**A close up of a logo

Description automatically generated**

**LD7083**

**Computing and Digital Technologies Project**

**Dissertation**Marine Water Quality Analysis and Predicting Optimal Fishing Grounds Using Machine Learning and Remote Sensing Technology.

by

W22072142

JOHN RAJ THANKARAJ

Supervisor Name: Rejwan Bin Sulaiman

A Dissertation

Submitted to Department of Computer and Information Sciences

Northumbria University

In Partial Fulfilment of the Requirements

For Master of Science in Big Data & Data Science with Advanced Practice

submission date: 15/01/2025

Declaration

I declare the following:

1. that the material contained in this dissertation is the end result of my own work and that due acknowledgement has been given in the bibliography and references to ALL sources be they printed, electronic or personal.
2. the Word Count of this Dissertation is **14981**
3. that unless this dissertation has been confirmed as confidential, I agree to an entire electronic copy or sections of the dissertation to being placed on Blackboard, if deemed appropriate, to allow future students the opportunity to see examples of past dissertations. I understand that if displayed on Blackboard it would be made available for no longer than five years and that students would be able to print off copies or download. The authorship would remain anonymous.
4. I agree to my dissertation being submitted to a plagiarism detection service, where it will be stored in a database and compared against work submitted from this or any other Department or from other institutions using the service.   
   In the event of the service detecting a high degree of similarity between content within the service this will be reported back to my supervisor and second marker, who may decide to undertake further investigation that may ultimately lead to disciplinary actions, should instances of plagiarism be detected.
5. I have read the UNN/CEIS Policy Statement on Ethics in Research and Consultancy and I confirm that ethical issues have been considered, evaluated and appropriately addressed in this research.

**Signature : JOHN RAJ THANKARAJ**

**Date : 15/01/2025**

Acknowledgements

I would like to express my sincere gratitude to Professor **Rejwan Bin Sulaiman**, his advice and assistance have been crucial to finishing this study. His knowledge, tolerance, and support have greatly impacted my academic career. I am incredibly grateful for his guidance, as his insightful criticism and constant assistance have been a continual source of inspiration.

I want to express my sincere gratitude to my friends for their constant company and support. The journey has been made pleasant and pleasurable by unending tolerance, encouragement during difficult times, and joyous presence during glorious moments. I appreciate your unwavering support and belief in me, even when I doubted myself.

I will always be thankful for my family's unwavering love and support. The motivation behind my accomplishments has come from your support and faith in my skills. My dreams have been made possible by your understanding, tolerance, and sacrifice, and I am incredibly grateful that you are in my life.

Table of Contents

[Declaration 2](#_Toc187746444)

[Acknowledgements 3](#_Toc187746445)

[Abstract 8](#_Toc187746446)

[Introduction 9](#_Toc187746447)

[1.1. General 9](#_Toc187746448)

[1.2. Problem statement 10](#_Toc187746449)

[1.3. Motivation of this study 10](#_Toc187746450)

[1.4. Aim 10](#_Toc187746451)

[1.5. Objective 11](#_Toc187746452)

[1.6. Scop 11](#_Toc187746453)

[1.7. Research Question 11](#_Toc187746454)

[1.8. Research Contribution 12](#_Toc187746455)

[Literature review 13](#_Toc187746456)

[2.1. Introduction 13](#_Toc187746458)

[2.2. Incorporating Summary Table 24](#_Toc187746459)

[2.3. Conclusion 27](#_Toc187746460)

[Research Methodology 28](#_Toc187746461)

[3.1. Study Area 28](#_Toc187746463)

[3.2. Influencers of Study Area 29](#_Toc187746464)

[3.3. Data & Data collection Instruments 30](#_Toc187746465)

[3.4. Data Acquisition Process 32](#_Toc187746466)

[3.5. Analysis Tools and materials 33](#_Toc187746467)

[3.6. Data Analysis and techniques 34](#_Toc187746468)

[3.7. Limitations 36](#_Toc187746469)

[Data Analysis and Implementation 37](#_Toc187746470)

[4.1. Introduction 37](#_Toc187746472)

[4.2. Procedure and Execution 39](#_Toc187746473)

[4.3. Libraries and functions 41](#_Toc187746474)

[4.4. Data Preprocessing procedures 44](#_Toc187746475)

[4.5. Implementation of integrated analysis system 44](#_Toc187746476)

[4.6. Python implementation and derived task 45](#_Toc187746477)

[4.7. Validation process 48](#_Toc187746478)

[4.8. Implementation of Machine Learning Technology Random Forest for Potential Fishing Zone (PFZ) Prediction 49](#_Toc187746479)

[Results 50](#_Toc187746480)

[5.1. Introduction 50](#_Toc187746482)

[5.2. Marine water quality Indicator visualisation 50](#_Toc187746483)

[5.3. Evaluating Marine Water Health: Critical Quality Indicators 58](#_Toc187746484)

[5.4. Scatter Plot Visualization of Marine Parameter Interactions 59](#_Toc187746485)

[5.5. Extraction of scatter plots for Ideal ranges for marine water quality parameters 64](#_Toc187746486)

[5.6. Summary of Result 67](#_Toc187746487)

[Discussion and Evaluation 68](#_Toc187746488)

[6.1. Evaluation of existing studies with this study 68](#_Toc187746490)

[6.2. Interpretation and evaluation of main key findings 69](#_Toc187746491)

[6.3. Evaluation of marine water Health with Ideal range 70](#_Toc187746492)

[6.4. Prediction of potential fishing zones 70](#_Toc187746493)

[6.5. Contributions of this study 71](#_Toc187746494)

[6.6. Challenges of this study 72](#_Toc187746495)

[6.7. Limitations of this study 72](#_Toc187746496)

[Conclusion and Recommendations 74](#_Toc187746497)

[7.1. Conclusion 74](#_Toc187746499)

[7.2. Limitations 75](#_Toc187746500)

[7.3. Recommendations 76](#_Toc187746501)

[7.4. Final summary 77](#_Toc187746502)

[Reference 78](#_Toc187746503)

[Appendix A 81](#_Toc187746504)

[Student Project Approval Form 81](#_Toc187746505)

[Appendix - B 86](#_Toc187746506)

[Source code link: 86](#_Toc187746507)

[Appendix - C 87](#_Toc187746508)

[Meeting Logs 87](#_Toc187746509)

List of Tables

[Table 1. Incorporation of Literature Review 27](#_Toc187674572)

[Table 2. Scatter Plot Visualization of Marine Parameter Interactions 65](#_Toc187674573)

List of Figures

[Figure 1. South Coast of UK 28](#_Toc187674574)

[Figure 2. Auto download data in python programme 32](#_Toc187674575)

[Figure 3. Manual downloading data interface 33](#_Toc187674576)

[Figure 4. Flow chart of Methodological implementation 38](#_Toc187674577)

[Figure 5. Interface of SeaDas Programme 40](#_Toc187674578)

[Figure 6. List of Used libraries in python 41](#_Toc187674579)

[Figure 7. Authentication procedure in earthaccess used python for access data online 46](#_Toc187674580)

[Figure 8. Data downloading with using Python 47](#_Toc187674581)

[Figure 9. Validation process by SeaDAS Programme 48](#_Toc187674582)

[Figure 10. Random forest implementation in Python 49](#_Toc187674583)

[Figure 11. Chlorophyll-a after image process from SeaDAS 51](#_Toc187674584)

[Figure 12. Chlorophyll-a pixel frequency after image process from SeaDAS 51](#_Toc187674585)

[Figure 13. Final output for visualisation of chlorophyll-a concertation 51](#_Toc187674586)

[Figure 14. SST after image process from SeaDAS 52](#_Toc187674587)

[Figure 15. SST pixel frequency after image process from SeaDAS 52](#_Toc187674588)

[Figure 16.Final output for visualisation of SST variation 52](#_Toc187674589)

[Figure 17. PIC after image process from SeaDAS 53](#_Toc187674590)

[Figure 18. PIC pixel frequency after image process from SeaDAS 53](#_Toc187674591)

[Figure 19. Final output for visualisation of PIC concertation 53](#_Toc187674592)

[Figure 20. POC after image process from SeaDAS 54](#_Toc187674593)

[Figure 21. POC pixel frequency after image process from SeaDAS 54](#_Toc187674594)

[Figure 22. Final output for visualisation of POC concertation 54](#_Toc187674595)

[Figure 23. Turbidity (Kd\_490) after image process from SeaDAS 55](#_Toc187674596)

[Figure 24 Turbidity (Kd\_490) pixel frequency after image process from SeaDAS 55](#_Toc187674597)

[Figure 25. Final output for visualisation of Turbidity (Kd\_490) 55](#_Toc187674598)

[Figure 26. Relationship between chlorophyll-a and SST 59](#_Toc187674599)

[Figure 27. Relationship between POC and SST 59](#_Toc187674600)

[Figure 28. chlorophyll-a and PIC 60](#_Toc187674601)

[Figure 29. chlorophyll-a and POC 60](#_Toc187674602)

[Figure 30. PIC and SST 61](#_Toc187674603)

[Figure 31. Chlorophyll-a and Kd\_490 nm 61](#_Toc187674604)

[Figure 32. Kd\_490 nm and SST 62](#_Toc187674605)

[Figure 33. Kd\_490 nm and POC 62](#_Toc187674606)

[Figure 34. PIC and POC 63](#_Toc187674607)

[Figure 35. Kd\_490 nm and PIC 63](#_Toc187674608)

[Figure 36. Final output for visualisation of PFZ prediction 66](#_Toc187674609)

Abstract

However marine ecosystems are important to the economy and food supply, human activity and climate change are creating an increasing risk to them. By utilizing developments in remote sensing and machine learning, this study fills data gaps and promotes sustainable fisheries management by monitoring marine water quality and locating ideal fishing grounds. Utilizing remote sensing methods like satellite images, the project collects crucial water quality data. At the same time, machine learning algorithms—specifically, the random forest method—are used to find patterns and abnormalities in big datasets, improving data integration and decision-making.

The main goal is to identify the major environmental elements affecting the quantity and distribution of fish species. A prediction model is created to identify potential fishing grounds and suggests a practical method for precisely evaluating markers of marine water quality. By monitoring and managing water quality, the study uses statistical, geospatial, and time series analytics to better understand problems with water quality, pinpoint sources of contamination, and evaluate health concerns. For sustainable water management, this integrated approach helps create data-driven planning and early warning systems.

The study uses 10 years of data and MODIS satellite images to investigate the biological and geological features of the Cornish coast in the United Kingdom. The research analyses trends and correlations between water quality measurements using statistical and visual techniques including scatter plots and graphs. The results highlight how crucial it is to preserve water quality in order to sustain healthy marine ecosystems and give fishermen precise information on where to fish. This creative method indicates how machine learning and remote sensing could be combined to effectively manage resources and save marine ecosystems.

This study shows how well machine learning and remote sensing technologies work together to monitor marine ecosystems. The method supports sustainable fisheries and marine habitat protection by determining the best places to fish and evaluating the quality of the water, which helps make well-informed decisions for the long-term productivity and health of the ecosystem.

Chapter-1

# Introduction

## General

The world's economy and food chain depend on marine ecosystems, which are increasingly under stress from human activity and climate change. Sustainable resource management requires an understanding of these ecosystems and regular monitoring of their health. In addition to being costly and time-consuming to perform, traditional techniques of assessing the quality of marine water are sometimes limited in their coverage of time and location.

Improvements in machine learning and remote sensing technology in recent years have created new opportunities for precise and effective maritime monitoring. Techniques for remote sensing, such satellite photography and airborne surveys, can yield important data on water quality indicators like sea surface temperature, turbidity, and chlorophyll-a concentration. On the other hand, machine learning algorithms are able to find patterns, trends, and errors through processing big datasets from remote sensing and other resources.

The purpose of this study is to identify the best fishing locations by utilizing machine learning and remote sensing to evaluate the quality of marine water. We as humans want to create a strong and effective system that can assist with marine habitat preservation and sustainable fisheries management by combining these technologies.

Using remote sensing and machine intelligence to identify the finest fishing spots, as well as concentrating on marine water quality, marine studies research is extensive. The purpose of this work is to better understand marine water quality by combining oceanographic elements, Python, machine learning, and satellite data. It examines a number of studies published in the past five years that demonstrate how these technologies can evaluate data on water quality with the goal of improving precision and effectiveness via the integration of several data sources.

Recent research articles over the past five years have placed a lower priority on the volume of data required for in-depth analysis. Using additional historical data can increase the accuracy of the results. I use data from the past decade in my technique. Previous research sometimes only examined three or four oceanic components, therefore excluding a wider variety of ocean variables. The goal of my research is to use cutting-edge technology like machine learning to better assess marine water quality by including more ocean elements. This will improve the efficiency and accuracy of the study.

## Problem statement

The unpredictable behaviour of fish stocks and the difficulty of locating profitable fishing sites sometimes result in significant time and financial difficulties for fishing communities. A decrease in catch rates, an increase in operational expenses, and the potential for fishermen to lose motivation are all consequences of such conditions.

## Motivation of this study

Because of my close connections to the fishermen community, I have an intimate understanding of the challenges that fishermen encounter, particularly the inefficiency and unpredictability involved in locating productive fishing grounds. Seeing the time and money lost as a result of this issue has inspired me to look for a solution. I think that by enhancing the sustainability and efficiency of fishing communities' activities, technology that forecasts possible fishing zones (PFZs) will significantly enhance their quality of life. My personal goal is to assist people who have influenced my life via this study.

## Aim

My goal is to actively support conservation efforts and sustainable fishing practices to enhance marine resource management. Become more productive in this area. The long-term sustainability and health of our oceans and marine ecosystems will be promoted by fusing innovative scientific research with state-of-the-art technological developments.

## Objective

* Finding the fundamental environmental factors that significantly affect fish species' distribution and abundance in the ecosystem is the primary objectives.
* to create a predictive model that is both precise and effective in identifying areas that have a significant probability of fishing.
* To accurately and reliably assess marine water quality indicators, develop an effective strategy.

## Scop

* By eliminating other coastal locations for a more thorough investigation, the research effort concentrated only on examining the biological dynamics and geological formations of the Cornish coast in the United Kingdom.
* The study's scope was restricted to data and conclusions from the past ten years, with an exclusive focus on research completed in the recent ten years.
* In this case, no particular kind of fish was picked out or recognized as unique or exceptional.
* Only the accurate satellite data obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS) was used to conduct a thorough investigation of the environmental changes in the region.

## Research Question

* What are the main environmental variables that have a significant impact on the distribution and abundance of fish populations in this study?
* What are the specific ranges of important parameters that correspond to high-quality marine water in this study?
* How can the forecast of potential fishing zones be combined with water quality monitoring?

## Research Contribution

This research greatly advances my understanding by using a unique integrated systems approach to predict potential fishing zones (PFZs) with previously unheard-of accuracy. One of the most significant innovations is the use of an advanced multi-parameter study that explores the combined impacts of five important oceanographic parameters: chlorophyll-a, SST, turbidity (Kd-490), PIC, and POC. The intricate relationships controlling fish distribution patterns are successfully reflected by taking seriously a wide range of factors.

Additionally, by adding a temporal component to the prediction models, time series analysis takes responsibility for changing seasons and small fluctuations in fish behaviour as well as environmental elements. By adding a dynamic element that can adjust to shifting circumstances throughout time, this temporal component strengthens and improves the predictions.

Furthermore, PFZ predictive modelling has advanced significantly with the use of state-of-the-art machine learning Random Forest method. This study has overcome the drawbacks of conventional statistical methods by utilizing the strength of these sophisticated algorithms, producing forecasts that are more dependable and precise than those of the industry.

It is predicted that the fishing industry's use of advanced prediction models would help fishermen much in improving the effectiveness of their fishing operations. Fishermen may strategically optimize their operations and reduce fuel usage by utilizing this cutting-edge technology, which helps to lessen their ecological impact. Fishermen can compare their activities with sustainable fishing principles by using these accurate and comprehensive Potential Fishing Zone (PFZ) projections to make educated judgments. As a result, in addition to helping individual fisherman, the broad use of these innovative solutions also supports environmental conservation initiatives and the larger-scale promotion of sustainable fishing methods. By using precise PFZ forecasts, fishermen may practice environment friendly fishing, which will benefit marine ecosystems and advance the larger goals of marine conservation programs.

Chapter-2

# Literature review



## Introduction

While there is many various research being conducted in the broad field of marine studies these days, I have chosen to focus on a specific aspect of the subject. I begin by outlining my topic, which is the investigation of marine water quality and the use of machine intelligence and remote sensing technology to forecast the best fishing grounds.

In order to gather sufficient knowledge about remote sensing technologies and marine water quality parameters, I chose to compile combinations of three distinct sector studies, such as oceanographic parameters, satellite remote sensing data, and Python and machine learning. I also used various methodologies, such as machine learning and various deep learning models.

The primary goal is to gain a basic understanding of this topic, which can assist this study produce more accurate and understandable results. In the end, I made the decision to compile the articles from the previous five years that focused with this subject; each publication used a different set of parameters and methods to analyse the quality of sea water. I looked up several relevant papers and talked about them below.

1. The geographic information systems (GIS) and remote sensing (RS) may greatly improve the identification of Potential Fishing Zones (PFZs), which have a much greater fish capture rate than other locations. To identify PFZs in the Dakhla Oued Eddahab region of Morocco, the study effectively employed MODIS satellite data and analysed important oceanographic indicators, including Particulate Organic Carbon (POC), sea-surface temperature (SST), and chlorophyll-A (Chl-A) concentrations. MODIS (Moderate Resolution Imaging Spectroradiometer) is a crucial imaging capability on NASA's Terra and Aqua satellites, which monitor the Earth's atmosphere and surface to detect changes in the environment. This method worked successfully since the results of the research matched fishing statistics from Morocco's Ministry of Agriculture.(Younes Oulad Sayad, 2023)

**Literature Gap**

**Geographic and Temporal Constraints:** The study focuses only on one location (southern Moroccan waters) and covers only a specific timeframe (September to December 2021). This limited scope raises questions about the consistency and applicability of this approach across other regions or during different seasons.

**Parameter Limitation:** The study relies on POC, SST, and Chl-A as indicators. While these are important, other factors like **water salinity, ocean currents, or dissolved oxygen** could also impact fish distribution and may improve accuracy if included.

**Validation Limitations:** Although the study’s PFZ predictions were compared to recorded fish catches, additional validation approaches (such as real-time catch data or vessel tracking systems) could provide stronger support for the model's reliability.

**Addressing the Gaps**

**Expand Regional and Seasonal Coverage:** This will enable us, by trying this method for a number of different coastal regions and seasons, to establish how well the model is performing under different environmental conditions and what changes may be needed.

**Incorporate Additional Indicators:** Marine water quality is determined by evaluating physical, chemical, and biological factors; this is important for both human activity and the health of marine ecosystems.

**Oceanographic indicators:** The physical characteristics of water include colour, TSS (Total Suspended Solids), DO (Dissolved oxygen), turbidity, salinity, and temperature. Salinity affects density and circulation, temperature affects metabolism and species distribution, turbidity affects light penetration, colour signals organic matter or pollution, and TSS and DO levels affect aquatic life and water quality.

Many chemical elements affect the quality of marine water. pH, eutrophicating nutrients, dissolved gasses that indicate pollution, chlorophyll-a for phytoplankton, and hazardous heavy metals are important components. Ion content is determined by salinity.

Biological parametersassess marine life presence and diversity. Phytoplankton and zooplankton reflect ecosystem health, while bacterial counts determine water safety. Biodiversity monitoring reveals pollution effects. Biological Oxygen Demand indicates organic pollution. Monitoring macroalgae and coral helps evaluate habitat quality.

**Why the inclusion of additional parameters is essential?**

Because marine ecosystems are complex, it is essential to analyse additional parameters in marine water quality. This aids professionals in better comprehending the system, spotting dangers, and putting management plans into action.

**Evaluate Ecosystem Health:** Recognize the condition and evolution of marine environments.  
**Encourage Fisheries Management:** Maintain healthy fish populations and ecological equilibrium.

**Detect Pollution:** Determine the origins and effects of contaminants such as heavy metals, oil, and nutrients.

**Preserve public health:** Making sure that drinking water, seafood, and swimming are safe.

**Strengthen Validation Techniques:** Introducing real-time fishing data and vessel tracking systems could provide more immediate and precise validation, improving the model’s accuracy and practical utility.

1. It is evident that there is increasing interest in mapping coastal fishing zones using remote sensing data, particularly in identifying high-potential fishing regions using the ZPPI model. This study employed the Single Image Edge Detection (SIED). Image boundaries are detected with Single Image Edge Detection (SIED). It is necessary for tasks like image segmentation and object detection. In this study SIED to use sea surface temperature (SST) to identify thermal fronts and chlorophyll-a in conjunction with using GHRSST, SNPP-VIIRS, and MODIS-Aqua data from September 2018 and 2019. By mapping locations 4–12 nautical miles offshore, they were able to categorize them as either Low Potential (LP) or High Potential (HP) for fishing near Nias Island.(Purwanto et al., 2020)

**Literature Gap**

The study shows that using both chlorophyll-a and sea surface temperature data can enhance identifying productive fishing areas. However, the research is based on data from only specific times and one place (Nias Island), leading to concerns about the model's reliability in various conditions. Additionally, although the SIED method is beneficial, other algorithms and data sources could be further investigated for better outcomes.

**Addressing the Gap**

Expand the Timeframe and Locations: Applying this method in other coastal regions and seasons would allow us to assess if the ZPPI model's efficacy is constant or whether adjustments are required.

Integrate Additional Oceanographic Indicators: While chlorophyll-a and SST are essential, including other indicators, such as particulate organic carbon (POC) or salinity, might enhance the model's accuracy.

Test Alternative Algorithms: Exploring additional machine learning models could refine zone identification, provide a more detailed categorization of fishing possibilities.

1. In order to determine the best places to fish for anchovies, the study used catch per unit effort (CPUE), the study analysed capture statistics and remote sensing data, including salinity, chlorophyll-a, and sea surface temperature (SST). According to the results, September and October were the best times to fish during seasonal changes. Interestingly, fishing spots selected using SST did not always correspond with the PFZs determined using salinity and chlorophyll-a data. The Pesisir Selatan Regency had PFZs spread out over it that were predicted using composite data.(Ariana et al., 2020)

**Literature Gaps**

**Mismatch with Fishermen’s Knowledge:** The study found that fishermen’s fishing areas based on SST didn’t match the PFZs found using salinity and chlorophyll-a data. This suggests that there is a gap between fishermen's experience and scientific data, and more research could be done to combine both to improve fishing zone predictions.

**Limited Environmental Factors:** The study focused on salinity, chlorophyll-a, and SST. Other factors like ocean currents and water clarity were not considered but could improve predictions.

**Address These Gaps**

**Combine Local Knowledge with Remote Sensing:** I could work with fishermen to understand where they fish and compare their knowledge with scientific data to improve the model.

**Use More Environmental Factors:** I would include ocean currents, water clarity, and other factors to make the predictions more accurate.

1. The study used remote sensing data to link oceanographic factors to fish catches in the Java Sea. A modelling technique called GAM (Generalized Additive Model) is used to determine how fish catch and oceanographic conditions are related. Its nature is both non-parametric and semi-parametric. They used a statistical model GAM to examine how variables like sea temperature, depth, and chlorophyll-a impact fish harvests. Chlorophyll-a concentration was found to be the main factor affecting fish catches. Fish were observed in waters with specific salinity, depth, chlorophyll-a levels, and sea surface temperature. The study identified Potential Fishing Zones in the Java Sea that changed from March to June 2021, highlighting productive fishing areas in northern Madura, southern Borneo, and Pulau Laut.(Sasmito et al., 2022)

**Literature Gap:**

**Limited Oceanographic Variables:** The study examines SST, depth, and chlorophyll-a's impact on fish distribution. Other factors like salinity, currents, and water turbidity could also influence. Including more environmental variables may enhance PFZ forecasts.

**Temporal Changes:** The study identifies PFZs for specific moths in 2021 but does not consider how zones may change over seasons or years. Environmental fluctuations should be studied to understand long-term trends in fishing zones.

**Local Knowledge Integration:** The study doesn't use the observations and expertise of local fishermen; instead, it depends data from remote sensing. Information from fishermen may enhance satellite data.

**Address the Gaps:**

**Include Additional Oceanographic Variables:** An in-depth knowledge of fish habitats can be obtained by doing research on a variety of environmental factors, including salinity, ocean currents, and water clarity.

**Examine Temporal Changes Over Multiple Years:** Examine data from several years to track how PFZs change in response to changing conditions and improve seasonal fishing management methods.

**Integrate Local Knowledge with Remote Sensing Data**: To improve fishing techniques and combine scientific models with local experiences, fishermen's knowledge is being integrated with remote sensing data.

1. Harmful Algal Blooms (HABs) in Tolo Harbour, Hong Kong, where over 30% of HABs occur. They used Support Vector Machines (SVM) and Artificial Neural Networks (ANN) to model water quality and predict HAB outbreaks. SVM is better at finding optimal models but takes longer, while ANN provides quick predictions and accurate real-time responses, beneficial for analysing 30 years of data. These methods aid in understanding HAB causes, improving water quality forecasts, and preventing negative impacts in the area.(Deng et al., 2021)

**Literature Gaps:**

**Modelling Accuracy vs. Speed**: According to the study, ANN provides quicker but less accurate predictions than SVM, which takes longer to learn but produces accurate long-term forecasts.

**Limited Focus on Other Variables**: The study uses historical data to forecast HABs, but it does not examine other factors that need more research, such as nutrients loading, water pollution, or climate change.

**Integration with Environmental Management:** Based to the study, while machine learning is capable of predicting HABs, there may be a problem combining these predictions into practical management systems for efficient decision-making.

**Address These Gaps:**

**Improve Model Efficiency**: concentrate on creating a hybrid machine learning model that enhances processing speed and prediction accuracy for real-time applications by combining SVM and ANN.

**Incorporate More Variables**: Predictive models for environmental conditions can be improved by include human impact (e.g., pollution), oceanic changes, and climatic variability.

**Better Integration with Management Practices**: To effectively avoid HAB, investigate ways to improve the integration of machine learning in coastal management.

1. The study highlights the significance of Potential Fishing Zone (PFZ) advisories for Indian fishermen. These advisories use chlorophyll and sea surface temperature data to locate areas in the Indian Ocean with the best chances of catching pelagic fish. ESSO-INCOIS issues daily PFZ recommendations based on satellite observations, categorizing PFZs using a decision tree method with inputs like chlorophyll, persistence, and SST gradient. Higher catches are expected in areas with poor persistence and high SST gradient.(Majumder et al., 2021)

**Literature Gaps:**

**Lack of Direct Fish Quantity Data**: The present advisories contain particular fish quantity data and instead rely on environmental variables. This variation affects fishermen's capture efficiency by delaying precise forecasts of fish abundance.

**Model Generalization Across Seasons:** The model is only designed for the present. Fish distribution changes over time may have an impact on how accurate PFZ on recommendations.

**Limited Real-Time Adaptability:** Daily PFZ recommendations are not subject to rapid changes in the environment. The PFZ model's accuracy could be improved by real-time updates based on changes in the ocean.

**Address These Gaps:**

**Incorporate Direct Fish Quantity Data**:Future studies should include data on fish quantity from local fishing observations to enhance the accuracy and usefulness of PFZ predictions.

**Adapt the Model for Seasonal Variability**: To improve PFZ advisories' precision throughout different seasons, a model that considers seasonal variations in ocean conditions and fish behaviour should be developed**.**

**Implement Real-Time Updates:** Introducing real-time that adjust to sudden environmental changes could enhance PFZ advisories' responsiveness and benefits for fishermen.

1. Sea surface temperature (SST) fronts, winds, eddies, and chlorophyll (CHL) satellite data are used in a novel framework for identifying Potential Fishing Zones (PFZs) in the Bay of Bengal (BoB). Even in foggy situations, this method seeks to deliver constant PFZ predictions. High fish capture zones are confirmed to be in line with characteristics such as chlorophyll patches, SST fronts, and cyclonic eddies based on validation using Fishery Survey of India (FSI) data from 2016 to 2018. According to the study, the largest link between fish capture rates and chlorophyll-a concentration—a measure of food availability—is followed by SST gradients.(Sarangi et al., 2023)

**Gaps in this Study**

**Reliance on Historical Data for Validation:** Although historical data from 2016 to 2018 was used to evaluate the model, it might not accurately represent the current oceanic changes brought on by climate change.

**Limited Exploration of Multi-Year Variability:** The Bay of Bengal's long-term climatic and environmental changes may be missed by the existing model, which only examines a few years' worth of data.

**Challenges with Cloudy Condition Prediction:** Predicting cloud cover is challenging because satellite-based measurements provide difficulties.

**Suggestions for Gaps:**

Future research should use recent validation data to test model robustness. Annual updates are also recommended for ecosystem changes. Expanding the model with a decade-long dataset will enhance trend analysis. Integrate alternative data sources like microwave sensors for cloud-free data and predictive gap-filling techniques for improved consistency.

1. An innovative artificial intelligence (AI) model combines filter bat-optimized long short-term memory (FB-LSTM) networks with deep convolutional neural networks to forecast Potential Fishing Zones (PFZs) along India's coastline. This model improves existing techniques by using satellite imagery and information such as GPS coordinates, sea surface temperature, and chlorophyll levels to predict fish-gathering locations with 99% accuracy.(Sivasankari et al., 2022)

**Main Literature Gap:**

This study's primary flaw is that it depends on powerful computers to handle data and run the model efficiently. This amount of processing effort may restrict the model's applicability for real-time forecasts for fishing communities in remote or under-resourced locations.

**How to Address It:**

A cloud-based platform or simpler version can make the model useful and available to fishermen. For real-time data processing in low-resource environments, use edge computing.

1. The content discusses using remote sensing technologies to monitor ocean water quality indicators like turbidity and chlorophyll-a. Sensors like altimeters and optical sensors are effective, with optical and passive microwave sensors being particularly useful. Semi-empirical algorithms are successful for monitoring water quality, but there are challenges in maintaining consistent spatial and temporal coverage.(Mohseni et al., 2022)

**Main Literature Gap**

The study's primary problem is that surrounding influences and sensor limitations cause OWQ results to be inconsistent and inaccurate. Despite their effectiveness, semi-empirical algorithms are still limited by sensor capabilities.

**How to Address It**

Integrate data from several sensors and modify algorithms for local conditions to improve OWQ evaluations. This can increase accuracy, particularly in a variety of conditions when a single sensor or algorithm would not be sufficient.

1. By contrasting pond and cage aquaculture locations, this study examines the water quality in marine aquaculture zones. We looked at important water quality indicators, such as dissolved oxygen, temperature, salinity, pH, and antibiotic resistance genes (ARGs). According to principal component analysis (PCA), aquaculture type affected the variables affecting water quality. Ponds required salinity, dissolved oxygen, and ARGs, but cage habitats required chlorophyll a, salinity, and dissolved oxygen. Terrestrial sources may contribute to ARGs in cages, as pollution maximal values were seen in August for ponds and November for cages.(Zhang et al., 2020)

**Main Literature Gap**

Lack of specialized, ongoing water quality monitoring for various aquaculture kinds and their unique environmental challenges, such as seasonal variations in pollutants and land runoff influencing ARG levels, is the primary problem.

**How to Address It:** Seasonal variations and pollution sources, like as runoff, can be better understood with a thorough strategy for ongoing monitoring. The control of water quality can be enhanced by combining data from various farms.

## Incorporating Summary Table

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| No | Research Paper | Area of study | Ocean Parameters | Methodology and Model | Limitations of study | Data and Source | Time Frame |
| 1 | (Younes Oulad Sayad, 2023) | Dakhla Oued Eddahab (Morocco) | Chl-A, SST & POC | Image Processing & Mapping | Small time frame | MODIS  (Moderate Resolution Imaging Spectroradiometer) | SEP-DEC (2021) |
| Covers only Southern Moroccan waters |
| Limited parameters |
| Validation Limitation |
| 2 | (Purwanto et al., 2020) | Nias Island | Chl-A & SST | ZPPI Model (Potential Fishing Zone) & SIED (Single Image Edge Detection) | Limited time- frame, place & parameters | GHRSST, SNPP, VIIRS & MODIS Aqua Data | SEP (2018-2019) |
| 3 | (Ariana et al., 2020) | Pesisir Selatan Regency | Chl-A, SST & Salinity | CPUE Analysis  (Catch per Unit Effort) | Fisherman’s PFZ statistical data did not match with SST and chl-A, Limited parameter and time-frame | Fisherman traditional data & MODIS, ocean color  (In-situe) | SEP-OCT (2014-2018) |
| 4 | (Sasmito et al., 2022) | Java Sea, Northern Madhurai, southern Borneo & Pulan | Chl-A, SST, Depth & Salinity | GAM (Generalized Additive Model) & Statistical Model | Limited parameters & Temporal changes | MODIS & CMEMS Data portal | MAR-JUN (2021) |
| 5 | (Deng et al., 2021) | Tolo Harbour, Hong Kong | Chl-A, BOD, TIN, DO & Ph | SVM & ANN  (Machine Learning Models) | It’s not included other factor like nutrient loading water pollution or climate changes | In-situ data | 30 years of data |
| 6 | (Majumder et al., 2021) | India Ocean | Chl-A & SST | Decision Tree Method | Less parameter | NOAA,  AVHRR & MODIS | JAN- 2015- DEC 2016 |
| Naïve Bayes |
| Hybrid Decision Tree |
| 7 | (Sarangi et al., 2023) | Bay of Bengal | Chl-A, SST, Winds & Eddies | Noval Frame Work | Historical data not currently accurate cloud cover is difficult to prediction | Fishery survey of India | 2016-2018 |
| CPUE |
| 8 | (Sivasankari et al., 2022) | Indian Coast | Chl-A & SST | Novel Hybrid Prediction | Less parameters | NOAA-17,  INCOIS,  METOP 1& 2,  MODIS | JAN- 2011- APR 2021 |
| FB-LSTM (Filtter Bat Optimized long-short-term memory) | It depends on powerful machines and systems |
| 9 | (Mohseni et al., 2022) | N/A  General Study | Chl-A, SST,DOC,Turbidity & Secchi disk depth | Semi Empirical Algorithm | Sensor limitation & this algorithm is still limitations | NWQD,  NOAA | APR- 2017-FEB -2022 |
| 10 | (Zhang et al., 2020) | Zhuanghe  (China) | Chl-A, SST, DO, Salinity, Ph & ARGs (Antibiotic Resistance Genes) | PAC (Principal Component Analysis) | Various aqua culture has unique environmental challenges and which is not applicable in another situation | In-situ Data | MAY, AUG, NOV (2018) |

Table 1. Incorporation of Literature Review

## Conclusion

These articles provide an incredible amount of useful information about the features of oceanic water quality, including insights that may be used to the evaluation of marine water quality. It was shown that one of the most successful and economical ways to carry out these kinds of studies is using satellite remote sensing. A wide variety of data processing approaches were investigated during these in-depth investigations, including the application of machine learning techniques using the Python programming language. Every research used unique approaches and algorithms designed to achieve their particular goals. Each study contributed significantly and uniquely to the general understanding of water quality, resulting in an extensive variety of discoveries. While some of the studies focused on particular areas of water quality, others aimed to identify important knowledge gaps that may lead to future study and development. The main goal of this research project is to improve the project's correctness and efficiency by combining data from several sources and using a consistent and logical methodological framework. The project seeks to produce a thorough and reliable study of marine water quality that captures the knowledge gained from these many studies and methodologies by combining various data streams and standardizing the analytical procedures.

Chapter-3

# Research Methodology



## Study Area

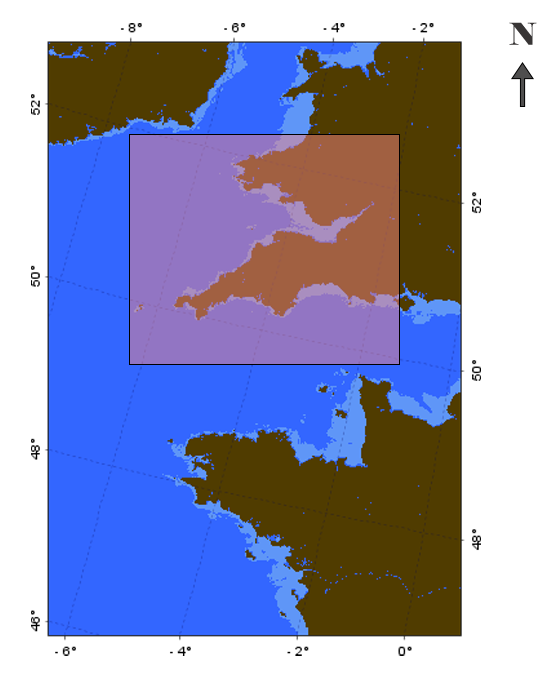
This project's primary goals are to analyse marine water quality and identify potential fishing zones (PFZ) along the UK's south coast, particularly the **Cornish Coast**, and I select ten years of data **(2013 Jan–2023 Dec)** to assess this area for research. This region attracts a lot of attention due to its rich marine biodiversity and importance as a hub for the UK fishing industry. Due to its influence from both local environmental factors and Atlantic Ocean currents, the Cornish Coast is an ideal example for studying marine water quality and fishing zone prediction.

Figure 1. South Coast of UK

People have lived in Cornwall for thousands of years, with back to the Stone Age or Palaeolithic. On Cornwall's beaches and in its shallow waters, early hunter-gatherers discovered a variety of marine life. Since the first-time people lived on the peninsula, fish and shellfish have been vital foods for the people of Cornwall, and they continue to be so today.

* Pilchards (Sardines)

For the Cornish, fishing has always been essential to their survival. The Pilchard fishery is Cornwall's oldest, most extensive, and best-documented fishery. Despite being renamed as Cornish sardines in recent years, pilchards are still the same small, delicious fish that are rich in omega-3 fatty acids.(“Cornish Fishing History, Cornwall Good Seafood Guide,” n.d.)

## Influencers of Study Area

Numerous marine habitats, such as rocky shorelines, sandy beaches, estuaries, and offshore waters, can be found along the Cornish Coast. Marine mammals, fish, shellfish, seabirds, and other marine life are all supported by these habitats. The Fal Estuary and the Isles of Scilly are two significant marine conservation areas in the area.

Marine water quality along the Cornish Coast is impacted by several factors, such as:

* **Ocean currents:** The Cornish Coast receives warm, nutrient-rich water from the Atlantic Ocean currents, which supports marine life.
* **Local environmental factors:** The quality of sea water can also be impacted by local environmental variables such agricultural runoff, sewage effluent, and river discharge.
* **Human activities:** Fishing, shipping, and tourism are examples of human activities that can affect the quality of sea water.

## Data & Data collection Instruments

The main source of data used in this study is satellite data, which is saved in NetCDF file format. To better grasp the depth and emphasis of the research, a number of data elements and objectives will be looked at and explained in detail in the section that follows.

* **Ocean Color Web**

NASA uses a variety of Earth observation tools to detect the color, or spectral character, of water. In particular, NASA collects, stores, and makes available to the public this type of information from a number of sources, such as remote sensing ocean color instruments on satellite and aircraft platforms, ships, long-duration autonomous in situ platforms, and Earth system model outputs.

* **MODIS Instrument**

One of the main instruments aboard NASA's Terra and Aqua satellites is the Moderate Resolution Imaging Spectroradiometer (MODIS). It offers detailed images of the Earth's surface, including its seas. Potential fishing zone (PFZ) forecasting and marine water quality assessments both greatly benefit from this data.

* **Level 2 Data**

A Level-1A or Level-1B data product can be used to create a Level-2 data product. The geophysical values for each pixel, obtained from the Level-1 radiance by applying the sensor calibration (for Level-1A), atmospheric corrections, and geophysical parameter algorithms, are the primary data contents of the product. Each Level-2 product is stored in a single physical NetCDF file and has geographic coverage (scan-line and pixel extent) that is precisely the same as that of its parent Level-1 product.

* **Data Objects**

Scan-Line Attributes, Geophysical Data, and Navigation are the three data object classes that include data that are functions of scan lines. They are dimensioned number\_of\_lines because each data object in these groups contains data for every scan line. In order to obtain all the information associated with a particular scan line, n, it would be necessary to read the nth values of every data item in these four groups. Every data object has the following standard attributes: \_FillValue, valid\_min, valid\_max, units, and long\_name.

1. **Scan-Line Attributes**

Information applicable to a whole line is contained in the data objects in the group scan\_line\_attributes.

1. **Geophysical Data**

The sensor science data is the source of the data objects in the group geophysical\_data. The data quality flags are also included in this group. The dimensions of each data item are number\_of\_lines x pixels\_per\_line. The majority of parameters are saved as integers that are scaled based on the data object's scale\_factor and add\_offset properties.

Wavelength-specific parameters (such Rrs) are derived using distinct data items for each band; the Mission and Sensors section contains a list of wavelengths for each sensor. With their corresponding quality fields, the SST and SST4 are kept in distinct archive products. Each sensor's Level-2 File Format includes a list of the particular fields. The L2gen User Guide provides a comprehensive list of parameters that the software can generate.

1. **Navigation**

The geolocation data for the data product is contained in the data objects in the group navigation\_data. This comprises the G-ring coordinates for the swath, the latitudes and longitudes of the observed sites, and the control point indices (required if the latitudes and longitudes are subsampled and not full-resolution). The tilt angle is also supplied for tilting sensors.

* **Data Format and Data Extensions**

NetCDF (Network Common Data Form)

Complex scientific data is often stored in NetCDF (Network Common Data Form) files, which are very flexible. They are essential tools for data administration and analysis, particularly in the field of science, because they are made especially to facilitate the production, access, and exchange of array-oriented scientific data. The strong structure of NetCDF files is evident when working with large and complex datasets, such those used in oceanographic research, allowing researchers to effectively store, retrieve, and analyse the large amount of multidimensional data that is essential to their work.

## Data Acquisition Process

* + **Auto Data Downloading process**

Python Library access data through google earthaccess:

Earthaccess is a Python package that makes it easier to find, download, and stream NASA Earth science data. By offering a higher-level abstraction for NASA's Search API (CMR), it facilitates data searches using a more straightforward syntax rather than low-level HTTP queries.

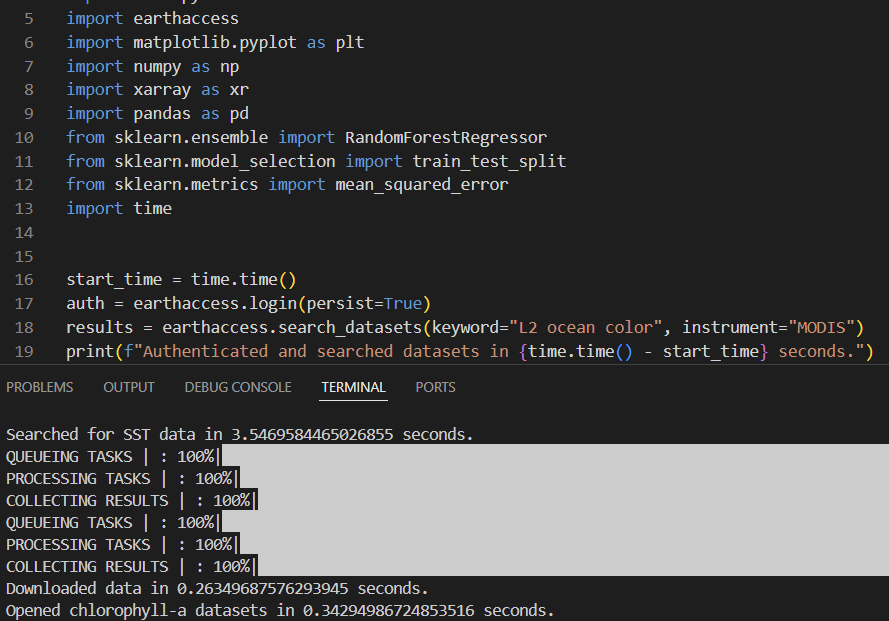


Figure 2. Auto download data in python programme

* + **Manual Data Downloading process**

The NASA Ocean Color Level-1&2 Browser allows us to access ocean color level 2 data:

Regional satellite data packages are available on this page from a number of sensors, including as SeaWiFS, MODIS Aqua and Terra, MERIS, OCTS, CZCS, and Aquarius. Data may be found by sensor, time period, and area. A variety of data formats, such as NetCDF and HDF are available.

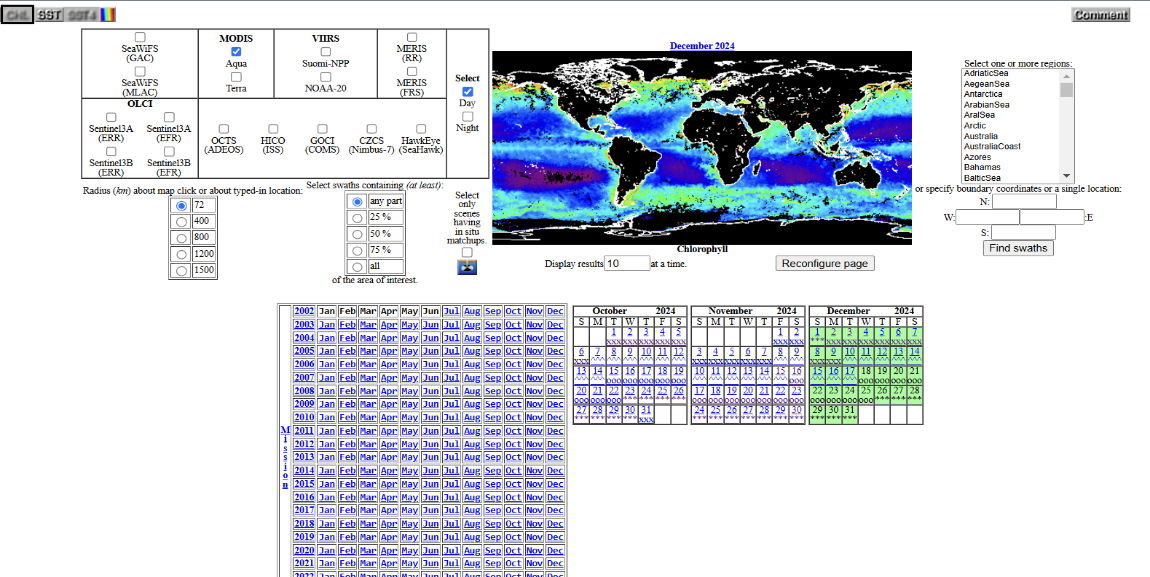


Figure 3. Manual downloading data interface

## Analysis Tools and materials

* + **Python and Python Libraries**

Python provides a robust environment for analysis, visualisation, and manipulation making it essential for data analysis. Python cleans and filters data, maintains large databases rapidly, and facilitates calibration and validation. It facilitates the creation of features, offers tools for model evaluation and training, and uses machine learning to provide predictions. Python creates instructional graphics and makes interactive data exploration easier. Python improves each stage of the process, from data processing to visualisation, and is essential for PFZ prediction and oceanographic data interpretation.

* + **SeaDAS (SeaWiFS Data Analysis System) Software Package**

A software program called SeaDAS (SeaWiFS Data Analysis System) is used to handle and examine data on ocean color. It is essential for determining fishing zones and researching the quality of ocean water. It obtains important water quality measures, manages the preprocessing of satellite data to guarantee its correctness, and displays data in plots and maps. In order to help machine learning algorithms anticipate fishing zones, SeaDAS also extracts information such as the concentration of chlorophyll-a. It also combines several data sources to provide a more comprehensive picture of the maritime environment.

I used this powerful software program to do thorough data validation tests and guarantee the highest level of correctness while confirming the project's finished product.

## Data Analysis and techniques

For the purpose to analyse marine water quality and estimate potential fishing zones, I have integrated statistical, geographic, and time series analyses. By combining these three analytical techniques, an accurate understanding of the dynamics of water quality can be obtained. Gaining important insights into water quality analysis through the successful integration of various analytical techniques can improve monitoring, management, and fishing zone forecast.

* **Statistical Model**

Utilizing statistical techniques is crucial for evaluating the data gathered in water quality analysis. Descriptive statistics, correlation analysis, and hypothesis testing are important methodologies. In order to identify patterns and anomalous outcomes, descriptive statistics display the averages and variances in water quality measurements. In order to find potential links, correlation analysis examines the relationships between different factors. Using statistical tests to assess the significance of observed variations, hypothesis testing allows the development and assessment of hypotheses regarding trends or differences in water quality.

* **Geospatial Analysis**

By using geographical information systems (GIS) approaches to show and assess spatial patterns in water quality indicators, geospatial analysis, on the other hand, provides an additional layer of knowledge. This method assists in recognising regions and possible to find bio rich by allowing the production of detailed maps that graphically show the geographical distribution of parameters. Geospatial analysis also enables the estimate of parameters at unsampled sites and the identification of spatial clusters of different water quality levels using techniques like spatial statistics and spatial interpolation, clarifying the spatial dynamics affecting water quality.

* **Time Series Analysis**

Time series analysis is essential for monitoring the changes in many parameters in water quality studies. Trend analysis shows the effects of human activity and climate change and finds increases, declines, or stability in water quality indicators. While time series modelling is essential for developing predictive models to anticipate future trends in water quality and evaluate the impacts of different management approaches, seasonal analysis investigates variations and patterns.

* **Integrated Analysis System**

It can better monitor, manage, and safeguard important water resources by integrating these analytical methods statistical, geospatial, and time series analyses to gain a greater understanding of complicated water quality problems. It can investigate the complex dynamics of water quality using this integrated approach, which gives the ability to make educated decisions that can improve the management of water resources. We may better understand the many variables impacting water quality changes across various geographical and temporal scales by combining statistical approaches with geospatial and time series analysis. This comprehensive knowledge helps to detect underlying trends and patterns, identify possible pollution sources, assess ecological and public health, hazards, and optimize monitoring techniques for efficient data collecting. Furthermore, using these analytical tools makes it easier to create evidence-based plans, assess intervention efforts, and set up proactive early warning systems all of which are essential for maintaining sustainable water quality management, protecting the wellbeing of aquatic ecosystems, and especially forecasting potential fishing zone identification (PFZ). The machine learning system, which uses the effective random forest approach, is the main driver of the integrated analytic system at its core. The foundation for the integrated combination of all the many outputs produced by statistical, geospatial, and time series techniques is this strong thing. The system efficiently gathers and analyses the various data streams by applying the random forest approach, producing a thorough and perceptive end result. By evaluating, interpreting, and refining the knowledge obtained from each technique, the machine learning system improves this integration process significantly. The technology guarantees a smooth fusion of data insights across this complex process, providing the ability to examine the studied data broadly and make well-informed decisions based on the combined results.

## Limitations

The main method used in this work is unable to directly control the processing and analysis of raw data; rather, it depends on obtaining input from three different models: the statistical model, the geographical model, and the time series model. The final product is created through a multi-step, complex procedure that greatly increases the amount of time required to get the desired outcome. One of the main disadvantages of this method is the length of time it takes to process and generate output. In addition to adding to the algorithm's time-consuming nature, the requirement to go through many phases of data processing and model integration also draws attention to a fundamental flaw in its structure.

Chapter-4

# Data Analysis and Implementation



## Introduction

I begin the study by extracting useful information from MODIS imagery from satellites that are available via NASA's ocean color data website, a crucial resource for evaluating ocean health indicators, in order to thoroughly evaluate the quality of marine water. These imagery from satellites offer a thorough look at a number of factors that are essential for water quality monitoring. To guarantee the integrity of the dataset, a number of complex procedures are performed as part of the first stages of data processing, including data cleaning, cloud masking to remove interference, land masking to concentrate only on marine regions, and sub setting to isolate certain locations. After the data has undergone these initial steps of refinement and purification, I proceed with the crucial process of extracting important oceanographic characteristics, which helps to provide important insights into the condition of the marine environment. I can find underlying patterns and identify new trends that might direct the study of environmental changes and effect assessments by displaying these oceanographic parameters, a procedure essential for interpreting complicated data sets.

I perform an accurate validation procedure that systematically collects ocean parameters in addition to using the highly effective program SeaDAS for validation. This program works differently, using a thorough statistical method to effectively validate parameter data. The method advances smoothly to guarantee the production of accurate outcomes. Importantly, a key component that connects all phases of this ocean parameter validation process is the use of a machine learning procedure. The validation process is enhanced by the use of SeaDAS, which allows a detailed and reliable analysis of data in a specialized framework. In order to guarantee the correctness of the acquired ocean parameters, the program replicates a systematic technique that makes use of statistical precision.

As it progresses, it goes through several steps to carefully verify every parameter, which improves the data validation procedure as overall. Additionally, this validation process becomes more efficient by utilizing machine learning techniques to automatically extract relevant insights from the data. In the end, this combination of machine learning and statistical techniques leads to a thorough and precise method of validating parameter data, demonstrating the creative potential that comes with using cutting-edge software tools like SeaDAS.

I use the Random Forest machine learning method, a powerful computational tool that can identify complex associations in the information, to improve the analysis even further and advance toward predictive modelling. I make use of a wide range of Python libraries during this complex analytical process, each of which has a unique function in assisting with data manipulation, visualization, and modelling tasks. This brings me one step closer to creating predictive models that can identify possible fishing areas that are essential for the sustainable exploitation of marine resources.

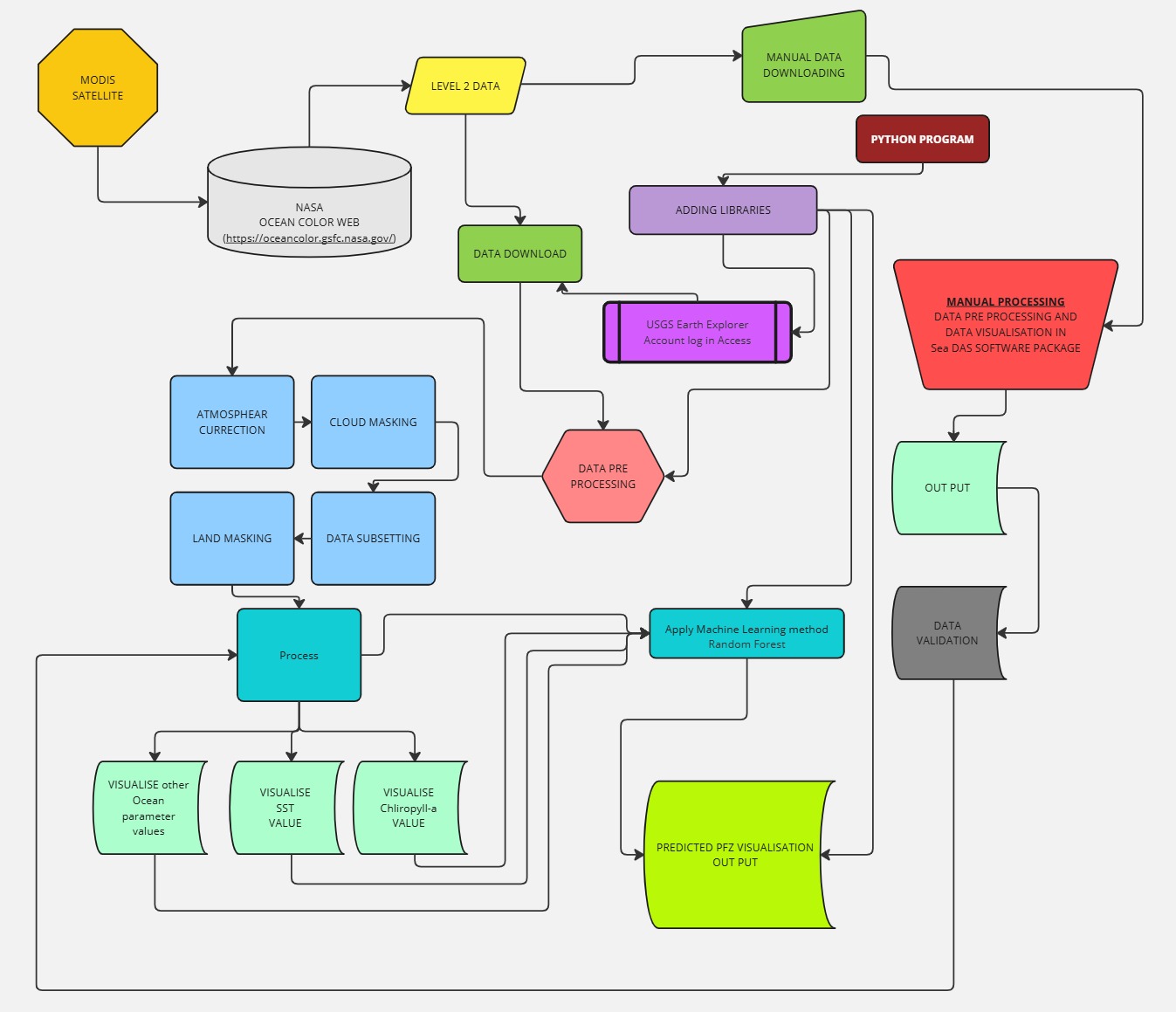


Figure 4. Flow chart of Methodological implementation

## Procedure and Execution

I want to start by introducing the tools that were used to do this study. As a powerful and adaptable programming language, Python was essential and shown exceptional performance in a wide range of everyday applications. For essential tasks like pre-processing, statistical analysis, and geospatial analysis, I used SeaDAS, an image processing program, combined with Python. The data validation process was greatly strengthened by the use of SeaDAS, which increased its accuracy and efficiency. These technologies served as a basis of my main data processing approach. The basic processing queue emerges when I examine the program's syntax in further detail.

Utilizing this application to carry out the primary analytics activities and pre-processing operations is part of my strategy goal. I hope to optimise the analysis process and optimise the insights obtained from the data by defining discrete stages that include pre-processing and thorough analytics.

I purposefully decided to use remote sensing satellite data for the duration of the project before really looking into the data. This required accessing NASA's ocean color data library to retrieve an extensive collection of satellite data. It is significant to note that, this data falls into the level-2 category after undergoing initial processing by NASA's Ocean Biology Processing Group (OBPG). These pre-processing steps comprised crucial activities including cloud masking, georeferencing, calibration and validation, atmospheric correction, and algorithm application.

The next step was to carefully examine all of the data using SeaDAS's help after obtaining the required information from the ocean color web. My focus then turned to optimising the process by using Python and importing necessary libraries to enable the easy completion of every project phase. I wanted to guarantee the correctness and integrity of the data processing cycle by following this systematic methodology, which would provide a strong basis for the next analyses and insights that the project will likely result in.

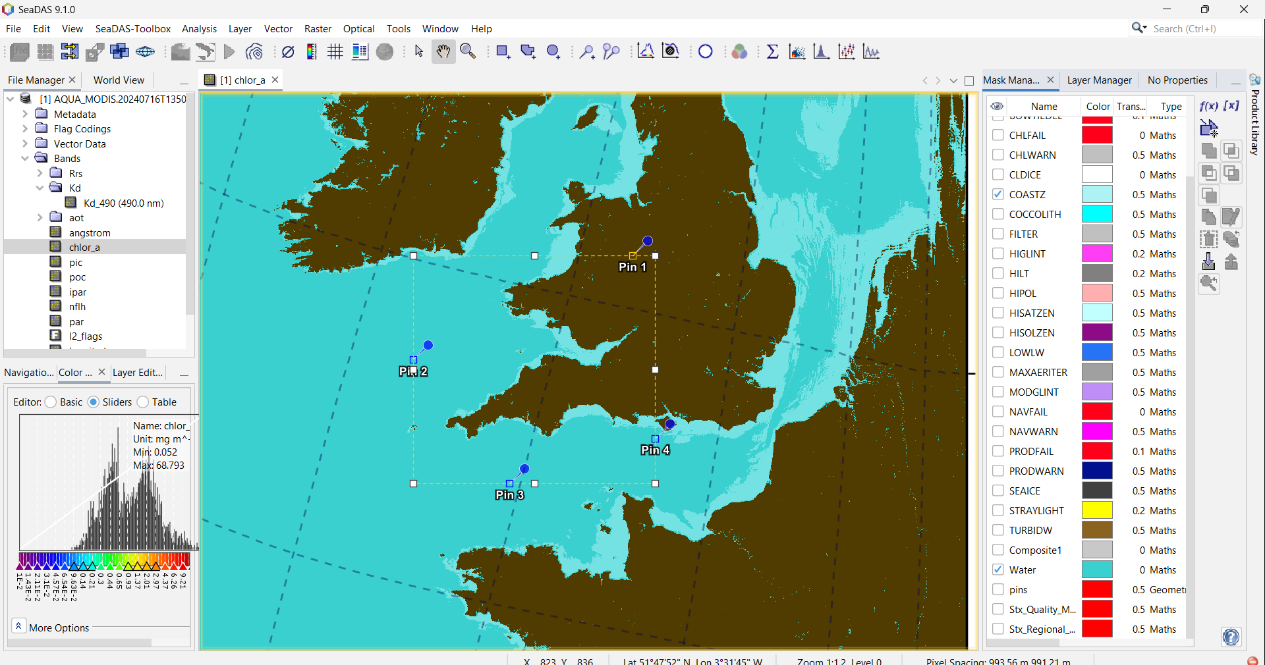
****

Figure 5. Interface of SeaDas Programme

At this crucial point in my arrangement, I feel needed to go into great detail about a few essential libraries that have greatly influenced and improved the functionality of the project I was working on. The outcome and effectiveness of the endeavour were greatly impacted by these libraries, which were carefully chosen and carefully included into the framework. At this point, it is crucial to highlight the significant influence of those libraries had on the project's general growth and eventual success.

Every library has a specific function and contributes its own set of features and capabilities to the project's overall scope. It is clear from a thorough examination and investigation of these libraries that their use has been crucial in improving and maximizing the project's functionality and performance, taking it to a completely new level of complexity and effectiveness. A greater understanding of these libraries' relevance and value within the project is gained by analysing their features and complexities, highlighting their essential role in the overall architecture and design.

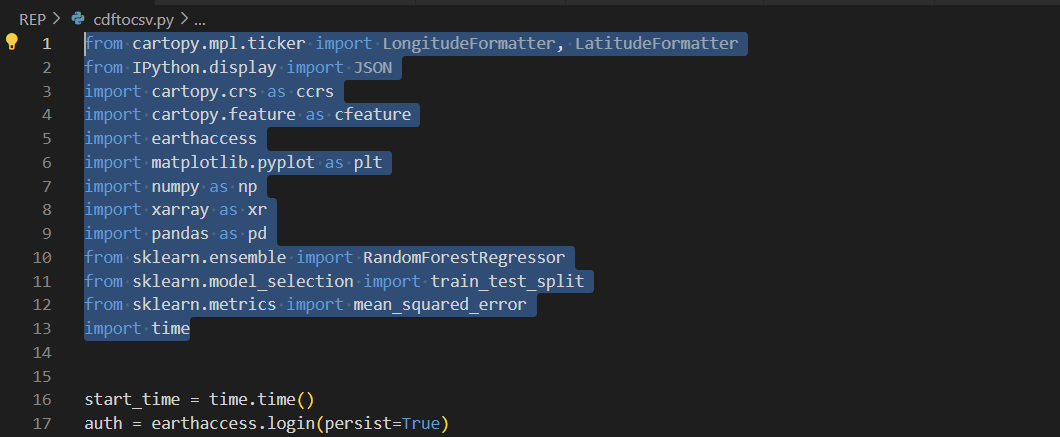
****

Figure 6. List of Used libraries in python

## Libraries and functions

**1. cartopy.mpl.ticker:**

* **LongitudeFormatter, LatitudeFormatter:** On maps created with Cartopy, these classes are used to format the tick labels for latitude and longitude. They enhance readability and guarantee that coordinates are shown correctly.

**2. IPython.display:**

* **JSON:** Although it is probably not utilized in the main analysis, this function (from IPython.display) might be useful for showing findings as JSON objects in a Jupyter Notebook context.

**3. cartopy:**

* **ccrs (cartopy.crs):** This submodule offers a variety of Cartopy map projections. The purpose of the visualization and the research region would determine which projection was used.
* **cfeature (cartopy.feature):** This submodule provides tools for creating context on maps by including geographical characteristics such as landmasses, coasts, and political borders.

**4. earthaccess:**

* This library was created especially for handling Earth science data from the Earth Observing System (EOS) of NASA. For datasets that are relevant to the study, such oceanographic data from MODIS or other EOS satellites, it makes data access, processing, and visualization easier.

**5. matplotlib.pyplot (plt):**

* This is a basic Python charting library. In order to examine the oceanographic data and model outcomes, it is probably utilized to create a variety of visualizations, like as maps, scatter plots, and time series plots.

**6. numpy (np):**

* NumPy is a fundamental Python package for scientific computing. It offers effective array data structures for data processing and numerical calculations. For array operations on oceanographic data (such as temperature and chlorophyll-a), NumPy is probably going to be used.

**7. xarray (xr):**

* For working with labelled multi-dimensional data, such as geo-referenced scientific datasets, Xarray is an additional robust library. With its geographical dimensions (latitude, longitude), it provides a practical means of storing, modifying, and analysing your oceanographic data.

**8. pandas (pd):**

* Pandas is a well-known Python package for data manipulation and analysis. Working with tabular data—such as water quality values from in-situ sampling—is one of its common uses. It might be applied here to combine or process data from many sources.

**9. sklearn.ensemble (from scikit-learn):**

* **RandomForestRegressor:** The Random Forest regression algorithm, a machine learning method frequently applied to prediction applications, is implemented in this class. Potential fishing zones (PFZs) are predicted by a model that is probably built in my project using oceanographic data (features) together with historical fish capture data (target variable).

**10. sklearn.model\_selection:**

* **train\_test\_split:** For machine learning models, this feature assists in separating my data into training and testing sets. The model is trained using the training set, and its performance on unseen data is assessed using the testing set.

**11. sklearn.metrics:**

* **mean\_squared\_error:** With this function, the mean squared error (MSE) between the actual and anticipated values is computed. It's a standard statistic for assessing how well regression models like Random Forest in my case perform.

**12. time:**

* This module is commonly used to implement time-related features in the analysis or data processing phases or to time the execution of code.

All things considered, these libraries offer an extensive toolbox for PFZ prediction machine learning model construction, oceanographic data analysis and visualization, and maybe performance evaluation.

## Data Preprocessing procedures

For efficient reading and loading of satellite data in NetCDF format, specialized libraries such as xarray and netCDF4 are essential. The tools required to manage the complex structures of NetCDF files are provided by these libraries.

Two essential components of data processing are data cleaning and filtering. Any incorrect or missing data points are eliminated from the dataset in this stage. For future research, it's also critical to arrange the data according to certain time periods or geographical regions.

Data sub setting, which includes extracting pertinent variables for analysis, is another way to further refine the dataset. To do this, temporal or geographical subsets can be created for targeted analysis of particular data points. Focusing on certain factors and subgroups allows researchers to obtain more significant findings.

Workflow for data processing requires validation and calibration. Data correctness is checked through validation by comparing to established reference datasets or measurements. In order to ensure that the data is reliable and consistent with benchmarks, calibration modifies the data according to standards or fixes inconsistencies discovered during validation.

## Implementation of integrated analysis system

I have assembled a wide range of analytical techniques in an effort to efficiently assess marine water quality and identify potential fishing grounds. In order to fully analyse the complexity involved in water quality dynamics, I have specifically combined statistical analysis, geographic research, and time series evaluations. A thorough and accurate understanding of the role that water quality plays in marine ecosystems may be attained by combining and using these many analytical techniques.

This integrated approach improves the accuracy of fishing zone predictions, the effectiveness of monitoring systems, and the application of sustainable management methods in addition to facilitating a greater understanding of the variations and patterns of water quality.

The relevance of using multiple analytical tools for the thorough research of marine habitats is shown by the combined effect of integrating these unique techniques. In the end, this complete plan strengthens the information and resources available for protecting our marine ecosystems and guaranteeing the sustainable use of marine resources, in addition to enhancing the ability for scientific study.

I used Python programming and the SeaDAS package to integrate statistical, spatial, and time series analysis to obtain crucial insights. Monitoring, management, and fishing zone predictions have greatly benefited from the improved water quality analysis brought about by this efficient fusion of several analytical techniques.

## Python implementation and derived task

The first and most important step is to install all the necessary libraries and import those needed for different process executions in order to build up the Python program. It follows that data processing is a major program priority. I have written prescriptive code that this Python script uses to find and obtain data on its own. In this regard, it is essential to emphasize the need to register and authenticate with USGS Earth Explorer. This is made possible by the earthaccess library, which simplifies authentication, search capabilities, and the acquisition of necessary information all along the way.

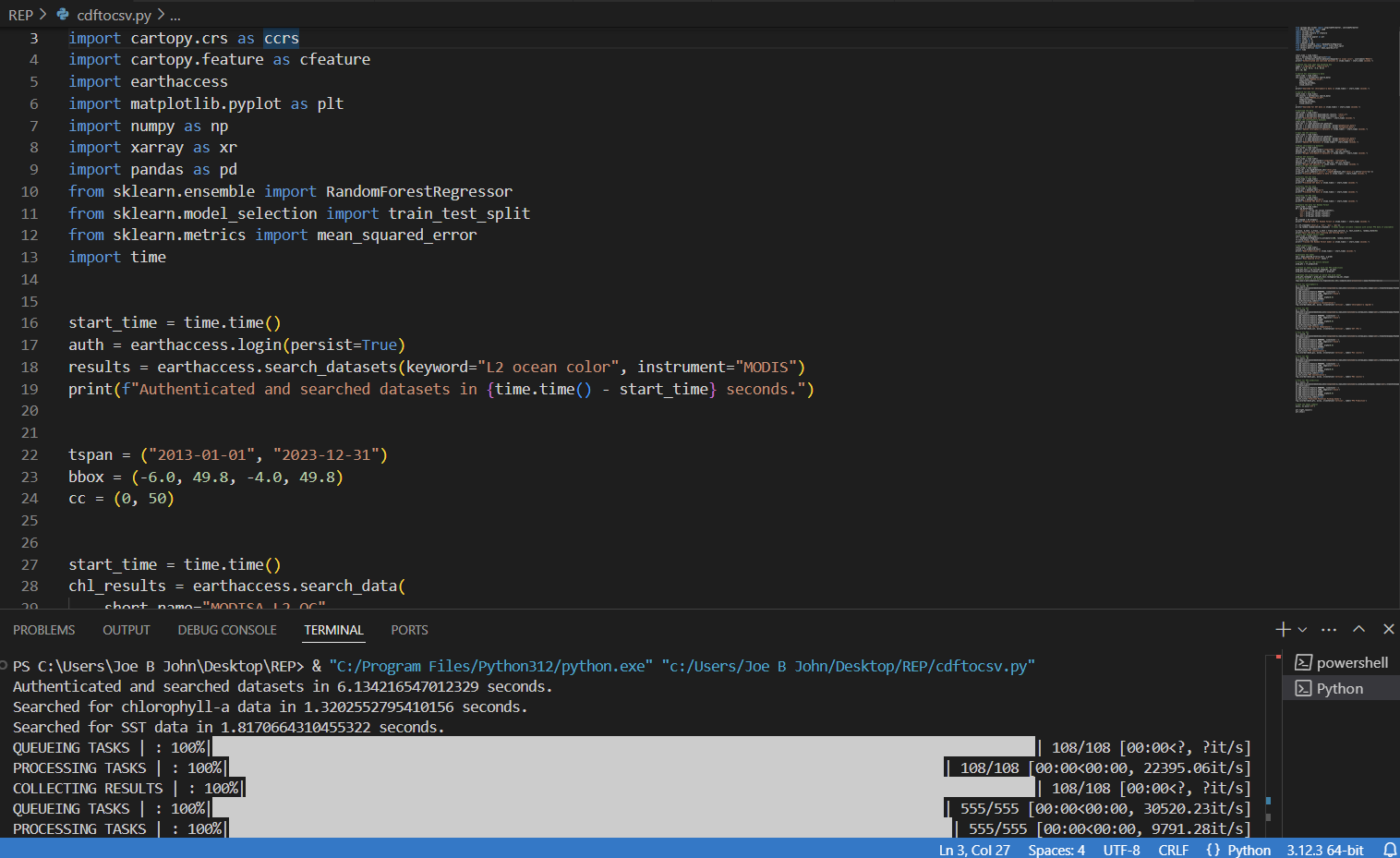


Figure 7. Authentication procedure in earthaccess used python for access data online

The next step is to open the downloaded data file and carefully look for the particular ocean parameter in that dataset. Each parameter is intelligently retrieved separately by the system, which then arranges them in a stack for visual representation. The ability to precisely identify and calibrate geographical coordinates, such as latitude and longitude, makes this stacking and merging procedure more useful. Additionally, it guarantees the efficient use of the time parameter during this entire process, improving the accuracy and consistency of the parameter visualization process. A thorough analysis that permits a deeper comprehension of the marine phenomena being studied is made possible by the methodical setting of the parameters with the aid of the stacking merge, which promotes a smooth integration of the data. The process of locating, extracting, and displaying the required ocean parameter is greatly simplified by using these advanced methodologies, which eventually enables more complex and perceptive analyses of the information.

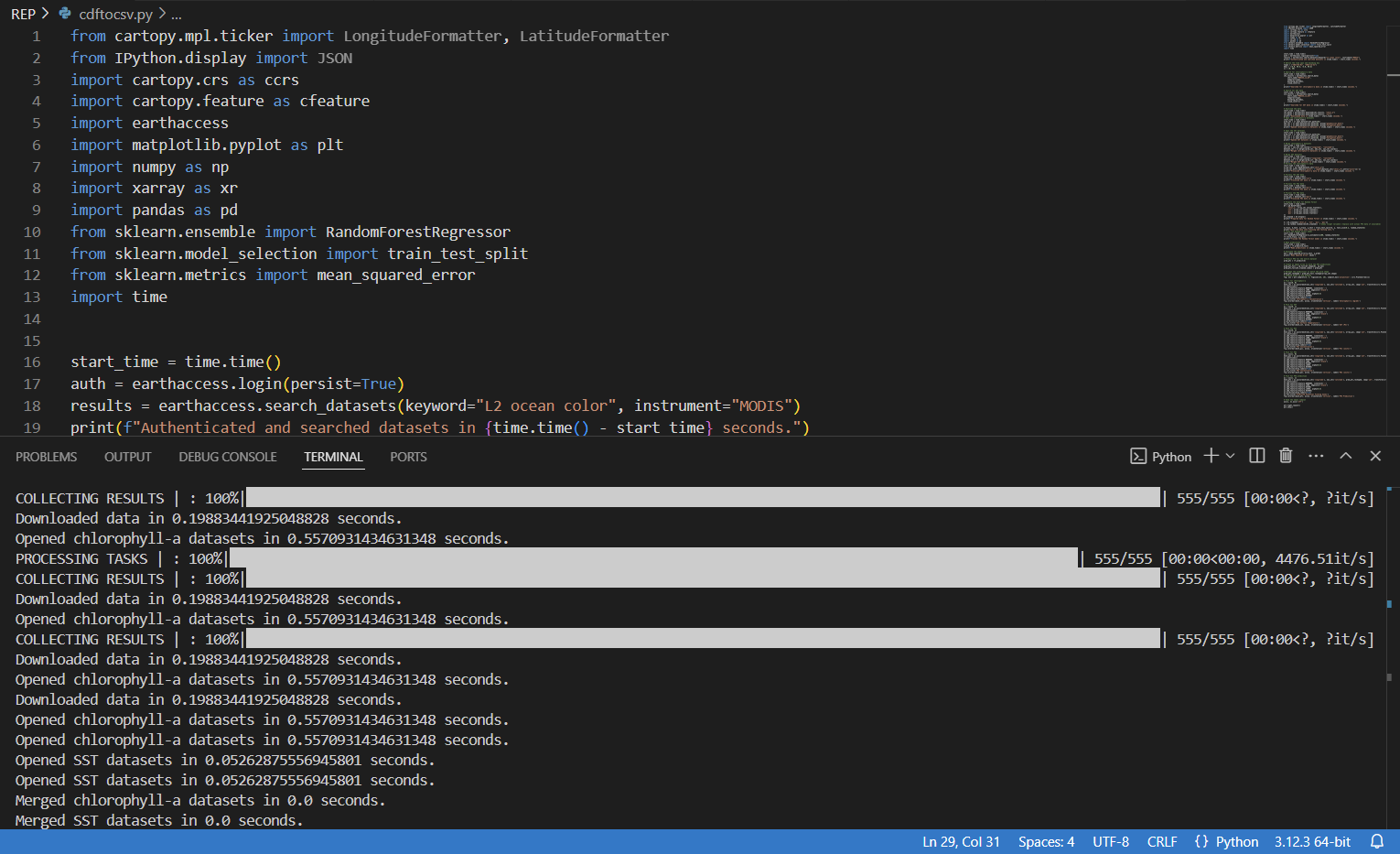


Figure 8. Data downloading with using Python

Each parameter's visualization is one of the main procedures that will take place in this Python application. Statistical analysis, geographic analysis, and time series analysis are just a few of the analytical methods that will be used concurrently during this procedure. By combining these studies, an Advanced Analysis System will be produced that is especially made for efficiently showing the parameters. Geospatial analysis will give a spatial viewpoint that highlights geographical patterns and linkages, while statistical analysis will shed light on the distribution and trends of the data. The temporal component of time series analysis will also be emphasized, capturing the evolution of the parameters across time. Combining these analytical techniques will produce a thorough and perceptive parameter visualization. A deeper comprehension of the underlying facts and patterns will be made possible by the accurate and educational visual depiction of the parameters provided by the Integrated Analysis System, which is executed in the background.

## Validation process

To achieve accurate findings, I simultaneously used SeaDAS software to manually replicate the image processing approach. Next, in this complex procedure, I personally obtained the required data from NASA's color ocean web, which is an essential step. A number of manual pre-processing techniques are then applied to the obtained data, such as cloud masking, land masking, atmospheric adjustments, georeferencing, and sub setting all of which are necessary to provide reliable findings. In order to extract important ocean characteristics and provide insightful information, the processed data is then statistically examined. During the data validation phase, this output is crucial since it is cross-referenced and compared with the Python output to ensure correctness and consistency. The integrity and dependability of the final dataset, which is essential for additional study and analysis in the context of oceanographic studies, are ensured by this difficult approach.

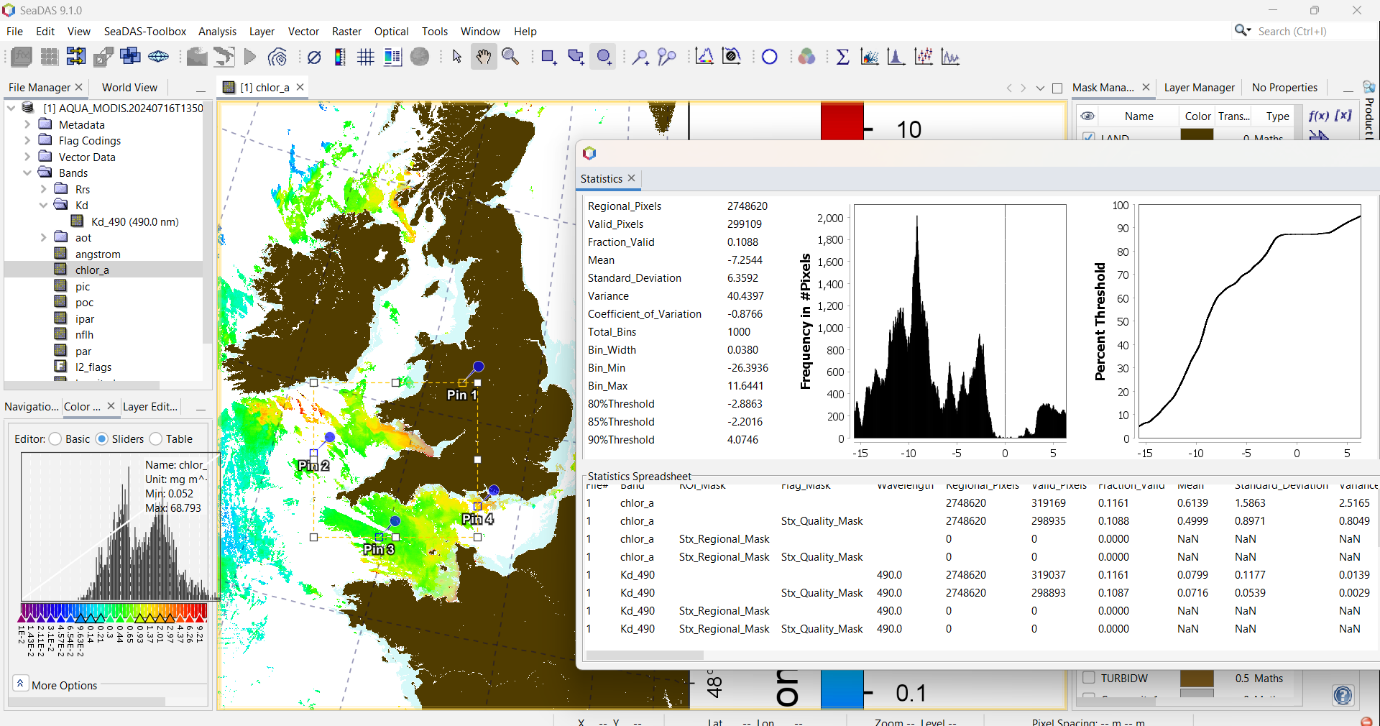


Figure 9. Validation process by SeaDAS Programme

## Implementation of Machine Learning Technology Random Forest for Potential Fishing Zone (PFZ) Prediction

In order to estimate possible fishing zones, the results of integrated analysis are essential, providing important information for fisheries management. Using advanced machine learning technology, specifically a Random Forest regression method, greatly improves the process of producing predictions for possible fishing areas. The task of accurately predicting potential fishing areas falls to this advanced algorithm, which helps the fishing sector make better decisions. Both the training and testing stages are effectively carried out to maximize the prediction models for locating profitable fishing places by using Random Forest regression. This strong strategy builds the groundwork for realistic, data-driven actions to identify the best places to fish, giving that useful knowledge to improve fishing methods and resource use. More efficient methods for maximizing fishing operations and guaranteeing sustainable practices in the future are made possible by the methodical integration of technology and data analysis in this area.

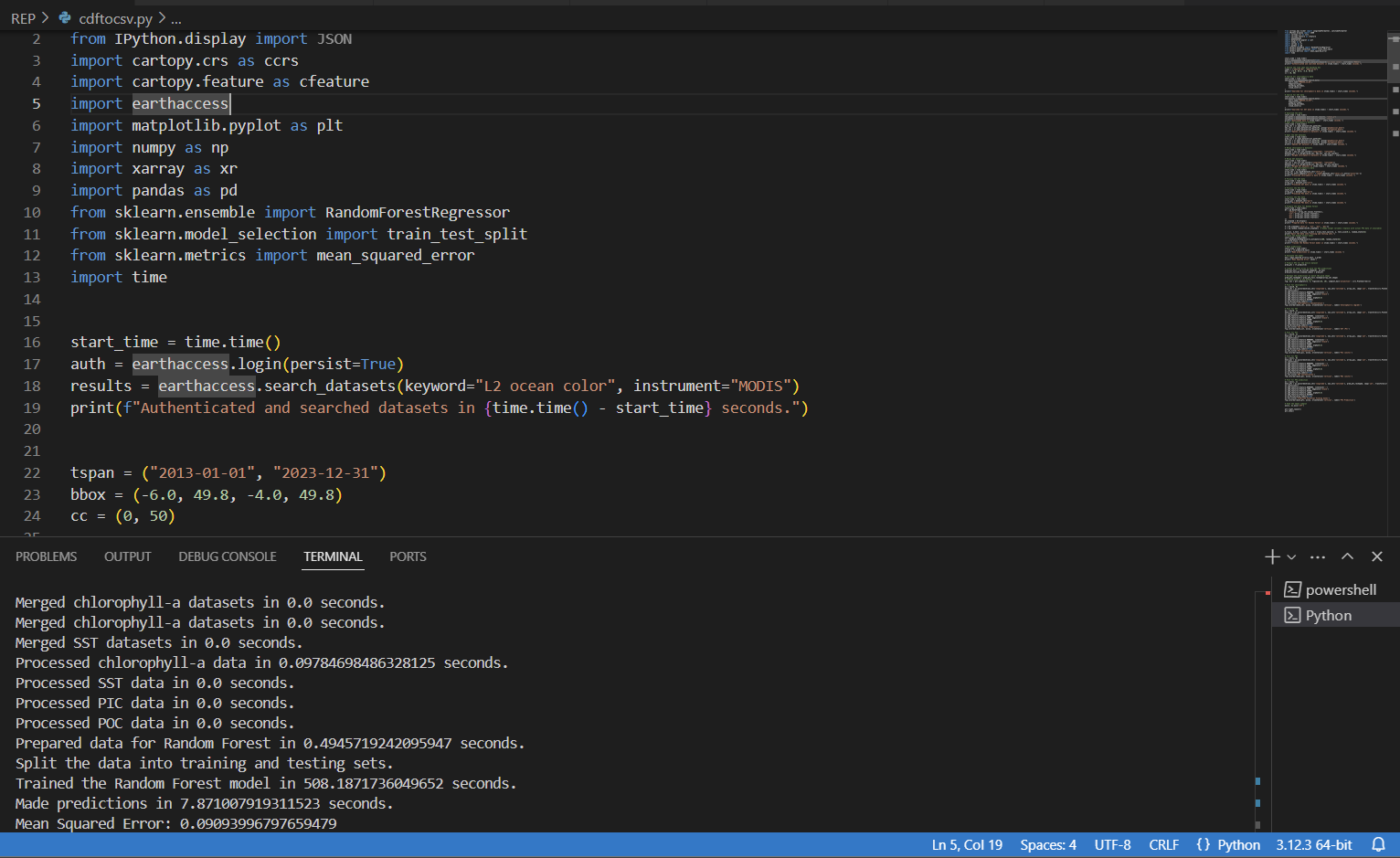


Figure 10. Random forest implementation in Python

Chapter-5

# Results

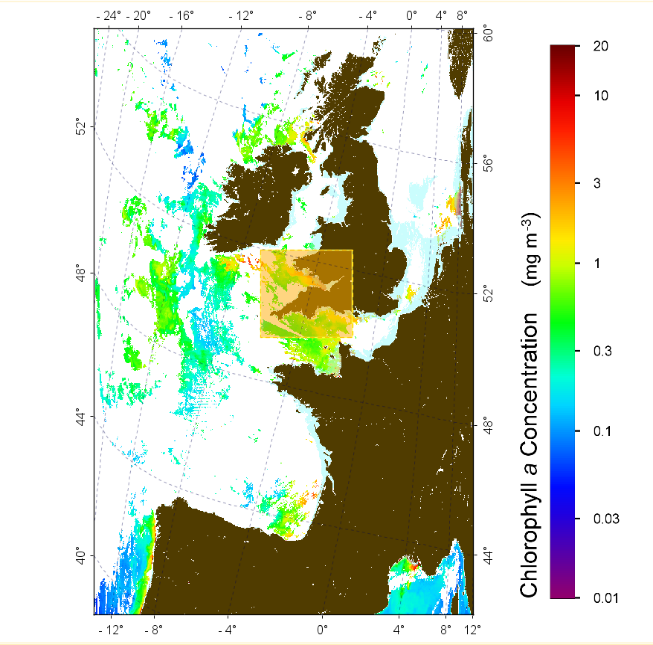


## Introduction

I describe the key findings of the study in the results section, which includes displaying the parameters of marine water quality and pointing out important ranges for high-quality marine water. In order to facilitate the evaluation of patterns in the datasets, I employ scatter plots to illustrate the correlations between parameters. It also includes a visual representation of the predicted potential fishing zones. I also use random forest as a machine learning method and talk about using an integrated model. These techniques support data processing and insight extraction to better understand marine ecosystems for environmental management and conservation.

## Marine water quality Indicator visualisation

With a concentrate on the extraction and display of ocean water characteristics, I provided a detailed explanation of the significant analytical findings in this section. The technique has two visual elements: a statistical model comes after the first, which is based on the result of the SeaDAS processing. To ensure that the findings are not only correct but also dependable for additional analysis and interpretation, complete commitment to the process greatly improves the final output's validity and accuracy. To encourage further research and scientific endeavours in this area, I want to make recognition of the difficulties involved in determining the characteristics of ocean water. The second visualization part presents the finished output in a thorough and detailed manner using Python programming. The SeaDAS output verification procedure generates and validates a precise and targeted outcome by utilizing direct access to satellite data. Because it derives from a complete framework that includes many models, particularly statistical, time series, and geospatial models, the Python-generated result is very impressive. In addition to enhancing the reliability and accuracy of the results generated, this integration ensures an integrated and multiple approach.

1. **Marine Water Quality Parameter Visualisation**
2. **Chlorophyll-a concentration**

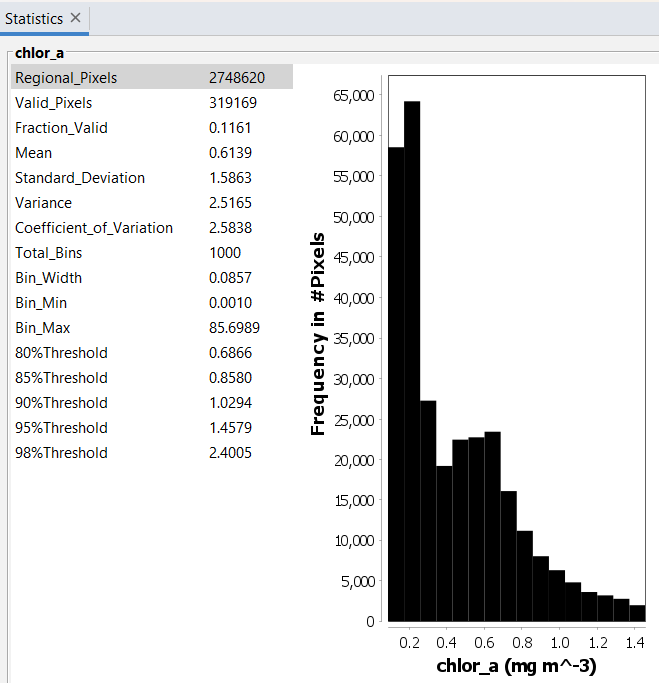


Figure 11. Chlorophyll-a after image process from SeaDAS

Figure 12. Chlorophyll-a pixel frequency after image process from SeaDAS

**Output from Python program**

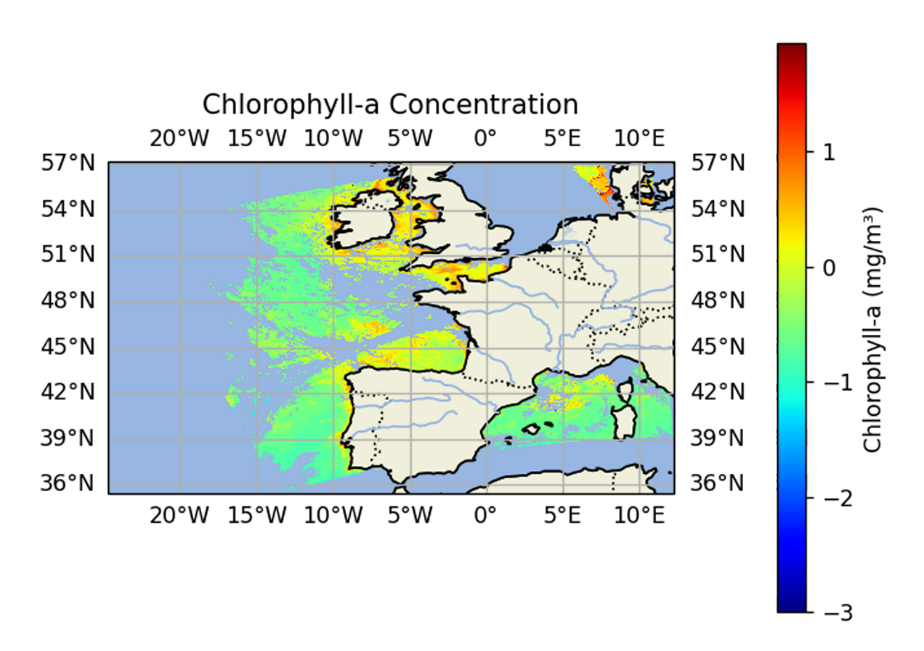


Figure . Final output for visualisation of chlorophyll-a concertation

1. **Sea surface Temperature (SST)**

Figure . SST after image process from SeaDAS

* + **Output from SeaDAS**

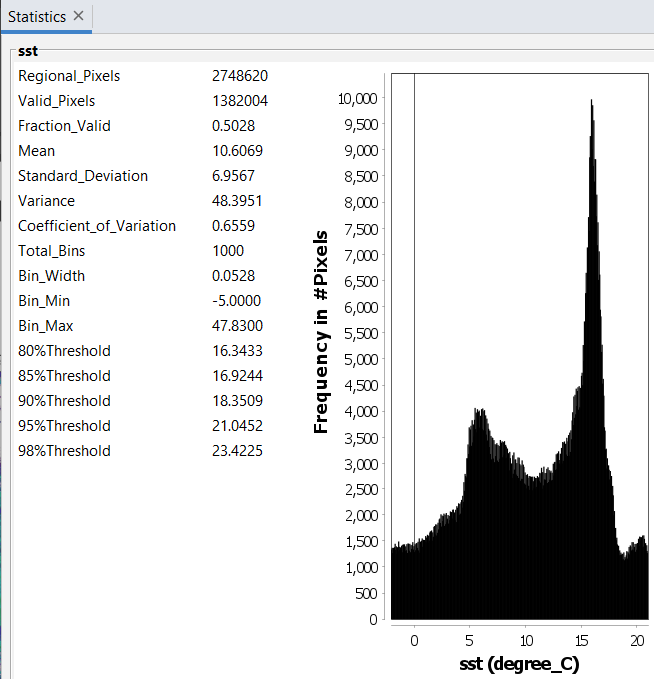


Figure 15. SST pixel frequency after image process from SeaDAS

* + **Output from Python program**

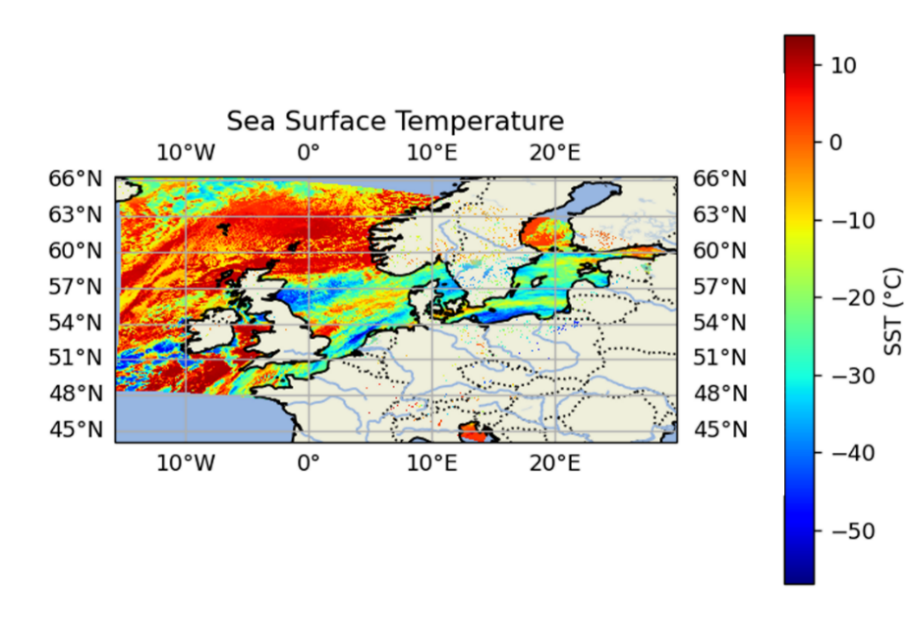


Figure 16.Final output for visualisation of SST variation

1. **Particulate Inorganic Carbon (PIC)**
   * **Output from SeaDAS**

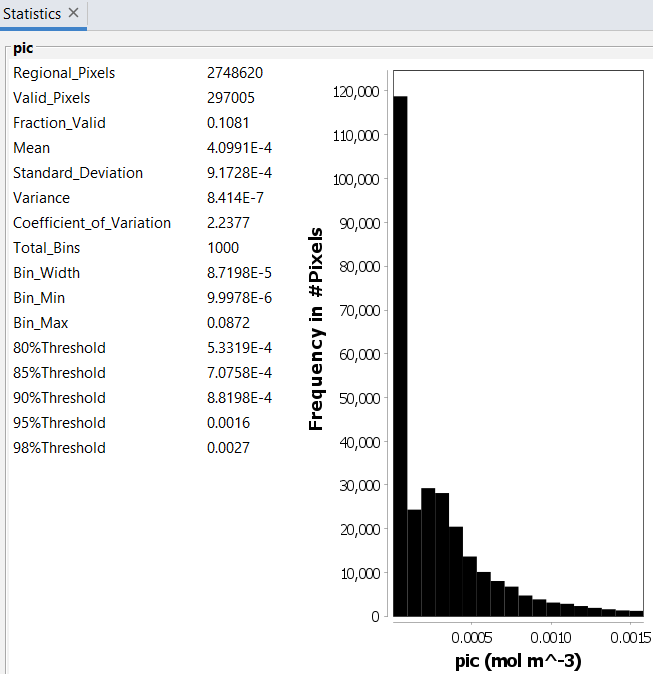
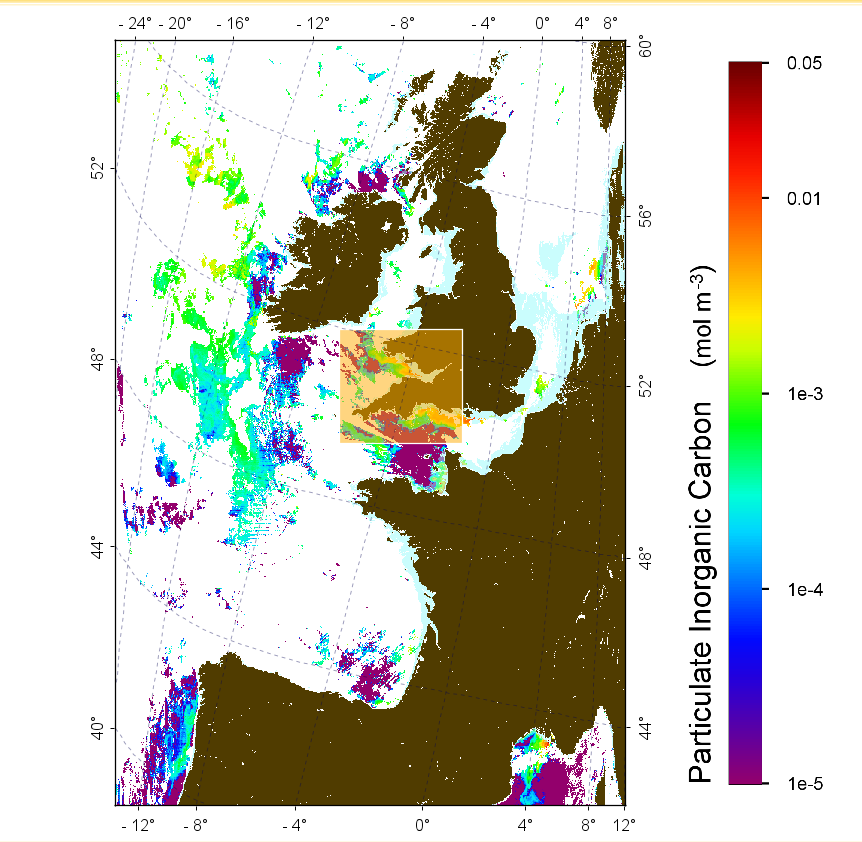


Figure 17. PIC after image process from SeaDAS

Figure 18. PIC pixel frequency after image process from SeaDAS

* + **Output from Python program**

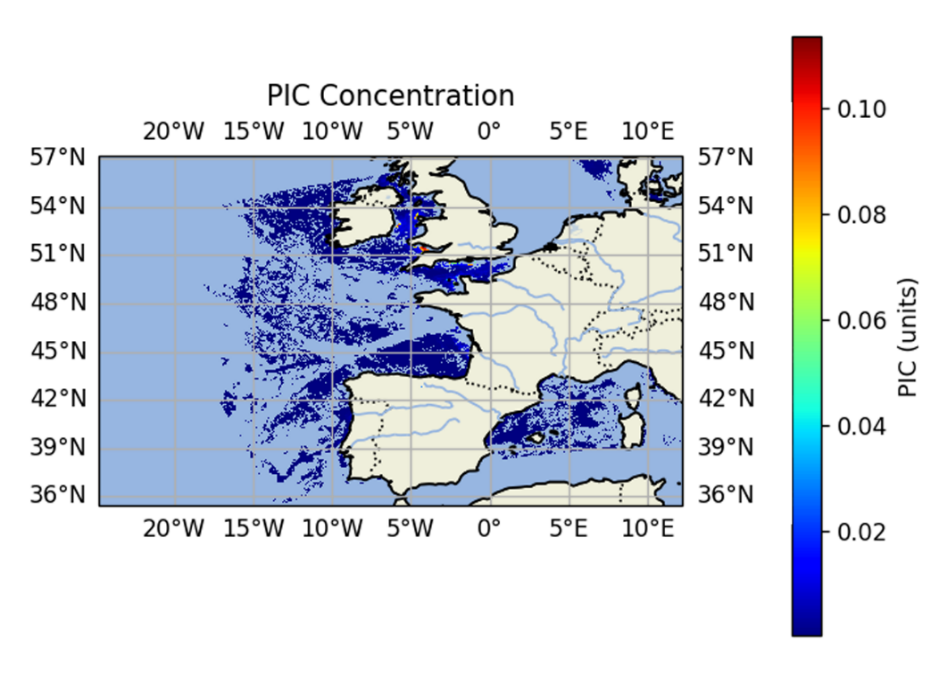


Figure 19. Final output for visualisation of PIC concertation

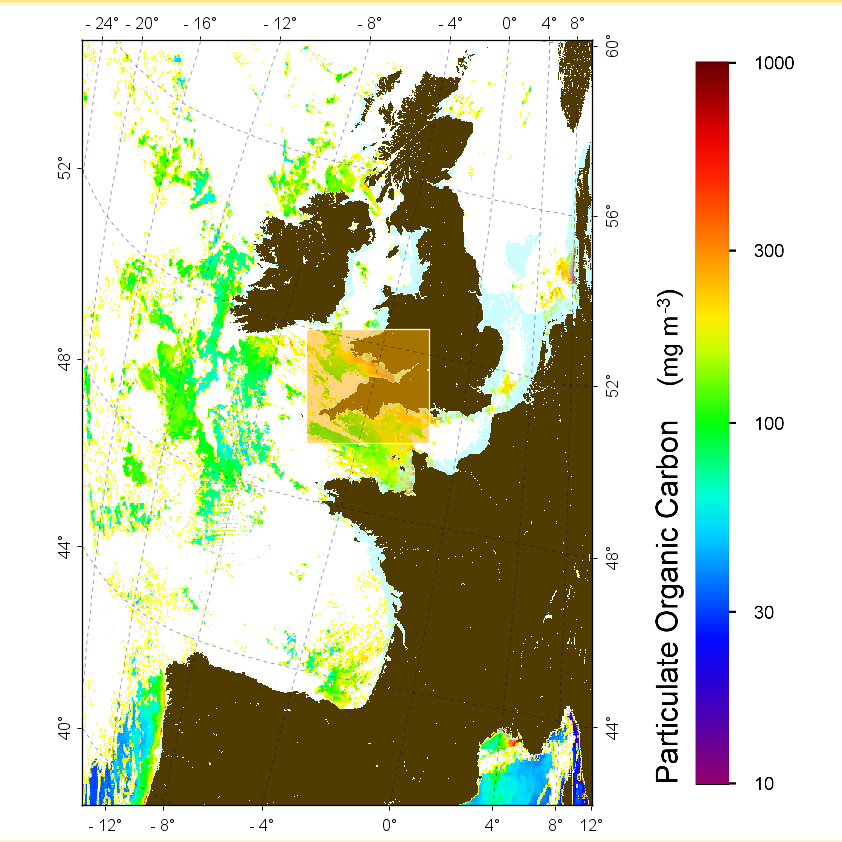
1. **Particulate Organic Carbon (POC)** 
   * **Output from SeaDAS**

Figure 20. POC after image process from SeaDAS

* + **Output from Python program**

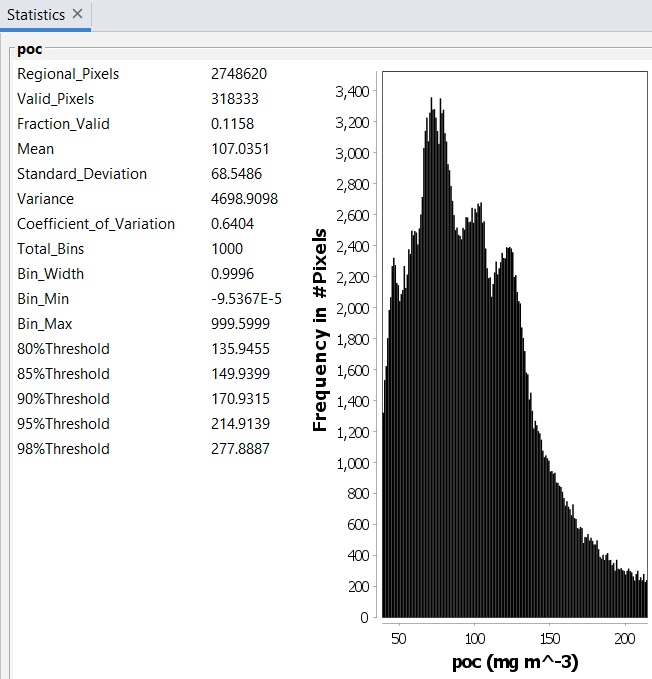


Figure 21. POC pixel frequency after image process from SeaDAS

* + **Output from Python program**

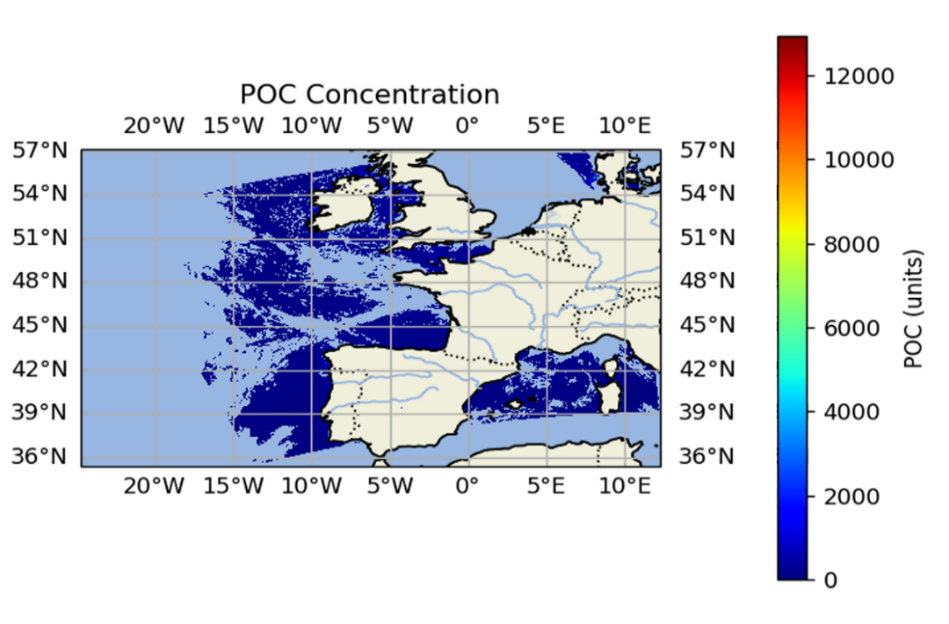


Figure 22. Final output for visualisation of POC concertation

1. **Diffuse Attenuation Coefficient (Kd\_490) - Turbidity**
   * **Output from SeaDAS**

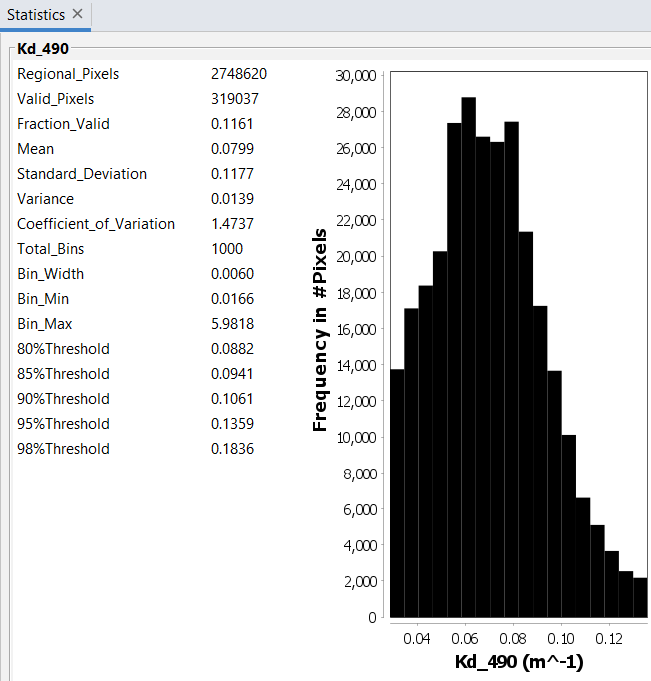
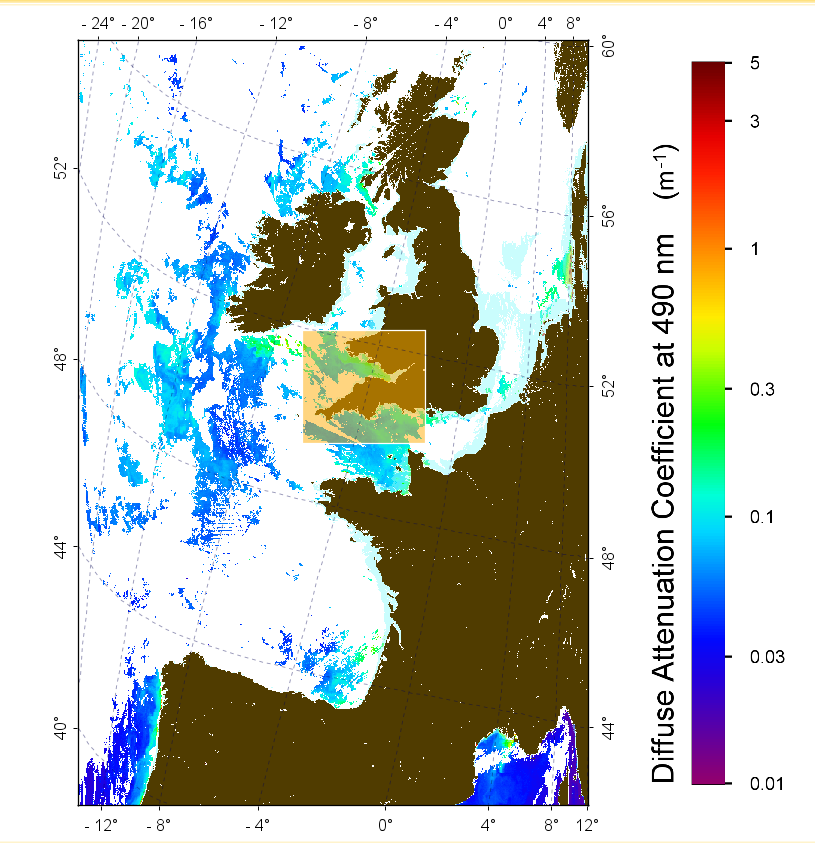


Figure 23. Turbidity (Kd\_490) after image process from SeaDAS

Figure 24 Turbidity (Kd\_490) pixel frequency after image process from SeaDAS

* + **Output from Python program**

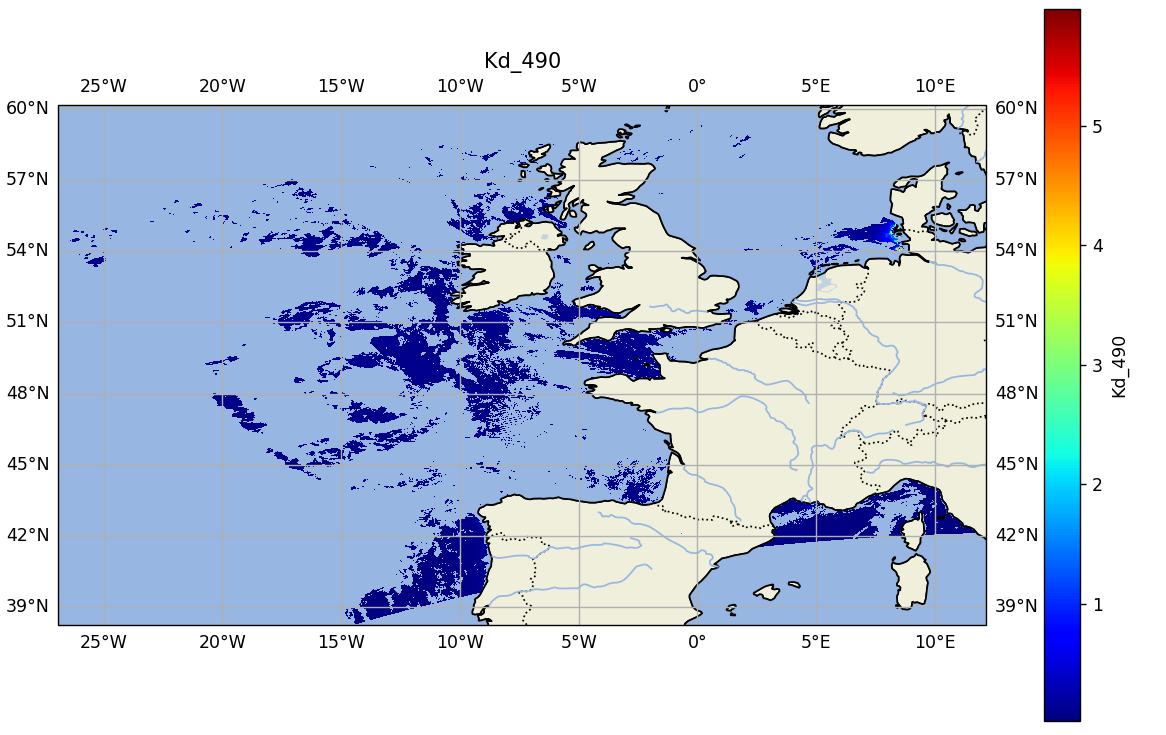


Figure 25. Final output for visualisation of Turbidity (Kd\_490)

* 1. **Chlorophyll-a**

Monitoring marine environments requires the statistical analysis and display of chlorophyll-a data. Higher amounts of chlorophyll-a are shown by warmer colors, while lower levels are indicated by cooler colors in the representation that is presented. This allows the rapid identification of phytoplankton-rich regions, which is essential for managing ecosystem health and marine productivity. The frequency distribution and statistical analysis of chlorophyll-a data provide information on the distribution patterns of phytoplankton. This makes identifying places that are productive and less productive easier. Continuous monitoring informs choices on fisheries and conservation tactics by identifying long-term patterns and evaluating the effects of human activity and climate change.

* 1. **Sea surface Temperature (SST)**

The precise statistical assessment of the sea surface temperature (SST) distribution seen in the picture is provided by the histogram and statistics taken together. It can find out more about the whole range of SST values, identify the dominant temperature ranges, and determine the distribution pattern whether it is distortion or shows features of a normal curve by looking at these graphic and numerical illustrations. Utilizing this statistical data also makes it possible to identify unusual hot or cold locations in the oceanic data, which has significant effects for a variety of fields, including oceanography, meteorology, and fisheries management.

* 1. **Particulate Inorganic Carbon (PIC)**

According to the results and their interpretation, there are notable variations in the regional distribution of the ocean's Particulate Inorganic Carbon (PIC) concentrations. The color-coded map, in which various hues correspond to varying PIC concentrations, clearly shows this difference. The connection between greater PIC levels and biologically active locations, including upwelling zones and coastal areas, is especially significant. By providing a frequency analysis of PIC values and showing an increase of lower concentrations with a noticeable tail toward higher values, the accompanying histogram enhances this visual representation and suggests the significant influence that individual pixels have on the overall variance in PIC concentration. Additionally, comprehensive statistical information such as the mean, standard deviation, and different criteria that are shown in the accompanying table helps us better understand the features of the output.

* 1. **Particulate Organic Carbon (POC)**

The visualization of the POC concentration throughout the area is given by the geographical distribution map, which also provides onto the different amounts of organic carbon present. In order to better understand marine ecosystems and carbon cycling processes, it clearly displays the regions with elevated POC concentrations, indicating locations that could be undergoing significant biological activity or organic matter growth. By providing a thorough frequency distribution of POC readings and highlighting the presence of particular concentration ranges in the area, the histogram enhances this. A thorough picture of the variability and distribution patterns of POC is provided by the statistical summary, which dives into quantitative data such as variance, mean concentration, and standard deviation.

* 1. **Diffuse Attenuation Coefficient (Kd\_490) – Turbidity**

The visualization, which is most likely based on satellite data analysed using SeaDAS software, shows the degree of turbidity or water clarity in the area shown by the Diffuse Attenuation Coefficient at 490 nm (Kd\_490). This illustration effectively conveys turbidity, which is the haziness or cloudiness of a fluid caused by suspended particles. The color-coded map, which shows areas with both greater (warmer colours) and lower (cooler colours) levels, successfully illustrates the geographic diversity in turbidity. Coastal currents, sediment resuspension, and river flow are some of the variables that affect this fluctuation. By providing a clear frequency distribution of Kd\_490 values in the area, the accompanying histogram enhances the map by highlighting common turbidity levels as well as condition variances. The visualization helps better understand the dynamics of water clarity and its effects on the marine environment, making it a useful tool for fisheries management and water quality monitoring.

* **Final Python Program Outputs of all ocean parameters.**

Information obtained from MODIS satellite data, a source known for its precise remote sensing of ocean color, must be communicated using Python output visualizations. The first step is transforming this data using Python algorithms to guarantee accuracy and quality control, with SeaDAS acting as a reliable tool recognized by NASA for thorough data validation. Python's many libraries, which are well-known for their extensive capabilities in data handling, analysis, and visualization, are integrated in the following approach. Time-series analysis, geospatial assessments, and statistical techniques are all seamlessly integrated by these libraries to provide thorough insights into important ocean parameters like sea surface temperature and chlorophyll-a concentration, as well as essential variables like PIC, POC, and necessary Kd\_490 coefficient. The efficient analysis of these complex datasets is ensured by the iterative deployment of Python programs, which turn raw data into showing and visually stimulating representations that are essential for environmental monitoring and scientific research.

## Evaluating Marine Water Health: Critical Quality Indicators

Using scatter plots to illustrate the links between all the parameters as I present the results of identifying the optimal range of marine water quality parameters. Here is a list of the plots that are produced when one parameter is plotted against another. The systematic process for generating these plot pairs helps in the analysis of general trends and patterns in the data sets in addition to offering a clear visual representation of the relationships between the different parameters. I carefully use scatter plots to illustrate the relationships in order to fully present the connections between different parameters. The systematic construction of these plot pairs makes it easier to evaluate broad patterns in the datasets and offers a clear illustration of the relationships between various parameters.

The data that is provided includes the extraction and display of five important parameters: turbidity, Particulate Inorganic Carbon (PIC), Particulate Organic Carbon (POC), Sea Surface Temperature (SST), and chlorophyll-a. To ensure a comprehensive evaluation and analysis, each parameter has to go through a repeated procedure.

## Scatter Plot Visualization of Marine Parameter Interactions

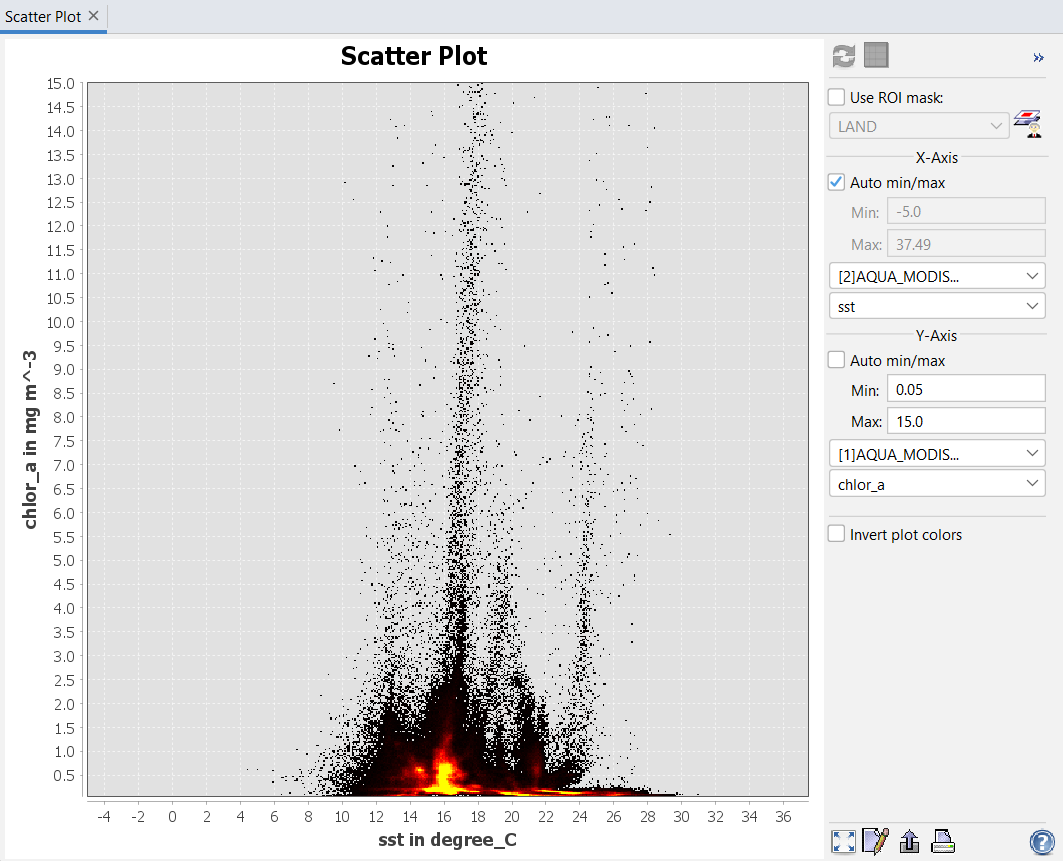


Figure 26. Relationship between chlorophyll-a and SST

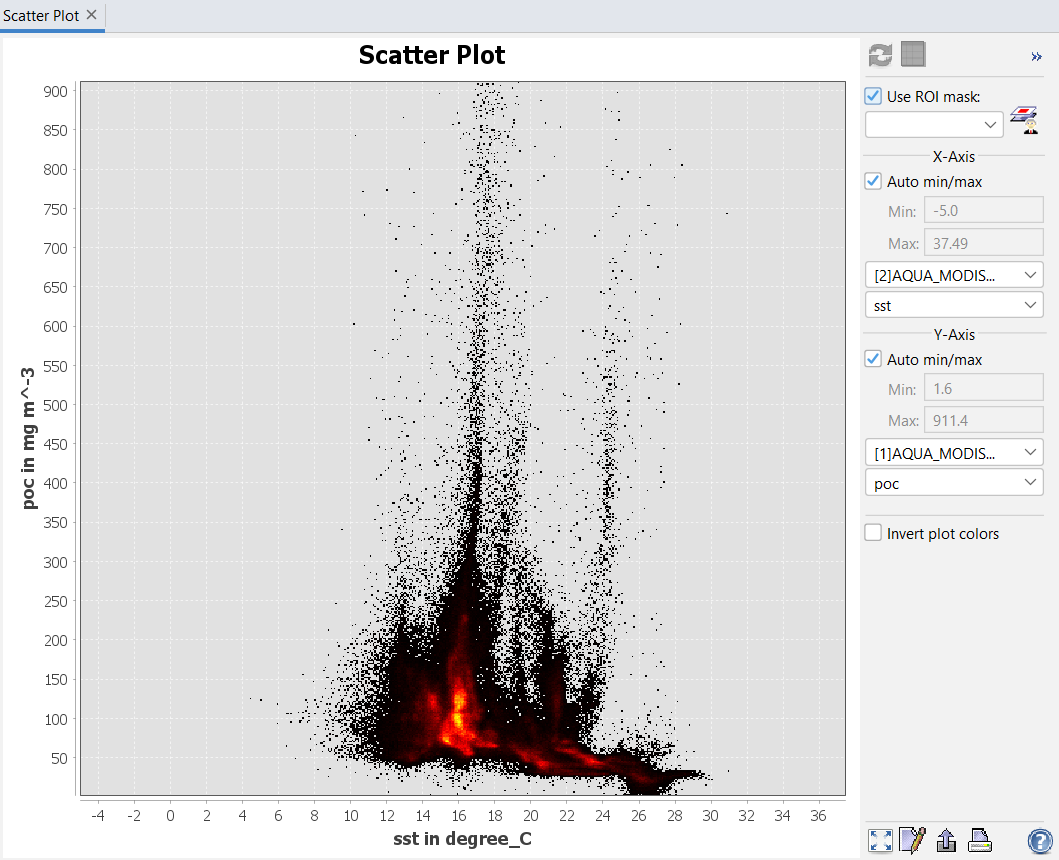


Figure 27. Relationship between POC and SST

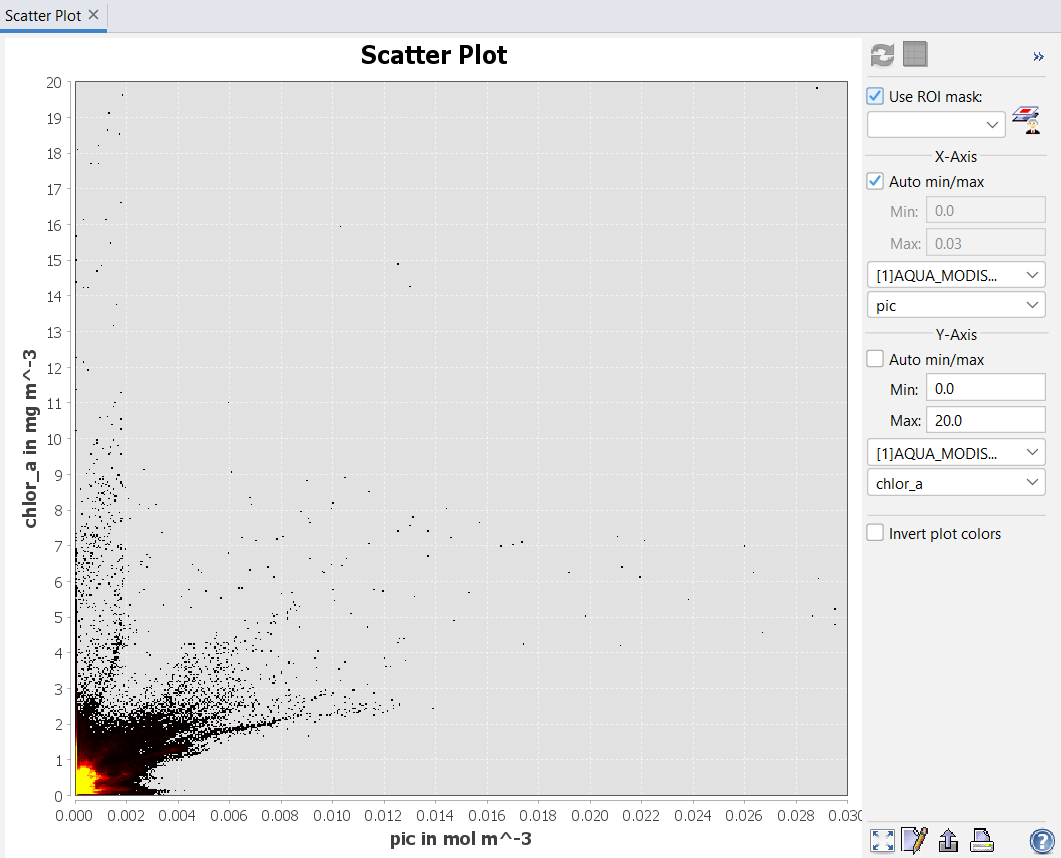


Figure 28. chlorophyll-a and PIC

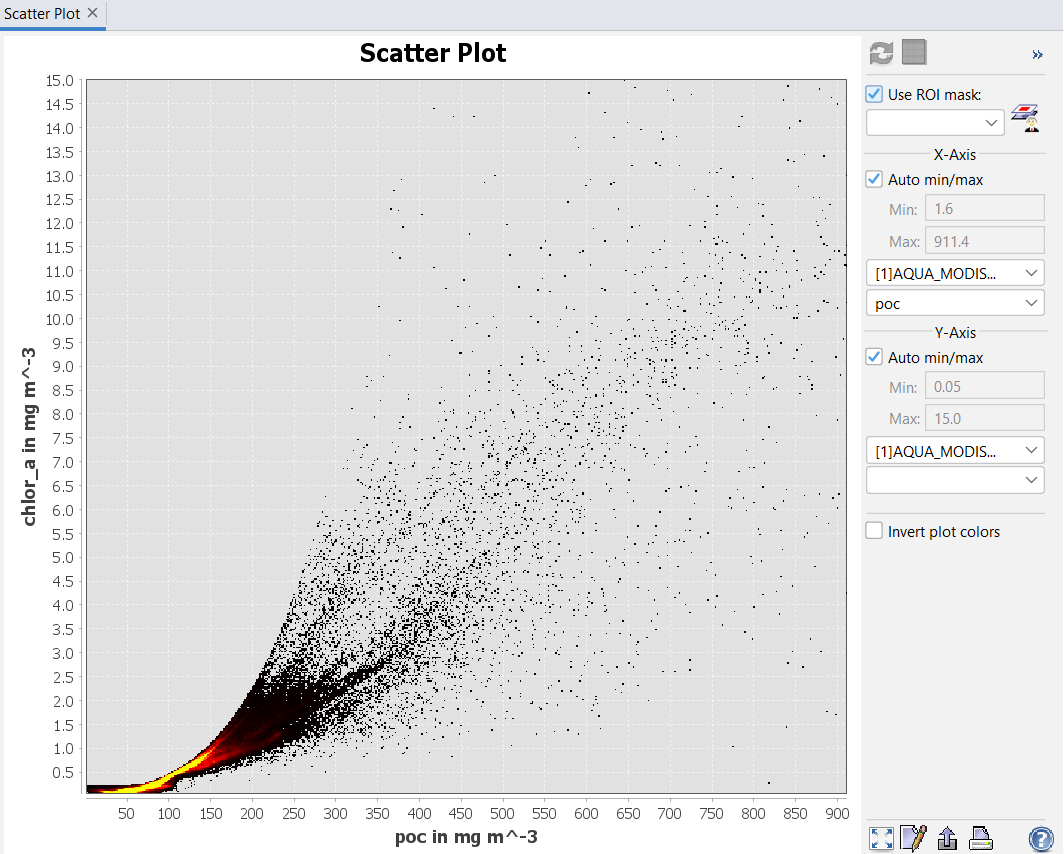


Figure 29. chlorophyll-a and POC

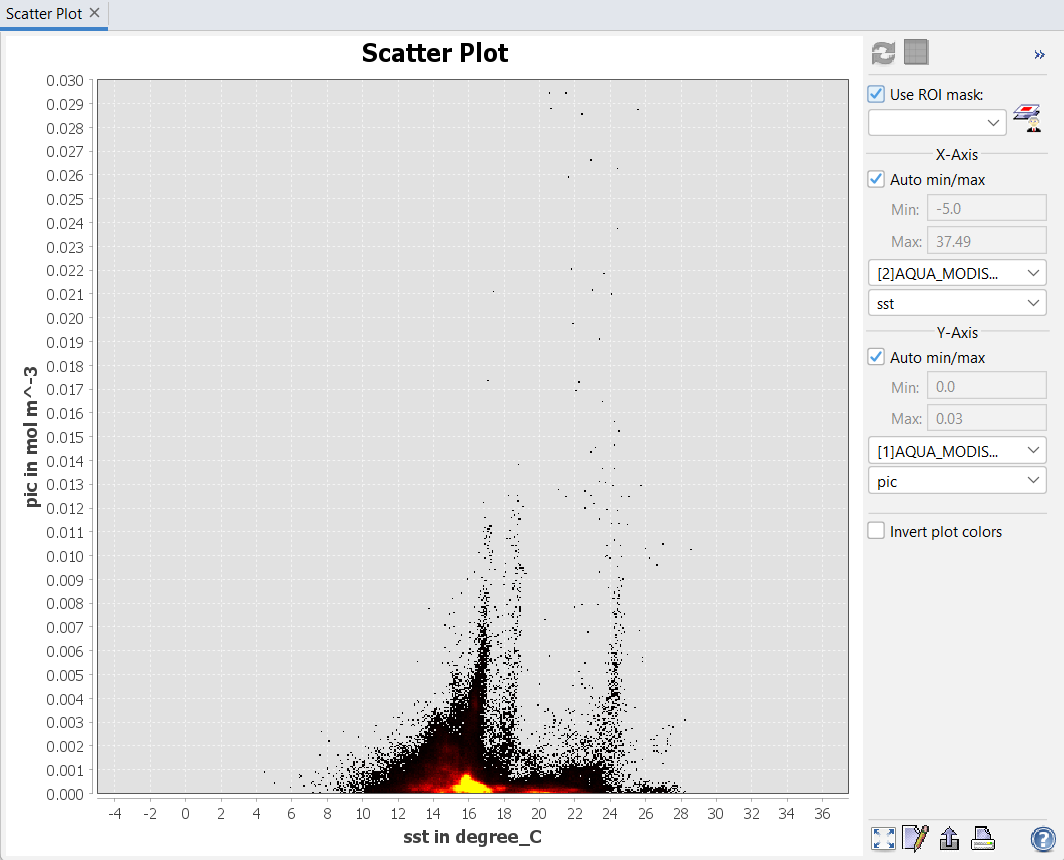


Figure 30. PIC and SST

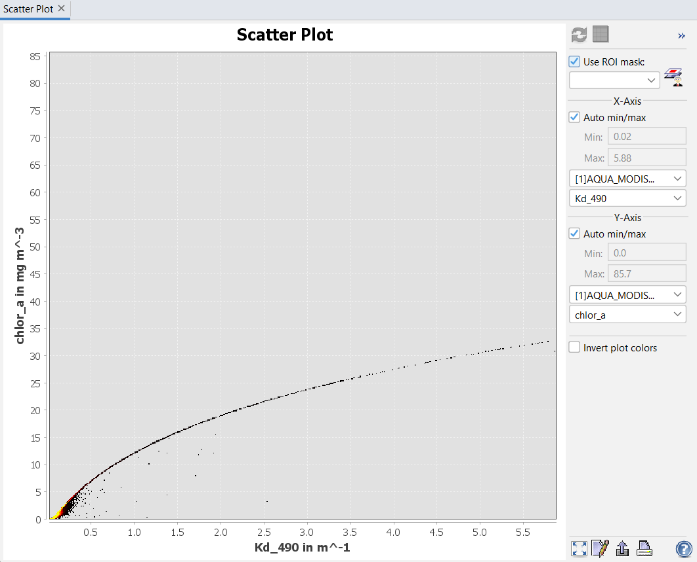


Figure 31. Chlorophyll-a and Kd\_490 nm

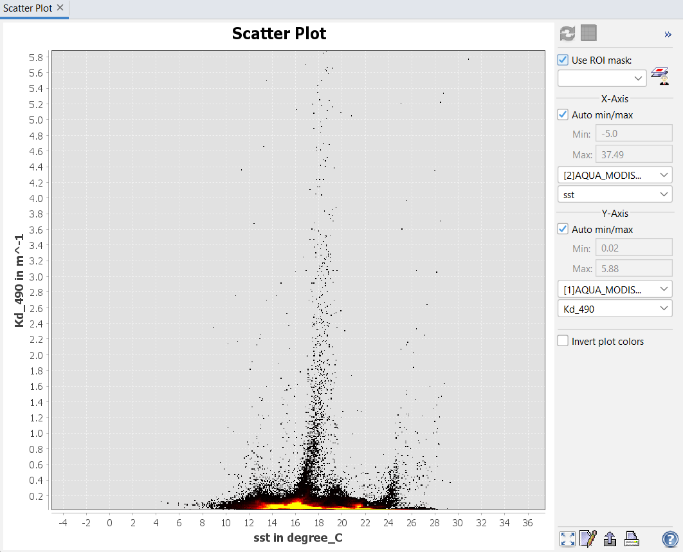


Figure 32. Kd\_490 nm and SST

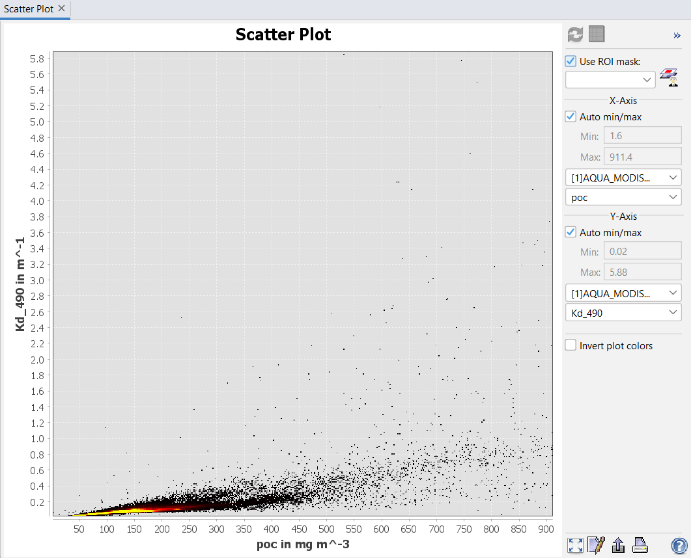


Figure 33. Kd\_490 nm and POC

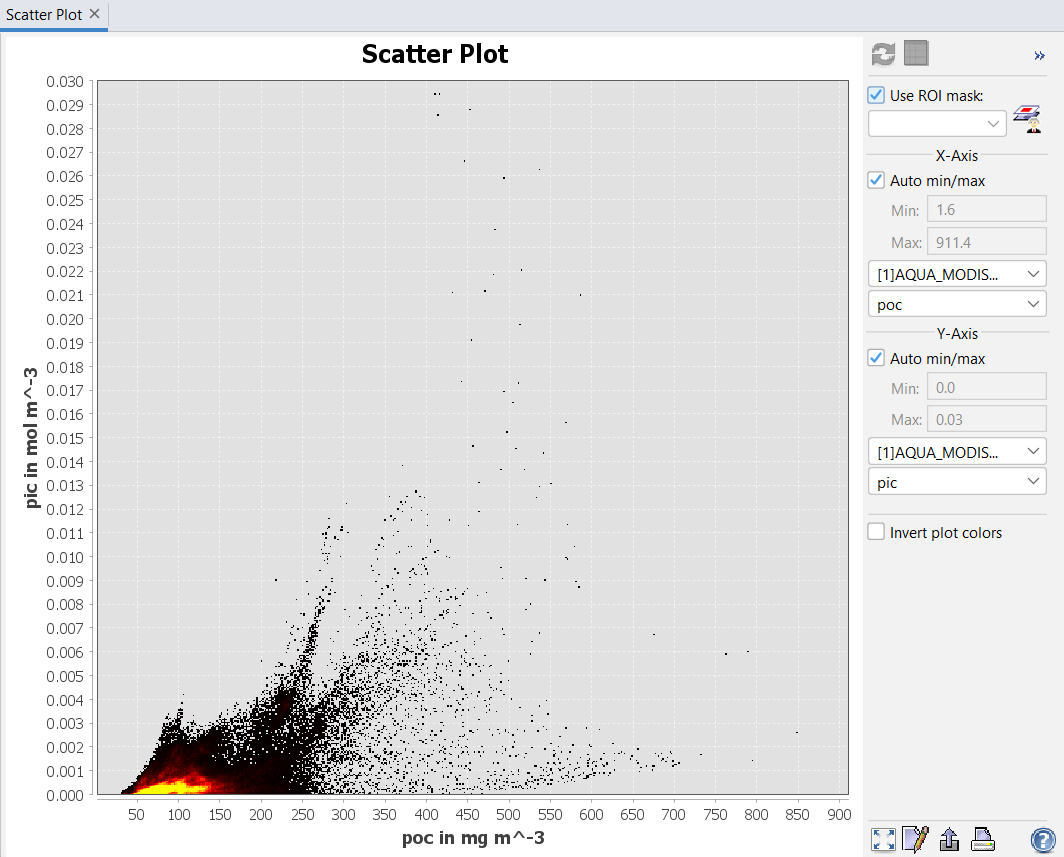


Figure 34. PIC and POC

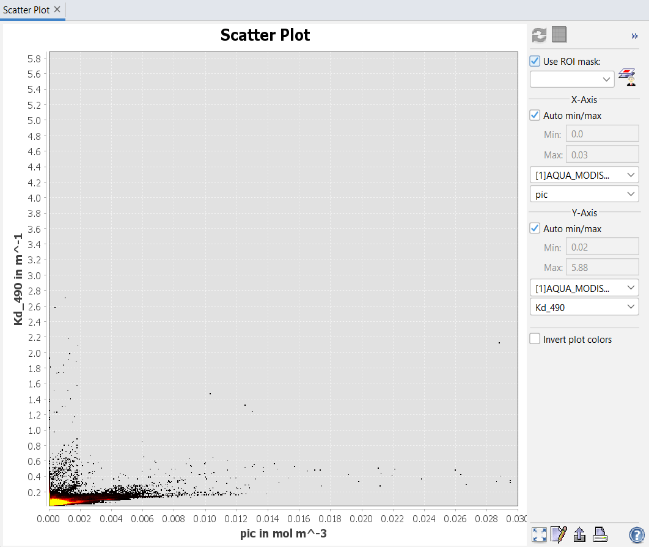


Figure 35. Kd\_490 nm and PIC

## Extraction of scatter plots for Ideal ranges for marine water quality parameters

The ideal ranges for marine water quality parameters, such the content of chlorophyll-a, are essential for recognizing and tracking the condition of marine water quality. The range for chlorophyll-a concentration that was found in the plot is important because it is a measure of phytoplankton levels, which are essential for maintaining the marine food web. Typically, this range is between 0.5 and 5.0 mg/m³. This range ensures sufficient primary productivity without running the risk of dangerous algal blooms that might disrupt the delicate balance in the environment, in addition to indicating the existence of healthy phytoplankton. In addition, considering sea surface temperature directly affects the metabolic rates of different marine life forms, it is essential to keep it within the ideal range of around 12°C to 25°C in order to support their overall health. In the plot's particular temperature range helps to prevent stressful situations that may lead to ecological imbalances by promoting thermal stability that is suitable for the various demands of marine life. The water quality and balance of marine habitats gets successfully protected by following the suggested ranges for important metrics such as particulate inorganic carbon, diffuse distortion coefficient, and particulate organic carbon. For the health and productivity of marine ecosystems, these ranges provide a reference for maintaining a balanced chemical composition and ideal circumstances. In the face of the challenges caused by pollution and climate change, maintaining and improving marine environments needs constant monitoring and constant adherence to these standards.

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Output Range** | **Implications** |
| **Chlorophyll-a Concentration (chlor\_a)** | 0.5 mg/m³ to 5.0 mg/m³ | Healthy primary production without toxic algal blooms is indicated by low to moderate levels. |
| **Sea Surface Temperature (SST)** | 12°C to 25°C | (differs by area) Diverse marine life is supported by suitable temperatures; high temps may be a sign of discomfort of marine primary productivity |
| **Particulate Inorganic Carbon (PIC)** | 0.0 to 0.01 mol/m³ | Reduced pH imbalance in the ocean hazards is indicated by less pollution from inorganic carbon sources. |
| **Particulate Organic Carbon (POC)** | 1.6 to 200 mg/m³ | shows that the organic matter in the water is in a healthy balance, which supports the marine food web. |
| **Diffuse Attenuation Coefficient (Kd 490)** | 0.02 to 1.0 m⁻¹ | Low turbidity and clear water enable deeper solar penetration and promote photosynthesis. |

* **Table of parameter Ideal Range from extracted from the scatter plots**

Table 2. Scatter Plot Visualization of Marine Parameter Interactions

Potential fishing zone (pfz) forecast and integrated evaluation is a difficult procedure that necessitates a thorough examination of several variables. A thorough examination of certain factors, such as sea surface temperature, chlorophyll content, and ocean turbidity, is crucial to the correct forecasting of pfz. Machine learning methods, namely the Random Forest approach, are used to create a strong prediction model in order to improve the accuracy of this study. This prediction model provides a thorough and precise picture of potential fishing regions by combining different data, models, and characteristics. In addition to streamlining decision-making procedures, this integrated method increases the accuracy and efficiency of locating optimal fishing zones. By combining several models and data sources, this approach improves the ability to identify the best places to fish, supporting sustainable resource use and fisheries management.

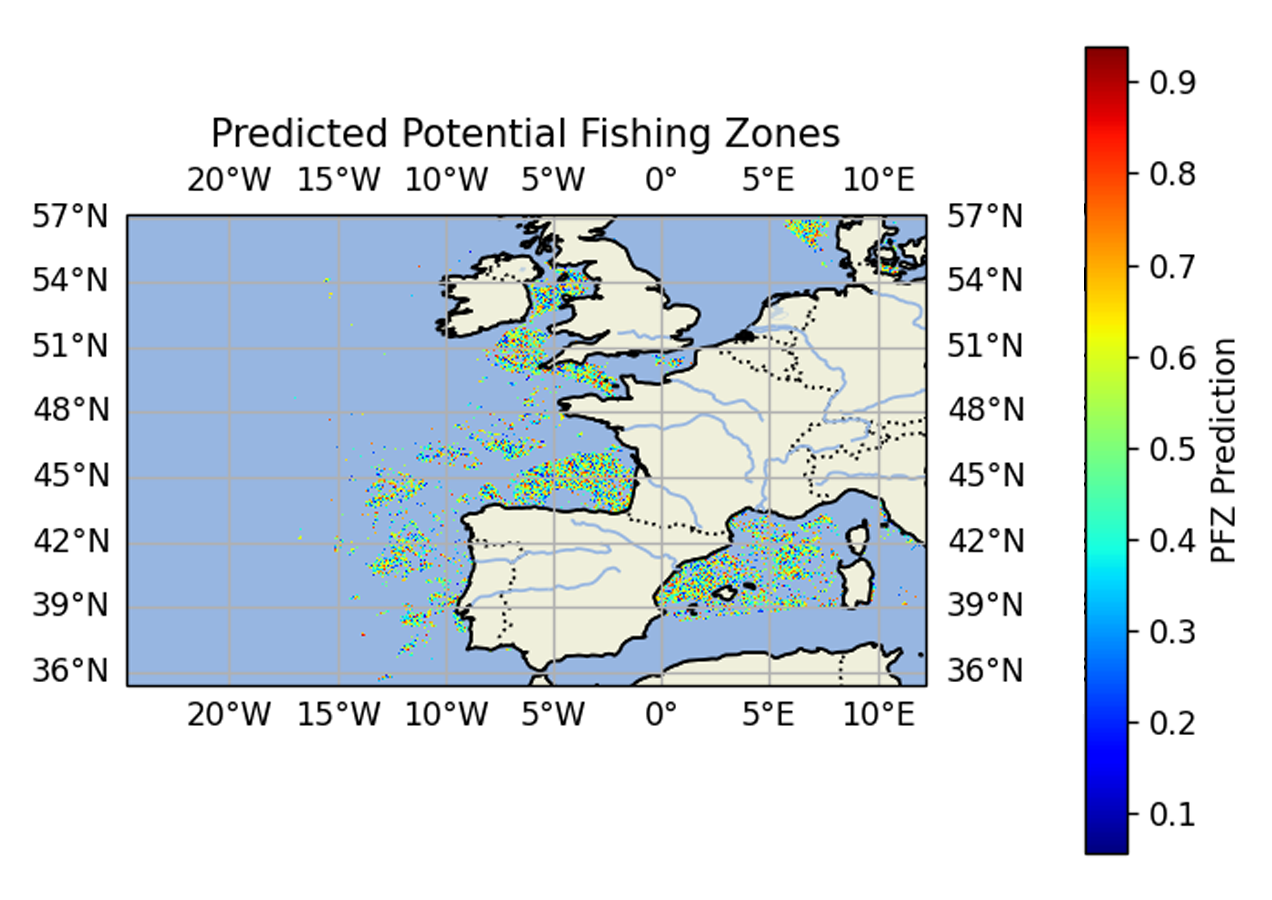


Figure 36. Final output for visualisation of PFZ prediction

The visualisation demonstrates the major effects of the machine learning model that utilizes the random forest approach. By using data from the MODIS satellite, it offers precise information on measurements of marine water quality that are essential to ocean biodiversity. Sea surface temperature, which affects fish behaviour and distribution, and chlorophyll-a, which is essential for phytoplankton resources that marine species depend on, are only two examples of the important parameters. The predictive power of the model increases with particle organic carbon and particulate inorganic carbon, which indicate the availability of organic matter and water clarity, respectively. The attenuation coefficient (490 nm) also assesses how light absorption affects primary production and the depth of light penetration.

If one examines the visual interpretation closely, the representation of different coloured dots over a geographical area, provides an interesting viewpoint. The colour spectrum for the "PFZ Prediction," which shows the potential of good fishing locations, from warmer hues denoting higher possibilities to cooler hues denoting lower projections, conveys a feeling of strategic awareness. Interestingly, the irregular arrangement of dots that suggests some areas with higher chances of productive fishing operations illustrates the model's complex predictive character. The spectrum of knowledge is expanded by taking into consideration the feasible results in for dot distribution, which include environmental factors such as ideal water quality compositions that support rich marine ecosystems and fish behaviour patterns impacted by these conditions.

## **Summary of Result**

In summary, a thorough presentation of the key conclusions. These findings include identifying particular parameter ranges essential for ideal marine water conditions and graphically presenting parameters essential to marine water quality. Effectively illustrate the relationships between various parameters with scatter plots to provide a thorough grasp of their correlations. By systematically constructing these plot pairs, it is possible to better analyse broad patterns in the datasets and demonstrate how different parameters are related to one another. The study also explores the prediction and visual representation of possible fishing zones. Furthermore, using the random forest machine learning technique and talking about the integrated model are essential for improving data processing and gathering valuable knowledge that greatly advances our understanding of marine ecosystems for environmental management and conservation. In the end, the findings highlight how important it is to keep water quality within the established ranges in order to maintain a healthy marine ecosystem, taking into various factors like stable temperatures, healthy phytoplankton levels, lower pollution levels, and clear water for successfully supporting marine life.

Chapter-6

# Discussion and Evaluation

This portion of the study explores important material that forms the basis of the investigation. It specifically explores the contrast between the current research and the past studies, offering a thorough examination of the parameters utilized in the earlier researches as opposed to those used in this study. It also carefully analyses the models and approaches used, highlighting how they vary or connect with those of earlier studies. The data collecting period is another important topic covered, with careful consideration given to how the time range may affect the study's findings. In addition to detailing the complex relationships between every aspect examined, the result section outlines the specific features that the research revealed. The part also critically examines the study's essential limitations, offering insights into the parameters that regulate its operations. Additionally, it provides strong suggestions for future initiatives, outlining potential paths for additional investigation and improvement in the field of study. This part provides a thorough and informative explanation of the study's methodology, conclusions, and suggestions for further research, guaranteeing a strong basis for the study's contributions to the area.



## Evaluation of existing studies with this study

Multiple ocean parameters are analysed in this study, which is crucial to the investigation. Sea surface temperature (SST), particle inorganic carbon (PIC), particulate organic carbon (POC), Kd\_490, and chlorophyll-a are the five main ocean metrics that are the subject of this study. These parameters serve as the study's foundation and produce significant results. This study differs from others in that it looks at a different set of parameters. This examination spans a longer duration and a wider variety of parameters than previous studies, which examined fewer factors and shorter timeframes. This study attempts to improve knowledge of the complexities associated with these marine factors by utilizing datasets that cover a minimum of 10 years.

The research methodology is very different from previous studies. used advanced tools like Python and SeaDAS to apply a variety of models, including statistical, geographic, temporal, and integrated models. In order to assist this study, machine learning models are essential. Deeper analysis and new perspectives on the subject are made possible by the unique methodology, particularly blends and differentiates these models and approaches. Attempt to handle complicated problems and make a significant contribution to the field's existing body of knowledge using an integrative strategy.

Overcoming the disadvantages of earlier studies has improved in recent years. The careful assessment of several ocean parameters, which improves the accuracy of results, is one significant development. Furthermore, processing output data for additional analysis and prediction has become easier through the integration of a greater range of historical data. Numerous previous studies missed the crucial task of forecasting future changes in water quality and potential fishing zones for the purpose of concentrating exclusively on water quality study.

This study, differentiates itself as a more forward-thinking and creative endeavour in the field of ocean environmental research by including predictive elements into its framework in addition to doing a comprehensive water quality examination.

## Interpretation and evaluation of main key findings

* 1. **Marine water quality parameters visualization**

To properly assess the quality of sea water, a thorough selection procedure was used to identify the criteria that would provide the most useful information. Using the variety of ocean data at present to choose which ocean metrics to focus was a crucial part of this procedure. Chlorophyll-a, sea surface temperature, poc, pic, and turbidity are among the critical factors that have been determined to be vital for the assessment of marine water quality by a comprehensive analysis of previous studies and research. In order to guarantee an effective assessment, these particular ocean characteristics were specifically chosen. Additionally, the data was presented in an understandable and instructive way through in large part to the visualization component. A complex visualization process was created to offer an enhanced and perceptive picture of all five ocean characteristics by creating algorithms and procedures utilizing tools like SeaDAS and Python programming. The result of the visualization process was a visually appealing representation that provided an extensive understanding of the marine water quality evaluation.

## Evaluation of marine water Health with Ideal range

This study's second focus examines the evaluation of marine water health in respect to the defined optimal range for a healthy ocean ecosystem. By improving the general comfort and wellbeing of marine life, this evaluation seeks to support a healthy and stable marine ecosystem. The analytical method mostly uses statistical techniques that take in to consideration a number of factors that are important in assessing the quality of water. A reliable indicator of the marine water quality in the particular research region is provided by the values derived from these several measures. A detailed scatter plot examination of these elements' interactions reveals a clear picture of their linkages, illustrating the complexity of the marine environment. The final objective of this assessment procedure is to reach firm conclusions on the general water quality in the research region. According to the research, most of the areas have extremely excellent water quality standards, which suggests that marine life and the ocean ecosystem as a whole are growing and surviving. The importance of healthy marine ecosystems in promoting flexible ecological systems and sustaining marine biodiversity is highlighted by this confirmation of high-water quality.

## Prediction of potential fishing zones

In the course of this study, one of the pivotal discoveries that emerged at the climax was the ability to accurately predict potential fishing zones. This prediction holds utmost importance in the realm of this study, serving as the central focus. All findings and analyses conducted throughout the research process converge to bolster and refine this final predictive capability. Particularly, the core elements contributing to this predictive model revolve around the productivity of fish populations, which plays a vital role in the identification of potential fishing zones. This productivity is intricately linked to the abundance and distribution of phytoplankton, serving as the foundational basis for ecosystem dynamics in marine environments. It is feasible to predict and support fish population expansion by comprehending and utilizing these relationships, creating a strong and long-lasting ecosystem in marine environments. Consistent surveillance and preservation of marine water quality, as well as careful attention to critical criteria that determine the stability and health of marine habitats, are essential components of this process. A self-sustaining marine ecosystem is created when these factors are maintained at their ideal levels, supporting a dynamic food chain that sustains a wide variety of marine species. The richness of marine life in certain regions is ultimately sustained and preserved via the effective identification and maintenance of these flourishing ecosystems, improving the environment and the standard of living of people who depend on these marine resources.

## Contributions of this study

An important advancement in the field is the creation and application of advanced prediction models in fisheries management. Armed with these state-of-the-art instruments, fishermen are able to obtain extremely precise information on the likely locations of fish populations, which transforms their tactics and results. These models are essential for reducing fishermen's ecological mark in addition to enabling them to make wise and informed decisions that maximize their catches and operational effectiveness.

Fishermen can improve their operations and make their business more environmentally friendly and sustainable by utilizing advanced prediction models. Their productivity and resource utilization are improved, and their quality of life is improved while marine ecosystems and ocean life are preserved.

Combining these current forecasts with traditional fishing techniques can strengthen the fishing industry's durability and improve the interaction between human activity and the environment. The advantages of these models go in addition their effectiveness; it is also inspiring a new strategy focused on environmental preservation and improved fish resource management.

## Challenges of this study

One of the biggest challenges I encountered while studying was a shortage of fresh research on this subject from the previous five years. However, I pushed and discovered some appropriate studies for insightful knowledge. Several fields of knowledge are combined in this study. It makes use of remote sensing to gather ocean data, marine biology to investigate fish behaviour and habitats, data science to analyse and visualize data, and Python programming for a variety of functions. It takes time, experience, and a broad perspective to grasp the links between multiple fields in order to become effective in them. Due to data gaps caused by frequent cloud cover and severe weather, collecting accurate data was extremely difficult. To overcome missing data, advanced approaches and more data sources were required. Careful thought was needed to integrate satellite photos, environmental data, and sparse historical fishing records; this frequently involved manual work.  and Python programming. It was necessary to test several Python methods and optimize models for dependable predictions in order to create appropriate models for Potential Fishing Zones (PFZs).

The multidisciplinary character of the study and the inherent difficulty of gathering, collecting, and evaluating marine data were the main causes of its complex set of obstacles. Through careful data selection, rigorous preprocessing, and iterative model construction, the researchers successfully addressed these issues, enabling them to carry out their study and make a valuable contribution to the area of fisheries management.

## Limitations of this study

This study has a number of issues. Important elements such as depth data, in-situ data, and historical fish capture statistics are absent. To properly verify Potential Fishing Zones (PFZ), these components are required. Understanding marine ecosystems, particularly the biological diversity in deeper oceans, might be improved by using depth data. To increase accuracy, historical records and in-situ data are required. Although satellite data is useful, actual samples might provide more accurate results. Filling in these gaps is necessary for a thorough evaluation.

This research focused primarily on satellite data and only employed one sensor, the MODIS (Moderate Resolution Imaging Spectroradiometer). The absence of numerous sensor data might increase accuracy by obtaining more marine water information. More accurate predictions of potential fishing spots as well as improved overall analysis and data presentation may result from the use of many sensors. Research that incorporates multiple marine water properties from several sensors may help us better understand the complicated marine environment.

The research concentrated on the Cornish Coast in the United Kingdom for a ten-year period, from January 2013 to December 2023. Although the findings offer insights specific to this area and period, precise approximations depend on the availability and quality of the data. The study's techniques could be useful in various situations, but long-term data gathering and thorough validation would be required for predictions.

In the prediction stage, the current method fails to correctly categorize certain fish species and their names. Only the general existence of fish groups is taken into account. Because of this limitation, it is challenging to identify particular fish species in fishing grounds. Future studies should concentrate on including species identification because accurate categorization is crucial for both commercial and ecological reasons. A thorough categorization of species would improve forecasts, which would help conservation and fishery management initiatives.

In conclusion, it is critical to address the shortcomings found and develop the available data in order to improve the findings' reliability, particularly the accuracy of Potential Fishing Zones. Incorporating missing data, such as historical fish capture records and depth data, can improve our awareness of marine habitats. Furthermore, the precision of the results may be increased by utilizing several satellite sensors rather than just one.

Because the study was carried out on the Cornish Coast of the United Kingdom between 2013 and 2023, concern must be used when extending its findings to other regions in the absence of reliable data. The prediction capabilities of the current approach are limited since it can't differentiate between different kinds of fish. In order to improve fisheries management and conservation, future studies should concentrate on species-level identification. It is imperative that these data gaps be filled in order to evaluate Potential Fishing Zones and create sustainable fishing methods.

Chapter-7

# Conclusion and Recommendations

This section mostly highlights the study's main findings, including its conclusions, limitations, and suggestions. It mainly provides an overview of the goal and key conclusions of the study. Finally, it explores suggestions, pointing out how the research might go in the future and recommending the incorporation of innovative elements for future investigations. By focusing these important elements, the section aims to give an in-depth summary of the study's findings, highlighting its impact and suggesting directions for future exploration in the area.



## Conclusion

Marine ecosystems are essential to the world's economy and food supply, currently they are being challenged by both human activity and climate change. For sustainable management, it is critical to keep an eye on these ecosystems. New technologies in machine learning and remote sensing provide better monitoring compared to expensive and constrained traditional approaches. By evaluating marine water quality using machine learning, Python, and satellite data, this study seeks to fill up existing data gaps and identify the best fishing locations.

This portion of the results focuses on the significant analytical results related to ocean water parameters. Two visual components make up the method: one uses data processed by SeaDAS, and the other is based on statistical models. For future study, this method improves the findings' accuracy and validity. The second section displays Python-generated output that has been cross-checked with satellite data and merged with several models to improve reliability. By creating graphs that connect each parameter for trend analysis, scatter plots are utilized to investigate the connections between marine water quality metrics. To determine the ideal ranges for each parameter such as turbidity, SST, POC, PIC, and chlorophyll-a repeated analysis is essential for keeping an eye on marine health and avoiding ecological imbalances.

Certain characteristics help to keep the ecosystem balanced against pollution and climate change. It is difficult to forecast the potential fishing zone (PFZ) and involves looking at variables such SST, turbidity, and chlorophyll content. The Random Forest machine learning technique uses a variety of data to generate an accurate model for identifying fishing spots. In addition to taking into consideration environmental elements that impact fish behaviour and marine ecosystems, it displays fishing productivity.

The analysis of ocean water conditions and the measurement and visualization of key water quality indicators are the main topics of the study. It helps to understand the links between different water characteristics by displaying them using scatter plots. In order to enhance data processing, the study uses the random forest machine learning approach to forecast fishing zones. A healthy marine ecology depends on maintaining water quality within advised range, which take pollutants and stable temperatures into consideration. The study provides information about ideal oceanic conditions, particularly machine learning for forecasting fishing zones. It indicates how crucial maintaining water quality is to sustaining marine habitats. These models assist fisherman improve their strategies and outcomes by giving them accurate information on fish locations. Better productivity and less ecological damage result from this, enhancing livelihoods, preserving marine environments, and promoting sustainable fishing methods.

## Limitations

This section of the investigation highlights a number of important weaknesses. Important data, such as depth data, in-situ data, and thorough historical fish capture statistics, are missing. Accurately evaluating Potential Fishing Zones (PFZ) requires these elements. Particularly in deeper waters, depth data can improve our comprehension of marine ecosystems. Historical fish records and in-situ data are also essential for increasing the precision of the research. The MODIS sensor provided most of the satellite data. The precision and quality of the data might be increased by using additional sensors. The research focused on the south coast of the United Kingdom, specifically the Cornish Coast, and collected data over a ten-year period, from January 2013 to December 2023. As a result, the results are restricted to this region and period, and it will take time to validate these forecasts.

Additionally, the study does not categorize specific fish species, which restricts accurate estimates of fish species in the PFZ. For improved ecological and commercial understanding, species-level identification should be a part of future studies. For the study's conclusions to be strengthened and sustainable fisheries management to be advanced, it is important that these data gaps be filled.

## Recommendations

As I look for suggestions, my main priorities are getting beyond present limitations and improving the abilities which were gained from this research. Adding more important criteria is one important technique to open up possibilities for further research that will increase the accuracy of the results. This will be accomplished by using data from numerous satellite sensors, which will make it easier to integrate different datasets and include a wider variety of ocean characteristics. Another important strategy is to use past fish capture data to create prediction models for potential fishing areas, which is an essential first step in our future research projects. The exploration of traditional fishing knowledge obtained from the local fishing community in the research region will be intimately linked to this endeavour.

Last but not least, one of the study's main strengths is its in-depth examination of fish species classifications, which calls for the use of innovative algorithms and techniques to further research. By adopting these cutting-edge methods, we stand to expand our knowledge of the topic and open up new paths for investigation and learning.

Based on my observations, these advices are invaluable. Their main goal is to use modern technology to address and overcome the limitations of the current study. By doing this, I hope to improve the precision of forecasting possible fishing grounds and actively promote conservation initiatives and ethical fishing methods. This endeavour will increase production in this area and support the sustainable management of marine resources. Additionally, advancing the long-term sustainability and well-being of our oceans and marine ecosystems will be greatly aided by the combination of innovative scientific research and modern technology. In addition to placing a major priority on innovation, this strategic approach highlights how crucial it is to manage resources responsibly in order to preserve the marine environment for coming generations.

## Final summary

This section summarizes the study's key conclusions, restrictions, and suggestions. In order to identify the best fishing locations, ocean parameter visualisation, the study uses machine learning and remote sensing to assess the quality of the marine water. To increase accuracy, it makes use of satellite data from the MODIS sensor. The findings highlight the necessity of using the Random Forest approach to forecast fishing zones and manage water quality standards. The study's primary shortcomings are the absence of historical fish records and depth data, as well as its ten-year focus on the Cornish Coast without species-level identification. To promote sustainable fisheries management and conservation, recommendations include utilizing additional satellite sensors, including data from fish catch, and utilizing advanced algorithms for species identification.

# Reference

Adjovu, G.E., Stephen, H., James, D., Ahmad, S., 2023. Overview of the Application of Remote Sensing in Effective Monitoring of Water Quality Parameters. Remote Sensing 15, 1938.

Arabi, B., Salama, Mhd.S., Pitarch, J., Verhoef, W., 2020. Integration of in-situ and multi-sensor satellite observations for long-term water quality monitoring in coastal areas. Remote Sensing of Environment 239, 111632.

Ariana, M., Suyasa, I.N., Simbolon, D., 2020. Remote sensing for assessing the potential anchovy fishing ground in the Pesisir Selatan regency, West Sumatra, Indonesia. Aacl Bioflux.

Cherukuru, N., Martin, P., Sanwlani, N., Mujahid, A., Müller, M., 2021. A Semi-Analytical Optical Remote Sensing Model to Estimate Suspended Sediment and Dissolved Organic Carbon in Tropical Coastal Waters Influenced by Peatland-Draining River Discharges off Sarawak, Borneo. Remote Sensing 13, 99.

Cornish Fishing History, Cornwall Good Seafood Guide [WWW Document], n.d. URL https://www.cornwallgoodseafoodguide.org.uk/cornish-fishing/history-of-the-cornish-fishing-industry.php (accessed 12.7.24).

Deng, T., Chau, K.-W., Duan, H.-F., 2021. Machine learning based marine water quality prediction for coastal hydro-environment management. J Environ Manage 284, 112051.

Deshpande, A., Priyadarshini, E., More, A., Pate, S., 2024. A new framework for accessing and visualizing Ocean Color Data for Water quality parameter analysis. Journal of Integrated Science and Technology 12, 707–707.

Gómez, D., Salvador, P., Sanz, J., Casanova, J.L., 2021. A new approach to monitor water quality in the Menor sea (Spain) using satellite data and machine learning methods. Environmental Pollution 286, 117489.

Majumder, S., Maity, S., Balakrishnan Nair, T.M., Bright, R.P., Nagaraja Kumar, M., Shwetha, N., Kumar, N., 2021. Potential Fishing Zone Characterization in the Indian Ocean by Machine Learning Approach. Advances in Intelligent Systems and Computing 1393, 43–54.

Mohseni, F., Saba, F., Mirmazloumi, S.M., Amani, M., Mokhtarzade, M., Jamali, S., Mahdavi, S., 2022. Ocean water quality monitoring using remote sensing techniques: A review. Marine Environmental Research 180, 105701.

Pardeshi, S., Gandre, P., Poojari, N., Pansare, S., Alte, B., 2023. Water Quality Analysis from Satellite Images. 2023 International Conference on Data Science and Network Security (ICDSNS) 1–6.

Purwanto, A.D., Prayogo, T., Marpaung, S., Suhada, A.G., 2020. ANALYSIS OF POTENTIAL FISHING ZONES IN COASTAL WATERS: A CASE STUDY OF NIAS ISLAND WATERS. IJReSES 17, 9.

Rostam, N.A.P., Malim, N.H.A.H., Abdullah, R., Ahmad, A.L., Ooi, B.S., Chan, D.J.C., 2021. A Complete Proposed Framework for Coastal Water Quality Monitoring System With Algae Predictive Model. IEEE Access 9, 108249–108265.

Sarangi, R.K., Jishad, M., Sharma, R., Das, A., Mali, K., Ramalingam, L., Kizhakudan, S.J., Kumar, A.S., Samal, R.N., Prakash, C., 2023. Multiple ocean parameter-based potential fishing zone (PFZ) location generation and validation in the Western Bay of Bengal. Environ Monit Assess 196, 98.

Sasmito, B., Bashit, N., Arinda, B.R., Sukmono, A., 2022. Application of Generalized Additive Model for Identification of Potential Fishing Zones Using Aqua and Terra MODIS Imagery Data. JAGI 6, 583–591.

Sent, G., Biguino, B., Favareto, L., Cruz, J., Sá, C., Dogliotti, A.I., Palma, C., Brotas, V., Brito, A.C., 2021. Deriving Water Quality Parameters Using Sentinel-2 Imagery: A Case Study in the Sado Estuary, Portugal. Remote Sensing 13, 1043.

Sivasankari, M., Anandan, R., Chamato, F.A., 2022. HE-DFNETS: A Novel Hybrid Deep Learning Architecture for the Prediction of Potential Fishing Zone Areas in Indian Ocean Using Remote Sensing Images. Computational Intelligence and Neuroscience 2022, 1–10.

Tian, S., Guo, H., Xu, W., Zhu, X., Wang, B., Zeng, Q., Mai, Y., Huang, J.J., 2023. Remote sensing retrieval of inland water quality parameters using Sentinel-2 and multiple machine learning algorithms. Environ Sci Pollut Res 30, 18617–18630.

Topp, S.N., Pavelsky, T.M., Jensen, D., Simard, M., Ross, M.R.V., 2020. Research Trends in the Use of Remote Sensing for Inland Water Quality Science: Moving Towards Multidisciplinary Applications. Water 12, 169.

Younes Oulad Sayad, 2023. Mapping Potential Fishing Zones Using Remote Sensing Data and GIS: A Case Study of Moroccan Waters. MR 6, 1–13.

Zainab, S., Handajani, N., Wibisana, H., Wardhani, P.C., 2022. Distribution of Water Quality Concerning Sea Surface Temperature and Dissolved Oxygen Using Landsat 8 Satellite Imagery on the Tuban Coast, East Java 7.

Zhang, X., Zhang, Y., Zhang, Q., Liu, P., Guo, R., Jin, S., Liu, J., Chen, L., Ma, Z., Liu, Y., 2020. Evaluation and Analysis of Water Quality of Marine Aquaculture Area. IJERPH 17, 1446.

# Appendix A

## Student Project Approval Form

|  |  |
| --- | --- |
| **Supervisor sign off** | |
| Ethics form complete |  |
| Ethical concerns acknowledged |  |
| Research tool(s) checked |  |
| All relevant forms included (consent etc.) |  |
| Is not high risk |  |

**LD7083/ Computing and Digital Technologies Project: Student Project Approval Form**

You should use this document if you intend to use one of the existing module level approval ethics applications. Please complete this document and discuss your study with your supervisor before you collect any data. *Failure to complete this document and have all aspects signed off and approved by your supervisor risks a notable deduction in your grade and may risk a case of Academic misconduct. Please see the module Bb site for more details.*

Please ensure that your project meets the conditions of the existing ethics application (available on Module Bb site). ***If it does not, then you will need to submit a full ethics application instead***.

|  |  |
| --- | --- |
| Student Name: | JOHN RAJ THANKARAJ |
| Project Title: | Marine Water Quality Analysis and Predicting Optimal Fishing Grounds using Machine Learning and Remote Sensing Technology. |
| Supervisor Name: | Rejwan Bin Sulaiman |
| Ethics application you are amending (check box): | Low-risk Lab-based research  Low Risk Secondary Data Science project  Medium Risk Secondary Data Science project from the private domain required membership  Questionnaire/ survey Study  Interview Study or other Usability Study |

**Introduction to the project:** *Treat like an introduction to the study. Why is your proposed study important? What has already been done on the topic? How does your proposed study ‘fit’ with the current literature and what does it add? What is the aim of the proposed study? Make reference to appropriate studies.*

|  |
| --- |
| Introduction  The world's economy and food chain depend on marine ecosystems, which are increasingly under stress from human activity and climate change. Sustainable resource management requires an understanding of these ecosystems and regular monitoring of their health. In addition to being costly and time-consuming to perform, traditional techniques of assessing the quality of marine water are sometimes limited in their coverage of time and location.  Improvements in machine learning and remote sensing technology in recent years have created new opportunities for precise and effective maritime monitoring. Techniques for remote sensing, such satellite photography and airborne surveys, can yield important data on water quality indicators like sea surface temperature, turbidity, and chlorophyll-a concentration. On the other hand, machine learning algorithms are able to find patterns, trends, and errors through processing big datasets from remote sensing and other resources.  The purpose of this study is to identify the best fishing locations by utilizing machine learning and remote sensing to evaluate the quality of marine water. We want to create a strong and effective system that can assist with marine habitat preservation and sustainable fisheries management by combining these technologies. |

**Methodology:** *Please complete the table below, using the following info to guide you.**Write this as a future tense method. Describe the* ***participants*** *that you will recruit, how many you are going to recruit, and indicate if you have any additional exclusion criteria. Include the* ***research design*** *(e.g. randomised/repeated measures/quantitative/qualitative/case study etc) and detail of your proposed* ***procedures*** *(i.e., how are you collecting the data?). Include information on all of the equipment you plan to use. If this is a low-risk study, outline how you will extract data and list the criteria you will use to do this. Somebody should be able to read this and replicate it. Describe all planned* ***data analysis*** *for both quantitative (e.g. t-tests, ANOVA, correlation etc.) and qualitative (content analysis, thematic analysis etc.) data. If doing a low-risk study explain how you intend to analyse the data you have collected. Use literature to justify your method.*

|  |  |
| --- | --- |
| 1. Is this a low-risk secondary data or lab-based study? If Yes please go to questions 6 and 7. | YES  NO |
| 1. Who are your participants and what is the inclusion criteria? | N/A |
| 1. How many will you recruit and from where? | N/A |
| 1. Are there any exclusion criteria (reasons why people should not participate)? | N/A |
| 1. Research design: | N/A |
| 1. Procedures (describe what you will do to collect data, include all equipment/methods you plan to use). | MODIS data may be downloaded for free. NASA makes a wide range of MODIS products available from a number of portals and data centres. The land, ocean, atmosphere, and cryosphere are only a few of the Earth scientific fields that are covered by these data packages.  OceanColor Web: This website specializes on ocean color information, such as sea surface temperature, chlorophyll-a concentration, and other oceanographic factors.  NASA uses a range of satellite-based sensors to gather data on Oceancolor. Phytoplankton is the main material that affects the colour of the ocean's surface, which is measured by these sensors. Scientists may learn about the production, health, and other essential elements of the ocean through studying these colour changes. |
| 1. Data analysis methods: | The analysis of remote sensing data for forecasting fishing grounds and marine water quality is covered in the content. It starts with data collection, referencing satellites such as Sentinel-2, Landsat, and MODIS. Preprocessing of data includes geometry correction, atmospheric adjustment, radiometric calibration, and data quality evaluation. Spectral index calculations, texture analysis, and Principal Component Analysis are all part of feature extraction. Both supervised learning (classification and regression) and unsupervised learning (clustering) are included in machine learning model creation. Splitting data, training the model, and verifying it are all part of the model training and validation process. Developing a predictive model, creating prediction maps, and finding important environmental elements are all part of predicting the best fishing locations. |
| 1. Additional information: | N/A |

**Health and Safety:**  *Relevant risk assessments are listed in the ethics application. If your project needs additional risk assessments, then you will need to submit a new ethics application. Please identify the elements of the listed risk assessment that are relevant for your study and the risk assessment(s) you are working with.*

Please check the relevant boxes\*:

HL\_RISK\_173 Testing in an external environment

HL\_RISK\_722 face to face interview

HL\_RISK\_727 Group interview

|  |  |  |
| --- | --- | --- |
| **Areas of potential risk**  *Please indicate how you will eliminate, or as a minimum ameliorate, the following areas of potential risks throughout the processes of research design, data generation, data analysis and dissemination* | | |
| **Area of risk** | **Questions relating to this risk** | **How will you mitigate against this risk?** |
| Avoiding harm to all involved in or potentially affected by the research | How will you ensure that your participants/ respondents come to no harm (psychological; emotional; physical). e.g. not subjecting them to questioning about sensitive issues without advance agreement? | N/A |
| How will you ensure your own safety (beyond just physical) in undertaking the Enquiry? | N/A |
| Ensuring the anonymity of all participants/respondents | How will you ensure anonymity in collecting/generating data | N/A |
| How will you ensure anonymity in reporting the data? | N/A |
| Gaining informed consent from all participants / respondents | How will you ensure respondent/participant consent in advance? You should provide a copy of the necessary consent form/s with this document | N/A |
| (How) might participants/respondents be able to withdraw their data? | N/A |
| Avoiding deception | How will you how you promote accuracy in recording, analysis, reporting of the data/findings? | N/A |
| Data storage and destruction | How will you transport and store your data securely (e.g. password protected; cloud storage) | N/A |
| How will you destroy the data and when? | N/A |
| Secondary data sets | *Is your data set(s) from a domain requires membership?* | No |
| *Does this data set can be used for educational or academic research purpose?* | Yes |

Please check this box after you have read and understood ethics and health and safety information.

I confirm I have read the University’s health and safety policy and ethics policy. I have read and understood the requirement for the mandatory completion of risk assessments and that my study does not deviate from the module level approval ethics forms on Blackboard.

**Further information (add below, if applicable)**

* Consent forms
* Participant information sheet
* Debrief form
* Recruitment materials
* Permission letters
* Data collection tools

|  |  |
| --- | --- |
| **Student’s Name and sign**    **JOHN RAJ THANKARAJ** | **Date 07/11/2024** |
| **Supervisor’s name and sign**  **A blue text on a black background  Description automatically generated**  **(Name)** | **Date: 12/11/24** |

# Appendix - B

Source code link:

<https://drive.google.com/file/d/1Sn_YfnJIkJFK6POyPdz8KEo6RTO_S83l/view?usp=drivesdk>

# Appendix - C

## Meeting Logs

|  |  |  |
| --- | --- | --- |
| **Meetings Date** | **Discussion about topics** | **Updates about tasks** |
| 22/11/2024 | Literature review and methodology development | Gathering and evaluating research publications. created a table outlining the main conclusions, methods, and restrictions. improved search phrases for literature. |
| 29/11/2024 | Data processing and methodology refinement | Improved the data processing codes. adjusted for better outcomes after testing other approaches. using reliable techniques. |
| 06/12/2024 | Discussion about study area investigation and actual data preprocessing | looked into the history of the subject area to have a better grasp of the implications for future research.  Python was used to collect and handle the data, improving its quality for analytical preparation. |
| 13/12/2024 | Outcome validation and writing process | assessed the accuracy and reliability of the program's results. continued writing the sections of the study paper. |
| 06/01/2024 | Discussion about final referencing and citations | Cited and referenced sources accurately. Double-checked reference list completeness. |