THE ECONET PROJECT: USE OF AI FOR SURFACE WATER MONITORING WITH SATELLITE AND GROUND SENSOR DATA

Valeria La Pegna[1], Fabio Del Frate[1], Davide De Santis[1], Dario Cappelli[1], Martina Frezza[1], Roberto Dragone[2], Gerardo Grasso[2], Daniela Zane[2], Bruno Brunetti[2], Sabrina Foglia[2], Giorgio Licciardi[3], Patrizia Sacco[3], Deodato Tapete[3]

[1]Dipartimento Ingegneria Civile ed Ingegneria Informatica, Università degli Studi "Tor Vergata", Via del Politecnico 1, 00133 Rome, Italy
[2]Istituto per lo Studio dei Materiali Nanostrutturati, Consiglio Nazionale delle Ricerche (CNR-ISMN), P.le Aldo Moro 5, 00185 Rome, Italy
[3]Agenzia Spaziale Italiana (ASI), Via del Politecnico snc, 00133 Rome, Italy

ABSTRACT

The activities undertaken within the EcoNet project aim at the design and development of an integrated system for the monitoring of changes in surface waters natural status based on different sensoristic techniques. The proposed integration approach combines ground measurements and hyperspectral satellite images. The promising dialogue that occurs between these two multi-sensoristic technologies requires the implementation of appropriate tools for data handling and analysis which in this work are represented by Artificial Intelligence (AI), particularly suitable to retrieve very subtle relationships among the data. This integration can open enormous potential for overcoming the limits of traditional environmental monitoring and diagnostic techniques.

Index Terms— Hyperspectral remote sensing, bio/chemosensoristic devices, Artificial Intelligence, water quality

1. INTRODUCTION

Maintenance of the natural state of freshwater is threatened by urban settlements due to related anthropogenic activities (agriculture, industry, etc.). Various xenobiotic contaminants can be introduced into the receiving waters, with hardly predictable effects on freshwater ecosystem homeostasis. Existing monitoring systems can hardly understand complex spatial and temporal alteration in ecosystems, and do not allow early actions to counter anthropic impacts.

A novel and relatively unexplored area of research involves the integration of ground bio/chemosensoristic devices with remote sensing systems [1]. In recent years, Earth Observation (EO) multi/hyperspectral optical imaging technologies have proved their effectiveness as affordable and practical support for more traditional ground-based methodologies. Advantages include: provision of an independent characterization based on electromagnetic principles; wide range of spatial/temporal resolution, making the integration a valuable way for monitoring aquatic status and allowing early identifications of hazards and alert procedures. On the other hand, the contribution of satellite data for water quality monitoring is a consolidated technique [2-4]. Data from sensor instruments on board PRISMA, Landsat 8/9 and Sentinel-2 missions are nowadays exploited for monitoring inland waters and their alterations [5, 6].

Ground (bio)sensoristic devices as on site and cost effective analytical systems can provide real time information about the presence of bioactive compounds in environmental matrices like surface waters. Ground (bio)sensoristic devices can therefore be employed for a rapid diagnostics and early corrective actions. This proves advantageous in real world scenarios where preliminary screening is necessary to direct subsequent confirmatory and advanced analyses [7].

This paper reports a selection of the results achieved in the EcoNet project [8] after one year of work. By integrating EO, in situ bio/chemosensoristic devices and AI, EcoNet pursues: *i*) association and correlation between two or more indices measured continuously by remote sensing and/or ground sensors; *ii*) complementarity of ground and satellite sensor systems.

2. METHODOLOGY

2.1. Test areas

To validate the designed methodology, the test areas include the following natural reserves of the Natura 2000 network, hosting rare and threatened species for which the prime concern is to ensure long-term subsistence of the habitats: Riserva Naturale Regionale "Selva del Lamone" (Farnese, Viterbo, Italy); Riserva Naturale Regionale "Nazzano Tevere-Farfa" (Nazzano, Roma, Italy); Lago di Piediluco (Piediluco, Terni, Italy).

All the test areas share common factors of anthropogenic pressures deriving from punctual and diffuse sources of pollution: peri-urban agro-zootechnical productions and discharges from wastewater treatment plants. Local stakeholders responsible for managing the natural reserves and water resources are involved in the project, providing support for the identification of surface water sampling points and field activities, including the sampling itself.

2.2. Water quality parameters

Concentrations of relevant aquatic parameters from satellite data is carried out starting from the most significant remote sensing techniques available in literature [9].

Chlorophyll-a (Chl-a) was firstly identified, being a specific form of chlorophyll used in the photosynthesis process that allows an indirect estimation of phytoplankton biomass. It provides the main photosynthetic pigment present in microalgae and constitutes one of the main indicator parameters of the nutritional status and organic pollution in waters. As for Total Suspended Matter (TSM) concentration, usually associated with water turbidity, it is used in determining water quality after wastewater treatments. It indicates the quantity of solids, measured in units of mg/l, present in suspension. In addition, the concentration of Colored Dissolved Organic Matter (CDOM), was also analyzed. It represents an optical measure of the absorption of light from organic materials coming from tannins production by the decomposition of organic matter which absorbs strongly at blue wavelengths (450-485 nm). Also Total Phosphorus (TP) and Total Nitrogen (TN) are considered, both representing chemical parametric indicators of emission of urban waste water discharges and strictly linked to the presence of suspended organic particles.

2.3. Surface water sampling and ground sensor analysis

Surface water samples were weekly/monthly collected by each test area in surface water sampling points located in the waterways. The methods of collection and conservation of surface water samples were obtained starting from the legislation reference (UNI EN ISO 5667-1:2022 Water quality - Sampling - Part 1: Guidance on the design of sampling programs and sampling techniques).

In the project field measurements are carried out by the CNR-ISMN research group through the implementation of "Snoop" prototype, a CNR-ISMN European patent coupled with other sensoristic probes. "Snoop" prototype measurements were carried out for the assessment of level of general or integral toxicity levels, linked to the presence of contaminating bioactive substances (e.g. pesticide residues and heavy metals) in surface water samples.

More specifically, "Snoop" prototype was configured to the use a microalgal sensitive material for the measuring of the chlorophyll-a (Chl-a) fluorescence kinetics curves [10]. Various substances such as pesticides, heavy metals, as well as some pharmacological active ingredients are able to interfere with the photochemistry of the cellular processes involved in the production of the biooptical signal by the microalgal sensitive material [11]. Studying the changes in parameters and the profile of the Chl-a fluorescence kinetic curves generated by the material sensitive microalgal sensitive material exposed to surface water samples it is possible to determine the levels of bioactive contaminants. The percentage index of photosynthetic interference (%A) was calculated by comparing the values of the normalized areas (Sm) of the OJIP curves obtained from the microalgal suspensions exposed to surface water samples (Sm_{exp}) and the control microalgal suspensions (Sm_{blk}):

$$\%A = [(Sm_{exp} - Sm_{blk})/(Sm_{blk})] \times 100$$

"Snoop" was used coupled with other electrochemical and optical probes for the following parameters: temperature (°C), dissolved O_2 (mg/L), dissolved CO_2 (mV), redox potential (mV), pH, ionic conductivity (mS/cm), free inorganic ion fraction (Ca^{2+}, K^+, Na^+, NH_4^+, NO_3^-, Cl^- and F^-). Visual Basic for Applications has been used within Excel to automate and enhance sensoristic data analysis processes.

2.4. Satellite Dataset

ASI's PRISMA (*PRecursore IperSpettrale della Missione Applicativa*) hyperspectral images are used. The retrieval algorithm relies on the reflectance in the visible spectrum (range 438-754 nm) given the strong sensitivity to the chosen water quality parameters.

Target values are obtained via physical models derived with visible bands from Landsat 8/9 and Sentinel-2 images, which are both characterized by a higher temporal frequency of data acquisition, maintain similar spatial resolution (30 m) as the PRISMA data, but lack in spectral resolution [12]. There is a maximum of 3 days between the two types of satellite acquisitions, allowing the spectral information coming from the sensors to be compared, assuming minimal changes into the water environments. The time series analysis based on Landsat 8/9 and Sentinel-2 has shown clear similarities between the two satellites measurements.

Reflectance values are extracted at pins equally and widely distributed within the surfaces of the water bodies, maintaining the geographical information of the sampling points of the in-situ data measurements. The use of data acquired in-situ by the CNR-ISMN group is carried out by including these concentration values as additional example data to be inserted into the AI models to increase the density and accuracy of the proposed examples.

2.5. AI algorithms development

Multi-Layer Perceptron (MLP) models were chosen, being artificial neural network made up of multiple layers of nodes internally connected to each other, generally used for retrieval prediction analysis [13]. Given the variability of the natural reserves and the variety of biological and physical parameters to estimate, a single AI algorithm for each variable and each area of interest was developed. The AI models present similar network topology characteristics, with one hidden layer assumed for each application and sigmoidal activation function. The input layer, consisting of reflectance values at sensitive channels, the amount of neurons in the hidden layer and the training epochs are determined depending on water parameter and density of the input dataset. All simulations were performed in Python environment with the use of the Keras, Scikit-learn and TensorFlow libraries, characterized by open access licenses.

Following the training phase, the outputs of the networks represent the estimated concentration values of every water quality parameter. The trained model is then extended to the entire surface of the water body to observe spatial distribution of significant quality indices.

For each AI model, the initial input dataset was divided with 60% of the available dataset dedicated to the training phase, while the remaining 40% was further divided into testing (22% of the total dataset) and validation (remaining 18 % of the dataset).

The AI models accuracy and ability to predict the parameters of interest is evaluated through the coefficient of determination (R^2), the Root Mean Squared Error (RMSE) and the Mean Absolute Error (MAE), as they currently represent the most used metrics to determine the precision obtained by AI retrieval algorithms.

3. RESULTS AND DISCUSSION

The results currently derived from the AI models development proved to be interesting for estimating parameters assessing the quality status of a water basin. In particular, for each test sites, R^2 values of above 0.67 up to 0.95 were obtained in the comparison between the estimations retrieved through the AI algorithm model and the ones obtained with the physical-mathematical model from the multispectral sensors. As an example, Fig.1 and Fig.2 show the results achieved at the confluence of the Tiber and Farfa rivers in the Nazzano reserve for TSM.

As refinement strategy, data augmentation techniques are being considered to broaden the statistical representativeness of the PRISMA input dataset. Data volume enlargement approach could allow to improve the statistical robustness of the initial dataset, while increasing the estimation capability of the AI algorithms.

As an example, Fig. 3 shows the results from “Snoop” measurements of ammonium ion (NH_4^+) concentration in water samples collected from Selva del Lamone. Ammonium ions (NH_4^+) can be present in water, often resulting from the breakdown of organic matter or the discharge of nitrogen-containing compounds.

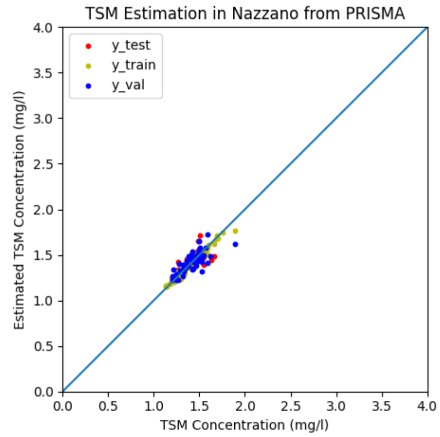

Fig. 1 - Scatterplot of TSM concentration derived from Landsat 8 physical model vs. TSM estimated with AI model in Nazzano reserve.

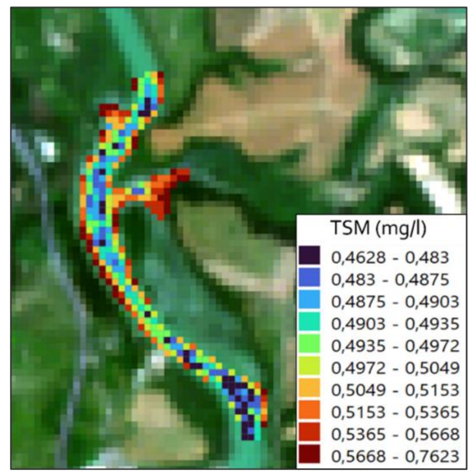

Fig. 2 - Estimated concentration of TSM at the confluence of Tevere and Farfa rivers (Nazzano reserve) derived from PRISMA image of 27/06/2023.

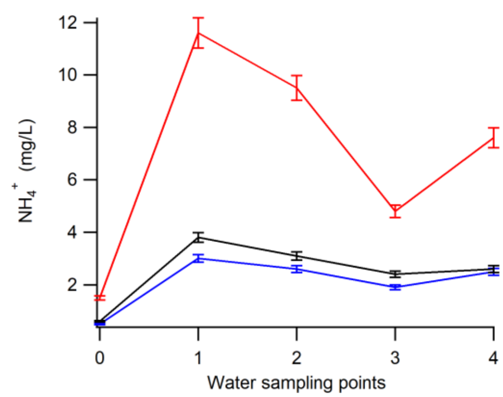

Fig. 3 - Ammonium ion (NH_4^+) concentration (mg/L) measured on water samples collected from Selva del Lamone reserve. Blue line, black line and red line refer to 14 July 2023, 31 July 2023 and 28 August 2023 water sampling campaigns, respectively. Surface water sampling points on X-axis: 0 (Mezzano lake), 1 (Valle Latera), 2 (S. Maria di Sala bridge), 3 (Salabrone waterfalls), 4 (Farnese water treatment plant). %RSD ≤ 5%.

High levels of ammonium ions in surface water may be indicative of recent pollution from various sources: agricultural runoff (fertilizers and animal waste), (peri) urban runoff/wastewater discharge (i.e., effluents from sewage treatment plants and runoff from urban areas where nitrogen-based fertilizers are used) and/or natural processes (by decomposition of organic matter, such as dead plants and animal waste). One main difference between the water sampling campaigns whose results are reported in Fig. 3 is that water samples of 28 August 2023 have been collected following overnight heavy rainfall. The NH_4^+ concentration in surface waters can increase due to soil runoff following heavy rainfall. When rainwater interacts with soil, it can mobilize and carry with it dissolved and suspended substances, including ammonium ions leading to increased levels in nearby water bodies [14]. This phenomenon may have contributed to the elevated ammonium ion concentrations in water samples collected on 28 August 2023.

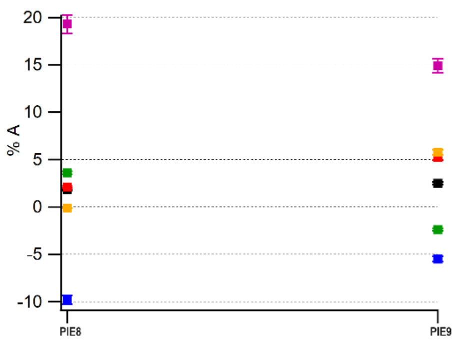

Fig. 4 - Percentage index of photosynthetic interference (%A) measured on water samples from Piediluco lake, 2 sampling points PIE8 (coord. 42.5337, 12.7559) and PIE9 (coord. 42.5240, 12.7669). Black dots (May 2023), purple dots (June 2023), blue dots (July 2023), green dots (August 2023), red dots (September 2023) and orange dots (October 2023). %RSD ≤ 5%.

Fig. 4 shows the percentage index of photosynthetic interference (%A) calculated after "Snoop" biosensoristic measurements on water samples collected monthly (May-October 2023) from Piediluco lake, 2 sampling points PIE8 (located in the middle of Piediluco lake) and PIE9 (located in the southern branch of Piediluco lake). Differences in %A values can be observed for water samples collected on the same day but in different sampling points. Except for samples collected in June and July 2023, all the %A values are within a range of ± 5%. Confirmatory analysis on the same samples have been performed by the Regional Agency for the Protection of the Environment (ARPA) of Umbria and results are currently under evaluation.

4. CONCLUSIONS

The processing of the sensoristic data obtained so far offers interesting prospects. AI algorithms to identify possible association and correlation between two or more indices measured by remote sensing and/or ground sensors are currently being developed and tested. The integration of ground measurements with satellite data can enable the reliability of water quality monitoring since multi sensor data approach can overcome the limitations of single acquisitions techniques, providing also the chance to gain higher spatial and temporal resolutions. EcoNet therefore contributes to the ongoing efforts to optimize the early identification of possible hazards in water quality management.

Finally, EcoNet includes specific educational sessions leading to build qualified personnel on water monitoring through analytical and satellite measurements. Training involves an e-learning course and a technical training mainly focused on aquatic ecosystems environmental monitoring.

5. ACKNOWLEDGEMENT

The EcoNet project is funded by the Italian Space Agency (ASI), agreement n. 2022-32-HH.0, as part of the program "Innovation for Downstream Preparation for Science" (I4DP_SCIENCE) [15].

6. REFERENCES

[1] G. Grasso et al., "Field and Remote Sensors for Environmental Health and Food Safety Diagnostics: An Open Challenge", *Biosensors*, pp. 1-3, 2022.

[2] A. Gitelson et al., "Quantitative remote sensing methods for real-time monitoring of inland waters quality", *International Journal of Remote Sensing*, 14:7, pp. 1269-1295, 1993.

[3] M.H. Gholizadeh et al., "A Comprehensive Review on Water Quality Parameters Estimation Using Remote Sensing Techniques", *Sensors*, 16:8, pp. 1298, 2016.

[4] J.C. Ritchie et al., "Remote Sensing Techniques to Assess Water Quality", *Photogrammetric Engineering & Remote Sensing*, 6:10, pp. 695-704, 2003.

[5] K. Toming et al., "First Experiences in Mapping Lake Water Quality Parameters with Sentinel-2 MSI Imagery", *Remote Sensing*, 8, pp. 640, 2016.

[6] M. Niroumand-Jadidi et al., "Water Quality Retrieval from PRISMA Hyperspectral Images: First Experience in a Turbid Lake and Comparison with Sentinel-2", *Remote Sensing*, 12:23, pp. 3984, 2020.

[7] R. Dragone et al., "Portable bio/chemosensoristic devices: innovative systems for environmental health and food safety diagnostics", *Frontiers in Public Health*, 5:80, 2017.

[8] V. La Pegna et al., "The Econet Project: AI Based Satellite and Ground Sensor Analysis for Surface Waters Protection," *IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, pp. 2580-2583, 2023.

[9] H. Yang et al., "A Review of Remote Sensing for Water Quality Retrieval: Progress and Challenges", *Remote Sensing*, 14(8):1770, 2022.

[10] A. Stirbet, "On the relation between the Kautsky effect (chlorophyll a fluorescence induction) and Photosystem II: Basics and applications of the OJIP fluorescence transient", *Journal of Photochemistry and Photobiology B: Biology*, 104:1-2, pp. 236-257, 2011.

[11] G. Grasso et al., "Microalgae-Based Fluorimetric Bioassays for Studying Interferences on Photosynthesis Induced by Environmentally Relevant Concentrations of the Herbicide Diuron", *Biosensors*, 12:67, 2022.

[12] S. Poddar et al., "Estimation of Chlorophyll-a in Northern Coastal Bay of Bengal Using Landsat-8 OLI and Sentinel-2 MSI Sensors", *Frontiers in Marine Science*, 6:598, 2019.

[13] A. Juna et al., "Water Quality Prediction Using KNN Imputer and Multilayer Perceptron", *Water*, 14(17):2592, 2022.

[14] N. Akhtar et al., "Various natural and anthropogenic factors responsible for water quality degradation: A review", *Water*, 13(19):2660, 2021.

[15] D. Tapete and A. Coletta, "ASI's roadmap towards scientific downstream applications of satellite data", EGU General Assembly 2022, EGU22-5643, Vienna, 2022.