

**BATHYMETRIC ANALYSIS USING MULTI-SPECTRAL SATELLITE
DATA THROUGH MACHINE LEARNING ALGORITHMS IN
VARYING COASTAL REGIONS OF INDIA**

A thesis submitted to
University of Petroleum and Energy Studies

For the Award of
Doctor of Philosophy
In

Earth Sciences

By
Mohammad Ashphaq

January, 2023

SUPERVISOR (s)

Dr. Pankaj Kumar Srivastava
Dr. Debashis Mitra



DEPARTMENT OF PETROLEUM ENGINEERING & EARTH SCIENCES
ENERGY CLUSTER
UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
DEHRADUN-248007, UTTARAKHAND, INDIA

**BATHYMETRIC ANALYSIS USING MULTI-SPECTRAL SATELLITE
DATA THROUGH MACHINE LEARNING ALGORITHMS IN
VARYING COASTAL REGIONS OF INDIA**

A thesis submitted to
University of Petroleum and Energy Studies

For the Award of
Doctor of Philosophy
in
Earth Sciences

By
Mohammad Ashphaq
(SAP Id.: 500071972)

January, 2023

SUPERVISOR (s)

Dr. Pankaj Kumar Srivastava (Internal Guide)
Professor, Energy Cluster, UPES, Dehradun

Dr. Debashis Mitra, (External Guide)
Head, Marine and Atmospheric Sciences Department
Indian Institute of Remote Sensing ISRO, Dehradun



DEPARTMENT OF PETROLEUM ENGINEERING & EARTH SCIENCES
ENERGY CLUSTER
UNIVERSITY OF PETROLEUM AND ENERGY STUDIES
DEHRADUN-248007, UTTARAKHAND, INDIA

January, 2023

DECLARATION

I hereby declare that the thesis entitled “**BATHYMETRIC ANALYSIS USING MULTI-SPECTRAL SATELLITE DATA THROUGH MACHINE LEARNING ALGORITHMS IN VARYING COASTAL REGIONS OF INDIA**” has been prepared by me at Department of Petroleum Energy & Earth Sciences, School of Engineering, University of Petroleum & Energy Studies under the guidance of Dr. Pankaj Kumar Srivastava, Professor of Energy Cluster, UPES, and Dr. D Mitra, Head, Marine and Atmospheric sciences Department, IIRS. No part of this thesis has formed the basis for the award of any degree or fellowship previously.



Mohammad Ashphaq

Energy Cluster

University of Petroleum & Energy Studies

Energy Acres Building, Bidholi,

Dehradun-248007 Uttarakhand, India

Date: 21.07.2023

THESIS COMPLETION CERTIFICATE

This is to certify that the thesis entitled “**BATHYMETRIC ANALYSIS USING MULTI-SPECTRAL SATELLITE DATA THROUGH MACHINE LEARNING ALGORITHMS IN VARYING COASTAL REGIONS OF INDIA**” submitted by MOHAMMAD ASHPHAQ (SAP ID: 500071972) in fulfilment of the requirements for the award of the degree of DOCTOR OF PHILOSOPHY in Earth Sciences is an original work carried out by him under my supervision and guidance.

It is certified that the work has not been submitted anywhere else for the award of any diploma or degree of this or any other University.


Dr. Pankaj Kumar Srivastava

Professor

Energy Cluster, SoE

UPES

Dehradun-248007

Date: 24.7.2023

भारत सरकार
अंतरिक्ष विभाग
भारतीय सुदूर संवेदन संस्थान
4, कालिदास मार्ग, पो. बाक्स सं. 135
देहरादून-248001, भारत
दूरभाष : +91-135-2524399
फैक्स : +91-135-2741987, 2748041

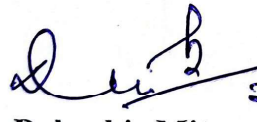


Government of India
Department of Space
Indian Institute of Remote Sensing
4, Kallidas Road, P.B. No. 135,
Dehradun - 248001, India
Telephone : +91-135-2524399
Fax : +91-135-2741987, 2748041

THESIS COMPLETION CERTIFICATE

This is to certify that Mr. MOHAMMAD ASHPHAQ (SAP ID: 500071972) has prepared his thesis entitled “**BATHYMETRIC ANALYSIS USING MULTI-SPECTRAL SATELLITE DATA THROUGH MACHINE LEARNING ALGORITHMS IN VARYING COASTAL REGIONS OF INDIA**”, for the award of PhD degree of the University of Petroleum & Energy Studies, under my guidance. He/she has carried out the research work at the Marine & Atmospheric Sciences Department, Indian Institute of Remote Sensing-Indian Space Research Organization, Dehradun under my supervision.

External Supervisor

 21/7/23

Dr. Debashis Mitra,
Head, Marine and Atmospheric Sciences Department
Indian Institute of Remote Sensing-ISRO, Dehradun
Dehradun-248007 Uttarakhand, India

Date:

ABSTRACT

Satellites have provided us with a view of the ocean's surface over the last five decades, allowing us to accurately measure temperature, wind, and other meteorological phenomena. Although satellite-derived bathymetry (SDB) has been studied for four decades, its application in marine navigation remains limited due to accuracy constraints. Only a few of the SDB algorithms developed to date have been validated in various geographic regions. To add to the complexity, very few algorithms have been reported to be usable in turbid water, and this is one of the challenges this research tries to address.

The study's objective is to utilise satellite data and machine learning techniques to verify the viability of SDB. The high-density in-situ data offers the potential to create a new, improved model to estimate depths in complex and challenging regions with significant turbidity. The primary goal of this study is to develop a better SDB algorithm that will be useful for operational maritime navigation skills. The goal of the research was to create an SDB algorithm that could be used in very complex turbid environments with some degree of accuracy. This could aid in charting relatively large and remote shallow water areas without sacrificing life or resources. INCOIS-SATCORE data and high-density bathymetric data from NHO were used to achieve this, providing a rare opportunity to find alternative solutions to existing SDB models that are based on fewer, sparsely distributed parameters, with the potential to establish robust models over variable shallow waters.

The required satellite imagery Landsat satellite data (Landsat-5-8) temporally proximal to INCOIS SATCORE data collected during 2008 to 2017 for validation of OAS was downloaded from USGS earth explorer with filter of date and less than 10 percent cloud coverage. The imagery with glint were avoided for further processing. The next step in the process of developing Algorithm was delineation of water region using NIR band of satellite data. The pre-processing has been carried using the ACOLITE open-source module. The comparative analysis of most cited SDB model with ML algorithm using single spectral band as well as multispectral satellite data

was carried in exploratory study. The exploratory study proved the efficacy of RF machine learning algorithm as in comparison to any other SDB model in complex and more turbid water of Vengurla and Mormugao with MS bands. The OAS, Satellite bands in visual spectrum with bathymetry were used in further processing by predictor analysis of variables. The study on preliminary analysis of OAS by the predictor analysis indicated that Bandwidths 0.561 μm & 0.479 μm of Landsat-7 & 8, Chlorophyll, TSM, and Turbidity are vital variables in bathymetry estimation. The three algorithms for Chlorophyll- Molkov et al., 2019; TSM- Molkov et al., 2019; and Turbidity-Nechad et al., 2009 have been used to derived OAS from the same satellite imagery used for SDB. The processing was carried out separately for raster data in ENMAPBox plugin of QGIS and using Jupyter Notebook python IDL for .csv data. The resultant RF algorithm were saved as .pkl file and then the merging of algorithm was carried using python script. All raster dataset was also extracted in .csv format for creating an algorithm based on merged dataset of 5 test sites. The final MRF algorithms has been cross-validated in different geographical areas with reasonable accuracy. The SDB derivation in a vast region of 100 Sq Km to 1400 Sq. km is rare in SDB which has been achieved. Besides, this study has applied SDB derivation up to depth of 90 m with reasonable accuracy in Gopalpur site. A generic Merged RF algorithm for recursive bathymetry mapping in coastal region of India has been tested and validated. A generic Merged RF algorithm which includes three additional parameters; TSM, Chl, & Turbidity derived from the same satellite imagery for recursive SDB is proposed, tested & validated in different coastal region of India.

This thesis is dedicated
to my late father

Mr. Mohammad Sharif

Who always dreamt for me to excel in all
sphere of life

ACKNOWLEDGEMENTS

At the outset, I would like to express my heartfelt gratitude to my Research Guide **Dr. Pankaj Kumar Srivastava**, Professor, Department of Petroleum Engineering and Earth Sciences, UPES, Dehradun and my Co-Guide **Dr. D Mitra**, Scientist-SG, IIRS-ISRO, for their invaluable guidance and encouragement throughout the tenure of this research work. Especially, I would like to thank Dr. D Mitra who introduced me to Marine applications of remote sensing during my NNRMS course at IIRS, Dehradun and made me take up this exciting and demanding topic for my research. I owe my deepest gratitude to both the guides for their invaluable contributions and guidance throughout my research.

I am extremely grateful to Director, Indian National Centre for Ocean Information Services (INCOIS), Hyderabad, Dr. Aneesh Lotliker of INCOIS, Hyderabad, and especially entire team of Satellite Coastal and Oceanographic Research (SATCORE) programme whose data from year 2008-2017 has been used in this study. I am also thankful to National Hydrographic Office, Dehradun for allowing use of data for this purpose of research. I am also indebted to many faculties of IIRS who have helped me in developing algorithms for this study.

I will always be grateful to Dr. Pooja Purang and Dr. Meenakshi Gupta, my M Phil research guides at HSS department, IIT, Mumbai, who taught me the rigorous research process and academic writing.

My eternal gratitude goes to my late father, who always believed in my abilities and did everything he could to get me to this point in life. I am grateful to my mother, who dedicated her entire life to my care and the pursuit of my dreams. May her eyes light up with triumph on the day I receive my doctorate. I am grateful to my beloved wife, my children, my brother, and my entire family for their unwavering support, patience, and faith in me. Finally, I am thankful to my all friend, colleagues, Seniors, and everyone who was directly or indirectly part of this academic journey.

TABLE OF CONTENTS

ABSTRACT	i
ACKNOWLEDGEMENTS	iv
TABLE OF CONTENTS	v
LIST OF FIGURES	viii
LIST OF TABLES	x
ABBREVIATION	xi
CHAPTER: 1 INTRODUCTION	1
1.1 GENERAL OVERVIEW	1
1.2 MOTIVATION	3
1.3 OBJECTIVES	3
1.4 OVERVIEW OF RESEARCH.....	4
1.5 CONTRIBUTION OF RESEARCH.....	7
1.6 OUTLINE OF THESIS CHAPTERS.....	8
CHAPTER: 2 LITERATURE REVIEW	9
2.1 SPACE TECHNOLOGY & BLUE ECONOMY.....	9
2.2 SATELLITE DERIVED BATHYMETRY.....	11
2.2.1 SDB Methods.....	12
2.2.2 Optical SDB.....	13
2.2.3 SDB Algorithms.....	14
2.2.4 SDB Classification.....	14
2.3 OPTICAL SDB APPROACHES	16
2.3.1 Statistical Approach.....	16
2.3.2 RT Physics Based Approach.....	16
2.3.3 Models of RT Physics Based Approach.....	18
2.4 SDB METHODS.....	19
2.4.1 Empirical SDB.....	20
2.4.2 Semi-Empirical SDB.....	24
2.4.3 Analytical SDB.....	26
2.4.4 Semi Analytical SDB.....	29
2.4.5 Quasi Analytical (QA) SDB.....	30

2.5 Summary.....	31
CHAPTER 3: METHODOLOGY.....	34
3.1 MATERIALS AND METHODS.....	34
3.1.1 Study Sites.....	34
3.1.2 Data	41
3.1.3 Pre-processing.....	44
3.1.4 Performance Evaluation.....	45
3.2 RESEARCH STAGES.....	46
3.2.1 SDB Algorithms in Turbid Coastal Water.....	47
3.2.2 Univariate ML Algorithms for SDB.....	48
3.2.3 Multivariate ML Algorithms for SDB.....	49
3.2.4 Optically Active Substance & Bathymetry.....	51
3.2.5 Influence of Chlorophyll, TSM, & Turbidity on SDB.....	54
3.2.6 Merged Random Forest (MRF) Model of SDB.....	57
CHAPTER 4 : RESULT	62
4.1 SDB ALGORITHMS & TRANSFORMS	62
4.1.1 Log-Ratio Transformation.....	62
4.1.2 Log-Linear Transformation.....	63
4.1.3 Semi-Automated Tools.....	64
4.1.4 Machine Learning.....	65
4.1.5 Comparison of Satellite data and Technique of SDB.....	67
4.2 ML UNIVARIATE	68
4.2.1 Result of Univariate ML Algorithms.....	69
4.2.2 Linear Vs Robust-Linear Algorithms.....	70
4.2.3 Linear Vs Robust-Linear Algorithms.....	72
4.2.4 Sensitivity Analysis: Sample Size	74
4.2.5 Linear, Robust and Non-linear Algorithms.....	74
4.2.6 Summary of Univariate SDB Analysis.....	75
4.3 MULTIVARIATE ANALYSIS.....	76
4.4 BATHYMETRY- SATCORE DATA ANALYSIS.....	83
4.4.1 Descriptive Statistics.....	83
4.4.2 Correlation Matrix.....	83
4.4.3 Regression Modelling of all Parameters.....	84
4.4.4 Predictor Analysis.....	85

4.4.5 Summary.....	86
4.5 SDB AND OAS ANALYSIS.....	87
4.5.1 Descriptive Statistics.....	87
4.5.2 Correlation Matrix.....	88
4.5.3 Result of SDB Algorithms and Cross-validation.....	88
4.5.4 Validation of Satellite-Derived OAS.....	90
4.5.5 SDB Residuals and OAS.....	92
4.6 DEVELOPMENT OF MRF.....	97
CHAPTER 5 : DISCUSSION.....	99
5.1 VARIOUS TRANSFORMS AND SDB ALGORITHMS	99
5.2 UNIVARIATE ML REGRESSION ANALYSIS	100
5.3 MULTIVARIATE ANALYSIS	103
5.4 BATHYMETRY-SATCORE DATA ANALYSIS	103
5.5 SDB AND OAS ANALYSIS	104
5.6 DEVELOPMENT OF MRF	106
5.7 COMPARISON WITH GEBCO DATABASE DERIVED MAPS.....	107
CHAPTER 6: CONCLUSION	112
6.1 Future Research and Challenges.....	114
REFERENCES.....	116
APPENDICES.....	122
APPENDIX-I: TABLES.....	122
APPENDIX-II: FIGURES.....	128
APPENDIX-III: ML SYNTAX.....	133
APPENDIX-IV: LIST OF PUBLICATION.....	146
APPENDIX-V: THESIS PLAGIARISM CERTIFICATE	147
APPENDIX-VI: FIRST PAGE OF PLAGIARISM REPORT	148

List of Figures

Figure 1.1: Overview of research.....	05
Figure 2.1: SDB Classification Framework.....	16
Figure 2.2: Empirical SDB	21
Figure 2.3: SDB ML algorithms.....	23
Figure 2.4: SDB Studies: SA	30
Figure 3.1 Study Area Vengurla & Mormugao.....	35
Figure 3.2 INCOIS SATCORE Data collection sites.....	37
Figure 3.3 Study Area -Kochi.....	38
Figure 3.4 Study Area Chennai.....	39
Figure 3.5 Study Area-Okha.....	40
Figure 3.6 Study Area Gopalpur.....	41
Figure 3.7: The Flowchart of pre-processing of data for exploratory study	45
Figure 3.8: The Flowchart of various stages of research process	46
Figure 3.9: Flowchart of exploratory study to compare SDB	47
Figure 3.10: The Flowchart of Univariate SDB ML algorithms	48
Figure 3.11: Flowchart of the Multivariate SDB ML algorithms	50
Figure 3.12: The Flowchart of Pre-processing with INCOIS-SATCORE data.....	53
Figure 3.13: Flowchart of the SATCORE data analysis.....	54
Figure 3.14: Flowchart for SDB error analysis	56
Figure 3.15 Flowchart of MRF.....	60
Figure 4.1: Depths vs SDB by Log-Ratio transformation.....	63
Figure 4.2: Depths vs SDB using Envi 5.3.....	64
Figure 4.3 (a): y_{test} vs $y_{predict}$ for ASTER.....	66
Figure 4.3 (b): y_{test} vs $y_{predict}$ for LANDSAT-8.....	66
Figure 4.3 (c): y_{test} vs $y_{predict}$ for SENTINEL-2.....	67
Figure 4.4: Linear & Robust Linear Regression – Mormugao.....	71
Figure 4.5: Linear & Robust Linear Regression - Vengurla	72
Figure 4.6: Varying ratio of training and validation	73
Figure 4.7: The scatterplot of Actual bathymetry vs SDB of Vengurla.....	77
Figure 4.8. Density Distribution Vengurla	78
Figure 4.9: The scatterplot of Actual bathymetry vs SDB of Mormugao	79

Figure 4.10. Density Distribution Mormugao	80
Figure 4.11: The comparison of R2, RMSE, & MAE for Vengurla	81
Figure 4.12: The comparison of R2, RMSE, & MAE for Mormugao.....	82
Figure 4.13: Result of SDB derivation using Linear, RF, And SVR algorithms.....	90
Figure 4.14: Graphs showing validation of Satellite-Derived OAS.....	91
Figure 4.15: The IDW interpolated verses Satellite-Derived OAS.....	92
Figure 4.16: Distribution of SDB residuals and concentration of OAS.....	94
Figure 4.17: OAS Classification and SDB residuals.....	95
Figure 4.18: Histogram of three classes of OAS with SDB residuals.....	96
Figure 5.1: SDB compared with GEBCO derived Map-Chennai.....	108
Figure 5.2: SDB compared with GEBCO derived Map-Okha.....	109
Figure 5.3: SDB compared with GEBCO derived Map-Kochi.....	110

List of Tables

Table 2.1: Summary of Bathymetry (Modified after Jawak, Vadlamani, & Luis, 2015).....	13
Table 2.2: SDB selection matrix.....	32
Table 3.1: Satellite Data used in the study.....	42
Table 3.2: The Multivariate ML Algorithms used in this study.....	51
Table 3.3: Details of SATCORE and Satellite Data	55
Table 4.1: Bathymetry Correlation Log-Ratio transform.....	63
Table 4.2: Bathymetry Correlation Log-Linear transform.....	64
Table 4.3: SDB Estimation using ML Algorithms.....	65
Table 4.4: Univariate-Regression ML Algorithms.....	68
Table 4.5: ML Regression Algorithms in Site A, Vengurla.....	69
Table 4.6: ML Regression Algorithms in Site B, Mormugao.....	70
Table 4.7: ML Regression Algorithms.....	76
Table 4.8: Descriptive Statistics of data used in study.....	88
Table 4.9: Correlation Matrix of OAS.....	88
Table 4.10: Result of SDB Algorithms.....	89
Table 4.11: Analysis of categorical variable OAS and SDB Residuals.....	93
Table 4.12 Implementation of Merged RF algorithm.....	98
Table 5.1 Comparison between GEBCO & MRF Map.....	107

ABBREVIATIONS

ANN, Artificial Neural Network
AOP, Apparent Optical Property
AUV, Autonomous Underwater Vehicle
CDOM, Coloured Dissolved Organic Matter
Chl, Chlorophyll
GIS, Geographic Information System
HS, Hyper- Spectral
IHO, International Hydrographic Office
INCOIS, Indian National Centre for Ocean Information System
IOCCG, International Ocean Colour Coordinating Group
IOP, Inherent Optical Properties
LUT, Look-Up-Table
MBES, Multi Beam Echo-Sounder
ML, Machine Learning
MLR, Mul-tiple Linear Regression
MS, Multi-Spectral
NIR, Near-Infrared
PCA, Principal Component Analysis
QA, Quasi-Analytical
RF, Random Forest
RMSE, Root Mean Square Error
RS, Remote Sensing
Rrs, Remote Sensing Reflectance
RT, Radiative Transfer
SA, Semi-Analytical
SATCORE, SATellite Coastal and Oceanographic Research
SBES, Single Beam Echo-Sounder
SE, Semi-Empirical
SVM, Support Vector Machine

CHAPTER 1: INTRODUCTION

This chapter in its initial sections introduces the general overview of the coastal region of states, its significance in the Blue Economy, and the importance of Geographic Information Systems & Remote sensing in the development of the coastal region. Thereafter the next section discusses the motivation for selecting this topic for research in this study. The further section elaborates on the objective of the research, an overview of methodological procedures, and finally the contribution of this research in the domain of knowledge has been discussed.

1.1 GENERAL OVERVIEW

Coastal regions are home to more than 40% of the world's population. The coastal zone is important for many reasons, such as navigation, coastal life, coastal morphology, etc. Sustainable use of marine resources, food provision for over-stretch coastal areas, and continuous monitoring of the oceans are among the priority areas set by the United Nations Oceans Decade 2021-2030 Program set forth by the United Nations (Ryabinin et al., 2019). The growth of the green economy of any country depends mainly on the coastal regions. However, it is a very surprising fact that modern technologies have evolved to map the surface of Mars and the moon better than mapping the hydrosphere of the Earth. More than 80% of the Earth's seafloor remains unexplored. Although, over the past 5 decades, satellites have provided us with a glimpse of the ocean's surface, helping us to measure temperature, wind, and various meteorological phenomena very accurately. Several Ocean Color Monitoring (OCM) satellites provide continuous observations of the ocean by estimating water quality indicators like chlorophyll, SST, ocean color, waves, salinity, wind, etc., and offer several derivative operating products. But even today, the big challenge is looking deep into the sea and mapping its representation in three dimensions (Purkis & Chirayath, 2022).

Knowledge of water mass mapping lies within the field of hydrography. Charts are the product of hydrographic surveys to satisfy maritime needs and are

therefore essential to more than 95% of the commerce of any country. A chart (electronic navigational chart or paper chart) is a cartographic representation of an area's depth, and navigational hazards, and an important support for navigational information gathered during navigation. hydrographic survey. Hydrographic surveys use two different methods, the SBES and the MBES to determine the depth and topography of the seabed. Recently, several emerging technologies are used to measure the depth of the water masses. This technology includes LIDAR, the use of ROV, and AUV to determine effective depth in coastal waters. All of these techniques are very expensive due to high purchase and maintenance costs, and limited repetitive and frequent readings. In addition, there is a risk of life and loss of resources in complex traffic areas such as tidal estuaries, coves, and complex reefs.

Remote sensing as well as GIS techniques are widely used for studies of coastal regions because of their advantages of reproducibility, multispectral, and generalizability. Satellite-derived depth measurement has been studied for four decades, but its application in marine navigation is still lacking due to accuracy limitations. Although SDB was developed in the late 1970s, it hardly had any application in operational-depth research until the last decade. contemporary advances in satellite or space technology in terms of resolution, multi-spectral band, open-source availability, etc. enhance its potential for use as a source of hydrological data. The use of satellite data is increasing throughout hydrology as a low-cost data source for coastline delineation. It is currently being accepted not only as an operational probe but also as an advanced technique capable of providing navigators with calibrated and confirmed depths with the use of very few resources.

Several SDB algorithms have been developed to date, but only a few have been validated in different geographic regions. Adding to the complexity, very few algorithms have been reported to be usable in turbid water and this is one of the future challenges that need to be addressed with satellite imagery (Purkis & Chirayath, 2022). Among the two mainstream approaches in SDB, one is an analytic approach that depends upon two or more water column variables, and the other is an empirical approach using sparsely distributed depth samples to establish an empirical relationship between the reflectance and the corresponding depth point. Several studies have discussed the fusion of analytical and empirical methods as a semi-analytic approach, which also has some inherent limitations. This study aims to fill a research gap by combining analytical and experimental SDB methods based on high-

resolution data related to water column parameters and high-density In-situ bathymetry applied to high-resolution Satellite data via machine learning.

1.2 MOTIVATION

This work envisages validating the practicality of SDB based on Satellite Data and Machine Learning algorithms. The high-density in-situ data has the potential to establish a new improved model to estimate depths in complex highly turbid areas. The goal of this research is to develop an improved algorithm of SDB so that it has utility in operational maritime capabilities in navigation.

- i) The motivation of the research is to develop an algorithm of SDB that can be employed with reasonable accuracy in highly complex turbid areas. This may help in charting relatively large and remote shallow water areas without compromising with loss of life and resources.
- ii) The availability of the Indian National Centre for Ocean Information Services (INCOIS) project namely ‘Satellite Coastal and Oceanographic Research’ (SATCORE) data and high-density bathymetric data provides a rare opportunity to find alternative solutions to existing SDB models which are based on fewer, sparsely distributed parameters, thereby having the potential to establish robust model over variable shallow waters.
- iii) Providing low-cost resources for nautical charting as an alternative to high-cost hydrographic surveying to establish high-density data in coastal regions of India up to the maximum achievable depth in shallow water regions.
- iv) Overcoming the challenges of shallow waters SDB where creeks and estuaries, huge hidden reefs, tidal bores, and surge occurs.
- v) Envisage to improve relatively accurate and reliable bathymetry by multispectral images that may be applied to Sediment transport, dredging, coastal engineering (port, constructions, etc.), Coastal fisheries and aquaculture, Coastal Tourism, and exploratory studies. Also, Addressing the Data gaps in existing bathymetry data, Reconnaissance of previous surveys, and Change detection and mapping.

1.3 OBJECTIVES

The present study has the following objectives to be considered to derive operational SDB.

- 1) Analysis of existing SDB algorithms in varying coastal regions of India.
- 2) Determining the best-suited machine learning algorithm in complex coastal turbid water.
- 3) Deriving the relationship between SATCORE parameters & reflectance of satellite data to develop the empirical model by empirical analysis.
- 4) Developing SDB model based on the remote sensing reflectance and SATCORE data through machine learning algorithm and its validation

The results will be compiled to elaborate on the procedures to operationalize the SDB model for nautical charting.

1.4 OVERVIEW OF RESEARCH

The literature has been reviewed for the past 5 decades in the SDB field. Variables such as depth, sensor resolutions, and also the dynamics of inherent coastal optical properties (IOPs) are some barriers to performing comparative analysis as per previous studies. A systematic framework for classifying SDB studies over the past 5 decades has been proposed based on a literature review of previous studies. This study also identifies areas for additional SDB research where there are obvious hurdles and knowledge gaps. The recent increase in SDB research has concentrated primarily on shallow water depth estimate, which faces numerous difficulties because confounding variables are dynamic and active models for coastal areas must be developed.

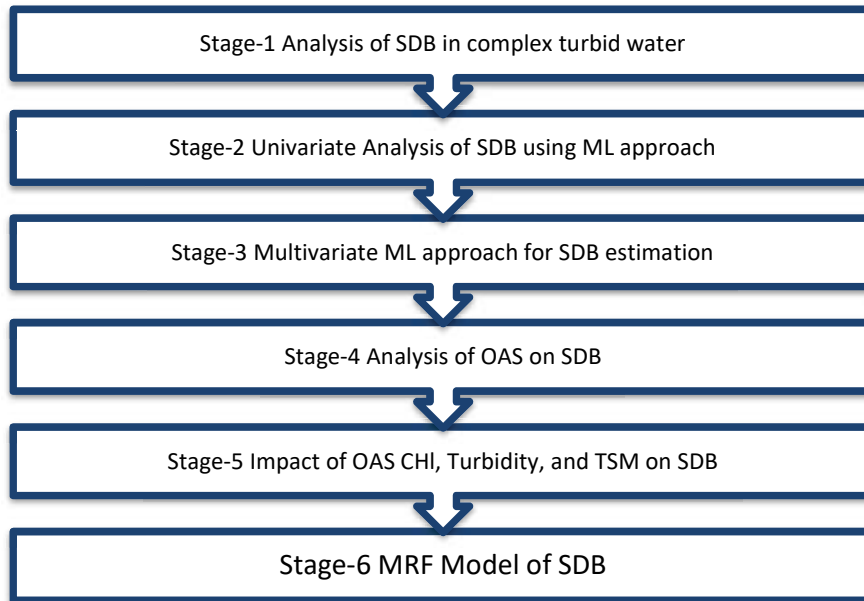


Figure 1.1- Overview of Research

The exploratory portion of the study includes an assessment of satellite imagery suitable for obtaining the SDB of the study area characterized by turbidity and complexity. To achieve this goal, several commonly used SDB algorithms were used to generate SDB. The techniques included; i) Log Ratio algorithm (Stumpf et al., 2003); ii) log-linear models (Lyzenga et al., 2006); and iii) The tools in the Envi 5.3 suite SDB measurement model, SPEAR (Spectral Processing Mining and Analysis Software) use the Log Scale algorithm. In addition, ML techniques which include, Linear regression, RF regressor, and SVR were evaluated to obtain SDB using the high-resolution satellite data.

The above stages have proven the superiority of ML over traditional algorithms. In the next step, the performance of univariate ML methods viz. i) Linear, ii) Robust Linear, and iii) non-linear ML algorithms in SDB derivation using LANDSAT-8, ASTER, & SENTINEL-2A spectral bands and two different high-resolution in-situ bathymetric datasets of turbid water (Ashphaq et al., 2022). A univariate technique was used based on the SDB literature as most existing studies recommend a wavelength band of 0.5–0.6 μm for estimating SDB in coastal waters (Stumpf et al., 2003). We examined the depth-dependent reflectance for spectral bands and computed empirical correlations between remotely sensed reflectance and satellite bands. The selected study sites were heterogeneous as well as navigably

complex, so LR, three Robust LR, and nonlinear GPR algorithms were compared for SDB estimation.

This study then evaluated the effectiveness of the SDB method using multivariate ML regression techniques. The goal of this phase was to validate the use of several ML regression algorithms for SDB and investigate the most appropriate techniques for predicting his SDB for multivariate dependent variables. This phase demonstrates multivariate ML regression techniques MLR, GPR, MARS, DTR, DTR-K, and RF at two different sites using three satellite datasets and high-resolution in-situ bathymetry.

Based on the accuracy of results achieved in previous phases, identifying variables that limit the use of a single scene for deriving SDBs across coverage areas, or different remote sensing reflections for different images of different dates were considered essential to variations in the IOP of the water column. To study this, we used coastal seawater longitudinal data, remote sensing reflectance data, and bathymetric data provided by the INCOIS-SATCORE project to study the relationship between them. This study analyzed the relationship between bathymetry and each SATCORE parameter distribution in the study area. Regression modeling analyzed the causal relationship of the dependent variable 'bathymetric measurements' and the independent variables 'remote sensing reflectance', and OAS.

Results from the last stage showed that bathymetric topography only affected the spatial distribution of chlorophyll, TSM, as well as turbidity in coastal waters. Therefore, the goal of the next step was to improve the accuracy of SDB by understanding the spatial distribution of chlorophyll, total suspended solids, and turbidity of the coastal waters. At this stage, a preliminary survey was conducted using numerical analysis between Landsat 7 and 8 spectral bands, OAS parameters, and bathymetry. SDB was derived using three machine learning algorithms. Linear, RF, and SVM regression. The resulting SDS residues were analyzed for chlorophyll, turbidity, and TSM in coastal seawater. In this study, we found that most of the erroneous SDB residuals and peaks were distributed in high- or medium-rich OAS regions. The results of this study pave the way for further improving the accuracy of SDB estimation by considering the above three OAS effects in coastal waters.

The RF-SDB model showed considerable accuracy in deriving SDB at the last stage. Further evaluation of the algorithm in SDB estimation in different scenes is necessary to demonstrate the robustness of the practical model. Results from previous

research stages showed that most of the erroneous SDB residuals and extremes were distributed in high- or medium-enriched OAS regions. Therefore, the three OAS Chl, TSM, and turbidity were included in further data processing to improve the accuracy of the SDB estimation by considering the effects of the above three OAS in coastal waters. The aim is to develop a more robust predictive SDB model for practical application by merging algorithms developed at different sites and datasets from all sites to derive the Merged RF algorithm. Validation and cross-validation were performed and performance evaluations were elaborated using in situ and bathymetry maps derived from the GEBCO database.

1.5 CONTRIBUTION OF RESEARCH

This study has contributed to the SDB domain in several ways as follows:

1. Firstly, the study examined the contemporary, most cited SDB model with ML algorithm using single spectral band as well as multispectral satellite data.
2. The study proved the efficacy of ML algorithms compared to other SDB models in highly complex and very turbid water in the study site of Vengurla Rocks.
3. Further, the study has compared several ML algorithms for SDB derivation, and based on higher efficiency, the less time-consuming and computationally extensive model was suggested.
4. This study was one of the pioneering efforts in examining the role of OAS on SDB derivation.
5. The first time in SDB such vast longitudinal data was used for preliminary analysis of OAS on SDB.
6. The INCOIS SATCORE data have been collected from 2008 to 2016 and have been shared for the research to identify the moderating or mediating effect of OAS on SDB derivation. This data has been collected with the use of great amounts of effort and huge resources.
7. The study on preliminary analysis of OAS concluded only three parameters in the water column have an effect on SDB derivation and the same were used in further processing.
8. The SDB derivation for such a vast region of 100 Sq Km to 1400 Sq. km is rare in SDB, Besides, this study has applied SDB derivation up to a depth of 90 m with reasonable accuracy in the Gopalpur site.

9. A generic Merged RF algorithm which included three additional parameters Chl, TSM, and Turbidity derived from the same imagery was developed, tested, and validated for recursive bathymetry mapping in the coastal region of India.

1.6 OUTLINE OF THESIS CHAPTERS

The thesis is compiled into six chapters wherein each chapter comprises of Section, subsection, and paragraphs. The First Chapter INTRODUCTION briefly introduces the importance of bathymetry, the need for research on SDB, motivation for undertaking this study. This chapter also provides an overall summary of the research work. The second chapter LITERATURE REVIEW provides the importance of Satellite Oceanography in Blue Economy development and the chronological progress of SDB studies. This chapter also introduces all important theories, approaches, models, methods, and techniques in SDB. The chapter provides details of literature progressively from Western to Indian studies chronologically. The third Chapter MATERIALS AND METHODS includes details of Study Sites at various stages of research, Data used in the study, Pre-processing steps from Atmospheric correction to Surface reflectance used in this study, and Performance Evaluation matrices used in this study. Further, this chapter includes details of various stages of research. The fourth chapter RESULT provides details of all the stages of research. All the results of transforms & algorithms have been shown in the form of tabular data as well as resultant bathymetric maps. The fifth chapter DISCUSSION explains the performance evaluation of retrieval ML algorithms of SDB, validation of chlorophyll, TSM, & Turbidity and their relationship with SDB retrieval, and the development of a Merged RF model for SDB. The Sixth chapter is the summary and conclusion chapter. This is followed by References and additional appendices which include the syntax of various ML algorithms used in this study.

CHAPTER 2: LITERATURE REVIEW

The potential of the Blue Economy has been widely discussed over the past decade for the country's economic growth, improving resource use, creating livelihoods, and protecting oceans through sustainable use consistent with the goals of the United Nations Ocean Decade (Lecours et al., 2022). Satellite oceanography for a sustainable Blue Economy has been studied for the past five decades. The space technology boom of the past decade for coastal regions has been fuelled by new high-resolution satellite sensors and advances in the processing capabilities of modern-generation computers, especially in applying machine learning algorithms to large amounts of geospatial data. Extensive data processing capabilities give the advantage of aggregating existing knowledge across multiple domains into a unified approach that can drive the blue economy through accurate depth prediction, estimation column parameter and bottom composition as well as derivatives based on WQP. Advances in recent years have strengthened the prospect of more accurate, real-time products based on satellite oceanography. However, relatively few applications have been operated based solely on satellite oceanographic data due to the barriers of on-site data to validation and limitations caused by sensors and techniques in coastal waters.

2.1 SPACE TECHNOLOGY & BLUE ECONOMY

Recent satellite oceanographic studies on the recovery of WQPs, their application in Blue Economic activities and the various challenges in creating WQPs for a sustainable blue economy, mapping Habitats and other coastal applications were reviewed to identify knowledge gaps and challenges. Areas of opportunity such as diving operations, military applications, etc., where less theoretical knowledge is available, are analyzed based on the WQP relationship. It is seen that the need to integrate data from various sources is emerging, to overcome the challenges and limitations of WQP retrieval to operate blue economic products.

Advances in satellite oceanography over the past few years have reinforced the prospect of more accurate, real-time, and operational products for a thriving blue economy. Various national and international agencies have identified the need to converge data from multiple sources into a unified database to improve the availability and sharing of marine data with all end users. Satellite oceanography has

the potential to provide products that are sourced, even to any remote location in the world. Spatial variability along coastlines, monitoring of complex and dynamic estuarine systems, and management of navigable harbour channels require immediate attention to data integration. This section discussed coastal water remote sensing theory and its application areas, including some lesser-known applications such as diving operations, depth measurement and detection of small/large objects in the incident area.

Ocean color sensors mostly have a coarse spatial resolution, making them unsuitable for coastal, estuarine, and estuarine regions. Ground-based sensors with higher spatial resolution are used to overcome spatial resolution issues, but the resolution is low. The highly variable optical properties of coastal waters complicate the distinction between and within water bodies based on varying concentrations of sediment and plankton. The ability to estimate phytoplankton size, type, and physiology from remote sensing data is relatively underdeveloped, and many methods are still not widely validated. Coastal areas, especially in bays, inlets and estuaries, have less geographic coverage/width. The coarse resolution of HS sensors is below the spatial detection of single pixels. Most of the shoreline detection & delineation methods use on-the-shelf commercial tools making techniques accessible to only a few. The progress of automated procedures for shoreline detection & delineation is in its infancy and needs further development. The existing algorithms and methods are limited to specific satellites, and common algorithms are not available independent of the application or image. Thus, efforts are needed in the direction of complete automation applied to very high-resolution satellite imagery to get relevant products for the blue economy.

The contemporary literature reveals the surge in studies focusing on satellite oceanography or coastal water remote sensing in the last decade. However, several challenges are apparent in developing operational products for coastal regions for the reason of the dynamic variables of the water column. The synergy of the existing knowledge established from the research of the last four decades to create operational products for coastal regions demonstrates efficacy to provide appropriate and precise information for a sophisticated operation in ports & harbours, navigable channels, creeks, estuaries, and sensitive marine parks & protected areas. Coastal water remote sensing is highly dependent on the retrieval of WQP and application areas. Bathymetry retrieval and its relationship to WQP have been one of the unexplored

areas of research in this domain. A further literature review is guided by focusing on different bands, different algorithms, and a detailed discussion of RT equations, variables and their estimation by several researchers. Along with this, research has also highlighted research areas having challenges and substantial knowledge-gap existence for a sustainable and prosperous blue economy.

2.2 SATELLITE DERIVED BATHYMETRY

Accurate bathymetry information is crucial for a variety of applications, including maritime navigation, harbour construction, the laying of underwater cables and pipes, and more. In coastal areas, socio-economic activities like fishing, maritime trade and transport, tourism, coastal aquaculture, offshore exploration, experimental exploration of alternative renewable energy, marine ecosystems, and such other applications have grown over the past decade, which require bathymetry information primarily. Bathymetric mapping also has wide application in evaluating coastal processes, which include sea-level changes, littoral drift, erosion and accretion of shorelines, coastal vegetation and wetlands, sediment concentration, coastal currents, coastal habitats, etc.

Bathymetric surveys mostly employ sonic echo-sounding technology to acquire data. Single Beam Echo Sounding (SBES), which offers less cover and spatial resolution, wherein Multi Beam Echo Sounding (MBES) has wide coverage and completely insonifies the region, are the two methods used for acoustic echo-sounding. MBES also depicts undersea topography. The bathymetry of the ocean is now being determined using a variety of contemporary methods, viz. LIDAR, ROVs and AUVs (fitted with SBES, MBES, and or LIDAR) for accurate depth estimation in coastal areas. The substantial operational costs of hydrographic surveying prevent them from being repeated and very routine and frequent in any region of interest. Due to the risk to human life and material damage, conducting hydrographic surveys in certain isolated regions, such as large hidden reefs, and estuaries, tidal bores & surge zones, is extremely difficult.

The purpose of the study is to objectively evaluate the applicability and flaws of the evolution of SDB algorithms during the past fifty years. Based on statistical, bio-optical, and physio-optical properties, algorithms have been classified. The purpose of this literature review is to demonstrate the applicability of SDB algorithms in various coastal areas.

2.2.1 SDB Methods

SDB is among the possible options that have been examined for the past 50 years and successfully offer a crucial answer to coastal regions that are characterised by rapid changes in the bottom and complicated terrain (Ashphaq, Srivastava, & Mitra, 2021). However, depending on already-existing in-situ bathymetric data, these approaches have the potential to produce outcomes. The SDB Method was created to estimate SDB from optical RS data utilising the concepts of underwater reflectance, and underwater optics (Polcyn, 1969 & Colleagues). In sites where high energy waves occur with a number of limiting criteria like the swell magnitude, and interaction between swell & sea waves, synthetic aperture radar data is utilised to estimate coastal bathymetry.

Despite the SAR dataset having a modest resolution & coverage, obtaining the necessary input parameter and putting an algorithm into practice are difficult tasks. In shallow coastal seas, this method is unreliable (Wiehle et al., 2019). Another team of researchers developed a lesser-resolution bathymetry of the seabed using satellite altimeter readings and sparse in-situ bathymetry. For use in coastal locations, the satellite altimetry dataset has comparably extremely poor accuracy & resolution (Smith & Sandwell, 2004). The capacity to map the topography of the seabed has significantly improved because of the advent of high-resolution optical sensors and advancements in algorithms. Over the last 50 years, several researchers have employed optical data in developing algorithms. The current techniques for hydrographic survey and SDB are listed in Table 1 below, with each technique's reasonable depth, precision, robustness, and restrictions shown.

Table 2.1: Summary of Bathymetry (Modified after Jawak, Vadlamani, & Luis, 2015)

Method	System	Sensible depth	Accuracy	Affecting factors	Strengths	Limitations	Applications
Ship based Systems Echo Sounders	Singlebeam ES	Shallow to deep	High	Single Footprint	Highly reliable Wide depth range	Expensive, High cost of operation	Diverse environments as per the IHO Standards
	Multibeam ES	Shallow to deep	Very high	Swath, Heave	High Precision Wide depth range	Expensive High cost of operation	Diverse environments with Very High Resolution
Non-imaging Active RS	LiDAR	Up to 70 m	Very high	Water clarity, bottom material, surface state, background light	Wide depth range; concurrent measurement not essential	Expensive Limited swath width	Clear waters with Very High Resolution
	Radar Altimetry	Beyond 40 km From the coast	Very Low	Elastic thickness Of lithosphere and/or Crustal thickness, sediments	Global coverage, needs Only simple altimetry with no iono / troposphere measurement	Possible over a limited Wavelength band	Coarse bathymetry Derivation in open ocean Deep seas & Oceans with accuracy of ± 50 m
Imaging Active RS	Microwave / SAR Spaceborne	Shallow to deep	Low	Image resolution slicks, fronts, weather condition (eg. waves)	Over large areas Not subject to cloud cover	Complex and not so accurate	Open, coastal and oceanic waters but Unreliable, Low accuracy
Imaging Passive RS	Optical – analytical	Up to 30 m	High	Water quality, atmospheric conditions	Based on physical Process, Accurate	Complex as several input parameters required Concurrent Sea truth essential	Turbid and shallow inland waters, estuary and river Nearshore and coastal waters
	Optical – empirical	Up to 30 m	Varying	Atmospheric calibration, water turbidity Bottom reflectance	Simple to implement Accurate at certain depth	Limited depth Accuracy lower at a depth Concurrent Sea truth essential	Theoretically, the 0.48–0.60 μm radiation is able to penetrate clear, calm sea water up to 20 m.
	Video	Tidal height	High	Image resolution	Able to reveal minor bathymetric change	Restrictive area, Bathymetry along profiles	Shallow water with vegetation; accuracy not yet established

2.2.2 Optical SDB

SDB depends upon radiant energy reflected from the visible spectrum of the EMR, and its intensity decreases with the depth because of the IOP of the coastal water column. Signal attenuation caused by the atmosphere, the characteristics of the water column, the characteristics of the bottom, and the depth of the water are some of the confounding variables that restrict reflectance. The fundamental idea behind remote sensing techniques is to use known values, which are approximately gathered through field observations, to identify the majority of the unknown parameters. To calculate the water depth, we need to find values for a few unknown parameters. One way of doing this is by using optical bandwidth as found in work done by Polcyn and Rollin back in 1969. Another method developed at around the same time was based on reflection rates within two spectral bands of visible light Polcyn & Rollin, (1969). This has been followed up with more recent research that aims to reduce the number of variables needed for an accurate estimation. Experimental methods have also continued playing an important role in SDB studies over the years, allowing us to get closer and closer towards solving its mysteries.

In order to accurately measure the water column and depths below, a new method was developed by Lee et al., (1998). This system involves measuring both the depth and IOPs. Over the past decade, research has been devoted to correlating these parameters with one another. A large database of marine IOPs and oceanographic data has also been collected, making it possible for more accurate measurements in the future. In addition to LUT techniques that have recently emerged, this information will help us better understand how our environment is changing over time. The development of powerful computers with large data processing and storage capabilities has led to the application of ML algorithms in SDB. In addition, this technology has also introduced a number of ML techniques into the world of database management.

2.2.3 SDB Algorithms

The literature on SDB for almost five decades is classified mainly based on the method to estimate SDB derivative coefficients in analytical, SA and experimental methods (Jawak et al., 2015; Misra et al., 2018; Traganos et al., 2018). Others have mentioned similar studies under the classification of physical or statistical-based methods (Dekker et al., 2011). In addition, similar studies were referenced differently among the above categories depending on the search context. SDB experts, especially from technical backgrounds, regularly and interchangeably employed phrases such as approach, tool, method, technique, & model without comprehending the philosophical worldview behind these concepts (Ashphaq et al., 2021). Despite the great differences between these terms, none of them have an advantage over the others but relate to completely different purposes, have distinct meanings, and should be used consistently in an appropriate way. To systematically categorise any issue, attention should be made to the topic's specifics and agreement among stakeholders for the appropriate application of the requirements. Based on the extant literature on SDB during the last five decades, this article provides an overview of the SDB classification based on the research philosophy.

2.2.4 SDB Classification

In language learning theories, Anthony (1963) separated the words method, approach, and technique. The word approach refers to the underlying philosophy or ideology

associated with a certain objective that focuses on addressing a problem about the set of assumptions inferred from a group of theories and concepts (Andiappan & Kin, 2020). The SDB literature primarily discusses two approaches; i) statistical which unconsidered the physical characteristics for light transmission in the water and optical properties features of the water column, is used to determine the relationship between R_{rs} and bathymetry measurements; and ii) physical properties-based emphasizing propagation of light in the water column and also attenuation of light caused by water column composition and other environmental factors.

A model specifies a broad conceptual framework or approach to problem-solving. When describing a model as a collection of approaches utilised in accordance with predetermined principles, the terms methodology and framework are sometimes used interchangeably. The better the reasonableness in determining a sample, the more accurate a method is chosen. The physics-based RT approach can be of two categories: the bio-optical model & the optical-physical model. The bio-optical model is predicated on the idea that the optical properties are primarily controlled by biological substances in the water column, particularly phytoplankton & also its derivatives (Smith and Baker, 1977), whereas the other model explains the reflectivity of R_S as a function of depth, water quality, and bottom reflectance (Lee et al., 1999; Hedley et al., 2009).

A method is a description of an actual approach to problem-solving. Empirical methods, which include a variety of statistical techniques, can be used to implement statistical methods, which are very straightforward. Semi-empirical (SE) and analytical approaches can be used to carry out optical-physical and bio-optical models in this context. The empirical validation of converted data to field data is used in the SE, which explicitly assumes light's RT and also its attenuation in the medium it is propagating. For the RT of light in a body of water, analytical methods are physics-based algorithms based solely on the photo-physiological-biological properties of the water components. However, when applied in practice, the analytical method has been regarded as a difficult and complex problem (Werdell et al., 2018). Several theoretical assumptions have been used to address the limitations and complexity of analytical methods. These assumptions divide analytical methods into semi-analytical (SA) and analytical methods (QA). Lee et al. (1998; 1999), developed the SA methods, and transformed (Maritorena, Morel, & Gentili, 1994; Mobley, 1994) using the optical properties in the analytical RT equation using spectral fusion without topographic

data, on column depth d and seafloor reflectivity as a function of water backscatter and light absorption properties. Among other scientists, the QA method was created by Lee, Carder, and Arnone, (2002) to decide the absolute ingestion α (which can be deteriorated into phytoplankton take-up coefficient α_ϕ and gelbstoff α_g), and backscatter coefficient b_b , given the connection between far off regions utilizing the RT radiation equation to identify water reflection and IOP.

According to Andiappan & Kin (2020), the technique is a series of actions taken to observe and measure phenomena as well as the collection, processing, and analysis of data regarding the outcome. In any case, not all issues should be settled in each class, consequently proposing a "Hybrid of a methodology, model, strategy or procedure. As a result, it has been suggested to group studies on SDB according to the diagram in Figure 2.1, which explains how the aforementioned concepts relate to SDB.

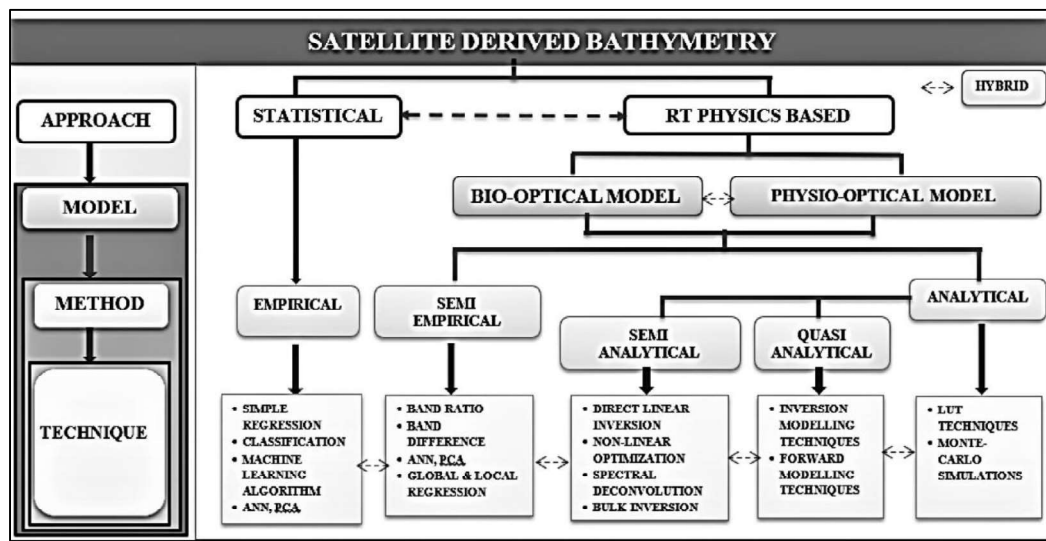


Figure 2. 1: SDB Classification Conceptual Framework

2.3 OPTICAL SDB APPROACH

2.3.1 Statistical Approach

Estimates, or coefficients of determination, for the same satellite imagery from which they are drawn, are the primary objective of a statistical method that is based on a statistical relationship. This relationship does not take environmental, spectral, or radiometric parameters into account. The accuracy of statistical methods solely depends on the data on which the statistical coefficients are based or the physical

model from which they are trained. The ease with which statistical methods can be used, the availability of tools for data processing and analysis, and recent advancements in advanced machine learning techniques have increased processing efficiency are the main benefits. According to Dörnhöfer & Oppelt (2016), the limitations include the requirement of on-site data, the specific adaptation to the same image and location, and the difficulty of porting to other locations. Few studies have used "spatial transfer" (applying empirical methods to other neighbouring locations) and "temporary transfer" (different times of ground data acquisition on the same study site) for SDB studies (Danilo & Melgani, 2019).

2.3.2 Radiative Transfer Approach

The radiative Transfer (Physics-based) approach is based upon the exponential attenuation of EMR or light with an increasing water depth of the water column and also its reflection either from the bottom or water column (Bramante, Raju, and Sin, 2013). Blue and green are often the bands chosen for SDB since they have the lowest light absorption rates. The physical attributes of spectra, the water column, and environmental factors, including Chl concentration, TSM, the concentration of turbidity, spectral shape, absorption & backscatter coefficients, and depth, may all be estimated and interpreted using physics-based methods (Brando et al., 2009). Estimating the physical factors that influence spectrum observations, whether they include or exclude in-situ data, is the intrinsic strength of physics-based techniques. Though difficult to execute realistically, physics-based methods of RT are known as complicated mathematical problems (Mouw et al., 2015). Additionally, in order to explain the simulated link, RT physics-based techniques need an in-depth theoretical understanding of unaccounted-for components.

Therefore, the use of physics-based RT techniques should only be promoted in the presence of complete knowledge of the biological and physical processes occurring in water. Beer's Law is theoretically used to describe the intensity of radiation dispersed and absorbed by water, leading to a physics-based approach to SDB. Since then, several methods for estimating SDB have been created under the names of physics-based models or bio-optical models.

2.3.3 Models in RT Approach

Under two separate but complementary sets of assumptions for reflection data, a physics-based approach is used. The first, known as the bio-optical model based on the concept that the optical characteristics of water are governed by biological substances in the water column, especially phytoplankton as well as its other derivatives (Smith and Baker 1977). It employs a forward model that explains R_{rs} as a consequence of water condition, depth of waterbody, and bottom reflection and yields depth estimations when inverted (Lee et al., 1999; Hedley et al., 2009).

2.3.3.1 Physio-Optical Models

Sir Rayleigh (1899) proposed a theory that diffuse reflection, as well as transmission of sunlight, causes molecular scattering in the atmosphere. However, Chandrasekhar (1950) gave a solution to the Rayleigh issue in the book RT (Suomi & Haar, 1970).

Chandrasekhar defined "RT" as the mathematical solving of RT equations in absorbing, emitting, and scattering media. RT theory, which is the most extensively used approach for reliably extracting features of the Earth and atmosphere from satellite data, gives the justification for causal links between data received by instrument sensors and the physical course that created the signals. It has evolved into a useful tool. The scattering behaviour of RT theory in different transmission media and its representation of the RT equation in solution techniques used for atmospheric adjustments, studies of airborne particulate matter and clouds, water bodies, flora, etc. have all been the subject of many types of research.

2.3.3.2 Bio Optical Model

The term "bio-optical" was used by Smith and Baker (1977) to describe the within-water optical state which is primarily influenced by the optical characteristics of the biological components (Chl, plankton, sediments, etc.) in the water column. Since the term has been applied to describe bio-optical models in a variety of ways (Ogashawara, 2015). Radiometric variables IOPs & AOPs such as downwelling as well as upwelling spectrum irradiance as well as the absorption & scattering characteristics of components of the water column are the foundation of bio-optical models. These spectrum features may be quantified numerically and calculated at an individual cell level (using physical variables like cell, dispersion size, chemical

composition, etc.) (Morel, 2001). The very first bio-optical model developed correlations between the AOPs & IOPs of water using a Monte Carlo simulation for each RT equation. The wide and inconsistent use to designate ‘Bio-Optical’ models was also highlighted (Ogashawara, 2015). Such models are dependent upon radiometric quantities like IOPs and AOPs, such as the downward and upward spectral irradiance as well as the absorption and scattering properties of elements in the water column (Ashphaq et al., 2021). These spectral features can be defined at the level of a cell (using physical structures such as cell size, size distribution, chemical composition, etc.) amount for that population of cells (Morel, 2001).

2.4 SDB METHODS

Wherein, SE & analytical methods may be implemented in the Bio-optical model and the Physio-optical model, respectively. The SE techniques calibrate field data empirically which has been converted and is dependent on the explicit assumption of RT and attenuation of it through the propagating medium. Analytical approaches presume algorithms based only on the bio-physio-optical characteristics of water components and solely grounded in the physics of RT of light inside a water body (Ashphaq et al., 2021). However, the analytical approach has been viewed as a contentious mathematical topic that is difficult to implement practically (Mouw et al., 2015; Werdell et al., 2018). Few theoretical presumptions, based on that analytical techniques get divided into SA & QA Methods, addressed these analytical method limits and complexity issues. The development of SA techniques may be credited to Lee et al. (1998; 1999), who adapted the analytical RT equation (Maritorena et al., 1994) to determine the optical properties, water depth, and seafloor reflectance based on the water absorption & light-scattering characteristics without the requirement for field data. Lee, Carder, and Arnone (2002) developed the QA method to derive total absorption and back-scattered coefficients, depending on interactions between R_{rs} with the IOPs for a given column of water by application of the RT equation.

2.4.1 Empirical SDB

According to Ashphaq et al., (2021), the statistical estimators developed from ground truthing data are the single basis for empirical approaches. A dataset of ground observations and reflectance values of suitable bands from satellite photography is

used to construct the empirical method. Without taking into account physio-optical characteristics, the empirical approach estimates the strongest possible relationship among the reflectance and other parameters of interest using techniques like linear regression, least squares, non-linear regression, maximum likelihood, neural networks, etc. These algorithms have the benefit of processing vast volumes of data quickly, easily, and efficiently. After reviewing a number of empirical researches on remote sensing in coastal waters, Matthews (2011) concluded that, if one is willing to accept a significant level of error, empirical approaches can yield a great deal of useful information. Empirical methods are site-specific and time-dependent since they are based on the premise that optically homogenous surroundings exist in a single scene. But empirical techniques utilising ML and multitemporal data have assisted in overcoming these restrictions (Salameh et al., 2019). Below is a discussion of significant empirical method-based strategies used in SDB investigations.

2.4.1.1 SDB Techniques: Empirical

The majority of empirical approaches use regression tools to analyse data utilising spectral values from one or more bands along with ground-truthing data to determine the coefficients parameters of regression. Although regression approaches have been seen to be the utmost practical answer to large dataset processing, they may lead to algorithm results failure in areas with varying seabed conditions (Doxani et al. 2012). Gao, (2009) reported that regression resulted in coefficients worsening in heterogeneous bottom types; consequently, a new regression technique needs to be created for individual bottom types comprising different natures of bottom and underwater flora of the region.

Gao (2009) claims that mixed bottom types have worse regression coefficients; as a result, various regression algorithms may be created for each kind of bottom, which can include the vegetation and bottom characteristics of the location. Prior to using band ratio methods, the bathymetry and bottom in satellite pictures are classified using the supervised/unsupervised classification approaches, either alone or in a hybrid combination with SE. Clark, Fay, and Walker (1988) employed a supervised cluster and the maximum likelihood before using a band-ratio technique on clustered satellite datasets. They used this technique to identify the bathymetry, and spectral properties, of the seabed. This is a key concept of classification algorithms. After classification, statistical bathymetry variables from training areas

are used to calibrate the picture (Correa & Avila, 2002). Radiometric reflectance is divided into broad groups via unsupervised classification, which is then utilised to produce areas that may be regressed versus bathymetric measured data (Collet et al., 2000; Correa 2002; Mavraeidopoulos et al., 2019). The effectiveness of grouping pixels into subgroups prior to applying the SDB algorithm has been determined by a number of unsupervised classification techniques, including (Poursanidis et al., 2019) Iso-Cluster unsupervised classification and K-Mean unsupervised classification (Geyman & Maloof, 2019).

Other methods include PCA, which correlates water depth using the first component and all three bands (transformed) (Gholamalifard et al., 2013). Mohamed et al. (2016) employed PCA to find SDB, and as in-situ water depths and the main components of the log-transformed reflectance had a linear relationship, this allowed for a superior SDB estimate. In a few investigations, SDB and seabed categorization were also calculated using Maximum Likelihood techniques (Zhou, 2011). Jay and Guillaume (2014) suggested a maximum-likelihood (non-stationary) estimate method for satellite bathymetry and water quality using HS data to achieve better results. SDB also employs a number of other empirical & image-based methods, including object-based image analysis (Hedley et al., 2018).

The results of a few empirical SDB research, the depth of the study area, and the satellite data utilised are shown below.

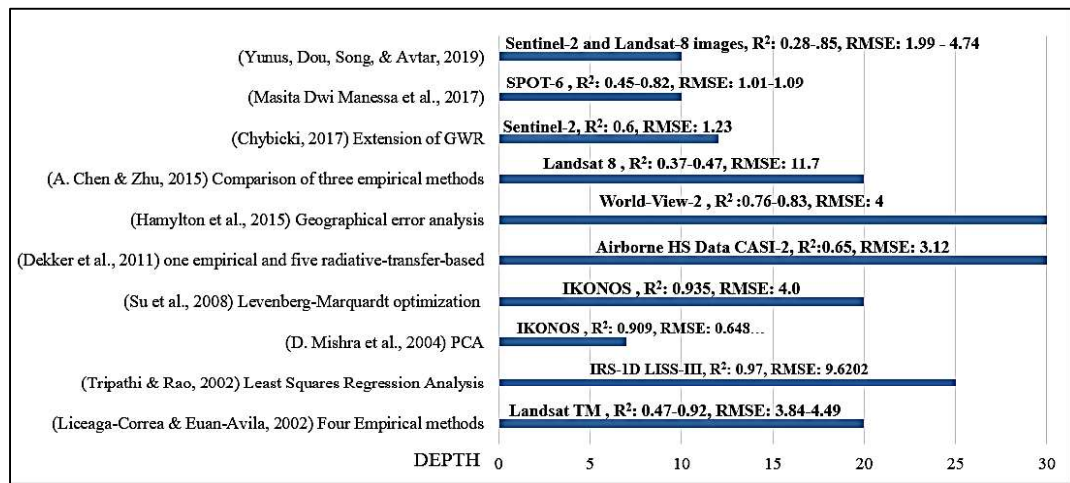


Figure 2.2: Empirical SDB

In the limited empirical studies, the effects of a few environmental confounding variables on SDB estimates have been considered. Tripathi & Rao (2002) researched the effect of turbidity on SDB using IRS-1D (LISS-III) Band 1 (0.52 - 0.59 nm) and

developed an adjustment feature called ‘Turbidity Influence Factor’ minimising inaccuracy. Although predicted SDB's RMSE was high, the Least Squares Regression method between band 1 reflectance with a water depth of 25 m generated a substantially high R^2 of 0.97 when employing TIF. Mishra et al. (2004) studied bathymetry in Honduras for a lesser depth of 7 m but varying inconsistent bottom types using PCA on IKONOS data (seagrass, sand, and coral). They achieved an R^2 value of 0.90 and with a standard error of 0.64 m.

Some researchers have equated strategies to find the finest appropriate algorithm developed to date for their study region. Four empirical methods were used with the Landsat TM dataset in the Alacranes Reef (Gulf of Mexico) for a depth range of 20 m: a linear regression with the first principal component, an MLR, a non-supervised (two-step) classification with the MLR, and a supervised classification. The results show RMSE of 4.1 m, 3.8 m, 3.8 m, and 4.4 m, respectively, demonstrating that the two-step non-supervised yield low RMSE (Correa & Avila, 2002). Using Landsat-8 data, Chen and Zhu (2015) investigated three empirical techniques to estimate SDB at Pratas Is. They concluded that regression was not robust due to significant outliers.

The empirical findings indicate that SDB achieved by numerous studies have up to 60 percent accuracy at validation sites where RMSE is also less than 1 m, while some studies projected data have high errors and several of them have RMSE more or less about the depth of the region. This makes it obvious why empirical approaches to determining SDB have only occasionally been used. In addition, it is suggested that empirical approaches be used with caution in coastal turbid water due to larger errors based on research in the turbid region.

2.4.1.2 SDB: Machine Learning

With more flexibility in approaches to analyse enormous volumes of data, ML is becoming an extensively acknowledged tool for research used by scientists in remote sensing & GIS investigations. In research using remote sensing, machine learning has been widely accepted, particularly when processing temporal high-resolution spatial information or high-resolution ground truth data. In a pioneering study that applied machine learning to SDB, Ceyhun and Yalçin (2010) used the ANN ML algorithm to analyse ASTER & QUICKBIRD satellite data at Foca, Izmir, Turkey, where a depth

of 45 m, and they were able to derive substantially accurate SDB estimates with a coefficient of determination 0.80.

With the use of the SVM technique, Eugenio et al. (2015) were able to get R^2 values between 0.93 & 0.94 and Residuals between 1.20 & 1.94 m for the coastal regions of the Canary Is. using World-View 2 data and WorldView-2 & WorldView-3 data in Cape Rodney of New Zealand, Kibele & Shears (2016) applied non parametric ML technique nearest neighbour regression. They compared the results with the Lyzenga SDB model and showed that the KNN approach outperformed Lyzenga's methodology. ANN ML algorithm was used to analyze IRS-P6 (LISS-IV) in the murky water for a depth range of 12 m at Bhopal City Lower Lake, attaining R^2 0.95 and RMSE of 1.61 m. This demonstrated that ANN ML approaches can also be employed without further adjusting for confounding environmental variables like vegetation and bottom (Patel, Katiyar, & Prasad, 2016). A few studies are shown below that used ML algorithms techniques to analyse SDB, along with information on the techniques used and the results obtained.

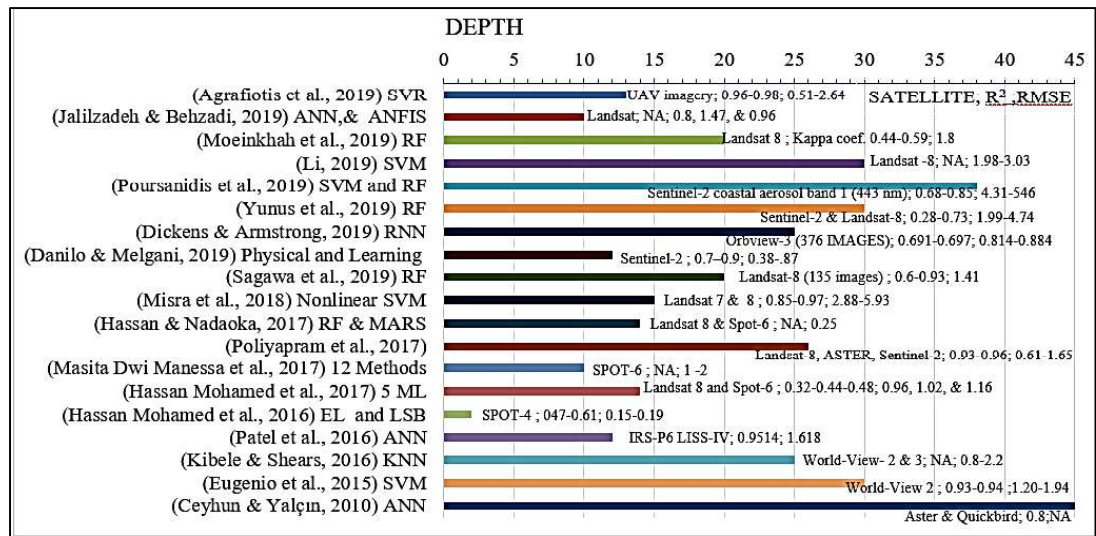


Figure 2.3: SDB using ML algorithms

SDB of five ML has been compared by Hassan et al. (2017) which includes SVR, Neural Network, and the GLM of Lyzenga, using satellite data of Landsat-8 & Spot 6 at for depth range of 10.5 m three distinct sites at Alexandria port of Egypt; upto 6 m at Lake Nubia, and up to 14 m at Ishigaki Island, Japan. Misra et al. (2018) used the Non-linear SVM ML technique in shallow water at Maarten Is., Netherlands, at depths ranging from 1 to 15 m. The SVM algorithm performed better in shallow

murky water, with an R^2 of 0.73. Another notable study was undertaken in 2019 by Sagawa et al., using multi-temporal Landsat images clustering 135 scenes at five different sites with depths ranging up to 20 m. The RMSE of the resultant SDB was found about 1.41 m in the five different regions. However, the derived SDB was in the different shallower water areas under highly open conditions. According to Dickens & Armstrong (2019) found that even SDB derived from deep learning approaches did not match IHO bathymetry standards when they used RNN to 376 ORBVIEW-3 multi-temporal satellite images acquired at 3 distinct sites in the water depth of 25 m.

The optimum band integration for SDB (1-4), according to Moein et al. (2019), has an RMSE & MAE of 1.25 m and 0.76 m, respectively, in the depth of 0-5 m. Landsat 8 data at a depth of 0-20 m were used in this research. However, the measurement inaccuracy increased dramatically after 10 m. Therefore, it has been concluded that ML algorithms did show themselves to be superior to traditional empirical methodologies. The most common algorithms, SVM & RF, considerably yield SDB for depths range of 10 m, but increasing depth, inaccuracy upsurges exponentially. In addition, relatively limited researchers have used SDB derivation in shallow as well as turbid water. One of the newest areas of research in SDB is the development of operational SDB models, which has a significant amount of potential.

2.4.2 Semi-Empirical SDB

According to (Ashphaq et al., 2021) Many scholars have endeavoured to examine the analytical model for SDB, but prior to the early 1990s, their efforts mostly consisted of matching image pixels with concurrently obtained in-situ hydrographic data, with very little attention paid to other environmental, water column, and physical factors. Without sufficient ground data, the estimate of SDB to an accuracy level of near about 70 percent was made possible by the causal hypothesis of band-ratio (particularly the blue & green band) as a substitute for attenuation coefficients (Polcyn, 1969 and associates; Paredes & Spero, 1983).

Similarly, other investigators proposed, even a single/unique band could be able to explain the exponential decrease of light under water if the disparity between a pixel's true radiance value and its deep water R_{rs} in the same imagery was transformed utilising logarithm function (Warne, 1978; Lyzenga, 1978). The log-

linear transformation function optimally was extended to dual-band (Lyzenga, 1985; Clark et al., 1987) as well as multiband (Lyzenga et al., 2006) channels for increased SDB precision. The improved log-ratio transformation had first been presented by Stumpf, Holderied, and Sinclair (2003) for dual-band, and it was later updated for use with an integration of various distinct bands (Kabiri, 2017).

RT approach increased the possibilities for using MS data for SDB evaluation by making an overt hypothesis (in empirical methods, the premise is intrinsic) in the model in which either Band-Ratio or Linear Regression Method were used to address exponential attenuation of EMR in water, minimizing the number of uncertain variables.

2.4.2.1 Band Ratio SDB

SDB Algorithm based upon Band Ratio approaches pioneered (Polcyn & Rollin, 1969) utilising 18-band data of MS instrument obtained from the aerial platform in the Gulf of Maine. This study has attained depths of 13 ft, which were enhanced up to 20 ft (Polcyn et al., 1970) by an MS instrument sensor in Caesar Creek of Florida. Polcyn and Lyzenga (1973) proposed a numerical approach for SDB demonstrating the correlation of depths up to 5 metres around Little Bahama Bank using Band 4 & 5 (ERTS-1) MSS data acquired in the month of October 1972. On the implicit premise that attenuation is persistent across the satellite scene, Paredes & Spero (1983) expanded the modified ratio assumption to multiband. A semi-empirical solution was created by Stumpf et al. (2003) utilising a reflectance ratio with only using two parameters, which may be used for low-albedo topographies. SDB using the ratio transform approach has shown to be reliable; it can recover data down to a depth of 25 metres, is stable, and gives a normalised RMS error of 30 percent up to that depth range. Nevertheless, this method solitary works in translucent, clear water with better accuracy. This method had been expanded to take into account how turbidity and chlorophyll affect SDB estimates (Caballero, Stumpf, & Meredith, 2019).

2.4.2.2 Linear Band SDB

Warne (1978) demonstrated that Landsat can estimate SDB up to 0-20 metres with a precision of 10% by utilising single-band linear method in Australia. Lyzenga (1981), assumed that the water's optical characteristics were consistent throughout a

particular satellite scene, discovered that this approach delivers correct results to a 15-meter depth in clear water at North CatCay in the Bahamas. To determine SDB at two locations in the Bahamas Islands at depths of 0–10 m, Lyzenga (1985) employed a hybrid airborne sensor that combines a LIDAR system with a passive MS. For depths of 8–10 m, RMSE was determined to be 0.928 m. Lyzenga persisted in working to enhance SDB and put out a multi-band linear approach (Lyzenga et al., 2006).

This technique's algorithm corrects deviations in both attenuation as well as seabed using a summation of log-transformed reflectance, mainly in the green or blue bands. The said model is useful in locations with homogeneous optical characteristics of water & uniform bottom reflectance. This technique also accounts for sun glare, provides flexibility in operations, improves seabed distinction, and improves performance by utilizing two or more than two satellite bands.

2.4.3 Analytical SDB

The physics of radiative transfer of EMR through a water body is fundamental for analytical procedures, which are based primarily on the physical characteristics of water elements including absorption, backscattering, and attenuation. Gordon and Morel (1983) claim that the analytical tools directly apply the RT theory by characterising the absorption as well as backscattering coefficients components of the water. Proper input of a collection of atmospheric effect-related parameters is also necessary for the implementation of analytical approaches. Even small mistakes in atmospheric correction can create considerable retrieval problems since atmospheric influences make up 90% of the total signal while water leaving radiance just makes up 10 percent of the total of it (Caballero et al., 2019). Another problem of the analytical technique is that proper water constituent prediction in coastal vibrant water requires the concurrent gathering of field data and data acquisition. Analytical approaches are computationally demanding and challenging to carry out since there are no perfect atmospheric correction tools that deliver actual water R_{rs} for shallower or optically very complex coastal region waters.

The major strategies used to resolve the RT problem analytically include inverse modelling, forward modelling, LUT based upon the forward as well as inverse modelling, or even any such combination of all three (Hodl et al., 2018). Modelling a

variety of possible models of Rrs as an interactive function of the water composition, bathymetry, & sea bottom reflectance is a method of analysis. By finding the simulated Rrs that correspond closely to the observed Rrs in each & every pixel the most closely, these models are then inverted to establish the constituents of the water as well as bathymetry (Hedley et al., 2009).

2.4.3.1 SDB: Forward/Inversion Modelling

The analytical method especially using ‘Forward modelling’ is applied predominantly to aid three functions; describe, forecast and/or model the inversion to describe the associations among bio-optical or physical parameters and RS parameters, as well as derivative variables such as AOPs and IOPs. Actual satellite imagery was used to elaborate on some occurrences using model simulations. The possibility of modifying the model to more accurately anticipate some occurrences seen in the actual data but not explained by it is presented (Verhoef, 1998). However, the forward model must be inverted in order to retrieve any constituent parameter.

Using identified remote sensing data, inversion modelling to RTE is used in analytical approaches to estimate the input parameters utilised in the model. However, its application is dependent on a variety of known and unknowable characteristics as well as the bands that are available in satellite imagery. Furthermore, the inversion of the model needs intensive computer work and numerous iterations of the forward model to draw conclusions (Verhoef, 1998). The common inversion techniques simulate constants and variables related to the modelling of AOP & IOP using spectral signatures to model parameters. Iteratively revising the variable factors until the disparity between the estimated and real image spectral signatures is just as little as feasible. Data recovery through inverse modelling is possible for water components, water depth, and seabed sediments. Inversion strategies for collecting water components in marine remote sensing have rapidly gained popularity (Hedley et al., 2009). Various inversion approaches, including linear & nonlinear inversion techniques, log-linear inversion technique, adaptive inversion, log-ratio inversion, and so forth have been employed in the SDB research.

2.4.3.2 SDB: using Look-Up Table

The LUT concept describes an extensive database of known component concentrations that comprises spectral signatures, IOPs, radiance leaving the water,

water depth, and sea-bottom properties. Choosing the best match across all of the criteria means assessing the spectral signatures of satellite images in addition to the LUT database (Dekker et al., 2011). An analytical approach in Mobley's (1994) to RTE generated the initial LUT system, named "Hydrolight," via the use of 'Forward Modelling' of satellite images Rrs (Mobley et al., 2005). LUT by spectral matching (Mobley et al. 2005) also known as 'Comprehensive Reflectance Inversion based on Spectrum matching and Table Look up' (CRISTAL) designed for simultaneous identification of bathymetry, benthic substrate, and IOP (Dekker et al. 2011). Another technique ALUT was recognized by researcher Hedley et al., (2009) based on LUT which showed SDB successfully derived for the depth up to 30 m significantly.

In turbid, muddy & shallow coastal waters of Singapore for depths of 4 m, Bramante, Raju, and Sin (2013) used MS data from Worldview-2 to compare traditional SDB and LUT approaches. Despite having a meticulousness of 0.64 m, LUT was constrained by a limited training dataset that had not accurately reflected the variation of the water column as well as benthic characteristics. However, the outcome of the derivation is dependent on the accuracy of the database of LUT, including whether it consists of IOP & AOP, the nature of benthic substrate & their spectral signature, along with water depth as in the region of the satellite imagery. LUT is widely used for SDB prediction as well as water component identification. According to Maritorena et al., (1994), suggested that validating the underlying approximations and quantifying their influence is necessary for the safer application of LUT algorithms. They also employed a two-step comparison process: first, they compared the analytical solution to the exact answer obtained through RT Monte Carlo simulations, and then they compared the spectrum reflectance in varying depth and bottom type to ground truthing data.

Despite significant progress has been made in analytical techniques for calculating marine IOPs using sensor radiance, these techniques still fall short of perfectly resolving to that of the analytical equation. Mouw et al., (2015), draw a similar conclusion describing analytical tools as a contentious scientific issue that requires ongoing improvement to translate from laboratory-based conclusions to field procedures. The said limits and difficulties of analytical tools & techniques were resolved by a limited theoretical convention, that served as the foundation for the division of analytical methods into SA and QA.

2.4.4 Semi-Analytical SDB

When used prior to 1998, the majority of SDB studies refer to SE techniques as SA techniques. The analytical RT formula (Mobley, 1994; Maritorena et al., 1994) was modified by Lee et al. (1999), who also developed the SA approach, to an explanation for the water-column optical properties, bathymetry, and seabed reflectance. This technique was created using spectral matching using the 'Levenberg-Marquardt' optimization algorithm based on the way that water absorbs and scatters light. The water quality parameters dataset is required to enhance the use of HS data to meet the spectral properties in the HS data.

According to Lee, Carder, and Arnone (2002), their proposed method does not require any field data for validation, as well as developed algorithms may be employed in a range of settings to predict the outcome more reliably than empirical methods (Sathyendranath, 2000). Several types of marine research, especially those requiring bathymetry & water-column characteristics, have seen a rise in the popularity of the semi-analytic approach. Though the said method was primarily established for HS data, it may be effective to analyse the MS dataset (Dekker et al., 2011). Utilizing HS satellite data, an optimization approach, and bathymetry in shallow regions, Lee et al. (1999) established a more accurate derivation.

A conventional SA method that relies on spectral matching involves three steps, as per McKinna et al. (2015): (i) calculated Rrs is analytically interpolated to the detected one via forward modelling, (ii) Exponential / power law function is used to describe the spectral forms of unknown quantities, and (iii) An inverse approach is then used to iteratively alter the spectral IOP's magnitude in the forward model. According to Werdell et al. (2018), there are four main categories into which SA-based techniques can be divided: a) Non-linear Spectral optimization, viz. Levenberg-Marquardt method (McKinna et al., 2015), b) Spectral-deconvolution, where the spectral profile is allotted by step-wise algebra; c) Direct-linear Inversion viz. linear-matrix inversion, d) Bulk-inversion which regulate IOPs at each individual band independently.

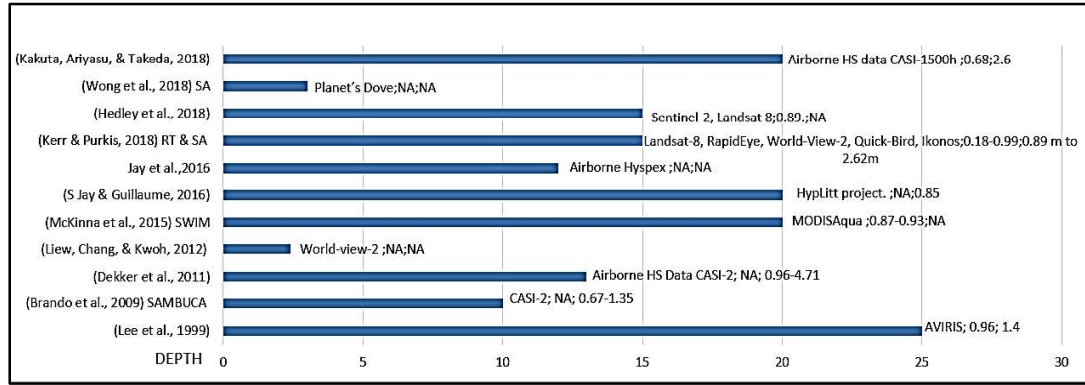


Figure 2.4: SA-based SDB Techniques

Lee and Carder (2000) stated that SA could retrieve SDB with an R^2 of 0.96, but the error grew worse as turbidity increased. Following on the optimization of Brando et al. (2009) developed the SAMBUCA approach, which was used to evaluate aerial HS images from the CASI 2 satellites at Moreton Bay, Australia, at depths ranging up to 10 m. The data shows optically deep water provided less output than shallow clear water. Dekker et al. (2011) equated five SA, including BRUCE, HOPE, CRISTAL, ALLUT, & SAMBUCA by use of CASI-2 at the Lee-Stocking Island, of Bahamas at depths ranging up to 13 m. Regardless of the fact that almost all techniques predict SDB & benthic substrate classes, the RMS error for the BRUCE approach was the lowest. Liew, Chang, and Kwoh (2012) studied the accuracy and limits of WorldView-2 imagery in Singapore Strait to recover SDB in turbid muddy coastal water employing SA. Although the study discovered that the Red & Yellow bands proved beneficial for shallow water SDB, the maximum depth of SDB had only been around 2.4 m. Notwithstanding, the many SDB researches stated above, SA algorithms have the ability to yield depths of up to 15-20 m. Few studies expanded on SDB retrieval even when there was no field data. This advantage can help to ensure data creation at any location and over a long time.

2.4.5 SDB: Quasi Analytical

Lee, Carder, and Arnone (2002) proposed a multi-band QA technique to find total backscattering & absorption coefficients using associations between R_{rs} & IOPs of water-column by adapting RT theory & equation. Since the QA technique has analytical nature, it may be used for both data from MS & HS satellite sensors. The

research also advocated for the adoption of retrieval approaches for quality control that are as accurate as optimization and as quick to calculate as empirical procedures.

In the QA equation, $b_b(\lambda)$ is calculated according to the RT equation (Gordon and Morel, 1983) using R_{rs} with a reference wavelength of 555 nm. QA techniques have been used (Zhou, 2011) employing HS data of EO-1 satellite to predict bathymetry and sub-bottom benthic features by means of MS satellite data (Huang et al., 2017; Eugenio et al., 2015); comparison among few SA methods in SDB derivation (McKinna et al., 2015); and to prove SDB without any ground-truth bathymetry (Chen et al., 2019).

2.5 Summary

SDB has gained a lot of scientific community acclaim for its synoptic coverage and ability to gather data from inaccessible places. Although active sensors are utilised to predict SDB, passive sensors are the most common. We have primarily considered passive sensors in this study. For the past 50 years, optical SDB has been compared both within and between approaches. The comparison, however, was constrained by characteristics or inconsistent environmental factors that were peculiar to the site. Comparative analysis was hampered by the research area's varied depths, different sensor resolutions, and the dynamics of coastal IOPs, to name a few. We also highlighted the areas in the SDB domain where there are difficulties and evident knowledge gaps that need to be filled in. With the fluctuating distribution of influencing variables for creating operational products for coastal areas, the current boom in estimating SDB research is particularly concentrating on shallow region water depth derivation, which presents various complications.

Based on an examination of the literature in the field, a matrix of SDB technique selection is provided. Future researchers may find this matrix useful in deciding how to proceed with SDB research based on the needs of the study region. For experts in the field, a ready reckoner of required satellite imagery, ground-truth bathymetry, methods, and SDB algorithms as per the required accuracy level.

Table 2.2: SDB selection matrix

Depth region	Accuracy required	Satellite resolution	In-situ resolution	Preferred method	Preferred Algorithms	Refer Studies	Data Resources	
							Satellite	BathyData
Shallow Turbid < 30 m	L	L	L	• Empirical Methods	• Regression, • Band Ratio	• (Caballero et al., 2019; South Florida); • (Hernandez & Armstrong, 2016; Puerto Rico)	Low: MODIS, MERIS SENTINEL 3	Low: SBES GEBCO DCDB#
	M	M, H	M, H	• Semi-Empirical	• Classification • ML			
	H	-	-	-	• Not Achievable			
Shallow clear < 30 m	L	M	M	• Empirical Methods	• Regression • Band Ratio • Band Diff. • ML	• IHO Cookbook (Pe'eri, Azuiké, & Parrish, 2013; US, Nigeria, and Belize) • (Sagawa et al., 2019; Japan, USA, Puerto Rico, Japan, Vanuatu) • (Masita Dwi Manessa et al., 2017; Indonesia)	Medium: LANDSAT SENTINEL 2 SPOT, ASTER	Medium: LIDAR
	M	M	H	• Semi-Empirical	• Forward/ inverse modelling • Optimization			
	H	H	H*	• QAA • SA*				
Shallow turbid, Creek & Estuaries < 15 m	L	M	M	• Analytical Methods*	• ML if High resolution in-situ data available ANN, Deep Learning • Forward/inverse modelling	• SAA SWIM (McKinna et al., 2015, GBR, Australia) • SA Optimization (Lee et al., 1999, Florida) • (Dekker et al., 2011, Australia & Bahamas)	High: IKONOS QUICKBIRD WORLDVIEW RAPIDEYE CARTOSAT	High: MBES SSS Depth-Profiler
	M	H	H	• SA*				
	H	H	H*					
Shallow	L	M, H	No In-situ data	QAA	4SM Method by Favoretto et al., (2017), California; and Kerr & Purkis, (2018), 5 Sites in Caribbean for depth up to 15 m. RMSE 2-5 m			

L:Low (>300m), M:Medium (300-10m), H:High (<5m) ; * indicates Water Quality Parameters -Required; GEBCO (General Bathymetric Chart of Ocean) & DCDB (Data Centre for Digital Bathymetry) are open source bathy database of IHO. # For Indian Coast data Refer Sindhu et al., (2007).

The present research on SDB is concentrated on how well it can provide operational goods, with the goal of using it in channels, ports, harbours, etc. in real-time. Some space satellite interventions are presently discussing the contemporary limitations of best-suited equipment and sensors for the precise collection of data in shallow coastal region waters. The cloud-based platforms as well as web-based solutions are now accessible to get over the restrictions of big-data processing abilities with or without downloading or acquiring data and have the advantage of thinking outside the boundaries of traditional algorithms while using the hybrid approach. Given the dynamic nature and heterogeneity of water column characteristics, several constraints—such as chlorophyll, turbidity, and other variables a substantial issue that has gone unresolved in SDB research for several years. These limitations demand a new perspective employing more complex data and models. SDB investigators now have a plethora of possibilities to investigate how dynamic components in the water column of the sea relate to SDB.

To sum up the literature and arrive at a research statement, it can be inferred from the literature that;

- i) The application areas of SDB are significantly rather exponentially growing. At the micro level, it emphasises the importance of ports, harbours, and navigable channels wherein at the macro level projects like Seabed 2030 (Collaboration between the Nippon Foundation & GEBCO) aim to map the Earth's seabed by the year 2030. The commercial use of SDB in navigation, submarine cable laying for internet & telephony, marine economic zones and spatial planning, military activities, fisheries and aquaculture, commercial exploitation of corals & benthic habitats, global climate change, rising sea levels and depleting coasts are also demanding revolution in SDB methods.
- ii) SDB have the potential to contribute hugely to the coastal database but still a major challenge to SDB methods is dealing with turbidity and deriving SDB in Case-II waters of coastal regions.
- iii) The highly turbid nearshore areas are very dynamic in nature with regard to the content of sediment and various other water column properties. In order to provide a reasonable SDB accuracy, conventional SDB algorithms need to be replaced with AI-based solutions to include water column properties data in SDB derivation.
- iv) The SDB algorithms with reasonable accuracy need to be validated in different geographical areas having to wide differences in optical properties of water columns for operational use as a cost-effective alternate for hydrographic surveys.

CHAPTER 3: METHODOLOGY

This chapter elaborates on the methodology implemented for research in this study. The initial section specifies the materials and methods used in the study specifically, study sites for exploratory, development of model, and validation stages of research. Thereinafter, the satellite data, bathymetric data, & INCOIS-SATCORE data used in this study have been discussed. The remote sensing data needs to be pre-processed before using for further analysis, the same has been discussed with the flowchart of steps followed in pre-processing of data. The further section discusses the research methodology and procedures adopted in this research.

3.1 MATERIALS AND METHODS

This section describes all the geographical areas/sites of the study selected for exploratory and validation of research. The further section describes data used in the study viz. hydrographic survey data, satellite data, and INCOIS-SATCORE data. Also, pre-processing of data and conversion of data to usable form has been described.

3.1.1 Study Sites

Exploratory

Preliminary analysis for assessing the accuracy of the SDB model was undertaken in two different study sites. The sites were chosen for their complexity, underwater terrain as well as extremely turbid nature as river mouths and also a confluence of rivers.

Vengurla: The study site Vengurla is located at a height of 1.2 meters from Mean Sea Level, situated in the Sindhudurg district of Maharashtra, India. The coastline is made up of two steep cliffs that surround a beach that is embayed in the middle. A Casuarina-type plantation may be found beyond the relieved sand dunes that run parallel to the coast. There are mangrove plantations on both sides of the Karli River's creeks. There are semi-diurnal tides that range in height from 1.3 to 2.3 meters in this location. During the monsoon occurrence of high waves, the Vengurla shoreline is severely subject to seasonal erosion. A group of two big hillocks and almost eighteen significant and several smaller rocks make up the archipelago. This location has become dangerous for navigation because of the abundance of visible and submerged

rocks as well as impediments. Southwest of Vengurla, these rocks are 8.3 km in width and 14 km in length. Several kinds of seabirds have their ecological habitation in this small region (Mahabal et al., 2007). The location was marked on the nautical map as being fouled for anchorage and dangerous for seagoing boats to pass through. The fishing activities are carried out by local fishermen in the area, who are highly acquainted with navigating hazards, and they are able to traverse their fishing boats effectively enough to avoid these submerged rocks (Ashphaq, Srivastava, & Mitra, 2022). The region has been preferred for investigation because of its complication for steering hydrographic surveys, the existence of several underwater structures, and the very turbid character of the sea water as a result of the Karli River's outflow carrying deluge of sediments. Due to these complications, the location is perfect for evaluating SDB in the region. To suit the research region, the ROI was clipped. The study's encompassed area is 208.46 square kilometers, which is around Twelve kilometers from the shore.

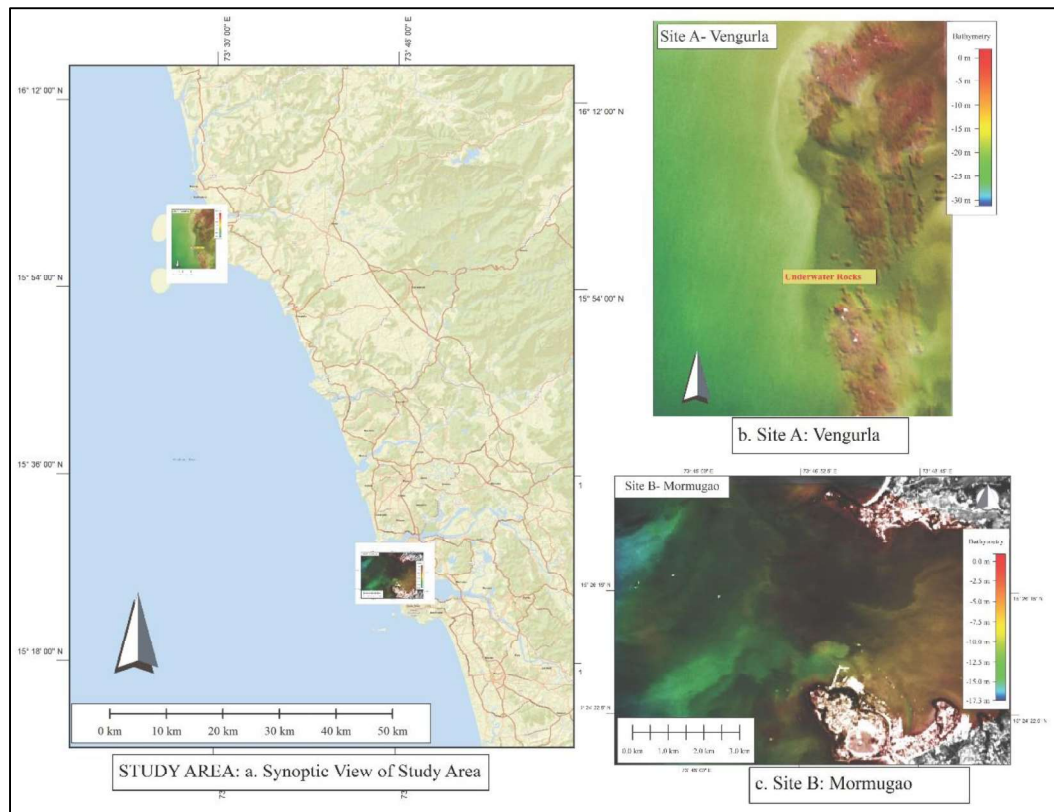


Figure 3.1 Study Area Vengurla & Mormugao

Mormugao: The study site Mormugao, it's a beautiful natural harbour and one of the notorious oldest ports on the Western Coast of India. Today it is one of the major

ports of India. This site of study is the confluence of the two big rivers, the Zuari and Mandovi Rivers. The research area is extremely north of the port's navigable channel, which is not included since dredging is conducted regularly and bathymetry may differ greatly from actual in-situ data. The bathymetric data collection using a hydrographic survey was carried out using an SBES with a frequency of 210 kHz. Total of 1241 depth samples have been used in the investigation, with training and testing data in the ratio of fifty percent over a depth range of 22 m for roughly 120 sq. km area. Figure 3.1 shows the overall coverage of the study site on the western coast of India.

INCOIS-SATCORE DATA Sites

INCOIS led SATCORE to the creation of novel bio-optical algorithms in Indian coastal waters and long-term observations of bio-physio-optical features for validating current OCM algorithms. Twelve time-series data stations for sampling transects throughout the east and west coasts of India were used by INCOIS to collect data for continual bio-optical parameter monitoring. These sites are shown below in Figure 3.2 and are also listed below;

- Junagadh Agricultural University (study site Off Okha)
- NIO, Goa (for site Off Goa), Goa University (study site Off Goa)
- Mangalore University (study area Off Mangalore)
- CIFT, Kochi (Off Kochi)
- Annamalai University (Off Parangipettai)
- IIT Madras (Off Chennai)
- Andhra University (Off Visakhapatnam),
- CSBoB, Andhra University (Off Visakhapatnam),
- Berhampur University (Off Gopalpur),
- Jadavpur University (Off Frazergunj),
- CARI (ICAR) (A & N Islands).

Among the above study sites, 4 sites were selected for the study **Kochi, Okha, Chennai, and Gopalpur** which were used for model development and testing the model.

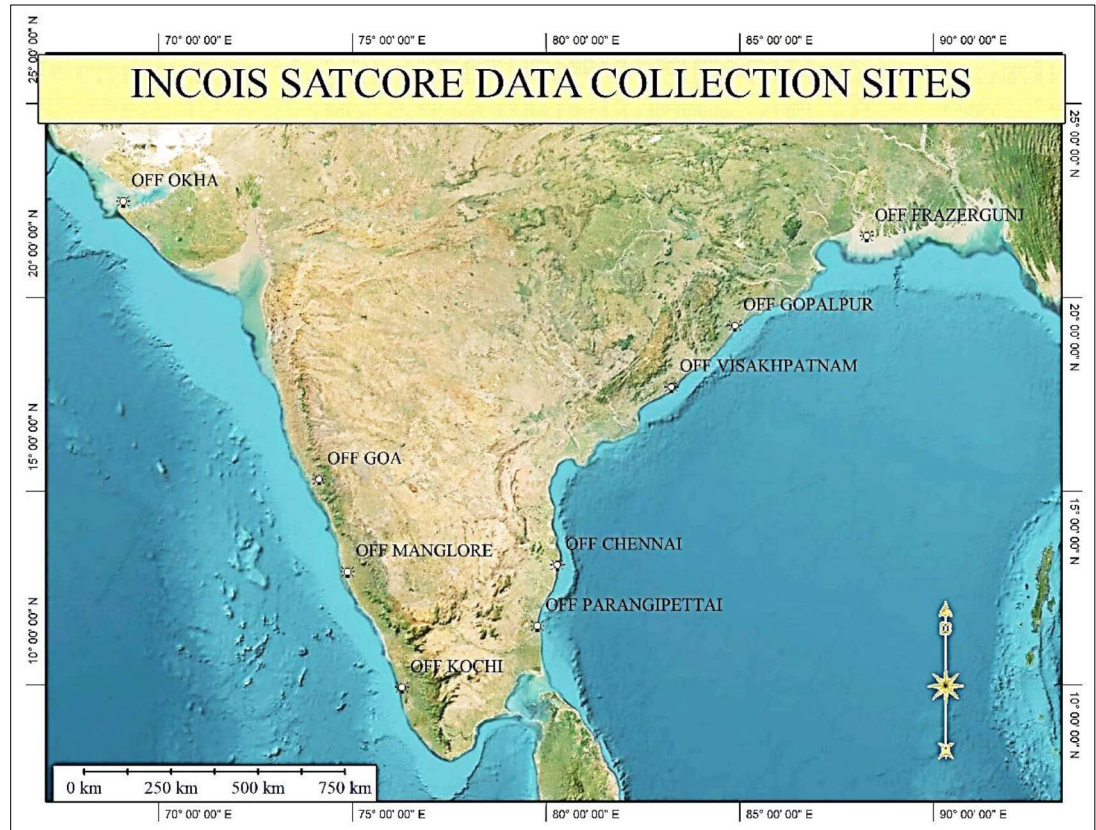


Figure 3.2 INCOIS SATCORE Data Collection Stations

Sites for Model Development

Kochi The study site Kochi is among the major Indian port and is situated along the southwest coast of Kerala state, India. This port has historical significance and is also known as the “**Queen of Arabian Sea**”. Kochi was selected due to the presence of a large influence of freshwater discharge, highly connected backwaters, and the occurrences of seasonal mudbanks (Minu, Lotliker, Shaju, & Santhoshkumar, 2014). Although the underwater terrain is smooth and depths increase gradually seaward, riverine OAS discharges along the coast often produce erroneous water constituents estimation from the satellite data (Chakraborty, Gupta, Lotliker, & Tilstone, 2016). The INCOIS-SATCORE data samples were collected at discrete stations off Kochi coastal waters from 2008 to 2016. However, this study has included data from only five days which are proximally near to satellite pass. The figure below shows the overview of the study area.

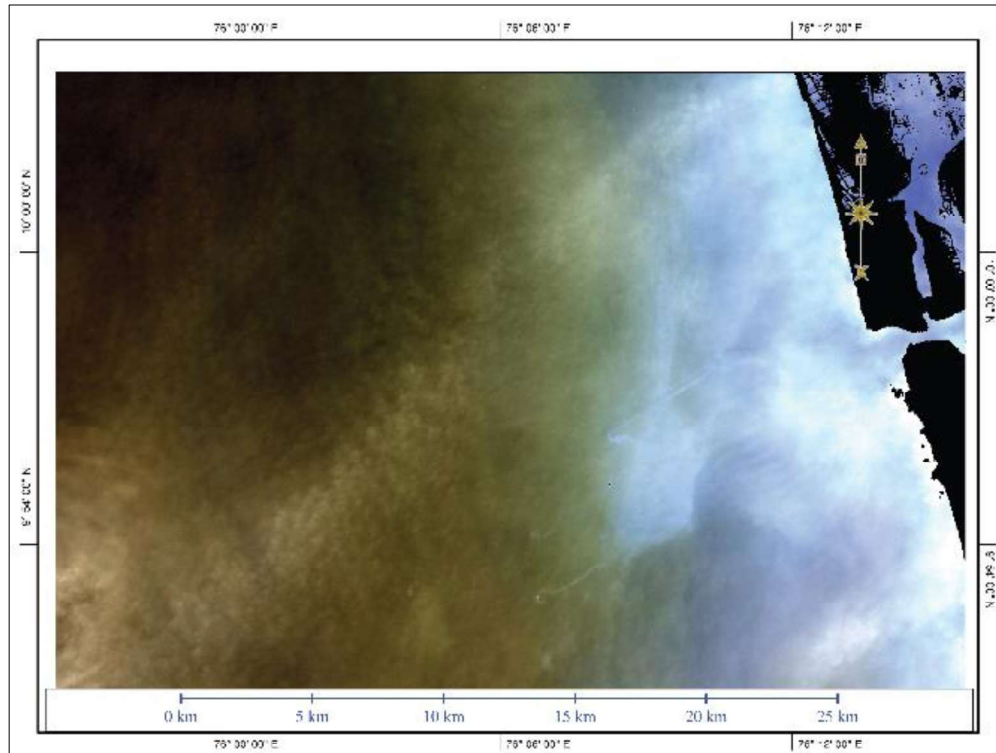


Figure 3.3 Study Area Kochi

Chennai: Chennai Port is India's second-largest cargo port. It is the third oldest port out of India's thirteen major ports as well as the biggest in the region of the Bay of Bengal. Chennai Port lies along the flat coastal strip recognized as the Coromandel Coast on the eastern coast of India. It slopes quite flatly at the bed. Due to its location along the shore and proximity to the thermal equator, the port doesn't experience large seasonal temperature variations. Because the tides are semi-diurnal, there are 2 high tides & 2 low tides each day across the port region. The covered area is more than 100 sq. km.

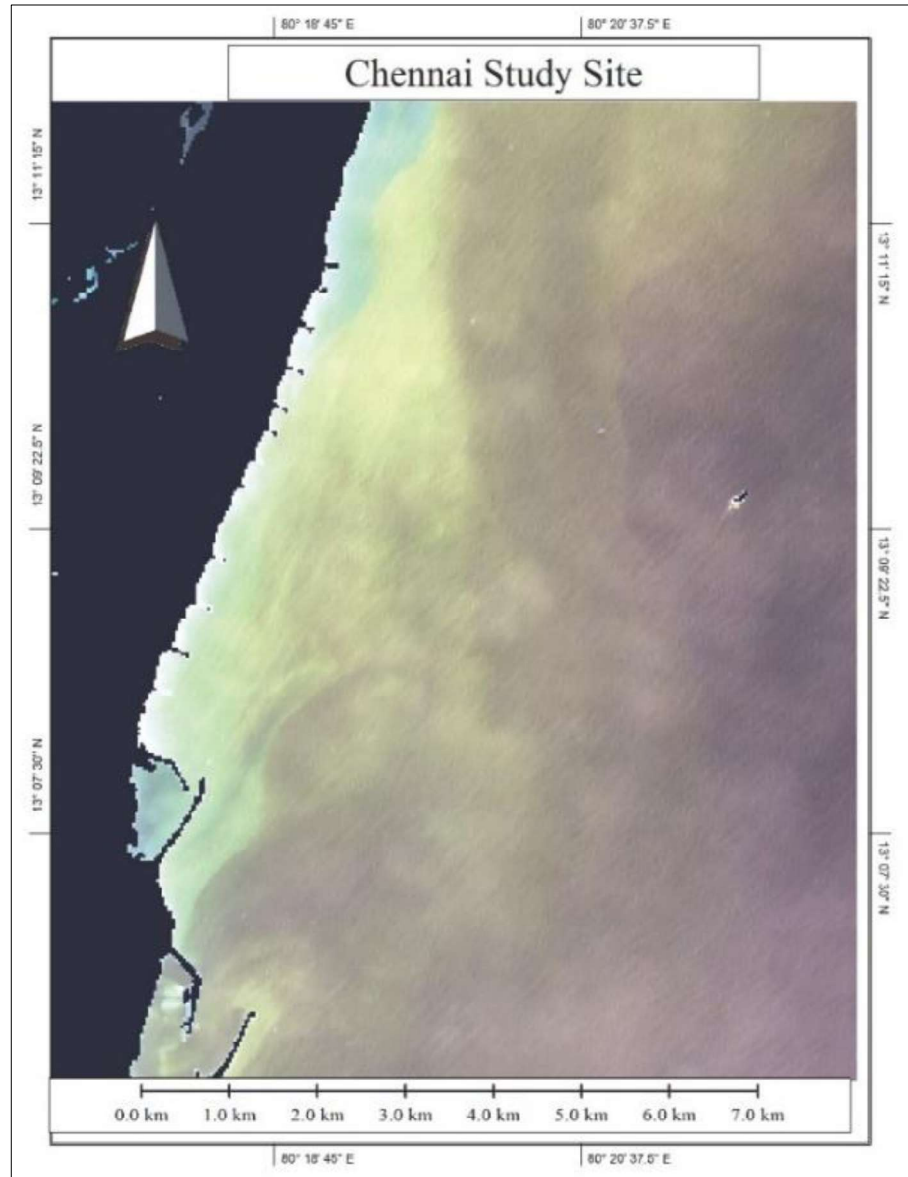


Figure 3.4 Study Area Chennai

Okha: OKHA is a medium port located opposite Bet Dwarka Island and on the approach to the boarding point for the ferry to Bet Dwarka at Okha Jetty. Okha is a busy and historically significant port in Gujarat that is ideally positioned. The covered area is having about 1030 sq. km coverage.

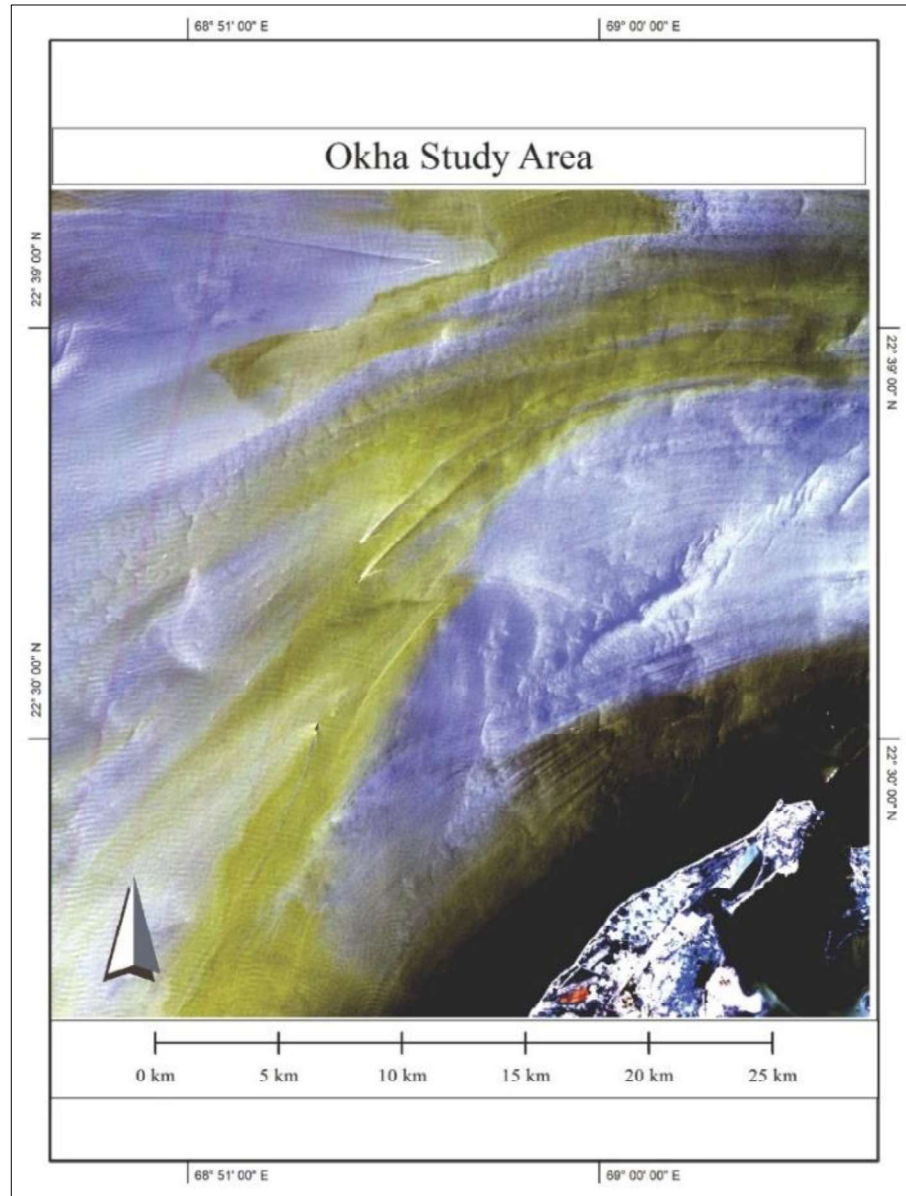


Figure 3.5 Study Area Okha

Gopalpur: Gopalpur port is the deep-sea port on the Bay of Bengal located in Gopalpur, Ganjam District, Odisha, India. The natural port harbor has a depth of 18.5m. Situated on the eastern coast, it is a natural deep-water port. Gopalpur is spread across on a 4 km stretch of desolate shore without any mangroves or tropical vegetation. The covered area is having coverage of about 1379 sq. km.

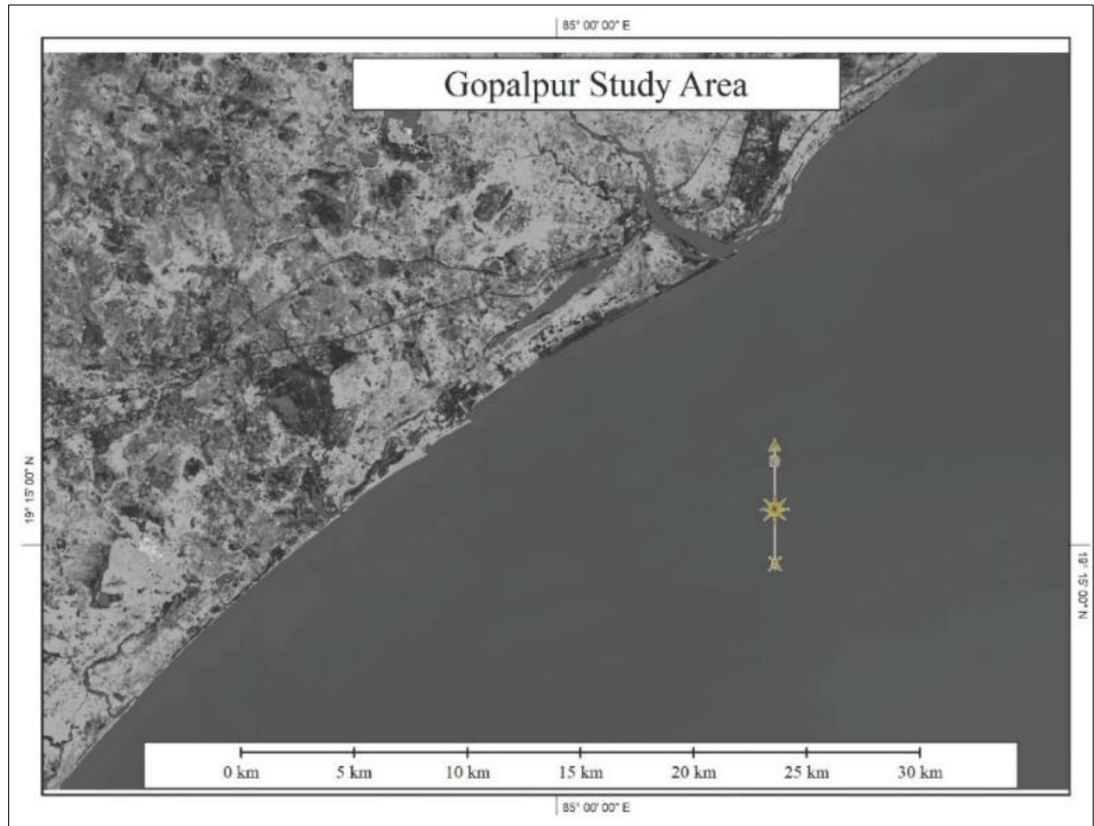


Figure 3.6 Study Area Gopalpur

3.1.2 Data

Hydrographic Survey Data: The cliffy submerged hills and underwater hazards in Vengurla rocks presented a complicated and difficult region for conducting hydrographic surveys in order to assure survey boat safety, instrument safety, and safety of hull-mounted transponders. Therefore, the site was selected for the exploratory part of the study along with another site Mormugao which is a river confluence of two major rivers and thus highly turbid in nature. Hydrographic data was collected utilizing a bathymetric survey in the region in 2018 using an SBES Deso-30 which propagates sound at 210 kHz. In the undersea rocky area, Edgetech, 4200 FSL, Side Scanner sonar was engaged to locate shoals and impediments for side-scanning. According to the order of the survey, the precision of vertical accuracy reached is 0.05 percent of maximum depth for depths ranging from 0 to 32m. The echo sounder instruments were calibrated using bar-check before and after the survey. The 'Differential Global Positioning System' achieves horizontal precision of 95 percent confidence level, (i.e. 5 m Plus 5 percent of water depth). Although the

hydrographic survey included huge depth data, data samples were chosen based on a 3% error margin as well as a 95% confidence level. The scikit-learn library's train test split function was employed to generate a randomized sample dataset.

For all other sites used for the development and validation of the model respective navigational charts were utilized to derive bathymetric data. All the largest scale charts updated with the latest hydrographic surveys by National Hydrographic Office for the study region were used. Although, bathymetric data was sparse, but was large enough for the development of the model.

Remote Sensing Data: The information on the satellite images is shown in Table 3.1 follows. The level-1 pre-processed data from the OLI sensor of Landsat-8, sensor MSI of Sentinel-2, as well as data of ASTER sensor onboard Terra satellite, images have been radiometrically corrected as well as orthorectified was downloaded from the USGS Earth Explorer application. All the above three sensor satellite datasets are publicly accessible which deliver adequate spatial resolution necessary for SDB research. The Landsat-8 instrument has a spatial resolution of 30 meters (15-meter Panchromatic) and a re-visit period of 16 days with 185 km swaths. The sentinel-2 is gathering data at 10 metre spatial-resolution with a re-visit duration of 05 days and a swath of 290 km. The ASTER has a spatial-resolution of 15 meters, and a temporal-resolution of 16 days, for a swath width of 60 km. Except for ASTER, the satellite imagery data was retrieved for the hydrographic survey's proximal date. The table below summarises the satellite image utilized in the research.

Table 3.1: Satellite Data in the Study

Band	Band name	Wavelength λ	Central λ	Bandwidth	Spatial Resolution in m
OLI/LANDSAT-8					
2	Blue	450–515	482	60	30
3	Green	525–600	561	57	30
4	Red	630–680	654	37	30
5	NIR	845–885	864	28	30
MSI/ Sentinel-2					
2	Blue	448–546	490	65	10
3	Green	538–583	560	35	10
4	Red	646–684	665	30	10
8	NIR	763–908	842	115	10
ASTER/Terra					
1	Green	520–600	556	9	15
2	Red	630–690	661	6	15
3N	NIR	780–860	807	10	15

Most SDB research has employed blue-green bands for research on SDB out of several many other spectral bands (Gao, 2009). The present study attempts to validate frequently used SDB techniques in the Vengurla Rocks' complicated and turbid water. The purpose is to use appropriate methods to get the optical satellite data image to correlate to depth points.

Satellite datasets were selected by their geographical coverage of the research region, cloud covering 10 percent below, and with temporal approximation to the in-situ field investigation. Satellite images had been processed to a radiometric resolution of 12 bits, which provides enough variation in remote sensing reflectance to predict SDB. ASTER/Terra imagery has an 8-bit radiometric resolution, which provides substantially less fluctuation in a dataset, particularly in the water area. ASTER/Tera dataset does not have a blue band, therefore only the green plus red bands were used. Because the green band's transect profile has greater fluctuation in reflectance than other bands including the red band, only the green band is used for the SDB in log-linear transform for ASTER.

INCOIS-SATCORE Data: 'Indian National Centre for Ocean Information System (INCOIS)' project 'SATellite Coastal and Oceanographic Research (SATCORE)', established eight different time-series stations in the Indian coastal water, for measurement of the bio-optical and physio-optical, and chemical parameters. All stations are geared with sophisticated Instruments like Integrating Sphere, Spectrophotometer, Sun-photometer, Weighing Balance, Fluorometer, Automatic Weather Station, Aspirator Pump, and Vacuum Filtration Unit, (INCOIS Annual Report, 2016). These instruments have collected time-series data for various bio-optical parameters, a few of which have been studied extensively in marine remote sensing studies. The SATCORE data includes Time-series data on TSM, Total Chlorophyll, Particle size (micro, nano, and pico), Temperature, Salinity, Alkalinity, Dissolved Oxygen, pH, Turbidity, Dissolved Fluorescence, and other chemical composition (like Nitrite, Nitrate, Phosphate, Total Phosphorous Silicate, Ammonium, etc.). The dataset was provided by the INCOIS-SATCORE project coordinator in .xlsx format. The dataset was having information collected from the

year 2008 to 2017 the for above stations. The missing data was filtered prior pre-processing of data.

3.1.3 Pre-processing

Top-of-Atmosphere radiance data is processed to compute the reflectance. The metadata file contains the vital parameters for reflectance calculation: i) Earth-Sun distance (in astronomical units), ii) at-satellite radiance, iii) mean solar exo-atmospheric irradiances, iv) sun elevation in degrees, and v) solar zenith angle. To obtain underwater reflectance and sub-bottom reflectance in shallow water, water column correction is also used in the computation of SDB (Ashphaq, Srivastava, & Mitra, 2022). Surface reflectance is converted from scaled TOA reflectance in Landsat imagery. Luca Congedo developed a Semi-Automatic categorization plug-in (SCP) for QGIS that has automated techniques for downloading, processing, and translating radiance data to surface reflectance data for different satellite images (Congedo, 2019). The ASTER imagery has been pre-processed using the QGIS SCP plugin. ACOLITE ‘Atmospheric Correction for OLI lite’, is an RT-based atmospheric correction tool that applies a standard NIR-SWIR technique for atmospheric correction, (Vanhellemont & Ruddick 2015). This tool is also extended to pre-process Sentinel-2 and Landsat-8 data. This technique assumes 0% SWIR reflectance and obtains water-leaving Rrs in the visible and NIR bands in coastal waters and inland seas (Vanhellemont, 2019).

ACOLITE performs well in coastal applications developed using Landsat-8 and Sentinel-2 (Salameh et al., 2019; Caballero & Stumpf, 2020). The DOS atmospheric correction has been used as it eliminates the impacts of gas particles in the atmosphere and also aerosol to determine BOA values of reflectance using an image-based approach. The DOS technique suggests that the black pixel in the image has no reflected light because of air dispersion. By taking the values of the dark pixels out of each pixel in the band images, the effect of air scattering is lessened. The derived surface reflectance images were masked in the land area using QGIS. The European Space Agency’s SNAP Desktop has been used to extract the reflectance values for the corresponding depth location from the corresponding pixel of the image. The data from all the satellite images were exported to tabular .csv format for

further processing in Jupyter Notebook using Python. Figure 3.7 below shows the steps involved in data preparation.

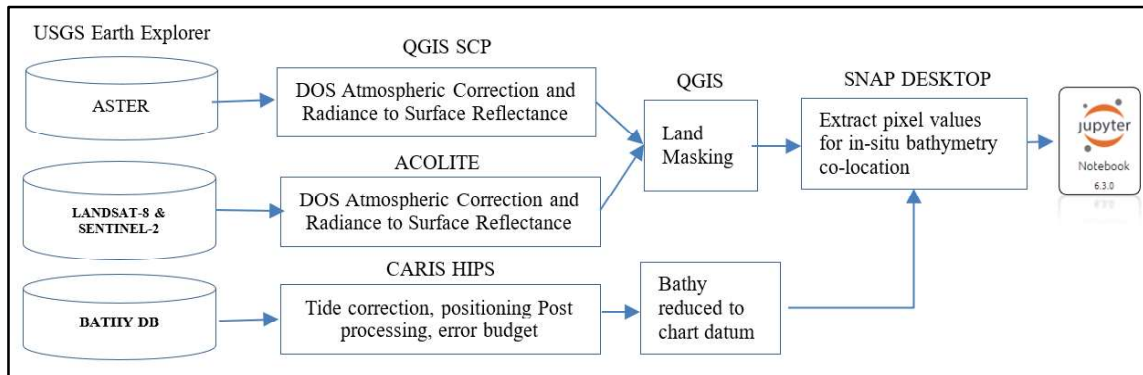


Figure 3.7: The Flowchart of pre-processing of data for exploratory study

3.1.4 Performance Evaluation

By analyzing the SDB and regressing the SDB dataset against the validation dataset, the perfection of the derived models was determined. The test bathymetric dataset was utilized for validation as well as for deriving statistical parameter values. For each technique, i) coefficient of determination, R^2 , ii) the Root Mean Square Error, RMSE, and iii) Mean Absolute Error, MAE were determined using the approach indicated in equations 1 and 2 below.

$$RMSE = \sum_{i=1}^n \sqrt{(x_i - y_i)^2 / n} \quad \dots\dots(1)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_0 - y_0| \dots\dots\dots(2)$$

Where, X_i is Satellite imagery-derived SDB, and X_0 is the average depth of SDBs

Y_i is the depth of in-situ data for validation, and y_0 is the average of in-situ depth

n is total depth of data points

A regression model's effectiveness can be quantified using R^2 , by comparing the predicted model to a consistent baseline by using data mean and thus, it indicates how

much superior the projected model is. R^2 values will always have a value less than or equal to 1.

The R^2 Formula is shown below in Equation 3,

$$R^2 = 1 - \frac{MSE(model)}{MSE(baseline)} \dots(3)$$

3.2 RESEARCH PROCESS

The research being exploratory in nature was undertaken in a phase-wise manner consisting of various stages as shown below in Figure 3.8. Each stage was feed-forward to the next stage and feedback was to address any shortcomings in the previous stage.

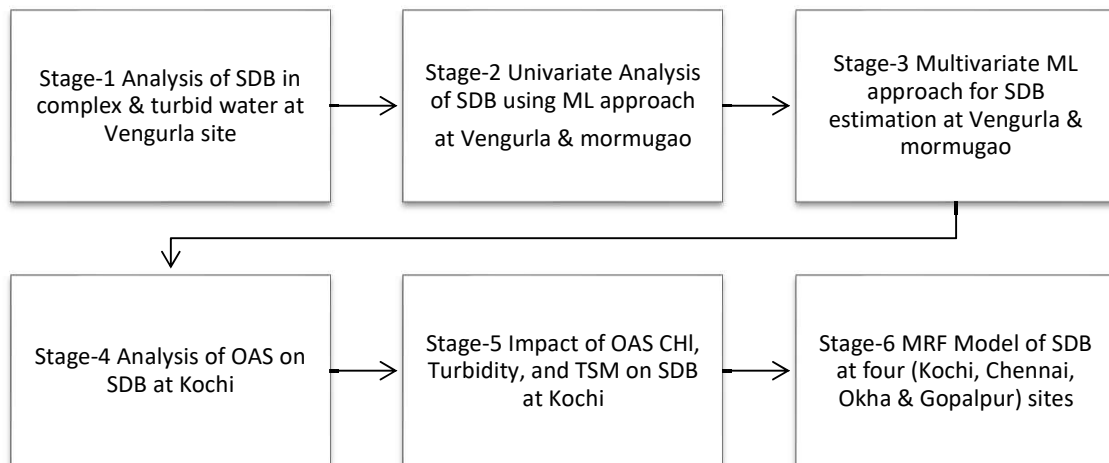


Figure 3.8: The Flowchart of various stages of the Research Process.

Each stage of research is discussed below with a separate flowchart of its execution. Stage one dealing with a comparison between conventional, semi-automatic, and ML algorithms was executed at the Vengurla site only; stage two (Univariate study) & stage three (Multivariate study) were focused on evaluating the most appropriate ML algorithm for SDB in coastal turbid water (employed at both Vengurla and Mormugao site); the stages four & five (employed at site Kochi only) were dealing with assessing influence of WQP for increasing accuracy of SDB algorithm as well as assessing the

impact of OAS with the help of INCOIS-SATCORE data. Wherein, stage six was to develop and validate the SDB model executed on all four study sites.

3.2.1 SDB Algorithms in Turbid Coastal Water

In this step, fully accessible free satellite data from the OLI instrument of satellite Landsat-8, the MSI onboard twin satellites Sentinel-2 (A and B), and ‘Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) onboard satellite Terra were analyzed to extract bathymetry in the navigably extremely compound and very muddy water of the Vengurla. The goal of the study was to assess the most effective method and dataset for deriving SDB in the area. A few regularly deployed SDB techniques were used for SDB derivation in order to achieve the goal of the study. SDB algorithms applied include Stumpf et al., (2003) developed Log-Ratio Algorithm; Lyzenga et al., (2006) Log-Linear Model; The relative bathymetry model utilizing the Log Ratio Algorithm is part of the SPEAR software suite, which is part of Envi 5.3. Additionally, ML methods SVR, LR, and RF Regression have been assessed to produce SDB. The results of processing as per depicted flowchart in Figure 3.9 is shown in section 4.1 of Chapter 4.

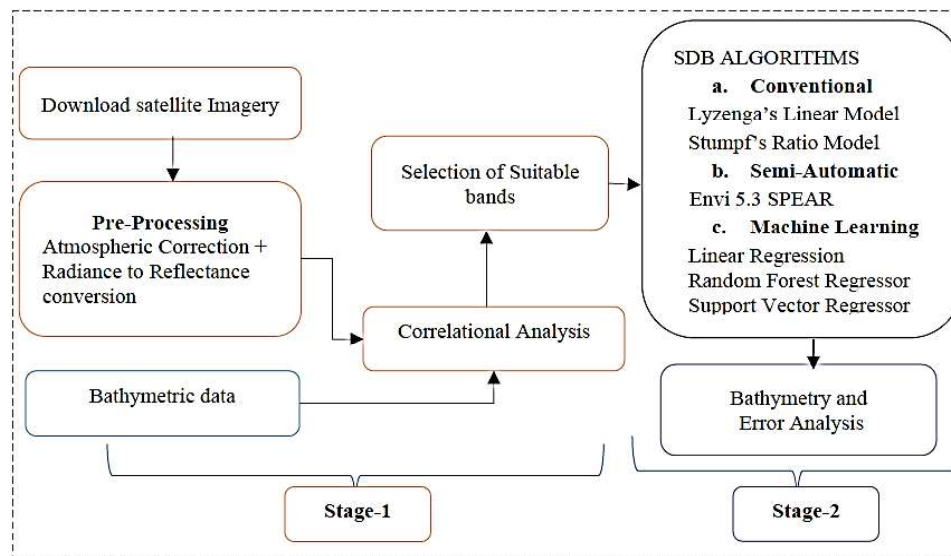


Figure 3.9 Flowchart of exploratory study to compare SDB

The ML syntax used in Python language codes implemented in Jupyter Notebook for LR, RF, and SVR is attached for further reference.

Refer Syntax-1 from Appendix-III :- Linear ML syntax

Refer Syntax-2 from Appendix-III :- RF ML syntax

Refer Syntax-3 from Appendix-III :- SVR ML syntax

3.2.2 Univariate ML Algorithms for SDB

Based on the inputs of the above step where ML proves superiority over conventional algorithms, this phase evaluates the effectiveness of a few uni-variate linear, non-linear, and robust ML regression methods in SDB estimation utilizing better resolution in-situ water depth data from two sites with complicated effluents and the spectral bands of SENTINEL-2A, ASTER, and LANDSAT-8. According to the SDB literature, where the majority of the study to date has shown the practicality of 0.5-0.6 m different wavelengths for SDB extraction in aquatic environments, the uni-variate method has been used (Stumpf et al., 2003). In addition to evaluating the empirical connection between Rrs and the appropriate satellite bands, examined depth-dependent Rrs of spectral bands. The linear ML, three different robust ML linear regressions, and a non-linear GPR ML algorithm were utilized for the relative analysis of SDB since both research sites are extremely unequal and complicated. The result of processing as depicted in the flowchart in Figure 3.10 has been shown in section 4.2 of Chapter 4.

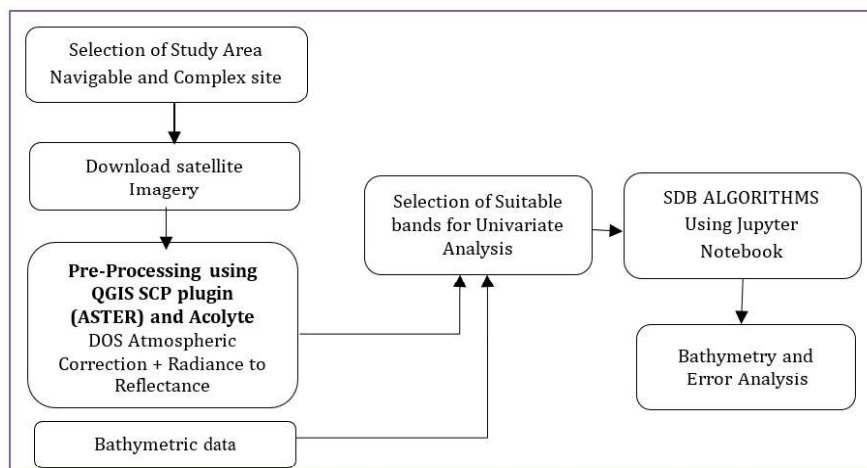


Figure 3.10: The Flowchart of Univariate SDB ML algorithms

The ML syntax used in Python language codes implemented in Jupyter Notebook for the following algorithms is attached for further reference.

Refer Syntax-4, Appendix-III :- Robust-regression Theil Sen regression

Refer Syntax-5, Appendix-III :- Robust-regression -Huber regression

Refer Syntax-6, Appendix-III :- Robust-regression -Random Sample Consensus

Refer to Syntax-7 from Appendix-III:- Compare Robust Regression Algorithms

3.2.3 Multivariate ML Algorithms for SDB.

In the previous stage, it was found that univariate Non-linear ML algorithms perform better than linear and robust-linear algorithms. Multispectral equipment mounted on aircraft driven with 18-channels in the spectral bandwidth of 0.3 to 15 μ , with 12 bands in the visible EMR spectrum (Polcyn & Rollin, 1969). Since then, various researchers have discovered that the bandwidth 0.55 to 0.58 μ is better for shallow water observation, the red region 0.62 to 0.66 μ for the shallowest depths, and in the blue region 0.40 to 0.44 μ (Ashphaq, Srivastava, & Mitra, 2021). However, bands which have lower light scattering than other bands are useful in SDB estimation. The notion of optically uniform conditions in a single satellite scene allows the utilization of all spectral bands in the study. Several studies have used multiple bands to build empirical relationships (Moeinkhah, Shakiba, & Azarakhsh, 2019; Salameh et al., 2019). Multivariate Regression is a supervised ML approach that analyses several data variables. One outcome variable and several independent variables make up a multivariate regression, which is a generalization of multiple regression. It makes an effort to anticipate the outcome utilizing the input of different independent factors. Causal relationships between input Remote sensing reflectance and output bathymetry are automatically learned by the ML algorithms.

In this step, we employed multivariate ML regression techniques to assess the efficacy of SDB techniques. Because the research sites are so complicated and unequal, non-linear, linear, and decision tree regression are utilized to compare SDB results. The aim of this study is to validate ML regression algorithms for the pragmatic derivation of SDB and to examine the best-suited technique for predicting SDB for multivariate dependent variables. This stage examines the performance of multivariate ML regression techniques MLR, GPR, MARS, DTR, DTR-K, and RF at two different sites, using high resolution ground-truth bathymetric data, spectral bands from LANDSAT-8, SENTINEL-2, and ASTER.

After pre-processing satellite imagery with the DOS method to remove the influence of the atmosphere, and changing from radiance to surface reflectance, the dataset has been converted into a tabular format for further analysis using the SNAP desktop tool. The ML algorithm processing was performed in Jupyter Notebook to determine the best suitable technique for SDB derivation. This was followed by performance evaluation, error analysis, and creating of a final SDB map from the best result achieved in the study. Figure 3.11 below depicts the flowchart for the applied methodology and the result of the same has been discussed in section 4.3 of Chapter 4.

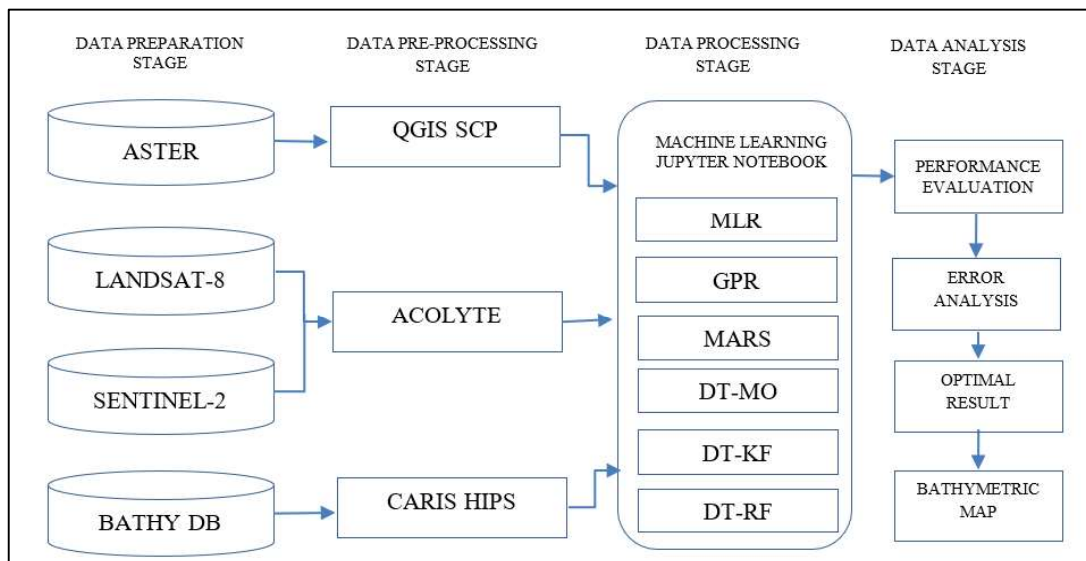


Figure 3.11: Flowchart of the Multivariate SDB ML algorithms

The various machine learning packages available in Anaconda Navigator's Jupyter Notebook-Python (using scikit learn libraries) are used to implement the ML in this study. ML regression approaches are highly influenced by the sample distribution used in training and testing.

The effectiveness of the ML is significantly impacted by the sample distribution. This is especially important at difficult research locations (like Vengurla Rocks), where bathymetry is scattered extremely unevenly geographically and the accuracy of the prediction depends heavily on the spread of the training dataset. To create a more reliable SDB model, the predicted accuracy in such a complicated

region also needs a high number of sample sizes. All datasets were randomly divided by using the scikit learn library into 50 percent for model training and 50 percent for model validation to verify the effectiveness of the ML algorithms. Table 3.2 below shows the algorithms utilized for SDB are given;

Table 3.2: Multivariate ML Algorithms

Classification	Method	Libraries	Estimation
Multivariate Regression	Multiple Linear Regression MLR	sklearn.linear_model LinearRegressor	r ² , RMSE, MAE, Scatter- plot, and density Distribution plot
Non-linear Multivariate Regression	Gaussian Process Regression GPR	sklearn.gaussian_process GaussianProcessRegressor	
	Support Vector Regression SVR	from sklearn.svm SVR	
Decision Tree	K-Neighbor Regression KNR	from sklearn.neighbors KNeighborsRegressor	
	Decision Tree Regression DT-MO	sklearn.tree DecisionTreeRegressor	
	Decision Tree Regression using k-fold cross-validation DT-KF	sklearn.tree DecisionTreeRegressor RepeatedKFold -10	
	Random Forest Decision tree DT-RF	sklearn.ensemble RandomForestRegressor	

The ML syntax used in Python language codes implemented in Jupyter Notebook for the following algorithms is attached for further reference.

Refer Syntax-8 from Appendix-III :- Multivariate RF

Refer Syntax-9 from Appendix-III :- Multivariate MARS

Refer Syntax-10 from Appendix-III :- Multivariate MLR

Refer Syntax-11 from Appendix-III :- Multivariate GPR

**Refer to Syntax-12 from Appendix-III:- Decision tree regression using k-fold
cross-
Validation**

**Refer to Syntax-13 from Appendix-III:- Multivariate Decision Tree for
Multioutput Regression**

3.2.4 Optically Active Substance & Bathymetry

Based on their contributions to Rrs, the water constituents have been divided into OAS (such as chlorophyll-a (unit-mg/L), TSM (unit-mg/L), CDOM (unit-mg/L), Turbidity (unit-NTU), Secchi Disk Depth (unit-m), Temperature (unit-°C), Total

Organic Carbon (unit-mg/L), Sea Surface Salinity (unit-PSU), Electrical Conductivity (unit- $\mu\text{s}/\text{cm}$) and optically in-active substances (like Dissolved Organic Carbon Total Phosphorus, Phosphate, Chemical Oxygen Demand, Biochemical Oxygen Demand, Nitrogen, etc.) (Gholizadeh et al., 2016). In accordance with environmental conditions and physical restraints, such as depth, beyond which these parameters cannot be recovered, OAS enables retrieval of water column data from remote sensing utilising the spectral properties of each parameter.

At this stage, we used longitudinal coastal seawater data, remote sensing reflectance data, and bathymetry data to examine the relationship between them. The study has been carried out in two phases; i) deals with pre-processing of data, and ii) analyzes the relationship between bathymetry with each SATCORE parameter distribution in the study site Kochi only. The regression model has been used to analyze the causal relationship of dependent variables i.e. ‘bathymetry’ and independent variables ‘remote sensing reflectance’, and OAS.

The objective of this step was to identify the influence and utility of INCOIS-SATCORE data in further processing and model development. This involved rigorous processing of longitudinal data to infer the result. Being a very lengthy and time-consuming execution of data processing, the processing of data was limited to only site Kochi. The selection of the Kochi site was for the reason of the availability of time synchronous INCOIS_SATCORE & Remote Sensing data.

3.2.4.1: Pre-processing of SATCORE Data

The INCOIS-SATCORE data was in .csv format collected for 10 discrete geo-locations points evenly distributed in the study area. The terrain in Kochi is very smooth and bathymetry gradually increases seawards. The influx of sediment in the channel is mapped by 03 samples at regular intervals (0m, 5m, & 10m) within the Kochi channel. The data were converted to raster images by using the ‘Inverse Distance Weighted (IDW) tool in QGIS to interpolate. IDW functions are useful to map the range of surface variation using a set of point data. A linear weight for a set of sample points is determined for calculating the raster cell values as a function of the distance between the input points and the output cell such that the more the distance, the less is influence on the output value (Morales, Stuart, Platt, &

Sathyendranath, 2011). The European Space Agency’s SNAP Desktop has been used to extract the reflectance values and OAS for the corresponding depth location from the corresponding pixel of the image. The data from all the satellite images were exported to tabular .csv format for further processing. Figure 3.12 below shows the steps involved in data preparation.

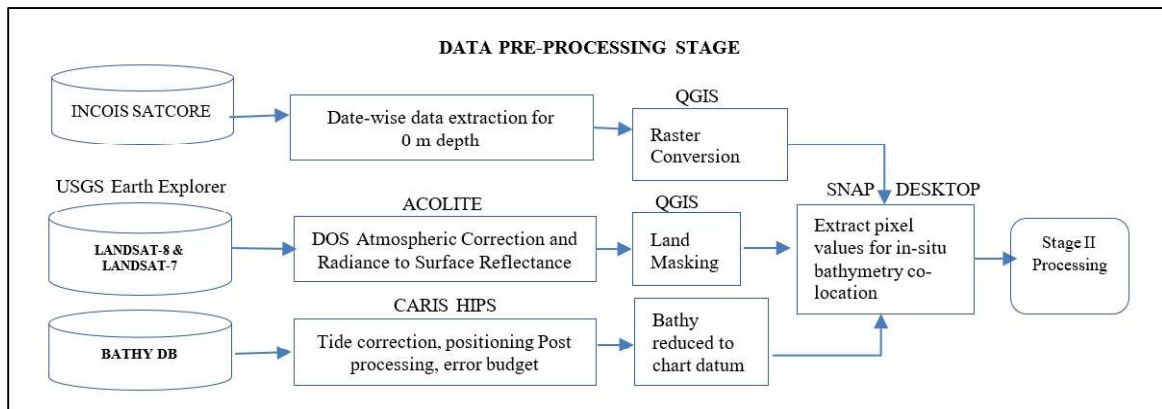


Figure 3.12: The Flowchart of Pre-processing with INCOIS-SATCORE data

3.2.4.2 Bathymetry-SATCORE Data Analysis

In the previous stage, the data has been converted into a tabular .csv format for further analysis using the SNAP desktop tool as an output of pre-processed data. In this stage, the statistical parameters were analyzed creating a descriptive statistical summary of data. The SATCORE data was collected in different units, therefore data were standardized to a scale of 0-1. The correlation matrix for each dataset was created to check the relationship among variables in each event. Each of the SATCORE parameters and each spectral band of satellite data was regressed against bathymetry to assess the bathymetry-dependant spatial variability and distribution of OAS as well as spectral bandwidth data using IBM SPSS Statistical software.

The multiple regression analysis with all the parameters was carried out for each event of SATCORE and all the important sub-set parameters contributing to the increasing SDB accuracy were identified using Minitab 18 software Regression > Best Subsets function. The R^2 (higher the value, the better the model and is always between 0-100%), Standard Error (standard deviation measured in response variable units, if S is lower indicating the better the model), Mallows' Cp criterion were used to analyze a different set of multi-variate regression model. An indicator called Mallows' Cp indicates the comparison of the accuracy and bias of the complete model

to models that include only some of the predictions, thus helping reduce the number of predictors of the model. Figure 3.13 below shows the flowchart of the methodology of this second stage. The result of the processing as per the flowchart below has been presented in section 4.4. of Chapter 4.

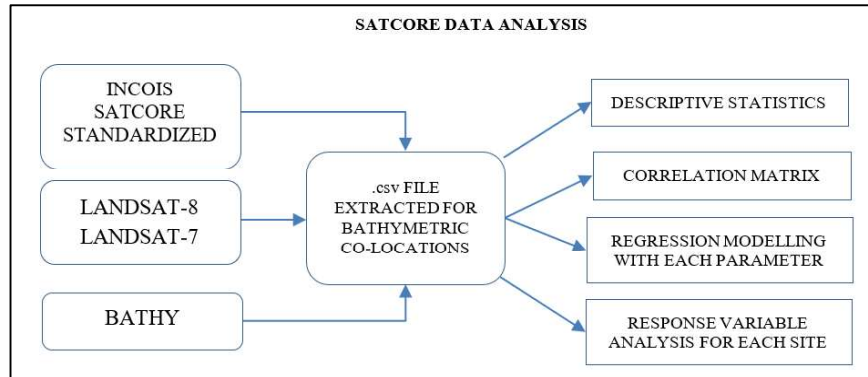


Figure 3.13: Flowchart of the SATCORE data analysis.

3.2.5 Influence of Chlorophyll, TSM, & Turbidity on SDB

The result of the previous stage shows that Bathymetry is only influenced by the spatial distribution of Total Suspended Material (TSM), Chlorophyll (Chl), and Turbidity in coastal water. Therefore, the goal of this stage was to improve the accuracy of SDB by understanding the spatial distribution of Total Suspended Material (TSM), Chlorophyll (Chl), and Turbidity in coastal water. A preliminary investigation using the numerical analysis between Landsat-7 & 8 spectral bands, OAS parameters, and bathymetry was carried out in this stage. SDB has been derived using three ML; Linear, RF, and SV Regression. Resulting SDB residuals have been analyzed in relation to Chlorophyll, Turbidity, and TSM in coastal seawater. The study found that most of the erroneous SDB residuals and extreme values were distributed in high or medium-concentrated OAS regions. The result of the study may further enhance SDB estimation accuracy by considering the influence of the above three OAS in coastal waters.

Table 3.3: SATCORE & Satellite Data

SATCORE Date	SATCORE Parameters	Satellite data	Sun Elevation	Earth-sun distance-AU
13 Feb 2015	CHL, Diss_Oxy, Salinity, Temp, TSM, Turbidity, Fdom, Ammonium Nitrite, Phosphate, PII, Silicate, Aot500, Nano, Pico	L8_OLI_2015_02_14_05_11_28_144053_L2R	52.98206259	0.9874698
* 16 Dec 2015	Chl, Diss_Oxy, Salinity, Temp, TSM, Turbidity, Fdom, Ammonium, Nitrite, Phosphate, Silicate, Nano Pico	L7_ETM_2015_12_23_05_12_57_144053_L2R	52.98206259	0.9874698
Units for SATCORE data: Total Chl- mg/m**3, TSM-mg/l, Particle size micro(20µm), nano(2µm), pico(0.2µm)-mg/m**3, Salinity-PSU, Dissolved Oxygen- mg/l, Turbidity-NTU, Dissolved Fluorescence (Fdom)- QSE, Nitrite, Nitrate, Phosphate, Silicate, Ammonium- mol/l,				
* Only used for SDB cross-validation as other meta information is almost equivalent for two Landsat products				

3.2.5.1 Cross Validation

This stage deals with the estimation of SDB from satellite data using Landsat-7 & 8 spectral bands. The bathymetric data were also transformed to a raster of 30-m spatial resolution to match Landsat data using the IDW interpolation method. Three different machine learning algorithms were applied for SDB estimation; LR, RF Regression, and SV Regression using QGIS ENMAP-Box plugin (EnMAP-Box, 2019). EnMAP-Box is a QGIS plugin providing a graphical user interface-based visualization and processing framework for vector and raster data. This processing framework has commonly used classification and regression algorithms in remote sensing and photogrammetry. Each algorithm was developed for an individual dataset of a particular scene, and then it was cross-validated to both scenes to estimate SDB and thus create a matrix of the result. Finally, SDB residuals were analyzed for spatial variability and distribution of OAS by converting Chl, TSM, and Turbidity into categorical variables. The detailed procedures adopted for Stage II are elaborated below in Figure 3.14.

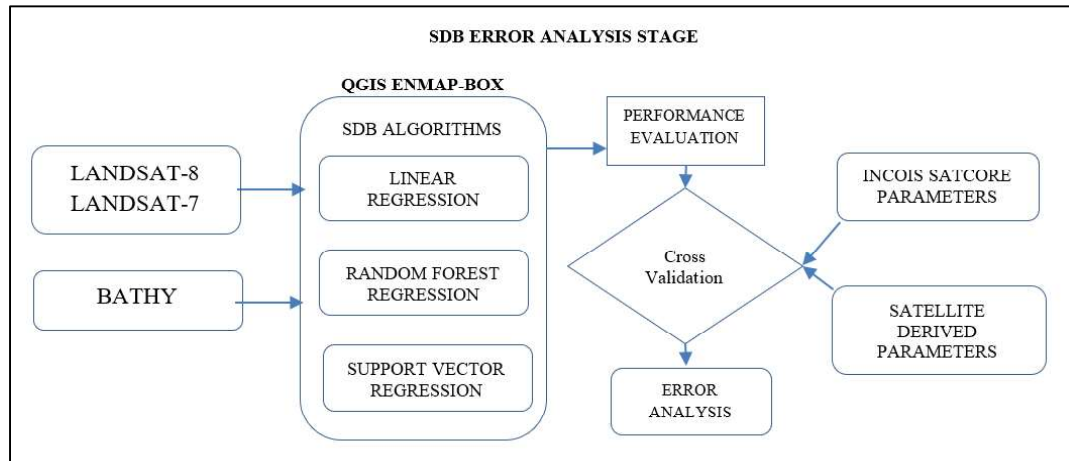


Figure 3.14: Flowchart for SDB error analysis

3.2.5.2 Validation of Satellite-Derived Chl, TSM, and Turbidity

The study aims for SDB analysis in highly turbid areas, as the higher level of sediments causes stronger backscattering. The time gap between satellite pass and SATCORE data acquisition was a major constraint to overcome in this study. To overcome this, retrieval of these parameters (Chl, TSM, and Turbidity) from satellite data using different regional and global algorithms and validating them against SATCORE data was a feasible solution. To achieve this aim the literature on retrieval of TSM, Chl, and Turbidity was examined. Saberioona et al., (2020) developed and tested more than ten algorithms for Chl and TSM prediction using Sentinel-2A satellite imagery in various water reservoirs of the Czech Republic, Central Europe and achieved an accuracy of predictive accuracy for both Chl (R^2 -0.85) and TSM (R^2 0.80). Wang et al., (2017) developed and tested a few algorithms for TSM and Chl on-field hyperspectral data collected in irrigation ponds with significant accuracy. Wang et al., (2017) used Landsat data for modelling the TSM concentration of highly turbid estuaries along with coastal China and tested more than 13 TSM algorithms from previous studies. Avinash et al., (2012) used IRS(P4) ‘Ocean Colour Monitor’ (OCM) dataset to develop a regional algorithm for TSM in coastal Karnataka with reasonable accuracy. Dogliotti et al., (2015) provided a single algorithm to predict turbidity in all coastal regions using a semi-empirical single band ($0.859 \mu\text{m} / 0.645 \mu\text{m}$) for turbidity retrieval. Molkov et al., (2019) developed more than ten regional Models for estimating Chlorophyll-a as well as TSM Concentrations using Sentinel-2 satellite data in the Gorky Reservoir.

The tested algorithms include more than 30 different algorithms; for Chl (Saberioon, Brom, & Soucek, 2020; Wang et al., 2017); TSM & Turbidity by Wang et al., (2017) who tabulated and tested more than 20 algorithms, TSM & Chl (Avinash et al., 2012; Molkov et al., 2019) and Turbidity (Dogliotti et al., 2015; Nechad, Ruddick, & Park, 2010). However, the final algorithms after validation with temporally proximal SATCORE data were as follows;

$$Chl = 2.910 \left(\frac{Rrs\ 661}{Rrs\ 835} \right) - 2.186.....(5) \text{ (Molkov et al., 2019)}$$

$$TSM = 2.208 \left(\frac{Rrs\ 835}{Rrs\ 661} \right) + 3.169(6) \text{ (Molkov et al., 2019)}$$

$$Turbidity = \left(\frac{247.10 \times Rrs\ 661}{1 - \frac{Rrs\ 661}{0.1697}} \right).....(7) \text{ (Nechad et al., 2009)}$$

For analysis, both OAS and SDB residuals into categorical variables. The standardized OAS were segregated into three categorical variables; Low (0.33), Medium (0.67 - 0.34), and High (0.67-1). The SDB residuals were classified into two categorical variables 1- SDB residuals within the RMSE limit of ± 2.5 m and 2-SDB residuals that are above or below the RMSE of ± 2.5 m. Thereinafter, SDB residuals were analyzed for spatial variability and distribution of OAS.

The result of the entire processing of this step has been discussed in section 4.5 of Chapter 4.

3.2.6 Merged Random Forest (MRF) Model of SDB

Different SDB models for enhanced SDB estimates have been described in earlier Chapters. Amongst them, the RF SDB algorithm has demonstrated greater accuracy in SDB determination. Further assessment of the RF algorithm in SDB estimation in different scenes is required to show the robustness of the model for practical use. The

strength of the RF algorithm for suitability in SDB derivation is further elaborated below.

Strong modelling methods like RF outperform other methods like single decision trees in aspects of persistence. To reduce overfitting as well as lessen bias-related errors, they combine multiple decision trees, producing insightful results. (Sagawa, et al., 2019). Numerous decision trees are developed utilizing randomly chosen samples in RF, which is a transformation on single Classification Trees. Due to the fact that RF techniques are an extended version of bootstrap aggregation 'bagging' DT, they are also known as ensembles of DT algorithms. The full set of data is used to produce each random subset, and each one has the same number of data points. The used data is added back to the entire dataset so that it can be used in additional trees. Every data point has to have an equal chance of being chosen with each successive random subset when using the "bagging" method that RF uses to select random subsets. In bagging, each tree grows on its own, and training samples are created from the original data using bootstrap sampling. The average of the model generated by each individual tree serves as the final estimation.

RF's greatest strength lies in the fact that it provides one of the most precise ML algorithms currently in use, able to handle a substantial number of predictors while also maintaining accurateness in the existence of a significant chunk of missing value.

In this step following execution was processed;

- i) In this stage of research, we have derived all three OAS as per algorithms mentioned in the previous section and the same was cross-validated for each event of SATCORE data collection date for each of the four study sites.
- ii) The result of previous stages of research has indicated that most of the erroneous SDB residuals and extreme values were distributed in high or medium-concentrated OAS regions. Therefore, three OAS Chl, TSM, and Turbidity were included in further data processing to enhance SDB estimation accuracy by considering the effect of the above three OAS in coastal waters.

- iii) The SDB derivation has been carried out using raster as well as vector data points. Two different software packages were used for the purpose. The Open Source EnMAP-Box plugin in QGIS for raster data and Jupyter Notebook-based Python IDL was used for .csv data.
- iv) The purpose of this phase is to create a much more robust prediction model to put the SDB into practice. Thereafter RF algorithm was utilized for SDB derivation in each site for each date of the SATCORE event.
- v) Thereinafter merging of the algorithms for each date was executed to create a single algorithm for each site (such as; there were five RF algorithm .pkl files for Kochi, which were merged to create a single .pkl file). Also, merging the datasets of each date of each site to derive the MERGED RF algorithm was executed simultaneously.
- vi) All single .pkl files for each site (merged for different dates), were now merged into a single .pkl file (final MRF model). The ML syntax for the same is attached for reference.

Refer Syntax-14:- Syntax for merging of RF .pkl files

3.2.6.1 Merged RF Model

The steps of development of the MRF model are shown in the figure below. The case studies for each site Kochi, Okha, Chennai, and Gopalpur as per the sequence of steps mentioned i) - vi) above were executed. The resultant algorithm files .pkl was derived by two different methodologies; 1) by merging all .pkl files for each date and each site developing a final single .pkl file and 2) by merging all the dataset and developing a final .pkl file. The first approach has resulted repeatedly in erroneous programming and therefore the second approach was accepted for the final model.

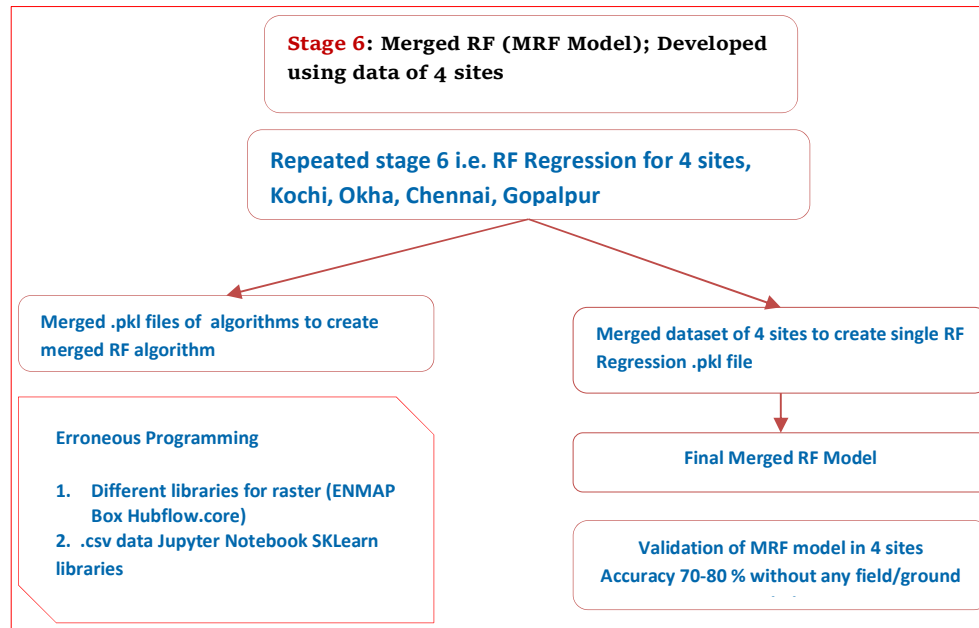


Figure 3.15 Flowchart of MRF

3.2.6.2 Methodology to Develop Merged RF Model

Summing up all the stages of exploratory research for the development of the MRF model, the steps of development of the MRF model from scratch are as follows:

1. The Landsat dataset is available for almost five decades, and searching the satellite images that perfectly match Zero-tide is very rare. The bathymetric data used in the study was already reduced to chart datum / Lowest Astronomical Tide. Therefore, the effect of tides on satellite imagery was neglected. The availability of INCOIS SATCORE data for validation of OAS used in the study paves the way for searching required satellite imagery temporally proximal to it. Accordingly, Landsat satellite data (Landsat-5-8) was downloaded from USGS earth explorer with the filter of date and less than 10 percent cloud coverage. The imagery with glint was avoided for further processing.
2. The next step in the process of developing the RF Algorithm was the delineation of the water region using the NIR band of satellite data. Although, pragmatically delineating water regions without a ground truth/ GPS coastline delineation may be erroneous and may produce severe errors

in intertidal areas. However, the NIR band provides a synoptical overview of the coastline when a very long coastline is under consideration for study.

3. The pre-processing has been carried out as discussed in the previous three chapters using the ACOLITE open-source module. The flowchart shown in the previous chapter shows the pre-processing stages.
4. Theoretical background of different ML algorithms is already been discussed and the previous chapters have shown RF algorithms have the best accuracy in estimation with very reasonable processing time. The various algorithm to derive OAS has already been discussed and the final three algorithms viz. for Chlorophyll- Molkov et al., 2019; for TSM- Molkov et al., 2019; and for Turbidity-Nechad et al., 2009 has been used by derived OAS from the same satellite imagery. The validation of OAS was carried out with available INCOIS SATCORE data.
5. The OAS, Satellite bands in the visual spectrum i.e., B2, B3, and B4 were used in further processing to develop the algorithm. The processing was carried out separately for raster data in the EnMAP-Box plugin of QGIS and using Jupyter Notebook python IDL for .csv data. The resultant RF algorithm was saved as a .pkl file and then the merging of the algorithm was carried out using Python script.
6. Besides, all raster dataset in step 5 above was also extracted in .csv format for creating an algorithm based on a merged dataset of 5 test sites.
7. Thus, a new generic MRF algorithm included three additional parameters viz. Chl, TSM, & Turbidity derived from the same satellite imagery used for SDB derivation was proposed.
8. The proposed algorithm was tested and validated in different geographic regions of India for a depth range of up to 90 meters and an area of coverage of more than 1000 sq km.

CHAPTER 4 – RESULTS

This Chapter summarises the results of research as elaborated in the methodology section of the previous chapter. All six stages of research and procedures to implement the same have been discussed in the previous chapter. The results achieved at each stage have been shown separately in each section below. The first section shows the results of various SDB algorithms and transformations processed in the exploratory stage in both tabular and graphical formats. Section two discusses the results achieved in univariate ML analysis. In Section 3 of this chapter, multivariate ML and its results have been discussed. The further section elaborates on the results of the OAS analysis and its distribution in the areas of interest in the study area. Finally, the development, testing, and validation of MRF have been shown.

4.1 SDB ALGORITHMS & TRANSFORMATION

The most cited SDB tools in the literature on the subject include Stumpf et al.'s (2003) log-ratio algorithm and Lyzenga et al.'s (2006) log-linear model. The relative bathymetry model, which is based on utilising the Log Ratio algorithm, is part of the SPEAR software suite in Envi 5.3. Additionally, the ML methods SVR, LR, and RF Regression have been assessed to produce SDB. The result of each step of processing is discussed below.

4.1.1 Log Ratio Transformation:

The Raster Calculator feature in QGIS is used to perform the Log-ratio transform on a raster dataset of satellite imagery and bathymetry. The table below shows the few band groupings that have been tested and are listed. The correlation coefficients of bathymetry to the product derived from the log-Ratio transform are shown below in Table 4.1. In comparison to other band groupings, the ratio transforms of the Blue and Green bands in the Sentinel-2 as well as Landsat-8 imagery have proven a closer relationship to the bathymetry of the area. For the purpose of deriving the coefficients, the better resultant transforms from each of three satellite data sets (B/G for Landsat-8 and Sentinel-2; R/G for the ASTER dataset) and in-situ water depth were regressed.

SDB was calculated using the empirical parameters that were obtained. The testing set has been used to validate the SDB that was derived. The three performance indicators R^2 , RMSE, and MAE are shown in Table 4.1 below, showing very poor performance of log-ratio transforms with very high RMSE (more than 5m) values for a depth range of 0–30 m.

Table 4.1: Bathymetry-Log Ratio Transform correlation

Depth VS Product	B/G	G/R	G/B	B/R	R/G	R/B	BG/R
Landsat-8	-0.530	0.315	0.433	-0.145	-0.436	0.143	-0.043
Sentinel-2	0.52	0.51	-0.51	-0.26	0.26	0.03	0.14
ASTER	-	0.430	-	-	-0.523	-	-
Calculation of Statistical parameters for Actual Depth vs. SDB for highlighted transform							
OLI/Landsat-8	B/G	R^2 : 0.41		RMSE: 5.821 m		MAE: 4.876 m	
MSI/Sentinel-2	B/G	R^2 : 0.2958		RMSE: 6.37 m		MAE: 5.455 m	
ASTER/Terra	R/G	R^2 :0.6415		RMSE: 5.929 m		MAE: 5.073 m	

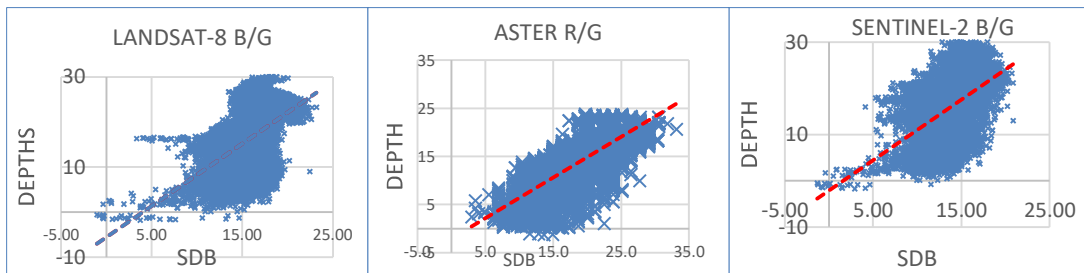


Figure 4.1: Depth vs SDB: Log-Ratio transformation

Figure 4.1 depicts the scatterplot of in-situ Bathymetry versus SDB derived from the log-Ratio transform for all three datasets at the Vengurla site. Except for the ASTER dataset, the distribution of the other two datasets shows very poor performance of SDB derivation due to huge outliers.

4.1.2 Log Linear Transform:

The visual spectrum bands of all satellite images were also applied with Lyzenga’s single-variable log-linear transform. Then every band's resulting transform was correlated with the satellite dataset, and the results are displayed below in Table 4.2. To determine SDB, bathymetry has been regressed against the transform with the closest correlation. The R^2 , MAE, and RMSE obtained as a consequence of validating the SDB against the test set of data are included in the table below.

Table 4.2: Bathymetry vs Log Linear transform

Depth VS Log-Linear transform	OLI/Landsat-8	MSI/Sentinel-2	ASTER/Terra
Blue	0.7299	0.4266	-
Green	0.3817	-0.08322	-0.124
Red	0.3879	-0.08320	-0.6343
Calculation of Statistical parameters for Actual Depth vs. SDB for highlighted transform			
R ²	0.691	0.21	0.556
RMSE	4.282 m	6.97 m	4.127 m
MAE	3.753 m	5.11 m	3.19 m

The correlation coefficients of bathymetry to the product derived from the Log-Linear transform are shown above in Table 4.2. In comparison to other bands, the Blue band for Landsat-8 and Sentinel-2 and the red band of ASTER have proven to have a closer relationship to the bathymetry of the area. The three performance indicators R², RMSE, and MAE are shown in the table above for a Log-Linear transform with three bands of Landsat-8 and Sentinel-2 and two bands of ASTER, showing poor performance of log-linear transforms with RMSE below 5m for a depth range of 0–30 m. However, the performance is slightly better in comparison to the log-ratio transform. The Landsat-8 dataset has provided an R² of 0.69 and an RMSE of 4.2, proving the utility of the log-linear transform in such turbid waters.

4.1.3 Semi Automated methods:

The Relative Water Depth (RWD) tool is available in the ENVI 5.3 Package and is used to create the semi-automated SDB output. The in-situ bathymetry has been examined to correspond with the Ratio method B/G. Because the blue band, which is necessary for such processing by the ML algorithm, is absent, ASTER imagery hasn't been analysed in this tool. The findings are displayed in Figure 4.2 by a graph showing the association between the predicted SDB and validation data. The calculated SDB was verified against ground data. The result from the execution of the algorithm through this semi-automated tool shows that an almost similar result has been achieved for the dataset.

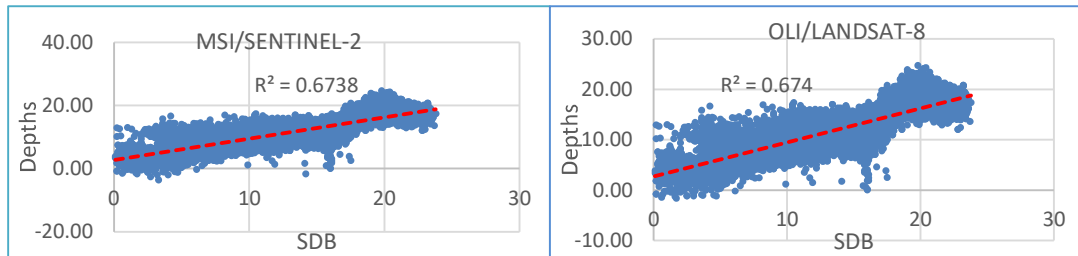


Figure 4.2: Depths vs SDB (ENVI 5.3)

4.1.4 Machine Learning:

Huge training database sets are necessary for applying ML algorithms to function efficiently. Bathymetry and relevant data from across all three satellites have been divided equally among the train & test datasets for this multi-variate ML in a ratio of 50% each. Scikitlearn libraries in Jupiter Notebook (python) were used to extract the empirical parameters. Python libraries used for processing include; i) sklearn.svm SVR for SV Regression, ii) for RF Regression sklearn.ensemble Random Forest Regressor, and iii) sklearn linear model for multiple regression. The table below displays the results of the algorithms that were used along with the results achieved. The result of processing shows that ML algorithms on all three datasets have provided better results than conventional and semi-automated tools. The Landsat-8 dataset applied with the RF algorithm has resulted in an R^2 of 0.88, RMSE of less than 2.8, and MAE of less than 1.9. This result is well within the acceptable limit of a 10% error margin for a recursive hydrographic survey.

Table 4.3: SDB by ML Algorithm

Satellite Data	Algorithm	R^2	RMSE	MAE
ASTER	Multiple Linear Regression	0.57	4.95	3.66
	Support Vector Regressor	0.73	4.19	3.27
	Random forest	0.75	3.97	2.97
LANDSAT-8	Multiple Linear Regression	0.74	4.05	3.07
	Support Vector Regressor	0.85	3.11	2.38
	Random forest	0.88	2.78	1.83
SENTINEL-2	Multiple Linear Regression	0.70	4.29	3.33
	Support Vector Regressor	0.79	3.71	2.82
	Random forest	0.80	3.65	2.60

The plugin "matplotlib.pyplot" has been used to plot the comparison of the y test and y predict data. The 'density distribution' plot of "Actual verses SDB" was also shown using the "sns.distplot" library, as seen in Figures 4.3(a) ASTER, 4.3(b) Landsat-8, and 4.3(c) Sentinel-2.

The scatterplot and density distribution plot as shown in figure 4.3(a) for the ASTER dataset (between y-test and y-predict) show that the MLR algorithm has produced huge outliers representing bathymetry up to -90 m, i.e., elevation above ground. Wherein, for the SVR algorithm, although there are no such outliers derived, the standard deviation from the best-fit line is enormous. The Density Distribution plot

for the RF algorithm shows both curves are following the trend except for a few spikes in the region of 10–15 m.

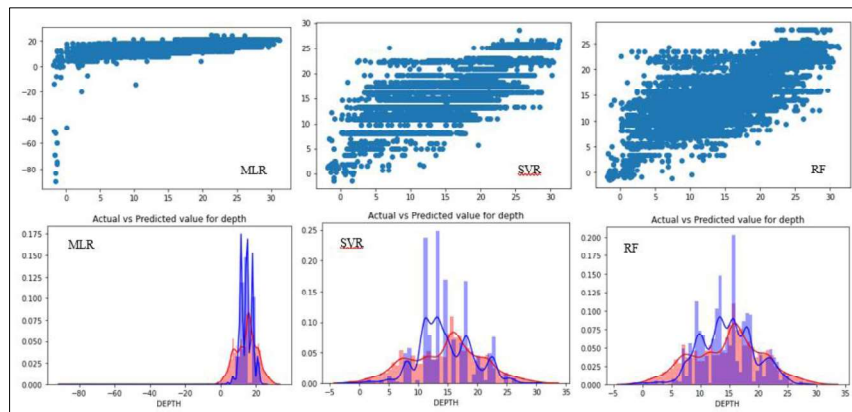


Figure 4.3 (a): y_{test} and $y_{predict}$ (Satellite data ASTER)

The scatterplot and density distribution plot as shown in figure 4.3(b) for the Landsat-8 dataset (between y_{test} and $y_{predict}$) show that the MLR algorithm has produced huge outliers representing bathymetry up to -30m. Wherein, for both the SVR and RF algorithms, there are no such outliers derived. Although the standard deviation from the best-fit line in the SVR scatter plot is smaller and more uniform through all depth ranges than that for the RF algorithm. The Density Distribution plot for SVR and RF algorithms follows a similar trend except in the depth region of 10-15m, where RF shows SDB and Bathymetry curves are following the trend except for a few spikes in the region of 15m.

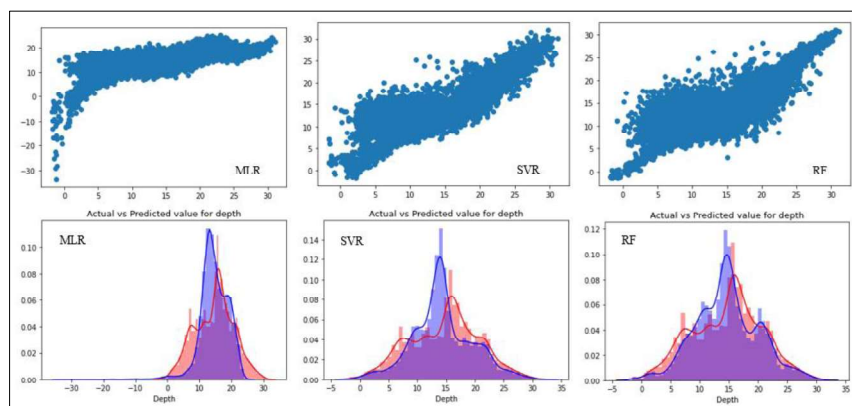


Figure 4.3 (b): y_{test} and $y_{predict}$ (Satellite data LANDSAT-8)

A similar result for Sentinel-2 data has been produced, as shown in figure 4.3(c) below. Except for the ASTER dataset, there were no outliers in the other two datasets, and the density distribution plot for the SVR and RF algorithms follows a similar trend, depicting vast errors in the depth region of 10–15 m.

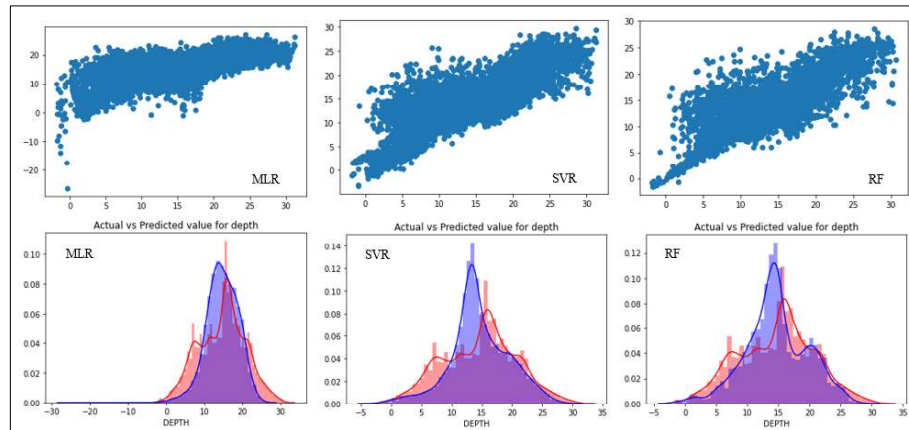


Figure 4.3 (c): y_{test} and $y_{predict}$ (Satellite data SENTINEL-2)

4.1.5 Comparison of Satellite data and Technique of SDB

As a whole, the results show that the Landsat-8 imagery, which uses the RF method, has the errors minimised to the lowest level and produces the best performance among the three-satellite data used. The least RMSE in the depth range of 0 to 32 m was 2.78 m, and the least MAE was 1.83 m, both of which were significantly less than 10% of the maximum depth. The verification with real contour maps is performed using the output product, the SDB map, using Landsat-8 and the RF algorithm model to build the SDB map. ML regression methods LR, RF, and SVR, as well as Log-ratio and Linear-Ratio transforms, were executed to create the SDB product. ML utilising Python script libraries in the Jupyter Notebooks yielded better results than previous approaches. Three index values, MAE, RMSE, and R^2 , were used to evaluate the precision of algorithms. In terms of the overall findings, Landsat-8 offers the best outcome out of the three datasets. In addition to showing promising results, the Sentinel-2 dataset also has an edge over Landsat-8 data in terms of spatial resolution. In comparison to the other two accessible datasets, the ASTER result was the least desirable.

4.2 ML - UNIVARIATE

The causative linkages involving input "Remote sensing reflectance" and output "bathymetry" may be autonomously learned by ML algorithms. The several ML libraries in Jupyter Notebook based on language Python by Anaconda Navigator are used to implement ML methods for this investigation. The "ten-fold cross-validation" approach is also used to evaluate the effectiveness of a few chosen ML algorithms. The correlation analysis between satellite data and bathymetry was executed to assess the better suitable bandwidth for applying the SDB algorithm. The chosen band based on the higher correlation coefficient for both sites is shown in Table 2. Table 2 below is a list of the procedures that were applied to the dataset. The table below is a list of the algorithms that were applied to the data.

Table 4.4: ML Algorithms - Univariate Regression

Site & Data	ML	Method	Libraries	Estimation
A: ASTER 0. 561 μm B: ASTER 0. 676 μm A: Landsat 0. 561 μm B: Landsat 0. 655 μm A: Sentinel 0. 560 μm B: Sentinel 0. 560 μm	Linear	Simple Linear Regression (SLR)	sklearn.linear_model LinearRegressor	R ² , RMSE, Standard Deviation and MAE
		Robust- Huber Regression (HR)	sklearn.linear_model HuberRegressor	
		Robust- random Sample Consensus Regression (RANSAC)	sklearn.linear_model RANSACRegressor	Box Plot and Combined MAE plot
		Robust- Theil Sen Regression (TSR)	sklearn.linear_model TheilSenRegressor	
	Non-Linear	Gaussian Process Regression	sklearn.gaussian_process GaussianProcessRegressor	R ² , RMSE, SD and MAE

The test and train sample distributions have a big influence on ML regression algorithms. Three different combinations of train and test data were utilised to assess the effectiveness of the ML algorithms and investigate the effect of the sample distribution. The sample is divided into three groups: 1) 50% train-50% test; 2) 70% train-30% test; and 3) 80% train-20% test using the sklearn package in Python train test split. The objective of using different combinations of test-train data was to assess the adequacy of test data in the execution of algorithms. The intent was to verify the hypothesis that there are no major changes in the accuracy of the algorithm with a change in the test-train ratio and then the maximum depth of samples required for processing.

4.2.1 Result of Univariate ML

According to the findings, the volume of training and testing affects how accurate ML algorithms are at estimating data. The tables below demonstrate how increasing the training size considerably improves the ability of algorithms to estimate the dependent variables. With a low RMSE of 1.83 m, this result shows that the SLR approach was able to predict efficiency with the ASTER dataset with a R^2 0.83. On this site, however, the GPR provided a slightly better prediction than the SLR algorithm. When estimating MAE from the same data, the Huber Regression performed only barely better than the SLR. Sentinel image bands in SLR performed better than those of other linear robust algorithms at site B, where R^2 0.87 and RMSE was 1.61 m. However, Huber regression proved to be superior at predicting the minimum MAE at this location. The GPR method generated the highest result with R^2 0.91 and the smallest error for this area. At site A, TSR had poor prediction accuracy, while RANSAC performed poorly at site B. GPR had the greatest accuracy in terms of MAE estimates in each location, trailed by HR & SLR. GPR gave one of the most comprehensive solutions, outperforming other robust linear algorithms based on the RMSE. ASTER produced the best results at site A, whereas Sentinel-2A fared better at site B. Tables below illustrate the efficacy of all algorithms used in research areas A and B, correspondingly. Each better-performing algorithm is highlighted in bold text.

Table 4.5: ML Regression results - Site A (Vengurla)

Data	Method	Train:Test 50:50 %			Train:Test 70:30%			Train:Test 80:20%		
		R ²	RMSE	MAE	R ²	RMSE	MAE	R ²	RMSE	MAE
ASTER	Linear Regression	0.638	4.5739	3.539	0.764	2.489	1.460	0.835	1.837	1.287
	Robust- Huber Regression	0.634	4.5953	3.473	0.760	2.496	1.452	0.833	1.840	1.282
	Robust- RANSAC Regression	0.626	4.6340	3.601	0.698	2.615	1.473	0.810	1.871	1.289
	Robust- Theil Sen Regression	0.600	4.753	3.609	0.736	2.547	1.468	0.811	1.871	1.293
	GPR	0.708	4.195	3.386	0.813	2.364	1.439	0.871	1.774	1.274
LANDSAT-8	Linear Regression	0.414	5.410	4.369	0.589	2.753	1.555	0.702	1.964	1.342
	Robust- Huber Regression	0.397	5.454	4.329	0.574	2.767	1.551	0.691	1.971	1.340
	Robust- RANSAC Regression	0.363	5.447	4.439	0.526	2.806	1.565	0.692	1.970	1.345
	Robust- Theil Sen Regression	0.397	5.454	4.333	0.574	2.767	1.552	0.689	1.972	1.340
	GPR	0.55	4.949	4.004	0.701	2.610	1.513	0.789	1.895	1.318
SENTINEL-2A	Linear Regression	0.468	5.252	4.035	0.634	2.705	1.519	0.738	1.941	1.321
	Robust- Huber Regression	0.335	5.598	3.735	0.519	2.810	1.484	0.646	1.991	1.301
	Robust- RANSAC Regression	0.28	5.70	3.878	0.529	2.984	1.505	0.654	2.182	1.323
	Robust- Theil Sen Regression	0.240	5.818	3.743	0.433	2.884	1.484	0.545	2.019	1.30
	GPR	0.284	5.698	4.627	0.470	2.840	1.577	0.604	2.009	1.355

Table 4.6: ML Regression results - Site B (Mormugao)

Data	Method	Train:Test 50:50 %			Train:Test 70:30%			Train:Test 80:20%		
		R ²	RMSE	MAE	R ²	RMSE	MAE	R ²	RMSE	MAE
ASTER	Linear Regression	0.459	4.161	3.400	0.627	2.352	1.442	0.732	1.768	1.276
	Robust- Huber Regression	0.426	4.240	3.276	0.599	2.379	1.426	0.710	1.782	1.267
	Robust- RANSAC Regression	0.374	4.345	3.400	0.568	2.404	1.427	0.686	1.794	1.268
	Robust- Theil Sen Regression	0.388	4.318	3.423	0.563	2.407	1.513	0.693	1.791	1.277
	GPR	0.551	3.908	3.097	0.699	2.265	1.400	0.788	1.725	1.251
LANDSAT-8	Linear Regression	0.587	3.793	3.084	0.726	2.225	1.401	0.808	1.704	1.252
	Robust- Huber Regression	0.586	3.795	3.084	0.726	2.226	1.401	0.807	1.704	1.252
	Robust- RANSAC Regression	0.558	3.888	3.237	0.665	2.311	1.416	0.750	1.756	1.264
	Robust- Theil Sen Regression	0.571	3.845	3.129	0.706	2.255	1.409	0.793	1.720	1.257
	GPR	0.670	3.476	2.711	0.786	2.112	1.346	0.85	1.64	1.219
SENTINEL-2A	Linear Regression	0.752	3.087	2.473	0.813	2.048	1.327	0.871	1.612	1.208
	Robust- Huber Regression	0.750	3.098	2.457	0.771	2.143	1.297	0.841	1.662	1.189
	Robust- RANSAC Regression	0.750	3.099	2.521	0.751	2.182	1.303	0.795	1.718	1.194
	Robust- Theil Sen Regression	0.740	3.150	2.466	0.709	2.251	1.300	0.793	1.720	1.191
	GPR	0.799	2.815	2.238	0.874	1.861	1.272	0.914	1.513	1.174

4.2.2 Linear Vs Robust-Linear Algorithms

The graph below shows a visual overview of the algorithm & accuracy achieved using the statistical indicator MAE for the three at each site. The scatterplot's x-axis represents Rrs, while the y-axis represents derived bathymetry. The scatterplot for the ASTER dataset shows a significant difference in the best-fit line for linear and other robust linear algorithms. Wherein, for the Landsat-8 and Sentinel-2 datasets, there was no significant difference in the best-fit line.

The boxplot in figure 4.4 below compares the MAE (for a test-train of 50 %) from all executed algorithms. The median MAE for both Landsat-8 and Sentinel-2 is around 3 m, whereas for the ASTER dataset it is 3.2m. The minimum and maximum MAE range between 2.6 and 3.8 m, as there are no extreme values derived by the algorithms applied. The range of MAE as depicted by the upper and lower quartiles is more or less similar for all datasets except for the RANSAC algorithm applied to the ASTER dataset of Mormugao.

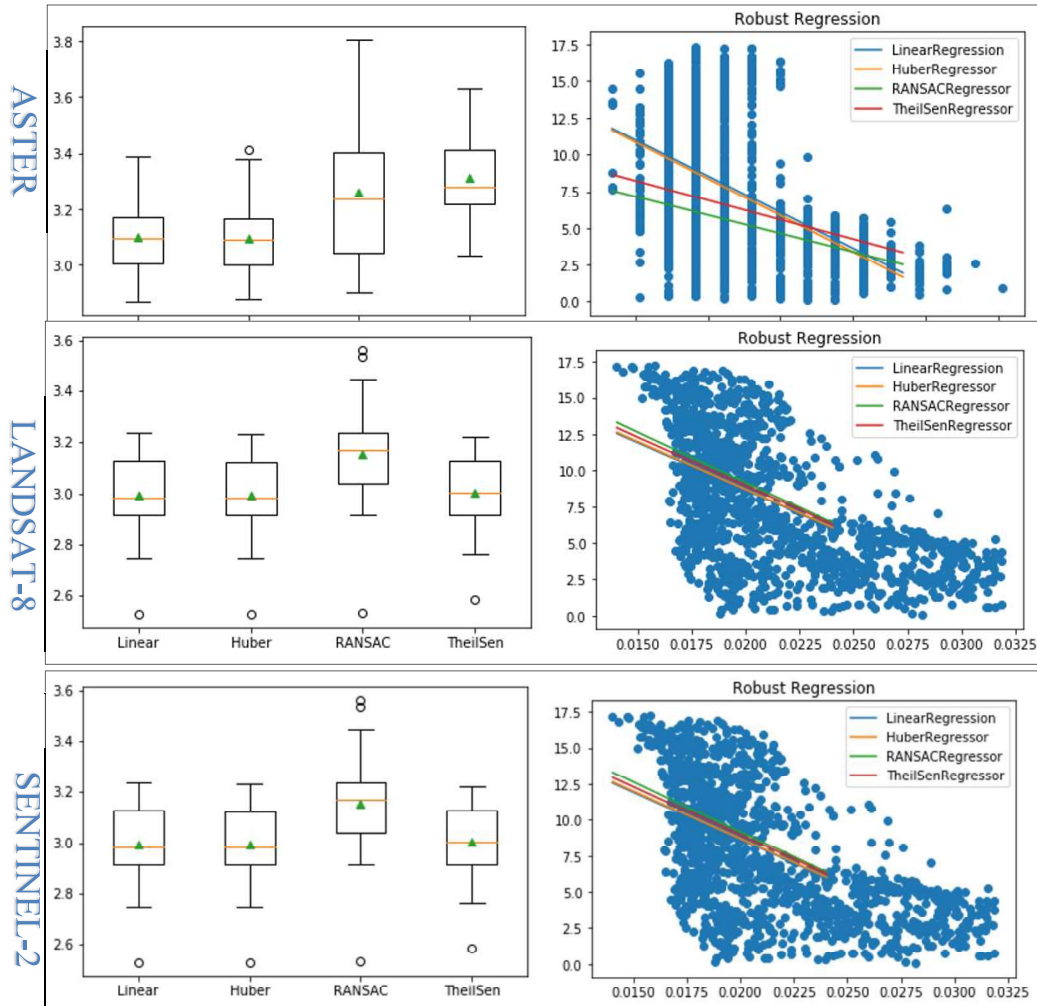


Figure 4.4: Linear & Robust-Linear Regression - Mormugao

Similarly, for the site of Vengurla, the scatterplot for the ASTER dataset shows a significant difference in the best-fit line for linear and other robust linear algorithms. Wherein, for the Landsat-8 and Sentinel-2 datasets, there was no significant difference in the best-fit line. The boxplot in figure 4.5 below compares the MAE (for a test-train of 50%) from all executed algorithms at Vengurla. The median MAE for both ASTER and Sentinel-2 is around 3.5 m, whereas the Landsat-8 dataset is 4.3m. The minimum and maximum MAE range between 3 and 5.8 m, as well as a few extreme values derived by the algorithms applied. The range of MAE as depicted by the upper and lower quartiles and median values for all three datasets have shown significant differences.

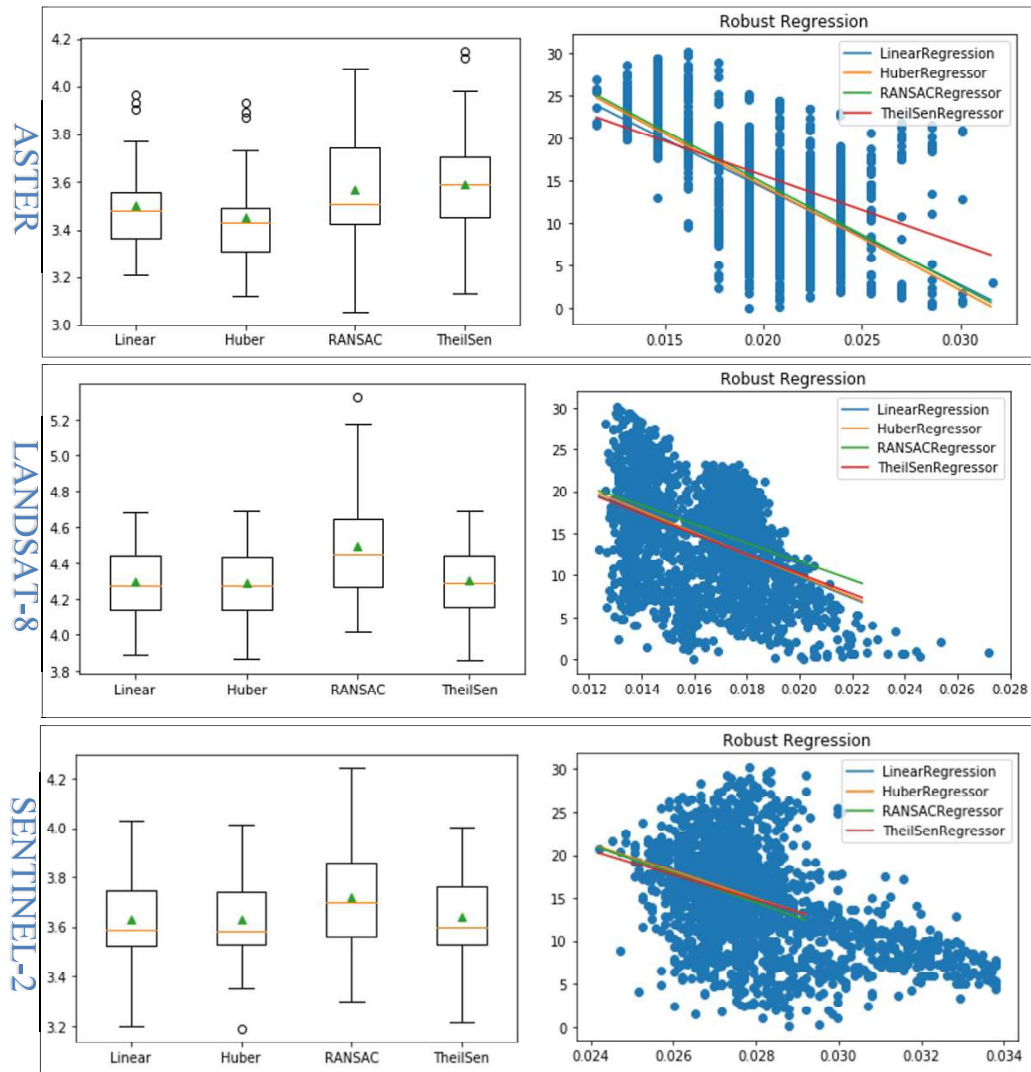


Figure 4.5: Linear vs Robust Linear Regression – Vengurla

4.2.3 Linear Vs Robust-Linear - Accuracy Assessment

The purpose of the research was to determine how well non-linear, simple, and robust regression algorithms performed while estimating SDB in rather complicated environments. This has produced some important findings that may be used for SDB estimation using ML algorithms. The effectiveness of the models has been significantly impacted by the sample distribution. Research's findings support the "Pareto Principle," according to which it is preferable to divide training and test samples into an 80:20 ratio, particularly in study areas that are diverse. This would be especially important in complicated research sites like Vengurla, in which the concentration of sounding is very dispersed spatially and the outcome prediction heavily depends on the spread of the spatial train dataset. In order to create a broad and more reliable SDB model with better predicted efficiency in such a complicated region, we also need a large sample size. As

per the findings reported by this study, as depicted in Figure 4.6 below, which shows the plotting of accuracy assessment for three diverse sample sizes and with four employed ML on three distinct datasets at two separate study sites, study findings suggest an 80:20 train-test division of the data split. With the use of ASTER and Landsat images, the findings indicate a considerable rise in R^2 with larger sample sizes; however, the Sentinel data shows just a slight increase. With such a variation in sample size, a significant variation is shown in the RMSE and MAE measurements. The sample size in training has been expanded, and both indications have significantly decreased. The size of the sample in the train dataset is extended, and both indications have significantly decreased.

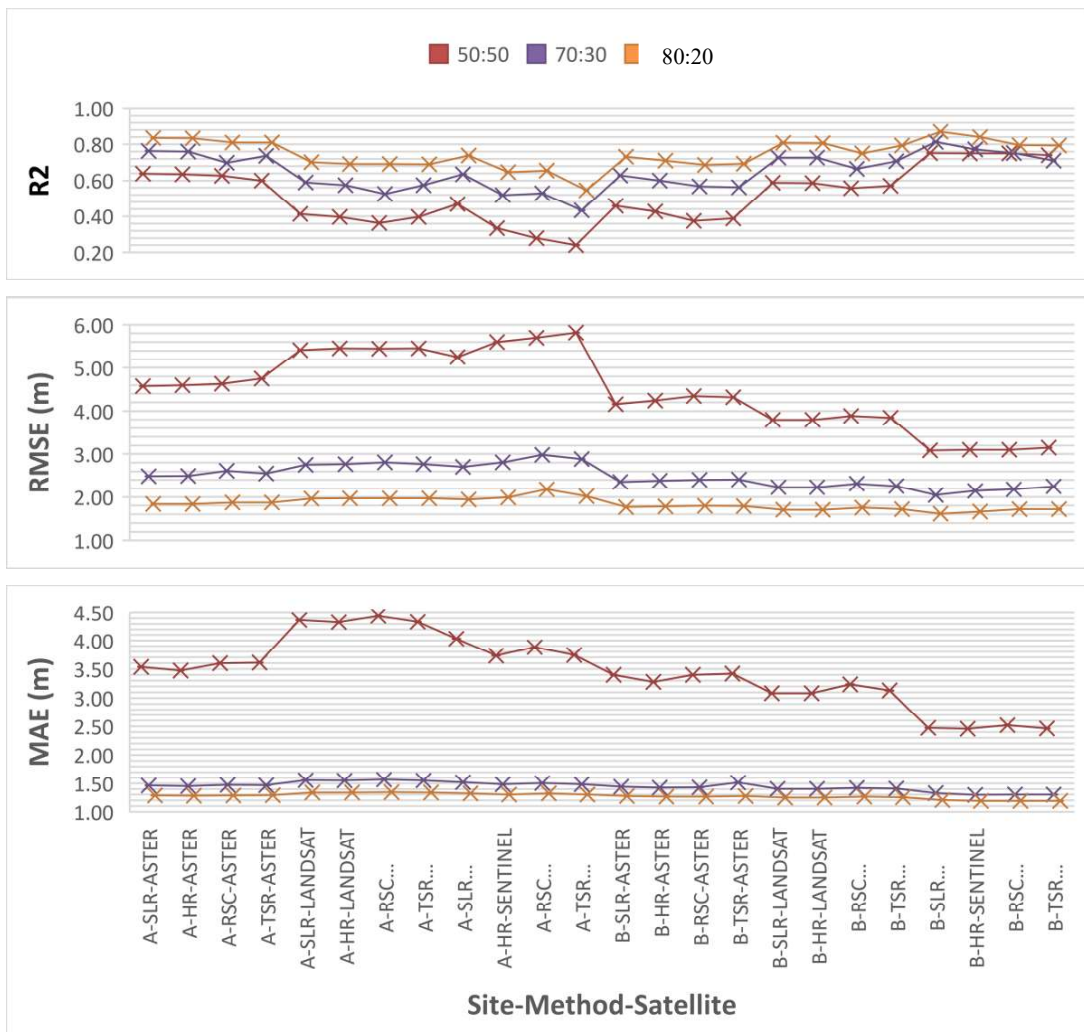


Figure 4.6: Performance on varying ratio of train and validation

4.2.4 Sensitivity Analysis (Sample Size)

A sensitivity analysis was carried out to determine the appropriate sample size needed to calculate SDB using depth data points regression with the Rrs of the single spectral band. The sensitivity study utilises the MAE as a performance indicator. For Mormugao and Vengurla, train and test data sizes were employed at a 1:3 ratio, gradually increasing for each. The results showed that increasing the representative sample for training above 200 depth values had no additional positive effects on the outcome. The study reveals that the statistical indices MAE seem sensitive to a good amount of sample train data up to a maximum of 200 depth values, and the GPR method works at its best in every situation with all satellite data. This suggests that non-linear techniques are required for SDB investigations, potentially as a result of the varied topography beneath the sea.

4.2.5 Linear, Robust-Linear and Non-linear ML Algorithms

SLR exhausted the other trustworthy methods in both research locations according to the measured R^2 for the used regression techniques. Although the GPR model worked well in both research locations, site A's findings using other ASTER data were R^2 of 0.87, RMSE of 1.77 m, and MAE of 1.27 m, and site B's findings using the Sentinel data were R^2 of 0.91, RMSE of 1.51 m, and MAE of 1.17 m. Several scientists have examined the SDB in complex seawater at relatively modest depths of less than 5 m using conventional methods (Bramante et al., 2013). Only a small number of previous studies have used ML approaches, and SDB has only recently started to use ML. Several studies have explored SDB in very complicated and turbid locations, but the results of the algorithms used, including RF, NN, and other ML algorithms, were quite unsatisfactory even for a lesser depth of 6 meters. Several studies have estimated bathymetry in an Indian Context applying the linear model of Lyzenga and the Ratio Model of Stumpf for a depth range of 6m (Jawak & Luis, 2015) and a depth range of 10 m (Pushparaj & Hegde, 2017), where measurement errors were much higher than in this work. We may thus suggest the use of ML algorithms over traditional methods based on the results of this study. This research, which is possibly the first of its kind, examines bathymetric mapping for determining the depth of coastal 'complex and turbid' waters. The research demonstrates that ML regression analysis can predict depths in even complicated waters with reasonable precision and has the possibility of

being employed as an alternative for reconnaissance hydrographic survey in hazardous non-navigational seas utilizing optical RS data. Nevertheless, using optical RS data necessitates at least a few in-situ depths for verification. As a recent replacement for the echo sounder depth required for validation, ATLAS onboard ICESat-2 (space-born LIDAR) may meet this requirement globally (Parrish et al., 2019). Considering increased data coverage throughout the world, this might be an alternative to field hydrographic survey, although application is restricted to clean and translucent water, as proven by Thomas et al. (2021), who demonstrated that ICESat-2 proved ineffective at identifying seabed in the dredging areas.

4.2.6 Summary of Univariate SDB Analysis

The main advantages of the ML algorithms used in this work are: i) a comprehensive examination of the impact of sampling size; ii) the existence of outliers and inliers in the data; and iii) the effectiveness of ML for SDB prediction in complicated coastal underwater terrain with variable water column constituents and submerged topography. SDB estimates in muddy, unreachable, and shallower water may be solved by the SDB techniques used in an affordable, precise, quick, and adaptable manner. The detailed examination of the train data sample size's sensitivity, the appropriateness of satellite spectral channels, and the analysis of outliers gave a strong comprehension of the used ML algorithms and also the outcomes of their use.

Since there is no single rapid technique that fits all datasets, it is quite customary in most ML research to apply multiple algorithms to obtain the best findings. According to the state of the science of SDB to date, no research has previously evaluated the application of robust-regression algorithms to evaluate SDB modelling for the estimation of SDB in coastal region water. The study described in this section reveals that non-linear algorithms, such as GPR, have the potential to identify depths in even complicated waters with accuracy and thus may substitute hydrographic surveys for such uncharted waters. In order to meet international charting criteria, very accurate depth data is needed that satisfies various IHO-CATZOC standards. Therefore, SDB is only useful in recursive or validation surveys since nautical charts require a far greater degree of accuracy.

4.3 MULTIVARIATE ANALYSIS

The performance assessment is based on estimation parameters shown in table 4.7 which demonstrates the efficacy of executed ML algorithms in the study for both the sites. The optimal results of applied algorithms are marked with bold text and poor performing in italics in the table below. The density distribution plot and scatterplot for test data y_{test} and predicted data y_{pred} have been presented in figures 4.7 & 4.8 for Vengurla site and figure 4.9 & 4.10 for Mormugao site. Figures portrays the visual synopsis of the accuracy achieved by applied algorithms for each dataset in both sites. The plot of regression, scatter plot of y_{test} versus $y_{predict}$, and distribution plot of actual & predicted bathymetry has been plotted for each algorithm.

Table 4.7: ML Regression Algorithms

Data	Method	VENGURLA			MORMUGAO		
		R ²	RMSE	MAE	R ²	RMSE	MAE
ASTER	Multiple Linear Regression	0.76	1.91	1.25	0.80	1.70	1.26
	Gaussian Process Regression	0.91	1.66	1.23	0.92	1.50	1.16
	k-Neighbor Regressor	0.89	1.72	1.25	0.90	1.56	1.18
	Support Vector Regressor	0.91	1.67	1.23	0.91	1.51	1.16
	Decision Tree for Multioutput Regression	0.90	1.67	2.83	0.90	1.53	2.21
	Decision Tree Regression k-fold cross-validation	0.91	1.67	2.83	0.90	1.53	2.22
	Random forest Decision trees	0.91	1.66	2.80	0.91	1.52	2.16
LANDSAT-8	Multiple Linear Regression	0.88	1.74	1.25	0.90	1.55	1.18
	Gaussian Process Regression	0.94	1.53	1.15	0.96	1.32	1.07
	k-Neighbor Regressor	0.94	1.55	1.15	0.93	1.47	1.84
	Support Vector Regressor	0.91	1.66	1.22	0.96	1.32	1.07
	Decision Tree for Multioutput Regression	0.83	1.82	2.77	0.96	1.34	1.10
	Decision Tree Regression k-fold cross-validation	0.85	1.80	2.76	0.92	1.50	1.84
	Random forest Decision trees	0.94	1.56	2.01	0.96	1.31	1.34
SENTINEL-2	Multiple Linear Regression	0.88	1.75	1.24	0.94	1.42	1.13
	Gaussian Process Regression	0.93	1.61	1.19	0.97	1.24	1.05
	k-Neighbor Regressor	0.92	1.65	1.20	0.96	1.32	1.08
	Support Vector Regressor	0.91	1.65	1.21	0.97	1.26	1.06
	Decision Tree for Multioutput Regression	0.81	1.85	3.13	0.93	1.46	1.81
	Decision Tree Regression k-fold cross-validation	0.82	1.84	3.06	0.93	1.44	1.75
	Random forest Decision trees	0.91	1.67	2.63	0.97	1.23	1.22

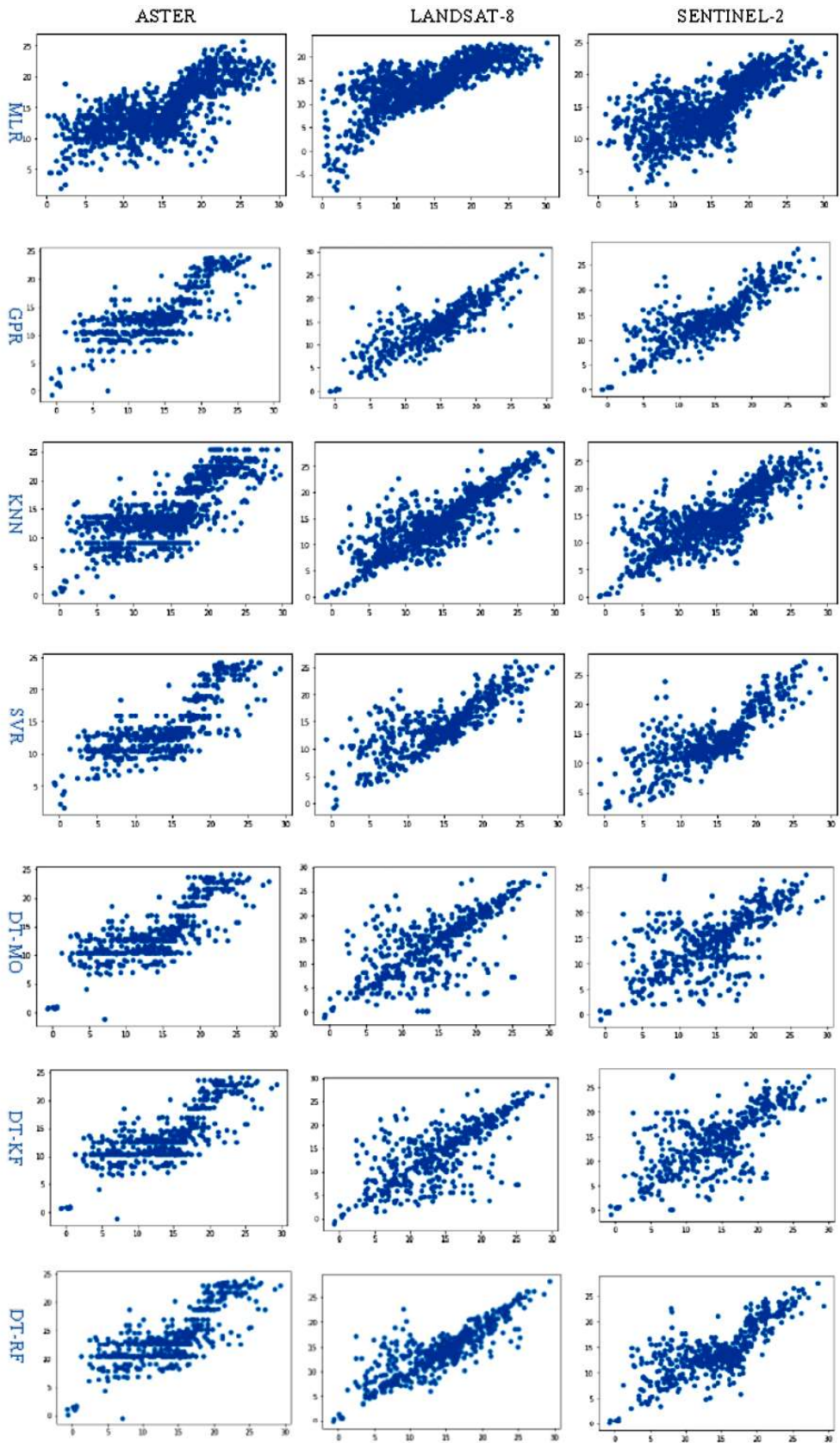


Figure 4.7: The scatterplot of Actual bathymetry vs SDB of Vengurla

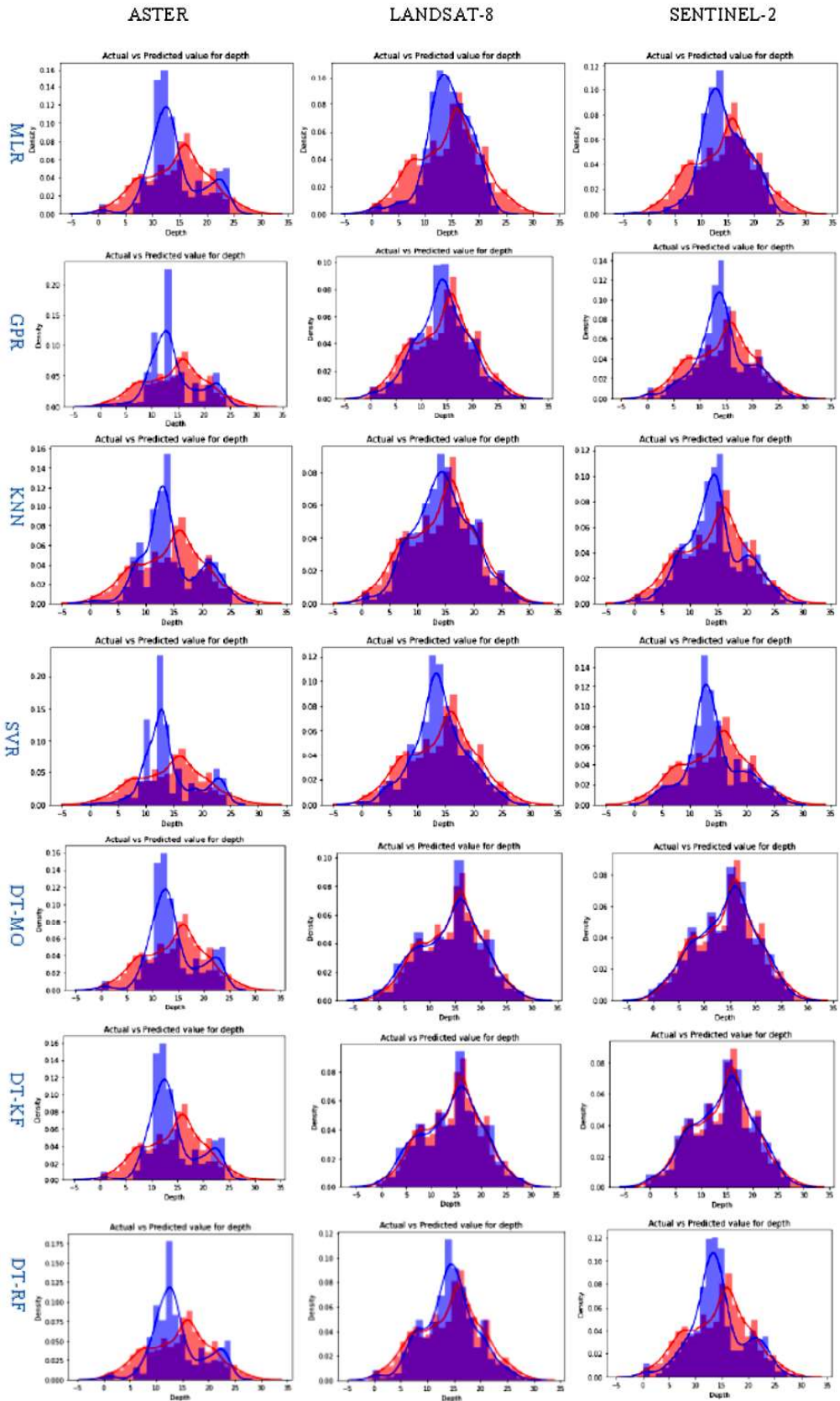


Figure 4.8. Density Distribution plot of Actual bathymetry (Blue) vs SDB (Red) of Vengurla

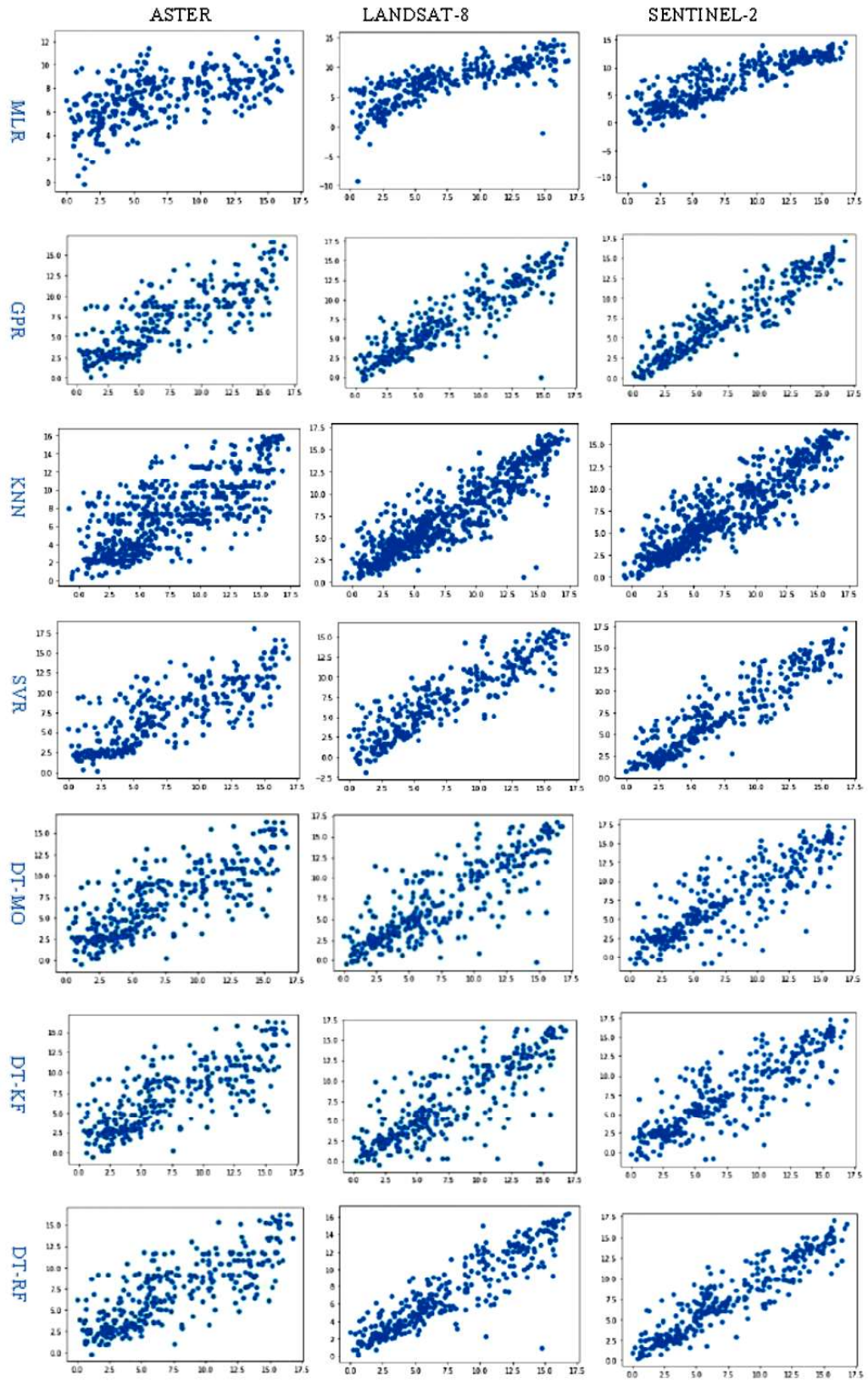


Figure 4.9: The scatterplot of Actual bathymetry vs SDB of Mormugao

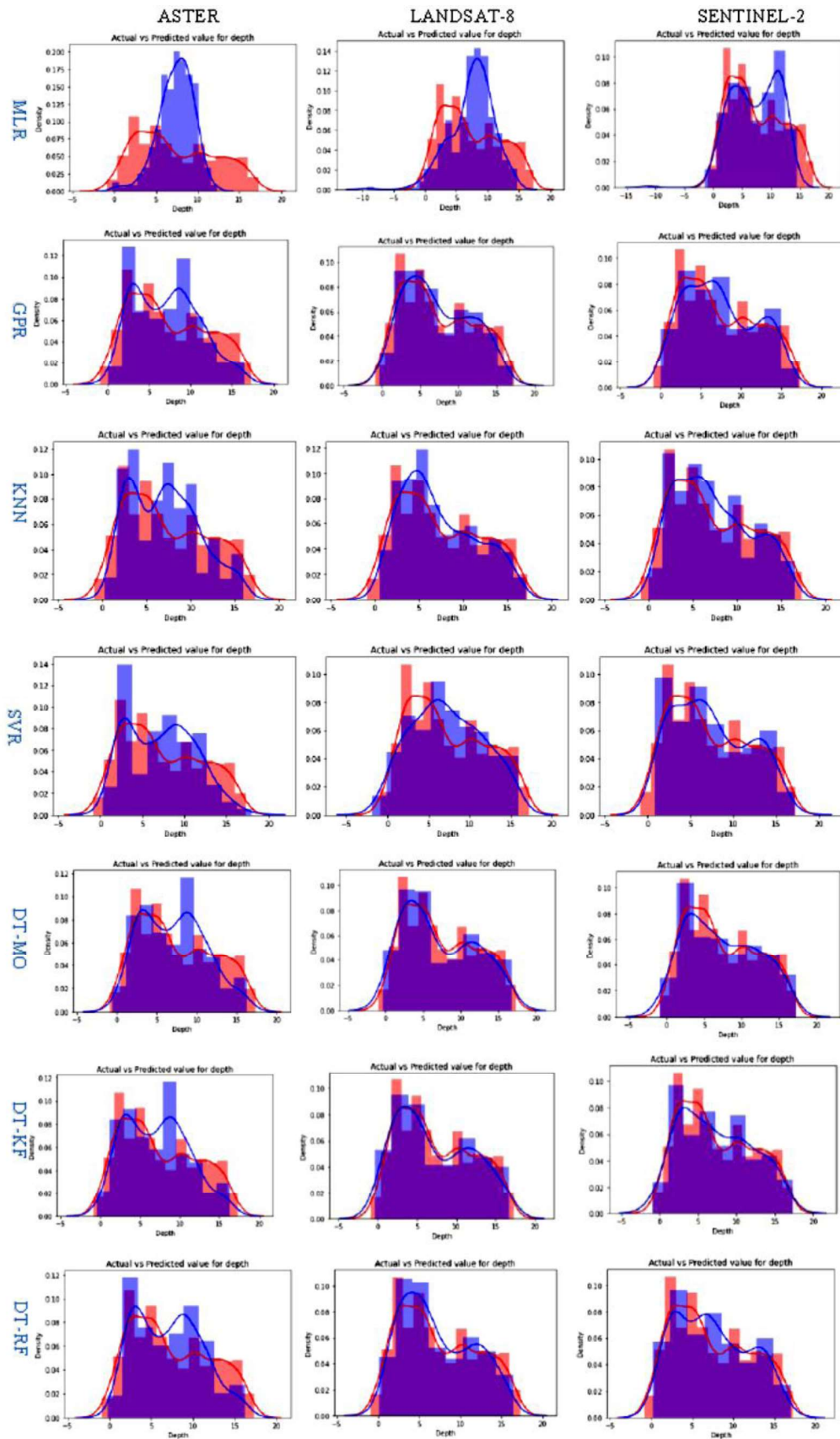


Figure 4.10. Density Distribution plot of Actual bathymetry (Blue) vs SDB (Red) of Mormugao

The research intended to examine the efficacy of multivariate ML algorithms for SDB derivation in highly challenging & complex water. Research findings revealed a few implications for using ML to predict SDB. Figure 4.11 & 4.12 below represents plots of accuracy assessment for all utilized algorithms of two different sites. The MLR and GPR have better performance in all three datasets in both sites.

The decision tree algorithms have proven to performed well except their one estimator MAE was substantially high for all three datasets except for RF in site Mormugao. The MLR and GPR are more robust for the spectral bands of these satellites in respect to both the error estimate RMSE and MAE. Overall, the accuracy of all the algorithms is substantial for such complex waters. The GPR and RF have outperformed providing the best predictive accuracy in both sites respectively. The Landsat-8 dataset with the application of the GPR algorithm result shows in R^2 of 0.94; RMSE with 1.53 m and MAE of 1.14 m using all the visible spectrum bands at Vengurla Site. Whereas at the Mormugao site Sentinel-2 dataset using all the visible spectral bands in the RF algorithm resulted in the predictive accuracy R^2 0.97, RMSE of 1.23 m and MAE with 1.21 m.

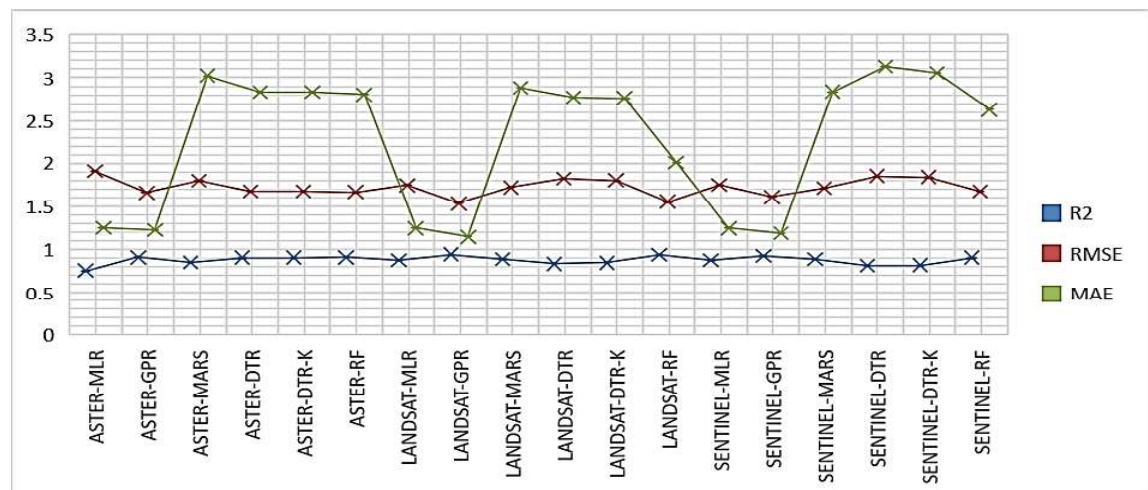


Figure 4.11: The comparison of R^2 , RMSE, & MAE for Vengurla

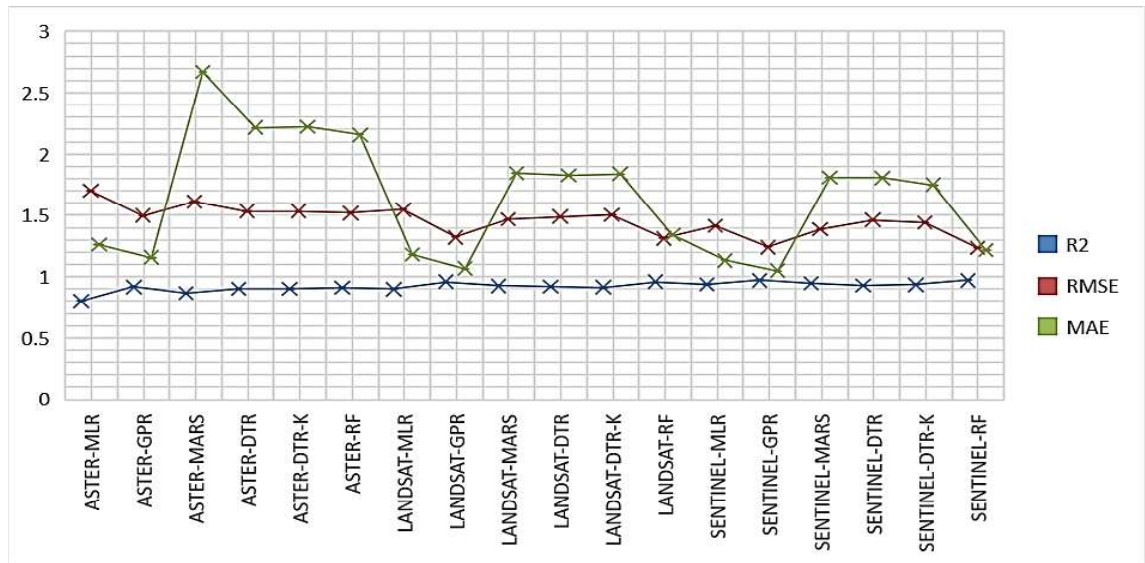


Figure 4.12: The comparison of R², RMSE, & MAE for Mormugao

In this study, we examined how the empirical ML algorithms applied to high-resolution satellite imagery can significantly improve SDB estimation, without the need for any biophysical parameter. This work demonstrated effectiveness of ML in estimating SDB in such complicated coastal waters with variations in water column characteristics and undersea terrain, resulting in a significant concentration of outliers in the data. This study reveals that out of six different algorithms implemented in this study GPR and RF show better predictive accuracy.

The prediction accuracy between the actual and predicted SDB achieved is very high at 0.97 and 0.94 for two complex sites. RF has been proved to be an optimal ML technique suitable for building regression models to estimate SDB. In comparison to other empirical models, RF, SVR and GPR have been found persistent to create more flexible and precise predictions. Few methodological limitations of research include, i) timespan disparity between in-situ & satellite data, ii) Tidal movement's influence on the nearshore region, ii) outliers' distribution and sediment trails and their effect on SDB. Although there is time gap between in-situ bathymetry and satellite data, may not have substantial impact on SDB as the later is reduced to chart datum and corrected for effect of tide.

4.4 BATHYMETRY- SATCORE DATA ANALYSIS

4.4.1 Data Descriptive Statistics

Table 1 represents the details of SATCORE and Satellite Data used in the study at the Kochi site, and Table 2 below demonstrates the descriptive statistics for all the parameters that were included in the study. The data for each event of SATCORE shows measurements of parameters: Minimum value of the parameter, Maximum value of the parameter, Mean and Median of each parameter of the dataset, Range, Standard Error (SE), and Standard Deviation (SD) for n number of measurements. The descriptive statistics help understand the total distribution and variability of parameters in the study area.

Refer Table 1 here from Appendix 1

Refer Table 2 here from Appendix 1

4.4.2 Correlation Matrix

The Correlation (r) measures the degree and path (positive/negative) relationship of two variables, which may range from -1 to 1. 'Correlation-coefficient' r indicates; (-0.3 to + 0.3) shows Weak relation, (- 0.5 to - 0.3 or 0.3 to 0.5) shows Moderate, (- 0.9 to - 0.5 or 0.5 to 0.9) shows Strong relation, wherein (- 1.0 to - 0.9 or 0.9 to 1.0) shows very strong association (Gujarati, 2003). The resultant correlation matrix is shown in Table 3 for each event of the SATCORE date.

The correlations among the variables indicate larger spatiotemporal variations in the dataset, except for a few of the parameters that have shown a consistent relationship to respective satellite scenes. The variables that were correlated to bathymetry in most of the events are; Chl, TSM, and Turbidity. The satellite data shows that bandwidth 561 has shown a very strong association with bathymetry in all scenes and is highest in all scenes, followed by bandwidth 479. All the correlations among variables are found to be very scene-specific and thus cannot be comprehensively commented on for the distribution in coastal seawater.

Refer Table 3 here from Appendix 1

4.4.3 Regression Modelling of all Parameters

Regression analysis provides a scientific explanation of the relationship among the variables. This method is used to forecast the dependent-variable (y) for the given independent-variable (x /univariate/simple linear regression) or the set of independent variables ($x_1, x_2, x_3 \dots x_n$ / Multivariate/multiple linear regression). It is also useful for identifying and controlling the confounding factors in the association of dependent & independent variables. Regression analysis defines ‘best fit’ line minimizing the residual errors (also called Sum of Squares) and verifies slope of the regression line is zero for each independent variable. Thus, these residuals may be construed as the discrepancy of an observed y & predicted \hat{y} by the regression (Gujarati, 2003). The different types of selecting regression depend on the data type, and nature of the problem; like linear regression (continuous variable), and logistic regression (categorical variable). We have used SPSS statistical package for applying regression models, namely; Linear, Logarithmic (initially growth/decay accelerates rapidly then slows down), Inverse (refers to inversely predicting independent variables), Quadratic (assess the equation of a parabola to best fit the dataset), Cubic (polynomial of degrees 3), Compound, Power (dependent variable is proportional to independent variable raised to a power), S (power to squared), Growth (based on the growth model), and Exponential (growth begins slowly and then accelerate till rapid decay). The regression models were applied using SPSS (ver. 25), and the resultant coefficient of determination r^2 (refers to the proportion of the variation that varies from 0 to 1 of the dependent variables explained by the model) has been shown in Table 4. The greater the coefficient r^2 , the superior the regression model. An overall score above 0.7 is usually considered a good prediction estimate.

The most basic interpretation of a regression model is more meaningful with a graphical representation of the model. The regression coefficients between bathymetry and all independent variables can be easily comprehended if depicted by the scatterplot. The scatterplot for each regression model is shown for all the variables in additional images placed in Appendix II. The result shows that bandwidth 561 has shown the highest and mostly more than 0.85 prediction accuracy using a polynomial (cubic) regression model. The bandwidths of 0.479 μm and 0.661 μm have also shown

predictive accuracy ranging between 0.5 to 0.7 for all the scenes. Among other SATCORE parameters, only a few have shown a consistent predictive relationship between 0.2 and 0.5 in most of the events, which include; Chl, TSM, and Turbidity.

Refer Table 4 Here from Appendix 1

Refer Images: Appendix II

4.4.4 Predictor Analysis

The satellite bandwidths having similar predictive accuracy may lead to multi-collinearity (referring to the presence of more than one precise linear association of a few or all independent-variables) in an applied regression model. While carrying out Multiple regression modelling, one of the assumptions is to avoid multi-collinearity among the explanatory variables. The best empirical analysis provides a solution to generate a model with the fewest required predictors. The predictor analysis is carried out using best subset regression in Minitab (v. 18) to compare different regression models. Minitab continuously selects the best-fitting models with one predictor, then two predictors, and so on for n predictors. The best-fitting models have the highest R^2 values. R^2 , SE, and Mallows' Cp criteria were used for assessing the best predictive accuracy of the model. Table 5 shows the results of multiple regression with R^2 and SE for each event. The predictor analysis shows that bandwidth 561 is a vital predictor of multiple regression, which has a predictive accuracy of more than 75% alone in all the scenes. The accuracy of the model increases with a decrease in values of Mallows' Cp and SE and an increase in variables in the model. In most of the scenes, the regression model was at saturation using 6-7 variables in the regression model. Further increasing the variables in the model marginally increased the value of R^2 and reduced the values of Mallows' Cp and SE. The most important parameters from the predictor analysis, in decreasing importance as per their frequency of occurrence in predictor analyses are; Bandwidth 0. 561 μm > Bandwidth 0.479 μm > Chl > TSM > Turbidity > Bandwidth 0.661 μm .

The adjusted R^2 also accounts for the 'degrees of freedom' and if the 'p-value' is lesser than 0.05, it indicates the relationship to be significant at the 95% confidence interval, as shown in the ANOVA result shown in Table 6. Table 6 shows the F-value, which indicates higher variations in a sample, such that the higher the F-value, the lower

is the p-value (p-value that decides whether the model as a whole and each of the predictor variables are significant). The result of the ANOVA shows the significance of all the variables in the model and also the significance of the overall regression model. Most of the variables in all the scenes have a p-value is less than 0.05, except for a few variables like bandwidth, which is 0.835 μm . Therefore, it can be concluded that most of the variables in the study influence bathymetry but need more detailed analysis to identify their confounding effect in regression analysis.

Refer Table 5 Here from Appendix 1

Refer Table 6 Here from Appendix 1

Though SDB was developed in the late 1970s, it was not thought to have any application in operational bathymetry retrieval until the last decade. Advances in space technology for higher resolution, MS bands, open-source availability, etc. have enhanced its potential to be used as source hydrographic data. The application of various remotely sensed data for coastline delineation is becoming more common in surveying as a cost-effective source of information. It is now getting more accepted not only as an operational exploration instrument but also as a sophisticated technology capable of giving calibrated and validated depths to mariners while utilising very few resources.

Most of the studies in the SDB literature have considered clear (CASE-I) waters where the impact of backscattering is minimal and EMR and light travel deeper in the water. However, the SDB model established in clear water cannot be directly applied to turbid (CASE-II) waters because of the difference in absorption and backscattering properties of water. This study was carried out in the highly turbid region of Kochi; the higher level of sediment in such a region causes stronger backscattering. The result indicated that among the other OAS, Chl, TSM, and Turbidity are consistent in predicting bathymetry along with the green and blue bands of Landsat data.

4.4.5 Summary

The following vital implications can be concluded for further progress of study from the results achieved in this section.

- i) To date, none of the SDB research has focused on assessing the influence of all the bio-optical parameters while the derivation of SDB. The primary reason for the dearth of such analysis is a rare opportunity to have data on bathymetry, sea-water constituents, and temporally proximal satellite data. The INCOIS, India SATCORE project has provided an opportunity to carry out an analysis of various OAS distribution and variability in the near-shore coastal region. Each OAS variable was regressed against the bathymetry to understand its dispersal throughout the study area. It was found that Landsat-7 & 8 visible spectrum bands are highly correlated to bathymetry as well as each other.
- ii) Therefore, to avoid multicollinearity among the satellite bands and OAS, the predictor analysis was carried out based on three important statistical criteria r^2 , Mallows' Cp, and SE. The regression analysis with each independent variable and predictor analysis indicated that Landsat-7 & 8 Bandwidths 561 & 479, Chlorophyll, TSM, and Turbidity are vital variables in bathymetry estimation.
- iii) The findings of this research are preliminary in nature on addressing the influence of Chl, TSM, and turbidity on bathymetry and, in turn, SDB. These findings can provide valuable input on the selection of the best scenes, developing SDB models grounded on weight or corrections based on Chl, TSM, and Turbidity over dynamic coastal waters.

The next stage of research focused hereinafter on only parameters; Bathymetry, B1_443, B2_483, B3_561, B4_655, B5_865, CHL, TSM, & TURBIDITY for developing the SDB model.

4.5 SDB AND OAS ANALYSIS

4.5.1 Descriptive Statistical Analysis

The table below demonstrates the descriptive statistics for all the data included in the study for the site Kochi as of February 13, 2015. This date dataset was chosen due to the most synchronous satellite pass and SATCORE data collection time. The data for each event of SATCORE shows measurements of parameters: Minimum value of the

parameter, Maximum value of the parameter, Mean and Median of each parameter of a dataset, Range, Standard Error (SE), and Standard Deviation (SD) for n number of measurements. The descriptive statistics help understand the total distribution and variability of parameters at the study site.

Table 4.8: Descriptive Statistics of data for site Kochi on 13.02.2015

<i>13022015-n=368</i>	<i>Bathy</i>	<i>B1_443</i>	<i>B2_483</i>	<i>B3_561</i>	<i>B4_655</i>	<i>B5_865</i>	<i>CHL</i>	<i>TSM</i>	<i>TURBIDITY</i>
Mean	19.555	0.075	0.074	0.062	0.043	0.026	1.062	3.428	3.337
SE	0.597	0.000	0.001	0.001	0.001	0.000	0.014	0.031	0.039
Median	18.699	0.074	0.073	0.057	0.039	0.024	0.996	3.351	3.275
SD	11.449	0.008	0.010	0.017	0.012	0.007	0.262	0.597	0.739
Range	41.091	0.037	0.044	0.064	0.052	0.045	1.207	2.515	3.392
Minimum	0.309	0.060	0.056	0.039	0.027	0.015	0.583	1.990	1.426
Maximum	41.400	0.097	0.100	0.103	0.079	0.060	1.791	4.505	4.818

4.5.2 Correlation Matrix

The resultant correlation matrix is shown in Table 4.9 for each event of the SATCORE data for 13.02.2015 at Kochi. The satellite data shows a bandwidth of 0.561 μm has shown a very strong correlation to bathymetry followed by 0.483 μm .

Table 4.9: Correlation Matrix for all the variables in the study

<i>13022015</i>	<i>Bathy</i>	<i>B1_443</i>	<i>B2_483</i>	<i>B3_561</i>	<i>B4_655</i>	<i>B5_865</i>	<i>Chl</i>	<i>TSM</i>	<i>Turbidity</i>
Bathy	1.000								
B1_443	-0.835	1.000							
B2_483	-0.906	0.985	1.000						
B3_561	-0.921	0.951	0.980	1.000					
B4_655	-0.858	0.937	0.947	0.978	1.000				
B5_865	-0.810	0.930	0.922	0.933	0.970	1.000			
Chl	-0.485	0.418	0.468	0.472	0.385	0.274	1.000		
TSM	-0.272	0.491	0.455	0.436	0.452	0.468	0.069	1.000	
Turbidity	-0.438	0.507	0.532	0.507	0.430	0.357	0.789	0.581	1.000

4.5.3 Result of SDB Algorithms and Cross-validation

The SDB algorithms; Linear, RF, and SVR were developed for each scene dated 13.02.2015 and 16.12.2015 using all four bands. However, each of the model was cross-validated in both scenes. Table 4.10 below shows the result of the cross-validation application of SDB algorithms across both scenes. The RF and SVR algorithms have shown better accuracy in terms of all three criteria R^2 , RMSE, and MAE than the linear algorithm. The SVR algorithm has shown predictive accuracy of R^2 0.92 with the lowest MAE of 2.07m. The Landsat-7 data for 16 Dec 2015 was used for cross-

validation of the developed algorithm. However, when cross-validated linear algorithms perform consistently across all the scenes, wherein both RF and SVR show inaccurate predictions with large errors. The SDB residuals were derived for each pixel for analysis with OAS.

Table 4.10: Result of SDB Algorithms

ALGORITHM	TESTED	16 Dec 2015			13 Feb 2015		
DEVELOPED	Method	R ²	RMSE	MAE	R ²	RMSE	MAE
16 Dec 15	LINEAR	0.86	4.01	3.24	0.69	6.01	4.52
	RF	0.90	3.28	2.21	0.47	7.84	6.28
	SVR	0.88	3.63	2.26	0.01	10.75	9.43
13 Feb 2015	LINEAR	0.79	4.89	3.98	0.85	4.18	3.27
	RF	0.38	8.49	7.50	0.90	3.28	2.26
	SVR	0.48	7.75	6.21	0.92	2.95	2.07

Figure 4.13 below depicts the bathymetric map of the area as well as the SDB map derived by applying Linear, RF, and SVR regression. The SVR regression has the best predictive accuracy of R² 0.92 with the lowest MAE and RMSE values. The general trends of changes in depth areas are well captured by the SDB algorithm using only 438 discrete point data points. The SVR-derived SDB map shows a very close resemblance to a bathymetric map of the area. However, The SVR regression algorithm was site-specific and performed very poorly when applied to other sites. wherein the RF algorithm, when applied to other sites, performed better than SVR. Therefore, the SVR algorithm was used for further development of the SDB model.

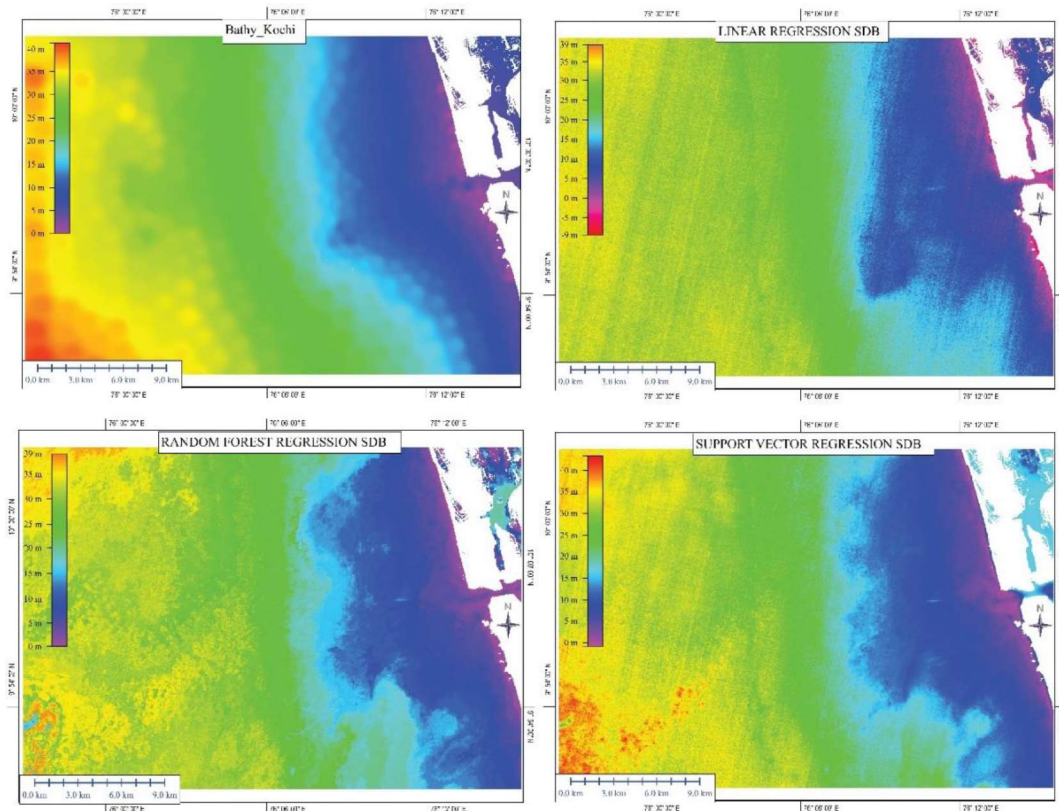


Figure 4.13: Result of SDB derivation using Linear, RF, And SVR algorithms

4.5.4 Validation of Satellite-Derived OAS

The coastal areas have a very dynamic nature, and the concentration of seawater constituents varies very rapidly. In-situ data reliability is very vital for the validation of satellite-derived products, e.g., Bathymetry. To overcome the time lag between in-situ SATCORE data and satellite pass time, various algorithms were tested for validating OAS with temporally proximal SATCORE data, including Chl (Saberioon, Brom, & Soucek, 2020; Wang et al., 2017); TSM & Turbidity by Wang et al., (2017), who tabulated and tested more than 20 algorithms, TSM & Chl (Avinash et al., 2012; Molkov et al., 2019); and Turbidity (Dogliotti et al., 2015; Nechad et al., 2010). The best-suited algorithm for the study site is for Chl with R^2 0.72 using Molkov et al., (2019), for TSM with R^2 0.69 using Molkov et al., (2019), and for turbidity with R^2 0.63 using Nechad et al., (2010). The results of the validation study reach the inference that Landsat data has the ability to provide a prediction of Chl, TSM, and Turbidity with significant accuracy. The scatterplot of in-situ SATCORE data (Kochi date: 13.02.2015) and satellite-derived OAS for Chl, TSM, and Turbidity is shown below in Figure 4.14.

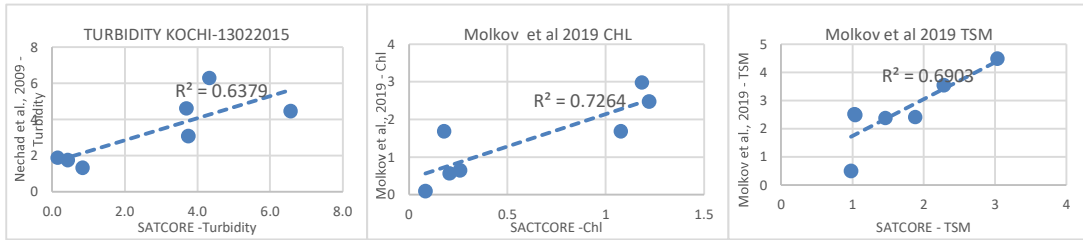


Figure 4.14: Graphs showing validation of Satellite-Derived OAS

The ultimate purpose of this study is to derive OAS for analysing the SDB residuals rather than developing or testing OAS algorithms, as there is a time lag between satellite image pass time and SATCORE data collection time. The coastal areas are very dynamic, and therefore the concentration of OAS varies rapidly. As the study area is a navigable channel and a creek, it has severe impacts from low and high tides, which are semidiurnal in nature at Kochi. The OAS collected during a particular time of day will vary with each of high and low water. Therefore, it was paramount to validate the OAS derivation with satellite data. The time-synchronous satellite data and SATCORE data on 13.02.2015 was better opportunity for this analysis.

Figure 4.15 depicts the map of interpolated and satellite derived Chl, TSM, & Turbidity. The map shows the derived OAS has better variation in distribution through the study area and, moreover, provides a better visual result than interpolation. Moreover, as a general fact, the values of turbidity, TSM are much higher along the coast, gradually decreasing from coast to seaward. Thus, the result of SATCORE's collected and satellite-derived OAS was accepted based on validation and visual interpretation.

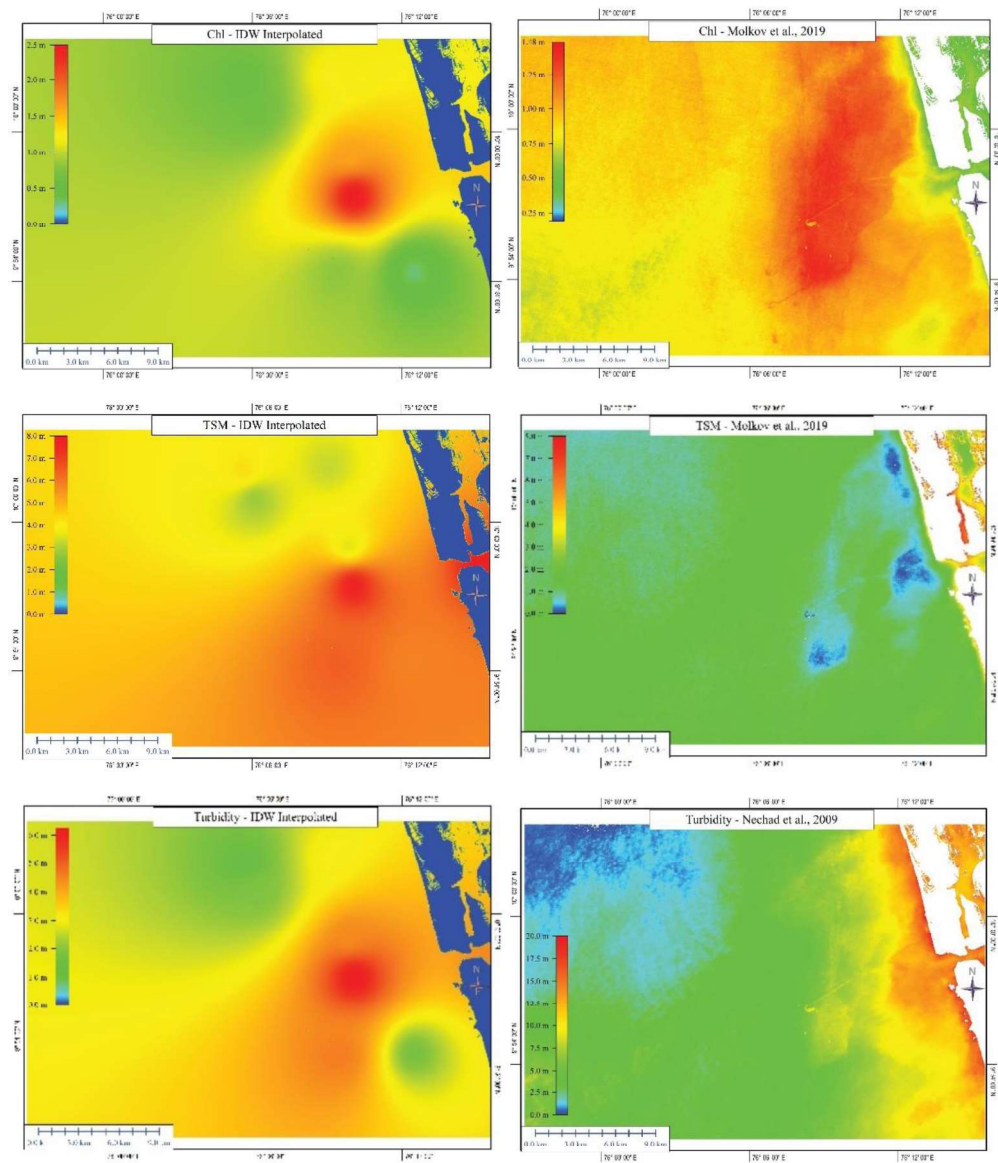


Figure 4.15: The IDW interpolated versus Satellite-Derived OAS distribution

4.5.5 SDB Residuals and OAS

The OAS has a severe impact on remote-sensing reflectance, which produces residual shoaling bias mainly caused by Chl, TSM, and turbidity (Ashphaq et al., 2023). The optimal scene selection for SDB is highly qualitative in nature (Caballero & Stumpf, 2020), and to overcome this limitation semi-empirical algorithms that consider the influence of OAS are needed. The primary aim of this study is the interpretation of errors i.e., SDB residuals that have been estimated above/below RMSE in relation to Chl, TSM, and Turbidity in order to develop a semi-automated SDB model. The SDB

residuals were derived for each pixel by validating against in-situ data. The SDB residuals were reclassified using the QGIS tool 'Reclassify by table' as per the classification values shown in Table 5 below. The SDB residuals were classified into two categorical variables 1- SDB residuals within the RMSE limit of ± 2.5 m and 2- SDB residuals which are above or below the RMSE of ± 2.5 m. The IHO S-44 standard specifies the depth accuracy IHO-CATZOC of C or lower as $2.0 \text{ m} + 5\%$ of depth, therefore 2.5m was used as a threshold for classification for depth range 0-41 m. The other OAS was categorized based on concentration in satellite-derived OAS images into three equal interval categories Low, Medium, and High as reflected in Table 4.11 below.

Table 4.11: Analysis of categorical variable OAS and SDB Residuals

All Categorical Variable Information				OAS Above/Below RMSE		
N = pixel in image (13022015)	N	Percent		N	Percent	
SDB Residual -10 < x ≤ -2.5 → Below RMSE -2.5 < x ≤ 2.5 → Within RMSE 2.5 < x ≤ 15 → Above RMSE	Within RMSE	625870	74.9%			
	Above/below RMSE	210020	25.1%	Above/below RMSE	210020 100%	
	Total	835890	100.0%			
TSM 0 < x ≤ 3 → LOW 3.1 < x ≤ 5 → MEDIUM 5 < x ≤ 10 → HIGH	Low	449148	53.7%	Low	98865 47.0%	
	Medium	377734	45.2%	Medium	102179 48.6%	
	High	9008	1.1%	High	8976 4.2%	
	Total	835890	100.0%	Total	210020 100%	
Turbidity 0 < x ≤ 6 → LOW 6 < x ≤ 10 → MEDIUM 10 < x ≤ 20 → HIGH	Low	32790	3.9%	Low	6093 2.9%	
	Medium	691602	82.7%	Medium	177724 84.6%	
	High	111498	13.3%	High	26203 12.4%	
	Total	835890	100.0%	Total	210020 100%	
Chl 0 < x ≤ 0.5 → LOW 0.5 < x ≤ 1 → MEDIUM 1 < x ≤ 2 → HIGH	Low	5915	0.7%	Low	5413 2.5%	
	Medium	596959	71.4%	Medium	160146 76.2%	
	High	233016	27.9%	High	44461 21.1%	
	Total	835890	100.0%	Total	210020 100%	

The evaluation of the classification of OAS in two different classes of SDB residual product is shown in Table 4.11 which shows that for Chl and Turbidity most of the SDB residuals estimated erroneously are distributed in areas of moderate and high concentration pixels. Wherein, TSM has been found distributed equally in low and moderate concentration areas. The majority of research in the SDB literature has focused on pure and transparent (CASE-I) waters, where there is less backscattering and radiation has an ability to penetrate much deeper into the sea water. Owing to the differences in absorption and backscattering characteristics of water, the SDB model

developed for high level of clarity of water cannot be easily applied to turbid (CASE-II) waters. This research focuses on a very turbid area, where the greater the concentration of sediments creates more backscattering. The few factors responsible for light backscattering and absorption have been investigated with bathymetry by researchers of this study and it was found that Chl, TSM, and Turbidity were the few factors that influence SDB derivation using INCOIS SATCORE data. This study examined the influence of three OAS and quantified their effect on erroneous SDB retrieval. The influence of OAS was categorized into three classes; Low/Medium/High; it was found that the most of the SDB residuals which are above or below RMSE level of 2.5 m were distributed within moderate and high OAS.

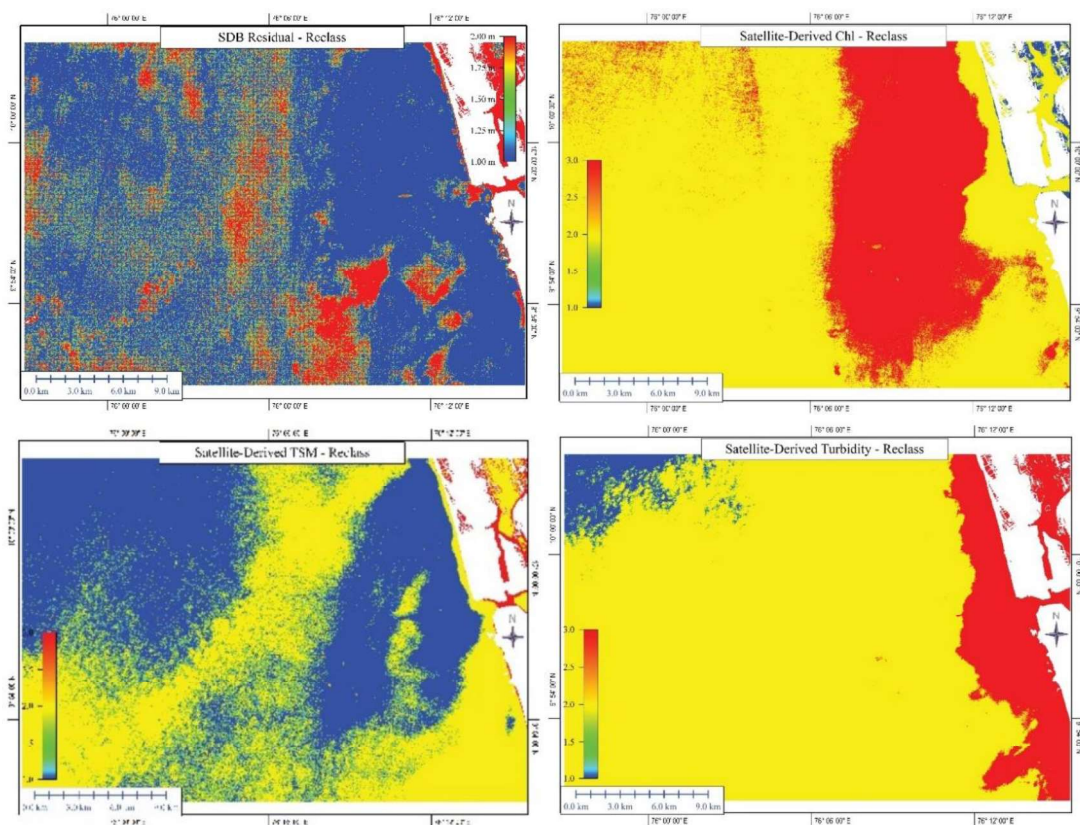


Figure 4.16: Distribution of SDB residuals and concentration of OAS (SDB: 1(Blue) -Within RMSE, 2 (Red) -Above/Below RMSE; OAS: 1 (Blue) -Low, 2 (Yellow)- Medium, 3 (Red) -High)

The figure above shows SDB residuals and OAS classified as per the details provided in Table 5. The SDB residual image depicts that the SDB derivation is more precise for a depth range of 0–20 m, beyond which errors in the derived SDB are unevenly distributed in the study area where the depth range is 20 m to 40 m. The concept of

‘Extinction Depth’ (Stumpf et al., 2003) seems more relevant when deriving SDB. Chl, TSM, and Turbidity were categorized into three classes as per the concentration in satellite-derived products. It is visible in all three images that the concentration range of medium and high OAS is highly related to overestimated/underestimated SDB values.

Figures 4.17 & 4.18 below show the frequency of OAS and the histogram of OAS distribution in three classes, of which most of the erroneous residuals are located in medium and high OAS. The high class of all three OAS has fewer pixels, but most of them have extreme outlier values. The histogram shows the presence of high values of TSM, and Turbidity is associated with a skewed underestimation of SDB up to -15 m.

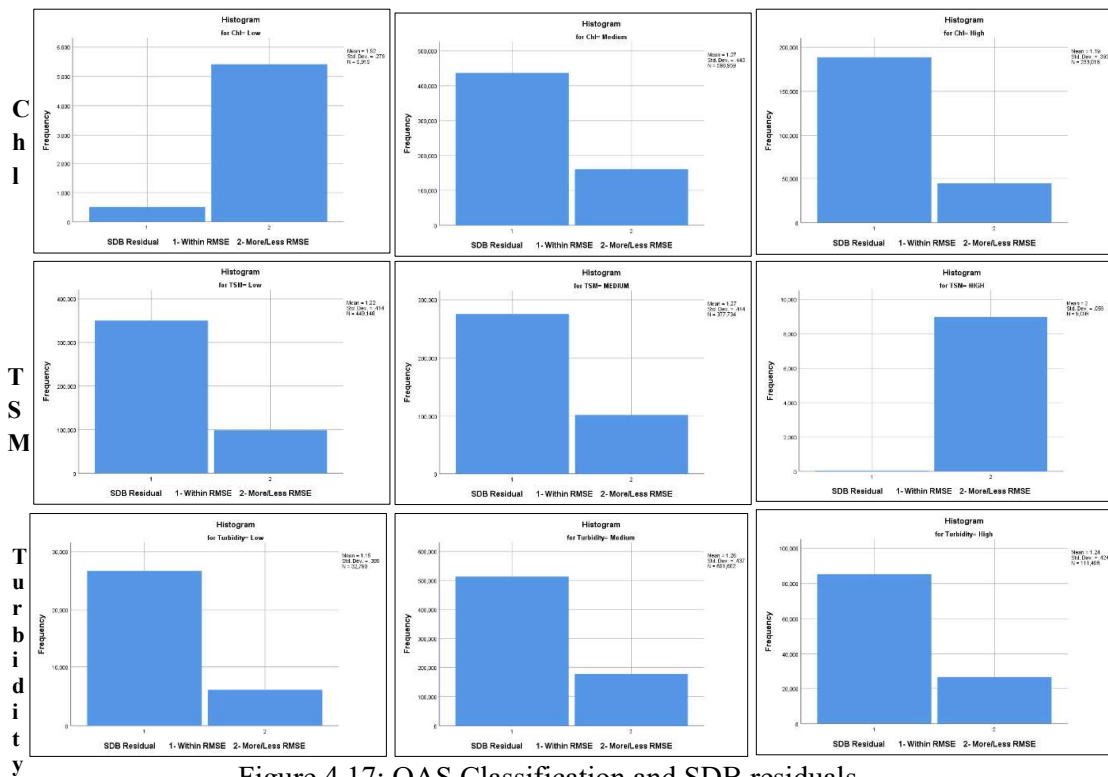


Figure 4.17: OAS Classification and SDB residuals

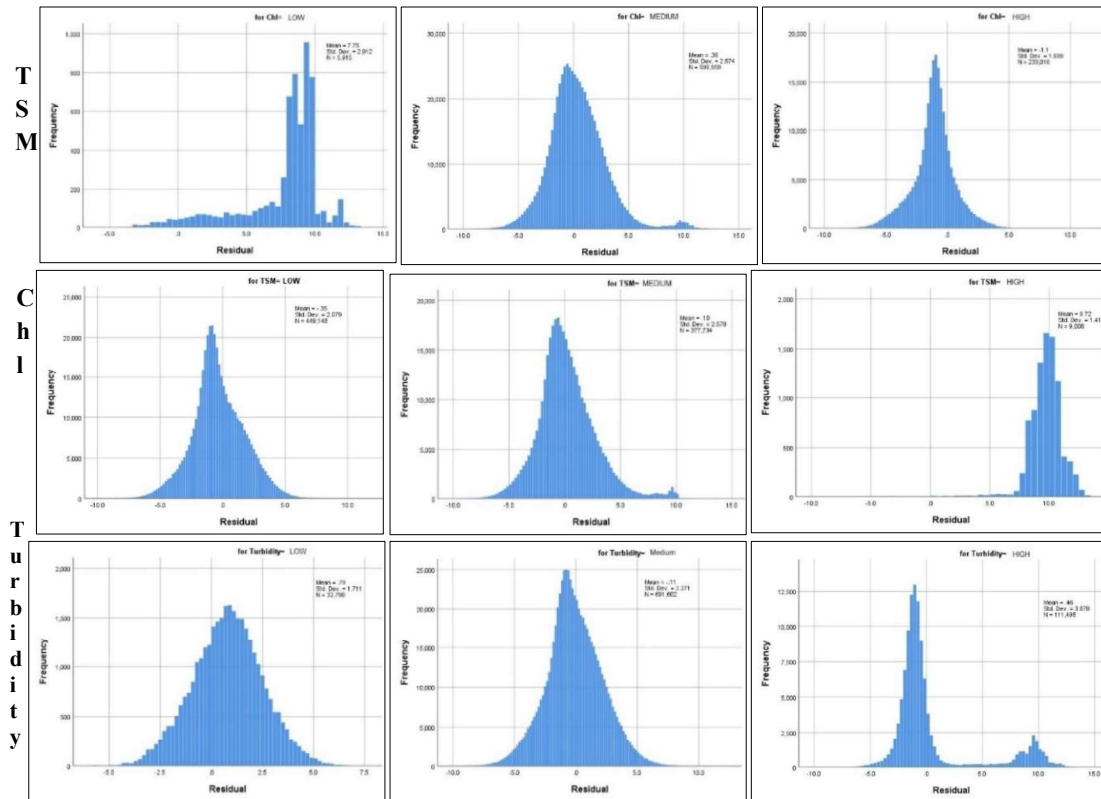


Figure 4.18: Histogram of three classes of OAS with SDB residuals.

This study is an important attempt to address the effect of OAS, especially Chl, TSM, and turbidity, on SDB. The ultimate study objective was to examine and quantify the presence of OAS to enhance SDB estimation. The study aims to quantify the errors in SDB prediction based on the OAS so that a more robust predictive model to operationalize the SDB in practice may be developed. The SATCORE parameters and satellite data have a time gap that was addressed before using the OAS data or analysis of errors in SDB estimation. This was achieved for only three OAS; Chl, TSM, and Turbidity. More than 30 different algorithms were tested to validate satellite-derived OAS products, and finally, three algorithms were identified that closely match in-situ validation data. The SDB was derived using three different ML algorithms; Linear, RF, and SV Regression. The SVR and RF algorithms provided optimal results for the given scene, with the lowest RMSE and MAE. However, RF was found to be more effective when applied to different sites than actually it was originally developed for. The SDB residuals were classified into two classes; within the RMSE limit of 2.5 m and erroneous above/below the RMSE limit of 2.5 m. The OAS concentration, categorized into low, medium, and high, was examined with respect to the SDB residual class. It

was found that most of the SDB residuals were distributed in areas with medium and high concentrations of OAS. Research findings suggest several valid implications for future research, which include; examinations of OAS influence on the selection of satellite imagery for SDB, assigning weight to the pixels whose reflectance has been diminished instigated by OAS, and further development of a bio-physio-optical model for SDB by collecting bio-physical parameter data concurrent with satellite pass time. This study advocates incorporating satellite-derived OAS data into SDB retrieval to enhance the ability to retrieve bathymetry and the utility of the RF algorithm to cross validate results in different scenes. The main limitation of this study is that it is purely empirical, and secondly, there is an absence of field spectral data to develop a semi-empirical SDB model.

4.6 DEVELOPMENT OF MRF

The following four case studies at sites in Okha, Gopalpur, Chennai, and Kochi were carried out in this stage of research to analyse the performance of the RF algorithm in SDB derivation. The OAS and satellite dataset were key inputs provided to the applied ML algorithm. We have derived all three OAS as per the algorithms mentioned in the previous section, and these three OAS (Chl, TSM, and Turbidity) were included in further data processing to enhance SDB estimation accuracy by considering the effect of the above three OAS in coastal waters. The SDB derivation has been carried out using raster as well as vector data points.

Table 4.12 below summarizes the performance of the individual RF algorithm in each site using an algorithm developed site-specific and followed by an MRF model created using merging RF algorithms (as explained in Section 3.2.6) and then cross-validating in each site. The table below compares the results using the site-specific RF algorithm and finally developed the MRF algorithm by merging all the .pkl files and evaluating them using the indices R^2 , RMSE, and MAE.

Table 4.12 Implementation of Merged RF algorithm

DATE	N	R ²	RMSE	MAE	MIN	MAX	For test data
Okha, Depth Range -1.56 TO 58.12							
12.02.2017	1211070	0.834	5.762	4.415	-1.34	55.37	0.911
MRF	1211070	0.24	6.44	4.97	-1.0	36.98	0.269
27.11.2015	1211070	0.749	6.232	4.57	-1.31	55.78	0.784
MRF	1211070	0.21	4.70	3.55	1.8	32.65	0.42
31.03.2015	1211070	0.695	7.748	6.094	-0.6	55.39	0.806
MRF	1211070	0.47	4.40	3.54	-1.2	50.90	0.58
GOPALPUR, Depth Range -1.29 TO 93.8							
17.03.2015	1617875	0.95	5.225	3.182	0.17	85.27	0.933
MRF	1617875	0.82	9.364	6.325	0.12	82.53	0.854
22.03.2013	1555667	0.735	11.551	7.305	0.0	85.09	0.688
MRF	1617875	0.524	20.77	16.56	0.17	87.23	0.59
25.01.2015	1617875	0.770	10.743	6.309	1.37	53.30	0.735
MRF	1617875	0.636	18.62	14.58	0.17	81.00	0.645
CHENNAI, Depth Range 0.0 TO 23.39							
08.11.2013	111136	0.939	1.857	1.239	0.0	22.99	0.951
MRF	111136	0.45	3.90	2.63	2.09	38.03	0.59
27.03.2014	111136	0.942	1.86	1.349	0.0	22.28	0.913
MRF	111136	0.515593	2.664	1.855	0.93	22.12	0.64
08.10.2013	111136	0.852	2.567	1.781	1.60	23.13	0.862
MRF	111136	0.740	3.824	3.224	0.3	35.69	0.689
KOCHI, Depth Range 0.14 TO 41.22							
02.11.2011	840907	0.898	3.321	2.125	0.72	38.95	0.932
MRF	903158	0.593	6.60	4.40	0.7	38.12	0.900
04.12.2013	840907	0.964	2.040	1.265	0.41	40.34	0.984
MRF	903158	0.559	9.43	7.10	0.98	39.07	0.765
13.02.2015	840907	0.902	3.174	2.291	0.46	39.47	0.957
MRF	903158	0.23	11.7	10.3	1.58	36.00	0.30
14.02.2012	840907	0.90	3.304	2.218	0.95	39.73	0.939
MRF	903158	0.51	5.309	3.88	11.02	35.26	0.782
16.12.2015	840907	0.924	2.86	1.757	0.54	38.00	0.980
MRF	903158	0.375	3.966	2.595	2.45	37.74	0.514

The area covered in each scene ranges from 100 Sq. Km to 1400 Sq. km, and the depth ranges from -1m to 90m. The study area considered in this study was relatively large compared to any previous SDB studies. Besides, there are hardly any studies that have advocated and tested SDB for depths greater than 50m. This study has considered a depth range up to 90m at the Gopalpur site. Besides, there was a huge dataset of ground-truth data available for training and testing algorithms. The processing time for each run of the RF algorithm was less than 5 seconds for the configuration of the machine (Intel Core i5 processor, 2.60 GHz processor speed, 8 GB RAM, 4 GB graphics, and 1 TB hard disc).

The results achieved by this study have been further discussed in light of existing literature in the next chapter.

CHAPTER 5 –DISCUSSION

For many Blue Economy prospects, a thorough understanding of bathymetry and seafloor topography is essential. However, as was previously said, due to the erratic nature of coast topography, conducting hydrographic/bathymetric surveys in complicated areas such as Vengurla is challenging and also requires significant valuable resources in addition to the danger to both people and equipment. Although the SDB is being studied for years as an alternate to hydrographic surveys, its application in muddy or turbid water & complicated locations is still difficult. This chapter highlights the discussion of the findings from each of the six research phases covered in the preceding section on methodology.

5.1 SDB ALGORITHMS & TRANSFORMS

This research sought to estimate SDB at a location that was both very difficult because there are abundant underwater rocks and nature is highly turbid in the Karli River's mouth. Finding an appropriate satellite band and precise transform for SDB estimate was the main goal of the research's early phase. By comparing the bathymetric information with each transform performed, a correlation analysis was conducted to determine the optimal band. After analyzing the coefficients of correlation for each band, it was discovered that the green band was far more strongly connected with bathymetry than all the other bands. This conclusion is in line with earlier work on remotely sensed that described how light is attenuated in water and how the green channel near 561 nm is useful for SDB (Caballero et al., 2019). In comparison to other bands, the green channel at (561 nm) has better relationships to bathymetry since its very sensitive to factors like CDOM and particulate in the water.

The SDB method was implemented using many transforms, including the two mostly applied Log-linear & Ratio-transform. Log-linear transformation is significantly more linked with bathymetry compared to the other satellite bands. Ratio-transform had poor marginal performance, showing that it is less useful in severely murky water. This proves compatible with the hypothesis of exponential degradation of EMR in water with variable dissolved substance and also turbidity

distribution. The ML was applied to all satellite datasets, and the results demonstrated that RF shows the highest predicted accuracy using Landsat-8 data.

This illustrates how ML algorithms are better than alternative approaches. Additionally, ML techniques are very beneficial for large datasets like the MBES dataset, which is larger in size than point samples in a given data set. Through using the GeoPandas package, the RF technique was used to produce the final SDB map from the data set. To compute the probability plot and histogram of in-situ bathymetry and SDB, all samples were retrieved. According to the probability plot and histogram, predictions are more accurate until a depth of 20 meters before they start to dramatically worsen. When the resulting SDB map has been compared to the area's real bathymetry, a few highly intriguing results were discovered.

Even though analysis revealed RMSE as well as MAE range between 02 to 03 m, in most of the geographical area, with the exception of the submerged rocks area and near the river mouth, revealed a variation in SDB and water depth of less than 2m. The undersea rocky section has demonstrated a variance of approximately 05m in estimate against real bathymetry. The estimating technique failed dramatically in the severely murky water of the Mouth of the river, with a range of more than 10m. This shows that SDB ML algorithms alone may not be sufficient in such areas of severely turbid water and that more examination of the water column features may be required.

5.2 UNIVARIATE ML

The study sought to assess the efficacy of basic, non-linear, and robust regression methods in SDB determination in rather complicated waters. The findings have shown significant observations for using ML in SDB prediction. The sample distribution has shown a substantial influence on the prediction model. The larger amount the of training data, the better the outcome. This study's findings support the 'Pareto Principle' of splitting training to the testing samples in 80 - 20 proportion, which is preferable in complicated study areas. This is especially important in complicated research sites such as site A, in which the distribution of depth sample is relatively uneven spatially and the estimation of results is heavily reliant on the equally distributed spatial datasets. Predictive performance in such a complicated region needs a high number of samples to create a broader and more robust SDB model. According to this study's findings, an 80 - 20 train-test ratio is recommended.

The results demonstrate that increasing the sample size increases R^2 significantly for ASTER & Landsat data, but just minimally for the Sentinel data for site B. With a modification in sample size, the RMSE & MAE measures show a significant variation. With an increment in the training sample, both indices have decreased significantly.

Sensitivity analysis is carried out to establish the ideal sample size needed to calculate SDB using regression between bathymetry and the Rrs of the spectral band. During the sensitivity analysis, the MAE is employed as a performance measure. At Mormugao & Vengurla, the testing and training data sizes were employed at a ratio of 1-3, with the training sample increasing eventually from 50-400 for each. The outcome showed that sample sizes greater than 200 depth values for training did not further improve the outcome. According to the sensitivity study, the indicators MAE was sensitive towards the best choice of a sample of the training data about 200 depth points, and the GPR method works at its best in all situations using various satellite data. This suggests that non-linear techniques are required for SDB investigations, maybe as a result of the variable character of undersea topography.

An unbiased regression calculation, and eliminating outliers have been advised in a few prior SDB research carried out by Favoretto et al., (2017); nevertheless, in complicated water regions, the potential of using predictive ML models to improve SDB estimate is lost. For the estimated SDB outlier analysis, SDB derived from the best-performing ML results was employed. WEKA tools were used, and the outliers & data were found in accordance with the procedure described previously. To evaluate the spatial geographical patterns of outliers, the detected outliers were mapped against the Sentinel's 0.665 band, which has been demonstrated to be successful in detecting sediment plumes in coastal seas. The outliers were mainly found in a particular region of the research area, on a sandy beach near Vengurla, and in a few plumes in the region (Ashphaq, Srivastava, & Mitra, 2022). This means that the intertidal zone, particularly beaches and shallower regions close to the coastline, needs special consideration while processing data because they were not disguised as land features due to lower Rrs. The bathymetric chart for the area was superimposed on the resulting image, and it was discovered that around 80% of soundings fell within the same depth zone. Other soundings, on the other hand, are dispersed unevenly around the chart, with the majority of them being concentrated in a single

location or depth zone. The regional patterns in bathymetry were followed by the general contour trends.

In both study sites, R^2 values for the implemented ML regression algorithms were evaluated by comparing, it was discovered that SLR significantly outperforms the other applied robust ML algorithms. However, the non-linear GPR has performed best in both study areas, with results for site A being R^2 of 0.87, RMSE of 1.77m, and MAE of 1.27m using the ASTER data, and site B being R^2 of 0.91, RMSE of 1.51m, and MAE of 1.17m by using Sentinel-2 data. Several researchers have used traditional techniques to examine the SDB for turbid water at quite shallow water below 5 meters (Bramante et al., 2013). Few prior researchers have employed ML techniques, and the usage of ML in SDB is only a few years old. Several researches have tried SDB in very turbid and complicated locations, however, the results of the used algorithms, like RF, NN, as well as other ML, were also quite unsatisfactory even at a depth of 6 meters (Hassan & Nadaoka, 2017). Some research has used well-known methods to determine bathymetry in the Indian context for depths up to a depth of 6 m (Jawak & Luis, 2015) as well as a depth of 10 m (Pushparaj & Hegde, 2017), where the error was noticeably higher than in our work. As a result, based on the results of this study, we may suggest using ML algorithms rather than traditional techniques. Moreover, the SLR technique yields superior results to commonly used linear univariate algorithms. The non-linear GPR approach, however, delivered the optimum outcome for this investigation. This research, which is possibly the first of its kind, examines bathymetric mapping for calculating depth in coastal waters. According to the study's findings, regression models have the ability to replace bathymetric surveys for navigable as well as non-navigable water when employing optical RS data to accurately measure depths in even complicated seas. However, to validate the use of optical satellite data, at minimum a few in-situ depth data are required. ICESat-2 Onboard ATLAS is a LIDAR which had recently been utilized as a replacement for echo-sounding depths for validation (Parrish et al., 2019; Forfinski-sarkozi & Parrish, 2016; Quilleuc et al., 2021). It could offer a solution to field measurements with growth in data coverage all across the world, but its utilization may be restricted to clean & transparent water as previously Thomas et al., (2021) demonstrated that the ICESat-2 dataset was unable to identify depth in the dredged sediment-laden channel.

5.3 MULTIVARIATE ANALYSIS

The model trained for best performing algorithm for each site was saved as a .pkl file to apply later on different estimation datasets. The Python Geospatial Data Abstraction Library (GDAL) was used to apply trained models on satellite images directly. The result was also saved as a .csv file and SDB maps were created using Global Mapper software. The SDB maps were draped on charts of the area to visually compare the SDB with charted depths. The visual depiction of SDB maps shows a very close resemblance to depth areas and depth contours of the navigational chart.

The study's use of ML algorithms has shown their effectiveness in predicting SDB at complicated coastal waters with variable water column characteristics and undersea terrain, which generally leads to a significant proportionate number of outliers in the data. In environments of turbid & shallow water, the applicable SDB techniques have the capability to provide affordable, exact, quick, and adaptable bathymetry estimate solutions. This study reveals that out of six different algorithms implemented in this study, GPR and RF show better predictive accuracy. The prediction accuracy between the actual and predicted SDB achieved is very high at 0.97 and 0.94 for two complex sites. RF has been proven to be an optimal ML technique suitable for building regression models to estimate SDB. In comparison to other empirical models, RF and GPR have been found persistent to create more flexible and precise predictions. The research has some serious weaknesses that will be dealt with in subsequent studies, such as the discrepancy in the timespan of the collected dataset, the influence of tide surge mostly in nearshore areas where a large number of the anomalies were dispersed, and also plumes of sediment and its impact on SDB computation.

5.4 BATHYMETRY-SATCORE DATA ANALYSIS

Though SDB was developed in the late 1970s, hardly it was thought to have application in operational bathymetry retrieval till the last decade. Contemporary advancements in space technology viz. higher resolution, more MS bands, open-source availability, etc. have enhanced its potential to be used as an alternative to hydrographic data sources. The use of satellite imagery as a cheap source of datasets for coastal delineation is also growing in the hydrographic domain. It is now being accepted not just as an operational research technique, but also as a sophisticated technology that offers validated and calibrated depth data to mariners while utilizing

very few resources. Most of the studies in SDB literature have considered clear (CASE-I) water where marginal backscattering occurs and EMR can penetrate deep inside the water. However, the SDB model established in clear water cannot be directly applied to turbid (CASE-II) waters because of the difference in absorption and backscattering properties of water. This study was carried out in the highly turbid region of Kochi, the higher level of sediments in such region causes stronger backscattering. The result indicated that among the other OAS; Chl, TSM, and Turbidity are consistent in predicting bathymetry along with green and blue bands of Landsat data.

To date, none of the SDB research has focused on assessing the influence of all the bio-optical parameters while the derivation of SDB. The primary reason for the dearth of such analysis is a rare opportunity to have data on bathymetry, sea-water constituents, and temporally proximal satellite data. The INCOIS, India SATCORE project has provided an opportunity to carry out an analysis of various OAS distributions and variability in the near-shore coastal region. Each OAS variable was regressed against the bathymetry to understand its dispersal throughout the study area. It was found that Landsat-7 & 8 visible spectrum bands are highly correlated to bathymetry as well as each other. Therefore, to avoid multicollinearity among the satellite bands and OAS, the predictor analysis was carried out based on three important statistical criteria r^2 , Mallows' Cp, and SE. The regression analysis with each independent variable and predictor analysis indicated that Landsat-7 & 8 Bandwidths 561& 479, Chlorophyll, TSM, and Turbidity are vital variables in bathymetry estimation. The findings of this research are preliminary on resolving the influence of Chl, TSM, and turbidity on bathymetry and in turn SDB. These findings can provide valuable input on the selection of the best scenes, and developing SDB models grounded on weight or corrections based on Chl, TSM, and Turbidity over dynamic coastal waters.

5.5 SDB AND OAS ANALYSIS

Even though SDB was invented in the late 1970s, it was not deemed to have any application in practical bathymetry retrieval until the last decade. The majority of research in the SDB literature has focused on pure and transparent (CASE-I) waters, where there is less backscattering and radiation can penetrate deep through the water.

For the reason of differences in absorption and backscattering characteristics of water, the SDB model developed for clear water may not be easily applied to turbid (CASE-II) water. This research focuses on a very turbid area, where the greater concentration of sediments creates more backscattering. The few factors responsible for light backscattering and absorption have been investigated with bathymetry by researchers of this study and it was found that Chl, TSM, and Turbidity were the few factors that influence SDB derivation using INCOIS SATCORE data. This study examined the influence of three OAS and quantified their effect on erroneous SDB retrieval. The influence of OAS was categorized into three classes; Low/Medium/High; it was found most SDB residuals that are above or below the RMSE level of 2.5 m were distributed within moderate and high OAS.

The SDB residual image depicts that the SDB derivation is more precise for a depth contour of 20 meters beyond that the SDB errors are unevenly distributed in the study area where the depth range is 20 m to 40 m. The concept of ‘Extinction Depth’ (Stumpf, Holderied, & Sinclair, 2003) seems more relevant when deriving SDB. The Chl, TSM, and Turbidity were categorized into three classes as per the concentration in satellite-derived products. It is visible in all three images that the concentration range of medium and high OAS is highly related to overestimated/underestimated SDB values. The high class of all three OAS has fewer pixels, but most of them have extreme outlier values. The histogram shows the presence of high values of TSM and Turbidity is associated with skewed underestimation of SDB up to -15 m.

This study is a vital step toward resolving the impact of OAS especially Chl, TSM, and turbidity on SDB. The ultimate goal of the study was to examine and quantify the presence of OAS to enhance SDB estimation. The study aims to quantify the errors in SDB prediction based on the OAS so that a more robust predictive model to operationalize the SDB in practice may be developed. The SATCORE parameters and satellite data have a time gap that was addressed before using the OAS data or analysis of errors in SDB estimation. This was achieved for only three OAS; Chl, TSM, and Turbidity. More than 30 different algorithms were tested to validate satellite-derived OAS products and finally, three algorithms were identified closely match in-situ validation data. The SDB was derived using three different ML algorithms; Linear, RF, and SV Regression. The SVR algorithm provided optimal results for the given scene with the lowest RMSE and MAE. The SDB residuals were

classified into two classes; within the RMSE limit of 2.5 m and erroneous above/below the RMSE limit of 2.5 m. The OAS concentration categorized into low, medium, and high was examined with respect to SDB residual class. It was found that most of the SDB residuals were distributed in the areas of medium and high concentrations of OAS. The research provides several novel implications for future research which include; examinations of OAS influence on the selection of satellite imagery for SDB, assigning weight to the pixels whose reflectance has been diminished instigated by OAS, and further development of a bio-physio-optical model for SDB by collecting bio-physical parameters data concurrent with satellite pass time. This study advocates incorporating satellite-derived OAS data into SDB retrieval to enhance the ability to retrieve bathymetry. The main limitation of this study is that it is purely empirical and secondly the absence of field spectral data to develop a semi-empirical SDB model.

5.6 DEVELOPMENT OF MRF

Numerous SDB models for enhanced SDB estimates have been described in preceding Chapters. Among these, the RF SDB model put forth in the previous chapter has demonstrated notable accuracy for SDB derivation. Further examination of the RF algorithm in SDB estimation in different scenes was required to show the robustness of the model for practical use. The result of previous stages of research has indicated that most of the erroneous SDB residuals and extreme values were distributed in high or medium-concentrated OAS regions. Therefore, three OAS Chl, TSM, and Turbidity were included in further data processing to enhance SDB estimation accuracy by considering the effect of the above three OAS in coastal waters. The goal of this section was to develop a more robust SDB predictive model to operationalize SDB for practical purposes. In this stage of research, we have derived all three OAS as per algorithms mentioned in the previous chapter, and the same was cross-validated for each event of SATCORE data collection date for each study site. Thereafter RF algorithm was utilized for SDB derivation in each site. The SDB derivation has been carried out using raster as well as vector data points. Two different software packages were used for the purpose. The Open Source ENMAPBox

plugin in QGIS for raster data and Jupyter Notebook-based Python IDL was used for .csv data.

Several case studies at different study sites have been conducted using a variety of satellite images to assess the *RF* SDB algorithm. The first case study was Vengurla, followed by Mormugao, Kochi, Chennai, Okha, Gopalpur, and finally Chennai. However, the same has not been discussed with reference to each site, but an overall analysis of using the RF algorithm and thereafter merging the algorithm and merging the dataset to derive the merged RF algorithm is discussed. The area covered in each scene ranges from 100 Sq. Km to 1400 Sq. Km area and depth range varies from -1m to 90m. The study area considered in this study was relatively much larger than any previous SDB studies. Besides, there are hardly any studies that have advocated and tested SDB for depths more than 50m. This study has considered a depth range of up to 90m in the Gopalpur site. Besides, there were available huge datasets of ground-truth data for training and testing algorithms, which has helped to validate the algorithm in such turbid waters with reasonable accuracy.

5.7 Comparison with GEBCO database-derived Maps

‘General Bathymetric Chart of the Ocean’ (GEBCO) is a huge dataset prepared from bathymetric contours of the oceans originally at a 1:10 million scale. GEBCO Grid is Compiled at 30 arc seconds (926.1 meters at the equator) and has Negative meters units with WGS-84 as a Horizontal datum & Mean Sea Level as a Vertical datum. The derived SDB maps from MRF models were compared against the GEBCO-derived maps & Bathymetric maps for three sites Chennai, Okha, And Kochi. The summary of regression with indices r^2 is shown below

Table 5.1: Summary of comparison between GEBCO and MRF maps

Comparison	Chennai	Okha	Kochi
Bathy Vs GEBCO	$r^2= 0.707$	$r^2= 0.7318$	$r^2= 0.7503$
Bathy vs MRF	$r^2=0.802$	$r^2= 0.8725$	$r^2= 0.865$
GEBCO vs MRF	$r^2= 0.703$	$r^2= 0.6522$	$r^2= 0.7606$

The depiction of the comparison between the maps is shown below in Figure 5.1-5.3 as shown below.

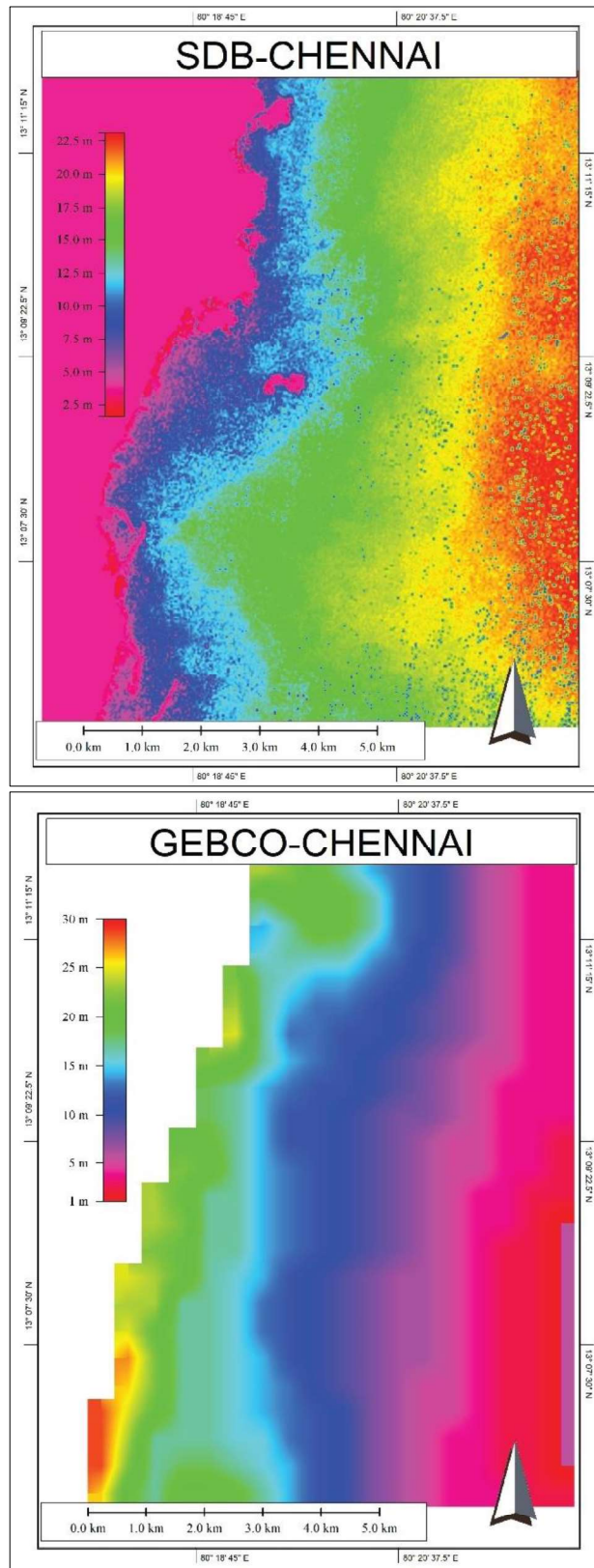


Figure 5.1 SDB compared with GEBCO-derived Map-Chennai

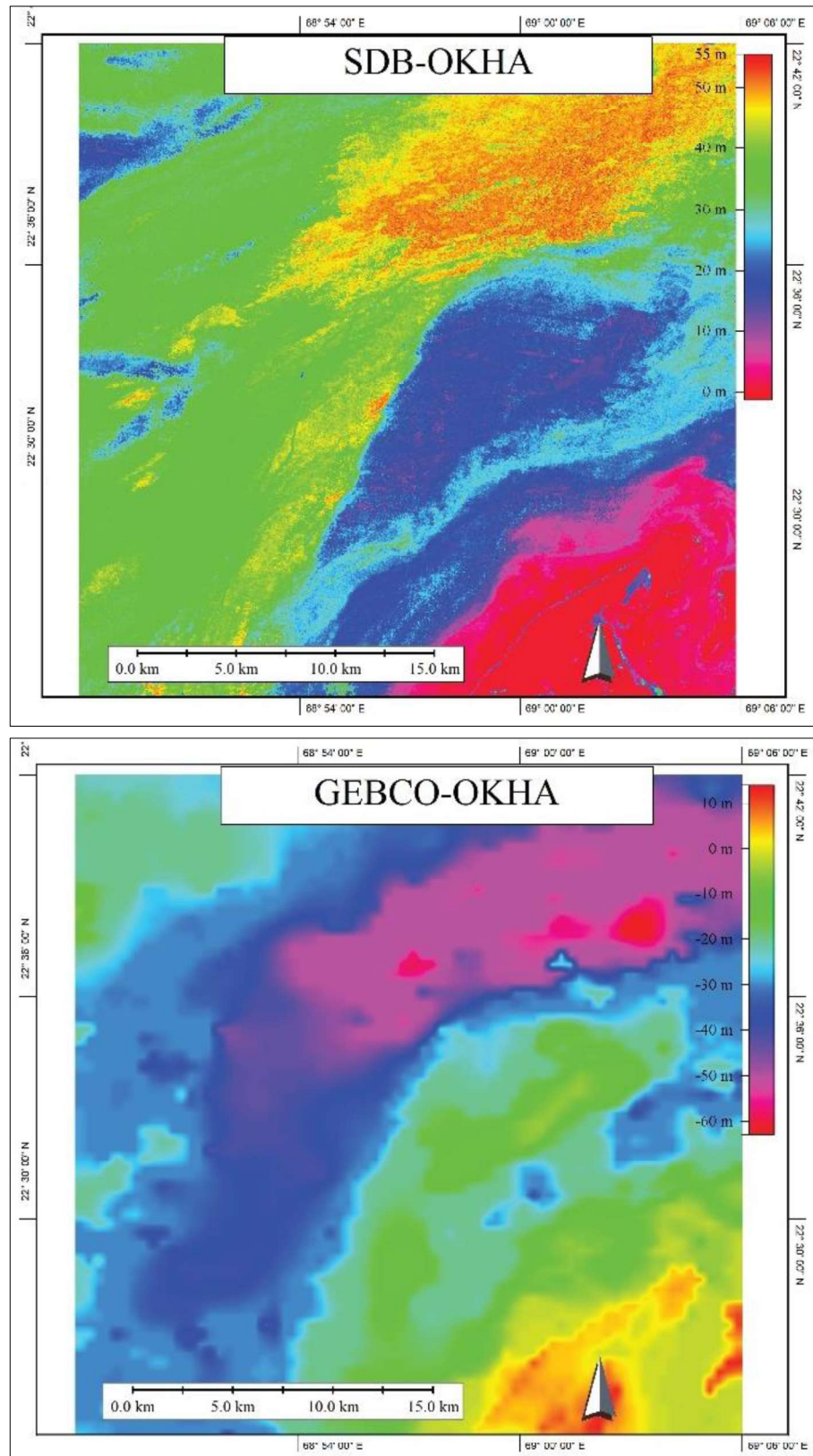


Figure 5.2 SDB compared with GEBCO-derived Map-Okha

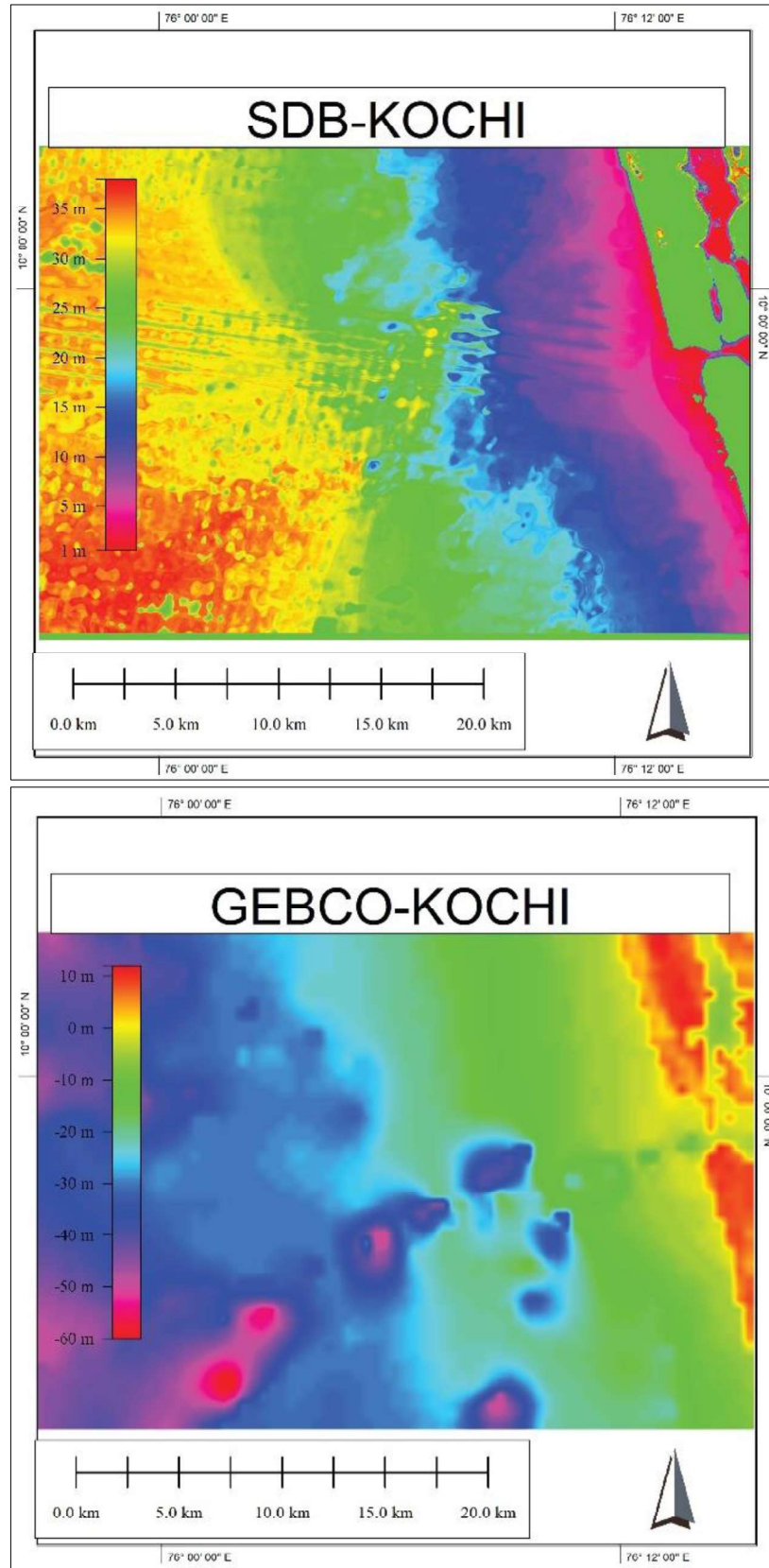


Figure 5.3 SDB compared with GEBCO-derived Map-Kochi

The figures 5.1 for Chennai, figure 5.2 for Okha and figure 5.3 for Kochi above shows the SDB maps generated using MRF algorithm and also by GEBCO derived bathymetry. The SDB maps of the area are having much better resolution than GEBCO based maps. Besides the coarse resolution GEBCO database, further few commercial companies are providing SDB maps on demand for any particular area. It also provides various other services related to mapping water bodies and commercially on-the-shelf software solutions for the same. EOMAP is a company that provides satellite-derived bathymetry (EOMAP official web page, 2022). EOMAP also uses optical bathymetry (Landsat 8) methodology, has a 15-meter resolution with Negative meters unit, WGS-84 horizontal datum, and Chart datum (or LAT) as a vertical datum. However, the biggest drawback of EOMAP is the non-consideration of factors like Turbidity, water column corrections, etc. which this study has tried to overcome with an AI-based approach.

CHAPTER 6: CONCLUSION

This research has contributed in several ways to the knowledge domain of SDB. The results and discussion chapters have already elaborated on the findings of this research. The results achieved from the research has proven few of the implications for the practical utilization of SDB for operational purpose. This study has achieved all four objectives mentioned in Chapter 1. The major contribution of this research to the existing body of SDB knowledge is as follows:

1. Firstly, the study examined the contemporary, most cited SDB model with ML algorithm using single spectral band as well as multispectral satellite data. Other researchers also made a similar attempt but limited their work to comparison rather than further development of the model. In this research, each step was used as a feed-forward for further development of the model and errors were computationally used as an opportunity for further improvement of the model.
2. This study proved the efficacy of machine learning algorithms when compared to other SDB models in the highly turbid & complex water of Vengurla. The question of univariate versus multivariate data use for processing was addressed along with dealing with outliers in the data by robust algorithms. The adequacy of the sample size was addressed by sample sensitivity analysis. Moreover, the best distribution of the training and testing dataset was also evaluated using three different splits of sample data.
3. Then this study proved non-linear multivariate ML model like GPR performs better in such complex waters, but has a very high computation time. Wherein, models like RF & SVR perform slightly less accurately but have high computation speed. These results were validated in two different geographical areas of Vengurla & Mormugao using a single scene of the satellite image, which is very rare in SDB research to date. Further, the study has compared several ML algorithms for SDB derivation, and based

on higher efficiency, the less time-consuming and computationally extensive model was suggested.

4. This study was one of the pioneering efforts in examining the role of OAS on SDB derivation. This is the first time in SDB research where such vast longitudinal data was used for preliminary analysis of OAS on SDB. This unmatched opportunity was achieved because of the availability of the INCOIS SATCORE data collected from 2008 to 2016 and shared for the research to identify the moderating or mediating effect of OAS on SDB derivation. This data has been collected with the use of great amounts of effort and huge resources.
5. This study was exploratory in nature and on preliminary examination of OAS by parametric analysis concluded only three parameters in the water column have an effect on SDB derivation and the same were used in further processing for the development of the SDB model.
6. The four case studies have been conducted to develop and refine the algorithm at Okha, Kochi, Gopalpur & Chennai sites with vast geographical separation. But these sites almost represent variability in coastal region characteristics of India. The SDB derivation for such a vast region of 100 sq. Km to 1400 Sq. km is rare in SDB, which was achieved by this study. Besides, this study has applied SDB derivation up to a depth of 90 m with reasonable accuracy in the Gopalpur site.
7. A generic Merged RF algorithm was developed as an SDB model and evaluated in all study sites. A generic Merged RF algorithm was suggested for recursive bathymetry mapping in the coastal region of India by testing and validating in different geographical locations.
8. However, the most paramount contribution of this research is in achieving the SDB in highly turbid water by adding three parameters Chl, TSM & Turbidity derived from the same satellite imagery used for SDB. Previously, there were various efforts to correct turbidity's influence on SDB retrieval (Caballero & Stumpf, 2023). The different approaches have contributed significantly to cater for attenuation of EMR in the water column, but mostly without validating water column properties. INCOIS-

SATCORE longitudinal data helped to overcome these challenges and provided a solution to remove the turbidity factor to improve bottom detection.

9. Finally, this study has provided an algorithm to generate on-the-fly / real-time recursive bathymetry using Python script. However, converting the same into a console where if one has a satellite image, he can derive SDB is a work in progress and will be continued in the future.

The highlighted contribution of this study is notable, especially during this decade when most marine studies are advocating for the role of SDB in sustainable ocean observing systems, science & technology based ocean management systems, nourishing ocean health, sustainability of the marine ecosystem, and many such other applications (Ryabinin et al., 2019). The last decade has shown a surge in Satellite Derived Bathymetry studies. A study worth mentioning in concluding the thesis is a systematic review of Ledar et al., (2023) which proved the last phase of SDB studies (2017-till date) has majorly focused on improving the accuracy of models to meet the IHO criterion of accuracy. This study has also concluded, irrespective of the substantial advancement of sensors and processing tools, SDB results still do not meet the IHO criterion for accuracy. However, the contribution of this research specially to achieve higher accuracy in highly turbid areas needs to be highlighted.

The scope of SDB, its application areas, and further research to overcome challenges is very large, however, a few of the challenges or limitations and future research avenues are presented below.

6.1 Future Research and Challenges

The research provides several ways ahead for future research which include; examinations of OAS influence on the selection of satellite imagery for SDB, assigning weight to the pixels whose reflectance has been diminished instigated by OAS, and further development of a bio-physio-optical model for SDB by collecting bio-physical parameters data concurrent with satellite pass time. The few of the OAS like Chl are derived from HS data of ocean color sensors which have very coarse spatial resolutions unsuitable to meet requirement for coastal areas. Therefore, validation of such data is very essential prior further use in SDB research. Future

research may focus on utilizing Ocean Color products of various agencies for developing SDB algorithms.

This study advocates incorporating satellite-derived OAS data into SDB retrieval to enhance the ability to retrieve bathymetry. The huge dataset of the study can also be used for creating a LUT database with a Python console for SDB derivation in the future. Achieving the IHO S-44 accuracy level by applying the above methodology may help overcome these limitations of accuracy to utilize SDB as an alternate to MBES and SBES in the near future. The influence of Chlorophyll, TSM, & Turbidity on SDB has been only studied very narrowly in this study limiting the scope of research. However, the application of this OAS data has a vast potential that will be immensely advantageous for future SDB studies.

Ship-borne deployment of SBES, MBES, SONAR, AUVs, and ROVs for hydrographic surveying has grown widely used, providing an effective method to meet IHO criteria. For routine surveys, airborne LiDAR is highly costly and susceptible to environmental variables. Satellite-based approaches, particularly optical SDB, have reduced operational costs but may have resolution and accuracy limitations that must be evaluated in comparison with data from MBES and SBES.

The methodological limitations in the current research include the time lag between in-situ & satellite data. Tidal movement's influence on the nearshore region may severely compromise the accuracy level, at study sites with high tide ranges. These limitations may be resolved with future space missions by obtaining synchronized datasets from all sources. Future studies may focus on these aspects to derive better solutions to the problem of accuracy.

References

- [1] Albert, A., & Mobley, C. D. (2003). An analytical model for subsurface irradiance and remote sensing reflectance in deep and shallow case-2 waters. *Optics Express*, *11*(22), 2873-2890.
- [2] Ashphaq, M., Srivastava, P. K., & Mitra, D. (2021). Review of near-shore satellite derived bathymetry : Classification and account of five decades of coastal bathymetry research. *Journal of Ocean Engineering and Science*, *6*, 340–359.
- [3] Ashphaq, M., Srivastava, P. K., & Mitra, D. (2022). Role of space technology in synoptical analysis of opportunity and challenges for revolutionizing the blue economy. *Journal of East China University of Science And Technology*, *65*(4), 75–91.
<https://doi.org/10.5281/ZENODO.7234787>
- [4] Ashphaq, M., Srivastava, P. K., & Mitra, D. (2022). Evaluation and performance of satellite-derived bathymetry algorithms in turbid coastal water : a case study of Vengurla rocks. *Indian Journal of Geo-Marine Sciences*, *51*(04), 310–321.
- [5] Ashphaq, M., Srivastava, P. K., & Mitra, D. (2022). Analysis of univariate linear , robust-linear , and non-linear machine learning algorithms for satellite-derived bathymetry in complex coastal terrain. *Regional Studies in Marine Science*, *56*(102678).
<https://doi.org/10.1016/j.rsma.2022.102678>
- [6] Ashphaq, M., Srivastava, P. K., & Mitra, D. (2023). Preliminary examination of influence of Chlorophyll , Total Suspended Material , and Turbidity on Satellite Derived-Bathymetry estimation in coastal turbid water. *Regional Studies in Marine Science*, *62*, 102920. <https://doi.org/10.1016/j.rsma.2023.102920>
- [7] Avinash, K., Jena, B., Vinaya, M. S., Jayappa, K. S., Narayana, A. C., & Bhat, H. G. (2012). Regionally tuned algorithm to study the seasonal variation of suspended sediment concentration using IRS-P4 Ocean Colour Monitor data. *The Egyptian Journal of Remote Sensing and Space Sciences*, *15*(1), 67–81.
<https://doi.org/10.1016/j.ejrs.2012.05.003>
- [8] Bramante, J. F., Raju, D. K., & Sin, T. M. (2013). Multispectral derivation of bathymetry in Singapore 's shallow , turbid waters. *International Journal of Remote Sensing*, *34*(6), 2070–2088. <https://doi.org/10.1080/01431161.2012.734934>

- [9] Caballero, I., & Stumpf, R. P. (2023). Confronting turbidity, the major challenge for satellite-derived coastal bathymetry. *Science of the Total Environment*, 870. <https://doi.org/10.1016/j.scitotenv.2023.161898>
- [10] Caballero, I., & Stumpf, R. P. (2020). Towards Routine Mapping of Shallow Bathymetry in Environments with Variable Turbidity : Contribution of Sentinel-2A / B Satellites Mission. *Remote Sensing*, 12(451). <https://doi.org/10.3390/rs12030451>
- [11] Caballero, I., Stumpf, R. P., & Meredith, A. (2019). Preliminary Assessment of Turbidity and Chlorophyll Impact on Bathymetry Derived from Sentinel-2A and Sentinel-3A Satellites in South Florida. *Remote Sensing*, 11(645), 1–20. <https://doi.org/10.3390/rs11060645>
- [12] Chakraborty, K., Gupta, A., Lotliker, A. A., & Tilstone, G. (2016). Estuarine , Coastal and Shelf Science Evaluation of model simulated and MODIS-Aqua retrieved sea surface chlorophyll in the eastern Arabian Sea. *Estuarine, Coastal and Shelf Science*, 181, 61–69. <https://doi.org/10.1016/j.ecss.2016.08.002>
- [13] Dogliotti, A. I., Ruddick, K. G., Nechad, B., Doxaran, D., & Knaeps, E. (2015). A single algorithm to retrieve turbidity from remotely-sensed data in all coastal and estuarine waters. *Remote Sensing of Environment*, 156, 157–168. <https://doi.org/10.1016/j.rse.2014.09.020>
- [14] Favoretto, F., Morel, Y., Waddington, A., Lopez-calderon, J., Cadena-, M., Cadena-roa, M., ... Blanco-jarvio, A. (2017). 4SM Method Tested in the Gulf of California Suggests Field Data are Not Needed to Derive Satellite Bathymetry. *Sensors*, 17(2248), 1–23. <https://doi.org/10.3390/s17102248>
- [15] Forfinski-sarkozi, N. A., & Parrish, C. E. (2016). Analysis of MABEL Bathymetry in Keweenaw Bay and Implications for ICESat-2 ATLAS. *Remote Sensing*, 8(772). <https://doi.org/10.3390/rs8090772>
- [16] Gao, J. (2009). Bathymetric mapping by means of remote sensing : methods , accuracy and limitations. *Progress in Physical Geography*, 33(1), 103–116. <https://doi.org/10.1177/0309133309105657>
- [17] Gujarati, D. N. (2022). *Basic econometrics*. Prentice Hall
- [18] Hassan, H. M., & Nadaoka, K. (2017). Assessment of machine learning approaches for bathymetry mapping in shallow water environments using multispectral satellite images.

International Journal of GeoInformatics, 13(2).

- [19] Hernandez, W. J., & Armstrong, R. A. (2016). Deriving Bathymetry from Multispectral Remote Sensing Data. *Journal of Marine Science and Engineering*, 4(8).
<https://doi.org/10.3390/jmse4010008>
- [20] Jawak, S. D., & Luis, A. J. (2015). Spectral information analysis for the semiautomatic derivation of shallow lake bathymetry using high-resolution multispectral imagery : A case study of Antarctic coastal oasis. *Aquatic Procedia*, 4(Icwrcoe), 1331–1338.
<https://doi.org/10.1016/j.aqpro.2015.02.173>
- [21] Kerr, J. M., & Purkis, S. (2018). An algorithm for optically-deriving water depth from multispectral imagery in coral reef landscapes in the absence of ground-truth data. *Remote Sensing of Environment*, 210(March), 307–324.
<https://doi.org/10.1016/j.rse.2018.03.024>
- [22] Lecours, V., Disney, M., He, K., Pettorelli, N., Rowcliffe, J. M., Sankey, T., & Scales, K. (2021). Remote sensing and the UN Ocean Decade : high expectations , big opportunities. *Remote Sensing in Ecology and Conservation*, 1–5.
<https://doi.org/10.1002/rse2.241>
- [23] Ledar, T. D., Baucic, M., Leder, N., & Gilic, F. (2023). Optical Satellite-Derived Bathymetry : An Overview and WoS and Scopus Bibliometric Analysis. *Remote Sensing*, 15(1294), 1–25. <https://doi.org/10.3390/rs15051294>
- [24] Lee, Z., Carder, K. L., Mobley, C. D., Steward, R. G., & Patch, J. S. (1999). Hyperspectral remote sensing for shallow waters : 2 . Deriving bottom depths and water properties by optimization. *Applied Optics*, 38(18), 3831–3843.
- [25] Lyzenga, D. R. (1978). Passive remote sensing techniques for mapping water depth and bottom features. *Applied Optics*, 17(3), 379–383.
- [26] Lyzenga, D. R., Malinas, N. P., & Tanis, F. J. (2006). Multispectral Bathymetry Using a Simple Physically Based Algorithm. *IEEE Transactions on Geoscience and Remote Sensing*, 44(8), 2251–2259.
- [27] Mahabal, A., Pande, S., Sharma, R. M., & Pednekar, S. N. (2007). *Status Survey of Indian Edible-nest Swiftlet Colloalia unicolor (Jerdon) in Western Ghats , West Coast and Islands in Arabian Sea , India*. Kolkala.
- [28] Masita Dwi Manessa, M., Haidar, M., Hastuti, M., Kresnawati, D. K. K., Dwi, M.,

- Manessa, M., ... Sensing, R. (2017). Determination of the best methodology for bathymetry mapping using spot 6 imagery : a study of 12 empirical. *International Journal of Remote Sensing and Earth Sciences*, 14(2), 127–136.
- [29] McKinna, L. I. W., Fearn, P. R. C., Weeks, S. J., Werdell, P. J., Reichstetter, M., Franz, B. A., ... Feldman, G. C. (2015). A semianalytical ocean color inversion algorithm with explicit water column depth and substrate reflectance parameterization. *JOURNAL OF GEOPHYSICAL RESEARCH:Oceans*, 120, 1741–1770.
<https://doi.org/10.1002/2014JC010224>.Received
- [30] Minu, P., Lotliker, A. A., Shaju, S. S., & Santhoshkumar, B. (2014). Effect of optically active substances and atmospheric correction schemes on remote-sensing reflectance at a coastal site off Kochi. *International Journal of Remote Sensing*, (September), 37–41.
<https://doi.org/10.1080/01431161.2014.926420>
- [31] Moeinkhah, A., Shakiba, A., & Azarakhsh, Z. (2019). Assessment of Regression and Classification Methods Using Remote Sensing Technology for Detection of Coastal Depth (Case Study of Bushehr Port and Kharg Island). *Journal of the Indian Society of Remote Sensing*, 0123456789. <https://doi.org/10.1007/s12524-019-00959-x>
- [32] Molkov, A. A., Fedorov, S. V, Pelevin, V. V, & Korchemkina, E. N. (2019). Regional Models for High-Resolution Retrieval of Chlorophyll a and TSM Concentrations in the Gorky Reservoir by Sentinel-2 Imagery. *Remote Sensing*, 11(1215), 2–29.
- [33] Morales, J., Stuart, V., Platt, T., & Sathyendranath, S. (2011). *Handbook of Satellite Remote Sensing Image Interpretation : Applications for Marine Living Resources Conservation and Management*. Dartmouth, Canada: EU PRESPO and IOCCG.
- [34] Nechad, B., Ruddick, K. G., & Neukermans, G. (2009). Calibration and validation of a generic multisensor algorithm for mapping of turbidity in coastal waters. In *SPIE European International Symposium on Remote Sensing, Berlin* (Vol. 7473, pp. 1–11).
<https://doi.org/10.1117/12.830700>
- [35] Nechad, B., Ruddick, K. G., & Park, Y. (2010). Calibration and validation of a generic multisensor algorithm for mapping of total suspended matter in turbid waters. *Remote Sensing of Environment*, 114(4), 854–866. <https://doi.org/10.1016/j.rse.2009.11.022>
- [36] Noujas, V., Kankara, R. S., & Selvan, S. C. (2019). Shoreline management plan for embayed beaches : A case study at Vengurla , west coast of India. *Ocean and Coastal*

- Management*, 170(January), 51–59. <https://doi.org/10.1016/j.ocecoaman.2019.01.001>
- [37] Parrish, C. E., Magruder, L. A., Neuenschwander, A. L., Forfinski-sarkozi, N., Alonzo, M., & Jasinski, M. (2019). Validation of ICESat-2 ATLAS Bathymetry and Analysis of ATLAS ' s Bathymetric Mapping Performance. *Remote Sensing*, 11(1630), 2–19.
- [38] Pe'eri, S., Azuike, C., & Parrish, C. (2013). Satellite-Derived Bathymetry a Reconnaissance Tool for Hydrography. *Hydro International*, 16-19
- [39] Purkis, S., & Chirayath, V. (2022). Remote Sensing the Ocean Biosphere. *Annual Review OfEnvironment and Resources*, 47, 823–847.
- [40] Pushparaj, J., & Hegde, A. V. (2017). Estimation of bathymetry along the coast of Mangaluru using Landsat-8 imagery. *The International Journal of Ocean and Climate Systems*, 8(2), 71–83. <https://doi.org/10.1177/1759313116679672>
- [41] Quilleuc, A. Le, Collin, A., Jasinski, M. F., & Devillers, R. (2022). Very High-Resolution Satellite-Derived Bathymetry and Habitat Mapping Using Pleiades-1 and ICESat-2. *Remote Sensing*, 14(133), 2–23.
- [42] Ryabinin, V., Barbière, J., Haugan, P., Kullenberg, G., Smith, N., McLean, C., Troisi, A., Fischer, A. S., Aricò, S., Aarup, T., Pissierssens, P., Visbeck, M., Enevoldsen, H., & Rigaud, J. (2019). The UN decade of ocean science for sustainable development. *Frontiers in Marine Science*, 6(JUL). <https://doi.org/10.3389/fmars.2019.00470>
- [43] Saberioona, M., Brom, J., Nedbal, V., Souček, P., & Císař, P. (2020). Chlorophyll-a and Total Suspended Solids Retrieving and Mapping Using Sentinel-2 and Machine learning for Inland Waters. *Ecological Indicators*, 113(106236). <https://doi.org/10.1016/j.ecolind.2020.106236>
- [44] Sagawa, T., Yamashita, Y., Okumura, T., & Yamanokuchi, T. (2019). Satellite Derived Bathymetry Using Machine Learning and Multi-Temporal Satellite Images. *Remote Sensing*, 11(115), 2–19.
- [45] Salameh, E., Almar, R., Baptista, P., Heygster, G., Lubac, B., Raucoules, D., ... Laignel, B. (2019). Monitoring Beach Topography and Nearshore Bathymetry Using Spaceborne Remote Sensing : A Review. *Remote Sensing*, 11(2212), 1–32.
- [46] Sindhu, B., Suresh, I., Unnikrishnan, A. S., Bhatkar, N. V., Neetu, S., & Michael, G. S. (2007). Improved bathymetric datasets for the shallow water. *Journal of Earth System Science*, 116(3), 261–274.

- [47] Smith, W. H. F., & Sandwell, D. T. (2004). Conventional bathymetry, bathymetry from space, and geodetic altimetry. *Oceanography*, 17(SPL.ISS. 1), 8–23. <https://doi.org/10.5670/oceanog.2004.63>
- [48] Stumpf, R. P., Holderied, K., & Sinclair, M. (2003). Determination of water depth with high-resolution satellite imagery over variable bottom types. *Limnol. Oceanogr.*, 48, 547–556.
- [49] Thomas, N., Pertiwi, A. P., Traganos, D., Lagomasino, D., Poursanidis, D., Moreno, S., & Fatoyinbo, L. (2021). Space-Borne Cloud-Native Satellite-Derived Bathymetry (SDB) Models Using ICESat-2 And Sentinel-2. *GEOPHYSICAL RESEARCH LETTERS*, 48, 1–11. <https://doi.org/10.1029/2020GL092170>
- [50] Wang, C., Chen, S., Li, D., Wang, D., Liu, W., & Yang, J. (2017). A Landsat-based model for retrieving total suspended solids concentration of estuaries and coasts in China. *Geosci. Model Dev*, 10, 4347–4365.
- [51] Wang, Z., Kawamura, K., Sakuno, Y., Fan, X., Gong, Z., & Lim, J. (2017). Retrieval of Chlorophyll- a and Total Suspended Solids Using Iterative Stepwise Elimination Partial Least Squares (ISE-PLS) Regression Based on Field Hyperspectral Measurements in Irrigation Ponds in. *Remote Sensing*, 9(264), 1–14. <https://doi.org/10.3390/rs9030264>
- [52] Wiehle, S., Pleskachevsky, A., & Gebhardt, C. (2019). Automatic bathymetry retrieval from SAR images. *CEAS Space Journal*, 0(0), 0. <https://doi.org/10.1007/s12567-018-0234-4>

APPENDICES

APPENDIX - I: TABLES

Table 1: Details of SATCORE and Satellite Data used in the study

SATCORE Date	SATCORE Parameters	Satellite data	Sun Elevation	Earth-sun distance-AU
02 Nov 2011	Chl Diss_Oxy, TSM, Nitrite, Phosphate, PH, Silicate	L7_ETM_2011_11_10_05_05_10_144053_L2R	54.55864699	0.9903976
14 Feb 2012	Chl, Diss_Oxy, TSM, Nitrite, Silicate	L7_ETM_2012_02_14_05_05_47_144053_L2R	51.79002560	0.9874163
04 Dec 2013	Chl, Temp, TSM, Turbidity, Ammonium, Silicate, Nano, Pico, Micro	L7_ETM_2013_12_17_05_07_59_144053_L2R	48.32981249	0.9840676
13 Feb 2015	CHL, Diss_Oxy, Salinity, Temp, TSM, Turbidity, Fdom, Ammonium Nitrite, Phosphate, PH, Silicate, Aot500, Nano, Pico	L8_OLI_2015_02_14_05_11_28_144053_L2R	52.98206259	0.9874698
16 Dec 2015	Chl, Diss_Oxy, Salinity, Temp, TSM, Turbidity, Fdom, Ammonium, Nitrite, Phosphate, Silicate, Nano Pico	L7_ETM_2015_12_23_05_12_57_144053_L2R	52.98206259	0.9874698
Units for SATCORE data: Total Chl- mg/m**3, TSM-mg/l, Particle size micro(20µm), nano(2µm), pico(0.2µm)-mg/m**3, Salinity-PSU, Dissolved Oxygen- mg/l, Turbidity-NTU, Dissolved Fluorescence (Fdom)- QSE, Nitrite, Nitrate, Phosphate, Silicate, Ammonium- mol/l,				

Table 2: Descriptive Statistics

16 Dec 2015 (Landsat-7), n=264																						
n=264	Bathy	B1_479	B2_561	B3_661	B4_835		CHL	DISS_OXY	Salinity	Temp	TSM	Turbidity	Fdom	Ammonium	Nitrite	Phosphate	PH	Silicate	AOT500	Micro	Nano	Pico
Mean	19.556	0.043	0.036	0.022	0.014	X	1.876	3.859	34.480	X	25.793	1.202	1.512	2.738	2.307	0.269	8.167	14.716	X	0.533	0.601	0.350
SE	0.597	0.000	0.001	0.000	0.000	X	0.033	0.007	0.038	X	0.199	0.018	0.014	0.029	0.016	0.003	0.001	0.131	X	0.004	0.014	0.005
Median	18.700	0.041	0.030	0.021	0.012	X	1.865	3.883	34.676	X	25.874	1.159	1.493	2.739	2.321	0.257	8.166	14.410	X	0.513	0.561	0.340
SD	11.454	0.006	0.011	0.006	0.003	X	0.535	0.117	0.621	X	3.231	0.299	0.231	0.474	0.267	0.044	0.018	2.129	X	0.066	0.232	0.076
Range	41.100	0.032	0.054	0.037	0.020	X	2.754	0.681	3.563	X	13.310	1.536	1.464	2.085	1.497	0.230	0.100	9.550	X	0.325	1.023	0.404
Minimum	0.300	0.033	0.019	0.012	0.008	X	0.830	3.469	32.158	X	19.822	0.693	1.042	1.665	1.678	0.184	8.119	11.125	X	0.395	0.212	0.213
Maximum	41.400	0.065	0.074	0.049	0.028	X	3.584	4.150	35.721	X	33.132	2.228	2.506	3.750	3.175	0.415	8.219	20.674	X	0.720	1.235	0.618
13 Feb 2015 (Landsat-8), n=368																						
n=368	Bathy	B1_443	B2_483	B3_561	B4_655	B5_865	CHL	DISS_OXY	Salinity	Temp	TSM	Turbidity	Fdom	Ammonium	Nitrite	Phosphate	PH	Silicate	AOT500	Micro	Nano	Pico
Mean	19.556	0.075	0.074	0.062	0.043	0.026	1.062	6.507	34.699	28.993	3.428	3.337	0.850	9.442	0.056	2.073	8.191	7.360	1.178	0.386	0.845	0.012
SE	0.597	0.000	0.001	0.001	0.001	0.000	0.014	0.013	0.017	0.015	0.031	0.039	0.007	0.117	0.001	0.034	0.002	0.061	0.002	0.007	0.011	0.000
Median	18.700	0.074	0.073	0.057	0.039	0.024	0.996	6.572	34.802	29.083	3.351	3.275	0.818	9.216	0.053	1.969	8.180	6.973	1.166	0.373	0.815	0.012
SD	11.454	0.008	0.010	0.017	0.012	0.007	0.262	0.246	0.321	0.291	0.597	0.739	0.141	2.241	0.019	0.645	0.036	1.166	0.040	0.127	0.213	0.002
Range	41.100	0.037	0.044	0.064	0.052	0.045	1.207	1.549	1.638	1.358	2.515	3.392	0.630	14.910	0.105	2.811	0.151	5.738	0.215	0.673	0.989	0.010
Minimum	0.300	0.060	0.056	0.039	0.027	0.015	0.583	5.783	33.557	28.293	1.990	1.426	0.633	2.413	0.021	0.799	8.131	4.827	1.085	0.127	0.453	0.007
Maximum	41.400	0.097	0.100	0.103	0.079	0.060	1.791	7.333	35.195	29.652	4.505	4.818	1.263	17.322	0.126	3.611	8.282	10.565	1.300	0.800	1.442	0.017
02 Dec 2013 (Landsat-7), n=434																						
n=434	Bathy	B1_479	B2_561	B3_661	B4_835		CHL	DISS_OXY	Salinity	Temp	TSM	Turbidity	Fdom	Ammonium	Nitrite	Phosphate	PH	Silicate	AOT500	Micro	Nano	Pico
Mean	17.719	0.049	0.049	0.032	0.019	X	2.357	X	X	25.940	5.985	5.751	X	4.006	X	X		8.052	X	0.875	1.185	0.128
SE	0.555	0.000	0.001	0.000	0.000	X	0.040	X	X	0.012	0.106	0.105	X	0.052	X	X		0.078	X	0.012	0.012	0.004
Median	14.700	0.049	0.048	0.029	0.020	X	2.046	X	X	26.002	5.594	5.452	X	3.797	X	X		8.088	X	0.787	1.157	0.094
SD	11.572	0.006	0.012	0.008	0.003	X	0.824	X	X	0.256	2.209	2.186	X	1.077	X	X		1.615	X	0.245	0.245	0.091
Range	41.100	0.026	0.045	0.040	0.016	X	4.658	X	X	1.155	11.454	11.877	X	6.152	X	X		8.276	X	1.316	1.850	0.397
Minimum	0.300	0.036	0.032	0.021	0.016	X	0.782	X	X	25.343	2.186	1.297	X	1.094	X	X		4.337	X	0.490	0.343	0.031
Maximum	41.400	0.062	0.077	0.061	0.032	X	5.440	X	X	26.498	13.640	13.174	X	7.246	X	X		12.613	X	1.806	2.193	0.428
14 Feb 2012 (Landsat-7), n=435																						
n=435	Bathy	B1_479	B2_561	B3_661	B4_835		CHL	DISS_OXY	Salinity	Temp	TSM	Turbidity	Fdom	Ammonium	Nitrite	Phosphate	PH	Silicate	AOT500	Micro	Nano	Pico
Mean	17.677	0.051	0.047	0.032	0.019	X	0.443	4.315	X	X	12.219	X	X	X	0.537	X	X	7.522	X	X	X	X
SE	0.556	0.000	0.001	0.000	0.000	X	0.005	0.010	X	X	0.159	X	X	X	0.006	X	X	0.087	X	X	X	X
Median	14.600	0.050	0.047	0.030	0.019	X	0.416	4.359	X	X	11.341	X	X	X	0.494	X	X	6.979	X	X	X	X
SD	11.591	0.006	0.013	0.009	0.004	X	0.102	0.202	X	X	3.324	X	X	X	0.130	X	X	1.818	X	X	X	X
Range	41.800	0.040	0.059	0.047	0.031	X	0.674	1.324	X	X	22.712	X	X	X	0.813	X	X	11.389	X	X	X	X
Minimum	-0.400	0.039	0.026	0.017	0.011	X	0.117	3.369	X	X	6.011	X	X	X	0.305	X	X	3.566	X	X	X	X
Maximum	41.400	0.079	0.085	0.064	0.042	X	0.791	4.693	X	X	28.723	X	X	X	1.119	X	X	14.955	X	X	X	X
02 Nov 2011 (Landsat-7), n=433																						
n=433	Bathy	B1_479	B2_561	B3_661	B4_835		CHL	DISS_OXY	Salinity	Temp	TSM	Turbidity	Fdom	Ammonium	Nitrite	Phosphate	PH	Silicate	AOT500	Micro	Nano	Pico
Mean	17.709	0.050	0.050	0.035	0.023	X	1.168	4.668	X	X	X	5.698	X	X	0.616	0.843	8.403	1.327	X	X	X	X
SE	0.558	0.000	0.000	0.000	0.000	X	0.006	0.009	X	X	X	0.101	X	X	0.006	0.013	0.002	0.031	X	X	X	X
SD	11.603	0.005	0.008	0.006	0.003	X	0.129	0.197	X	X	X	2.091	X	X	0.135	0.269	0.045	0.650	X	X	X	X
Range	41.800	0.023	0.037	0.032	0.019	X	0.893	1.359	X	X	X	8.296	X	X	0.765	2.133	0.274	4.268	X	X	X	X
Minimum	-0.400	0.040	0.034	0.025	0.016	X	0.831	3.731	X	X	X	1.202	X	X	0.217	0.122	8.284	0.597	X	X	X	X
Maximum	41.400	0.063	0.071	0.056	0.035	X	1.723	5.089	X	X	X	9.497	X	X	0.983	2.255	8.558	4.866	X	X	X	X

Table 3:Correlation Matrix

16 Dec 2015	Bathy	B1_479	B2_561	B3_661	B4_835	std_chl	std_DISS_OXY	std_SALINITY	std_TSM	std_TURBIDITY	std_Fdom	std_AMMONIUM	std_NITRITE	std_PHOSPHATE	std_SILICATE	std_NANO	std_PICO
Bathy	1.00																
B1_479	-0.79	1.00															
B2_561	-0.86	0.88	1.00														
B3_661	-0.68	0.78	0.87	1.00													
B4_835	-0.49	0.60	0.67	0.72	1.00												
STD_CHL	-0.35	0.15	0.24	0.16	0.04	1.00											
STD_DISS_OXY	0.19	-0.17	-0.23	-0.16	-0.04	-0.22	1.00										
STD_SALINITY	0.32	-0.12	-0.15	-0.03	0.03	-0.82	0.57	1.00									
STD_TSM	0.29	-0.08	-0.18	-0.15	-0.03	-0.96	0.77	0.76	1.00								
std_TURBIDITY	0.25	-0.32	-0.28	-0.20	-0.14	0.69	-0.76	-0.50	-0.74	1.00							
STD_Fdom	0.32	-0.05	-0.08	0.04	0.03	-0.78	0.61	0.80	0.77	-0.61	1.00						
STD_AMMONIUM	0.00	-0.15	-0.09	-0.04	-0.07	0.79	-0.74	-0.65	-0.87	0.90	-0.70	1.00					
STD_NITRITE	0.21	0.06	0.01	0.08	0.17	-0.79	0.46	0.66	0.84	-0.57	0.70	-0.67	1.00				
STD_PHOSPHATE	-0.09	0.02	0.05	0.05	0.12	0.60	-0.76	-0.42	-0.55	0.62	-0.62	0.51	-0.33	1.00			
STD_SILICATE	-0.04	-0.02	0.06	0.11	0.10	0.81	-0.89	-0.49	-0.81	0.88	-0.62	0.82	-0.53	0.80	1.00		
STD_NANO	-0.07	0.00	0.10	0.14	0.09	0.87	-0.86	-0.58	-0.85	0.87	-0.61	0.83	-0.62	0.71	0.97	1.00	
STD_PICO	-0.30	0.14	0.23	0.16	0.05	0.98	-0.82	-0.77	-0.93	0.71	-0.72	0.74	-0.77	0.60	0.82	0.89	1.00

13 May 2015	Bathy	B1_443	B2_483	B3_561	B4_655	B5_865	std_chl	std_DISS_OXY	std_SALINITY	std_TEMP	std_TSM	std_TURBIDITY	std_Fdom	std_AMMONIUM	std_NITRITE	std_PHOSPHATE	std_SILICATE	std_NANO	std_PICO	
BATHY KOCHI	1.00																			
B1_443	-0.84	1.00																		
B2_483	-0.91	0.99	1.00																	
B3_561	-0.92	0.95	0.98	1.00																
B4_655	-0.86	0.94	0.95	0.98	1.00															
B5_865	-0.81	0.93	0.92	0.93	0.97	1.00														
std_chl	-0.44	0.39	0.43	0.43	0.35	0.25	1.00													
std DISS OXY	0.35	-0.52	-0.48	-0.51	-0.58	-0.60	-0.08	1.00												
std SALINITY	0.62	-0.72	-0.71	-0.76	-0.81	-0.78	-0.37	0.76	1.00											
std TEMP	0.18	-0.40	-0.36	-0.33	-0.35	-0.42	0.19	0.60	0.32	1.00										
std TSM	-0.21	0.43	0.39	0.37	0.40	0.42	0.06	-0.73	-0.46	-0.88	1.00									
std TURBIDITY	-0.27	0.37	0.38	0.34	0.30	0.26	0.71	-0.35	-0.34	-0.42	0.63	1.00								
std Fdom	-0.60	0.77	0.74	0.76	0.81	0.83	0.20	-0.84	-0.88	-0.67	0.71	0.40	1.00							
std AMMONIUM	-0.30	0.41	0.39	0.39	0.39	0.35	0.56	-0.48	-0.52	-0.40	0.63	0.69	0.59	1.00						
std NITRITE	-0.45	0.57	0.55	0.56	0.57	0.51	0.63	-0.65	-0.72	-0.31	0.64	0.67	0.70	0.85	1.00					
std PHOSPHATE	-0.53	0.71	0.68	0.67	0.68	0.70	0.25	-0.69	-0.68	-0.75	0.79	0.52	0.91	0.74	0.73	1.00				
std SILICATE	-0.70	0.75	0.76	0.79	0.79	0.75	0.63	-0.62	-0.91	-0.17	0.34	0.45	0.79	0.59	0.78	0.67	1.00			
std NANO	-0.32	0.29	0.33	0.31	0.23	0.12	0.96	-0.04	-0.24	0.13	0.17	0.79	0.12	0.61	0.65	0.23	0.49	1.00		
std PICO	-0.59	0.68	0.69	0.72	0.72	0.69	0.61	-0.68	-0.87	-0.42	0.52	0.66	0.79	0.60	0.72	0.69	0.86	0.51	1.00	

02 Dec 2013	Bathy	B1_479	B2_561	B3_661	B4_835	STD_CHL	STD_TEMP	STD_TSM	STD_TURBIDITY	STD_AMMONIUM	STD_SILICATE	STD_NANO	STD_PICO	STD_MICRO
Bathy	1.000													
B1_479	-0.781	1.000												
B2_561	-0.909	0.873	1.000											
B3_661	-0.751	0.760	0.883	1.000										
B4_835	-0.339	0.388	0.454	0.661	1.000									
STD_CHL	-0.018	-0.225	-0.107	-0.077	-0.009	1.000								
STD_TEMP	0.199	-0.050	-0.116	-0.078	0.005	-0.762	1.000							
STD_TSM	0.043	-0.017	0.056	0.184	0.269	-0.097	0.182	1.000						
STD_TURBIDITY	0.082	0.001	0.044	0.183	0.280	-0.229	0.293	0.974	1.000					
STD_AMMONIUM	-0.363	0.209	0.291	0.294	0.225	-0.041	0.225	0.106	0.168	1.000				
STD_SILICATE	-0.258	0.090	0.253	0.266	0.162	0.612	-0.508	-0.156	-0.270	-0.184	1.000			
STD_NANO	0.189	-0.269	-0.297	-0.261	-0.086	-0.167	0.439	0.255	0.312	0.547	-0.721	1.000		
STD_PICO	-0.307	0.011	0.217	0.226	0.145	0.806	-0.692	0.136	0.002	0.164	0.724	-0.279	1.000	
STD_MICRO	0.041	-0.278	-0.171	-0.138	-0.054	0.960	-0.744	-0.251	-0.375	0.008	0.624	-0.188	0.777	1.000

14 Dec 2012	Bathy	B1_479	B2_561	B3_661	B4_835	STD_CHL	STD_DISS_OXY	STD_TSM	STD_NITRITE	STD_PHOSPHATE	STD_PH	STD_SILICATE
Bathy	1.000											
B1_479	-0.654	1.000										
B2_561	-0.901	0.849	1.000									
B3_661	-0.750	0.849	0.915	1.000								
B4_835	-0.532	0.805	0.724	0.817	1.000							
STD_CHL	-0.495	0.434	0.557	0.519	0.324	1.000						
STD_DISS_OXY	0.569	-0.584	-0.662	-0.726	-0.572	-0.720	1.000					
STD_TSM	-0.468	0.487	0.542	0.625	0.518	0.584	-0.910	1.000				
STD_NITRITE	-0.553	0.487	0.590	0.613	0.509	0.451	-0.745	0.840	1.000			
STD_SILICATE	-0.643	0.617	0.722	0.742	0.589	0.736	-0.863	0.762	0.847	1.000		

02 NOV 2011	Bathy	B1_479.tif	B2_561.tif	B3_661.tif	B4_835.tif	STD_CHL	STD_DISS_OXY	STD_TSM	STD_NITRITE	STD_PHOSPHATE	STD_PH	STD_SILICATE
Bathy	1.000											
B1_479.tif	-0.372	1.000										
B2_561.tif	-0.603	0.499	1.000									
B3_661.tif	-0.423	0.522	0.876	1.000								
B4_835.tif	-0.242	0.379	0.641	0.753	1.000							
STD_CHL	-0.150	0.295	0.160	0.261	0.365	1.000						
STD_DISS_OXY	0.489	-0.437	-0.377	-0.402	-0.361	-0.573	1.000					
STD_TSM	-0.399	0.081	0.179	0.082	-0.110	-0.349	-0.396	1.000				
STD_NITRITE	-0.560	0.258	0.326	0.206	0.014	0.056	-0.538	0.664	1.000			
STD_PHOSPHATE	-0.187	0.358	0.205	0.277	0.295	0.751	-0.801	0.051	0.385	1.000		
STD_PH	0.080	-0.026	-0.115	-0.175	-0.229	-0.297	0.481	-0.023	0.201	-0.377	1.000	
STD_SILICATE	-0.492	0.453	0.373	0.395	0.351	0.746	-0.898	0.288	0.551	0.807	-0.298	1.000

Table 4: Regression Modelling of all the parameters with Bathymetry

16 Dec 2015	B1_479	B2_561	B3_661	B4_835	std_chl	std_DISS_OXY	std_SALI_NITY	std_TEMP	std_TSM	std_TURBI_DITY	std_Fdom	std_AM-MONIUM	std_NIT-RITE	std_PHOS-PHATE	std_SILI-CATE	std_NANO	std_PICO	
Linear	.630	.738	.465	.241	.119	.035	.102	X	.085	.063	.105	.000	.042	.008	.002	.005	.090	
Logarithmic	.638	.778	.495	.231	.167	.001	.071	X	.000	.001	.004	.016	.000	.072	.065	.	.140	
Inverse	.634	.787	.484	.200	.133	.001	.039	X	.041	.042	.037	.033	.022	.102	.103	.	.095	
Quadratic	.644	.794	.508	.242	.208	.139	.131	X	.225	.121	.270	.071	.210	.143	.182	.156	.183	
Cubic	.646	.794	.513	.259	.222	.143	.167	X	.261	.151	.357	.084	.214	.155	.182	.164	.187	
Compound	.676	.845	.721	.435	.092	.050	.048	X	.061	.048	.028	.000	.006	.012	.014	.026	.078	
Power	.652	.810	.675	.378	.121	.003	.028	X	.000	.001	.018	.014	.012	.061	.072	.	.104	
S	.619	.753	.593	.301	.088	.001	.016	X	.037	.029	.034	.027	.031	.069	.073	.	.062	
Growth	.676	.845	.721	.435	.092	.050	.048	X	.061	.048	.028	.000	.006	.012	.014	.026	.078	
Exponential	.676	.845	.721	.435	.092	.050	.048	X	.061	.048	.028	.000	.006	.012	.014	.026	.078	
Dependent variable is Bathy_Kochi																		
Relationships are significant (less than 0.05) at 95 % Confidence Level																		

13 May 2015	B1_443	B2_483	B3_561	B4_655	B5_865	std_chl	std_DISS_OXY	std_SALI_NITY	std_TEMP	std_TSM	std_TURBI_DITY	std_Fdom	std_AM-MONIUM	std_NIT-RITE	std_PHOS-PHATE	std_SILI-CATE	std_NANO	std_PICO
Linear	0.70	0.82	0.85	0.74	0.66	0.19	0.13	0.38	0.03	0.04	0.08	0.36	0.09	0.20	0.28	0.48	0.11	0.35
Logarithmic	0.69	0.81	0.86	0.76	0.70	x	x	x	x	x	x	x	x	x	x	x	x	x
Inverse	0.68	0.80	0.84	0.76	0.69	x	x	x	x	x	x	x	x	x	x	x	x	x
Quadratic	0.70	0.82	0.86	0.77	0.71	0.22	0.30	0.39	0.12	0.23	0.10	0.37	0.21	0.22	0.34	0.50	0.15	0.37
Cubic	0.70	0.82	0.86	0.77	0.71	0.29	0.31	0.45	0.13	0.23	0.10	0.43	0.27	0.32	0.44	0.63	0.20	0.45
Compound	0.66	0.75	0.84	0.83	0.79	0.17	0.21	0.57	0.03	0.05	0.06	0.46	0.10	0.23	0.27	0.59	0.07	0.45
Power	0.64	0.72	0.80	0.80	0.75	x	x	x	x	x	x	x	x	x	x	x	x	x
S	0.62	0.69	0.75	0.74	0.68	x	x	x	x	x	x	x	x	x	x	x	x	x
Growth	0.66	0.75	0.84	0.83	0.79	0.17	0.21	0.57	0.03	0.05	0.06	0.46	0.10	0.23	0.27	0.59	0.07	0.45
Exponential	0.66	0.75	0.84	0.83	0.79	0.17	0.21	0.57	0.03	0.05	0.06	0.46	0.10	0.23	0.27	0.59	0.07	0.45
Dependent variable is Bathy_Kochi																		
Relationships are significant (less than 0.05) at 95 % Confidence Level																		

02 Dec 2013	B1_479	B2_561	B3_661	B4_835	std_chl	std_DISS_OXY	std_SALI_NITY	std_TEMP	std_TSM	std_TURBI_DITY	std_Fdom	std_AM-MONIUM	std_NIT-RITE	std_PHOS-PHATE	std_SILI-CATE	std_NANO	std_PICO	Std_Micro
Linear	.630	.826	.564	.115	.000	X	X	.040	.002	.007	X	.132	X	X	.066	.036	.094	.002
Logarithmic	.638	.851	.609	.103	.053	X	X	.000	.029	.021	X	.193	X	X	.128	.029	.096	.035
Inverse	.634	.858	.624	.090	.070	X	X	.004	.048	.039	X	.094	X	X	.073	.061	.065	.069
Quadratic	.644	.859	.644	.124	.138	X	X	.194	.116	.155	X	.391	X	X	.269	.291	.095	.131
Cubic	.646	.859	.646	.123	.169	X	X	.202	.123	.155	X	.443	X	X	.301	.332	.132	.179
Compound	.676	.781	.758	.343	.002	X	X	.025	.023	.016	X	.155	X	X	.108	.043	.127	.001
Power	.652	.748	.742	.302	.049	X	X	.003	.056	.048	X	.151	X	X	.115	.015	.099	.033
S	.619	.707	.696	.259	.049	X	X	.007	.037	.035	X	.059	X	X	.052	.039	.045	.048
Growth	.676	.781	.758	.343	.002	X	X	.025	.023	.016	X	.155	X	X	.108	.043	.127	.001
Exponential	.676	.781	.758	.343	.002	X	X	.025	.023	.016	X	.155	X	X	.108	.043	.127	.001
Dependent variable is Bathy_Kochi																		
Relationships are significant (less than 0.05) at 95 % Confidence Level																		

14 Dec 2012	B1_479	B2_561	B3_661	B4_835	std_chl	std_DISS_OXY	std_SALI_NITY	std_TEMP	std_TSM	std_TURBI_DITY	std_Fdom	std_AM-MONIUM	std_NIT-RITE	std_PHOS-PHATE	std_SILI-CATE	std_NANO	std_PICO	Std_Micro
Linear	.427	.811	.562	.283	.245	.323	X	X	.219	X	X	.132	.306	X	.413	X	X	X
Logarithmic	.444	.848	.629	.310	.118	.154	X	X	.071	X	X	.193	.127	X	.202	X	X	X
Inverse	.454	.848	.663	.321	.014	.025	X	X	.000	X	X	.094	.001	X	.020	X	X	X
Quadratic	.461	.863	.669	.319	.250	.326	X	X	.224	X	X	.391	.309	X	.414	X	X	X
Cubic	.461	.865	.675	.321	.314	.402	X	X	.349	X	X	.443	.450	X	.495	X	X	X
Compound	.477	.778	.702	.440	.252	.549	X	X	.458	X	X	.155	.528	X	.596	X	X	X
Power	.473	.744	.700	.435	.124	.334	X	X	.196	X	X	.151	.258	X	.260	X	X	X
S	.463	.685	.662	.406	.016	.086	X	X	.009	X	X	.059	.017	X	.027	X	X	X
Growth	.477	.778	.702	.440	.252	.549	X	X	.458	X	X	.155	.528	X	.596	X	X	X
Exponential	.477	.778	.702	.440	.252	.549	X	X	.458	X	X	.155	.528	X	.596	X	X	X
Dependent variable is Bathy_Kochi																		
Relationships are significant (less than 0.05) at 95 % Confidence Level																		

02 NOV 2011	B1_479	B2_561	B3_661	B4_835	std_chl	std_DISS_OXY	std_SALI_NITY	std_TEMP	std_TSM	std_TURBI_DITY	std_Fdom	std_AM-MONIUM	std_NIT-RITE	std_PHOS-PHATE	std_SILI-CATE	std_NANO	std_PICO	Std_Micro
Linear	.138	.363	.179	.058	.022	.239	X	X	.159	X	X	X	.314	.035	.242	X	X	X
Logarithmic	.130	.395	.190	.055	.001	.077	X	X	.177	X	X	X	.254	.000	.173	X	X	X
Inverse	.121	.418	.197	.051	.000	.005	X	X	.072	X	X	X	.112	.001	.007	X	X	X
Quadratic	.174	.442	.196	.062	.085	.265	X	X	.215	X	X	X	.383	.112	.300	X	X	X
Cubic	.173	.442	.196	.061	.119	.286	X	X	.235	X	X	X	.384	.215	.310	X	X	X
Compound							X	X		X	X	X				X	X	X
Power							X	X		X	X	X				X	X	X
S																		
Growth																		
Exponential																		
Dependent variable is Bathy_Kochi																		
Relationships are significant (less than 0.05) at 95 % Confidence Level																		

Table 5: Predictor Analysis (Response is Bathymetry)

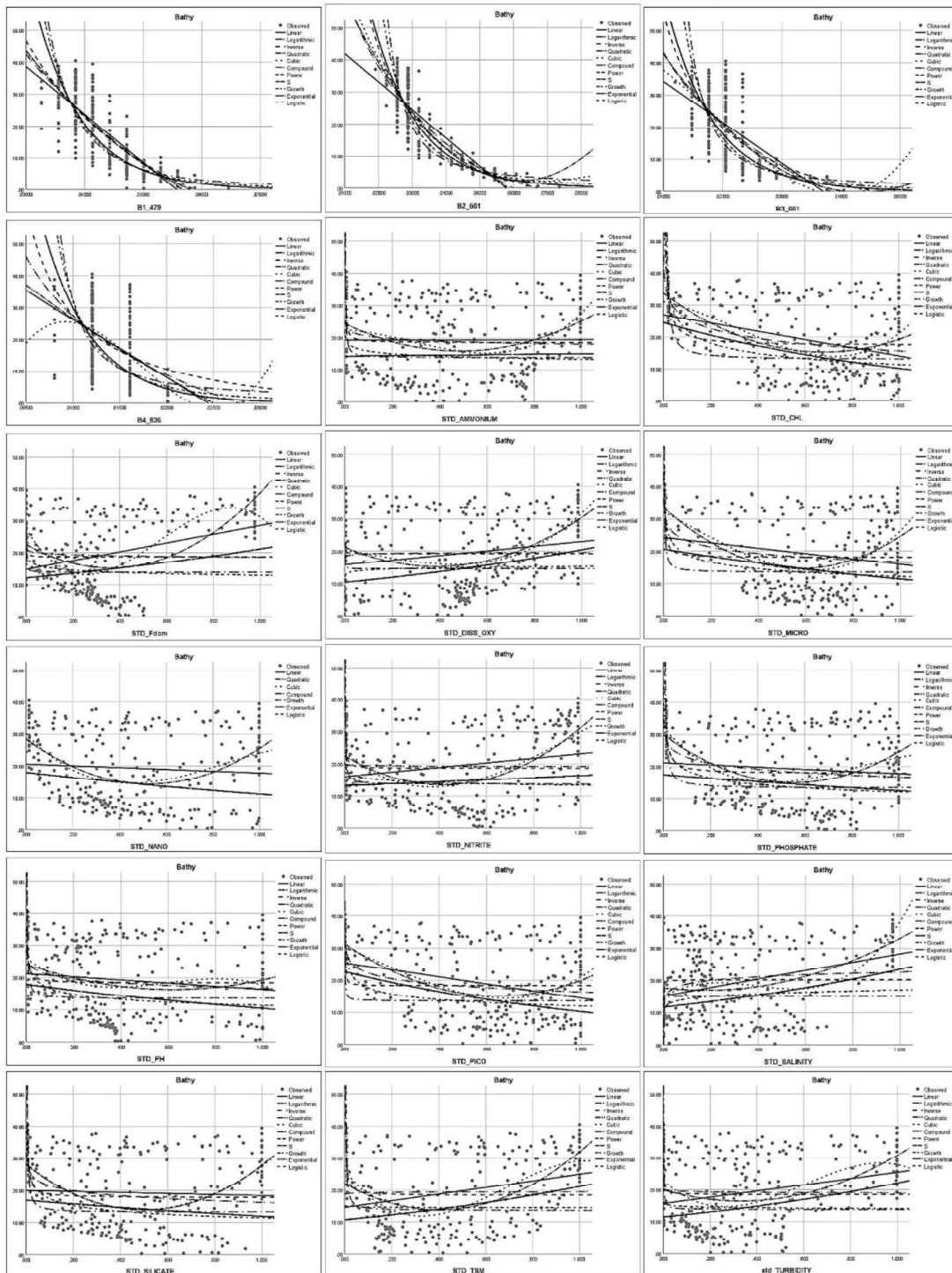
16 Dec 2015 (Landsat-7)																						
Vars	R-Sq	R-Sq (adj)	R-Sq (pred)	Mallows Cp	S	B1_479	B2_561	B3_661	B4_835	std_chl	std_DISS_OXY	std_SALI_NITY	std_TSM	std_TURBI_DITY	std_Fdom	std_AM-MONIUM	std_NIT-RITE	std_PHOS_PHATE	std_SILI_CATE	std_NANO	std_PICO	
1	73.8	73.7	73.4	678.8	5.9408		X															
3	85.1	84.9	84.5	280.2	4.5052		X				X				X							
4	87.3	87.1	86.8	202.3	4.1636		X							X	X			X				
4	87.2	87.0	86.6	204.7	4.1744		X							X	X	X						
6	88.4	88.1	87.7	165.9	3.9898		X					X		X	X	X						X
6	88.4	88.1	87.8	166.8	3.9943		X			X				X	X	X						X
7	89.7	89.4	89.0	121.2	3.7662		X			X		X		X	X	X						X
7	89.6	89.3	89.0	124.4	3.7822		X					X	X	X	X	X						X
8	91.1	90.8	90.4	74.1	3.5130		X			X		X	X	X	X	X						X
9	91.7	91.4	91.0	54.7	3.3999		X			X		X	X	X	X	X					X	X
10	92.4	92.1	91.7	31.9	3.2615		X			X		X	X	X	X	X	X			X		X
10	92.3	92.0	91.6	34.8	3.2789		X			X		X	X	X	X	X	X	X				X
11	92.7	92.4	92.0	23.3	3.2040	X	X			X		X	X	X	X	X	X		X			X
11	92.7	92.3	91.9	24.2	3.2091		X			X		X	X	X	X	X	X	X			X	X
12	93.0	92.6	92.2	14.4	3.1430	X	X			X		X	X	X	X	X	X	X			X	X
13	93.0	92.7	92.2	14.1	3.1346	X	X			X		X	X	X	X	X	X	X			X	X
13 Feb 2015 (Landsat-8)																						
Vars	R-Sq	R-Sq (adj)	R-Sq (pred)	Mallows Cp	S	B1_479	B2_561	B3_661	B4_835	std_chl	std_DISS_OXY	std_SALI_NITY	std_TSM	std_TURBI_DITY	std_Fdom	std_AM-MONIUM	std_NIT-RITE	std_PHOS_PHATE	std_SILI_CATE	std_NANO	std_PICO	
1	84.9	84.8	84.7	889.3	4.5				X													
1	82.1	82.1	82.0	1113.7	4.8		X															
2	93.4	93.4	93.3	183.1	2.9	X	X															
2	94.5	94.4	94.3	99.7	2.7	X	X															
3	93.9	93.9	93.8	141.5	2.8	X	X							X								
3	94.6	94.6	94.4	88.4	2.7	X	X												X			
5	94.9	94.8	94.7	70.5	2.6	X	X											X			X	
5	95.0	94.9	94.8	62.9	2.6	X	X		X	X								X				
5	95.1	95.0	94.8	55.4	2.6	X	X				X							X				
6	95.1	95.0	94.9	55.9	2.6	X	X		X	X								X				
6	95.2	95.1	95.0	47.4	2.5	X	X	X	X	X			X					X				
6	95.2	95.1	94.9	47.5	2.5	X	X		X	X								X				
6	95.4	95.3	95.1	33.6	2.5	X	X	X	X						X							X
7	95.3	95.2	95.1	37.4	2.5	X	X	X	X	X		X			X							
7	95.5	95.4	95.2	27.4	2.5	X	X	X	X	X				X				X				
8	95.7	95.5	95.4	15.5	2.4	X	X	X	X	X	X	X			X							
8	95.6	95.5	95.3	16.9	2.4	X	X	X	X	X		X			X							
11	95.8	95.6	95.4	13.5	2.4	X	X	X	X	X	X	X			X		X	X			X	X
12	95.8	95.6	95.3	13.3	2.4	X	X	X	X	X	X	X		X	X	X	X	X			X	X
02 Dec 2013 (Landsat-7)																						
Vars	R-Sq	R-Sq (adj)	R-Sq (pred)	Mallows Cp	S	B1_479	B2_561	B3_661	B4_835	std_chl	std_DISS_OXY	std_SALI_NITY	std_TSM	std_TURBI_DITY	std_Fdom	std_AM-MONIUM	std_NIT-RITE	std_PHOS_PHATE	std_SILI_CATE	std_NANO	std_PICO	
1	82.6	82.6	82.5	509.3	4.8086		X															
1	61.0	60.9	60.7	1680.7	7.2103	X																
1	84.1	84.1	83.9	431.3	4.6042		X															
2	85.9	85.8	85.7	335.6	4.3395		X							X								
2	85.7	85.6	85.4	350.3	4.3811		X			X												
3	88.1	88.0	87.8	217.9	3.9880		X							X		X						
3	88.0	87.9	87.6	227.7	4.0183		X			X												X
4	89.0	88.9	88.7	172.3	3.8415		X						X	X		X						
4	88.7	88.5	88.3	192.2	3.9056		X	X						X		X						
5	90.1	89.9	89.4	117.7	3.6580		X			X				X								X
5	89.8	89.7	89.4	132.7	3.7089		X						X	X		X						
5	91.2	91.0	90.3	61.3	3.4573		X			X				X							X	X
8	92.2	92.0	91.4	11.3	3.2603		X		X	X				X	X					X	X	X
14 Feb 2012 (Landsat-7)																						
Vars	R-Sq	R-Sq (adj)	R-Sq (pred)	Mallows Cp	S	B1_479	B2_561	B3_661	B4_835	std_chl	std_DISS_OXY	std_SALI_NITY	std_TSM	std_TURBI_DITY	std_Fdom	std_AM-MONIUM	std_NIT-RITE	std_PHOS_PHATE	std_SILI_CATE	std_NANO	std_PICO	
1	81.1	81.1	80.9	208.0	5.0359		X															
2	85.6	85.5	85.3	60.3	4.4094	X	X															
2	84.5	84.4	84.2	97.4	4.5741		X	X														
3	87.0	86.9	86.7	15.1	4.1956	X	X	X														
3	86.2	86.1	85.9	42.7	4.3248	X	X		X													
4	87.2	87.1	86.9	8.5	4.1592	X	X	X					X									
4	87.1	87.0	86.8	12.6	4.1787	X	X	X											X			
5	87.3	87.2	86.9	7.7	4.1505	X	X	X		X				X								
5	87.3	87.2	86.9	7.9	4.1514	X	X	X	X					X								
6	87.4	87.2	87.0	5.9	4.1371	X	X	X	X	X				X								
02 Nov 2011 (Landsat-7)																						
Vars	R-Sq	R-Sq (adj)	R-Sq (pred)	Mallows Cp	S	B1_479	B2_561	B3_661	B4_835	std_chl	std_DISS_OXY	std_SALI_NITY	std_TSM	std_TURBI_DITY	std_Fdom	std_AM-MONIUM	std_NIT-RITE	std_PHOS_PHATE	std_SILI_CATE	std_NANO	std_PICO	
1	36.3	36.2	35.7	342.4	9.2408		X							X								
2	45.1	44.8	44.3	238.7	8.5927		X															
3	53.5	53.2	52.5	139.2	7.9176		X	X					X				X					
7	64.5	63.9	63.1	14.4	6.9520		X	X		X	X			X			X	X				
8	64.9	64.2	63.3	11.2	6.9182	X	X	X		X	X			X			X	X				
8	64.8	64.2	63.2	12.2	6.9263		X	X	X	X	X			X			X	X				
9	65.2	64.4	63.3	9.7	6.8981	X	X	X	X	X	X						X	X				
Total Occurrences						31	63	23	17	31	9	14	24	24	20	18	14		19	4	19	

Table 6: Test of Significance

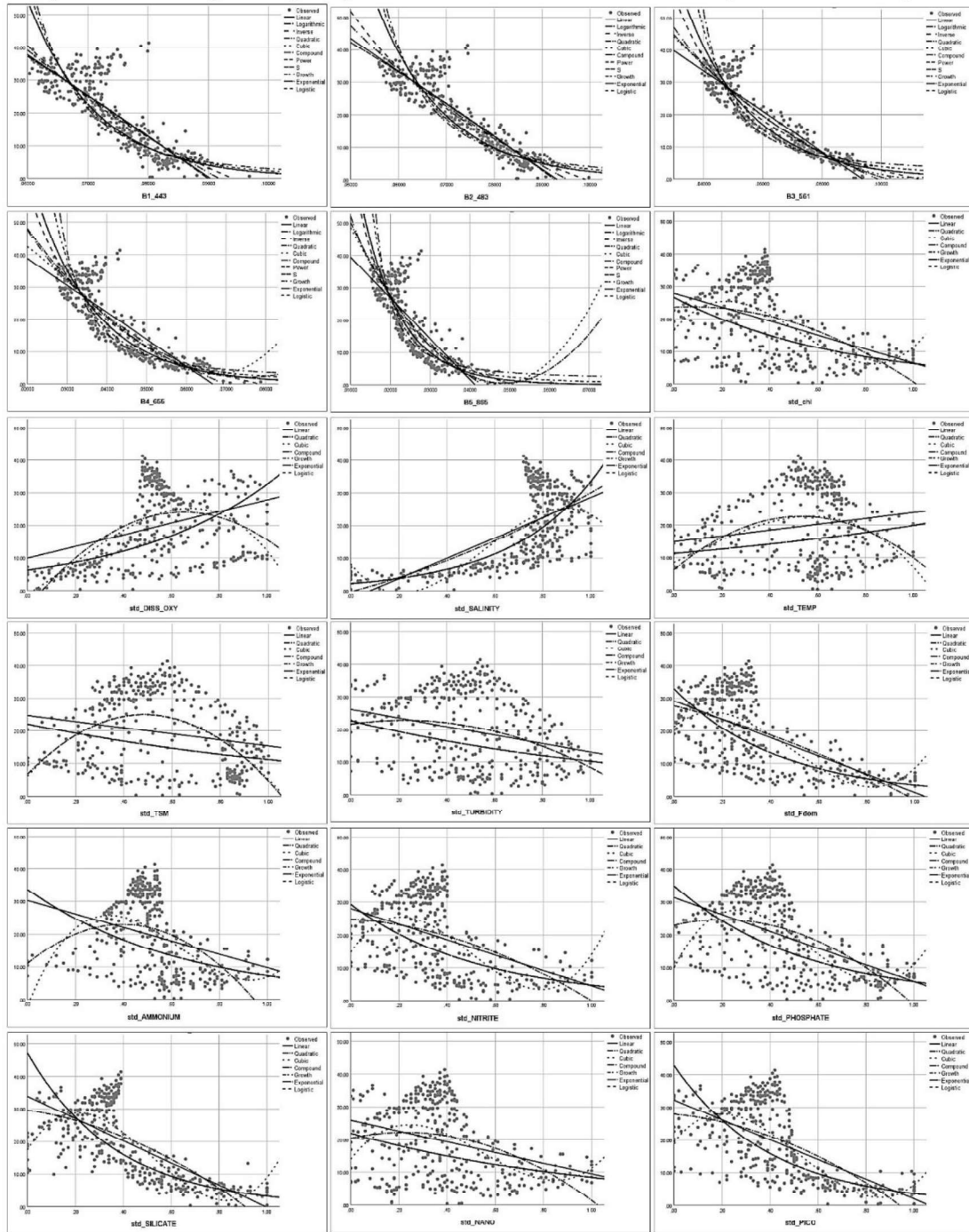
16022015 (DF-16) Landsat-7			13022015 (DF-16) Landsat-8			13022015 (DF-11) Landsat-7			14022012 (DF-9) Landsat-7		02112011 (DF-11) Landsat-7	
	F-Value	P-Value		F-Value	P-Value		F-Value	P-Value	F-Value	P-Value	F-Value	P-Value
Regression	208.50	0.000	Regression	377.23	0.000	Regression	248.678	.000	328.28	0.000	71.78	0.000
B1_479	9.61	0.002	B1_443	259.93	0.000	B1_479	1.702	.193	52.78	0.000	3.39	0.066
B2_561	13.59	0.000	B2_483	166.47	0.000	B2_561	156.712	.000	747.34	0.000	106.14	0.000
B3_661	0.68	0.412	B3_561	23.49	0.000	B3_661	2.051	.153	33.09	0.000	20.86	0.000
B4_835	2.08	0.150	B4_655	33.89	0.000	B4_835	.122	.728	4.44	0.036	3.45	0.064
STD_CHL	39.11	0.000	STD_CHL	8.53	0.004	STD_CHL	47.012	.000	4.91	0.027	3.37	0.067
STD DISS_OXY	1.35	0.246	STD DISS_OXY	3.90	0.049	STD DISS_OXY	X	X	0.12	0.731	14.05	0.000
STD_Salinity	54.93	0.000	STD_Salinity	1.66	0.199	STD_Salinity	X	X	X	X	X	X
STD_TSM	87.64	0.000	STD_TSM	0.853	0.03	STD_TSM	71.078	.000	0.72	0.397	1.18	0.279
STD_Turbidity	154.96	0.000	STD_Turbidity	0.80	0.037	STD_Turbidity	126.751	.000	X	X	X	X
STD_Fdom	147.88	0.000	STD_Fdom	4.93	0.027	STD_Fdom	X	X	X	X	X	X
Std_Ammonium	162.79	0.000	Std_Ammonium	4.13	0.043	Std_Ammonium	69.620	.000	X	X	X	X
STD_Nitrite	23.45	0.000	STD_Nitrite	1.22	0.270	STD_Nitrite	X	X	0.02	0.891	44.56	0.000
STD_Phosphate	8.99	0.003	STD_Phosphate	3.65	0.057	STD_Phosphate	X	X	X	X	96.45	0.000
STD_Silicate	2.34	0.128	STD_Silicate	0.01	0.943	STD_Silicate	32.346	.000	0.92	0.337	1.47	0.226
STD_Nano	10.30	0.002	STD_Nano	4.66	0.032	STD_Nano	28.777	.000	X	X	X	X
STD_Pico	91.03	0.000	STD_Pico	1.16	0.283	STD_Pico	2.040	.154	X	X	X	X
R-Square	0.931		R-Square	0.958		R-Square	.923		R-Square	0.874	R-Square	0.653
SE	3.13		SE	2.41		SE	3.25		SE	4.14	SE	6.1
Dependent variable is Bathy_Kochi												
Relationships are significant if less than 0.05 at 95 % Confidence Level												

APPENDIX-II: FIGURES

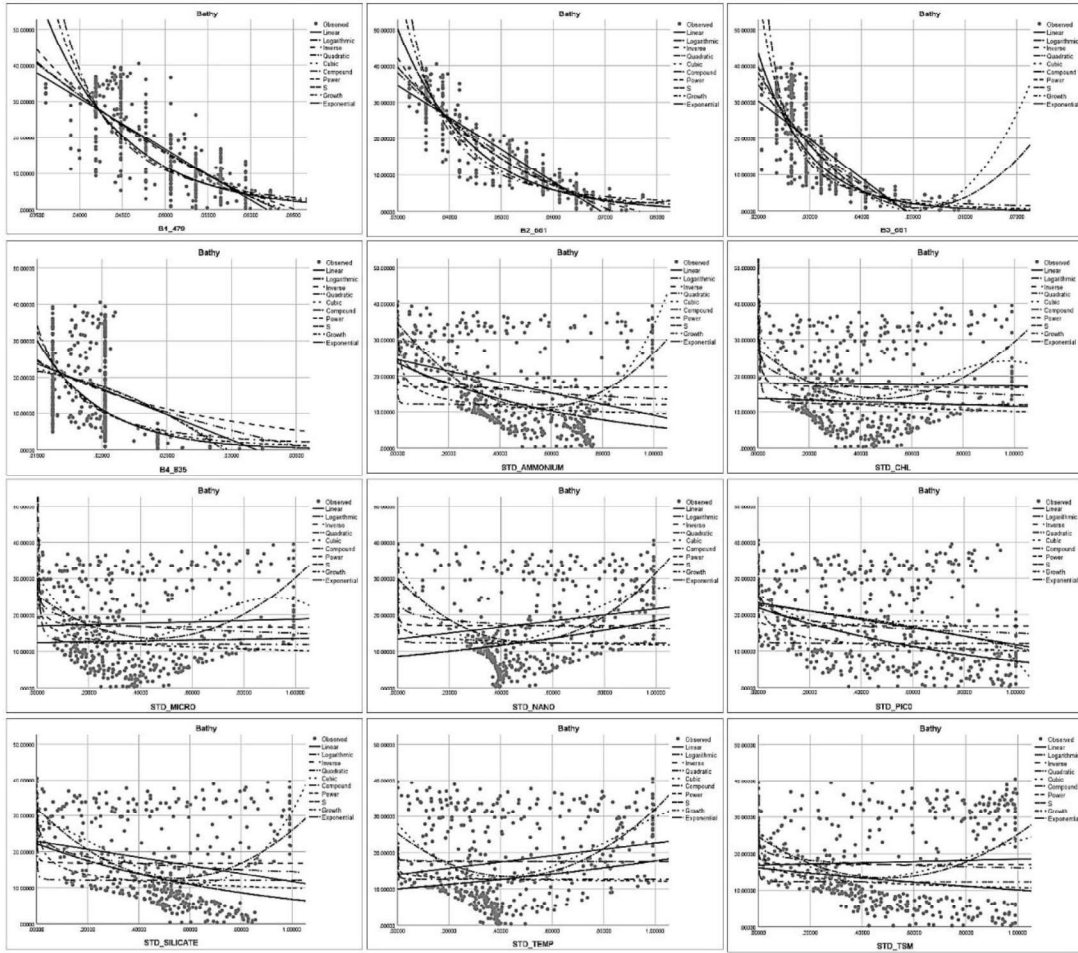
Regression modelling of Bathymetry verses Landsat-7 Bands and SATCORE parameters – 16 Dec 2015



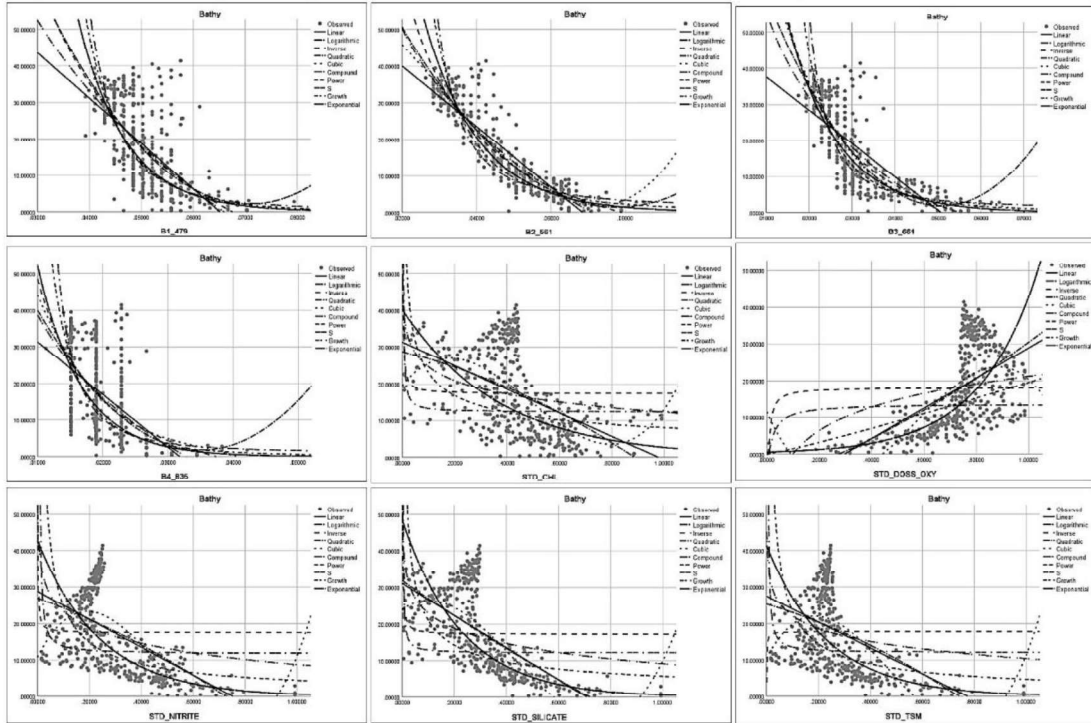
Regression modelling of Bathymetry verses Landsat-8 Bands and SATCORE parameters – 13 May 2015



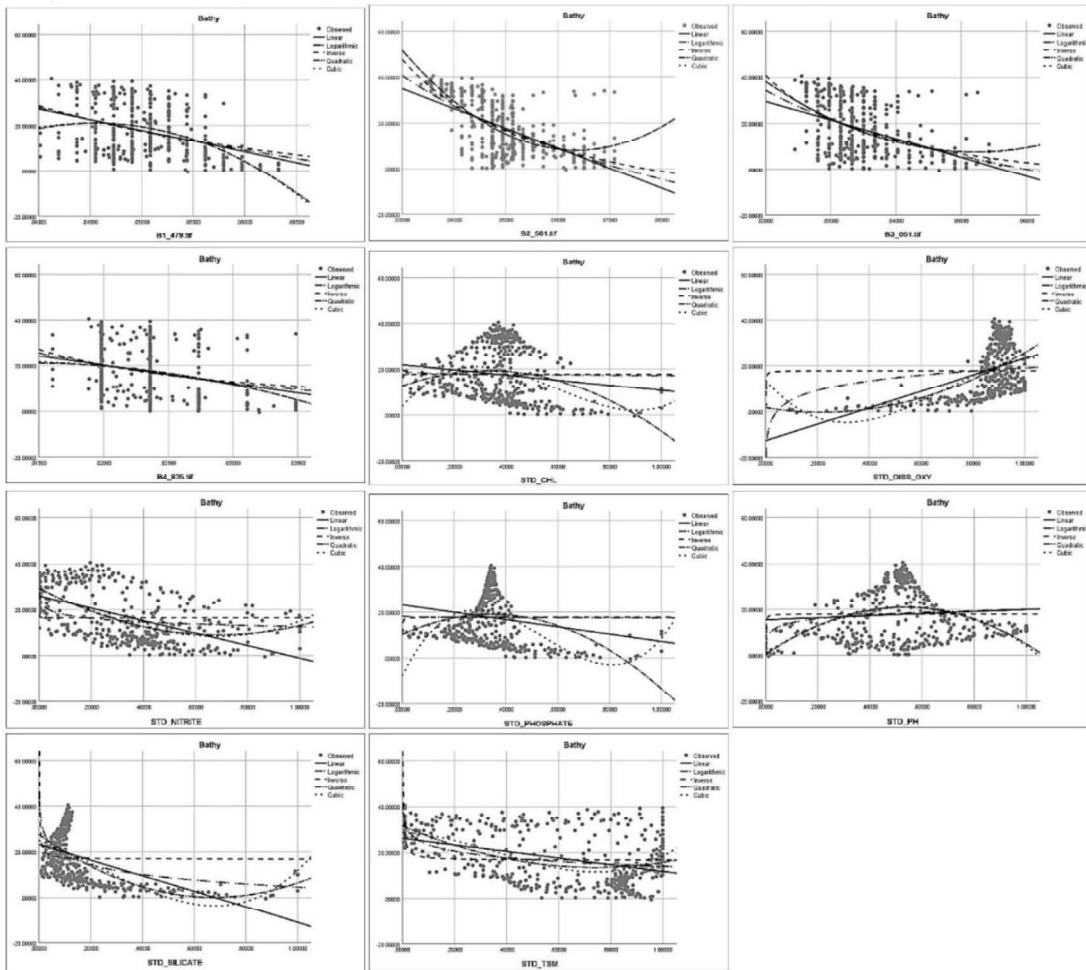
Regression modelling of Bathymetry verses Landsat-7 Bands and SATCORE parameters – 04 Dec 2013



Regression modelling of Bathymetry verses Landsat-7 Bands and SATCORE parameters – 14 Feb 2012



Regression modelling of Bathymetry verses Landsat-7 Bands and SATCORE parameters – 02 Nov 2011



APPENDIX-III: ML SYNTAX

```
In [ ]: # Decision Tree for Multioutput Regression
from sklearn.datasets import make_regression
from sklearn.tree import DecisionTreeRegressor
from sklearn.datasets import make_regression
from sklearn.neighbors import KNeighborsRegressor
from sklearn.model_selection import train_test_split
from sklearn.datasets import make_regression
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import RepeatedKFold
from sklearn.metrics import mean_absolute_error
from sklearn.ensemble import RandomForestRegressor
from sklearn.datasets import make_regression

In [ ]: import pandas as pd
df=pd.read_csv('VENGURLA-LANDSAT.csv')
# split into input (X) and output (y) variables
df.head()

In [ ]: X= df.iloc[:, 3:8].values
y= df.iloc[:,2].values
# split into train test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.5, random_state=

In [ ]: from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
model = LinearRegression()
model.fit(X_train, y_train)
print('\nCoefficient of model :', model.coef_)
print('\nIntercept of model',model.intercept_)
# make prediction on test set
yhat = model.predict(X_test)

In [ ]: # calculate error
score = mean_absolute_error(y_test, yhat)
print(score)

In [ ]: from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score
rmse_test = mean_squared_error(y_test, yhat)**(0.5)
print('\nRMSE on test dataset : ', rmse_test)

In [ ]: r= r2_score(y_test, yhat)**(0.5)
print(r)

In [ ]: from matplotlib import pyplot
pyplot.scatter(y_test, yhat)
pyplot.show()

In [ ]: import seaborn as sns
ax = sns.distplot(df['Depth'], color="r", label='Actual')
sns.distplot(yhat, color='b', label="predicted values", ax = ax)
pyplot.title('Actual vs Predicted value for depth')
pyplot.show()
```

Syntax-1:- Linear ML syntax

```

In [ ]: # Decision Tree for Multioutput Regression
from sklearn.datasets import make_regression
from sklearn.tree import DecisionTreeRegressor
from sklearn.datasets import make_regression
from sklearn.neighbors import KNeighborsRegressor
from sklearn.model_selection import train_test_split
from sklearn.datasets import make_regression
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import RepeatedKFold
from sklearn.metrics import mean_absolute_error
from sklearn.ensemble import RandomForestRegressor
from sklearn.datasets import make_regression

In [ ]: import pandas as pd
df=pd.read_csv('VENGURLA-ASTER.csv')
# split into input (X) and output (y) variables
df.head()

In [ ]: X= df.iloc[:, 3:6].values
y= df.iloc[:,2].values
# split into train test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.5, random_state=1)

In [ ]: # define model
model = RandomForestRegressor()
regr = RandomForestRegressor(max_depth=2, random_state=0)
regr.fit(X, y)
model.fit(X_train, y_train)
# make prediction on test set
yhat = model.predict(X_test)

In [ ]: # calculate error
score = mean_absolute_error(y_test, yhat)
print(score)

In [ ]: from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score
rmse_test = mean_squared_error(y_test, yhat)**(0.5)
print('\nRMSE on test dataset : ', rmse_test)

In [ ]: r= r2_score(y_test, yhat)**(0.5)
print(r)

In [ ]: from matplotlib import pyplot
pyplot.scatter(y_test, yhat)
pyplot.show()

In [ ]: import seaborn as sns
ax = sns.distplot(df['DEPTH'], color="r", label='Actual')
sns.distplot(yhat, color='b', label="predicted values", ax = ax)
pyplot.title('Actual vs Predicted value for depth')
pyplot.show()

```

Syntax-2:- RF ML syntax

```

In [ ]: from sklearn.datasets import make_regression
        from sklearn.preprocessing import MinMaxScaler
        from sklearn.model_selection import train_test_split
        from sklearn.svm import SVR
        from sklearn.metrics import mean_absolute_error

In [ ]: import pandas as pd
        df=pd.read_csv('VENGURLA-LANDSAT.csv')
        # split into input (X) and output (y) variables
        df.head()

In [ ]: X= df.iloc[:, 3:8].values
        y= df.iloc[:,2].values
        n_features = X.shape[1]

In [ ]: # split into train test sets
        X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.33, random_state=1)

In [ ]: # reshape target variables so that we can transform them
        y_train = y_train.reshape((len(y_train), 1))
        y_test = y_test.reshape((len(y_test), 1))

In [ ]: # scale input data
        trans_in = MinMaxScaler()
        trans_in.fit(X_train)
        X_train = trans_in.transform(X_train)
        X_test = trans_in.transform(X_test)

In [ ]: # scale output data
        trans_out = MinMaxScaler()
        trans_out.fit(y_train)
        y_train = trans_out.transform(y_train)
        y_test = trans_out.transform(y_test)

In [ ]: # define model
        model = SVR()

In [ ]: # fit model on the training dataset
        model.fit(X_train, y_train)

In [ ]: # make prediction on test set
        yhat = model.predict(X_test)

In [ ]: # invert transforms so we can calculate errors
        yhat = yhat.reshape((len(yhat), 1))
        yhat = trans_out.inverse_transform(yhat)
        y_test = trans_out.inverse_transform(y_test)

In [ ]: # calculate error
        score = mean_absolute_error(y_test, yhat)
        print(score)

In [ ]: from sklearn.metrics import r2_score
        r= r2_score(y_test, yhat)**(0.5)
        print(r)

In [ ]: from sklearn.metrics import mean_squared_error
        from sklearn.metrics import r2_score
        rmse_test = mean_squared_error(y_test, yhat)**(0.5)
        print('\nRMSE on test dataset : ', rmse_test)

In [ ]: import seaborn as sns
        ax = sns.distplot(df['Depth'], color="r", label='Actual')
        sns.distplot(yhat, color='b', label="predicted values", ax = ax)
        pyplot.title('Actual vs Predicted value for depth')
        pyplot.show()

```

Syntax-3:- SVR ML syntax

```

In [ ]: from random import random
        from random import randint
        from random import seed
        from numpy import arange
        from numpy import mean
        from numpy import std
        from numpy import absolute
        from sklearn.datasets import make_regression
        from sklearn.linear_model import TheilSenRegressor
        from sklearn.model_selection import cross_val_score
        from sklearn.model_selection import RepeatedKFold
        from matplotlib import pyplot

In [ ]: import pandas as pd
        df=pd.read_csv('ASTER-.csv')
        # split into input (X) and output (y) variables
        df.head()
        X= df.iloc[:, 3:4].values
        y= df.iloc[:,2].values
        print(X)
        print(y)

In [ ]: # evaluate a model
        def evaluate_model(X, y, model):
            # define model evaluation method
            cv = RepeatedKFold(n_splits=10, n_repeats=3, random_state=1)
            # evaluate model
            scores = cross_val_score(model, X, y, scoring='neg_mean_absolute_error', cv=cv, n_jobs=-1)
            # force scores to be positive
            return absolute(scores)

        # plot the dataset and the model's line of best fit
        def plot_best_fit(X, y, model):
            # fit the model on all data
            model.fit(X, y)
            # plot the dataset
            pyplot.scatter(X, y)
            # plot the line of best fit
            xaxis = arange(X.min(), X.max(), 0.01)
            yaxis = model.predict(xaxis.reshape((len(xaxis), 1)))
            pyplot.plot(xaxis, yaxis, color='r')
            # show the plot
            pyplot.title(type(model).__name__)
            pyplot.show()

        # define the model
        model = TheilSenRegressor()
        # evaluate model
        results = evaluate_model(X, y, model)
        print('Mean MAE: %.3f (%.3f)' % (mean(results), std(results)))
        # plot the line of best fit
        plot_best_fit(X, y, model)

In [ ]: from sklearn.metrics import mean_squared_error
        from sklearn.metrics import r2_score
        predict_test = model.predict(X)
        rmse_test = mean_squared_error(y,predict_test)**(0.5)
        print('\nRMSE on test dataset : ', rmse_test)

In [ ]: r= r2_score(y,predict_test)**(0.5)
        print(r)

In [ ]: pyplot.scatter(predict_test, y)
        pyplot.show()

In [ ]: import seaborn as sns
        ax = sns.kdeplot(df['Depth'], color="r", label='Actual')
        sns.kdeplot(predict_test, color='b', label="predicted values", ax=ax)
        pyplot.title('Actual vs Predicted value for depth')
        pyplot.show()

```

Syntax-4:- Robust-regression Theil Sen regression


```

In [ ]: from random import random
        from random import randint
        from random import seed
        from numpy import arange
        from numpy import mean
        from numpy import std
        from numpy import absolute
        from sklearn.datasets import make_regression
        from sklearn.linear_model import HuberRegressor
        from sklearn.model_selection import cross_val_score
        from sklearn.model_selection import RepeatedKFold
        from matplotlib import pyplot

In [ ]: import pandas as pd
        df=pd.read_csv('ASTER-GOA.csv')
        # split into input (X) and output (y) variables
        df.head()
        X= df.iloc[:, 3:4].values
        y= df.iloc[:,2].values
        print(X)
        print(y)

In [ ]: # evaluate a model
        def evaluate_model(X, y, model):
            # define model evaluation method
            cv = RepeatedKFold(n_splits=10, n_repeats=3, random_state=1)
            # evaluate model
            scores = cross_val_score(model, X, y, scoring='neg_mean_absolute_error', cv=cv, n_jobs=-1)
            # force scores to be positive
            return absolute(scores)

        # plot the dataset and the model's line of best fit
        def plot_best_fit(X, y, model):
            # fit the model on all data
            model.fit(X, y)
            # plot the dataset
            pyplot.scatter(X, y)
            # plot the line of best fit
            xaxis = arange(X.min(), X.max(), 0.01)
            yaxis = model.predict(xaxis.reshape((len(xaxis), 1)))
            pyplot.plot(xaxis, yaxis, color='r')
            # show the plot
            pyplot.title(type(model).__name__)
            pyplot.show()

        # define the model
        model = HuberRegressor()
        # evaluate model
        results = evaluate_model(X, y, model)
        print('Mean MAE: %.3f (%.3f)' % (mean(results), std(results)))
        # plot the line of best fit
        plot_best_fit(X, y, model)

In [ ]: from sklearn.metrics import mean_squared_error
        from sklearn.metrics import r2_score
        predict_test = model.predict(X)
        rmse_test = mean_squared_error(y,predict_test)**(0.3)
        print('\nRMSE on test dataset : ', rmse_test)

In [ ]: r= r2_score(y,predict_test)**(0.3)
        print(r)

In [ ]: pyplot.scatter(predict_test, y)
        pyplot.show()

In [ ]: import seaborn as sns
        ax = sns.kdeplot(df['Depth'], color="r", label='Actual')
        sns.kdeplot(predict_test, color='b', label="predicted values", ax=ax)
        pyplot.title('Actual vs Predicted value for depth')
        pyplot.show()

In [ ]: # Mean Absolute Error
        from sklearn.metrics import mean_absolute_error
        errors = mean_absolute_error(y,predict_test)**(0.3)
        print(errors)

```

Syntax-5 :- Robust-regression -Huber regression

```

In [ ]: from random import random
        from random import randint
        from random import seed
        from numpy import arange
        from numpy import mean
        from numpy import std
        from numpy import absolute
        from sklearn.datasets import make_regression
        from sklearn.linear_model import RANSACRegressor
        from sklearn.model_selection import cross_val_score
        from sklearn.model_selection import RepeatedKFold
        from matplotlib import pyplot

In [ ]: import pandas as pd
        df=pd.read_csv('LANDSAT-GOA.csv')
        # split into input (X) and output (y) variables
        df.head()
        X= df.iloc[:, 7:8].values
        y= df.iloc[:,2].values
        print(X)
        print(y)

In [ ]: # evaluate a model
        def evaluate_model(X, y, model):
            # define model evaluation method
            cv = RepeatedKFold(n_splits=10, n_repeats=3, random_state=1)
            # evaluate model
            scores = cross_val_score(model, X, y, scoring='neg_mean_absolute_error', cv=cv, n_jobs=-1)
            # force scores to be positive
            return absolute(scores)

        # plot the dataset and the model's line of best fit
        def plot_best_fit(X, y, model):
            # fit the model on all data
            model.fit(X, y)
            # plot the dataset
            pyplot.scatter(X, y)
            # plot the line of best fit
            xaxis = arange(X.min(), X.max(), 0.01)
            yaxis = model.predict(xaxis.reshape((len(xaxis), 1)))
            pyplot.plot(xaxis, yaxis, color='r')
            # show the plot
            pyplot.title(type(model).__name__)
            pyplot.show()

        # define the model
        model = RANSACRegressor()
        # evaluate model
        results = evaluate_model(X, y, model)
        print('Mean MAE: %.3f (%.3f)' % (mean(results), std(results)))
        # plot the line of best fit
        plot_best_fit(X, y, model)

In [ ]: from sklearn.metrics import mean_squared_error
        from sklearn.metrics import r2_score
        predict_test = model.predict(X)
        rmse_test = mean_squared_error(y,predict_test)**(0.5)
        print('\nRMSE on test dataset : ', rmse_test)

In [ ]: r= r2_score(y,predict_test)**(0.5)
        print(r)

In [ ]: pyplot.scatter(predict_test, y)
        pyplot.show()

In [ ]: import seaborn as sns
        ax = sns.kdeplot(df['Depth'], color="r", label='Actual')
        sns.kdeplot(predict_test, color='b', label="predicted values", ax=ax)
        pyplot.title('Actual vs Predicted value for depth')
        pyplot.show()

```

Syntax-6:- Robust-regression -Random Sample Consensus

```
In [ ]: from random import random
        from random import randint
        from random import seed
        from numpy import mean
        from numpy import std
        from numpy import absolute
        from sklearn.datasets import make_regression
        from sklearn.model_selection import cross_val_score
        from sklearn.model_selection import RepeatedKFold
        from sklearn.linear_model import LinearRegression
        from sklearn.linear_model import HuberRegressor
        from sklearn.linear_model import RANSACRegressor
        from sklearn.linear_model import TheilSenRegressor
        from matplotlib import pyplot
```

```
In [ ]: import pandas as pd
        df=pd.read_csv('ASTER-VEGURLA.csv')
        # split into input (X) and output (y) variables
        df.head()
        X= df.iloc[:, 3:4].values
        y= df.iloc[:,2].values
        print(X)
        print(y)
```

```
In [ ]: import pandas as pd
        df=pd.read_csv('ASTER-VEGURLA.csv')
        # split into input (X) and output (y) variables
        df.head()
        X= df.iloc[:, 3:4].values
        y= df.iloc[:,2].values
        print(X)
        print(y)
```

```
In [ ]: # same plot in the context of a scatter plot of the entire training dataset.
        from random import random
        from random import randint
        from random import seed
        from numpy import arange
        from sklearn.datasets import make_regression
        from sklearn.linear_model import LinearRegression
        from sklearn.linear_model import HuberRegressor
        from sklearn.linear_model import RANSACRegressor
        from sklearn.linear_model import TheilSenRegressor
        from matplotlib import pyplot
```

```
In [ ]: # dictionary of model names and model objects
        def get_models():
            models = list()
            models.append(LinearRegression())
            models.append(HuberRegressor())
            models.append(RANSACRegressor())
            models.append(TheilSenRegressor())
            return models

        # plot the dataset and the model's line of best fit
        def plot_best_fit(X, y, xaxis, model):
            # fit the model on all data
            model.fit(X, y)
            # calculate outputs for grid across the domain
            yaxis = model.predict(xaxis.reshape((len(xaxis), 1)))
            # plot the line of best fit
            pyplot.plot(xaxis, yaxis, label=type(model).__name__)

        # define a uniform grid across the input domain
        xaxis = arange(X.min(), X.max(), 0.01)
        for model in get_models():
            # plot the line of best fit
            plot_best_fit(X, y, xaxis, model)
        # plot the dataset
        pyplot.scatter(X, y)
        # show the plot
        pyplot.title('Robust Regression')
        pyplot.legend()
        pyplot.show()
```

Syntax-7:- Compare Robust Regression Algorithms

```

In [ ]:
from sklearn.datasets import make_regression
from sklearn.tree import DecisionTreeRegressor
from sklearn.datasets import make_regression
from sklearn.neighbors import KNeighborsRegressor
from sklearn.model_selection import train_test_split
from sklearn.datasets import make_regression
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import RepeatedKFold
from sklearn.metrics import mean_absolute_error
from sklearn.ensemble import RandomForestRegressor
from sklearn.datasets import make_regression

In [ ]: import pandas as pd
df=pd.read_csv('LANDSAT-VENGURLA.csv')
# split into input (X) and output (y) variables
df.head()
X= df.iloc[:, 3:8].values
y= df.iloc[:,2].values
# split into train test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1)

In [ ]: # define model
model = RandomForestRegressor()
regr = RandomForestRegressor(max_depth=2, random_state=0)
regr.fit(X, y)
model.fit(X_train, y_train)
# make prediction on test set
yhat = model.predict(X_test)

In [ ]: # calculate error
score = mean_absolute_error(y_test, yhat)
print(score)

In [ ]: from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score
rmse_test = mean_squared_error(y_test, yhat)**(0.2)
print('\nRMSE on test dataset : ', rmse_test)

In [ ]: r= r2_score(y_test, yhat)**(0.2)
print(r)

In [ ]: from matplotlib import pyplot
pyplot.scatter(y_test, yhat)
pyplot.show()

In [ ]: import seaborn as sns
ax = sns.distplot(df['Depth'], color="r", label='Actual')
sns.distplot(yhat, color='b', label="predicted values", ax = ax)
pyplot.title('Actual vs Predicted value for depth')
pyplot.show()

```

Syntax-8:- Multivariate RF

```

In [ ]: # evaluate multivariate adaptive regression splines for regression
        from numpy import mean
        from numpy import std
        from sklearn.datasets import make_regression
        from sklearn.model_selection import cross_val_score
        from sklearn.model_selection import RepeatedKFold
        from pyearth import Earth
        from sklearn.model_selection import train_test_split

In [ ]: import pandas as pd
        df=pd.read_csv('LANDSAT-VENGURLA.csv')
        # split into input (X) and output (y) variables
        df.head()

In [ ]: X= df.iloc[:, 3:8].values
        y= df.iloc[:,2].values
        # split into train test sets
        X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1)

In [ ]: model = Earth()
        model.fit(X_train, y_train)
        # define the evaluation procedure
        cv = RepeatedKFold(n_splits=10, n_repeats=3, random_state=1)
        # evaluate the model and collect results
        n_scores = cross_val_score(model, X, y, scoring='neg_mean_absolute_error', cv=cv, n_jobs=-1)
        # report performance
        print('MAE: %.3f (%.3f)' % (mean(n_scores), std(n_scores)))

In [ ]: yhat = model.predict(X_test)

In [ ]: # calculate error
        from sklearn.metrics import mean_absolute_error
        score = mean_absolute_error(y_test, yhat)
        print(score)

In [ ]: from sklearn.metrics import mean_squared_error
        from sklearn.metrics import r2_score
        rmse_test = mean_squared_error(y_test, yhat)**(0.2)
        print('\nRMSE on test dataset : ', rmse_test)

In [ ]: r= r2_score(y_test, yhat)**(0.2)
        print(r)

In [ ]: from matplotlib import pyplot
        pyplot.scatter(y_test, yhat)
        pyplot.show()

In [ ]: import seaborn as sns
        ax = sns.distplot(df['Depth'], color="r", label='Actual')
        sns.distplot(yhat, color='b', label="predicted values", ax = ax)
        pyplot.title('Actual vs Predicted value for depth')
        pyplot.show()

```

Syntax-9:- Multivariate MARS

```

In [ ]: import pandas as pd
        from sklearn.linear_model import LinearRegression
        from sklearn.metrics import mean_squared_error
        from sklearn.metrics import r2_score
        import matplotlib.pyplot as plt
        import seaborn as sns
        from sklearn.model_selection import train_test_split
        from sklearn.preprocessing import StandardScaler
        from sklearn.linear_model import LogisticRegression
        from sklearn.neighbors import KNeighborsClassifier
        from sklearn.metrics import confusion_matrix
        from sklearn.metrics import accuracy_score
        from sklearn.metrics import f1_score
        import pandas as pd
        import numpy as np
        import geopandas as gpd
        import shapely.geometry.point as pt
        from shapely.geometry import Point
        %matplotlib inline

In [ ]: import pandas as pd
        df=pd.read_csv('LANDSAT-VENGURLA.csv')
        # split into input (X) and output (y) variables
        df.head()
        X= df.iloc[:, 3:8].values
        y= df.iloc[:,2].values
        # split into train test sets
        X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1)

In [ ]: from sklearn.linear_model import LinearRegression
        from sklearn.metrics import mean_squared_error
        model = LinearRegression()

        # define model
        model.fit(X_train, y_train)
        # make prediction on test set
        yhat = model.predict(X_test)

        print('\nCoefficient of model :', model.coef_)
        print('\nIntercept of model', model.intercept_)

In [ ]: from sklearn.metrics import mean_absolute_error
        # calculate error
        score = mean_absolute_error(y_test, yhat)**(0.2)
        print(score)

In [ ]: from sklearn.metrics import mean_squared_error
        from sklearn.metrics import r2_score
        rmse_test = mean_squared_error(y_test, yhat)**(0.2)
        print('\nRMSE on test dataset : ', rmse_test)

In [ ]: r= r2_score(y_test, yhat)**(0.2)
        print(r)

In [ ]: from matplotlib import pyplot
        pyplot.scatter(y_test, yhat)
        pyplot.show()

In [ ]: import seaborn as sns
        ax = sns.distplot(df['Depth'], color="r", label='Actual')
        sns.distplot(yhat, color='b', label="predicted values", ax = ax)
        pyplot.title('Actual vs Predicted value for depth')
        pyplot.show()

```

Syntax-10:- Multivariate MLR

```

In [ ]: # Decision Tree for Multioutput Regression
from sklearn.datasets import make_regression
from sklearn.tree import DecisionTreeRegressor
from sklearn.datasets import make_regression
from sklearn.neighbors import KNeighborsRegressor
from sklearn.model_selection import train_test_split
from sklearn.datasets import make_regression
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import RepeatedKFold
from sklearn.metrics import mean_absolute_error
from sklearn.datasets import make_regression
from sklearn.gaussian_process import GaussianProcessRegressor

In [ ]: import pandas as pd
df=pd.read_csv('LANDSAT-VENGURLA.csv')
# split into input (X) and output (y) variables
df.head()
X= df.iloc[:, 3:8].values
y= df.iloc[:,2].values
# split into train test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.5, random_state=1)

In [ ]: from sklearn.gaussian_process import GaussianProcessRegressor
from sklearn.gaussian_process.kernels import ConstantKernel, RBF

# Define kernel parameters.
l = 0.1
sigma_f = 2
sigma_n = 2

# Define kernel object.
kernel = ConstantKernel(constant_value=sigma_f,constant_value_bounds=(1e-3, 1e3)) \
* RBF(length_scale=l, length_scale_bounds=(1e-3, 1e3))

In [ ]: from sklearn.gaussian_process import GaussianProcessRegressor
# Define GaussianProcessRegressor object.
model = GaussianProcessRegressor(kernel=kernel, alpha=sigma_n**2, n_restarts_optimizer=10)

In [ ]: GaussianProcessRegressor(alpha=0.16000000000000003, copy_X_train=True,
kernel=1.41**2 * RBF(length_scale=0.1),
n_restarts_optimizer=10, normalize_y=False,
optimizer='fmin_l_bfgs_b', random_state=None)

In [ ]: # define model
model.fit(X_train, y_train)
# make prediction on test set
yhat = model.predict(X_test)

In [ ]: # calculate error
score = mean_absolute_error(y_test, yhat)**(0.5)
print(score)

In [ ]: from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score
rmse_test = mean_squared_error(y_test, yhat)**(0.5)
print('\nRMSE on test dataset : ', rmse_test)

In [ ]: r= r2_score(y_test, yhat)**(0.5)
print(r)

In [ ]: from matplotlib import pyplot
pyplot.scatter(y_test, yhat)
pyplot.show()

In [ ]: import seaborn as sns
ax = sns.distplot(df['Depth'], color="r", label='Actual')
sns.distplot(yhat, color='b', label="predicted values", ax = ax)
pyplot.title('Actual vs Predicted value for depth')
pyplot.show()

```

Syntax-11:- Multivariate GPR

```

In [ ]: from numpy import absolute
        from numpy import mean
        from numpy import std
        from sklearn.datasets import make_regression
        from sklearn.tree import DecisionTreeRegressor
        from sklearn.model_selection import cross_val_score
        from sklearn.model_selection import RepeatedKFold
        from sklearn.model_selection import train_test_split

        import pandas as pd
        df=pd.read_csv('LANDSAT-GOA.csv')
        # split into input (X) and output (y) variables
        df.head()
        X= df.iloc[:, 3:8].values
        y= df.iloc[:,2].values
        # split into train test sets
        X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.5, random_state=1)

        # define model
        model = DecisionTreeRegressor()
        # define the evaluation procedure
        cv = RepeatedKFold(n_splits=10, n_repeats=3, random_state=1)
        # evaluate the model and collect the scores
        n_scores = cross_val_score(model, X, y, scoring='neg_mean_absolute_error', cv=cv, n_jobs=-1)
        # force the scores to be positive
        n_scores = absolute(n_scores)
        # summarize performance
        print('MAE: %.3f (%.3f)' % (mean(n_scores), std(n_scores)))

In [ ]: # define model
        model.fit(X_train, y_train)
        # make prediction on test set
        yhat = model.predict(X_test)

In [ ]: # calculate error
        from sklearn.metrics import mean_absolute_error
        score = mean_absolute_error(y_test, yhat)
        print(score)

In [ ]: from sklearn.metrics import mean_squared_error
        from sklearn.metrics import r2_score
        rmse_test = mean_squared_error(y_test, yhat)**(0.5)
        print('\nRMSE on test dataset : ', rmse_test)

In [ ]: r= r2_score(y_test, yhat)**(0.5)
        print(r)

In [ ]: from matplotlib import pyplot
        pyplot.scatter(y_test, yhat)
        pyplot.show()

In [ ]: import seaborn as sns
        ax = sns.distplot(df['Depth'], color="r", label='Actual')
        sns.distplot(yhat, color='b', label="predicted values", ax = ax)
        pyplot.title('Actual vs Predicted value for depth')
        pyplot.show()

```

Syntax-12:- Decision tree regression using k-fold cross-validation


```

In [ ]: # Decision Tree for Multioutput Regression
from sklearn.datasets import make_regression
from sklearn.tree import DecisionTreeRegressor
from sklearn.datasets import make_regression
from sklearn.neighbors import KNeighborsRegressor
from sklearn.model_selection import train_test_split
from sklearn.datasets import make_regression
from sklearn.model_selection import cross_val_score
from sklearn.model_selection import RepeatedKFold
from sklearn.metrics import mean_absolute_error

In [ ]: import pandas as pd
df=pd.read_csv('LANDSAT-VENGURLA.csv')
# split into input (X) and output (y) variables
df.head()
X= df.iloc[:, 3:8].values
y= df.iloc[:,2].values
# split into train test sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.5, random_state=1)

In [ ]: # define model
model = model = DecisionTreeRegressor()
model.fit(X_train, y_train)
# make prediction on test set
yhat = model.predict(X_test)

In [ ]: # calculate error
score = mean_absolute_error(y_test, yhat)
print(score)

In [ ]: from sklearn.metrics import mean_squared_error
from sklearn.metrics import r2_score
rmse_test = mean_squared_error(y_test, yhat)**(0.5)
print('\nRMSE on test dataset : ', rmse_test)

In [ ]: r= r2_score(y_test, yhat)**(0.5)
print(r)

In [ ]: from matplotlib import pyplot
pyplot.scatter(y_test, yhat)
pyplot.show()

In [ ]: import seaborn as sns
ax = sns.distplot(df['Depth'], color="r", label='Actual')
sns.distplot(yhat, color='b', label="predicted values", ax = ax)
pyplot.title('Actual vs Predicted value for depth')
pyplot.show()

```

Syntax-13 :- Multivariate Decision Tree for Multioutput Regression

12/11/22, 3:54 PM

RF MERGE KOCHI AND OKHA - Jupyter Notebook

```

In [1]: import pickle, shutil, os

#combine list of pickle files
with open('OKHA_KOCHI.pkl', 'wb') as dst:
    for pickle_file in ['RF12022017.pkl', 'RF_02112011.pkl', 'RF27112015.pkl', 'RF_16122015.pkl', 'RF31032015.pkl', 'RF_13052015.pkl']:
        with open(pickle_file, 'rb') as src:
            shutil.copyfileobj(src, dst)

In [ ]:

```

12/11/22, 3:58 PM

RF MERGE KOCHI AND OKHA - Jupyter Notebook

```

In [1]: import pickle, shutil, os

#combine list of pickle files
with open('OKHA_KOCHI.pkl', 'wb') as dst:
    for pickle_file in ['RF12022017.pkl', 'RF_02112011.pkl', 'RF27112015.pkl', 'RF_16122015.pkl', 'RF31032015.pkl', 'RF_13052015.pkl']:
        with open(pickle_file, 'rb') as src:
            shutil.copyfileobj(src, dst)

In [ ]: import pickle, shutil, os

#combine list of pickle files
with open('OKHA.pkl', 'wb') as dst:
    for pickle_file in ['RF12022017.pkl', 'RF27112015.pkl', 'RF31032015.pkl']:
        with open(pickle_file, 'rb') as src:
            shutil.copyfileobj(src, dst)

```

Refer Syntax-14 :- Syntax for merging of RF .pkl files

APPENDIX-IV: LIST OF PUBLICATION

- Ashphaq, M., Srivastava, P. K., & Mitra, D. (2023).** Preliminary examination of influence of Chlorophyll, Total Suspended Material, and Turbidity on Satellite Derived-Bathymetry estimation in coastal turbid water. *Regional Studies in Marine Science*, 62, 102920.
- Ashphaq, M., Srivastava, P. K., & Mitra, D. (2022).** Analysis of univariate linear, robust-linear, and non-linear machine learning algorithms for satellite-derived bathymetry in complex coastal terrain. *Regional Studies in Marine Science*, 56, 102678.
- Ashphaq, M., Srivastava, P. K., & Mitra, D. (2022).** Evaluation and performance of satellite-derived bathymetry algorithms in turbid coastal water: a case study of Vengurla rocks. *Indian Journal of Geo-Marine Sciences (IJMS)*, 51(04), 310-321.
- Ashphaq, M., Srivastava, P. K., & Mitra, D. (2022).** Role of space technology in synoptical analysis of opportunity and challenges for revolutionizing the blue economy. *Journal of East China University of Science and Technology*, 65(4), 75-91.
- Ashphaq, M., Srivastava, P. K., & Mitra, D. (2021).** Review of near-shore satellite derived bathymetry: Classification and account of five decades of coastal bathymetry research. *Journal of Ocean Engineering and Science*, 6(4), 340-359.

APPENDIX V: THESIS PLAGIARISM CERTIFICATE



PLAGIARISM CERTIFICATE

1. We Pankaj Kumar Srivastava (Internal Guide), _____ (Co Guide/
External Guide) certify that the Thesis titled
Bathymetric Analysis Using Multi Spectral Satellite
Data Through Machine Learning in Varying Coastal Regions of
submitted by Scholar Mr/ Ms Mohammad Ashfaq having SAP ID 500071972 ^{India}
_____ has been run through a Plagiarism Check Software and the Plagiarism
Percentage is reported to be 5 %.
2. Plagiarism Report generated by the Plagiarism Software is attached .

Pankaj Kumar Srivastava

Signature of the Internal Guide

Signature of External Guide/Co Guide

Mohammad Ashfaq

Signature of the Scholar

BATHYMETRIC ANALYSIS USING MULTI-SPECTRAL SATELLITE DATA THROUGH MACHINE LEARNING ALGORITHMS IN VARYING COASTAL REGIONS OF INDIA

ORIGINALITY REPORT

4%

SIMILARITY INDEX

1%

INTERNET SOURCES

4%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

1

Mohammad Ashphaq, Pankaj K Srivastava, D Mitra. "Review of near-shore satellite derived bathymetry: classification and account of five decades of coastal bathymetry research", Journal of Ocean Engineering and Science, 2021

Publication

2%

2

Mohammad Ashphaq, Pankaj K. Srivastava, D. Mitra. "Evaluation and performance of satellite-derived bathymetry algorithms in turbid coastal water: A case study of Vengurla rocks", Journal of Sea Research, 2022

Publication

1%

3

Mohammad Ashphaq, Pankaj K. Srivastava, D. Mitra. "Analysis of univariate linear, robust-linear, and non-linear machine learning algorithms for satellite-derived bathymetry in complex coastal terrain", Regional Studies in Marine Science, 2022

Publication

1%