



**Deliverable D8.10 – Activity 8.2, CNR IRET LE
PRELIMINARY VERSION OF THE
BIOMASS VRE INCLUDING THE RELEVANT DATA
AND ANALYSIS/MODELLING TOOLS"**





Deliverable number:	D8.10
Work package:	WP8
Intermediate Objective:	IO8.7
Deliverable type:	<input checked="" type="checkbox"/> Document, report
	<input type="checkbox"/> Websites, patent filings, videos, etc.
	<input type="checkbox"/> Other: please specify
Dissemination level:	<input checked="" type="checkbox"/> Public
	<input type="checkbox"/> Restricted
Estimated delivery (bimester):	B10
Actual delivery date:	30.06.2024
Author(s) (Partner-OU):	AUTHORS: ALBERTO BASSET, TEODORO SEMERARO, JESSICA TITOCCHI, LORENZO LIBERATORE AND FLAVIO MONTI
Reviewed by:	Executive board

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System - CUP B53C22002150006 (D.D. n. 130/2022)
 Funded by EU - Next Generation EU
 Mission 4 “Education and Research” - Component 2: “From research to business” -
 Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”



Sommario

1.	<i>OVERVIEW OF THE WORK PACKAGE 8.2</i>	4
2.	<i>STATE OF THE ART OF BIOMASS VRE</i>	4
	2.1. Functional and technical requirements for VRE	5
	2.2. Workflows of Biomass VRE	6
	2.3. Dataset implementation	9
	2.4. The typologies of products and models	10
3.	<i>ANALYSIS CURRENTLY DEVELOPED</i>	10
4.	<i>CASE STUDY DEVELOPMENT</i>	16
	4.1. Phytoplankton production response to global warming: estimating patterns of change across latitudes	16



1. OVERVIEW OF THE WORK PACKAGE 8.2

The Work Package (WP) 8.2 - Virtual Research Environment for aquatic Biomass services (BIOMASS VRE) is part of WP 8 (Virtual Research Environments and Cross-disciplinary Activities) of the Italian Integrated Environmental Research Infrastructures System (ITINERIS). ITINERIS is a project funded by EU - Next Generation EU PNRR- Mission 4 “Education and Research” - Component 2: “From research to business” - Investment 3.1: “Fund for the realization of an integrated system of research and innovation infrastructures”.

The main goal of WP 8.2 is to strengthen the collaboration between the e-Science European infrastructure for biodiversity and ecosystem research of LifeWatch ERIC, through its national node of LifeWatch Italy, and the CNR-IRET with the creation of a specific Virtual Research Environment: Biomass VR. The planned activity is to design, implement and validate a VRE on the responses of aquatic ecosystems to climate change, based on earlier developments by LifeWatch ERIC. This will be achieved through a use case on aquatic biomass responses to global warming, which will integrate data resources available in existing repositories, as well as services that will be available and accessible for the users through developing specific modelling tools. The use case can be of significant scientific and socio-economic interest, as it enables the investigation of the impact of biomass on the release of various ecosystem services, including supporting, provisioning, regulating, and cultural services.

The **Biomass VRE** will be of significant interest to the scientific community engaged in biodiversity and ecosystem research, with potential applications extending to the economic and social sciences. By facilitating research activities, the VRE will contribute to achieving international sustainable development goals (SDGs, Agenda 2030). These include promoting sustainable consumption patterns (SDG 12), conserving marine diversity and regulating fishing practices (SDG 14), reversing deforestation and desertification (SDG 15), and strengthening global partnerships for sustainable development (SDG 17).

2. STATE OF THE ART OF BIOMASS VRE.

The initial phase of WP8.2 focused on establishing the analytical framework for the development of the Biomass VRE. This included identifying the essential data types and formats, analyses and models to be integrated into its workflows and used for validation in case studies. A comprehensive review of existing data and web services was conducted, both internally and in collaboration with WP6 researchers, to ensure alignment with existing data resources and to identify the specific requirements and needs for models and tools within the ITINERIS project.

Subsequently, a *Terms of Reference* document was developed to guide the development of the Biomass VRE, as an external contractor will carry out the full implementation of the Biomass VRE.

This document outlines the characteristics and structure of the workflows and the minimum functional and technical requirements necessary for the development of the VRE. It



describes in detail all activities required for the complete development of the VRE, including the creation of web interfaces for the VRE, the specification of workflows, a list of potential data sources, the types of analyses and models to be integrated, and the development of scripts and algorithms needed for web services.

In February, the tendering process was launched using the *Terms of Reference* to identify the corporate contractor. The procurement process is currently ongoing.

2.1. Functional and technical requirements for VRE

The VRE will leverage existing resources and web services within the LifeWatch Italy and ERIC e-Science research infrastructures. Specifically, it will interface with the LifeWatch Data Portal and Metadata Catalogue to facilitate the acquisition and harvesting of data resources. The VRE will also integrate with the EcoPortal platform, a repository of semantic resources in the biodiversity and ecological domains, to ensure data harmonization and semantic interoperability. Additionally, the VRE will connect with Data Labs, LifeWatch's collaborative coding platform, to provide open access to scripts and projects and enable the deployment of web services. (Figure 1).

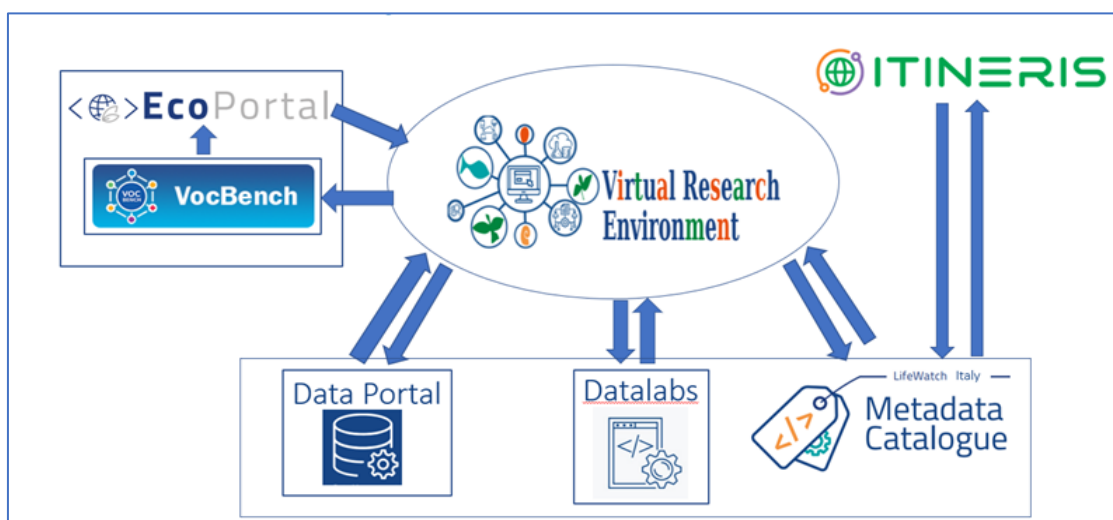


Figure 1. Current LifeWatch resources and their links in the Biomass-VRE.

The Biomass VRE will be implemented with a modular structure that includes data analysis and modelling services by automating computational operations in workflows in the cloud, without the need to install or manage softwares locally. In each block, service development can be supported by the R, Python and Matlab language programs. All workflows will be functional to return outputs that can be used for research reports.

The Biomass VRE will be developed using established technologies such as Tesseract and NaaVRE available from LifeWatch ERIC, with the aim of maximizing integration between these two platforms (Figure 2). The 'Biomass' VRE must guarantee the integrity and security of all input data throughout the utilization process.

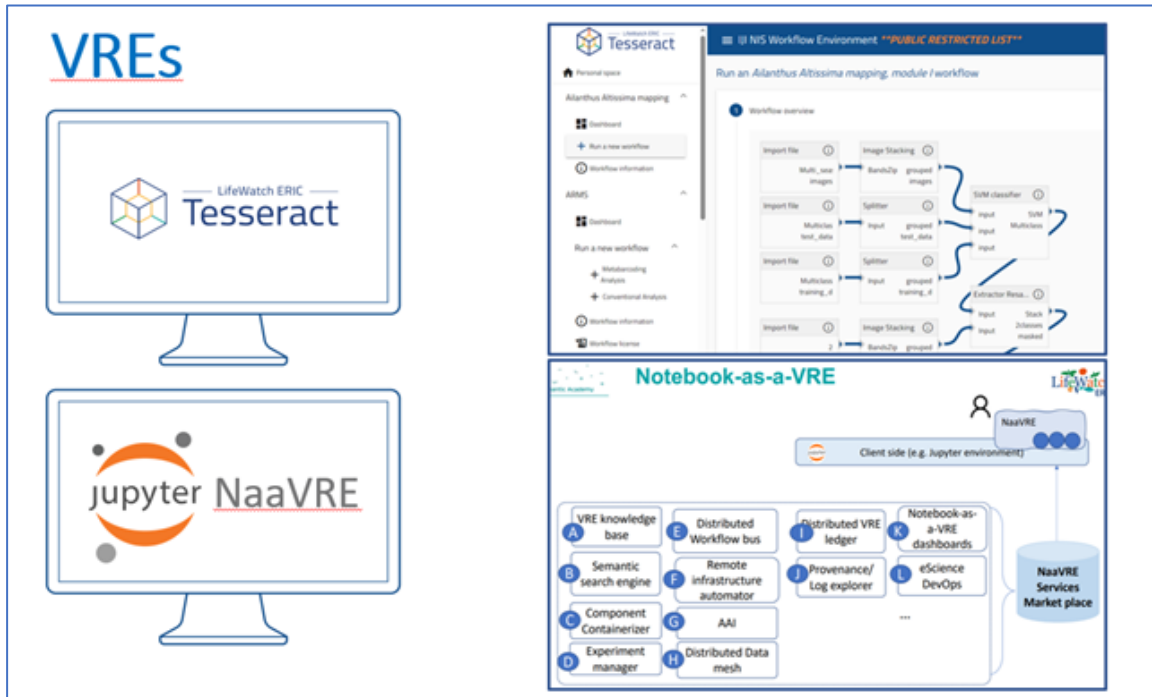


Figure 2. Tesseract and NaaVRE technology presentation.

The Biomass VRE will be accessible via a website which will be specifically created for this purpose within LifeWatch Italy and accessible from the CNR reference portal for the ITINERIS project (ITINERIS hub). The VRE interfaces web page will comply with the W3C accessibility standards. The portal will serve as a public forum for the dissemination of information pertaining to VRE, encompassing general data, the composition of the design team, a catalogue of published works, and an overview of associated projects. Furthermore, a private area will be provided, accessible via single sign-on authentication, for the management of content and the administration of the site.

2.2. Workflows of Biomass VRE

The Biomass VRE will include three main workflows dedicated to:

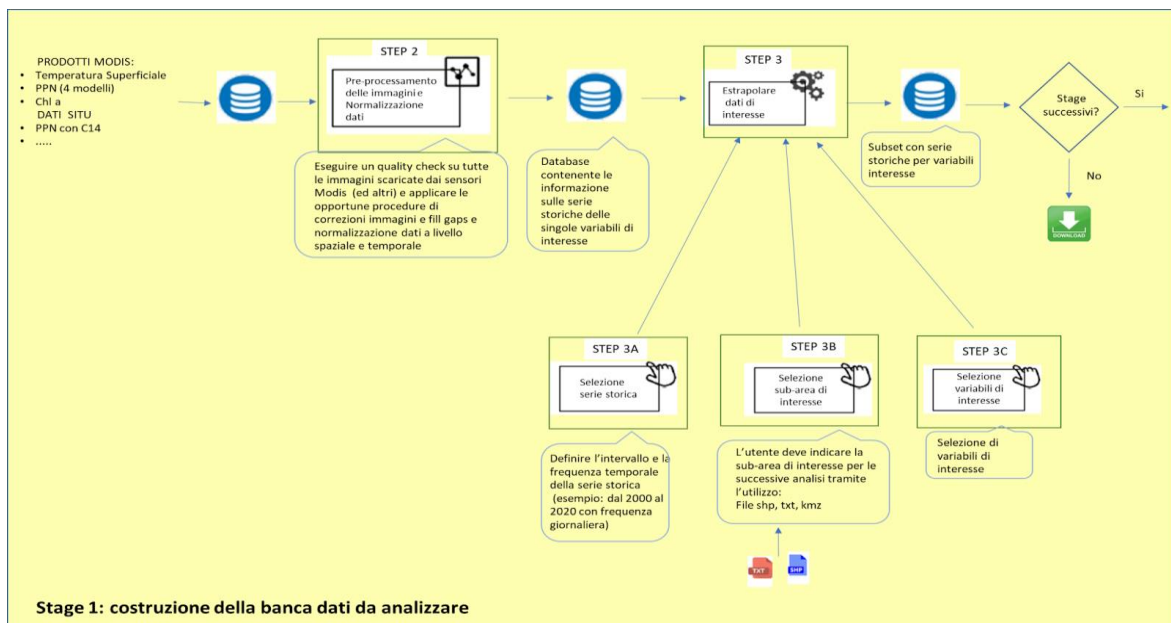
- Aquatic Biomass (Figure 3). This workflow has been designed to study the effects of thermal gradients on aquatic biomass at the producer level. It enables the modeling of time series for phytoplankton and net primary productivity in relation to sea surface temperature, integrating field data with remote sensing data. This analysis is useful for understanding and predicting how aquatic ecosystems respond to climate change and other disturbances. (Figure 3a)



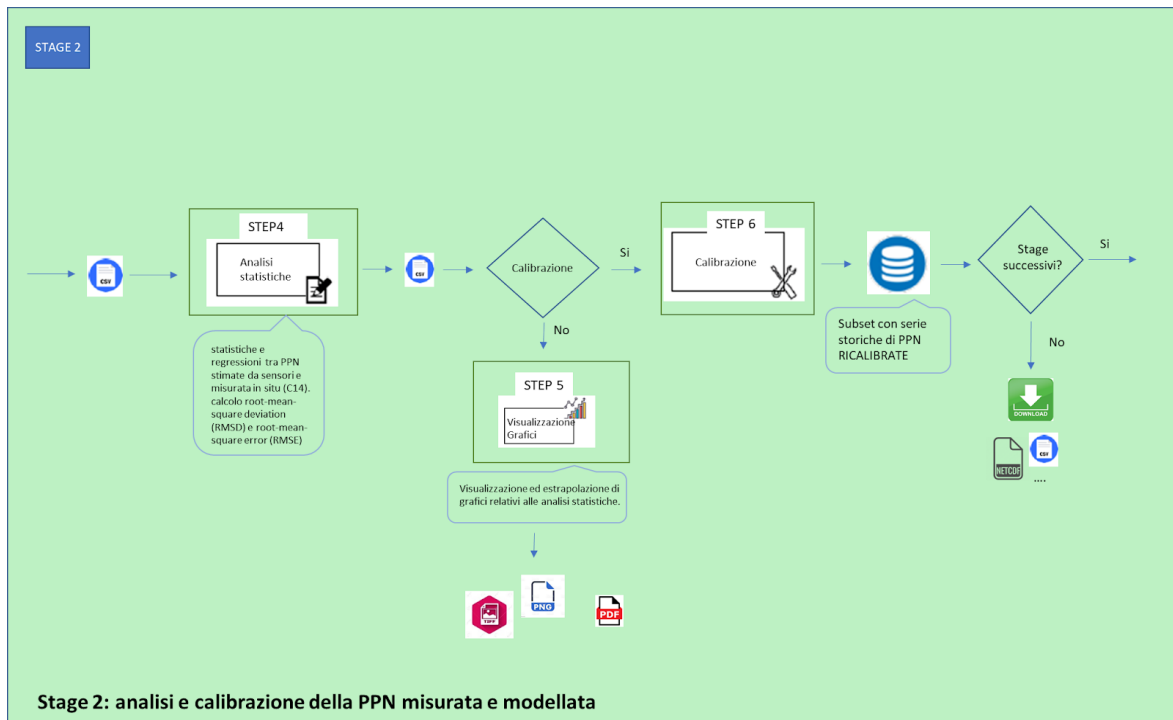
- Terrestrial Biomass (Figure 3). This workflow has been designed to examine the effects of thermal gradients on terrestrial biomass at the producer level. It allows for the modeling of time series for vegetation indices and net primary productivity in relation to land surface temperature. This analysis helps in understanding and predicting how terrestrial ecosystems will respond to climate change and other disturbances. (Figure 3b, c)
- Consumer Biomass (Figure 4). This workflow focuses on the effects of thermal gradients on biomass at the consumer level, specifically examining metabolic rate and home-range parameters. The workflow allows the modeling of the responses of animal organisms to global warming in terms of metabolism and spatial use. (Figure 4)

Each workflow will be designed with a modular structure that integrates services for data discovery, harmonization, interoperability, analysis and modeling. These workflows will be incorporated into the ITINERIS Hub via LifeWatch Italy and will be accessible through APIs from the web portal.

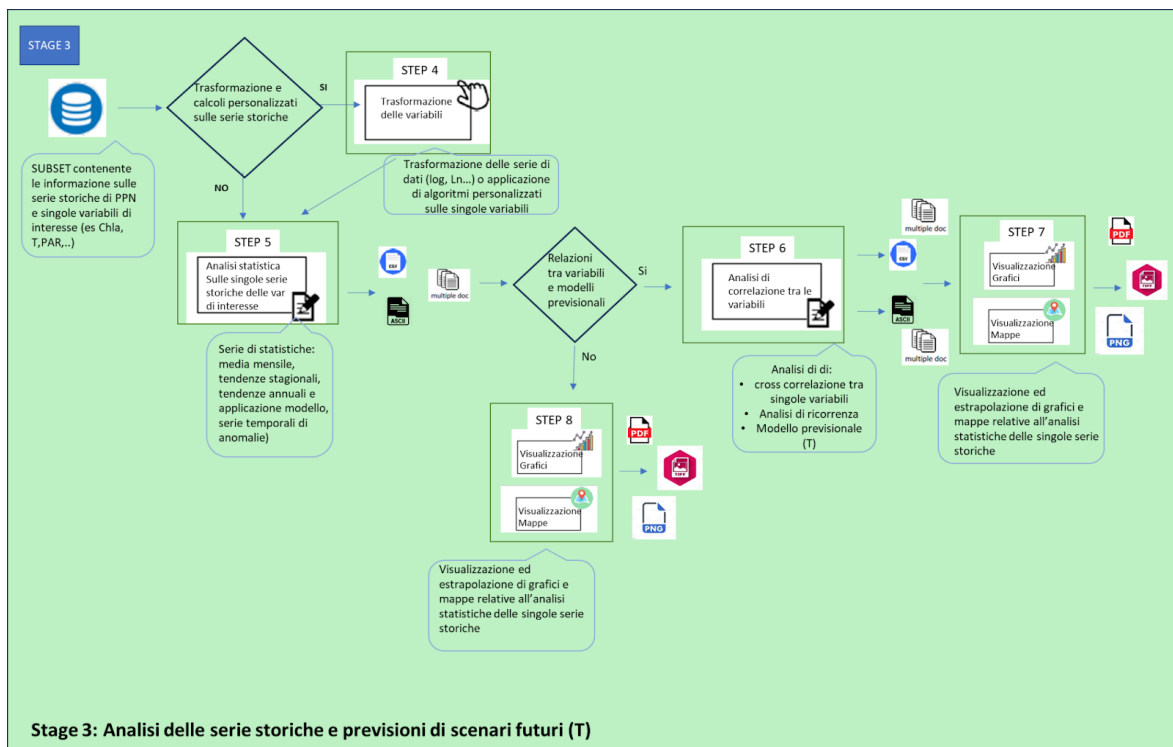
The input data types for these analyses include text or tables in CSV or TXT format, raster data in HDF, Open NetCDF, or IMG formats, and vector data in SHP format. The outputs of the workflows are tables in CSV or TXT format, raster imagery of the newly derived products, and plot results, such as time series, in PDF, JPG, or PNG formats. The inputs and outputs can be adapted or modified during the development phase to accommodate specific needs, in collaboration with the scientific team at CNR-IRET-LE.



A



B



C

Figure 3. A-B-C Example of schematic representation of VRE biomass: aquatic and terrestrial environments (part extrapolated from the Term of Reference).

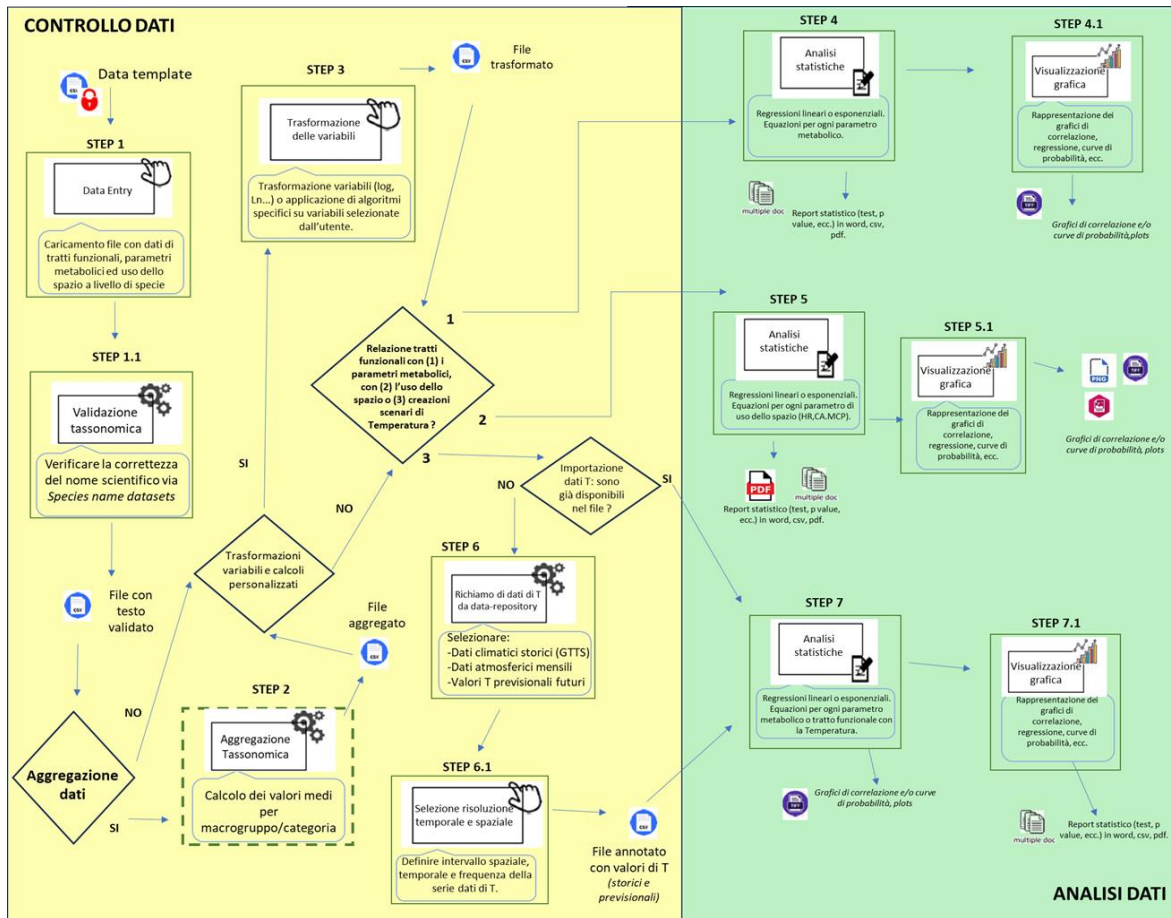


Figure 4. Example of representation of the biomass VRE at the level consumers (part extrapolated from the Term of Reference).

2.3. Dataset implementation

The VRE will be populated by creating databases from open-access data sources or by using available data from the DataPortal of LifeWatch Italy (<https://dataportal.lifewatchitaly.eu/data>).

The VRE is structured to combine and process, in distinct workflows, different typologies of data:

- Field data of Sea temperature, Chl-a and Net Primary Production (NPP) estimated with the C14 method and gathered from online repositories, such as Ocean Productivity (<http://sites.science.oregonstate.edu/ocean.productivity/field.data.c14.readme.php>) and PANGAEA (<https://doi.pangaea.de/10.1594/PANGAEA.932417>, <https://bats.bios.asu.edu/>). These data are downloaded and provided as table files, (in txt and/or .csv format). For the case study, we selected the data on the equatorial zone (From Lat 29, to Lat -29 and from Lon 178 to Lon-178).
- Trait based data (e.g. body size, primary production, metabolic rate and reproduction), metabolism and space use of both aquatic and terrestrial organisms and available in the form



of txt and/or .csv files. These data were acquired from literature sources, online data repositories, laboratory experiments and/or field monitoring.

- Ocean remote sensing imagery of Sea Surface Temperature (AQUA_MODIS) and Chl-a (AQUA_MODIS) from Ocean Color (<https://oceancolor.gsfc.nasa.gov/13/order/>) and Net Primary Production from Ocean Productivity (<https://sites.science.oregonstate.edu/ocean.productivity/>). The datasets are downloaded in the form of raster data (NetCdf, hdf format). For the case study, we used the global imagery cut for the equatorial zone, from the 2003 to 2023 dataset and they are cut using vectorial .shp file (From Lat 29, to Lat -29 and from Lon 178 to Lon-178).
- Terrestrial Remote Sensing data of Land Surface Temperature, Vegetation Indices, Gross primary production, Respiration, in the form of raster data in NetCdf, hdf, from EARTHDATA (<https://lpdaac.usgs.gov/products/mod16a2gfv061/>).

2.4. The typologies of products and models

The applications planned in the workflows include:

- Data aggregation and filtering.
- Raw variable transformation algorithms.
- Correlation and linear/non-linear regression models to study the data patterns between Chl-a, NPP and Sea Temperature in aquatic and terrestrial ecosystem.
- Allometric relationships between body size and energy consumption.
- Calculation services and analysis tools that relate species functional traits with i) metabolic parameters; ii) space use parameters (HR, MCP); and iii) metabolic and functional relationships with environmental parameters (e.g. temperature).
- Models for time series forecasting such as Seasonal Autoregressive Integrated Moving Average.
- Plot the results and download of the files.

3. ANALYSIS CURRENTLY DEVELOPED

A prototype representing part of the analysis processes of the Aquatic Biomass workflow has been developed on DataLabs by LifeWatch Italy, a virtual laboratory platform for collaborative coding and application management. This prototype included data acquisition activities to populate the models and perform the analysis processes. The analysis processes have been set up by developing some case studies: *“Phytoplankton production response to global warming: estimating patterns of change across latitudes.*

The input data of the workflow included collecting data from online repositories of field ocean data, such as sea temperature, Chl-a and Net Primary Productivity (NPP) estimated using the ¹⁴C method, along with MODIS remote sensing imagery, such as MODIS Chl-a and Sea Surface Temperature (SST) products (Figure 5).

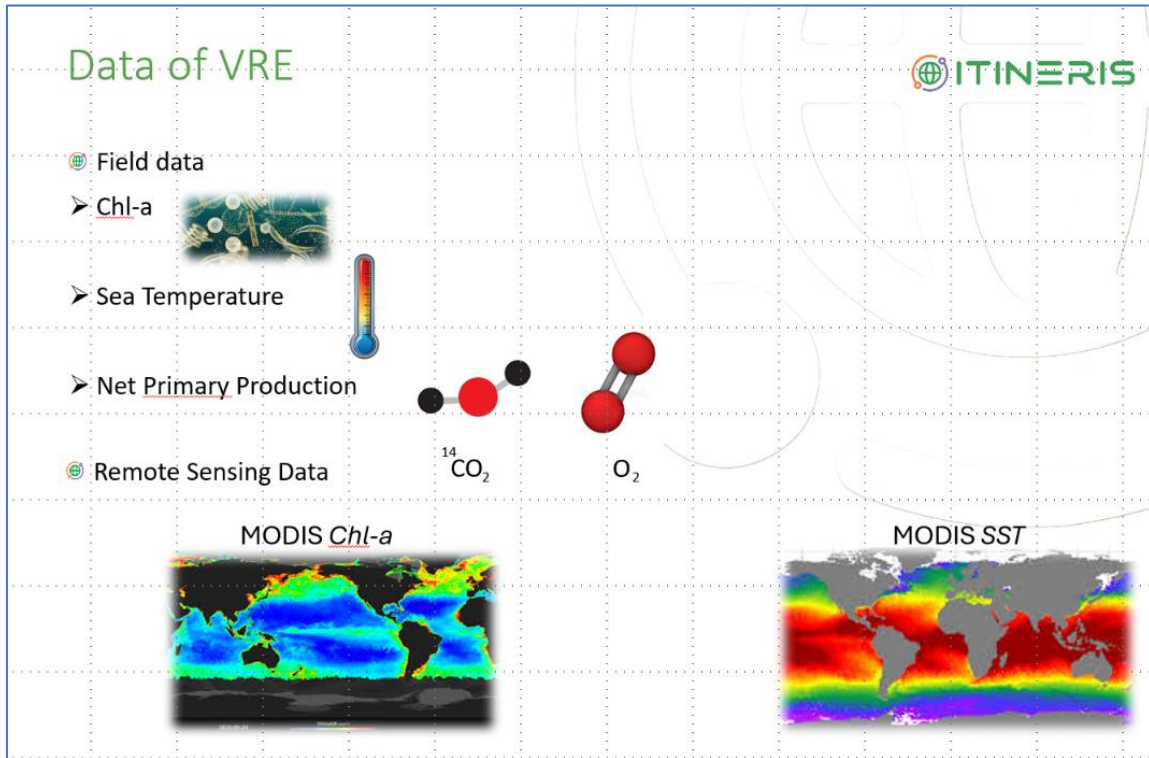


Figure 5. Representation of the dataset used in the prototype developed.

The Aquatic Biomass workflow prototype has been divided into four steps of analysis performed in DataLabs (Figure 6).

LIST OF PROJECTS							Create Project +
Title	Version	Programming language	Category	Status	Published		
WF_Biomasse_NPP	V2	r-4.3.0-cran	data processing	RUNNING	No	↕	
Risposta della NPP al riscaldamento globale: stima delle variazioni nella Produttività Primaria in ampie fasce latitudinali utilizzando SST e Chl-a come variabili predittive							
Creator: Teodoro Semeraro (teodoro.semeraro@cnr.it)				Open	Stop ■ ...	Actions ▾	
Aquatic_Biomass_Temporal_Forecast	V2	r-4.3.0-cran	data processing, data analysis, machine l...	RUNNING	No	↕	
Aquatic biomass prediction using sea surface temperature as an exogenous variable in the SARIMAX model. Applications of spatial correlations in Chl-a and SST to Remote Sensing data.							
Creator: Teodoro Semeraro (teodoro.semeraro@cnr.it)				Open	Stop ■ ...	Actions ▾	
WF_Biomasse	V2	r-4.3.0-cran	data processing	RUNNING	No	↕	
Risposta di Biomassa e produzione fitoplanctonica al riscaldamento globale: stima delle variazioni nella biomassa stabile in ampie fasce latitudinali							
Creator: Teodoro Semeraro (teodoro.semeraro@cnr.it)				Open	Stop ■ ...	Actions ▾	
Aquatic_Biomass_Temporal_Forecast	V1	r-4.3.0-cran	data processing, data analysis, machine l...	READY	No	↕	
Previsional_models	V2	r-4.3.0-cran	data processing, data analysis	RUNNING	No	↕	
Use of random forest and linear regression to estimate Chl-a and NPP from temperature. The models can simultaneously work with Chl-a and temperature (and other parameters) to estimate NPP							
Creator: Teodoro Semeraro (teodoro.semeraro@cnr.it)				Open	Stop ■ ...	Actions ▾	

Figure 6. The four analysis processes developed on DataLabs to create prototype workflow for Aquatic Biomass.



The first analysis process applies models to analyze the relationship between NPP and Chl-a with thermal gradients using field data (Figure 7). Mainly, starting from the table input file with the information of SST, Chl-a, NPP, and other information acquired by field data, it applies and compares the linear regression model with the random forest models highlighting the best model in representing the pattern variation of the Chl-a or NPP in relation of SST. Using the table of new predictable variables, it is possible to predict a new value of Chl-a and NPP.

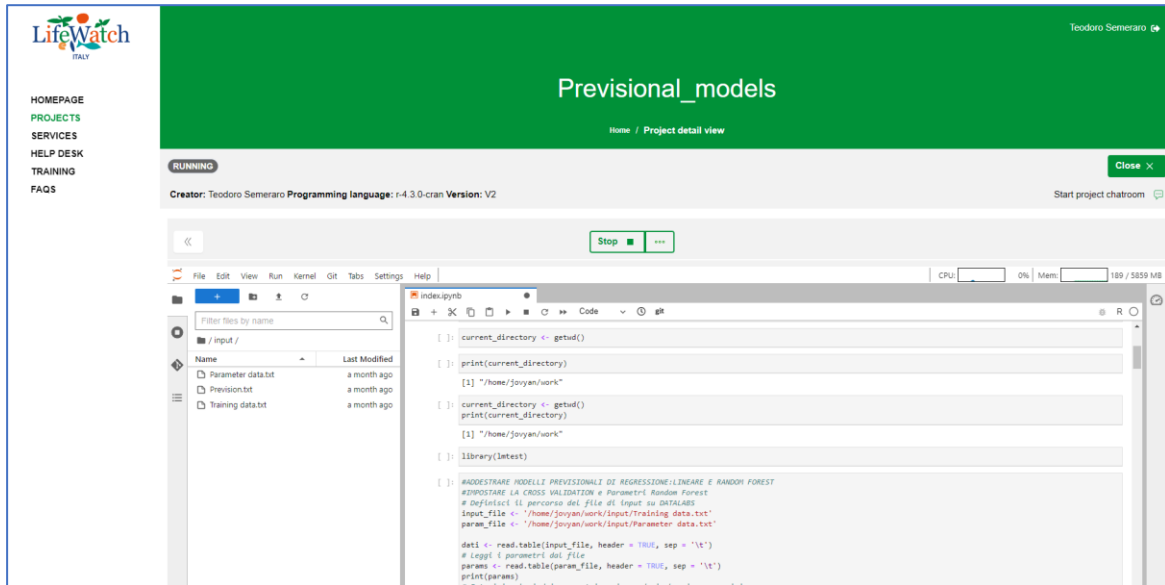


Figure 7. DataLabs development screen for analysis processes.

The second (Figure 8) and third (Figure 9) analysis processes apply regression models, populated with field data, to remote sensing imagery of sea surface temperature to produce new dataset products of Chl-a and NPP. From this new imagery dataset, it is possible to extrapolate the time series profile, trend analyze for the study area selected and create a potential future scenario of variation of Chl-a and NPP over time.

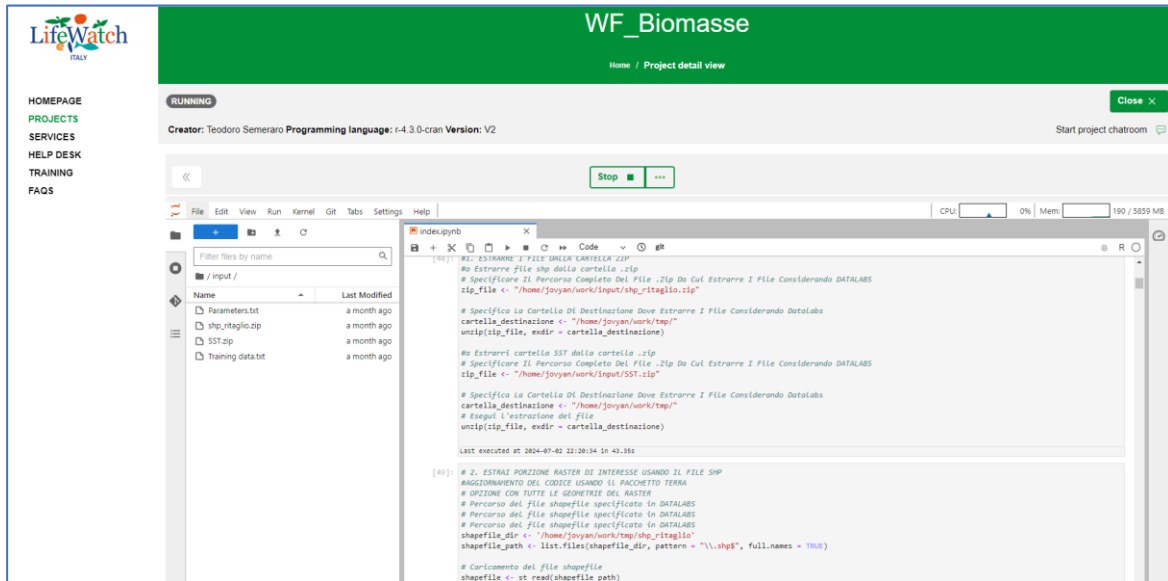


Figure 8. DataLabs development screen for analysis processes.

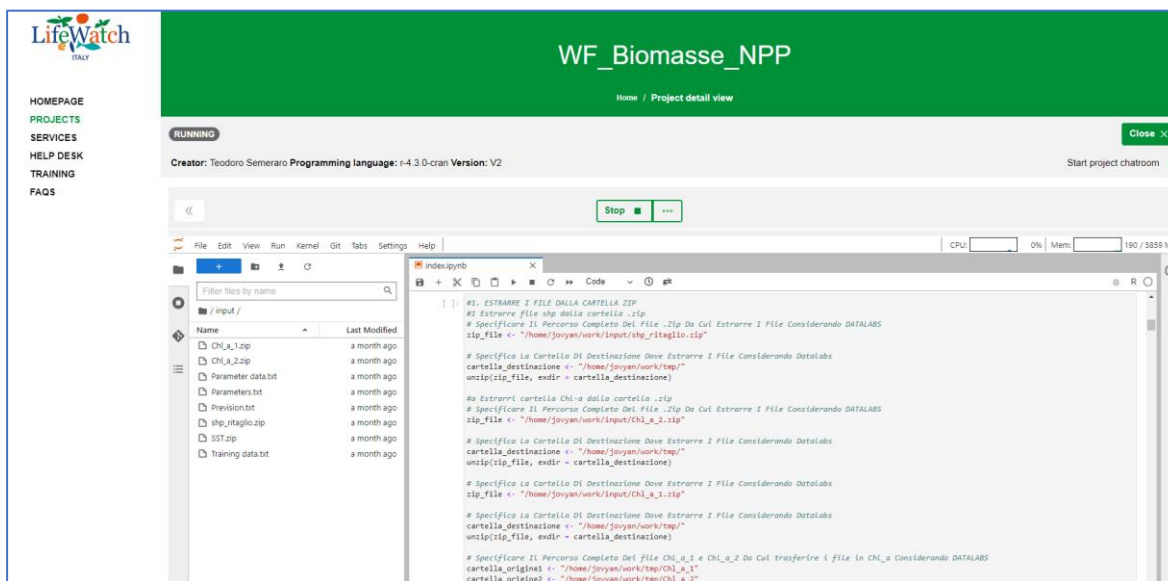


Figure 9. DataLabs development screen for analysis processes.

The fourth analysis process (Figure 10) investigates the spatial correlation between SST and Chl-a using remote sensing imagery and develops a temporal projection of Chl-a in consideration of the SST variability. Therefore, starting from the two datasets of SST and Chl-a imagery of remote sensing already existing, it is possible to calculate the spatial correlation of Chl-a in relation to SST for each corresponding frame and build a time series of Chl-a and SST for the study area. Applying the SARIMA-X model to Chl-a and SST time series, it is possible to forecast the projection of Chl-a overtime in function of SST variation.

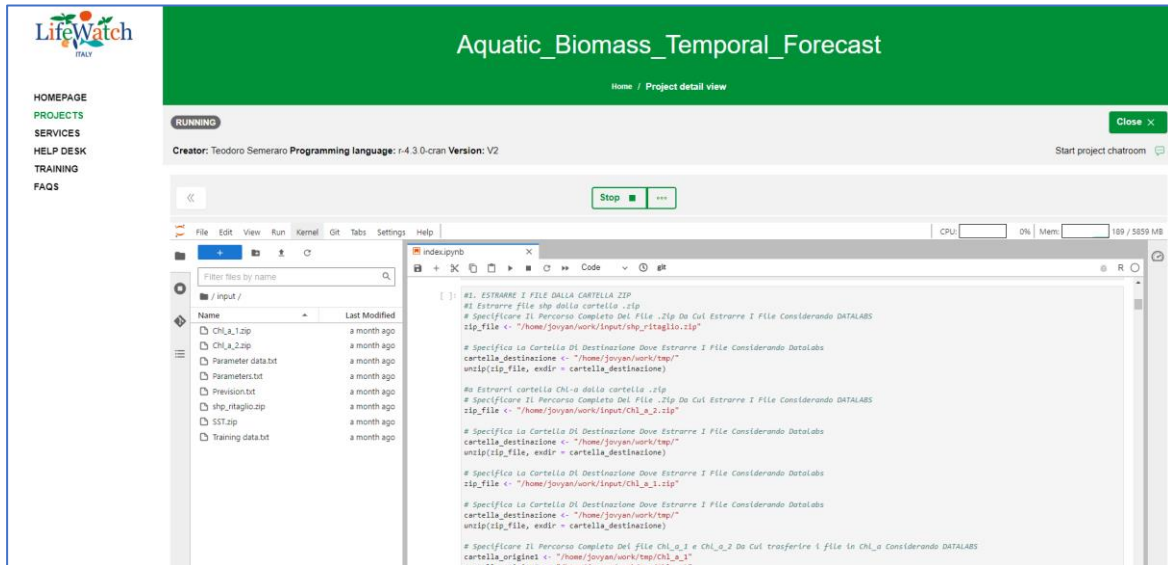


Figure 10. DataLabs development screen for analysis processes.

The individual process analyses have been developed in DataLabs with a modular structure, in accordance with the principles of Biomass VRE development. Each analysis process is composed of different blocks. Essentially, each block tool is represented by a discrete program code, and it is developed for the performance of a single operation with specific actions of uploading input and writing output files. For a process analysis to work like a workflow, the individual block tools are executed in sequence, with the output of one block tool serving as the input for the next block tool. This approach allows the output file from each block tool to be stored in a temporary or output folder within DataLabs. This allows a process analysis to be automated without the need for user intervention during process execution. The user simply needs to upload the files to initiate the analysis.

The main actions or tools of the analysis processes are (Figure 11):

- Extract the portions of the remote sensing imagery of interest related to the study area using a polygonal vectorial file with shp format.
- Define a good regression model (e.g. Random Forest) representative of the relation of Chl-a and NPP with the SST. The model has to be populated using field data of Chl-a, NPP and SST.
- Apply the regression model to the remote sensing imagery to produce new products of Chl-a and NPP.
- Build the time series of SST, Chl-a and NPP and perform statistical analysis.
- Realise temporal projections of time series to build future scenarios.

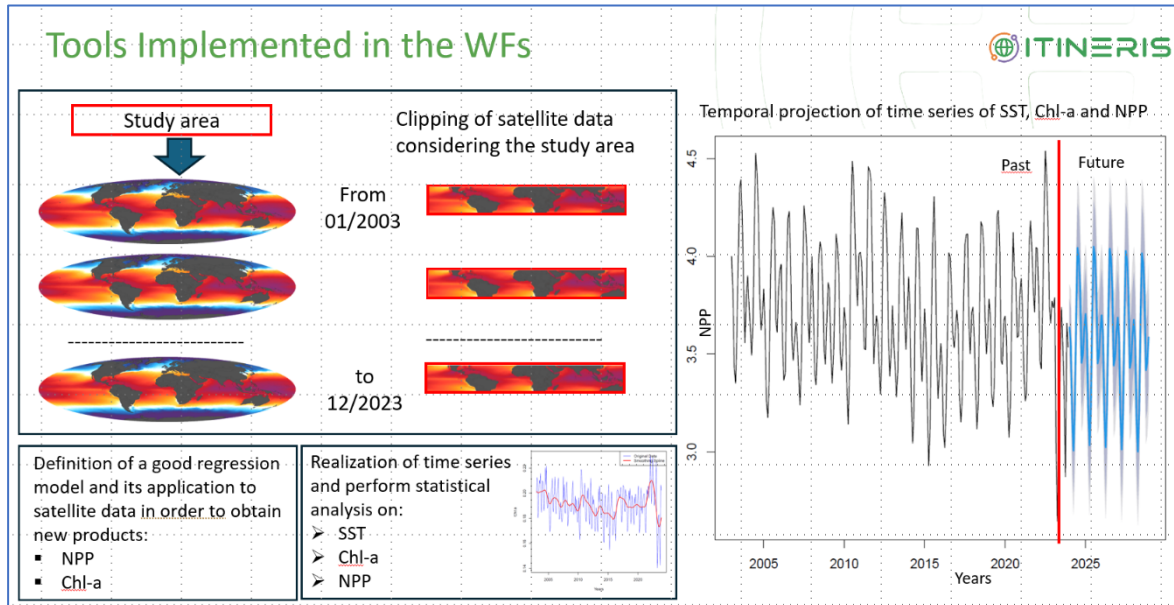


Figure 11. Representation of the tools of the workflow developed in the prototype developed.

The analysis processes begin by uploading table files with field data to populate the algorithms, shape files to define the study area and raster data derived from remote sensing imagery to obtain the Chl-a and NPP concentrations relative to the study area. The analysis processes culminate with the development of a time series of Chl-a and NPP, along with their temporal projections.

The workflow analysis results which can be downloaded include:

- Plots of time series of Chl-a and NPP with application of smoothing (PDF file). Smoothing is usually done to help us better see patterns and trends in time series.
- Tables of time series of Chl-a and NPP data with statistical information referred to the average value, standard deviation and moving average (CSV file).
- Details of the regression model applied such as Rsquared, RMSE and MAE (TXT File).
- Projections of time series in time to create future scenarios of variation for NPP and Chl-a with information on residuals (PDF File).

Part of the same analysis processes from above will be adapted for terrestrial biomass studies using the basic structure already completed. The analysis process for terrestrial biomass can be considered as completed, as it is enough to just modify the algorithms/code of algorithms to accommodate the new data used.

Currently, the process analysis codes have been developed in the R language but will be translated into the Python language.

4. CASE STUDY DEVELOPMENT

The Biomass VRE comprises the creation of case studies that are aligned with the workflow implementation. Furthermore, the case studies will be developed in conjunction with the activities of Work Package 6, which include the acquisition of data resources concerning the metabolism and spatial utilization of organisms in response to global warming.

Currently, the prototype development was realized using a case study and below, its abstract was reported.

4.1. Phytoplankton production response to global warming: estimating patterns of change across latitudes

Oceanic phytoplankton contributes to fundamental energy fluxes in marine ecosystems that support biomass and help mitigate climate change. Indeed, phytoplankton forms the base of the food chain in aquatic ecosystems and is responsible for absorbing CO₂ from the atmosphere. Phytoplankton photosynthetic and respiration processes are both influenced by the temperature; therefore, global warming is expected to have a strong impact on biomass, primary production and carbon dioxide absorbance.

This case study aims at developing an approach to study phytoplankton dynamics in the ocean, by combining data from field measures and remote sensing imagery. This case study has been developed to test the functionalities of the Aquatic Biomass workflow, included in the Biomass VRE. The methodology applied the random forest algorithm to study the pattern of existing data on Sea Surface Temperature (SST) and Chl-a. The Chl-a is used as a proxy for phytoplankton biomass production. The algorithm was then applied to monthly SST MODIS imagery to mapping Chl-a concentration in time, from 2003 to 2023. The analytical process was developed in R program on DataLabs by LifeWatch Italy (Figure 12).

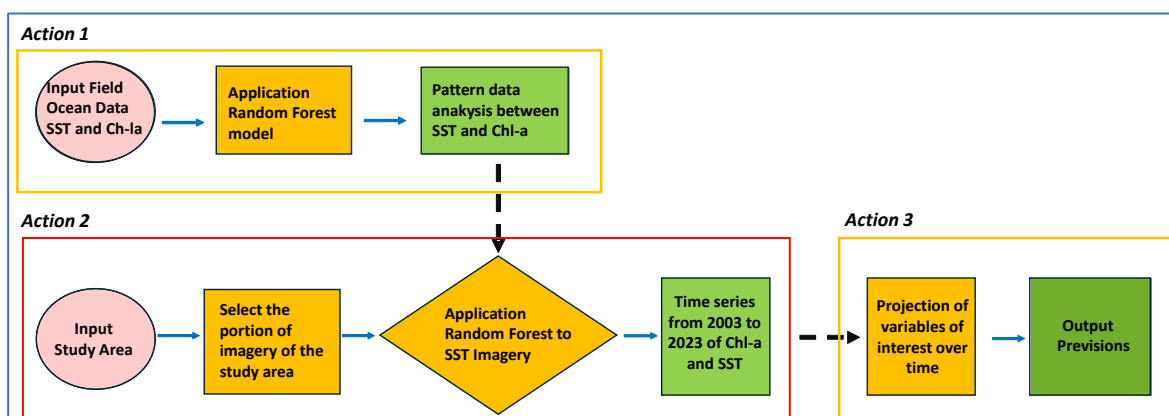


Figure 12. The following example illustrates the workflow analysis developed by DataLabs for the case study.



The results showed that the random forest, training with field ocean data, gave good representability of the capacity of the Chl-a to be described by the variation in SST.

The time series of SST showed a positive trend, whereas the Chl-a time series, build with application of random forest to SST MODIS imagery, showed a decreasing trend. The moving average applied to the Chl-a time series result showed two decreasing peaks between 2009-2010 and 2014-2016. Mainly, the variations of these peaks can be linked in time to the El Niño-Southern Oscillation (ENSO) process that affects the Pacific Ocean with the increase of sea temperature. The ENSO has a strong effect on the short-term variation of Chl-a, but less on the long-term variation.

This methodology highlighted the relation of the SST and Chl-a and can be used to estimate the potential impact of the processes that act also at a larger scale. The integrated approach of field and remote sensing data can help researchers working with field data or in the laboratory to extend their research to higher spatio-temporal scales.