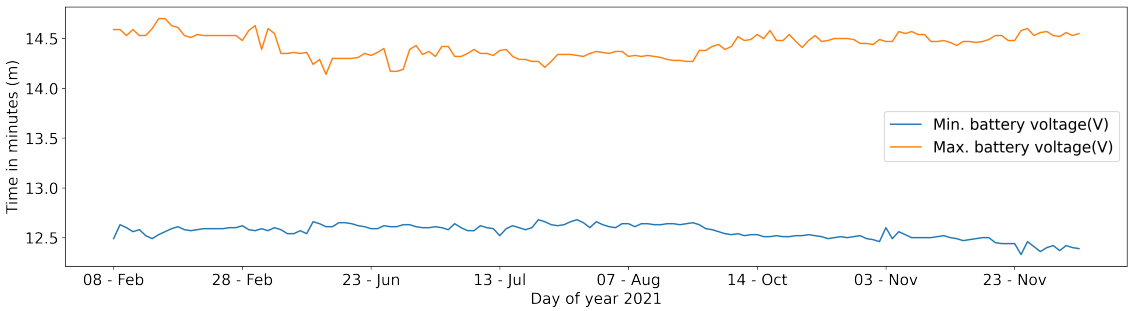

Figure 4.16: Minimum and maximum battery voltages.

4.4.5 Comparison between AHFM System and IDRONAUT's Probes

The developed system permits on one hand to collect data in real-time and on the other to import all the historical time series coming from the traditional campaigns in a unique entry point. To this end, the data management software, at the server side, helps with homogenizing data and metadata as well as in enabling the comparison and validation of the AHFM system dissolved oxygen data. In Figure 4.17, a general overview of the AHFM dissolved oxygen time-series is compared with the measurements taken with two types of probes. In fact, during the traditional campaigns performed in 2021, water profiles were taken using an IDRONAUT 316 Plus and an IDRONAUT 304 Plus probe. The first is a multi-parametric piece of equipment that can measure water temperature, water oxygen, Chlorophyll-a and Phycocyanin concentrations. The second is an older instrument which measures water temperature and dissolved oxygen. These instruments are used to make water profiles generally two times per month. The output is a time-series with a low temporal resolution but a high spatial resolution since data are recorded every 0.5 m by manually dropping down the probes. In Figure 4.17, a general overview of the comparison between IDRONAUT's probes and the AHFM system time-series is shown. Focusing on the first four graphs, which respectively concern data collected at 0.4, 2.5, 5 and 8 m depths, a similar trend can be highlighted. From January to February, a negative gradient of the water oxygen saturation is followed by an increment which was registered by all the sensors. This

event is followed by another, a more marked change, that began during the end of February. Two other peaks respectively near the end of March and the end of April follow. Such events can be well recognized at 0.4, 2.5, 5 and 8 m depth. From May until the beginning of July, a positive gradient is constantly registered from 0.4 to 8 m. It is more vigorous at 5 and 8 m depths. After that, the water oxygen saturation decreased until the second week of September. From 0.4 to 5 m depths, such trend is hindered by temporary increases in the values. In particular at 0.4 m, very high variations are measured with peaks that reached 200% of saturation. While at 0.4 and 2.5 m high variations are not associated with a general increment in the trend, at 8 m depth, a positive gradient starting from the second week of September until the beginning of November was detected. The last remaining part of the year is characterized by a negative gradient with the absence of sensible low and high peaks. At 12.5 and 20 m depth, the behaviour of the saturation of the dissolved oxygen is completely different from the previous description. At 12.5, there is a bell-shaped trend with a peak in May and a minimum in October. Afterwards, values are characterized by very high variations, but, in November, they became more stable and began to decrease at the beginning of December. At 20 m depth, the maximum is reached during April; then, a continuous decreasing of the saturation was detected till the end of November or the beginning of December, where, on the contrary, values started to increase to 70%. Values around 60% were then registered during the end of the year.

Figure 4.17: General overview of the time-series comparison between AHFM and IDRONAUT probes.

A point-to-point comparison between the data of the OPTOD chain of sensors and IDRONAUT 's probes is shown in Figures 4.18 and 4.19. The scatter plots are visualized together with the distribution of values. Ideally, if the observations coincided, the points would be arranged on a diagonal. Sensors from 0.4 to 5 m depths have a good correlation with IDRONAUT 316 plus. The dispersion tends to increase when higher values of saturation are registered. There is not a systemic overestimation or underestimation of values, which is therefore evident analysing the scatter plot at 8 m and 20 m depths. In those cases, OPTOD sensors generally underestimate values. The dispersion is higher at 8 m than at 20 m. The data distribution graphs show that values are ranged from about 40% to 160% from 0.4 to 8 m depths, while, at 12.5, from 60% to 100%, and, at 20 m, are even more concentrated

between 30% and 70%. While the correlation with IDRONAUT 316 Plus is quite high, the comparison with IDRONAUT 304 Plus does not show same evidence. In fact, a worse correlation must be highlighted. In Figure 4.19, a high data dispersion and very low correlation is shown at every depth. Such data are also confirmed looking at the boxplot in Figure 4.20, where variance is very high, median values span from −10% to more than −40% and the interquartile range is also generally quite high in particular at 12.5 m depth. A better performance can be mentioned for the values collected at 20 m. On the contrary, the differences comparison between the AHFM system data and the IDRONAUT 316 Plus shows median values very close to zero at 0.5, 2.5 and 5 m depths. The worse performances are registered at 8 and 12.5 m depths since the interquartile range is respectively about 20% and more than 30%.

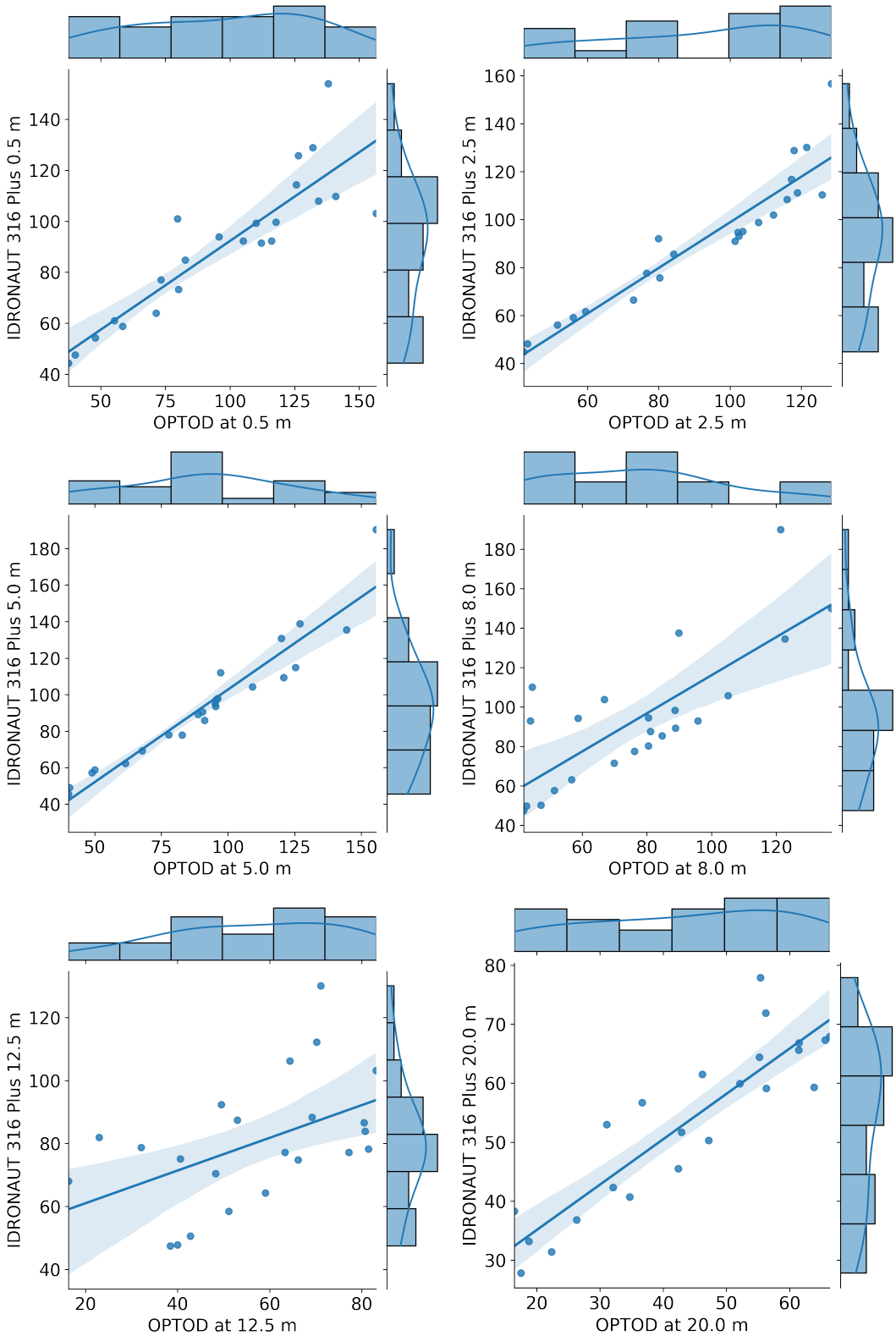

Figure 4.18: Point to point comparison of dissolved oxygen saturation time-series between AHFM and manual sampling with an IDRONAUT 316 Plus probe.

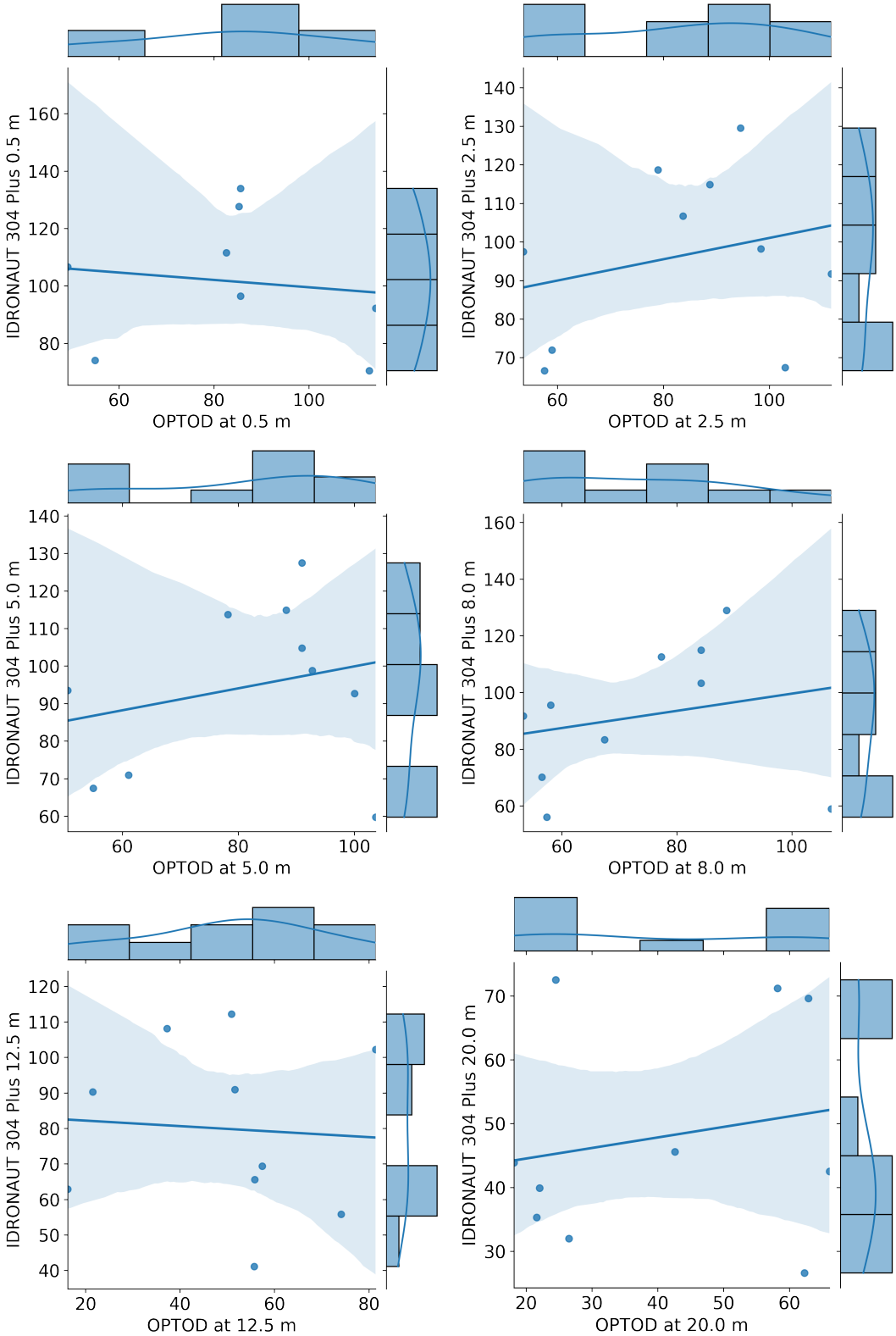

Figure 4.19: Point to point comparison of dissolved oxygen saturation time series between AHFM and manual sampling with an IDRONAUT 304 Plus probe.

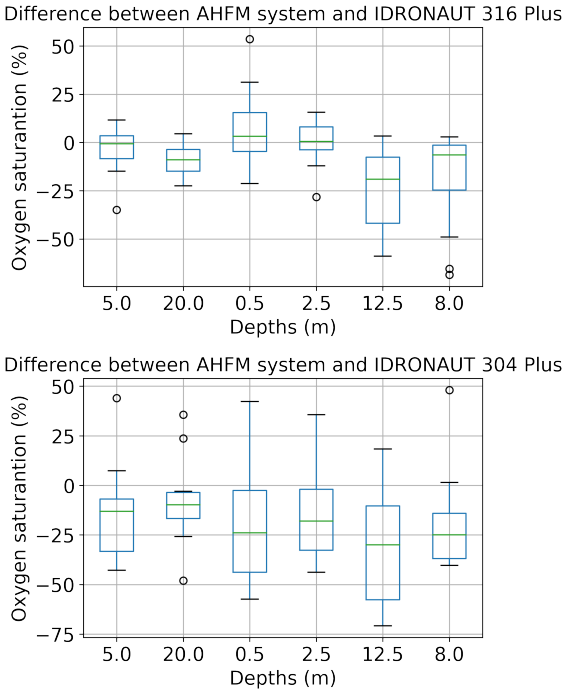

Figure 4.20: Comparison of FL dissolved oxygen saturation between AHFM time series and manual sampling with IDRONAUT probes.

A final consideration concerns the correlation about the water temperature values collected by the analysed systems. To this end, we chose to show just the boxplot representation (Figure 4.21), which, on one hand, confirms a good interrelation with IDRONAUT 316 Plus; on the other, it shows a much better correlation also with IDRONAUT 304 Plus. In the first case, water temperature differences span between ±1 Celsius degree, with median very close to zero at all depths. In the second, even if the interquartile range is very small and the median is much closer to zero than the previous, more outliers are detected.

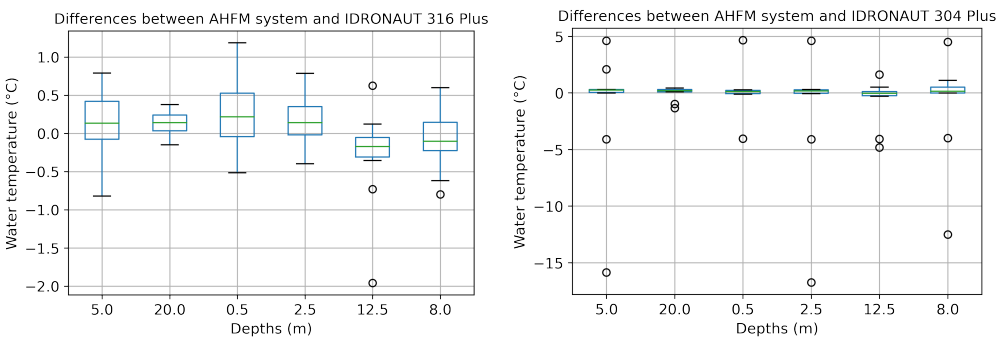

Figure 4.21: Comparison of water temperatures between AHFM and manual sampling with IDRONAUT probes.

4.5 Discussion and Conclusions

In recent years, the monitoring of water bodies is gaining renewed interest because, despite several mitigation actions (e.g., improving the wastewater treatment, banning phosphates from detergents, etc.), they are still susceptible to eutrophication due to climate change effects [64]. In fact, during the last few decades, many climatic parameters are deviating more and more from the average according to the mean values registered since consistent weather observations are available. The World Meteorological Organization (WMO) stated that the most recent years, 2015 to 2021, were likely the seven warmest years ever recorded [65]. In addition, for example, drought events will likely occur more frequently or with a higher intensity either in Europe or in Asia [66, 67]. Hence, the protection of water bodies, such as lakes, is crucial since the demand from various categories including drinking water, industrial use, artificial snow production, agriculture, ecological flow requirements, and hydropower production is increasing more and more [68]. To this end, in Europe and Switzerland, lake protection and water quality control are regulated respectively by the EU Water Framework Directive and by the CIPAIS. Only a low-frequency monitoring that in many cases is not enough to study short-term changes (e.g., algal blooms) is likely more frequent due to previously mentioned factors related to climate change [69] being required. In front of such needs, the digital revolution is an opportunity that can foster the study of new environmental dynamics thanks to the integration of different data sources, the adoption of new technologies for a smarter and high frequent data sampling as well as the breakdown of the digital barriers that still exist in many administrative sectors dedicated to the environmental risk management. In this context, the open digital ecosystem proposed in this article aims at verifying if a solution composed by a new data collection system vendor-agnostic connected to a data management architecture where different data sources are integrated can potentially foster the digitalization of the water sector and help with studying new water dynamics that facilitate decision makers. The data collection implemented concerns an AHFM system based on open technologies (software, hardware) to sample the dissolved oxygen and temperature at different depths. Such system was tested in a real case scenario on Lake Lugano and compared with the samples gained from a traditional campaign which are regularly performed by the local limnologists. Such system demonstrates good reliability and very promising results after

one entire year of testing, from 1 January 2021 till 31 December 2021, in harsh environmental conditions, since it was positioned in the middle of the southern part of Lake Lugano, a deep sub-alpine lake. The results of the data completeness showed a very reliable system with a percent of 97.3% with a daily data completeness above 99%. Therefore, the quality assurance performed on the time series collected highlighted how the system did not have critical biases in values or missing data. Almost 97% of the data demonstrated a very high daily quality after testing each value with an increasing level of restriction. However, in the future, more analysis is requested to better investigate the impact of the maintenance activity on the values. In fact, especially during the most productive period of the year (spring and summer), sensors were usually covered by an algae film which could somehow affect the dissolved oxygen measures. Due to the positive results reached, at the time of writing this article, the system was expanded with additional sensors in order to monitor the Chlorophyll-a and Phycocyian concentrations with UNILUX sensors by Chelsea Technologies, in order to have a wider overview on the eutrophication process of the lake and weather parameters with a compact Lufft station. Moreover, in the future, the data collected by this system will be further studied to implement possible new strategies to monitor the lake productivity, which at this moment is based on an effective but expensive and limited methodology, and algal blooms. The AHFM system is mutually connected to the server side open software suite composed of services with different scopes (see Section 4.3.4) that integrate different data sources, enabling the fully exploitation of the information thanks to the use of the SOS from OGC international standards. In this way, data integration will open up new opportunities for data validation and correlation, as well as new elaborations which can be done by cross analysing the observations. To this end, thanks to this system, it was possible to easily compare and validate the AHFM system with the data of the dissolved oxygen gathered with two types of probes (IDRONAUT 304 Plus, IDRONAUT 316 Plus) that are commonly used during the traditional campaigns performed on the lake. The comparison shows a very good correlation especially with the IDRONAUT 316 Plus showing, on one hand, that the system developed is collecting good quality of data, and on the other hand confirming that the other probe IDRONAUT 304 Plus had some issues probably caused by the membrane disk which needed to be replaced. The open source nature of the software suite developed permitted the adoption of such system by the CNR of Pallanza, a partner of the SIMILE project, which

allowed for further individuating bugs and developing new features. This is an important achievement that demonstrates how the openness philosophy can potentially decrease costs, share skills and increase collaborations. In conclusion, the open, cost-effective and replicable digital ecosystem, which was tested and validated for one year in a real case scenario, demonstrates how new technologies can potentially offer an open solution that can on one hand solve the interoperability and help in overcoming the limiting factors (see Section 4.3.4) that still permeate the monitoring in the water sector.

Author Contributions: Conceptualization, D.S., M.C., F.L. and C.C.; methodology, D.S. and M.C.; software, D.S.; validation, D.S., M.C., C.C. and F.L.; software testing and validation, A.L. and D.M.; formal analysis, D.S.; data curation, D.S.; writing—original draft preparation, D.S.; writing—review and editing, all authors; project administration, M.C., S.S.; funding acquisition, M.C., S.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research is carried out within the SIMILE project (informative System for the Integrated Monitoring of Insubric Lakes and their Ecosystems, ID: 523544) funded by the European Commission and the Swiss Confederation in the framework of the INTERREG Italy-Switzerland 2014-2021 programme.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: We would like to thank Stefano Beatrizzotti and Arturo Di Giacinto for the technical support they provided during the deployment of the AHFM system on the platform positioned in the middle of Lake Lugano.

Conflicts of Interest: The authors declare no conflict of interest.

5 Automatic assessment of lake status using an open source approach: Lake Lugano's case study

Daniele Strigaro [a,b,*], Massimiliano Cannata [b], Camilla Capelli [b], Fabio Lepori [b]

[a] Department of Earth and Environmental Sciences (DSTA), University of Pavia, Via Ferrata 9, 27100 Pavia, Italy

[b] Institute 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;

*Correspondence: daniele.strigaro01@universitadipavia.it

Strigaro, D., Cannata, M., Capelli, C., Lepori, F., 2022. AUTOMATIC ASSESSMENT OF LAKE STATUS USING AN OPEN SOURCE APPROACH: LAKE LUGANO'S CASE STUDY, in: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Presented at the ISPRS WG IV/4, III/10, IV/7, V/1 & ICWG IV/III Free and Open Source Software for Geopsatial (FOSS4G) 2022 - Academic Track - 22 / 28 August 2022, Florence, Italy, Copernicus GmbH, pp. 457-463. https://doi.org/10.5194/isprs-archives-XLVIII-4-W1-2022-457-2022

5.1 Abstract

Climate change and human activities are increasingly threatening water resources. In particular sub-alpine lakes are fundamental not only for tourism or other economical activities, but also as a source of water. In this context, there is a strong need to monitor such resources to understand, study and react to known and unknown impacts, so that appropriate mitigation actions can be taken. Unfortunately, although monitoring data already exist for many of these lakes, the information is archived in different formats and servers undermining the full exploitation of data and preventing a more efficient data management. The aim of this work is to improve this situation by implementing a system that integrates and standardizes data coming from different sources. In addition, the system integrates web based tools that estimate lake state indicators using open source software and standard. Thanks to this system, it will be possible to exploit the data potential more fully. This paper focuses on the achievements reached by the research carried out on Lake Lugano in the context of the project SIMILE after two years of work.

istSOS, SIMILE, SOS, open source, lake, open standard

5.2 Introduction

Lakes are a fundamental resource that provide a number of environmental benefits and services to the local economy and the quality of life. For example, they can store water, protecting floodplains from floods and supplying water during droughts. In addition, they help recharge aquifers [**cannata2018open**] and support biodiversity. From an economic point of view, they attract tourism and provide recreational and professional opportunities to residential populations. Unfortunately, climate change and other human impacts are threatening the general health status of lakes and the services they provide [8, 70, 71]. In this context, monitoring can play a key role in the preservation of lakes, in particular by providing insights that could promote timely mitigation actions. Lakes Lugano and Maggiore are two large transboundary Swiss-Italian lakes. They are systematically monitored by both Italian and Swiss administrative agencies. Lake Lugano, the case study chosen for this work, has been monitored since the early 80s with limnological campaigns being performed every two weeks in three stations: Gandria, Melide and Figino. As a result, for some parameters, 40 year old time-series are now available. These series

provide a precious set of data to investigate trends and changes over time. During last years, limnologists have been additional 'non-traditional' methods to measure lake parameters, often with the goal of increasing the spatial and temporal resolution. For example, remote sensing techniques permit to gather indirect information of many lake properties, i.e. water transparency, surface Chlorophyll-a concentrations, water temperatures, etc. [72]. However, remote sensing has intrinsic limits in temporal resolution [73]. The Internet of Things (IoT) revolution [74] and the increasing availability of sensors and low powered Microcontroller allow to build custom, low-budget solutions to measure lake properties using in situ, fixed, high-frequency monitoring systems. Such systems can improve the temporal resolution of the data, which is necessary to quantify highly-dynamic phenomena such as algal blooms [**marce2016automatic**]. This scenario highlights how many data sources could potentially be available to better study lake ecosystems. Unfortunately, for Lake Lugano, a unique, integrated automated monitoring system is still missing. At present, monitoring data is stored in local databases and in different formats, which undermines the full exploitation of the data and prevents a more efficient data management. Therefore, we identified a need to improve data interoperability and standardization, as well as to create a unique digitalized archive for all the available information. This paper presents the main achievements of two years of research aimed a standardizing and homogenizing data from Lake Lugano. The work was carried out in the context of the SIMILE project (System for the Integrated Monitoring of Insubric Lakes and their Ecosystems). The specific objective is to improve data interoperability by developing a system that integrates and standardizes data coming from different sources. The system will facilitate the use of the data by implementing a web based tool to calculate and display a set of key lake indicators using open source software and standard. The INTERREG project SIMILE, born from a collaboration between Italy and Switzerland, aims at developing an information system using an open source approach and innovative technologies to help decision makers in the management and evaluation of the status of the transboundary and sub-alpines lakes such as lakes Maggiore, Lugano and Como. To this end, the SIMILE project intends to intensify the monitoring of these lakes by implementing an open real-time monitoring system. In addition, it will promote the use of data from satellite and citizen science in order to increase the monitored properties of the lakes. Ultimately, SIMILE aims at helping decision makers by integrating data coming from different sources and

offering a single access to the elaborated information.

5.3 Materials and methods

5.3.1 Study area

Lake Lugano is a transboundary lake divided into two main basins, North and South, with an area of 27.5 Km^2 and 21.4 Km^2 and a maximum depth of 288 and 95 m, respectively. Lake Lugano was severely affected by eutrophication, which reached a peak in the 60s-70s. Thanks to an ongoing restoration programme, the state of the lake has been improving, and the current trophic state ranges between meso- and eutrophic.

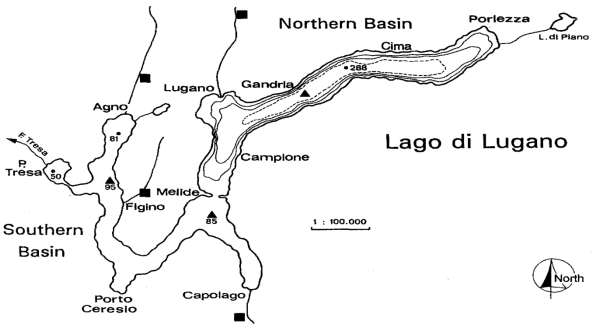

Figure 5.1: Lake Lugano map [**barbieri1992description**].

5.3.2 Data sources

The monitoring data available for Lake Lugano are collected in the South basin using three methods (see Figure 5.2). The first entails the use of multi-sensor CTD probes that are periodically (monthly or bi-monthly) lowered to the bottom of the lake. These sensors measure physical parameters such as water temperature, dissolved oxygen, conductivity, etc. (e.g. IDRONAUT 316 Plus or IDRONAUT 306) or derive biological information, as in the case of fluorimeter sensors, which use visible spectrum fluorescence to estimate algal biomass. The second method is based on the collection of water samples at discrete depths using Niskin bottles. The samples are then analysed in laboratory for an array of chemical and biological variables. The third method is based on a monitoring platform, i.e. an Automatic High Frequency Monitoring (AHFM) system mounted on a platform of 2.5 x 2.5 m deployed in the South basin

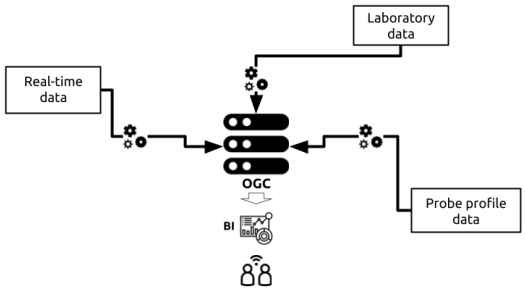

Figure 5.2: Graphical representation of main data sources available for Lake Lugano.

corresponding to the maximum depth. The system was developed during last two years in the context of the project SIMILE. The AHFM system collects real-time data at different depths and with a frequency of 10 minutes (raw data have a frequency of 1 minute, but then they are aggregated every 10 minutes after performing a quality test on each observation). The AHFM system is composed by:

- six optical sensors (OPTOD by Ponsel) to measure water dissolved oxygen (mg/L and %) and temperature (°C) positioned at 0.4, 2.5, 5.0 , 8.0 ,12.5 and 20.0 meters depth;
- four fluorimeters (UNILUX by Chelsea Technologies) to measure Chlorophyll-a and Phycocyanin concentrations (µg/L). Two sensors are positioned near the surface at 0.4 meter depth and the other two at 10 meter depth;
- a compact weather station (WS501 by Lufft) to measure air temperature (°C), relative humidity (%), pressure (hPa), wind velocity (m/s) and direction (°) and solar radiation (W/m^2).

Additional data are further elaborated to derive indexes of lake status (biovolume, LTECO and IPAM indexes).

Unfortunately all these data are currently archived in heterogeneous formats and databases (see Figure 5.3) and are archived in different servers, handled by respective administrative entities (SUPSI, ARPA, CIPAIS).

5.3.3 The information standardization

The missing in-data standardization is a critical issue because it highlights the current lack of data format uniformity, interoperability, ontology as well as an error-prone process, because data are accessible from different sources and a proper synchronization is missing. To fix these problems, international

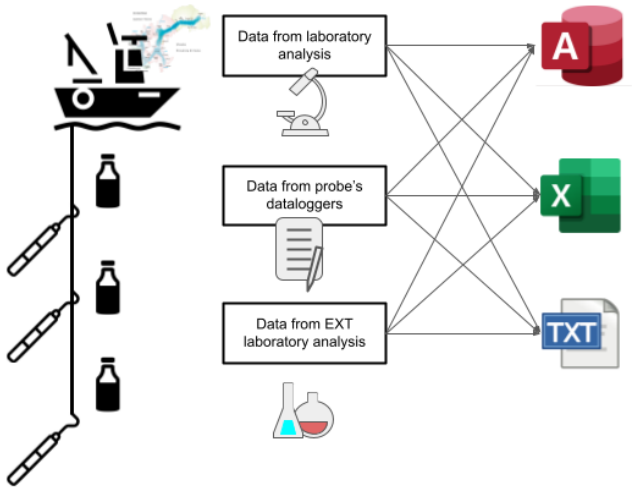

Figure 5.3: Data formats used to archive the historical data of Lake Lugano.

standards are available. They offer a solution to improve data homogenization and standardization [75, 76]. In environmental monitoring, the Sensor Thing API and the Sensor Observation Service of the Open Geospatial Consortium (OGC) are widely used. The first is strongly oriented to provide an open and unified framework for IoT devices interconnection [77] using JSON format. The second is a standard structured and based on XML format. It encodes sensor metadata using the Sensor Model Language (SensorML), and measured values in the Observations and Measurements (O&M) encoding format [57]. The istSOS software has been selected to manage data according to the SOS standard since it is already used in production for the management of the data coming from the Hydro-Meteorological network of Canton Ticino Switzerland [56]. istSOS allows to standardize the ontology of the observed properties and of the unit of measures. In addition, the definitions of common data quality flags/codes improved the use of the software to indicate the quality of the observations inserted. In fact, when quality tests are performed on the data, the software allows to associate the observations to a flag that indicates which tests have been performed [11]. In order to facilitate the import of historical data, a tool was implemented to extend the capabilities of istSOS (Figure 5.4). The importer tool, written in Rust, can decode a text file format and then permits to compile a web form to correctly individuate the data and columns. This tool takes advantage from the WebAssembly paradigm, that almost all browsers support. Hence, it is possible to run a script written with low level language directly on the client without using any server side process. Enabling the software user interface to interpret a WebAssembly script means to permit

an easy expansion of the capabilities. It also increases the performance of running scripts written with low level language.

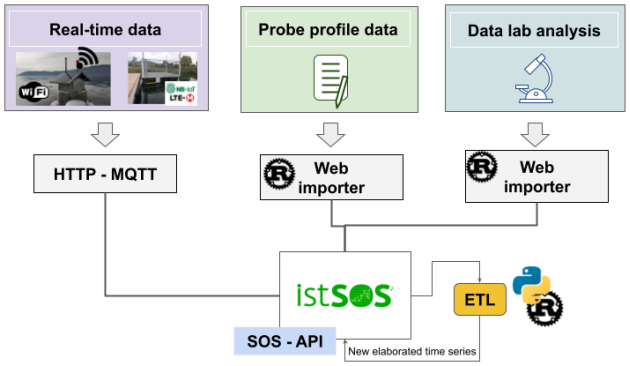

Figure 5.4: Data importing into istSOS.

5.3.4 Automatic Data elaboration

For monitoring the lake's health status, istSOS has been extended by enabling the possibility to write additional scripts, written using the Python programming language, that the software treats as plugins. Such scripts can retrieve observations using SOS standards or web API in order to extrapolate new time series by combining the already registered procedures with custom algorithms. To this end, a YAML configuration file must be structured to provide metadata explaining what the script does and describe the input data needed by the algorithm from which the web user interface can automatically generate a web form. Thanks to this new feature two main plugins were developed: Lake Analyzer and Data Analyzer plugins. The first is based on the LakeAnalyzer software [78], which originally written in MatLab, a tool that allows to calculate common metrics for lake physical states. Hence, some of these functions were translated into Python in order to better integrate the plugin in istSOS, which uses Python, and to offer software based on an Open Source language. In table 5.1, the functions developed at the moment of writing the article are listed along with the requested data inputs.

Name	**Main data inputs**
Schmidt stability	Water temperature profile; Bathymetry; Salinity
Epilimnion temperature	Water temperature profile; Bathymetry;
Thermocline depth	Water temperature profile; Minimum density gradient; Mixed cut-off
Buoyancy frequency	Water temperature profile
Water density	Water temperature profile; Water salinity profile; Salinity
Water density (cond)	Water temperature profile; Water conductivity
Metalimnion depth	Water temperature profile; Slope; Seasonal; Mixed cut-off
Mixed depth	Water temperature profile
Volume aggregation	Water parameter profile; Depths-Volumes values

Table 5.1: List of functions available on the Lake Analyzer plugin.

Schmidt stability Schmidt stability is an indicator of the resistance, or the stability, of a lake in terms of amount of work required to mix the water column uniformly over depth. The algorithm was theorized by Schmidt [79] and was improved and modified by Hutchinson and Idso [80, 81]. The result is a mixing energy required per unit area.

Epilimnion temperature In a stratified lake, the epilimnion is the top layer which has a nearly uniform temperature. The function developed within the LakeAnalyer plugin permits to calculate the average temperature from a profile of temperatures and the lake's bathymetry.

Thermocline depth The depth at which the downward change in temperature or density is greater than predefined thresholds (e.g. a rate of 0.5°C per meter) can be defined as the thermocline. Basically, it is the horizontal plane passing through the point of greatest change along a temperature or density vertical profile. The function adopted here uses the first derivative of the water density profile to estimate the thermocline depth.

Buoyancy frequency The buoyancy frequency, also called Brunt–Väisälä frequency, is a measure used to evaluate the stability of a fluid to vertical displacements such as those caused by convection. It can be calculated as in eq. 5.1 where g is gravity, i is depth, and ρ is water density.

$$N^2 = \frac{g}{\rho_i} \frac{(\rho_{i+1} - \rho_i)}{(d_{i+1} - d_i)} \tag{5.1}$$

Water density Temperature and substances dissolved in water influence the water density and so the capacity of a lake to be stratified or mixed. Therefore, water density is important to understand the physical dynamics of the lake. This function calculate the water density in Kg/m^3 using Martin and McCutcheon's method [82] or the one suggested by UNESCO [83] which is mostly used for seawater.

Metalimnion depths In a stratified lake, the metalimnion is the transition layer between the epilimnion and the hypolimnion. It is characterized by the steepest thermal gradient [84]. The function of this plugin permits to calculate the top and bottom depths of the metalimnion with correspond respectively to the bottom of the epilimmnion and the top of the hypolimnion.

Volume aggregation This function calculates a weighted average along a water profile based on the volume of a water layer thickness (eq. 5.2). It allows to assign different weights to data at various depths.

$$\frac{\sum_{d=i}^{N} a_i V_i}{V_{tot}} \tag{5.2}$$

The second plugin developed regards two main functions: a data aggregator and a data profile creator. The aggregator permits to select a *procedure* and one of its *observed properties* which will then aggregate on the basis of a time interval (ISO8601) and an aggregation function (e.g. AVG, SUM, COUNT,

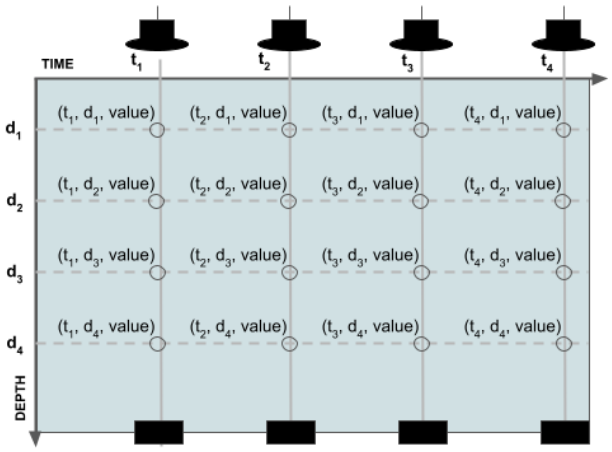

Figure 5.5: Schema of the profile type data composed by timestamp (t_n), depth (d_n) and value.

MAX, MIN). The data profile creator allows to integrate into a single *procedure* of type *profile* two or more *procedures* that share same *observed properties*. For example, the Automatic High Frequency Monitoring system positioned on Lake Lugano is composed by a chain of sensors. Each sensor is registered using a procedure of the type *insitu-fixed-point*, where at a timestamp corresponds to one or more observed properties. Thanks to this plugin, it is possible to create a single *profile procedure* in order to obtain per each timestamp the corresponding values of the *observed properties* at all depths (Figure 5.5).

Once a function is fired, all the information needed to run the algorithm are archived in the database. When an observation is inserted in istSOS, the PostgreSQL database lunches a notification which is listened by the Python server. Such notification carries information (e.g. the procedure name, the beginning and end time of the observations inserted, etc.) that are useful to recognize which procedures are updated and consequently triggers all the processes that have been created using those observations. Such processes will update the calculated procedure in order to update the values.

5.4 Results

The open data management system developed in the context of the SIMILE project is composed by a single instance of istSOS where three istSOS services have been created. A *service* is a isolated space with a dedicated database which is potentially independent by other services also in terms of user permissions.

Name	Description
air-pressure	Atmospheric pressure (hPa)
air-relative-humidity	Absolute humidity relative to the maximum for that air (%)
air-temperature	Air temperature at 2 meters above terrain (°C)
global-radiation	Solar radiation (W/m^2)
water-Chl-a	Chlorophyll-a concentration (µg/L)
water-Neph	Turbidity from Trilux Chelsea technologies (µg/L)
water-O2C	Oxygen concentration (mg/L)
water-O2D	Dissolved oxygen (ppm)
water-O2S	Oxygen saturation (%)
water-PAR	Photosynthetic active radiation (W/m^2)
water-PC	Phycocyanin concentration (µg/L)
water-SDT	Secchi Disk Transparency (m)
water-TSS	Total Suspended Solids

water-cond20	Conductivity at 20 °C (µS/cm)
water-cond25	Conductivity at 25 °C (µS/cm)
water-depth	Depth from surface (m)
water-pH	Water pH at a specific depth
water-pressure	Pressure at a specific depth (hPa)
water-temperature	Temperature of water at a specific depth (°C)
wind-direction	Wind direction (°)
wind-speed	Wind speed (m/s)
wind-speed-max	Wind gust (m/s)

Table 5.2: List of observed properties registered and standardized on istSOS.

A dedicated service, called *ceresio*, has been created to manage all the available historical raw data. It is composed by 26 *procedures* for a total of 22 *observed properties* (Table 5.2) and more than 2 million *observations*. Part of these observations derive from the raw data of the AHFM system which has a rate of sampling of 1 minute. Such data are gathered during maintenance since data are remotely sent to the server after aggregation. In addition, historical time-series, which come from the campaigns performed on the lake since 1972 till today, have been also imported into *ceresio service*. The second *service*, named *ceresiohourly*, is composed by 12 *procedures* that corresponds to the sensors installed on the AHFM system. The AHFM system transmits to the server 10 minutes frequency data in real-time using a MQTT service every 15 minutes. Such data are, in primis, used to offer an instantaneous overview on the real-time parameters measure on the lake, which in future can be used

to detect abrupt phenomena such has algal blooms, but also to check that the AHFM works correctly. In fact, in addition to the environmental parameter already described, observations about current voltage and power are archived in a dedicated *procedure* in order to check the state of the battery and voltages. Finally the last *service*, called *ceresiohourly*, has been created in order to offer data to external services. It comprises hourly aggregated data that thanks to the data aggregator was possible to automatically generate from the real-time data sampled by the AHFM and archived into the procedures registered on *ceresioremote service*. In addition, data from the CIPAIS (International Commission for the Protection of Italian-Swiss Water) annual reports and from ARPA monitoring campaign have been also added to this service in order to offer the public an interactive visualization of such data. A dedicated web user interface will be developed in the context of the project SIMILE. From the integration of the available data sources, it was possible to used the Lake Analyzer web tool in order to calculate indicators. Indicators such as Schimdt stability, Thermocline depth (Figure 5.7) and Mixed depth are useful to understand the physical status of the lake. For example, in Figure 5.6 the trend of the thermocline depth is shown. From 1988 until 2019 it is comprises between 0 and 20 m depth and only in 1991 the thermocline reached the 50 m depth. Of course when the lake is totally mixed the thermocline cannot be detected since it does not exist. The mixed depth is available in Figure 5.7. Since the South basin of Lake Lugano has a maximum depth of about 93 m, it can be highlighted that the water lake was mixed at least every year except for 1988, 2007, 2014, and 2016. In addition to that, the epilimnion depth (see Figure 5.8, ranged between 5 and 30 m. The seasonal behaviour can be appreciate from the graph. In the figures proposed, the dashboard for technical personnel has been showed. This interface permits to visualize data filtering on procedures, observed properties and time. Basically, two type of visualization are offered, *single stats* where there are available the plots for only one procedures and the mode *compare*, it allows to compare *procedures* with same observed properties.

5.5 Conclusions

The proposed paper wants to investigate a fully open web solution in order to calculate indicators that can help understand the health status of the lake and try to solve the limits that are currently affecting water monitoring. Such

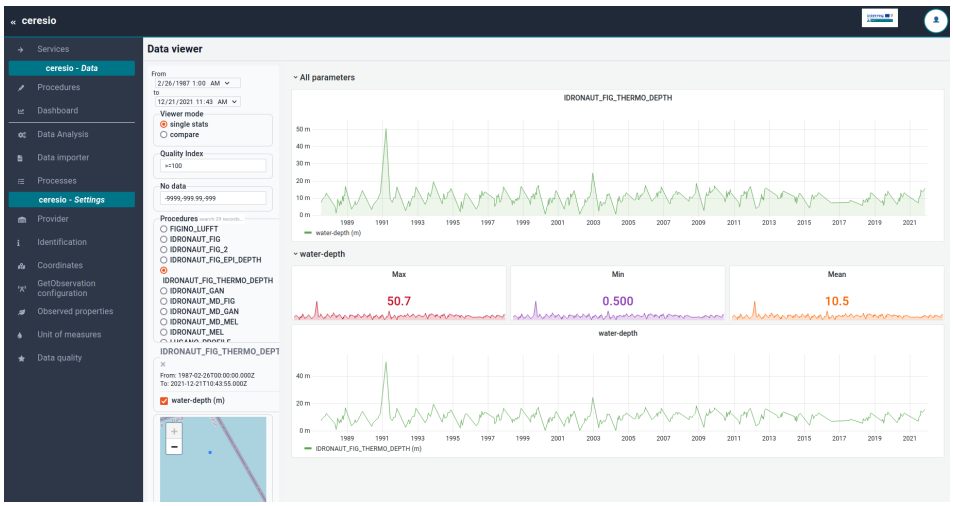

Figure 5.6: Thermocline depth calculated using the LakeAnalyer web plugin and the monthly data collected by the traditional campaigns, previously standardized and imported into istSOS.

Figure 5.7: Mixed depth calculated using the LakeAnalyer web plugin and the monthly data collected by the traditional campaigns, previously standardized and imported into istSOS.

an open platform uses open standards as the Sensor Observation Service (SOS) of the Open Geospatial Consortium (OGC) to integrate different sources of data and to offer the possibility to gather the information in a standardized way. To this end, istSOS, a Python implementation of the SOS standard, was selected as main data management system since it is open source and demonstrated good performance and stability in a production environment. In fact, it is

Figure 5.8: Epilimnion depth calculated using the LakeAnalyer web plugin and the monthly data collected by the traditional campaigns, previously standardized and imported into istSOS.

successfully used to manage the hydro-meteo network of Canton Ticino. Due to the open source nature of this software, it was possible to further extend its capabilities by implementing a system to manage custom scripts and treats them as plugins. In addition, thanks to the WebAssembly technology, it was possible to create an importer application, embedded into the User Interface (UI), to use clients system power to decode text file of profiles data and then trasmit such observations to the istSOS server instance. Thanks to these implemented tools, historical data were integrated into a single data source together with the data coming from the new AHFM system implemented during the project. The AHFM system collects information about temperature, dissolved oxygen, Chlorophyll and Phycocyanin concentrations and weather parameters (wind speed and direction, air temperature, air pressure, air humidity, global radiation). It transmits 10 minute frequency data in real-time every 15 minutes. The data standardization concerns also the ontology as for observed properties and unit of measure. This part was very important since the implemented solution will be adopted by other project partners for handling data of lakes Maggiore and Como. The database shaped by integrating almost all the data available on Lake Lugano support to the fully exploitation of the data for the scientists and researchers. The software is released under the GNU GPL v3 License (https://gitlab.com/ist-supsi/simile/istsos-suite). Due to sensibility of data, historical raw time series are available only upon specific requests. In

order to facilitate and offer common methods to calculate basic lake indicators, a module based on the LakeAnalyzer functions was implemented using Python. The scope is to validate and standardize the calculations and providing a common validated tool. The indicators produced using these tools are automatically updated when new observations are inserted into the procedures which were selected in order to make the computations of each indicators. This feature was implemented extending the capabilities of istSOS by implementing a tool which can take advantage of the LISTEN/NOTIFY feature of the PostgreSQL, the database technology on which the platform is based. In conclusion, an overview of the results about the improvements in data interoperability and standardization reached during these years of project were presented. The proposed solution demonstrated how open software and standard can fill the gap in data interoperability and standardization which still is affecting the lake monitoring. Such open technologies helped in developing an automatic system that can calculate indicators to help decision makers in managing the water resource and scientists to better study the new unknown dynamic and facing the new challenges to which lakes are exposed.

Acknowledgements

The presented research has been carried out in the context of the 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.

6 The Primary Production

6.1 Introduction

Eutrophication is a process that affects water bodies [85]. It consists in an overbundant presence of nutrients such as phosphorous and nitrogen. The high availability of nutrients encourages the abnormal growth of primary producer organisms. The primary producers can synthesize biomass from inorganic substances and light energy thanks to the photosynthesis process. The excess loads of nutrients is caused by antropic actions, it can derive from punctual source such as wastewater treatment discharges, industrial processes and aquaculture plants or from widespread sources (rainfall and runoff from agricultural land). In deep sub-alpine lakes like Lake Maggiore, Como and Lugano, this process is monitored since the 70s when the quality of the water of these lakes reached critical levels [86–88]. Such lakes are very important because they are a resource for local economy, tourism, and more and more used for supply drinkable water to households. Thus, the monitoring of the trophic status of lakes is one of the most important activity that is conducted to access the quality of their waters. The classification of lake trophic status is quite challenging and can be addressed using different methods and variables such as phosphorus loading [89, 90] or concentration [91], algal biomass and algal productivity. The last two methods listed are the most used for the evaluation of the trophic status of a lake. The estimation of the algal biovolume concerns the monitoring of some key parameters such as the concentration of limiting nutrient (phosphorous), chlorophyll and the transparency which can be expressed as Secchi depth, water turbidity or concentration of the total suspended solid. Algal productivity is another important measure to evalute the trophic state of the lake. Usually the productivity is calculated along a year which means an extensive sampling collection during the year. The most used and sensitive method for the estimation is based on the Carbon-14 technique [46, 92]. However, instead of using the Carbon as proxy to estimate the productivity it is possible to use the dissolved Oxygen in water. In fact, based on the photosynthesis process the oxygen is one of the main chemical element involved. Staehr [48] proposed a method based on the dissolved oxygen variations measurements collect from an Automatic High Frequency Monitoring system. This paper is focused on the comparison between these two methods with the aim to provide the basis for a real-time evaluation of the lake

metabolism if the results show a good correlation. Such evaluation has been perfomed considering Lake Lugano as study area. The productivity of this lake has been monitored since 80s using the C14 method. Monthly campaigns are performed to collect eleven water samples concentrated in the photic layer at depths of 0.4, 1.25, 2.5, 3.75, 5, 6.5, 8, 10, 12.5, 15, and 20 m. Inside the collected bottles, about 0.1 mL of diluted 14C stock solution (i.e., 218 kBq radioactivity per sample) is added. Then, the bottles are left at their collection depths for 4 h (from 10:00 to 14:00 h). Once the bottle are retrived they are sent to an external laboratory in order to measure the amount of β-activity (14C) assimilated by phytoplankton during the incubation time [47, 93] and the C assimilation rates. These rates are subsequently depth-integrated to calculate the hourly production rate per unit surface. In 2017 Franchini et al. [94] published a paper where they provided additional improvements to the algorithm used for the primary production estimation on Lake Lugano. However, even if this is the legally recognized method by the decision makers (Canton Ticino and Regione Lombardia) it still presents many challenges that the adoption of the proposed new mthodology can potentially addressed. The individuated negative aspects are:

1. use of radioactive element that is critical for the health of technicians;
2. low temporal resolution due to the fact that monthly campaign are performed;
3. water bottles are exposed from 10 a.m. to 14 p.m. This limited exposition do not cover the entire dayily period of production and in particular the night where the respiration happens;
4. the collected samples need expensive analysis which in the case of Lake Lugano are performed by external laboratories which caused an additional costs

Since November 2020 an Automatic High Frequency Monitoring System were deployed on the lake using an open source approach [95]. Such system is composed by six sensors which measure the variations of temperature and dissolved oxygen in water with a frequency of 10 minutes. Thanks to the data collected during the year 2021 it was possible to develop an algorithm based on the Staehr method and compare the results obtained between them. Due to these individuated limiting factors the proposed work aim at studying new

technologies to apply an open approach in order to provide an algorithm and a system that can be further and esasily calibrated, developed and shared.

6.2 Materials and Methods

6.2.1 The area of study

Lake Lugano is located in the insubric pre-Alps, at the North of the Po valley. It is a transbouandary Lake between Italy and Switzerland (Lombardy and Canton Ticino regions). Lake Lugano consists on three sub-basins which have different morphometric and hydraulic charateristics. the North Basin (between Melide and Porlezza) and the South basin (between Capolago and Agno) are separated by the dam bridge of Melide. Finally the little Ponte Tresa basin located near the outfall. The Northern portion of the lake is much deeper than the others with a chemical stratification (meromictic) which persist over the year and it has a very long water renewal time. The southern basins are prone to the complete mixing of the water column every year (monomictic). Lake Lugano is located in at the territorial border between border between Switzerland (Canton Ticino) and Italy (Lombardy Region) at an altitude of 271 m above sea level. The three basins present different morphological and hydrological characteristics (see Table 6.1). The North Basin is very deep (288 m) with a limited catchment area (270 km^2) compared to the water volume (4.68 km3), consequently it has a high theoretical turnover time, about 12 years. The South basin has a maximum depth of about 95 m and a basin area of 290.3 km^2.

	North basin	**South basin**
Catchment area (km^2)	269.7	290.3
Lake surface area (km^2)	27.5	20.3
Volume (km^3)	4.69	1.14
Maximum depth (m)	288	95
Average depth (m)	171	55
Water renewal (year^{-1})	0.081	0.67

Table 6.1: Lake Lugano morphometric characteristics.

The land surrounding the lake is typical of the subalpine Pre-Alps, with mountainous areas rising to over 1500 m above sea level, interrupted by wide

valleys of fluvio-glacial origin. The insubric climate and the morphology of the catchment area, typical of the pre-Alpine belt, give the tributary rivers typical torrential characteristics. Its emissary is the Tresa River, which flows out of the branch into Lake Maggiore. The lake basin is home to a population of about 195000 people, 73% of whom are Swiss; the total tourist movement exceeds 70000, so the overall population equivalent is just under 270000.

Lake Lugano is exposed to stressing factors which caused the degradation of water quality [96]. In particular, since 60s-70s, the eutrophication process raised because of the excess nutrient load, by means of excess input of phosphorous (P). From oligotrophic, it reached an eutrophic status which concerns many problems in term of water quality. P is the limiting factor for PP in many freshwater lakes [97, 98]. For lake lugano, the mitigation actions chosen and put into practice since the 1980s have involved monitoring external sources of phosphorous loading. The result of these actions has led to significant improvements in trophic status down to a mesotrophic level. Since the late 1980s, Lake Lugano has undergone research limnolgical surveies funded by the International Commision for the Protection of Italian-Swiss Waters (CIPAIS). Long time-series has been collected and help in gaining such good results. In addition to the eutrophication process, Lake Lugano, as well as in other deep sub-alpine lakes, is also affected by the spreading of potentially toxic cyanobacteria and appearence of superficial bloom [99], organic micronutrients contamination and finally microplastic pollution. Such emerging topics are gaining interest and are further studied by researchers [100–102].

6.2.2 The Automatic High Frequency Monitoring System

In order to monitor the productive surface layer, a chain of sensor was installed on a floating platform deplyed on the Lake Lugano at Figino (Lat. 45.95866, Lon. 8.89371, EPSG 4326). The chain is the sensor part of an Automatic High Frequency Monitoring (AHFM) system which was designed and developed using an open source approach [103] and positioned in the middle of the lake during November 2020. The AHFM system is collecting data continuosly from the 1st January 2021. The time serie has a frequency of 10 minutes and is transmitted to the server side every 15 minutes. The sensor chain is composed by six OPTOD sensors by Ponsel (see Table 6.2) that measure the concentration of the dissolved oxygen and the temperature. They are positioned in the photic area to get the oxygen variations caused by the

primary producers. In Figure 6.1, a simple schema of the AHFM is present. In addition to the OPTOD sensors, since March 2022, other sensors have been installed on the platform: a Lufft WS-501 station to gather weather parameters such as air temperature, humidity, pressure, global radiation and the wind speed and direction (see Table 6.4); four sensors from Chelsea Technologies UNILUX (see Table 6.3) which measure the Chlorophyll-*a* and Phycocyanin concentrations at the surface lavel and at 10 m depth. The position of the OPTOD sensors has been chosen to correspond to the samples taken during the campaign which are then used for the estimation of the PP using the recognized method.

Measure principle	Optical measure by luminescence
Measure ranges	0.00 to 20 mg/L 0.00 to 20 ppm 0-200%
Resolution	0.01
Accuracy	+/- 0.1 mg/L +/- 0.1 ppm +/- 1%

Table 6.2: OPTOD sensors technical specifications.

Measure principle	Highly sensitive miniature fluorometer
Measure ranges	Chlorophyll-*a* 0 to 400 μg/L Phycocyanin (Cyanobacteria) 0 to 400 μg/L
Limit of Detection	Chlorophyll-*a* <0.01 μg/L Phycocyanin (Cyanobacteria) <0.01 μg/L

Table 6.3: UNILUX sensors technical specifications.

6.2.3 Data Management System

The data used in this project were archived, managed and shared through a web platform composed by different services which were containerazied using the docker and docker compose technology [35]. Such software is based as core component on istSOS which implements the Sensor Observation Service (SOS) of the Open Geospatial Consortium (OGC). Two main services were created to manage raw data collected from the AHFM system and aggregated and

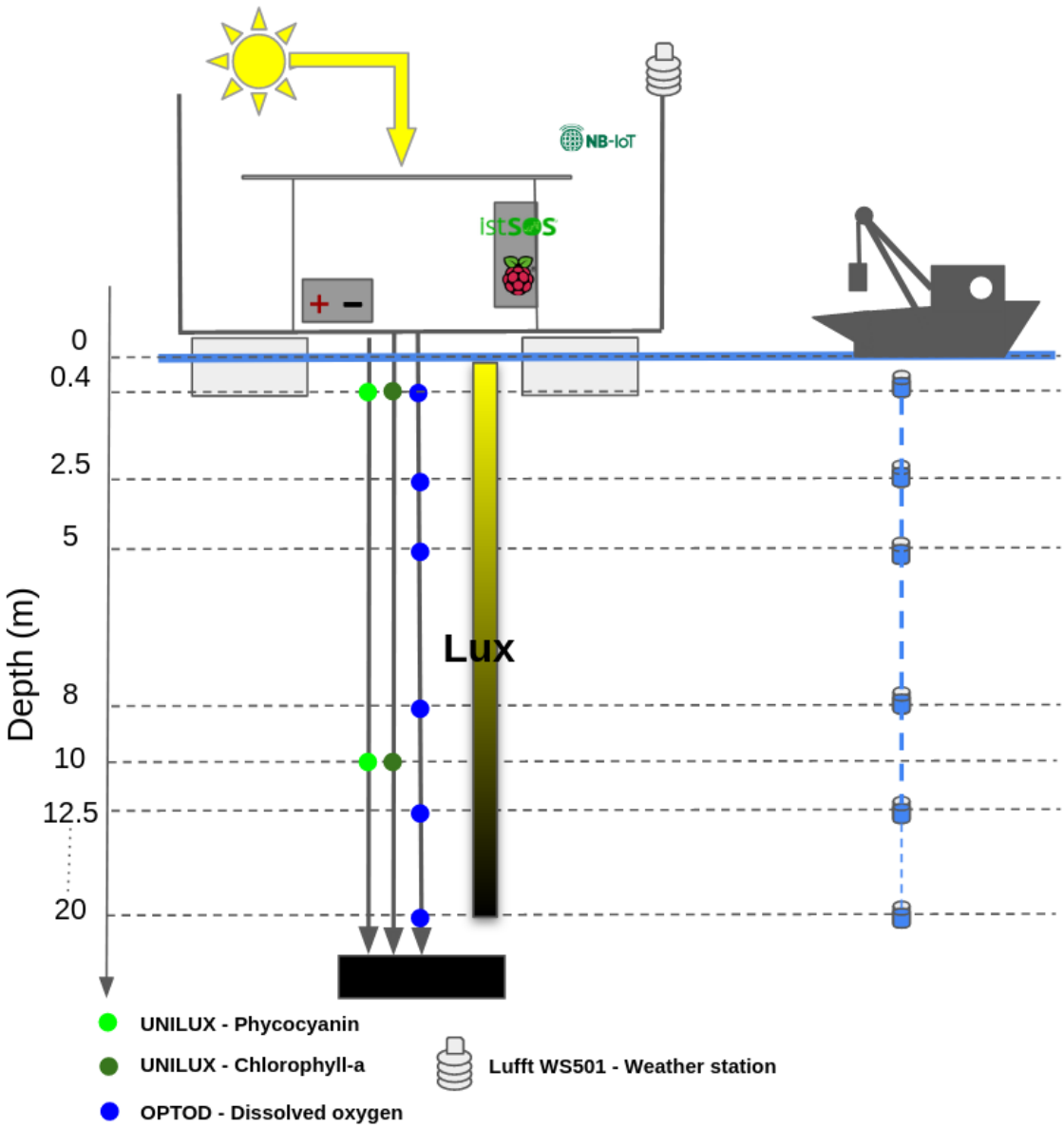

Figure 6.1: Schema of the AHFM system.

validated data with a frequency of 10 minutes. In istSOS, a service can be defined an instance where you can organize your sensor data. Every service has its own database schema independent from other instances. Hence, data and metadata are fully compatible with an international recognized standard so that they can be easily harvested, gathered and shared to other web services increasing the data interoperability since the SOS is open and adopted schema and formats [57].

6.2.4 The Primary Production estimation

The estimation of the lake metabolism is important to quantify the rate of production, understand the cycling of carbon in lakes and assess and manage the health of these ecosystems. Thus, an algorithm based on the methodology proposed by Staehr et al., 2010 [48] has been developed using an open source approach and the Python programming language. The methodology selected need the following input paramters to performed the elaboration:

- global radiation to discriminate between day and night;
- wind speed to calculate the piston velocity so that it is possible to estimate the micro-turbolence effect on the surface area;
- air pressure to correct the oxygen saturation;
- water dissolved oxygen to calculate the oxygen variation over a period of time;
- water temperature to calculate the mixing depth and Schmidt coefficient.

At the time of writing this thesis, the PP estimation was performed using the data collected during year 2021. Unfortunately, since the weather station was still not installed, the requested weather parameters, such as global radiation, wind speed and air pressure, were gathered using the nearest weather station available of the MeteoSwiss network. To this end, the Lugano station was selected as the most representative station of the area. The Lugano station dataset has a time resolution of 10 minutes. The other two parameters, temperature and dissolved oxygen of water, are measured by the developed AHFM system (see sec. 6.2.2) with a time resolution of 1 minute for raw data, which is the dataset used within to start the presented elaboration, and 10 minutes for aggregated data. In Table 6.5, all the equations

used in the developed algorithm are exposed. The eq. 6.1 is the governing equation of all the process. It puts in relation the differences of the oxygen concentration within a time interval with the Gross Primary Production (GPP), the Respiration (R) and the O_2 exhange with the atmosphere (F). Considering that the GPP is equal to zero during the night since no photosynthesis is possible without sunlight, R can be estimated by the changes in dissolved oxygen detected during this period. Thus, R is assumed to be equal during daylight. In Figure 6.3 the flowchart developed within the algorithm is shown. The method is based on the equations in the table 6.5 and it was developed from scratch using the Python programming language. The proposed flow is novel and adapted to the case of the Lake Lugano. It can be potentially used for real-time estimation of the Primary Production, which for robust results it needs of course to be validated and guided by expert in post-processing. However, it will allow to have real-time estimation of the trend of the lake production. In addition, the calculation of the PP is a novelty for the deep subalpine lakes and deep lakes in general like Lake Lugano. Starting from the beginning of the flowchart, the AHFM data are managed by removing values collected during the maintenance periods and performing the calibration correction when needed, then a gaussian filter is performed on the data to remove the noise due to unrelated dynamics such as internal lake motions and effects due to other physical processes (e.g turbulent diffusion, gas exchange, etc.). With regards the recalibration method, during the maintenance periods sensors were checked and recaibrated when needed. A coefficient is registered as the difference between the value before and after the calibration. The calibration of the OPTOD sensors is usually performed over the air. In this environmental conditions, OPTOD sesnors should measure, once dried, an oxygen saturation value near to 100%. While the algorithm is running, when a calibration factor is detected a linear function is applied on the values which fall in the period considered (see Figure 6.2).

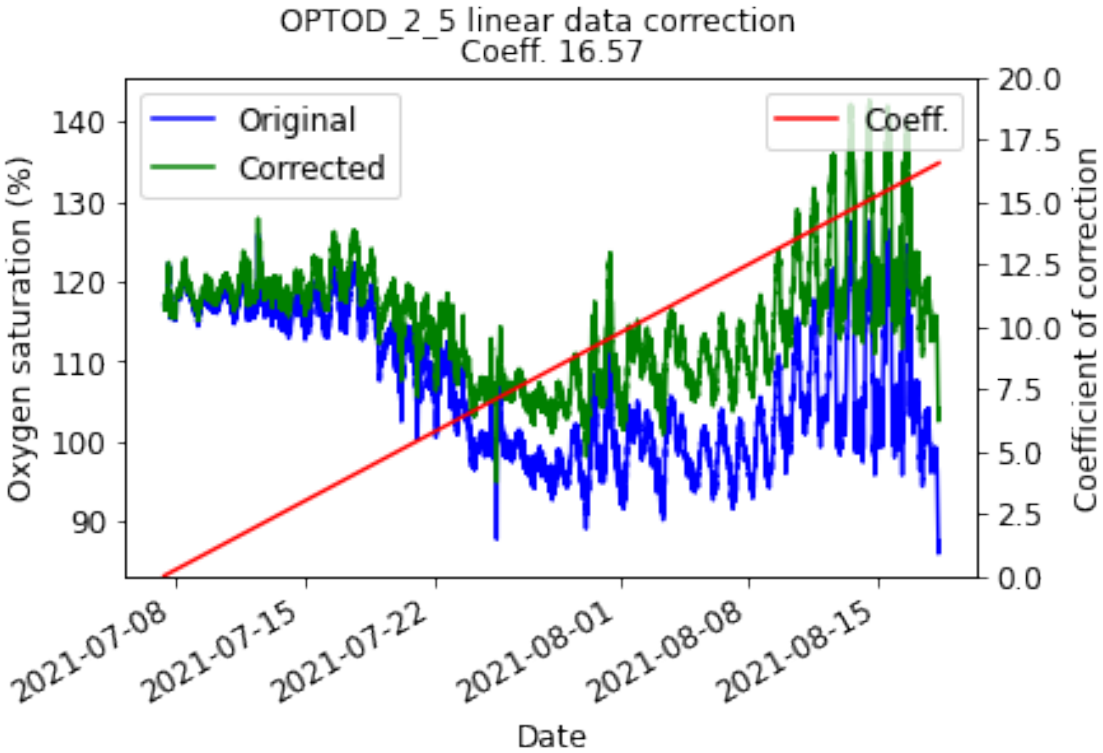

Figure 6.2: Example of a re-calibrated series of data.

Once the series were corrected, data of all the sensors, positioned at different depth, were hourly resampled and then merged with the Lugano weather station time series which were also previously resampled with the same frequency. The merged data formed a single DataFrame from which the PP elaboration begins. Before starting the calculation of the equation in the flowchart, the depth of the mixed layer (Z_{mix}) was calculated using the temperature values collected by the sencors chain and a critical threshold value of 0.5°C. The depth of the mixed layer is a parameter which is requested in eq. 6.6. The oxygen saturation is then barometrically correct using the formula in eq. 6.2. The Schmidt coefficient (Sc) is calculated as a function of the temperature and it determines the physicaly exhcange of oxygen with the atmosphere. Such coefficient is then used in the next eq. 6.4 where the Piston velocity in m/h is derived. The Piston velocity control the transport of dissolved gasses in water surface. Thanks to these parameters, the F can be estimated in order to remove all the physical effects on the variations of the dissolved oxygen (see eq. 6.5 and eq. 6.6). The dayfraction (eq. 6.7) has been derivated using the lighthours which are estimated considering the period during the day when the global radiation greater than zero. From the night and light hours, R, Net

Primary Production (NEP) and GPP have been estimated per day (see eq. from 6.8 to 6.13).

Id	Equation	Ref.
6.1	$$\frac{\triangle O_2}{\triangle t} = GPP - R - F \quad (6.1)$$	[104]
6.2	$$O_{2\ \text{sat corrected}} = O_{2\ \text{sat}} \times c \quad (6.2)$$ $$c = \frac{(pressure \times 0.00987) - 0.0112}{100}$$	[105]
6.3	$$Sc = 0.0476T^3 + 3.7818T^2 - 120.1T + 1800.6 \quad (6.3)$$	[106]
6.4	$$k(mh^{-1}) = k_{600} \times (Sc/600)^2 \quad (6.4)$$	[107]
6.5	$$F(gO_2m^{-3}h^{-1}) = k(O_{2\ \text{meas}} - O_{2\ \text{sat}}) \quad (6.5)$$	[108]
6.6	$$NEP_{\text{h}}(gO_2m^{-2}h^{-1}) = \triangle O_2 - (\frac{F}{Z_{\text{mix}}}) \quad (6.6)$$	[108]
6.7	$$dayfraction = lighthours/24h \quad (6.7)$$	
6.8	$$R_{\text{h}}(gO_2m^{-2}h^{-1}) = meanNEP_{\text{h}}duringdarkness \quad (6.8)$$	

6.9	$R_{\text{daytime}}(gO_2m^{-3}h^{-1}daylightperiod^{-1}) = R_{\text{h}} \times 24 \times dayfraction$ (6.9)	
6.10	$R_{\text{day}}(gO_2m^{-3}d^{-1}) = R_{\text{h}} \times 24$ (6.10)	
6.11	$NEP_{\text{daytime}}(gO_2m^{-3}h^{-1}) = meanNEP_{\text{h}}duringdaylight \times dayfraction \times 24$ (6.11)	
6.12	$GPP_{\text{daytime}}(gO_2m^{-3}h^{-1}) = NEP_{\text{daytime}} + R_{\text{daytime}}$ (6.12)	
6.13	$NEP(gO_2m^{-2}h^{-1}) = GPP - R_{\text{day}}$ (6.13)	

Table 6.5: List of equations based on the published paper by Staher [48].

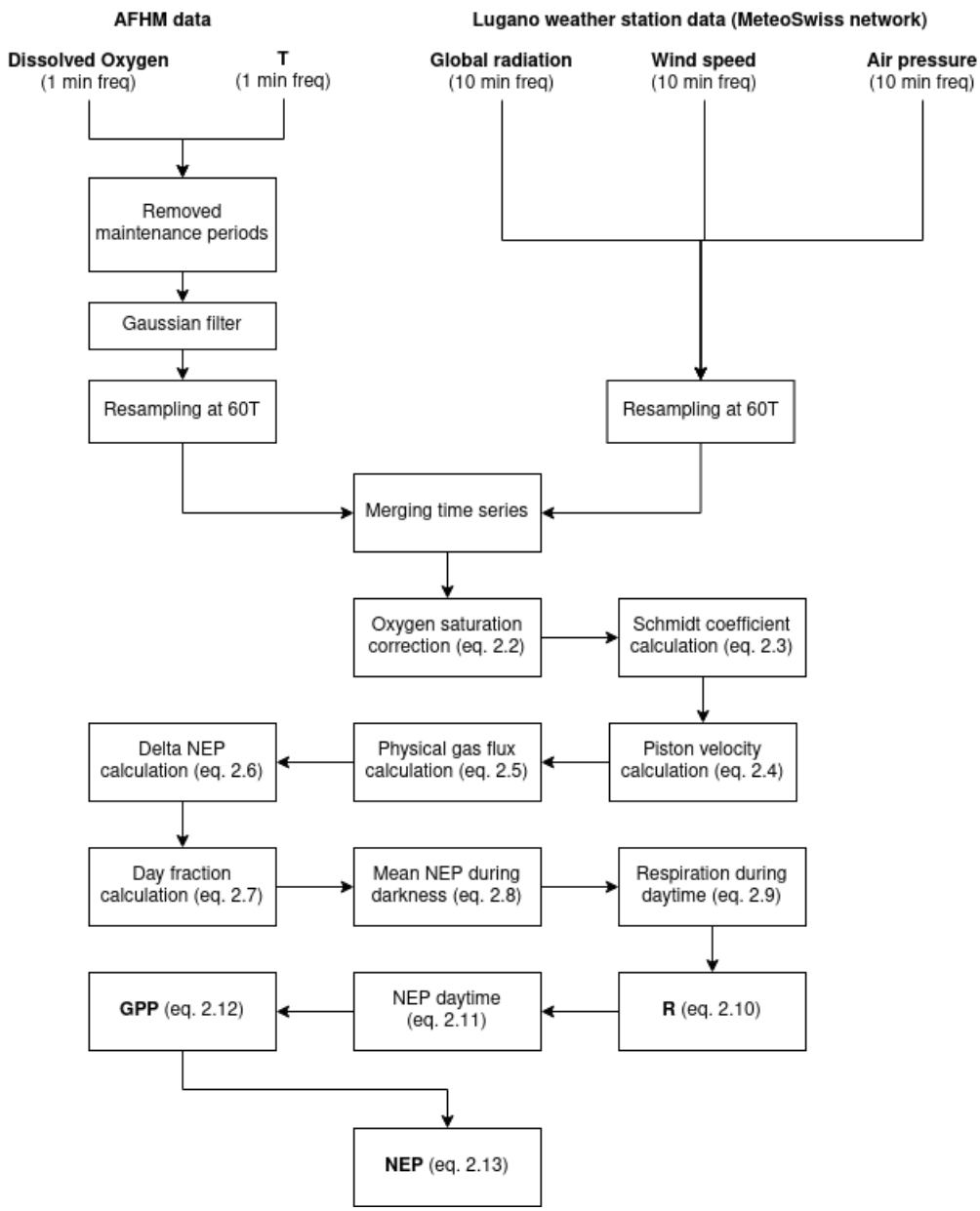

Figure 6.3: Flowchart of the algorithm.

The described algorithm was executed using six different time series collected by the sensors chain. The main result is the R, NEP and GPP trends over the year per each depth (0.4, 2.5, 5.0, 8.0, 12.5, 20.0 m). The step forward concerns the integration of the values obtained by performing a polynomial function (second degree) over the data profile hourly resampled. Once data are integrated over the profile, to get the total GPP and NEP, the hourly data were summed at a daily frequency and subsequently using a temporal monthly and yearly temporal window. The GPP and NEP daily timeseries weere useful

to make comparison also in terms of monthly daily mean between the two methods considered. At the time of writing this paper, the algorithm was developed using Jupyter Notebook and the Python programming language. The source code is licensed under the GNU General Public License version 3 and available at the following repository (Mettere repo).

6.3 Results

The temperature profiles of year 2021 are shown in Figure 6.4. The highest temperatures were measured during the summer period from July to Septmber where over 20°C were registered untill the depth of 8.0 m. During this period, in the first water layer (from 0.4 to 2.5 m), water reached over 27°C of temperature. The monitored water column (0-20 m) had the same temperature during January, February and in December 2021. Such behavior indicates the the water column considered during this periods was mixed. The mixing process began in October and ended in December where it is evident the omogenization of the temperatures from the top to the bottom.

In Figure 6.5, the profile of oxygen saturation highlights how the highest values of saturation were measured from 5 to 10 meters depth (over 145%). During January and December 2021 values were very silimar along the column otherwise generally the pick of oxygen were often concentratd at 5 meters or near the surface at the end of August and the beginning of September. Values around 10-20% were mearures at 12.5 meters depth betweeen October and November.

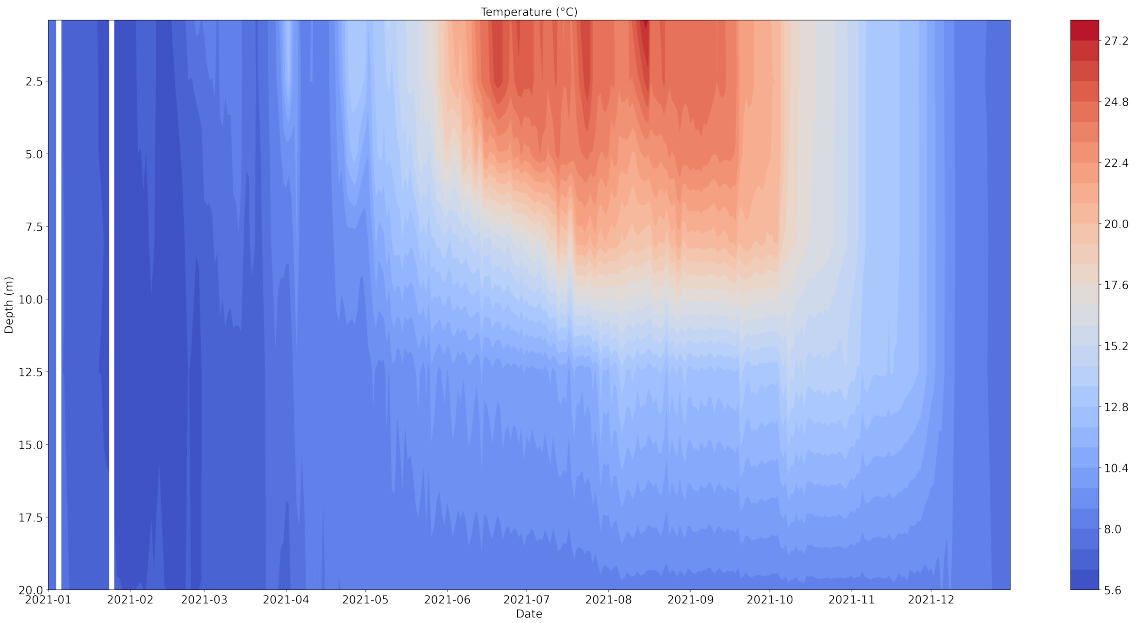

Figure 6.4: Temperature profiles in °C coming from the data collected using the chain of OPTOD by Ponsel sensors (year 2021).

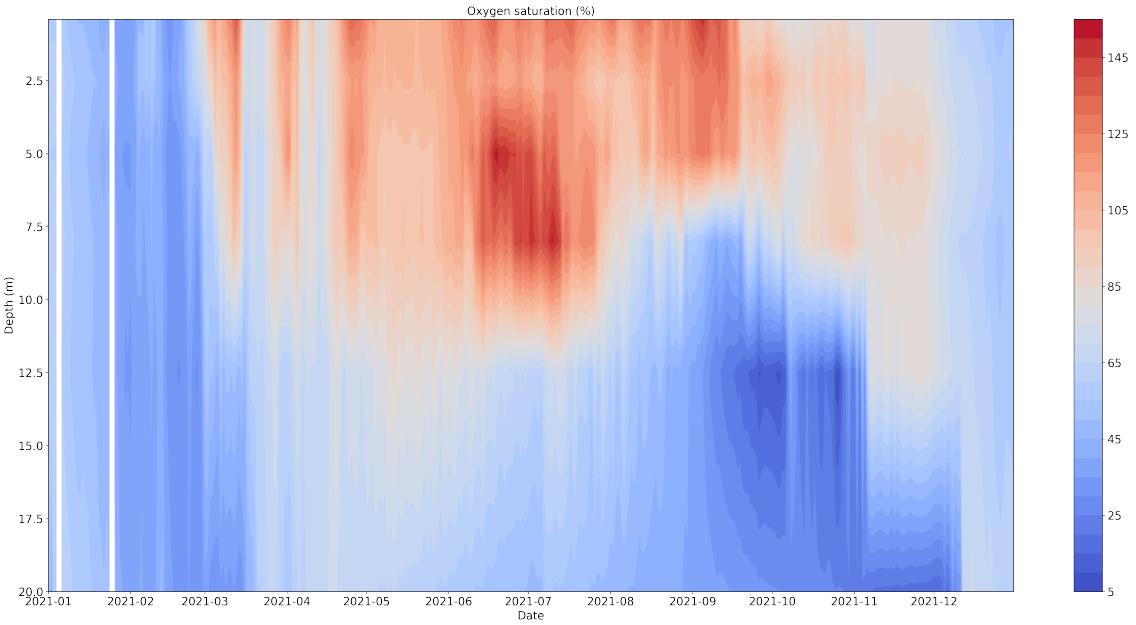

Figure 6.5: Oxygen saturation profiles in % coming from the data collected using the chain of OPTOD by Ponsel sensors (year 2021).

In Figures 6.6 and 6.7 the daily GPP, NEP and R values, expressed in $gCm^{-2}d^{-1}$, are showed in a series of plots per depth. The description of the results presented in this section are focused on the smoothed trend plotted

on the graphs which on one hand increase the readability of the values on the other it is a method used to removed the fluctions and variability of the values caused by other effects (e.g. physical effects) as already explained in subsection 6.2.4. Looking at the graphs, and in particular at the GPP trend, five quite distinct events can be highlighted. The first (E1) peaks in March. The E1 event is present at all depths with peak values around 0.5 $gCm^{-2}d^{-1}$, particularly in the surface layer up to a depth of 5-8 meters. During this event, the NEP is generally positive at almost all depths. The second event (E2), has maximum GPP values reached at the beginning of April to which corresponds a net production in positive balance with values reaching as high as 0.2/0.3 $gCm^{-2}d^{-1}$ at profundities of 0.4 and 2.5 m and then slowly decreasing down to the bottom depth where it still remains with positive values. The third event peaks at the end of APril and spans from the middle of the same month to the middle of the second week of May. The GPP values reached by this event are similar to those of E1 and E2 depending on the depths considered. However, compared to E2, the net balance is generally positive at all depths with lower values in almost all the depths. Between June and July there is evidence of another rise in GPP values at depths of 0.4, 2.5, 5.0 and 8.0 m (E4). E4 has the characteristic of having a more pronounced duration at the depths of 5.0 and 8.0 m. The peak values of GPP are always around 0.5 $Cm^{-2}d^{-1}$. The NEP significantly in positive at all these depths considered with values comprises between 0.1 and 0.3 $Cm^{-2}d^{-1}$. The E4 event has a different conformation at 12.5 m while absent at 20.0 m depth. At 12.5 m, GPP values remained between 0.4 and 0.5 g from May to August. Also during the same period and at the same depth, NEP values remain consistently positive. The fifth event has a peculiar pattern in that it presents peculiarities in all depths. The highest NEP values are recorded at 0.4 m while an almost flat trend is recorded at depths 2.5, 12.5, 20.0 m. The last two peculiarities concern a negative decrease in net production recorded around November at the depth of 12.5 m and during the first week of December at 20.0 m.

Figure 6.6: Daily values of GPP, NEP and R in $gCm^{-2}d^{-1}$ at depths 0.4, 2.5 and 5 m starting from the top.

Figure 6.7: Daily values of GPP, NEP and R in $gCm^{-2}d^{-1}$ at depths 8, 12.5 and 20 m starting from the top.

In Figure 6.8 the daily values of carbon per square meter was integrated over the profile from 0.4 to 20.0 m depth. The integration was performed calculating the best fit polynomial function which pass through the values which were hourly resampled. From these daily values the total sum of carbon per month and per year were calculated. Focusing on the NEP values they are close to zero during the in tail-end months of the year, with also negative values in November. Then, once the productivity starts, values raise over the zero for all the summer until October. The GPP trend reflects the described events E1, E2, E3, and E4 but starting from August the values regularly tend to near 0.0 $gCm^{-2}d^{-1}$. The GPP has maximum peaks of more than 6.0 $gCm^{-2}d^{-1}$, while the maximum daily NEP is registered during the summer with values of about 2.0 $gCm^{-2}d^{-1}$.

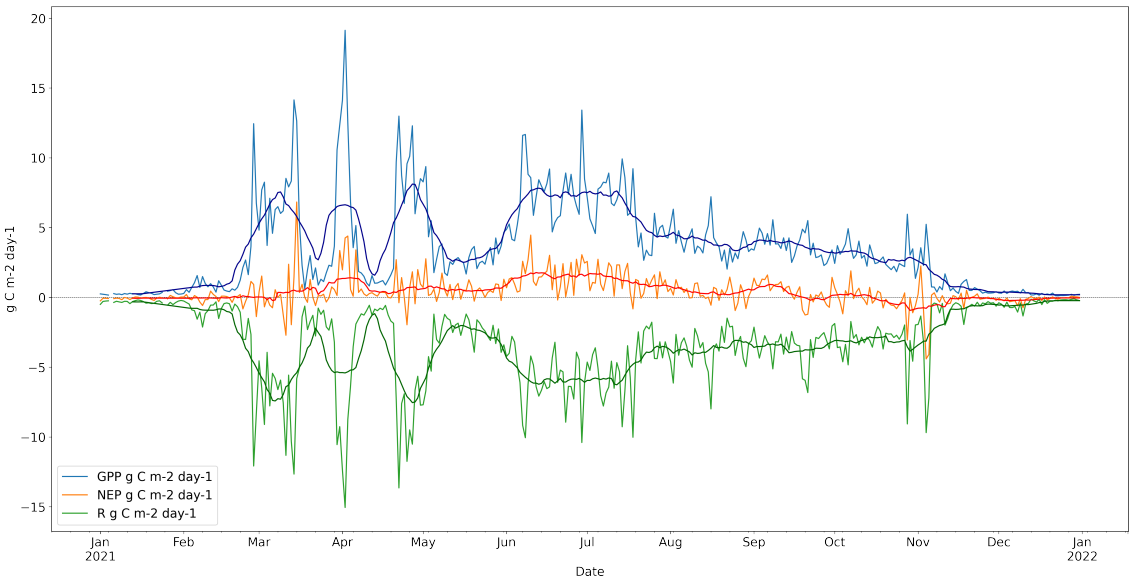

Figure 6.8: Daily values of GPP, NEP and R over the year 2021.

6.3.1 Results comparison between methods

The total Carbon calculated using the C14 method along the year 2021 is about 350 $gCm^{-2}d^{-1}$ while it is 140 $gCm^{-2}d^{-1}$ using the algorithm which implements the Staehr method. To the end of better understanding how the values are distributed along the year, in Figure 6.9 the calculated monthly sum are represented for both the cases. It is clear how the spring production of algae has been well modelled by the two methods with a different magnitude. The curve of the C14 has a peak of about 87 $gCm^{-2}d^{-1}$ while it is less than 20 $gCm^{-2}d^{-1}$ for the other method. While during the summer the net productivity

is linear with values higher than 20 $gCm^{-2}d^{-1}$, the new methodology register a peak of more than 40 $gCm^{-2}d^{-1}$ in June. In winter the values calculated with the new algorithm go down to near zero. This trend is also folows by the C14 method but with a minor changes netweeen the maximum and the minimum value.

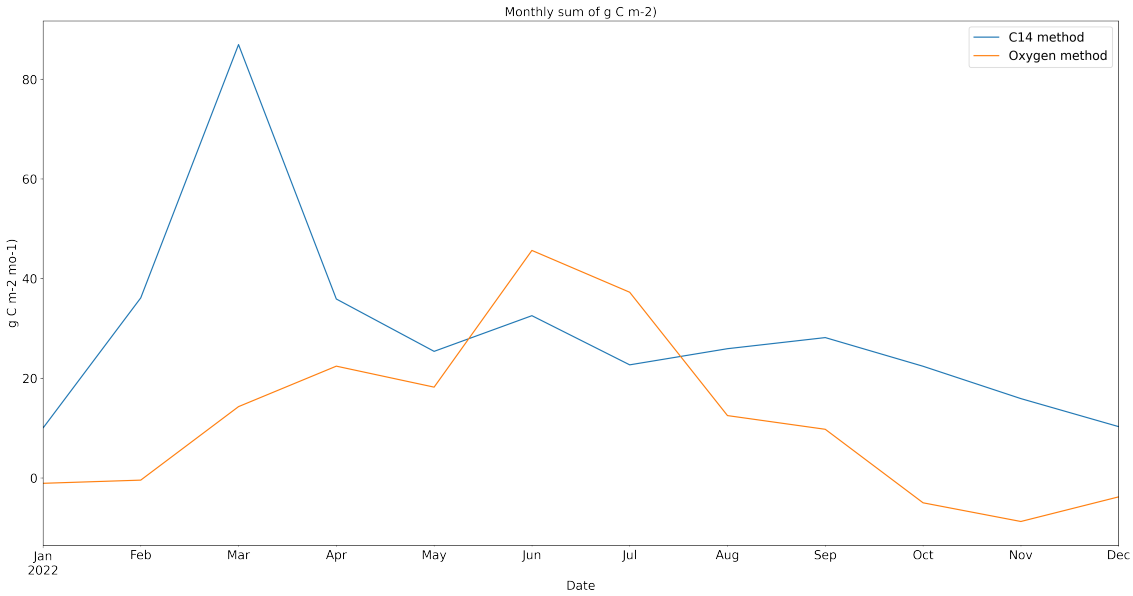

Figure 6.9: Comparison of the total monthly Carbon per square meter calcualted using the two method considered.

6.4 Discussion

Although the total value calculated by the analysis of Oxygen variations is lower by about 200 $gCm^{-2}d^{-1}$ from the comparison with the C14 method, the results are promising. This discrepancy is justified by three main factors. For primary production estimated using in-vitro method, it is necessary to expose light and dark bottles for a limited time. In the case of Lake Lugano, the exposure lasts 4 hours. In this time frame what is measured is somewhere between GPP and NEP [109]. Hence, that can explain the higher values of this method compared to the other. In addition, the AHFM system has a much higher resolution by having values with hourly frequency. In this case, both the process of oxygen production and respiration are monitored in more detail which means the detection of other minor phenomena of minor productivity that can affect the values. The second factor to consider is related to the fact that photosynthesis carried out by algal organisms is basically the main

process measured inside the bottles. This factor should not be underestimated because the in-situ measurements, by means of the sensor chain, are located in an open system where in addition to the photosynthetic activity of the algae there is the influence of all the biological community within the lake. Moreover there are additional effects related to chemical and physical processes. Hence, it is necessary to emphasize that while apparently the process monitored is the same for the two methodologies, namely photosynthesis, it is necessary to consider all the factors involved and the important difference in data acquisition and processing between the two methods. Moreover, to better understand and validate the trend described by the in-situ method results it is important to make a comparison with the concentrations of Chlorophyll-a and Phycocyanin using the profiles plotted in Figure 6.10. Such data are collected every two weeks using the Trilux by Chealsea Technologies sensor. The Chlorophyll-a plot shows three main events with different magnitude. The first is in March, the second in May and the last one in June and July. These events corresponds to cynobacteria blooms, but in this case, the Phycocyan concentration growns also in Autumn with higher values until the depth of 15 m. By comparing this trend with the graph in Figure 6.8, some correlations can be found between the trend of GPP and the Phycocyn concentrations. In fact, events E1, E2, E3 and E4 explain well the trend of productivity during these periods both going in terms of the time interval and along the depths. The pattern of the E5 event, which has different characteristics, can be exaplined by looking to both the plots about the concentrations of Chlorophyll and Phycocyanin. In fact, in that period small increment of Chlorophyll is followed by a cyanobacteria bloom. Concerning E5 and E6 events, it is important to consider that during those considered periods, mixing phenomena began and they gradually characterized the entire water column considered. This trend is evident by looking at the temperature profiles showed in Figure 6.4.

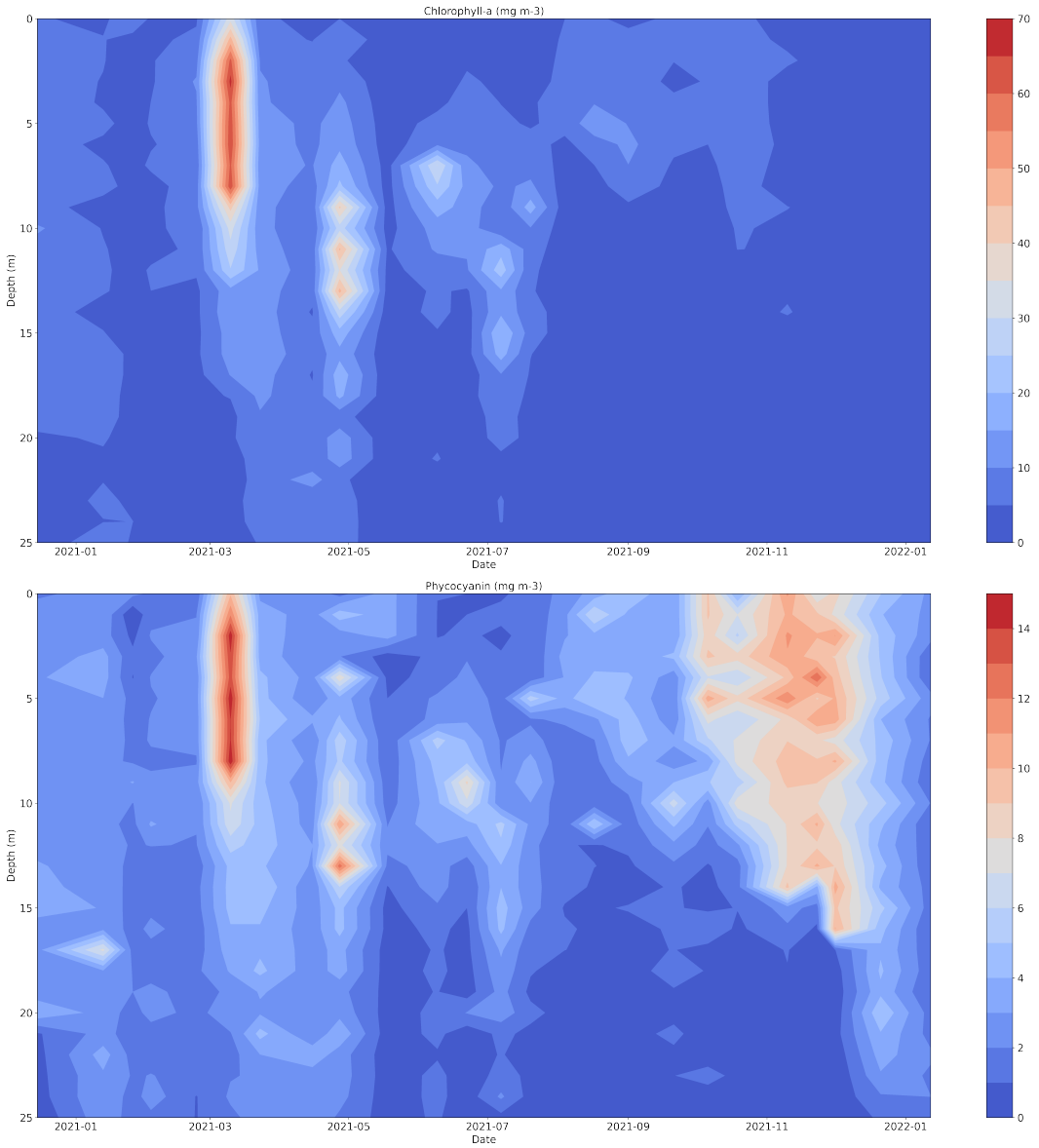

Figure 6.10: Profiles concentrations of Chlorophyll-a (upper plot) and Phycocyanin (bottom plot) in mg m-3.

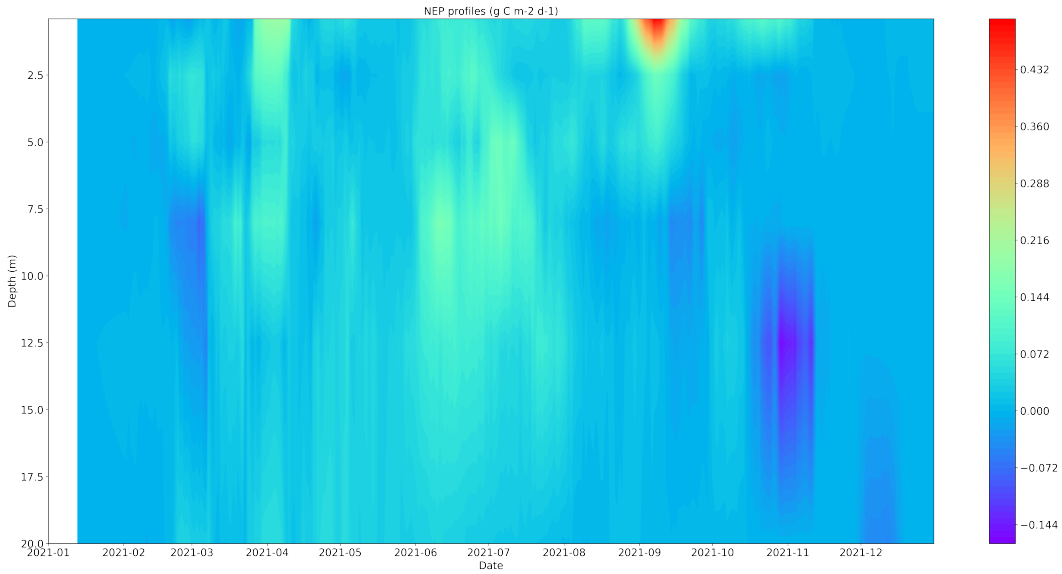

Figure 6.11: Profile concentrations of net primary productivity calculated by means of the algorithm developed within this project.

6.5 Conclusions

The presented study is focused on the Lake Lugano which has been strongly affected by eutrophication processes caused by human activities and thus mainly by excessive phosphorus and nitrogen loads. However, the effects of climate changes are also affecting the process of upwelling that began during the 80s. The Primary Production (PP) is a key process to monitor in order to estimate the health of a lake. The estimation of PP can be done through different methodologies. Two were considered within this study. The first calculates the metabolism of the lake based on the Carbon-14 technique [46, 92]. The second uses a methodology based on oxygen variations within a given time period [48]. For this last method, the data used comes from a high-frequency automatic system consisting of six sensors measuring the parameters of dissolved oxygen in water and temperature [95] positioned at the depth of 0.4, 2.5, 5.0, 8.0, 12.5 and 20.0 m. The in vitro method is based on the fixation of radioactive Carbon performed by algal organism through photosynthesis. Such method is based on the position of dark and light bottles at the depths of 0.4, 1.25, 2.5, 3.75, 5, 6.5, 8, 10, 12.5, 15, and 20 m exposed for 4 hours. This sampling locaiton is nearby to the AHFM system, 10 m maximum distance between one and the other. Within this work, the implementation of an algorithm based on the in-situ methodology was presented as well as the

comparison and validation process. The results obtained are very promising, since the productivity calculated by the new algorithm well correspond to the main bloom events happened during the year and are fully comparable in terms of the total grams of Carbon per square meter over the year, eventhough there are substancial methodological differences. It is also true, that this is a first step towards a fully validation of the new algorithm developed since critical aspects still exist and need to be further investigate, such as the greater influence that cyanobacteria seems to have over the DO measurements as well as the effects of fast and locally concentrated algal blooms that could potentially explain the peak of net productivity measured in September (see Figure 6.11). However, this leads to also very positive aspects such as the possibility to better study and discover unknown effects which can limit the lake recovering process. Once the algorithm and the method is fully validated, the developed algorithm, written using the Python programming language, can be used for real-time estimation of lake productivity thanks to the possibility to automatize the elaboration. This will potentially allow decision-makers to have a daily updated indicator to assess the trend of the trophic state of the lake and take new actions.

6.6 Acknowledment

Part of the data used in this study were collected as part of a research programme promoted by the International Commission for the Protection of Italian-Swiss Waters (CIPAIS). We gratefully acknowledge all the personnel that contributed to the monitoring program of Lake Lugano during the study period.

Temperature	
Principle	NTC
Measuring range	-50 to 60 °C
Unit	°C
Accuracy	±0.2 °C (-20 to 50 °C)
Relative humidity	
Principle	Capacitive
Measuring range	0 to 100 % RH
Unit	% RH
Accuracy	±2 % RH
Air pressure	
Measuring range	300 to 1200 hPa
Unit	hPa
Accuracy	±0.5 hPa (0 to 40 °C)
Wind direction	
Principle	Ultrasonic
Measuring range	0 to 359.9 °
Unit	°
Accuracy	< 3° RMSE > 1.0 m/s
Resolution	0.1
Wind speed	
Principle	Ultrasonic
Measuring range	0 to 75 m/s
Unit	m/s
Accuracy	±0.3 m/s or ±3 % (0 to 35 m/s) RMS
Resolution	0.1 m/s
Radiation	
Spectral range	300...2800 nm
Measuring range	2000 W/m²

Table 6.4: LUFFT WS-501 weather station technical specifications.

7 Conclusions

The Ph.D work was targeted at providing a fully open solution that can cover the entire pipeline involved in lake monitoring, from integrating new and different data sources to a software data management that can offer new opportunities for time series elaborations such as the developing of real-time estimation of Primary Production. The study was applied on a pilot study area, Lake Lugano, which is a deep sub-alpine lake stressed by human activities and climate changes effects, currently recovering from eutrophication. Related long time series data are available, being Lake Lugano monitored since 70s with campaigns performed fortnightly. The proposed solution stems from two issues: the first relates to the need to adopt new technologies for monitoring the lake ecosystem in order to identify, analyze, and study new dynamics that may influence the lake restoration process; the second issue relates to the lack of an integrated system to manage all available data on the lake ecosystem, which due to this issue are only partially exploited by researchers, decision makers, and stakeholders. Considering the problems highlighted, the Ph.D work was based on two main research questions:

1. Can the proposed open solution answer to these needs and be used in an applied operative monitoring at scale?

2. How might this system help in tackling the local effects of climate changes?

To this end, this thesis starts with an overview on the regulatory framework of the lake which was presented and described in Chapter 2. In fact, Lake Lugano is a transboundary lake between Italy and Switzerland, so it is regulated by two different national systems. In this chapter it was highlighted the importance of the CIPIAS commission, funded by the two nations in order to provide and guarantee a continuous monitoring system aimed at safeguarding the health of the lake.

Chapter 3 presents the first results obtained based on the design, development and implementation of an integrated software architecture for data management. The work that was carried out has shown how open source software tools on the market can be integrated through the use of new technologies that offer high replicability and scalability, guaranteeing some important aspects such as: security of the data; uniformity of the format in which the data are stored; data sharing through internationally recognised standard

(SOS-OGC); extensibility, by easily adding new functionalities thanks to the possibility of adding plugins. The system required the development of several components that allowed the integration of the various services selected and the creation of additional functions that can process time series to create new indicators that are useful for monitoring the health status of the lake ecosystem, in quasi real-time. Thanks to the international standard which was used to share information, the data archived using the software developed within this work can be easily harvested in other greater standard platforms (e.g. DataLakes portal https://www.datalakes-eawag.ch). In fact, the aim of this part of the work was to improve the Findability, Accessibility, Interoperability, and Reuse principles (FAIR, https://www.go-fair.org/fair-principles/) of the digital assets related to lake monitoring, and foster the digitalization. In this context, additional results were reached such as the software replication and installation on the server of the Water Research Institute of the National Research Council CNR IRSA of Pallanza, in order to manage data of other lakes e.g. Maggiore and Como.

In order to provide a solution to improve the adoption of new technologies in lake ecosystem monitoring, in Chapter 4 an open and cost effective AHFM system for lake water quality monitoring was presented. Such solution is composed by a photovoltaic system to power a core unit composed by a raspberry and a set of other hardware components (sensor units, voltage regulators, communication unit, etc.). The sensor unit is composed by a chain of six OPTOD by Ponsel sensors that measure water temperature and dissolved oxygen. The solution was tested and validated along more than one year after deploying it on a platform of 2.5 x 2.5 m on Lake Lugano. The validation process showed promising results with 99% of data completeness, very high quality of the data collected and a good relation between the time series and the data collected during traditional campaigns. In addition, the solution is SOS standard enabled since, in accordance with the edge computing paradigm, some elaboration processes were moved on the edge, so that data are archived using the standard directly in the AHFM system deployed in the middle of Lake Lugano.

Chapter 5 presents the main achievements of two years of research aimed a standardizing and homogenizing data from Lake Lugano. The system facilitates the use of data by implementing a web-based tool to import historical time series and calculate and display a set of key lake indicators using open source software and standard. More than 26 *procedures* for a total of 22

observed properties and more than 3 million *observations* has been collected.

Finally, the whole system has been tested in order to provide new frontiers in developing algorithm and elaborations that can have another applied impact on the lake monitoring. To this end, in Chapter 6 the estimation of Primary Production was calculated by means of the data collected from the AHFM system applying the Staehr method. Thanks to the chain of sensors installed, it was possible to use the variations in the dissolved oxygen in water to derive the Gross and Net Production in Lake Lugano. Such algorithm was developed using Python so that in the future it could be potentially applied in quasi real-time estimation of lake productivity. The results of this part of work are very promising since a good correlation was found between the Chlorophyll concentration and the results highlighted using the traditional method still used to evaluate the trophic status of the lake.

According to the presented results, there are many hints for planning possible future developments of this work. Leaving aside possible hardware and software improvements related to the products created during this project, there are three main points I would like to focus on for possible developments:

1. implement a standard system for real-time data quality control of both the AHFM system and measurement campaign data. Even though the quality of data was considered during this project, it highlighted the need of a system which could better suit the needs of the parameters concerning the lake water monitoring. The creation of new rules and tests to check the quality of data, although preliminary studied in Tiberti et al. 2021 [11], must be further improved and integrated in the software system for an automatic data validation;

2. envisage a miniaturisation of the high-frequency monitoring system in order to further reduce production costs and also improve spatial resolution. In particular, a miniaturised solution to monitor Chlorophyll and Phycocyanin could cover the needs of real-time algal bloom estimation, which is of increasing interest both for bathing and for water distribution companies, which increasingly build lake caption systems to provide water to households.

3. proceed with the validation of the new algorithm produced for the PP estimation by using the data collected during the year 2022. This further step of validation and calibration may bring real advantages in the future,

for the estimation of primary production as it can be easily integrated into the platform for a near real-time evaluation of lake produtivity trends.

Bibliography

(1) Du Plessis, A. In *Water as an Inescapable Risk: Current Global Water Availability, Quality and Risks with a Specific Focus on South Africa*, du Plessis, A., Ed.; Springer Water; Springer International Publishing: Cham, 2019, pp 13–25.

(2) Wang, D.; Hubacek, K.; Shan, Y.; Gerbens-Leenes, W.; Liu, J. A Review of Water Stress and Water Footprint Accounting. *Water* **2021**, *13*, 201.

(3) Woolway, R. I.; Kraemer, B. M.; Lenters, J. D.; Merchant, C. J.; O'Reilly, C. M.; Sharma, S. Global Lake Responses to Climate Change. *Nature Reviews Earth & Environment* **2020**, *1*, 388–403.

(4) Singh, S. P.; Singh, P. Effect of Temperature and Light on the Growth of Algae Species: A Review. *Renewable and Sustainable Energy Reviews* **2015**, *50*, 431–444.

(5) Straile, D.; Jöhnk, K.; Henno, R. Complex Effects of Winter Warming on the Physicochemical Characteristics of a Deep Lake. *Limnology and Oceanography* **2003**, *48*, 1432–1438.

(6) Wells, M. L.; Trainer, V. L.; Smayda, T. J.; Karlson, B. S. O.; Trick, C. G.; Kudela, R. M.; Ishikawa, A.; Bernard, S.; Wulff, A.; Anderson, D. M.; Cochlan, W. P. Harmful Algal Blooms and Climate Change: Learning from the Past and Present to Forecast the Future. *Harmful Algae* **2015**, *49*, 68–93.

(7) Jansson, P.; Hock, R.; Schneider, T. The Concept of Glacier Storage: A Review. *Journal of Hydrology* **2003**, *282*, 116–129.

(8) Lepori, F.; Roberts, J. J.; Schmidt, T. S. A Paradox of Warming in a Deep Peri-Alpine Lake (Lake Lugano, Switzerland and Italy). *Hydrobiologia* **2018**, *824*, 215–228.

(9) Ezeora, O. S.; Heckenbergerova, J.; Musilek, P.; Rodway, J. In *2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, 2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE), 2016, pp 1–4.

(10) Wüest, A.; Bouffard, D.; Guillard, J.; Ibelings, B. W.; Lavanchy, S.; Perga, M.-E.; Pasche, N. LéXPLORE: A Floating Laboratory on Lake Geneva Offering Unique Lake Research Opportunities. *WIREs Water* **2021**, *8*, e1544.

(11) Tiberti, R.; Caroni, R.; Cannata, M.; Lami, A.; Manca, D.; Strigaro, D.; Rogora, M. Automated High Frequency Monitoring of Lake Maggiore through in Situ Sensors: System Design, Field Test and Data Quality Control. *Journal of Limnology* **2021**, *80*, DOI: 10.4081/jlimnol.2021.2011.

(12) Olmanson, L. G.; Brezonik, P. L.; Bauer, M. E. In *Advances in Watershed Science and Assessment*, Younos, T., Parece, T. E., Eds.; The Handbook of Environmental Chemistry; Springer International Publishing: Cham, 2015, pp 111–140.

(13) Yang, C.; Huang, Q.; Li, Z.; Liu, K.; Hu, F. Big Data and Cloud Computing: Innovation Opportunities and Challenges. *International Journal of Digital Earth* **2017**, *10*, 13–53.

(14) Enguix, I. F.; Ruiz, P.; de la Torre, M. Online Digitalization Technologies for Monitoring Activities in the Marine Environment. *Proceedings* **2019**, *42*, 59.

(15) Marcon, É.; Marcon, A.; Le Dain, M.-A.; Ayala, N. F.; Frank, A. G.; Matthieu, J. Barriers for the Digitalization of Servitization. *Procedia CIRP* **2019**, *83*, 254–259.

(16) Van Kessel, R.; Wong, B. L. H.; Rubinić, I.; O'Nuallain, E.; Czabanowska, K. Is Europe Prepared to Go Digital? Making the Case for Developing Digital Capacity: An Exploratory Analysis of Eurostat Survey Data. *PLOS Digital Health* **2022**, *1*, e0000013.

(17) Lawrence, P. R. How to Deal With Resistance to Change. *Harvard Business Review* **1969**.

(18) Bidoglio, G.; Vanham, D.; Bouraoui, F.; Barchiesi, S. In *Encyclopedia of Ecology (Second Edition)*, Fath, B., Ed.; Elsevier: Oxford, 2019, pp 459–466.

(19) Cronache, R. Siccità, sul Po la più grave crisi degli ultimi 70 anni. Terna: «Poca acqua per raffreddare le centrali» Corriere della Sera, https://www.corriere.it/cronache/22_giugno_10/siccita-emergenza-bacino-po-la-piu-grave-crisi-ultimi-70-anni-0de23f50-e8b8-11ec-a288-5db7a6019886.shtml (accessed 06/13/2022).

(20) Ho, L. T.; Goethals, P. L. M. Opportunities and Challenges for the Sustainability of Lakes and Reservoirs in Relation to the Sustainable Development Goals (SDGs). *Water* **2019**, *11*, 1462.

(21) Lepori, F.; Capelli, C. Effects of Phosphorus Control on Primary Productivity and Deep-Water Oxygenation: Insights from Lake Lugano (Switzerland and Italy). *Hydrobiologia* **2021**, *848*, 613–629.

(22) Schindler, D. W. Recent Advances in the Understanding and Management of Eutrophication. *Limnology and Oceanography* **2006**, *51*, 356–363.

(23) Caspers, H. OECD: Eutrophication of Waters. Monitoring, Assessment and Control. — 154 Pp. Paris: Organisation for Economic Co-Operation and Development 1982. (Publié En Français Sous Le Titre »Eutrophication Des Eaux. Méthodes de Surveillance, d'Evaluation et de Lutte«). *Internationale Revue der gesamten Hydrobiologie und Hydrographie* **1984**, *69*, 200–200.

(24) Lepori, F.; Bartosiewicz, M.; Simona, M.; Veronesi, M. Effects of Winter Weather and Mixing Regime on the Restoration of a Deep Perialpine Lake (Lake Lugano, Switzerland and Italy). *Hydrobiologia* **2018**, *824*, 229–242.

(25) Xu, L. D.; Xu, E. L.; Li, L. Industry 4.0: State of the Art and Future Trends. *International Journal of Production Research* **2018**, *56*, 2941–2962.

(26) Zambon, I.; Cecchini, M.; Egidi, G.; Saporito, M. G.; Colantoni, A. Revolution 4.0: Industry vs. Agriculture in a Future Development for SMEs. *Processes* **2019**, *7*, 36.

(27) Ghobakhloo, M. Industry 4.0, Digitization, and Opportunities for Sustainability. *Journal of Cleaner Production* **2020**, *252*, 119869.

(28) Pozzoni, M.; Salvetti, A.; Cannata, M. Retrospective and Prospective of Hydro-Met Monitoring System in the Canton Ticino, Switzerland. *Hydrological Sciences Journal* **2020**, *0*, 1–15.

(29) Hill, A. P.; Prince, P.; Snaddon, J. L.; Doncaster, C. P.; Rogers, A. AudioMoth: A Low-Cost Acoustic Device for Monitoring Biodiversity and the Environment. *HardwareX* **2019**, *6*, e00073.

(30) Feenstra, B.; Papapostolou, V.; Hasheminassab, S.; Zhang, H.; Boghossian, B. D.; Cocker, D.; Polidori, A. Performance Evaluation of Twelve Low-Cost PM2.5 Sensors at an Ambient Air Monitoring Site. *Atmospheric Environment* **2019**, *216*, 116946.

(31) Ambrož, M.; Hudomalj, U.; Marinšek, A.; Kamnik, R. Raspberry Pi-Based Low-Cost Connected Device for Assessing Road Surface Friction. *Electronics* **2019**, *8*, 341.

(32) Alam, A. U.; Clyne, D.; Jin, H.; Hu, N.-X.; Deen, M. J. Fully Integrated, Simple, and Low-Cost Electrochemical Sensor Array for in Situ Water Quality Monitoring. *ACS Sensors* **2020**, *5*, 412–422.

(33) Sunny, A. I.; Zhao, A.; Li, L.; Kanteh Sakiliba, S. Low-Cost IoT-Based Sensor System: A Case Study on Harsh Environmental Monitoring. *Sensors* **2021**, *21*, 214.

(34) Strigaro, D.; Cannata, M.; Antonovic, M. Boosting a Weather Monitoring System in Low Income Economies Using Open and Non-Conventional Systems: Data Quality Analysis. *Sensors* **2019**, *19*, 1185.

(35) Cannata, M.; Strigaro, D.; Lepori, F.; Capelli, C.; Rogora, M.; Manca, D. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, ISPRS WG IV/4, WG IV/9 & WG V/8
FOSS4G 2021 – Academic Track - 27 September–2 October 2021, Buenos Aires, Argentina, Copernicus GmbH: 2021; Vol. XLVI-4-W2-2021, pp 25–29.

(36) IST-SUPSI, I. s. d. T. Ricerche Sull'evoluzione Del Lago Di Lugano. Aspetti Limnologici. Programma Triennale 2019-2021. Campagna 2020. *Commissione Internazionale per la Protezione delle Acque Italo-Svizzere (Ed.)* **2021**, 60.

(37) Seifert-Dähnn, I.; Furuseth, I. S.; Vondolia, G. K.; Gal, G.; de Eyto, E.; Jennings, E.; Pierson, D. Costs and Benefits of Automated High-Frequency Environmental Monitoring – The Case of Lake Water Management. *Journal of Environmental Management* **2021**, *285*, 112108.

(38) Laas, A.; de Eyto, E.; Pierson, D. C.; Jennings, E. NETLAKE Guidelines for Automated Monitoring System Development. **2016**.

(39) Sankaran, M. The Future of Digital Water Quality Monitoring Water Quality Products, `https://www.wqpmag.com/web-exclusive/future-digital-water-quality-monitoring` (accessed 08/24/2021).

(40) Dutta, S.; Geiger, T.; Lanvin, B. In *World Economic Forum*, 2015; Vol. 1, P80–85.

(41) Sabbagh, K.; Friedrich, R.; El-Darwiche, B.; Singh, M.; Ganediwalla, S.; Katz, R. Maximizing the Impact of Digitization. *The global information technology report* **2012**, *2012*, 121–133.

(42) Jones, M. D.; Hutcheson, S.; Camba, J. D. Past, Present, and Future Barriers to Digital Transformation in Manufacturing: A Review. *Journal of Manufacturing Systems* **2021**, *60*, 936–948.

(43) Gibert, K.; Horsburgh, J. S.; Athanasiadis, I. N.; Holmes, G. Environmental Data Science. *Environmental Modelling & Software* **2018**, *106*, 4–12.

(44) Sima, V.; Gheorghe, I. G.; Subić, J.; Nancu, D. Influences of the Industry 4.0 Revolution on the Human Capital Development and Consumer Behavior: A Systematic Review. *Sustainability* **2020**, *12*, 4035.

(45) Desai, J. R. Internet of Things: Architecture, Protocols, and Interoperability as a Major Challenge Handbook of Research on Implementation and Deployment of IoT Projects in Smart Cities, `https://www.igi-global.com/chapter/internet-of-things/www.igi-global.com/chapter/internet-of-things/233262` (accessed 12/13/2021).

(46) Nielsen, E. S. The Use of Radio-active Carbon (C14) for Measuring Organic Production in the Sea. *ICES Journal of Marine Science* **1952**, *18*, 117–140.

(47) Schindler, D. W.; Schmidt, R. V.; Reid, R. A. Acidification and Bubbling as an Alternative to Filtration in Determining Phytoplankton Production by the 14C Method. *Journal of the Fisheries Research Board of Canada* **1972**, *29*, 1627–1631.

(48) Staehr, P. A.; Bade, D.; de Bogert, M. C. V.; Koch, G. R.; Williamson, C.; Hanson, P.; Cole, J. J.; Kratz, T. Lake Metabolism and the Diel Oxygen Technique: State of the Science. *Limnology and Oceanography: Methods* **2010**, *8*, 628–644.

(49) OECD, *Eutrophication of Waters : Monitoring, Assessment and Control*; Organisation for Economic Co-operation and Development ; Sold by OECD Publications and Information Center] Paris : [Washington, D.C: 1982, 154 p. :

(50) Papazoglou, M.; Van Den Heuvel, W.-J. Service Oriented Architectures: Approaches, Technologies and Research Issues. *VLDB Journal* **2007**, *16*, 389–415.

(51) Aalst, W. M. V. D.; Beisiegel, M.; Hee, K. M. V.; Konig, D.; Stahl, C. An SOA-based Architecture Framework. *International Journal of Business Process Integration and Management* **2007**, *2*, 91.

(52) Granell, C.; Díaz, L.; Gould, M. Service-Oriented Applications for Environmental Models: Reusable Geospatial Services. *Environmental Modelling & Software* **2010**, *25*, 182–198.

(53) Morabito, R.; Cozzolino, V.; Ding, A. Y.; Beijar, N.; Ott, J. Consolidate IoT Edge Computing with Lightweight Virtualization. *IEEE Network* **2018**, *32*, 102–111.

(54) Chen, B.; Wan, J.; Celesti, A.; Li, D.; Abbas, H.; Zhang, Q. Edge Computing in IoT-Based Manufacturing. *IEEE Communications Magazine* **2018**, *56*, 103–109.

(55) Reynolds, C. S., *The Ecology of Phytoplankton*; Cambridge University Press: 2006; 437 pp.

(56) Cannata, M.; Antonovic, M.; Molinari, M.; Pozzoni, M. istSOS, a New Sensor Observation Management System: Software Architecture and a Real-Case Application for Flood Protection. *Geomatics, Natural Hazards and Risk* **2015**, *6*, 635–650.

(57) Bröring, A.; Stasch, C.; Echterhoff, J. *OGC Sensor Observation Service Interface Standard, Version 2.0.* Report; Open Geospatial Consortium, 2012.

(58) Salvati, M.; Brambilla, E.; Grafiche Futura, *Data Quality Control Procedures in Alpine Metereological Services*; Università di Trento. Dipartimento di ingegneria civile e ambientale: Trento, 2008.

(59) Zahumenskỳ, I. Guidelines on Quality Control Procedures for Data from Automatic Weather Stations. *World Meteorological Organization, Switzerland* **2004**.

(60) Shafer, M. A.; Fiebrich, C. A.; Arndt, D. S.; Fredrickson, S. E.; Hughes, T. W. Quality Assurance Procedures in the Oklahoma Mesonetwork. *Journal of Atmospheric and Oceanic Technology* **2000**, *17*, 474–494.

(61) Kunkel, K. E.; Andsager, K.; Conner, G.; Decker, W. L.; Hillaker, H. J.; Knox, P. N.; Nurnberger, F. V.; Rogers, J. C.; Scheeringa, K.; Wendland, W. M.; Zandlo, J.; Angel, J. R. An Expanded Digital Daily Database for Climatic Resources Applications in the Midwestern United States. *Bulletin of the American Meteorological Society* **1998**, *79*, 1357–1366.

(62) Schroeder, J. L.; Burgett, W. S.; Haynie, K. B.; Sonmez, I.; Skwira, G. D.; Doggett, A. L.; Lipe, J. W. The West Texas Mesonet: A Technical Overview. *Journal of Atmospheric and Oceanic Technology* **2005**, *22*, 211–222.

(63) Meek, D. W.; Hatfield, J. L. Data Quality Checking for Single Station Meteorological Databases. *Agricultural and Forest Meteorology* **1994**, *69*, 85–109.

(64) Bergkemper, V.; Weisse, T. Do Current European Lake Monitoring Programmes Reliably Estimate Phytoplankton Community Changes? *Hydrobiologia* **2018**, *824*, 143–162.

(65) Organization (WMO), W. M.; (WMO), W. M. O., *State of the Global Climate 2021 : WMO Provisional Report*; WMO; WMO: Geneva, 2021; 47 p.

(66) Spinoni, J.; Vogt, J. V.; Naumann, G.; Barbosa, P.; Dosio, A. Will Drought Events Become More Frequent and Severe in Europe? *International Journal of Climatology* **2018**, *38*, 1718–1736.

(67) Guo, H.; Bao, A.; Liu, T.; Jiapaer, G.; Ndayisaba, F.; Jiang, L.; Kurban, A.; De Maeyer, P. Spatial and Temporal Characteristics of Droughts in Central Asia during 1966–2015. *Science of The Total Environment* **2018**, *624*, 1523–1538.

(68) Brunner, M. I.; Björnsen Gurung, A.; Zappa, M.; Zekollari, H.; Farinotti, D.; Stähli, M. Present and Future Water Scarcity in Switzerland: Potential for Alleviation through Reservoirs and Lakes. *Science of The Total Environment* **2019**, *666*, 1033–1047.

(69) Marcé, R. et al. Automatic High Frequency Monitoring for Improved Lake and Reservoir Management. *Environmental Science & Technology* **2016**, *50*, 10780–10794.

(70) Fenocchi, A.; Rogora, M.; Sibilla, S.; Ciampittiello, M.; Dresti, C. 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* **2018**, *51*, 3521–3536.

(71) Free, G.; Bresciani, M.; Pinardi, M.; Ghirardi, N.; Luciani, G.; Caroni, R.; Giardino, C. Detecting Climate Driven Changes in Chlorophyll-a in Deep Subalpine Lakes Using Long Term Satellite Data. *Water* **2021**, *13*, 866.

(72) Dörnhöfer, K.; Oppelt, N. Remote Sensing for Lake Research and Monitoring – Recent Advances. *Ecological Indicators* **2016**, *64*, 105–122.

(73) Sagan, V.; Peterson, K. T.; Maimaitijiang, M.; Sidike, P.; Sloan, J.; Greeling, B. A.; Maalouf, S.; Adams, C. Monitoring Inland Water Quality Using Remote Sensing: Potential and Limitations of Spectral Indices, Bio-Optical Simulations, Machine Learning, and Cloud Computing. *Earth-Science Reviews* **2020**, *205*, 103187.

(74) Kumar, S.; Tiwari, P.; Zymbler, M. Internet of Things Is a Revolutionary Approach for Future Technology Enhancement: A Review. *Journal of Big Data* **2019**, *6*, 111.

(75) Giuliani, G.; Masó, J.; Mazzetti, P.; Nativi, S.; Zabala, A. Paving the Way to Increased Interoperability of Earth Observations Data Cubes. *Data* **2019**, *4*, 113.

(76) Liang, S. H. L.; Saeedi, S.; Ojagh, S.; Honarparvar, S.; Kiaei, S.; Mohammadi Jahromi, M.; Squires, J. An Interoperable Architecture for the Internet of COVID-19 Things (IoCT) Using Open Geospatial Standards—Case Study: Workplace Reopening. *Sensors* **2021**, *21*, 50.

(77) OGC SensorThings API Part 1: Sensing, Version 1.0. `https://repository.oceanbestpractices.org/handle/11329/914` (accessed 05/19/2022).

(78) Read, J. S.; Hamilton, D. P.; Jones, I. D.; Muraoka, K.; Winslow, L. A.; Kroiss, R.; Wu, C. H.; Gaiser, E. Derivation of Lake Mixing and Stratification Indices from High-Resolution Lake Buoy Data. *Environmental Modelling & Software* **2011**, *26*, 1325–1336.

(79) Schmidt, W. Über Die Temperatur- Und Stabili-Tätsverhältnisse Von Seen. *Geografiska Annaler* **1928**, *10*, 145–177.

(80) Hutchinson, G. E., *A Treatise on Limnology. Vol 1: Georgraphy, Physics and Chemistry*; John Wiley & Sons: 1957.

(81) Idso, S. B. On the Concept of Lake Stability 1. *Limnology and oceanography* **1973**, *18*, 681–683.

(82) Martin, J.; McCutcheon, S. Hydrodynamics and Transport for Water Quality Modeling, Lewis Publishers. *New York* **1999**, 1–794.

(83) Millero, F. J.; Poisson, A. International One-Atmosphere Equation of State of Seawater. *Deep Sea Research Part A. Oceanographic Research Papers* **1981**, *28*, 625–629.

(84) Wetzel, R. G., *Limnology: Lake and River Ecosystems*; Gulf Professional Publishing: 2001; 1023 pp.

(85) Smith, V. H.; Tilman, G. D.; Nekola, J. C. Eutrophication: Impacts of Excess Nutrient Inputs on Freshwater, Marine, and Terrestrial Ecosystems. *Environmental Pollution* **1999**, *100*, 179–196.

(86) Manca, M.; Torretta, B.; Comoli, P.; Amsinck, S. L.; Jeppesen, E. Major Changes in Trophic Dynamics in Large, Deep Sub-Alpine Lake Maggiore from 1940s to 2002: A High Resolution Comparative Palaeo-Neolimnological Study. *Freshwater Biology* **2007**, *52*, 2256–2269.

(87) Salmaso, N.; Mosello, R. Limnological Research in the Deep Southern Subalpine Lakes: Synthesis, Directions and Perspectives. *Advances in Oceanography and Limnology* **2010**, *1*, 29–66.

(88) Barbieri, A.; Mosello, R. Chemistry and Trophic Evolution of Lake Lugano in Relation to Nutrient Budget. *Aquatic Sciences* **1992**, *54*, 219–237.

(89) *Eutrophication: Causes, Consequences, Correctives*; The National Academies Press: Washington, DC, 1969.

(90) Odum, E. P. The Strategy of Ecosystem Development. *Science* **1969**, *164*, 262–270.

(91) CARLSON, R. In *Enhancing the State's Lake Management Programs: Proceedings of Conference, Chicago, 1991*, 1991, pp 59–71.

(92) Hama, T.; Miyazaki, T.; Ogawa, Y.; Iwakuma, T.; Takahashi, M.; Otsuki, A.; Ichimura, S. Measurement of Photosynthetic Production of a Marine Phytoplankton Population Using a Stable 13C Isotope. *Marine Biology* **1983**, *73*, 31–36.

(93) Gächter, R.; Mares, A.; tilzer, M. M. Determination of Phytoplankton Production by the Radiocarbon Method: A Comparison between the Acidification and Bubbling Method (ABM) and the Filtration Techn1que *. *Journal of Plankton Research* **1984**, *6*, 359–364.

(94) Franchini, F.; Lepori, F.; Bruder, A. Improving Estimates of Primary Production in Lakes: A Test and a Case Study from a Peri-Alpine Lake (Lake Lugano). *Inland Waters* **2017**, *7*, 77–87.

(95) Strigaro, D.; Cannata, M.; Lepori, F.; Capelli, C.; Lami, A.; Manca, D.; Seno, S. Open and Cost-Effective Digital Ecosystem for Lake Water Quality Monitoring. *Sensors* **2022**, *22*, 6684.

(96) Mosello, R.; Ambrosetti, W.; Arisci, S.; Bettinetti, R.; Buzzi, F.; Calderoni, A.; Carrara, E.; De Bernardi, R.; Galassi, S.; Garibaldi, L., et al. Evoluzione Recente Della Qualità Delle Acque Dei Laghi Profondi Sudalpini (Maggiore, Lugano, Como, Iseo e Garda) in Risposta Alle Pressioni Antropiche e Alle Variazioni Climatiche. **2010**.

(97) Rothe, M.; Frederichs, T.; Eder, M.; Kleeberg, A.; Hupfer, M. Evidence for Vivianite Formation and Its Contribution to Long-Term Phosphorus Retention in a Recent Lake Sediment: A Novel Analytical Approach. *Biogeosciences* **2014**, *11*, 5169–5180.

(98) Worsfold, P.; McKelvie, I.; Monbet, P. Determination of Phosphorus in Natural Waters: A Historical Review. *Analytica Chimica Acta* **2016**, *918*, 8–20.

(99) Tu, L.; Jarosch, K. A.; Schneider, T.; Grosjean, M. Phosphorus Fractions in Sediments and Their Relevance for Historical Lake Eutrophication in the Ponte Tresa Basin (Lake Lugano, Switzerland) since 1959. *Science of The Total Environment* **2019**, *685*, 806–817.

(100) Rotta, F.; Capelli, C.; Lepori, F., *Microplastic Variability in Subsurface Waters of a Deep Perialpine Lake (Lake Lugano, CH-IT)*, 2022.

(101) Lepori, F.; Capelli, C.; Foresti, D. Changes in Phytoplankton Composition Hinder the Recovery from Eutrophication in a Perialpine Lake (Lake Lugano, Switzerland and Italy). *Journal of Plankton Research* **2022**, *44*, 22–35.

(102) Lepori, F.; Capelli, C. Seasonal Variation in Trophic Structure and Restoration Effects in a Deep Perialpine Lake (Lake Lugano, Switzerland and Italy). *Journal of Great Lakes Research* **2020**, *46*, 870–880.

(103) Strigaro, D.; Cannata, M.; Lepori, F.; Capelli, C.; Lami, A.; Manca, D.; Seno, S. Open and Cost-Effective Digital Ecosystem for Lake Water Quality Monitoring. *Sensors* **2022**, *22*, 6684.

(104) Odum, H. T. Primary Production in Flowing Waters1. *Limnology and Oceanography* **1956**, *1*, 102–117.

(105) USGS WQ1981.11 WATER QUALITY–New Tables of Dissolved Oxygen Saturation Values, 1981.

(106) Wanninkhof, R. Relationship between Wind Speed and Gas Exchange over the Ocean. *Journal of Geophysical Research: Oceans* **1992**, *97*, 7373–7382.

(107) Jähne, B.; Münnich, K. O.; Bösinger, R.; Dutzi, A.; Huber, W.; Libner, P. On the Parameters Influencing Air-Water Gas Exchange. *Journal of Geophysical Research: Oceans* **1987**, *92*, 1937–1949.

(108) Cole, J. J.; Pace, M. L.; Carpenter, S. R.; Kitchell, J. F. Persistence of Net Heterotrophy in Lakes during Nutrient Addition and Food Web Manipulations. *Limnology and Oceanography* **2000**, *45*, 1718–1730.

(109) Fernández Castro, B.; Chmiel, H. E.; Minaudo, C.; Krishna, S.; Perolo, P.; Rasconi, S.; Wüest, A. Primary and Net Ecosystem Production in a Large Lake Diagnosed From High-Resolution Oxygen Measurements. *Water Resources Research* **2021**, *57*, e2020WR029283.

List of Figures

1.1 Comparison manual sampling and high frequency monitoring. . 3
1.2 Scheme of the pipeline of the proposed solution. 6

3.1 Design of the software architecture developed within this project. 14
3.2 Example of a text file format for data profile type. 25
3.3 Example of the dashboard created using the developed profile plugin visualization. 27
3.4 Dashboard for time series data. 28
3.5 Example of the importer form to make easier for the user the importing of the profile data. 28

4.1 Map of Lake Lugano, southern part of Switzerland. 36
4.2 The data flow architecture. 37
4.3 On the left, the design of the position of the chain sensors; on the right, the project of the platform designed and realized by Officine Ghidoni SA (Via al Pizzante 9, 6595 Riazzino, Switzerland). 38
4.4 The AHFM system architecture. 39
4.5 The power unit. 39
4.6 Comparison between AHFM and manual sampling depths. 40
4.7 The main unit components. 42
4.8 Data flow from buoy to server. 43
4.9 The system architecture. 45
4.10 Profile data type. 46
4.11 Daily data completeness heatmap based on the classification of the daily percent value of completeness in five FL classes: bad, poor, good, and high. 49
4.12 Daily aggregated data quality tests categorized into five FL discrete classes. 50
4.13 Data correction of time-series collected by the OPTOD sensors by Ponsel one positioned at 2.5 and another at 12.5 m depth. . . 52
4.14 Time in bulk, absorption and in float. 52

4.15 Energy yielded vs. energy consumed by the AHFM system. . . . 53
4.16 Minimum and maximum battery voltages. 54
4.17 General overview of the time-series comparison between AHFM and IDRONAUT probes. 56
4.18 Point to point comparison of dissolved oxygen saturation time-series between AHFM and manual sampling with an IDRONAUT 316 Plus probe. 58
4.19 Point to point comparison of dissolved oxygen saturation time series between AHFM and manual sampling with an IDRONAUT 304 Plus probe. 59
4.20 Comparison of FL dissolved oxygen saturation between AHFM time series and manual sampling with IDRONAUT probes. . . . 60
4.21 Comparison of water temperatures between AHFM and manual sampling with IDRONAUT probes. 60

5.1 Lake Lugano map [**barbieri1992description**]. 68
5.2 Graphical representation of main data sources available for Lake Lugano. 69
5.3 Data formats used to archive the historical data of Lake Lugano. 70
5.4 Data importing into istSOS. 71
5.5 Schema of the profile type data composed by timestamp (t_n), depth (d_n) and value. 74
5.6 Thermocline depth calculated using the LakeAnalyer web plugin and the monthly data collected by the traditional campaigns, previously standardized and imported into istSOS. 78
5.7 Mixed depth calculated using the LakeAnalyer web plugin and the monthly data collected by the traditional campaigns, previously standardized and imported into istSOS. 78
5.8 Epilimnion depth calculated using the LakeAnalyer web plugin and the monthly data collected by the traditional campaigns, previously standardized and imported into istSOS. 79

6.1 Schema of the AHFM system. 86
6.2 Example of a re-calibrated series of data. 89
6.3 Flowchart of the algorithm. 92
6.4 Temperature profiles in °C coming from the data collected using the chain of OPTOD by Ponsel sensors (year 2021). 94

6.5 Oxygen saturation profiles in % coming from the data collected using the chain of OPTOD by Ponsel sensors (year 2021). 94
6.6 Daily values of GPP, NEP and R in $gCm^{-2}d^{-1}$ at depths 0.4, 2.5 and 5 m starting from the top. 96
6.7 Daily values of GPP, NEP and R in $gCm^{-2}d^{-1}$ at depths 8, 12.5 and 20 m starting from the top. 97
6.8 Daily values of GPP, NEP and R over the year 2021. 98
6.9 Comparison of the total monthly Carbon per square meter calcualted using the two method considered. 99
6.10 Profiles concentrations of Chlorophyll-a (upper plot) and Phycocyanin (bottom plot) in mg m-3. 101
6.11 Profile concentrations of net primary productivity calculated by means of the algorithm developed within this project. 102

List of Tables

3.1 List of SOS requests with their short description. 25

4.1 Morphometric information about Lake Lugano. 36
4.2 Quality Indexes list. 47
4.3 Range test limits used at a data sensing layer. 47

5.1 List of functions available on the Lake Analyzer plugin. 72
5.2 List of observed properties registered and standardized on ist-SOS. 76

6.1 Lake Lugano morphometric characteristics. 83
6.2 OPTOD sensors technical specifications. 85
6.3 UNILUX sensors technical specifications. 85
6.5 List of equations based on the published paper by Staher [48]. . 91
6.4 LUFFT WS-501 weather station technical specifications. 104