Final project report

Machine Learning Model for the monitoring and detection of Harmful algal blooms

[BLUEDATA]

M3.2.DOC PROF/F/064/2020

Grant by

Fundo Regional da Ciência e Tecnologia (FRCT)

Coordinator Institution

[University of Azores]

Advisor:

Prof. Dr. Ana Maria de Pinho Ferreira da Silva Fernandes Martins

By

Dr. Issah Nazif SULEIMAN

November/2020 - [May/2022]

Scientific Research - [date/version]

14th July 2022

ACKNOWLEDGEMENTS

This research was supported by The Fundo Regional da Ciência e Tecnologia (FRCT) under grant number M3.2.DOC PROF/F/064/2020 and carried out at Eyecon group Software Ltd.

I would like to express my immense gratitude to the Lord most high (Allah) who has brought me this far and keeps blessing me with his favour and multitude of blessings. Thanks to Ana Maria de Pinho Ferreira da Silva Fernandes Martins who after my countless bothering as an advisor never gave up in her guidance and expert advice, Miguel Correia I should admit there are quite a number of people who have been very helpful in sharing their knowledge of the research, especially my oceanography team members. Last but not least, I would like to say this research wouldn't have been possible without the financial support of FRCT. The report should be read as the researchers input not necessarily as the thought of FRCT nor the individuals mentioned above.

Eyecon- group Software Ltd

Eyecon Group is a young but very successful startup company located in the Azores, Portugal. At Eyecon, we believe in keeping an eye on the environment and protecting it.We also believe that every data has a unique message which only when understood can solutions to industrial problems be formulated.Along the way, we developed applications such as **SOILRISK**, a predictive model application for landslides, **RIVERFLOW** an early warning system for floods, and we are creating **HABTRAIL**, a predictive model for toxic algal blooms…

TABLE OF CONTENT

1. EXECUTIVE SUMMARY

The continuous infestation of algae in coastal areas around the world has become a concern to stakeholders and researchers. As a result, scientists continuously work towards building much more improved monitoring techniques to measure the growth and the distribution of algae as a way of detection to give an early alert or warning to stakeholders to carry out mitigative measurement.

Harmful Algal Blooms (HABs) are a major environmental, economic, and public health threat to marine life and coastal communities. HABs occur naturally, but in recent years nutrient-rich agricultural runoff, contamination via ship ballast water, coastal aquaculture farms, and climate change appear to have contributed to an increase of HABs worldwide. Therefore, the detection of HABs is vital to preventing major environmental and human accidents. Accurately tracking and forecasting HABs can dramatically reduce their impacts on human health, fisheries, and economies. But growing pressure to cut costs means new ways must be found to achieve these goals.

In this research, we propose a deep learning algorithm for monitoring and prediction of HABs using a statistical dependency algorithm to quantify the association between chlorophyll-a and the cell counts or abundance and [2] a cell counts or abundance persistent. Our proposed data labelling approach uses a conditional probability Analysis approach which reveals a quantitative threshold above which and with the number of days of persistence or continuous observation of the cell counts, an alert base warning signal is given.

1.1.PROJECT RATIONALE AND OVERALL OBJECTIVE

Monitoring and detection of algal blooms over the years have seen much interest among researchers with the goal of improving the already available techniques in HABs detections. Remote sensing (RS) techniques among other approaches have become readily available for the study of large-scale biological processes and endangered environments including coastal areas. Earth observation data (EA) or remote sensing data has proven to be very useful in mapping algal blooms. Research reveals that the RS approach to algal detection provides a wide coverage zone with a high temporal resolution which in turn could be efficient for long-term monitoring.

The conventional methods of HAB detection from literature are mainly the optical and the physical approach. These techniques are used based on a binary tree question of Yes or No (Does the bloom change water colour or not). An optical approach is then deployed if the answer is yes while the physical forcing approach is used if the answer is No. The optical techniques (Empirical, Analytical and Semi-Analytical algorithms) have the ability to distinguish some blooms types in case 1 water although it appears to be difficult in coastal areas due to sediments and CDOM. The algorithms for the HABs detection using this technique use absorption, backscattering, relative patterns and spectral shape. The physical forcing technique defines a relationship between blooms and their physical features. It uses some environmental factors such as SST as a primary attribute such as fronts, unwillingness and wind events. It should be said however that it's difficult to view the extent of blooms using the physical forcing technique as no direct bloom detection is generally the case.

Almost all the existing ML algorithms use either Chl_a or the number of HAB cell count as a sole quantitative determinant of HABs and so build their algorithm using a known threshold for either Chl_a or HABs. With the

research-supported threshold and some other relevant pigmentations, an ML model is built to classify a suspected event as either HABs or No HABs. Although Chl_a as a biomass indicator and HABs cell abundance are both very important variables in the monitoring of HABs, it should also be noted that a high concentration of Chl_a does not always indicate the existence of HABs and vice versa. This is why algorithms built using these deterministic variables as discriminant variables give more false positives. HABtrail, therefore, is a bridge between the use of remote sensing data combined with an optimised ML algorithm based on the data labelling to classify and the suspected regions as HABs present or not.

Our proposed algorithm takes two different forms; [1] the first interface is referred to as the Satellite interface which takes into consideration satellite imagery data of algal growth measurement or algal growth triggering or favouring factors from the MODIS satellite which are then used to build a neural network to predict the occurrence of HABs. [2] The second interface (i.e. Citizen Scientist interface) is built on a multi-algorithmic scheme which takes in imagery data from digital cameras for example; phones, drones, and tablets. Using these algal growth variable measurements [1} and the digitally acquired image [2], our models are able to predict the temporal severity of HABs or the severity of HABs.

Project Objective Summary

i. To build and implement an effective ML algorithm that can be used to monitor and predict HABs in advance.

ii. Build free HABtrail- Citizen App available to the general public that gives a near to real-time prediction of a suspected region.

iii. Build a HABtrail-Web App available to civil authorities, academicians, and coastal and sea management agencies to help monitor a large area of a potential occurrence of HABs.

iv. Have an International Award with our Research through

   a. International Research Competition
   b. International Tender
   c. Journal publication
   d. Conference Paper
   e. Workshop Paper

1.2.Summary of progress

This research was divided into different stages from WP1 to WP4 and to ensure our research meets its objectives, an intensive study was carried out on various sensors and algorithms. We targeted some processing skills and chose an appropriate scheme for our said proposed algorithm based on some prior knowledge of HABs and Statistical approach application. We realised most of the false positives that the already existing remote sensing approach gives are as a result of either a poor estimation of some pigments including Chl_a or overestimation of such pigments.

After months of studies, we then proceeded into site selection due to the lack of in situ data in the Azorean Archipelagos, our algorithm was built on the south coast of Portugal (Algarve). The Algarve region occupies the southern part of Portugal with its south and west bounded by the North Atlantic Ocean. And considering the significant role of the coastal ocean in economies and society coupled with its location in the Atlantic Ocean, it was chosen as our study area. Data availability and acquisition happen to be common challenges researchers face in building an alert system. Nevertheless, we made use of the COPANUS system of the University of Azores (Department of Oceanography) which is a database platform which extracts data from MODIS satellite with a user-friendly and defined pre-processing feature to allow us make the necessary adjustment to suit our research objectives.

The early days of HAB detection using remote sensing make use of spatially isolated and single satellite data samples. However, there are various other approaches to the detection of HABs which take into consideration many satellite sensors and bands. Since the current remote sensing technology is unable to capture the ions of nitrate and phosphate, the amount of chlorophyll concentration is used as a proxy because photosynthetic activities are visible

under satellite imaging. It was therefore without a doubt that most of the algorithms in monitoring HABs both estimate Chl_a concentration and use that as a proxy variable for lake water and ocean. Nevertheless, we argue that Chl_a concentration can provide a good early indication of exceptional algal blooms which does not always imply toxicity. We also observed a high predictive relationship between Chl_a in the Algarve region and HABs using a stepwise regression model.

Using the acquired data and the knowledge gained from the pigment relationship between HABs and other important pigments using the MODIS data, we built our first ML model and accessed the model performances using data acquired using different satellite (Sentinel 2) and made some very important observation on why most of the existing early warning system built gives more false alarms.

1.3. Significant achievement since the last report

The identification and prediction of HABs has always been a concern in the scientific community which is as a result of the interaction between both biological and physical processes that causes HABs. In order to bridge this challenge, the use of remote sensing data provides significant information about the indicative pigments of HABs. The main causative factors of HABs are temperature, turbidity, wind, water current, light level and nutrient level and other pigmentations. Nutrients- mostly nitrogen and phosphorus are the most significant causes of HABs.

Arguably, the already existing monitoring system and early warning interfaces are mostly built on either using only Chl_a as HAB indicator or using HABs

cell abundance threshold to classify a suspected region as HAB or not. Although both Chl_a and the number of cell counts are key in the monitoring of HABs, it should also be noted that a high concentration of Chl_a is no indication of HABs visa via the use of a fixed research threshold of the number of cell counts.

We analysed the variable relationship and their corresponding predictive impact or importance to the occurrence of HABs and got some important insights.

a. Using only Chl_a concentration popularly known as the biomass indicator as a dependent variable to predict and monitor the occurrence or non-occurrence of HABs is not always a good algorithmic scheme and could lead to more false positives.
b. A high/low concentration of Chl_a is not enough to characterise a given region as HABs.
c. Using only the number of Cell abundance without considering the effect of the Chl_a and the persistence equally gives more false alerts.
d. In building a more reliable Machine learning model to monitor and predict the occurrence and non-occurrence of HABs it is important to use all these three variables in the data labelling (Chl_a, cell count/abundance and the persistence of the cell counts).
e. The relationship between Chl_a and SST although most thought to be linear, it was firmly seem to have a nonlinear relationship.
f. The availability of in situ data is always a great limitation in building a more robust machine learning algorithm.

1.4.Potential Impact and exploitation of results

The objective of this research was to build a more optimised Machine Learning algorithm that serves as a bridge between Remote sensing approach and Machine learning approach. With prior knowledge of the limitations associated with both schemes, HABtrail was built as an optimised tool for a more reliable predictive model for the monitoring and prediction of HABs.

While the already existing scheme discriminates or classifies an event as HABs or not based on either the concentration of Chl_a or the number of cell counts in a given region, HABtrail works in a different way. Our data labelling approach is based on three variables rather than the one used in both the ML approach and the Remote sensing approach.

1. The concentration of Chl_a
2. The number of cell counts or abundance (HAB)
3. The persistence of the number of cells counts

Using variables (1) and (2) we determine a quantitative probabilistic threshold such that, given that a known concentration of Chl_a has occurred what chances are there that, that concentration of Chl_a can trigger or cause a HABs to occur. With this quantitative value, we then carry out the first stage of our data labelling and then follow that up with the next step. Given that a region has been classified as HABs with a known concentration of HABs, we define a persistence feature to ensure that not all observations with a high value of cell counts are classified as HABs but those with a persistence occurrence.

As a result of this research, we have had two (2) important breakthroughs with which we have:

- Three (3) articles are in the pipeline for publication.
- Built a publicly available HABs detection App that the general public can use easily to access a suspected region for the existence of Dinoflagellate.
- We also know that Chl_a and SST are negatively correlated but they are not linearly related as theorised by most literature in the region of Algarve.
- We have a web interface that allows users to select their region of interest in access:
    i. The occurrence of HABs
    ii. The distribution of HABs

These findings, if improved as recommended, will have a great contribution in the monitoring of HABs and provide a much more reliable early warning to authorities and agencies responsible for coastal protection and management to have precision and saving life, economies and tourism.

1.5. Further opportunities for Research

HABtrail has shared both scientific and technical insight into some research areas where it could be applicable if further attention is given with the right set of teams. Amongst these opportunities that the study of HABtrail has shown to be useful in are areas such as:

01. Carbon harvesting
02. Ocean event mapping
03. Risk Assessment and Evaluation
04. Shellfish Harvesting

1.6. Summary of progress against Objectives

Summary of table

Table 1. Describes the work progress and workflow of our research work with their corresponding actions

	Date	Action
Phase 1/ WP1	16-11-2020 11-01-2021 28-02-2021	Task 1: Compilation of main RS techniques used to detect HABs Task 2: Compilation of main algorithms used to detect HABs Task 3: Learning about main processing RS techniques
Phase 2/WP2	01-04-2021 14-06-2021	Task 1: Site selection and main characteristics Task 2: Download of RS data for these regions and HAB bloom periods
Phase 3/WP3	01-07-2021 01-09-2021 11-10-2021	Task 1: Determination of main RS algorithms to be applied in this study Task 2: Processing of chosen HAB algorithms Task 3: Validation against in situ data in the case study regions
Phase 4/WP4	01-08-2021	Task 1: Copernicus Masters Challenge Task 2: Portuguese Space Challenge Task 3 - Evaluation of main results Task 4 - Final research report

		Task 5 - Dissemination of research study results in specialised forums Task 6: Final research report Task 7: Submission of an article in an International Peer Reviewed Journal

2. DEVIATION FROM PREVIOUS DESCRIPTION OF WORK

To meet the objective of this research, there have been some needed adjustments that were made to ensure we meet these goals. Some of these changes are:

i. Change in the study area from the Azorean Archipelagos to the Algarve Region
ii. The building of a public available App and make it free
iii. Our active involvement in the Copernicus Masters Challenge although yield a great success story but also caused us to deviate from our already stipulated deliverable schedule.

2.1 Achievements and Key dissemination Activities

The success of any research work is measured by many factors relative to the audience and industry however, HABtrail has seen its success combining many industrial recognition, Awards and approval from academics through to industrial solution tools.

As a partner to the University of Azores, our first success came when both the Department of Oceanography and Agriculture expressed their immense interest

to have a research partnership with our application and consultation. The relationship between HABtrail team and the University of Azores has since been a very strong research work that has created a bond of confidence in the University of Azores to trust our findings and consult when they need be.

Copernicus Masters Challenge is an EU organised competition that sees innovative minds and researchers bringing their innovation on stage and battling to the last for the crown of their solution being accessed and recognized as the best innovation each year. HABtrail went through each stage of the competition and came out the ***Winne***r of the 2021 edition visa via the Winner of the Portuguese Space Challenge 2021. We got a special guest invitation to be a part of the Atlantic innovation week which was held on 14th – 17th March in the Azores. We were also part of the Organizers of the ***Copernicus Masters Webinar: Innovation and Entrepreneurship for the Blue Economy***.

HABtrail took the front page of many newspapers and journals across Europe with its wave of success reaching the very door of the government house. A wave of success that even the house of parliament has to recognize, this got HABtrail an honorary acknowledgement from the house of parliament (https://videoalra.blob.core.windows.net/alraa-assets/167dee0d-89ec-4625-8b33-2dff93d958bd/video.mp4).

2.2 Table of Achievements

No.	Achievement
1	Acceptance to be a part of the Oceanography group in the University of Azores
2	Copernicus Masters Challenge Winner
3	Portuguese Space (PT-Space) Partner
4	Copernicus Masters Webinar Organisers
5	Successful building of HABtrail (Citizen App)

3. DISSEMINATION TOOLS /SOLUTIONS BUILT

There are two (2) dissemination HABtrail tools built, both of which are in active and prototype form now and undergoing robust optimization to make it available to the general public and to the agencies and authorities involved. These tools are both built on a state of the art Deep Learning architecture to serve as a health risk reducing tool and they are the Citizen App and the Web App.

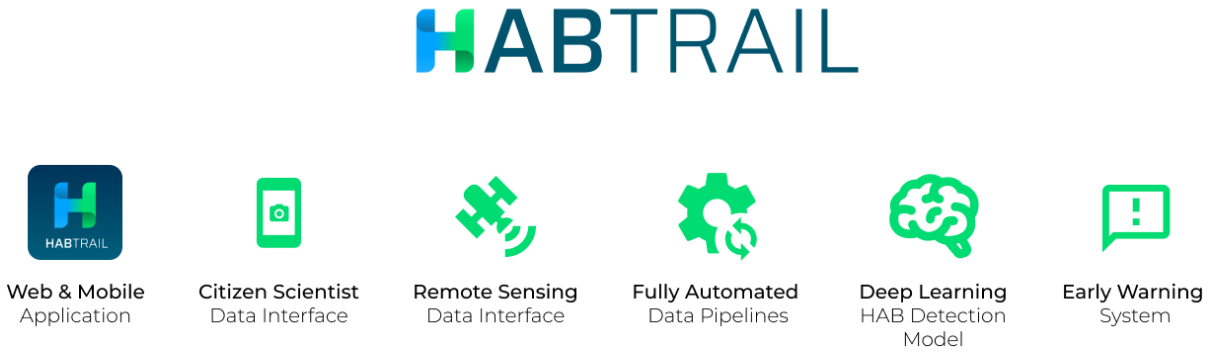

Fig: HABtrail automation tool structure

3.1 HABtrail (Citizen App)

The Citizen App is a free public available tool which when available to the public can easily be downloaded on smart devices namely phone and tablets. The user will be required to snap a suspected coastal surface based on some simple and user understanding instructions through the App which will then be uploaded instantly to give a near to real time prediction with a probabilistic value to give the user a fair assessment scale to help advice the user on the potential toxicity of the suspected coast.

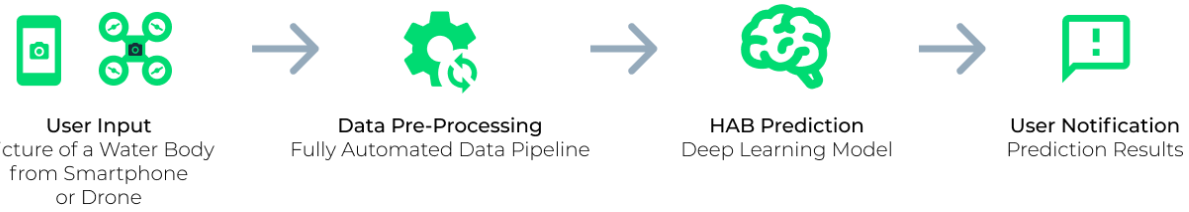

Fig: The above info-graphic shows the automation processing scheme of HABtrail in 4 simple steps.

3.2 HABtrail(Web App)

The Web App, although equally built on a neural network like its sister App (Citizen App), works in a slightly different way. It provides the user the chance to select an area of interest using a specific satellite through an automated scheme to monitor and predict the occurrence of HABs. The Web App unlike the Citizen App will only be accessible through a subscription. With the prototype version available, users will be able to have a fair understanding of the distribution of HABs and also be able to play with some pigments to have a

deeper understanding of some descriptive relationship between some variables of interest.

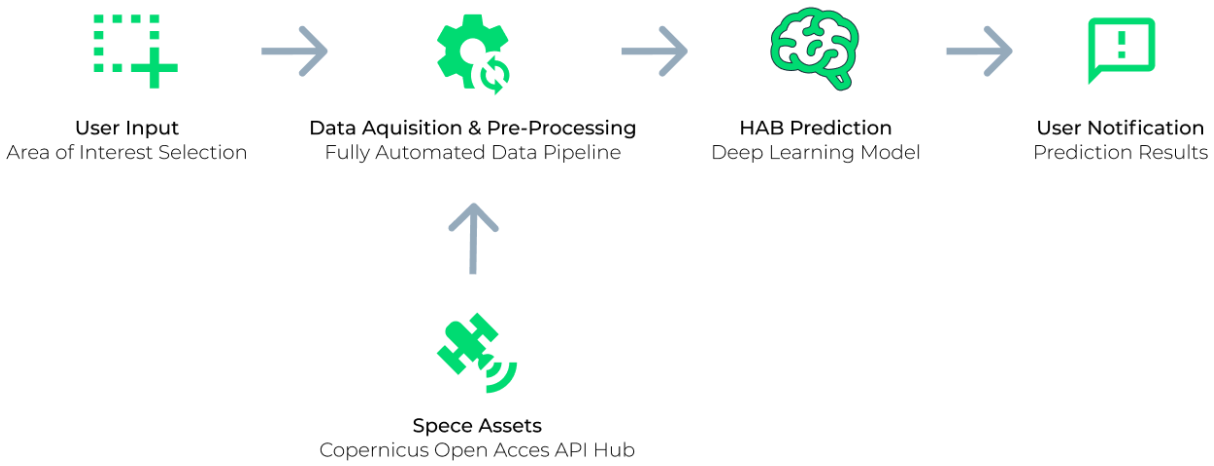

Fig: the figure above shows how the HABtrail Web App works from the area selection to an automated data acquisition and preprocessing which is passed through the built in ML algorithm to monitor and predict the occurrence of HABs.

4. SCIENTIFIC PUBLICATIONS

Journal papers

With two key findings through this research, presenting a major project results at the highest scientific standards and disseminating them to the scientific community is imminent. Since such publications take a long time-to-publish period due to the academic scrutiny which aims at substantial, matured and empirically verifying results and thus requires the acquisition of more in situ data to re-assess and confirm our findings. At the presents, only abstracts have been submitted and full papers are in full preparations. Our relevant target journals includes:

i. Fish and Fisheries

ii. Marine and Petroleum Geology

iii. Water Resources & Economics

iv. Water and Environment Journal

v. Environmental & Resources Economics

vi. Journal of Water Resources Planning and Management

vii. Water Resources Management

5. PRESS RELEASE

Updated list of press releases.

- Diário Insular
- Visão (Website)
- Correio dos Açores
- Tribuna das Ilhas
- Congratulatory acknowledgement from Assembleia Legislativa da Região Autónoma dos Açores