REMOTE SENSING OF DYNAMICS AND ABOVEGROUND BIOMASS OF SEAGRASS IN TAURANGA HARBOR (BAY OF PLENTY), NEW ZEALAND

A thesis

submitted in fulfilment

of the requirements for the degree

of

Doctor of Philosophy in Earth Sciences

at

The University of Waikato

by

Nam Thang Ha

THE UNIVERSITY OF
WAIKATO
Te Whare Wānanga o Waikato

2021

Abstract

Seagrasses are angiosperm plants that are completely adapted to life in seawater. They are distributed widely across the climatic regions, ranging from the tropics to temperate regions, in both inter-tidal and sub-tidal zones. Multiple ecosystem services are recognized as being supported by seagrass meadows, which recently has included appreciation role in carbon sequestration. Seagrass meadows, however, have been degraded in both terms of area and habitat quality across the globe, leading to a significant loss of ecosystem services and human livelihood support. This ongoing degradation has resulted in an urgent need to develop tools for assessing the temporal changes of extant meadows and accurate estimation of seagrass biological parameters, which will contribute to a sustainable conservation strategy into the future.

This thesis describes the use of a range of freely available Earth observation products, including multi-spectral imagery from Landsat and Sentinel-2, and synthetic aperture radar (SAR) products from Sentinel-1, coupled with a range of machine learning (ML) and meta-heuristic optimization algorithms, to develop novel and advanced techniques for remote sensing of seagrass. The work used field validation data from Tauranga Harbor, New Zealand, and specifically targeted mapping, change detection, and estimation of seagrass distribution and biomass.

The relatively small and patchy meadows of *Zostera muelleri* in this harbor can be mapped using a three-category classification (dense, sparse and none) with up to 91% accuracy for dense and 75% for sparse meadows using the machine learning algorithm Rotation Forest with Sentinel-2 imagery. Despite a slightly lower accuracy (90%), the algorithm Canonical Correlation Forest also shows merit for categorical seagrass mapping.

Historic Landsat multispectral satellite data used with ML models was able to map accurately the change in distribution of seagrass meadows over 29 years (1989 - 2019). For this binary mapping application (presence/absence) the CatBoost model obtained over 90% accuracy. Historic imagery indicates an approximately 50% of seagrass loss, from 2,424 ha in the year 1989 down to 1,184 ha in the year 2019 in Tauranga Harbor. Most of the early loss was from the northern and southern parts of the harbor and results were consistent with published estimates of change based on aerial photography.

In addition, a mapping scheme of seagrass distribution was developed from SAR data and a fusion of the multi-spectral and SAR data was developed for seagrass aboveground biomass (AGB) estimation. Optimal results were obtained using a combination of ML methods and meta-heuristic optimization. The seagrass distribution was mapped at an accuracy over 90% using the Extreme Gradient Boost (XGB) whilst the AGB map at 10 m spatial resolution was produced at 75% accuracy using the XGB model together with Sentinel-2 images and Particle Swarm Optimization (PSO).

The last part of the thesis describes the development of a web-based application to allow the advances in this research to be shared with a broader community and strengthen international and domestic collaboration in seagrass protection and conservation.

This study provides in-depth and advanced methods for seagrass resource inventory, maximizing the utilization of remotely sensed data, state-of-the-art ML and metaheuristic optimization algorithms to accurately map distribution and estimate desired biophysical parameters. The proposed methods are open-source and applicable across the globe, providing a complete toolset for both scientist and managers in aquatic resource management.

Acknowledgements

I would like to have my sincere appreciation to Prof. Ian Hawes and Prof. Merilyn Manley-Harris who are my excellent supervisors with a special enthusiasm for my PhD research. A special appreciation to my wife, Hoang Thi-Diem-Trinh, and my son, Ha Thanh-Binh who shared with me every happiness and challenge moment during my PhD journey. With respect, this thesis is a gift to Prof. Marnie Campbell who accepted my initial proposal and actively supported me during the application for doctoral scholarship at the University of Waikato, to my older brother, Dr Pham Tien-Dat, who gave me invaluable advice and motivation for each of research papers. This thesis would not have been completed without the sincere and enthusiastic supports of Holly Ferguson, Alice Morrison, Christine E.C. Gunfield, Matthew J. Finnigan, Stef Hendra, Shanelle Kivell in the Marine Field Station (University of Waikato) and my Vietnamese friends, Lam Suu, Hong Phuoc-Toan, To Dinh-Truong, Ho Thi-Hai-Ly, Truong Quoc-Cuong during the intensive field trip in Tauranga Harbor. Finally, I would like to present the PhD thesis to my father, Ha Thanh-Hao, and my mother Nguyen Thi-Chien who dedicated all their life to our higher education and our happiness.

Table of Contents

List of figures

List of tables

List of Abbreviations & glossary of technical terms

Abbreviation	Full
8A-V	8A vegetation
A	Accuracy
ACOLITE	Atmospheric correction for operational land imager (OLI) “lite” toolbox
AdaBoost	Adaptive boosting
ALOS-2 PALSAR-2	Advanced land observing satellite 2 (ALOS-2) phased array type L band synthetic aperture radar (PALSAR-2)
AGB	Aboveground biomass
AIC	Akaike information criteria
ANOVA	Analysis of variance
API	Application programming interface
ANN	Artificial neural network: A neural network-based remote sensing imagery classification technique
ALOS AVNIR	Name of a multispectral, medium spatial resolution remote sensing imagery
ASTER	Advanced spaceborne thermal emission and reflection radiometer
BI	Brightness index
BIC	Bayesian information criteria
BRI	Bottom reflectance index
CI	Color index
CASI	Compact airborne spectrographic imager
CB	CatBoost
CBERS	China – Brazil earth resources satellite
Case 1 water	A classification of water type for which phytoplankton control the optical properties
Case 2 water	A classification of water type for which phytoplankton, yellow sub-stance and mineral particles control optical properties

CCF	Canonical Correlation Forest
CCT	Canonical Correlation Tree
CCA	Canonical Correlation Analysis
CNN	Convolutional neural network
RNN	Recurrent Neural Network
DT	Decision tree: A remote sensing imagery classification technique in which a tree structure is built to present a decision of classification/regression
DART	Dropouts meet multiple additive regression tree
Endmember classification	A spectrum-based remote sensing imagery classification technique
EO	Earth observation
ESA	European space agency
ETM	Enhanced thematic mapper
EVI2	Enhanced vegetation index 2
Fuzzy membership	A membership function in fuzzy logic
FFA	Firefly algorithm
GA	Genetic algorithm
GB	Gradient boost
(GE)OBIA	(Geographic) object based image analysis: an object-based remote sensing imagery classification technique
GBDT	Gradient boosting decision tree
GOSS	Gradient-based one-side sampling
GPS	Global positioning system
GLOVIS	Global visualization viwer
GNSPI	(Geostatistical neighbourhood similar pixel interpolator): a gap filling technique for Landsat-7 ETM
GNDVI	Green normalized difference vegetation index
GPU	Graphical processing unit

GIS	Geographic information system
GTP	Ground truth point
g DW m^{-2}	Gram dry weight m^{-2}
Hyperspectral imagery	A remote sensing imagery with hundreds of spectral bands
HHO	Harris Hawks Optimization
IRS LISS	Indian remote sensing satellite linear imaging self scanner
ISODATA	(Iterative self-Organizing data analysis technique algorithm): an unsupervised classification technique for image pixel clustering
InVEST	A modelling tool that integrated valuation of ecosystem services and tradeoffs
Ikonos	Name of a multispectral, high spatial resolution remote sensing imagery
IRECI	Inverted red edge chlorophyll index
K-NN	K-nearest neighbor: An object-based classification technique of remote sensing imagery
K-means	An unsupervised classification technique for image pixel clustering
Kappa coefficient	A coefficient which measures the consistence of classified classes in remote sensing classification
Kompsat	Name of a multispectral, high spatial resolution remote sensing imagery
Landsat	Land remote-sensing system (System, Landsat)
LGBM	Light Gradient Boosting Machine
LMT	Logistic Model Tree
NIR	Near infrared
MCARI	Modified chlorophyll absorption in reflectance Index
MLC	Maximum likelihood classifier: a remote sensing imagery supervised classification technique in which a maximum likelihood pixel is classified into corresponding class

ML	Machine learning
Mahalanobis distance classification	A supervised classification technique of remote sensing imagery with an assumption of equal covariance among the classes
NDI45	Normalized difference index derived from band 4, 5
NDVI	Normalized difference vegetation index
OLI	Operational land imager
PCA	Principle component analysis
PSO	Particle Swarm Optimization
P	Precision
QUICKBIRD	Name of a multispectral, high spatial resolution remote sensing imagery
RMSE	Root mean squared error
$RMSE_{\%}$	Root mean squared error percent of mean
R^2	Coefficient of determination
R	Recall
RE	Red edge
RF	Random Forest
RoF	Rotation Forest
RI	Red index
RVI	Ratio vegetation index
SNAP	Sentinel application platform
SAVI	Soil adjusted vegetation index
Spectral angle mapper	A spectrum-based remote sensing imagery classification technique in which an angle between the spectra is calculated to match pixel with reference spectrum
Segmentation classification	A segment-based remote sensing imagery classification technique in which homogeneous pixels are grouped into segments
SVM	Support Vector Machine: A machine learning-based remote sensing imagery classification technique
Supervised classification	A group of remote sensing imagery classification

	techniques using ground truth data to train the algorithm
Sentinel	Name of a multispectral, medium spatial resolution remote sensing imagery
SI	Soil index
SPOT	Name of a multispectral, high spatial resolution remote sensing imagery
SAR	Synthetic aperture radar
SWIR	Shortwave infrared
THEOS	Name of a multispectral, medium spatial resolution remote sensing imagery
TOA	Top of atmosphere
Unsupervised classification	A group of remote sensing imagery classification techniques which are software-based automatic classification without ground truth data
UAV	Unmanned airborne vehicle or drone
UTM	Universal transverse mercator
VH	Dual-polarization vertical transmitting and horizontal receiving
VI	Vegetation index
VV	Dual-polarization vertical transmitting and vertical receiving
VH_cons	VH contrast
VH_diss	VH dissimilarity
VH_homo	VH homogeneity
VH_asm	VH asm
VH_energy	VH energy
VH_max	VH max
VH_entropy	VH entropy
VH_mean	VH mean
VH_variance	VH variance
VH_corr	VH correlation
VV_cons	VV contrast

VV_diss	VV dissimilarity
VV_homo	VV homogeneity
VV_asm	VV asm
VV_energy	VV energy
VV_max	VV max
VV_entropy	VV entropy
VV_mean	VV mean
VV_variance	VV variance
VV_corr	VV correlation
VHR	Very high resolution
Worldview	Name of a multispectral, high spatial resolution remote sensing imagery
WGS	World geodetic system
XGB	Extreme Gradient Boost
XGB-PSO	Extreme Gradient Boost with Particle Swarm Optimization for feature selection

Chapter 1 General introduction and overview

1.1. Introduction

Seagrass is an angiosperm that has adapted to complete its life cycle in seawater. It frequently grows in dense intertidal or subtidal meadows, which provide several valuable ecosystem services (Nordlund et al., 2016), including breeding and nursery grounds, water quality improvement, wave energy attenuation and coastline stabilization. In recent years, blue carbon initiative (Fourqurean et al., 2012; Macreadie et al., 2019), which indicates the sequestrated carbon inside such meadows, has provided a new perspective on the ecological importance of this ecosystem, in terms of mitigating global warming (Ramesh et al., 2018). The seagrass resource, however, is reported to be in significant decline globally (Waycott et al., 2009), which might lead to a mass loss of ecosystem services (Nordlund et al., 2016; Githaiga et al., 2019), enhanced CO_2 emission (Salinas et al., 2020), and a reduction in its contribution to climate change mitigation. The current degradation has led to a need for accurate mapping and monitoring tools (Unsworth et al., 2018; Dat Pham et al., 2019) to retrieve more detailed, real time information on seagrass distribution and biophysical parameters that will support monitoring, reporting, and verification (MRV), to inform sustainable restoration and conservation of this resource.

Several attempts have been made to map and monitor extant seagrass meadows (Dat Pham et al., 2019), with significant efforts on field data investigation, and accuracy improvement of retrieval models for seagrass detection and biophysical parameter estimation using remote sensing approaches. Traditionally, field-based sampling and analysis was preferred due to the high accuracy in measurement and estimation of selected parameters (Winters et al., 2017). This

approach, however, is time consuming, has high labor costs and is only applicable to a small spatial area. The application of space-born remote sensing, on the contrary, compensates for the drawbacks of the field-based approach, and has led to a rapid development of various types of image processing, classification, and modeling during the last decade (Devi et al., 2015; Joshi et al., 2016). Despite the successful employment of remote sensing-based mapping, numerous limitations have existed and are discussed in the literature (Xie et al., 2008; Kuenzer et al., 2011; Joshi et al., 2016). The main drawbacks include the poor spatial resolution of most satellite instruments, complexities in satellite image processing, impacts of the atmosphere and water column on the image pixel values, and an overall low accuracy of retrieval models, which leaves various gaps for further studies. The cost of many higher resolution satellite products is also a barrier to update application to the seagrass ecosystem.

The literature review in Chapter 2 revealed a limited number of research papers describing seagrass mapping from Sentinel-2 imagery (Hossain et al., 2015; Dat Pham et al., 2019), which is free-of-charge to access and offers up to 10 m spatial resolution. Even fewer studies used modern machine learning (ML) models to improve the accuracy of mapping. In addition, long-term observation of seagrass dynamics is not always conducted internationally and the 29 years of change detection (1989 - 2019) described in Chapter 4, is the first such study conducted for seagrass meadows in New Zealand. We also found that the estimation of *Z. muelleri'*s aboveground biomass (AGB) was either high cost in intensive field data collection/ using commercial satellite image or low accuracy in retrieval models (Roelfsema et al., 2014; Unsworth et al., 2018).

In this study, which was conducted in Tauranga Harbor (New Zealand), I attempted to develop novel, low cost, open-source and reliable methods for accurate mapping and long-term change

detection of smaller sized and patchy seagrass meadows, in parallel with accurate estimation of seagrass AGB. This was undertaken using a combination of freely available remotely sensed data, machine learning (ML) models, and meta-heuristic optimization. Landsat, Sentinel-2 multi-spectral and Sentinel-1 synthetic aperture radar (SAR) satellite images were the main data sources used. State-of-the-art and open source image processing and modeling frameworks were applied to deliver the most accurate and reliable mapping/ retrieval of selected parameters in the study. In addition, a demonstration of a web based geographical information system (web-GIS) application is presented to visualize the historical and near-real time seagrass data, which support collaboration, and provide readily accessible information source a better reference for policymakers.

The objectives of this research involved:

- Large-scale mapping and long-term change detection of *Z. muelleri* in Tauranga Harbor (New Zealand) using Landsat, SAR (Sentinel-1) and Sentinel-2 satellite images and machine learning.
- Estimation of aboveground biomass (AGB) of seagrass meadows in Tauranga Harbor (New Zealand) using fused multi-spectral (Sentinel-2) and SAR (Sentinel-1) images, machine learning, and meta-heuristic optimization.
- Creation of an historical database of seagrass distribution and biomass, toward a dynamic web-based system for visualization and query.

1.2. Study site

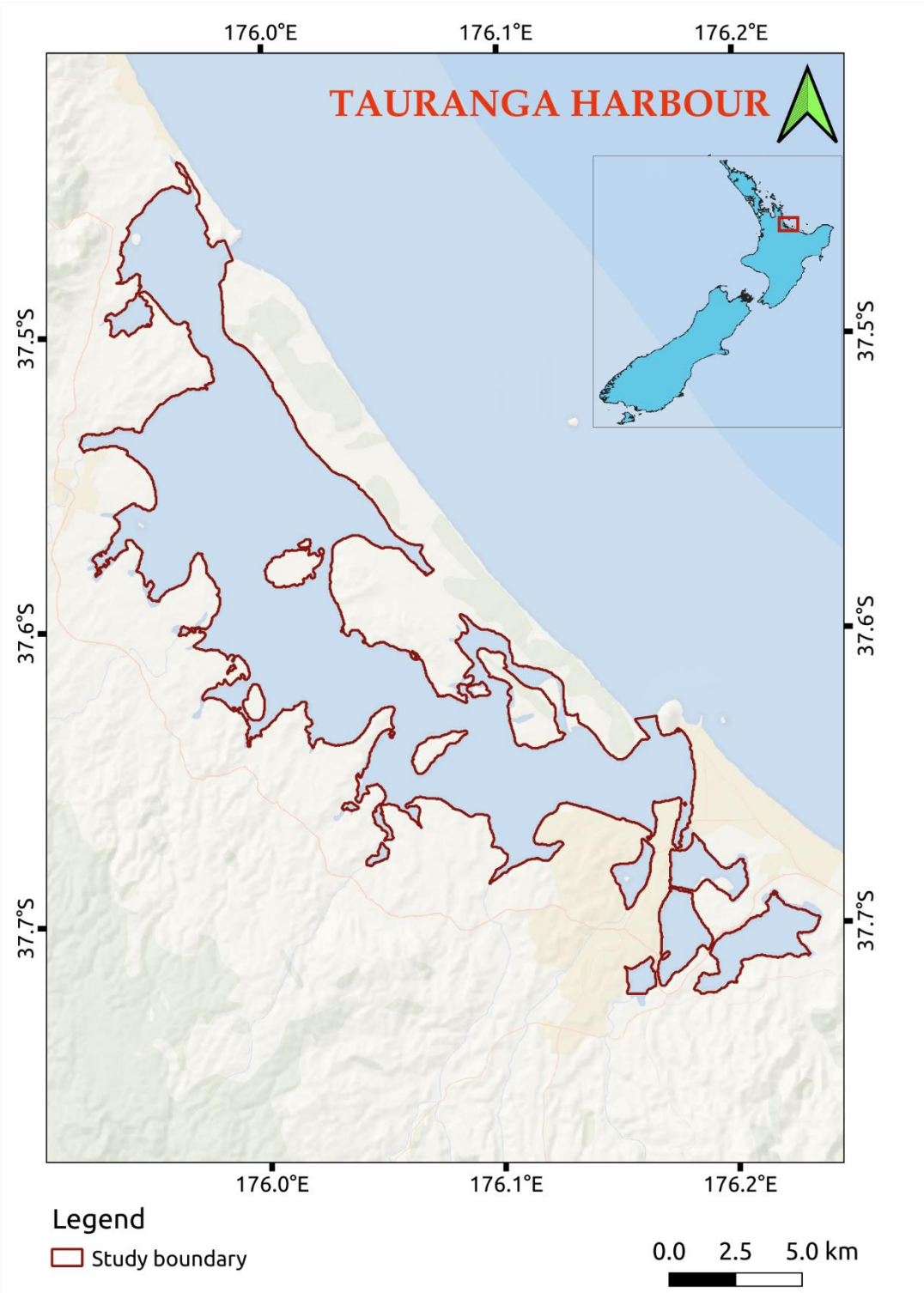


Figure 1.1. Tauranga Harbor - Study site

The study site is located in Tauranga Harbor (Bay of Plenty, New Zealand, Figure 1), a habitat that supports more than 1,184 ha (Ha et al., 2020) of seagrass. The site was selected due to the following reasons:

- Largest area of seagrass in New Zealand with extant patchy and scattered meadows
- Well-known tidal ranges
- Accessibility during low tide
- Soft substratum (better for soil core sampling)
- Availability of secondary environmental data and aerial images that is valuable for change detection and accuracy assessment
- A popular place for tourism, with agriculture, and horticulture activities leading to a potential risk of high sediment conditions.
- Homogeneous surface with known water depth, facilitate good field data collection for spatial analysis of seagrass biomass, and carbon content for relating to remote sensing data.
- Previous mapping has been conducted at different timelines, but with different sensors and processing techniques without a robust accuracy assessment (Park, 2011). This approach may lead to uncertainty in detection of change, and therefore motivated us to unify the sensors, spatial resolution, and image processing with state-of-the-art processing workflows to derive the most reliable and applicable approaches for seagrass mapping and retrieval of AGB in Tauranga Harbor.

1.3. Thesis structure

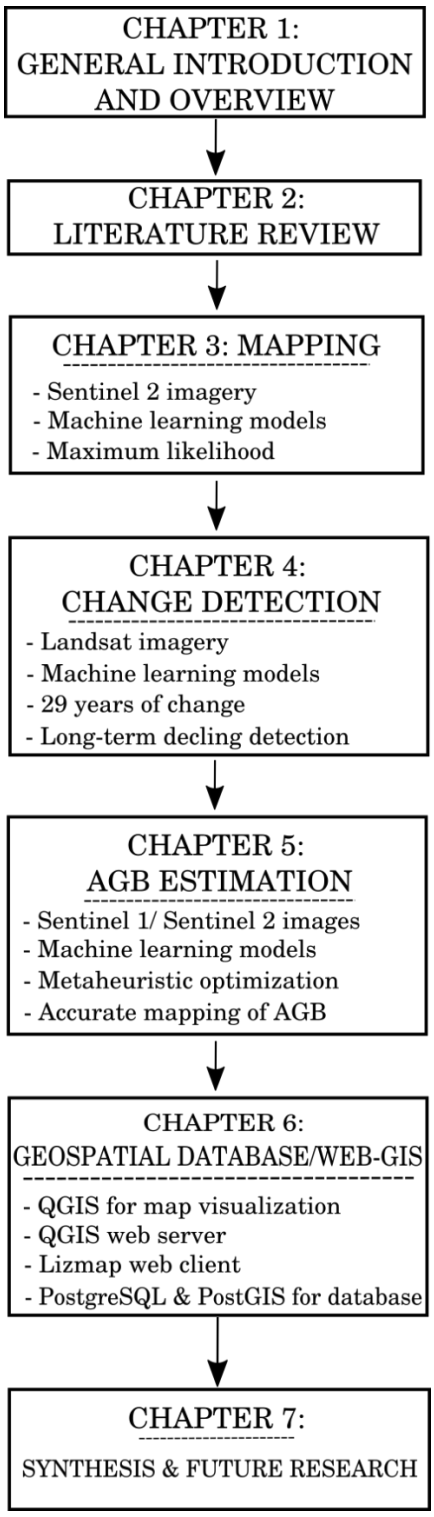


Figure 1.2. Thesis structure and flow-chart of the research

The thesis is structured as seven chapters, including the introduction (Chapter 1), literature review (Chapter 2), research findings (Chapter 3, 4, 5, 6), synthesis and future research (Chapter 7). The first two chapters introduce the general context, motivation of the research, current gaps in the literature, and research questions (Chapter 2). The following chapters emphasize contributions to the development of novel and open-source methods for accurate mapping of sparse, patchy seagrass meadows (Chapter 3), reliable and long-term change detection of seagrass meadows (Chapter 4), consistent and accurate mapping of seagrass distribution and AGB using fusion data of multi-spectral and SAR images (Chapter 5), geospatial data visualization and collaboration (Chapter 6). The final chapter concludes the main findings of the research and suggests cutting-edge future research topics, using remotely sensed data, ML models, and meta-heuristic optimization for accurate mapping and effective management of aquatic resources.

The next paragraphs briefly describe the content of each chapter in the thesis.

Chapter 1. General introduction and overview

Chapter 1 presents the necessity and state of research. In this chapter, the importance of the seagrass ecosystem is introduced, specifically the general current gaps for seagrass mapping, change detection, and estimation of seagrass AGB. Then, the research plan is presented together with the summary of the main finding of each research chapters.

Chapter 2. Literature review

Chapter 2 summarizes previously reported research for seagrass mapping and change detection, and biomass estimation using space-borne remote sensing data. The research methodology and results are analyzed, the limitations are discussed to determine the current gaps in the field of

study. Then, the following chapters attempt to answer the given questions in a format suitable for submission as research papers, two of which are already published.

Chapter 3. A comparative assessment of ensemble-based machine learning and maximum likelihood methods for mapping seagrass using Sentinel-2 imagery in Tauranga Harbor, New Zealand

This chapter has been published as a peer-reviewed paper in the Remote Sensing journal (https://www.mdpi.com/2072-4292/12/3/355/htm). The published paper is attached as Appendix 2.

Chapter 3 focuses on the mapping of *Z. muelleri* distribution using Sentinel-2 imagery and ML models. This paper develops a novel mapping technique using ML models (Random Forest, Rotation Forest, Canonical Correlation Forest) for the small size, patchy seagrass meadows in Tauranga Harbor, in comparison with the more traditional method (Maximum Likelihood). The results contribute novel, accurate mapping methods for seagrass which use open sources data and models, is reliable, and applicable to other species across various regions in the world.

Chapter 4. Detecting multi-decadal changes in seagrass cover in Tauranga Harbor, New Zealand, using Landsat imagery and boosting ensemble classification techniques

This chapter has been published as a peer-reviewed paper in the ISPRS International Journal of Geo-Information (https://www.mdpi.com/2220-9964/10/6/371). The published paper is attached as Appendix 4.

Chapter 4 attempts to understand spatial and temporal change of seagrass meadows in Tauranga Harbor, which is impacted by rapid development of agriculture in the northern part and urbanization in the southern part of the harbor. ML models were applied to historical Landsat

data over 29 years (1990 - 2019), and methods developed for fitting to historical Landsat images and patchy seagrass meadows in the study site. Our results detected a long-term decline of seagrass area, motivating the suggestion to use long-term Earth observation data (Landsat imagery) and ML models for large-scale mapping of temporal changes of seagrass ecosystem globally.

Chapter 5. The use of radar and optical satellite imagery combined with advanced machine learning and meta-heuristic optimization techniques to detect and quantify above ground biomass of inter-tidal seagrass in a New Zealand estuary

This chapter has been published as a peer-reviewed paper in the International Journal of Remote Sensing (http://dx.doi.org/10.1080/01431161.2021.1899335). The published paper is attached as Appendix 6.

Chapter 5, for the first time, attempts the binary mapping of seagrass meadows using SAR image, and accurately estimates the seagrass aboveground biomass (AGB) using a combination of SAR and multi-spectral remotely sensed data, ML models, and meta-heuristic optimization. We initially applied the SAR data (*i.e* Sentinel-1 imagery) to map the seagrass distribution in Tauranga Harbor. We then developed novel approaches integrating both multi-spectral and SAR data with the deployment of ML models and meta-heuristic optimization for the retrieval of the seagrass AGB in the study site. The paper contributes advanced and general methods for seagrass mapping from SAR data and accurate estimation of seagrass AGB in the field.

Chapter 6. A novel and open source web-GIS approach for seagrass data visualization and collaboration

In chapter 6, a local host demonstration of the open source geospatial database and web-based interface was introduced for seagrass distribution and AGB visualization, data query and collaboration. The motivation was to develop very low-cost approaches that will allow remote sensing to be used widely for seagrass mapping. It implements the QGIS desktop for geo-spatial data visualization, uses a QGIS server for QGIS desktop and web client connection, PostgreSQL for geo-spatial database management, PostGIS for data connection between QGIS desktop and PostgreSQL, and Lizmap to publish the spatial data to the world wide web. This approach is free to access, ease to implement, and is capable of further extension to various geo-spatial databases in the manner of a secure, elegant, and fast loading web-based service. This chapter contributes a very simple approach to seagrass mapping, which is suitable to different levels of management, is open-access, and stable in implementation.

Chapter 7. Synthesis and future research

Chapter 7 synthesize important results of the research and derives the essential recommendations. It synthesizes the key points of large-scale seagrass mapping and change detection using remote sensing data and ML models, that were developed in this work. The important findings for accurate mapping from SAR data and estimation of seagrass AGB are presented, and suggestions made for cutting-edge research topics in the future.

Chapter 2 Literature review

2.1. An introduction to seagrass in New Zealand

Seagrasses are a polyphyletic group of angiosperm plants that are adapted to grow in brackish and seawater and are widely distributed in intertidal and shallow sub-tidal waters around much of the world. Seagrass provides many valuable ecosystem services, with an estimated total economic value of approximately 1.9 trillion US dollar per year (Waycott et al., 2009). Currently, 24 ecosystem services have been described (Nordlund et al., 2016), including acting as an aquatic habitat, as a foodstuff, nursery, as a form of carbon sequestration, as a raw material, to aid coastal protection and sediment stabilization, with potential for mariculture, for water purification, and for more indirect uses as a cultural artifact, for education, and tourism.

In New Zealand, there is a single seagrass species, *Zostera muelleri* (Park, 2011; Wilton et al., 2016). It is found in both sub-tidal and intertidal zones, but mostly through intertidal areas, from Parengarenga harbor in the north to Stewart Island in the south (Green & Short, 2003; Tara et al., 2019). Table 2.1 and Figure 2.1 present the location and area of *Z. muelleri* in New Zealand, and, where possible, indicate recent change. Short-term monitoring shows no significant change in seagrass area. However, long-term mapping of seagrass distribution presents a different story in which seagrass area change was detected from aerial mapping over a period of 20 - 60 years (Park, 2011; Matheson & Wadhwa, 2012).

Table 2.1. Seagrass distribution and area in New Zealand

(Where data are available for more than one year, deviation of change is indicated as + (increased), - (decreased) or (0) no change)

Location	Year	Area (ha)	Change
Tauranga Harbor	1959[2]	4,424	
	2011[2]	2,744	-
Ōhiwa Estuary	2003[1]	107	
	2011[2]	100.5	-
South Bay (Slipper Island)	2004[2]	3	
Huruhi Bay and Parapara (Great Mercury Island)	2004[2]	7	
Mahurangi Harbor	2003[1]	3	
Whangateau Harbor	2003[1]	33	
Pahurehure Inlet	2003[1]	0	
Arm of Kaipara Harbor	2003[1]	0	
New River Estuary	2003[1]	94	
Matakana Harbor	2003[1]	0	
Whitianga Harbor	2003[1]	5	
	2008[2]	8.94	+
Tairua Harbor	2003[1]	125	
	2008[2]	130.08	+
Wharekawa Estuary	2003[1]	50	
	2008[2]	45.28	-
Otahu Estuary	2003[1]	0.2	
	2007[2]	0.6	0
Te Kouma Estuary	2003[1]	5.2	
	2009[2]	4.66	0
Firth of Thames	2003[1]	33	

1 Data was collected from World Atlas of Seagrass (Green & Short, 2003)

2 Data was collected from Waikato Region Council at https://www.waikatoregion.govt.nz/

Location	Year	Area (ha)	Change
	2006[1]	0	-
Waimea Estuary	2003[1]	28	
Havelock	2003[1]	0.9	
Whanganui Inlet	2003[1]	859	
Avon-Heathcote Estuary	2003[1]	13.7	
Kaikorai Estuary	2003[1]	0	
Harwood, Otago Harbor	2003[1]	82	
Coromandel Harbor	2009[2]	120.8	
Manaia Harbor	2003[2]	27	
	2008[2]	19.35	-
Tairua Harbor	2008[2]	130.8	
Whangamata Harbor	2003[2]	51	
	2007[2]	59.97	+
Whangapoa Harbor	2010[2]	202.71	
Wharekawa Harbor	2008[2]	45.28	
Whitianga Harbor	2009[1]	8.94	
Waikawau Estuary	2008[1]	0	
	2016[3]	0.01	0
Te Kouma Harbor	2009[1]	4.66	
Purangi Estuary	2012[1]	0	
Otama River Mouth	2010[1]	0.94	
Kennedy Estuary	2012[1]	0.25	
Aotea Harbor	2012[1]	584.16	
Kawhia Harbor	2012[1]	842.97	
Waikato River Estuary	2011[1]	0.82	
Colville Bay	2012[2]	0.25	
Kennedy Estuary	2012[2]	0.25	

3 (Robertson & Stevens, 2016)

Location	Year	Area (ha) Change
Tautane	2016[3]	0.8
Porirua Harbor	1942[4]	24
	2008[4]	54.8

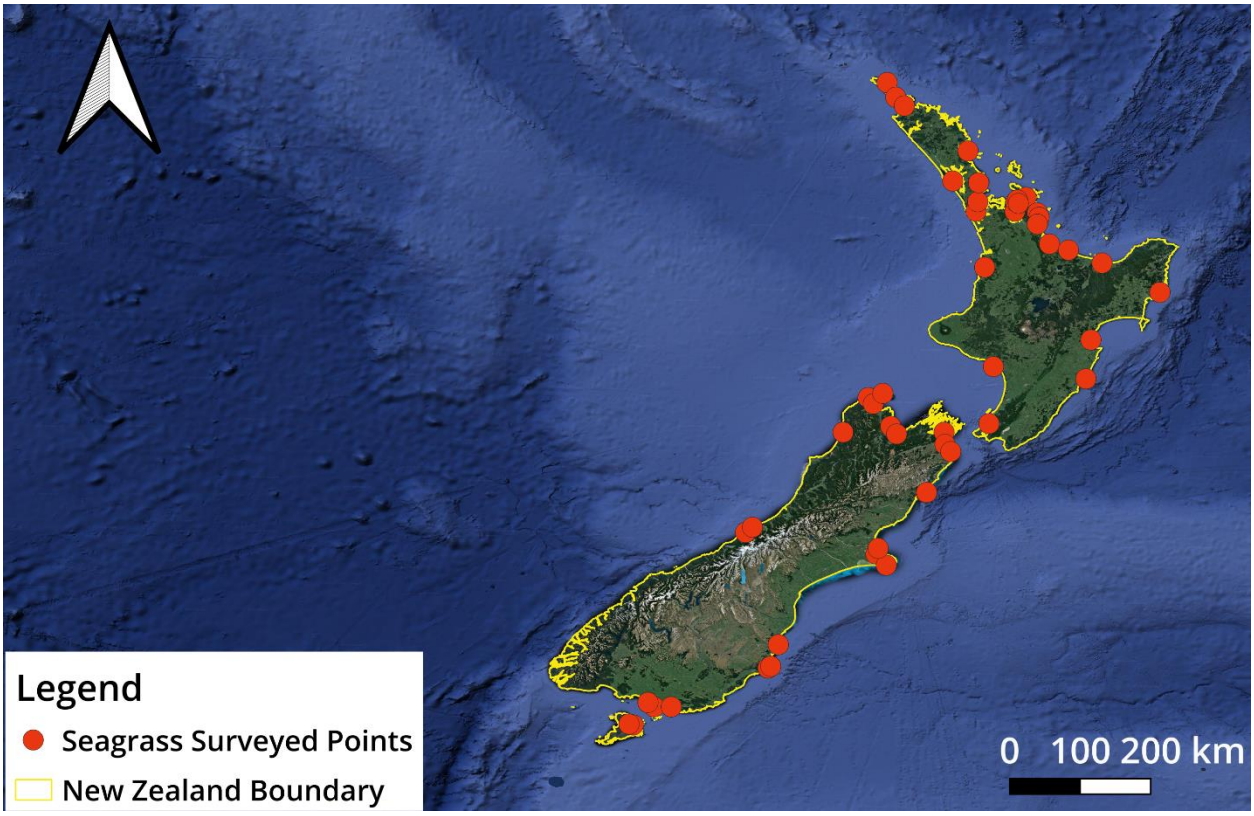


Figure 2.1. Map of locations of seagrass in New Zealand (satellite-based map from Google Earth). The location (orange dots) is referenced from the Global Seagrass Database (UNEP-WCMC, 2018)

Despite a wide geographical distribution of seagrass in coastal waters, there has been little research into mapping of seagrass dynamics in New Zealand (Dat Pham et al., 2019). Seagrass mapping has been conducted in only 22 out of 300 estuaries in New Zealand, representing less than 3% of total estuarine area (Tara et al., 2019). The number of mapped areas has increased

4 (Matheson & Wadhwa, 2011)

recently owing to airborne image-based mapping projects in several regions of New Zealand, though these tend to give limited detail on image processing and classification accuracy (Park, 2011; Matheson & Wadhwa, 2012). The total known area of seagrass in New Zealand to November 2018 was reported as 294 km^2 (Tara et al., 2019). Seagrass is mentioned in only 2 out of 16 marine reserve reports, with an area of 8.59 km^2 in Whanganui Inlet and Te Angiangi.

The most recent published papers on seagrass in New Zealand can be divided into four main topics: mapping using aerial imagery, ecology, restoration and ecosystem service and associated fauna (Table 2.2).

Table 2.2. Published papers on seagrass in New Zealand

Topic	Location	Citation
Ecology, restoration, associated fauna	Worldwide	(Green & Short, 2003)
Ecology, associated fauna	Slipper Island and Great Mercury Island, Coromandel Peninsula in New Zealand	(Schwarz et al., 2004)
Ecology, associated fauna	New Zealand	(Schwarz & Turner, 2006)
Ecology, associated fauna	Whangapoua Harbor, Wharekawa Harbor, Whangamata Harbor in Coromandel Peninsula, New Zealand	(Turner, 2007)
Ecology, associated fauna, ecosystem service	Mangawhai to Ahipara, North land, New Zealand; New Zealand	(Kerr, 2009; Wotton, 2009)
Mapping using aerial imagery	Tauranga Harbor, New Zealand; Porirua Harbor, New Zealand	(Park, 2011; Matheson & Wadhwa, 2012)
Ecology, associated fauna, ecosystem service	New Zealand	(Cole, 2014)

Topic	Location	Citation
Ecology, associated fauna, mapping using aerial imagery	Manawatu Estuary, New Zealand; New Zealand	(Robertson & Stevens, 2016; Wilton et al., 2016)
Ecology, associated fauna, ecosystem service	New Zealand	(Tara et al., 2019)

While acknowledging the important ecosystem services of seagrass, none of these New Zealand-based studies have provided a detailed description of either large-scale seagrass mapping and change detection using space borne satellite imagery or seagrass blue carbon assessment in New Zealand waters.

Large scale mapping of seagrasses in New Zealand is in its infancy. In 2009, a technical report presented, for the first time, a habitat map of Northland (Kerr, 2009). Aerial photographs and sonar data were interpreted to deliver bathymetry as well as a habitat map in tidal, sub-tidal and adjacent regions. According to the report, seagrass colonized 5,192 ha or 0.38% of the total area, but it was stated that further works were necessary to improve field survey data, map accuracy, and object interpretation from aerial imagery. In a later report, a detailed map of coastal habitats in several regions in New Zealand was released (DOC, 2011), but these lack seagrass, mangrove and salt marsh in the framework of classification and mapping. Other research has grouped seagrass and algae in one category to value ecosystem services in a marine protected area (Cole, 2014). The latest report (Tara et al., 2019) presented a systematic review on seagrass dataset collection and summarized the biogeographic distribution maps in Kaipara Harbor, Kawhia Harbor, Waitemata Harbor, Parengarenga Harbor, Rangaunu Harbor, Tairua Harbor, Tauranga Harbor, southern Farewell Spit, Golden Bay, and Porirua Harbor conducted by the SeaSketch project covering the period 1942 - 2016. However, the state of most recent seagrass conditions

(including the current distribution, biomass, carbon storage) is less documented, and where it is this is usually mapping at a local scale using aerial images. The declining trend of seagrass in New Zealand which was mentioned in the international (UNEP-WCMC, 2018) and local (Tara et al., 2019) reports, lacks a systematic assessment and a rich dataset on these topics is still not available.

In New Zealand, the largest threat to seagrass beds is from sedimentation due to urbanization, deforestation and rural land management practices (Schwarz et al., 2004). Additionally, nitrate and phosphate enrichment, recreational activities, wasting disease (a fungal disease which causes infection of seagrass tissue) together with grazing of the black swan threaten the integrity of seagrass meadows (Park, 2011). In many locations within New Zealand, declining trend in area is shown to have occurred during the 20th century. In the Bay of Plenty, for example, mapping in the years 1945, 1959, 1992, 1996, and 2011 showed a significant loss of seagrass area in the period of 1959 - 1996 (over 1000 ha) with a slower rate of loss to 2011. There was a 3.8%, which is equivalent to 43 ha, seagrass gain in the southern part of Tauranga harbor during the period 1996 - 2011. Grazing of the black swan and poor water quality were determined as the main reasons for this loss (Park, 2011). In Southern Bay, Slipper Island, seagrass area had lost approximately 65% by 2004, when compared to the year 1973 (Schwarz et al., 2004). In Whangamata harbor, aerial photos presented a roughly 28% increase of seagrass area during the period of 1944 - 1965, however, this reduced approximately 41% from 1965 - 1998 (Schwarz & Turner, 2006). Similarly, there has been a 40% loss between 1960 - 1980 in Porirua harbor (Matheson & Wadhwa, 2012).

This picture of declining areal coverage of seagrass beds in New Zealand is similar to other parts of the globe. To deal with identified problems, several policies have been introduced by the

Department of Conservation. This list involves the Wildlife Act (1953), Reserves Act (1977), Marine Reserves Act (1971), The Resources Management Act (1991), and The New Zealand Fisheries Act (1996) that all gave seagrass a better protection. Our review has determined three gaps regarding the overall management of seagrass resources that remains in New Zealand: (i) the lack of (large-scale) seagrass mapping and change detection tools; (ii) (large-scale) mapping of biophysical parameters of seagrass meadows is not available; (iii) and the lack of a web-based and geospatial tool to improve the public awareness of seagrass decline and foster community-based protection.

2.2. Remote sensing of seagrass mapping

Remote sensing refers to collection of information from the Earth's surface through the operation of instruments on board airborne or satellite platforms. In recent decades, the technology of remote sensing has developed dramatically due to the significant improvement of a range of supporting technologies: satellite construction and launching, database structure, image storage, processing and classification algorithms (Devi et al., 2015). In several fields, remote sensing provides essential mapping that significantly supports managers to make informed decisions. The following review focuses on remote sensing-based mapping with special links to seagrass ecosystems.

Imaging satellites receive reflectance signals from the Earth's surface at their sensors, usually divided across the electromagnetic spectrum into bands. Typical bands, for example used for creation of "normal" images can include ρ_{Red}, ρ_{Green} and ρ_{Blue} bands. The characteristics of the reflectance intensity across these bands provides data for the identification of different objects on the surface. In terms of remote sensing technique, this process is conducted with image processing and classification schemes. Use of satellite imagery has the advantages of low cost,

wide coverage and, for some platforms, the possibility of historical mapping as compared to classical, ground-based monitoring. However, this process is not simple due to the effects of the shape of the Earth, the atmosphere and water column on reflectance value. These factors may lead to a significant confusion in the classification process, particularly when applied to seagrasses at the land-water interface. Complex conditions such as unclear boundaries of aquatic ecosystem, the diversity of bottom types, large scale distribution of seagrass in different geographical regions and varying water depths mean that the challenges of image processing and classification for remote sensing of seagrasses requires advanced techniques for mapping and monitoring (Duarte et al., 2013). To date, gaps in the seagrass mapping database remain in several regions in the world, including Africa, South America, South East Asia and other parts of the Southern hemisphere. In the following sections, the most important factors that significantly contribute to effective seagrass mapping are explained.

2.2.1. Light attenuation in the water column and water column correction techniques for underwater remote sensing

Imaging the sea floor through an overlaying water column is particularly difficult. In the water environment, light is attenuated due to absorption, scattering, and reflection processes. Depending on the energy intensity, each spectral component is able to penetrate to a specific depth or be completely absorbed in the water. In shallow water, ρ_{Blue}, ρ_{Green}, ρ_{Red}, and near infrared (ρ_{NIR}) bands in visible and infrared spectrum penetrate reasonably well and provide useful information on substratum (Green et al., 1996). Other bands attenuate very fast with water depth due to strong absorption and scattering of the water column and therefore only leave noise in the image pixels. Light attenuation, fundamentally, depends on water depth. In the ocean (case 1 water), light is mostly absorbed by water and phytoplankton, whilst in the coastal zones (case 2

water), absorption is affected by colored dissolved organic matter (CDOM) and suspended material (ZhongPing Lee & Hu, 2006). The maximum penetration depth (case 1 water) of ρ_{Blue}, ρ_{Green}, and ρ_{Red} bands is 15 m, 10 m, and 5 m respectively. The ρ_{NIR} band is more strongly absorbed and only reliable at depths less than 2 m (Green et al., 1996). This leads to the necessity of water column correction if remote sensing data are to be applied to a submerged surface (Chavez, 1988; Mumby et al., 1998; Zoffoli et al., 2014). This process may include various techniques that compensate for the impacts of water column constitution and water depth on satellite image pixel value and these are mentioned in many research papers (Maritorena et al., 1994; Green et al., 1996; Zhongping Lee et al., 1999; Lyzenga et al., 2006; Sagawa et al., 2010; Yang et al., 2010; Brown et al., 2011; Ha et al., 2012; Manessa et al., 2014; Zoffoli et al., 2014; Manuputty et al., 2017).

Depending on the clarity of the water environment, either an empirical or an analytical approach is preferred for accommodating water column attenuation. Empirical models (Mumby et al., 1998; Lyzenga et al., 2006; Sagawa et al., 2010) are utilized in the case of limited ground truth data in both case 1 and case 2 water. However, when the water clarity is low, more comprehensive approaches known as semi-analytical, radiative transfer-based forward and look-up table spectrum matching inverse models provide an alternative solution. The latter approach is promising, but requires complex computation and intensive ground truth data. Empirical-based water column correction is, therefore, still a preferred solution for substratum reflectance retrieval. The method proposed by Lyzenga (1978, 2006) only requires the ρ_{Blue}, ρ_{Green}, ρ_{Red} bands of the image and the ground-truth data of the homogeneous substratum (Lyzenga, 1978; Lyzenga et al., 2006). An improved version involves the ρ_{NIR} band (Manessa et al., 2014) or bathymetry data (Sagawa et al., 2010) for additional correction. A third approach applies the

principal component analysis (PCA) technique (both of Inverse and Forward Principal Component Analysis) (Wicaksono, 2016; Manuputty et al., 2017) in parallel with a water column correction phase and empirical data. Under specific assumptions of water clarity, and at a limited water depth, this method is reliable and has been successfully applied in several regions for seagrass mapping (Green et al., 1996; Pasqualini et al., 2005; Dekker et al., 2006; Ha et al., 2012; Zoffoli et al., 2014).

2.2.2. Airborne and satellite imagery for seagrass mapping

In the realm of remote sensing, airborne and satellite platforms are the most commonly used (Table 2.3). Airborne imagery has the advantage of very high spatial resolution (with pixel size ranging from one centimeter to one meter), and very high spectral resolution across the ultraviolet to infrared range, referred to as hyperspectral (36 contiguous spectral bands in case of the widely used hyperspectral compact airborne spectrographic imager (CASI) sensor). Airborne sensing can avoid the complications of cloud cover, given that aircraft can fly below cloud or confine observations to cloud-free opportunities. Airborne imagery, however, covers a relatively small area, often with a low observing frequency, and may lead to challenges in larger scale monitoring or understanding comprehensive relationships between seagrass dynamics and external stressors (Hossain et al., 2015). In addition, airborne imaging requires a significant investment in equipment, image processing and data storage (Jia et al., 2020), leading to collaborative constraints in sharing or redistributing of data and can be impractical in developing countries. Observing from an extreme low altitude, unmanned airborne vehicle (UAV) has taken priority in recent years. Small in size, easy to control as well as providing a high spatial resolution image under different weather conditions, UAVs provide another option for seagrass mapping and monitoring at a very small scale (Konar & Iken, 2018; Tsouros et al., 2019; Yao et

al., 2019). Despite the advantages of high-quality hyperspectral images, high cost and small area coverage is obstructing further application of UAV.

On the other hand, satellite-based instruments represent a very different case. In recent decades, satellite imagery has been improving in spatial, spectral and temporal resolution as the result of technology evolution. Satellite-based instruments have advantages of low cost and wide coverage, but spectral and spatial resolution can be limited. For coastal habitat mapping, satellite sensors are categorized by spatial resolution larger than 100 m, 10 - 100 m interval, and below 10m. The above 100 m sensor measures sea surface temperature, phytoplankton biomass and provides information for fisheries management, toxic algal bloom monitoring, and anomaly warming phenomena resulting from climate variability. At an interval of 10 - 100 m, the image is most suitable for coastal management. High to very high spatial resolution image (below 10 m) provides the most reliable resolution for quantification of aquatic habitats, but few commercial sensors currently offer such resolution.

Table 2.3. Available airborne and satellite sensors for seagrass mapping

	Sensor	Acronym	Number of band	Pixel size (m)	Cost	Reference
1	Compact airborne spectrographic imager	CASI	36	4.4	Commercial sensor	(Phinn et al., 2008)
2	Unmanned aerial vehicle	UAV	3	< 0.1	Commercial sensor	(Yang et al., 2020)
3	Quickbird	Quickbird	5	0.6 - 2.9	Commercial sensor	(Satellite Imaging Corp, 2021e)
4	Worldview	Worldview	5	0.3 - 14.0	Commercial sensor	(Satellite Imaging Corp, 2021g)
5	Ikonos	Ikonos	5	0.8 - 4.0	Commercial sensor	(Satellite Imaging Corp, 2021c)

	Sensor	Acronym	Number of band	Pixel size (m)	Cost	Reference
6	ASTER	ASTER	14	15 - 90	Commercial sensor	(Satellite Imaging Corp, 2021b)
7	SPOT	SPOT	4	1.5 - 6	Commercial sensor	(Satellite Imaging Corp, 2021f)
8	Kompsat	Kompsat	6	0.5 - 5.5	Commercial sensor	(Satellite Imaging Corp, 2021d)
9	ALOS AVNIR	ALOS AVNIR	4	10	Commercial sensor	(Satellite Imaging Corp, 2021a)
10	Indian remote sensing	IRS	4	36 - 73	Commercial sensor	(ISRO, 2021)
11	Thailand Earth observation system	THEOS	5	2 - 15	Commercial sensor	(ESA, 2021c)
12	Landsat	Landsat		15 - 60	Free	(USGS, 2021c)
13	Sentinel	Sentinel	12	10 - 60	Free	(ESA, 2021b)
14	Earth observing one (EO-1) – advanced land imager (ALI)	EO1-ALI	10	10 - 30	Free	(USGS, 2021a)
15	Earth observing one (EO-1) - Hyperion	EO1-Hyperion	220	10 - 30	Free	(USGS, 2021b)
16	Synthetic aperture radar	SAR	2	22	Free	(ESA, 2021a)

In the case of seagrass mapping, Landsat and Sentinel images, with spatial resolution of 10 - 30 m, are suggested as the optimal choice to harmonize the cost and scale coverage, particularly given that the imagery is freely available (Dat Pham et al., 2019). At a higher spatial resolution, SPOT, Quickbird, Worldview, Ikonos all detect the distribution of seagrass more accurately than Landsat and Sentinel images, though, their spatial coverage is limited (Urbański et al., 2009). Dekker et al. (2003) utilized Landsat-5 TM and -7 ETM images to detect the variation of

seagrass area in Wallis Lake over 14 years (Dekker et al., 2003). Seagrass was also mapped with Landsat-8 OLI (Purnawan et al., 2016), ASTER, SPOT-4 and Kompsat-2 (Kim et al., 2015), SPOT-5 (Pasqualini et al., 2005). ALOS AVNIR-2 data provided good quality seagrass mapping in Tunisia, and Japan (Sagawa et al., 2010) as well as Brazil (Silva et al., 2017). Indian Remote Sensing (IRS) supplied 23.5 m spatial resolution images for seagrass mapping in Lakshadweep Island, India (Nobi & Thangaradjou, 2012) whilst THEOS data at 15 m supported a higher spatial resolution for this mapping in Phu Quoc island, Viet Nam (Nguyen et al., 2014). The very high spatial resolution (VHR) sensors may derive a higher accuracy of mapping, however there is a very high cost for purchasing the commercial imagery. Several research papers reported the uses of Worldview images (Tamondong et al., 2013; Roelfsema et al., 2014; Watkins, 2015; Baumstark et al., 2016; Candra et al., 2016; Manuputty et al., 2017; Su & Huang, 2019), Quickbird (Wolter et al., 2005; Urbański et al., 2009; Amran, 2010; Lyons et al., 2011; Roelfsema et al., 2014), Ikonos (Pu & Bell, 2017; Vela et al., 2008; Meyer et al., 2010) for seagrass mapping or to compare mapping performance among the VHR sensors (Wang et al., 2004; Kovacs et al., 2018).

In the realm of hyperspectral satellite imagery, Hyperion provides 30 m resolution across 242 bands from 400 to 2500 nm. These data are combined with other sensors with higher spatial resolution, involving Landsat TM, EO1-ALI, Ikonos (Meyer et al., 2010) or Landsat TM, EO1-ALI (Pu & Bell, 2013), for seagrass mapping. Other research papers include only Hyperion (Pu et al., 2012; Casal et al., 2013; Hedley et al., 2017). Most cases report a higher classification accuracy of Hyperion, both in terms of overall accuracy as well as Kappa coefficient (a consistence measurement of image classification results). This outcome is partly explained by the availability of a large number of spectral bands compared to multispectral sensors. These

hyperspectral sensors, however, are usually an on-demand request or in a specific geographic region (Heiden et al., 2017). Some studies focus on individual seagrass species mapping from hyperspectral imaging (Phinn et al., 2008; Valle et al., 2015; Pan et al., 2016; Hedley et al., 2017). Nonetheless, scattered and mixed species meadows, complex inherent optical properties of water and the unavailability of hyperspectral imagery for many locations remain significant obstacles for seagrass species detection using this technique.

In recent years, Landsat and Sentinel images have emerged as the most appropriate selection for seagrass mapping and monitoring at regional and worldwide scales. The spectral resolution, medium to high spatial resolution, and comprehensive global coverage enable efficient mapping and change detection of seagrass ecosystem for any sites of interest. Since the end of 2015, Sentinel-2 has officially provided qualified satellite imagery to end users free of charge. Sentinel-2 supports 10 m spatial resolution in the visible bands, and since 2017, the promising application of Sentinel-2 for seagrass mapping had been reported with data from the Mediterranean sea for *Posidonia oceanica* and *Cymodocea nodosa* species (Traganos and Reinartz, 2017), Indonesia for *Enhalus acoroides, Thalassia hemprichii, Cymodocea rotundata, Halophila ovalis* and *Halodule uninervis* (Fauzan et al., 2017), Australia for *Halophila ovalis, Halophila spinulosa, Halodule uninervis, Zostera muelleri, Cymodocea rotundata and Syringodium isoetifolium* (Kovacs et al., 2018).

During the course of this literature review, we noted little use of radar images, or synthetic aperture radar (SAR) specifically, for seagrass mapping and monitoring. SAR is available at medium spatial resolution (Ha et al., 2021), and is a type of active remote sensing, in which the sensor transmits a signal and records the back scattered signals from a combination of acquisition sequences. Hence, SAR is different from the passive optical sensors (*i.e* Landsat's Operational

Land Imager (OLI) or Sentinel-2's Multispectral Instrument (MSI)). It uses longer wavelength bands within the electromagnetic spectrum (termed X, C, L, P bands at the cm to m scale), which penetrate easier through the obstacles (cloud, for instance) (Elhefnawy & Sri Sumantyo, 2016). SAR imagery thus has several advantages over passive optical sensors in that it is possible to acquire images at night and through cloud cover. It also collects information at different polarizations (a combination of the horizontal (H) and the vertical (V) tranmission/ receiving), and polarization at a given band is sensitive to surfaces texture, and therefore adds a capability to collect spatial information on the interacting surface objects (*i.e* vegetation) (El Hajj et al., 2018; Blomberg et al., 2021). Sentinel-1, for example, provides freely available SAR data at C band, at up to 10 m resolution and the polarizations of VH and VV, might suitable for the monitoring of the low biomass vegetation like seagrass meadows. While there are examples of SAR use for more structured vegetation types, such as mangroves (Navarro et al., 2019; Pham et al., 2020) there is as yet no application of SAR imagery for seagrass mapping in various regions globally. While unsuited to characterize submerged seagrass meadows, it is potentially useful to use SAR data for mapping in the intertidal zones as it requires only image acquisition time to coincide with low tide, rahther than for overpass to coincide with low tide, suitable solar elevation and cloud free conditions that applies to optical sensors.

2.2.3. Image classification techniques for seagrass mapping

Classification technique is acknowledged as being the most important phase (Hossain et al., 2015) after image pre-processing, to successfully detect bottom covers (*e.g* sand, mud, seagrass, coral). Fundamentally, remote sensing classifications are divided into object-based and pixel-based techniques. The advantages, disadvantages and corresponding classifiers are summarized in Table 2.4.

Table 2.4. A simple comparison of remote sensing classification approaches

	Description	Advantages	Disadvantages	Classifier
Pixel based image classification	Spectral information from image pixel is the input to classifiers; Image pixel is assumed as pure information.	Fast and easy to implement; Suitable for medium to high spatial resolution imagery.	Can be unsuitable for high-resolution images due to spectral variation in each of the pixels; Contextual information is limited.	Unsupervised classifiers (k-means, ISODATA, SOM); Supervised classifiers (Maximum likelihood, Mahalanobis distance, machine leaning)
Sub-pixel-based image classification	Image pixel is assumed as mixed spectral information of different classes; A proportion is estimated for each sub-class in the pixel.	Suitable for medium to coarse spatial resolution imagery; Sub-classes can be accurately classified from a mixed information pixel.	Contextual information is limited.	Fuzzy classification; Regression modeling; Regression tree analysis; Spectral mixture analysis; Fuzzy spectral mixture analysis.
Image segmentation and object-based image analysis	Geographical objects are considered as the input to the classifier, not individual pixels; Multiple image pixels create an object.	More appropriate for very high spatial resolution imagery; Geographical objects may suggest a meaningful texture for the classification.	Generalizes the data; Difficult to compare image-objects in a time series data.	Image segmentation (region-growing, hierarchical algorithms); Object based image analysis techniques – OBIA – (Support vector machine, nearest neighbor classifier).

The object-based technique identifies similarity in shape and size of the objects and then combines them into different homogeneous groups. This process is known as segmentation classification and has been widely utilized for several decades (Urbański et al., 2009; Otukei & Blaschke, 2010; Roelfsema et al., 2014) prior to the introduction of geospatial object-based image analysis (GEOBIA). The pixel-based technique analyzes spectral information of each image pixel and then autonomously classifies them into model defined classes (the unsupervised method) or uses defined classes provided by the users to guide the classification model (the supervised method). The unsupervised technique has no requirement of ground truth data whilst supervised classification needs information from the field surveys as input for a training phase.

The most common unsupervised algorithms are K-means and ISODATA. Among supervised methods, Maximum Likelihood Classifier (MLC) is well-known and more widely applied than the K-Nearest Neighbor (KNN), Endmember method (Chen et al., 2016), Spectral Angle Mapper (SAM), Decision Tree (DT), Artificial Neural Network (ANN), and Mahalanobis distance methods (Kim et al., 2015). Recently, machine learning has attracted attention for supervised classification approaches due to its advantages in terms of edge discrimination between different classes (Wu et al., 2004; Nitze et al., 2012). For a wide range of remote sensing image classification, ML approaches include a group of bagging, boosting, and neural network-based techniques, which potentially produce higher classification accuracy (Mustapha et al., 2010; Colkesen & Kavzoglu, 2017; Ha et al., 2020). Recently, a suite of ML algorithms has been developed based on cutting-edge innovations from computer science. These new ML models work differently from the classical candidates, in that input data is not required to have a normal distribution, and the model is self-training through either a network of interpretation (neural network-based models) or a sequence of weak learner combinations (bagging/ boosting models)

(Thessen, 2016). These is a requirement of large dataset of input data for the training process, but ML is a practical approach where the user can collect large number of field data points and extract enough image pixels for the learning of ML models (Ha et al., 2020). In addition, the hyper-parameters (the parameters that define the structure of the model) can be optimized to make the ML model best fitting the local conditions and improve the accuracy (Ha et al., 2021). In this group of techniques, Support Vector Machine (SVM) and Random Forest (RF) have been introduced as effective methods for satellite imagery classification (Li et al., 2014; Ustuner, 2015). Numerous efforts have been made to compare the MLC and other methods for classification of land cover (Li et al., 2014; Ustuner, 2015), land cover change and crop type (Otukei & Blaschke, 2010). Most cases present an acceptable accuracy using the MLC technique. MLC accuracy, however, tends to be lower than machine learning or DT methods for the same classifications (Otukei & Blaschke, 2010; Li et al., 2014; Ustuner, 2015). To our knowledge, a similar comparison, including machine learning based classification, has not been reported for seagrass mapping (Dat Pham et al., 2019).

In other research, seagrass distribution has been converted to a biological index as a further interpretation for the post-classification process. Instead of discrete data, a continuous surface of seagrass abundance provides a more nuanced map of biological characteristics. These parameters are transferred to a range of 0 - 1 through a fuzzy membership procedure where 1 presents the highest index (Pu & Bell, 2013). The fuzzified map provides high quality input for deeper analysis and generation of a more informative map than a present/absent classification.

The techniques discussed above are all empirical approaches using ground truth data to interpret class features. More advanced classification techniques include the semi-analytical and semi-empirical procedure methods. To apply these models, a spectral library of bottom substratum

must be prepared prior to transfer to the programmed application. In the next step, a look-up table is used to match spectral data of the substratum with the library's references. This semi-analytical method potentially allows a higher accuracy than the empirical approach (Dekker et al., 2006; Wettle, 2006; Roelfsema et al., 2014; Hedley et al., 2017)). However, it also requires an intensive and inclusive spectral library of different bottom covers as input for the classification algorithm. This may lead to expensive field sampling and maintenance of the library in the case of large-area site monitoring.

It is evident that the standard techniques described above are not ideally and globally suited for an accurate mapping of (patchy) seagrass meadows. The diversity of water column constitution, seagrass species characteristics such as patchy growth and high-density variability, limitations of existing remote sensing sensors make this task more challenge and therefore, require an innovative approach to retrieve the highest accuracy mapping of patchy extant seagrass meadows at an acceptable cost.

2.3. Remote sensing for monitoring seagrass and detection of change

Change detection is a key process for understanding the dynamics of seagrass ecosystems in relationship to external stressors. Long-term monitoring from remote sensing is capable of describing the spatial variation of seagrass area and, with suitable environmental data collection, may allow the relationship between driving factors and seagrass degradation to be inferred. The remote sensing approach provides solid mapping and change detection compared to classical field-based monitoring due to the capacity of continuous, large scale mapping and spatial analysis at both meadow and regional scales (Ferwerda et al., 2007). In addition, several external stressors have been retrieved from satellite data that provide a direct connection to the variation of seagrass biomass and distribution. Among them, chlorophyll-a concentration, CDOM,

turbidity, total suspended sediment, sea surface salinity and land cover change are the most frequent stressors extracted from satellite imagery (Gholizadeh et al., 2016). The following review aims to describe work on the selection of satellite sensors and techniques for seagrass change detection.

2.3.1. Satellite sensors for seagrass change detection

To successfully detect spatial and temporal variation of seagrass meadows, sensors should have a long-time operation with qualified standards for storage, image pre-processing and processing as well as ease of data access. Currently, Landsat is the most popular satellite imagery for change detection not only for land use cover but also for aquatic resources (Zhu, 2017). The first generation of Landsat operated from 1973 and until 2013, Landsat-8 is the current sensor in the orbit. Notably, the scan line corrector (SLC) of Landsat 7 was interrupted on 31 May 2003 and released a zig-zag pattern or gaps on the image that leads to 22% loss of retrieved information (USGS, 2018). Users have to accept these gaps despite the development of gap filling algorithms. Other sensors are, however, available and are capable of providing an alternative source to maintain long-term monitoring. Nevertheless, the diversity of spatial, spectral and temporal resolution must be considered prior to any change detection procedures applied to mixed satellite data.

For long-term monitoring, the long-life operated Landsat family satellite imagery is a valuable dataset for seagrass dynamics assessment for eco-regional scale mapping (Wabnitz et al., 2008; Torres-Pulliza et al., 2013). Temporal change detection varies in durations in research papers across a range of 14 - 38 years (Dat Pham et al., 2019). Examples include the monitoring of the genera *Zostera*, *Posidonia*, *Ruppia* and *Halophila* in Wallis Lake, Australia using Landsat-5 TM and Landsat-7 ETM+ (Dekker et al., 2003), Moreton Bay, Australia with Landsat MSS, Landsat-

5 TM, Landsat-7 ETM+ from 1972 - 2010, Eastern Puerto Rico using Landsat-5 TM and Landsat-7 ETM+ in 1985 and 2000 (Shapiro & Rohmann, 2006). Chen et al. (2016) reported the mapping and change detection of seagrass (for a mixed group of *Enhalus acoroides, Halophila ovalis, Halophila minor, Thalassia hemprichii, Halodule pinifolia, Halodule uninervis*) for the period of 1996 - 2015 in Cam Ranh Bay (Viet Nam) from Landsat TM/ETM+/OLI (Chen et al., 2016). Landsat-5 TM and Landsat-8 OLI described the change of seagrass meadows between 2009 and 2013 (for a mixed group of *Halophila ovalis; Thlassia hemprichii; Enhalus acoroides; Halophila pinifolia, Halodule uninervis, Halophila decipiens*, *Cymodocea serrulata*, *Cymodocea rotundata*, *Halophila spinulosa*) in Malaysia (Misbari & Hashim, 2016). Applying Landsat TM and Landsat-7 ETM+ in Tanzania, the peak points of seagrass change (*Thalassia hemprichii*, *Enhalus acoroides*, *Halimeda* spp) were identified during the period 1986 - 2003 (Gullström et al., 2006). In another study, the temporal change of seagrass from 1991 - 2006 was detected through bio-optical models in Xincun Bay (Hainan, China) with integration of *in situ* hyperspectral observation and a range of satellite images, involving Quickbird, China-Brazil Earth Resources Satellite (CBERS) data and Landsat (Yang & Yang, 2009). The longest interval for seagrass change detection was a period of 40 years (from 1972 - 2013) in Spermonde Archipelago, Indonesia using Landsat-1 to Landsat-8 (Nurdin et al., 2014). Seagrass damage caused by three typhoons was also recorded within an interval of 24 years from Landsat and other sensors in Korea. In this case, remote sensing data successfully detected the great loss and enabled a more comprehensive understanding of the past change that may link to current dynamics of seagrass (Kim et al., 2015).

For short-term monitoring, commercial satellite imagery and small-scale mapping usually come together. Mediterranean seagrass (involving *Cymodocea nodosa* and *Posidonia oceania*) area

variation was assessed for 5 years from RapidEye time series data (Traganos & Reinartz, 2017). Quickbird and acoustic field survey data were also used to map seagrass change in 2004 and 2007 (Lyons et al., 2011). In a similar study, Quickbird-2, Ikonos and Worldview-2 allowed seagrass dynamics (*Cymodocea serrulata, Halophila ovalis, Halophila spinulosa, Syringodium isoetifolium, Halodule uninervis, Zostera muelleri*) to be followed from 2004 - 2013 in the Eastern Banks, Moreton Bay, Australia (Roelfsema et al., 2014) and from 10/1947 - 2/2003 in the Solomon Islands (Lauer & Aswani, 2010). Landsat TM, ALOS AVNIR-2 and THEOS images were another combination to trace the dynamics of marine habitat (seagrass involved) in Phu Quoc island (Nguyen et al., 2014). This combination, however, created challenges for the post-classification phase due to the differences in image processing of the sensors and it proved difficult to produce a continuous and precise comparison from classified maps during the years of the study.

2.3.2. Techniques for detecting change using remotely sensed data

Recently, different change detection approaches have been developed in several research fields. These fields range from land use cover change (Deilmai et al., 2014; Gómez et al., 2016; Karan & Samadder, 2016) to aquatic resources and more specially seagrass dynamics, involving spatial distribution, coverage, and aboveground biomass (Dekker et al., 2005; Shapiro & Rohmann, 2006; Urbański et al., 2009; Lyons et al., 2011; Misbari & Hashim, 2016; Deyanova et al., 2017; Zhu, 2017; Traganos & Reinartz, 2018). This has improved change detection algorithms, and accuracy of post-classification assessment and has resulted in a better understanding of the comprehensive relationships within aquatic ecosystems. As described below, several approaches have been found for change detection in remote sensing, however, no single technique has

emerged as optimal across diverse marine environments (Hossain et al., 2015; Dat Pham et al., 2019).

The change detection technique has a list of key requirements, involving sensor selection, image processing, image classification, post-classification, and detection of change in order to precisely detect the variation among classified maps. Importantly, change detection requires a consistency of sensor, date of acquisition, and geographical boundary. Unless these conditions can be reliably and repeatably satisfied, change detection remains questionable for accuracy (Liu & Zhou, 2004).

Tewkesbury et al. (2015) undertook an analysis of various techniques of change detection (Tewkesbury et al., 2015), using six approaches, involving layer arithmetic, post-classification, direct classification, transformation, change vector analysis, and hybrid. The advantages and disadvantages of these approaches are briefly summarized in Table 2.5.

Table 2.5. A simple comparison of change detection approaches (based on Tewkesbury et al. (2015))

	Description	Advantages	Disadvantages
Layer arithmetic	Satellite images (at radiance level) are numerically compared to determine the change	Easy to conduct	Difficult to illustrate types of change
Post-classification change	Classified images are compared to identify the change	Create change map with labels; Does not always require radiometric correction	Low accuracy of classified map will impact on the accuracy of change detection
Direct classification	A multi-layer image of different time points is classified and the change is directly	One-off classification stage with labeled classes; Can be suitable for	Difficult to apply training data for each layer, especially for a time series image.

	identified from these layers	time series analyses.	
Transformation	Apply mathematical transformation to reveal the variance among the images	Can be suitable for high dimensional data	Difficult to interpret transformed images to thematic change
Change vector analysis	The analysis of different vectors gives both magnitude and direction of the change	Discover different types of change	Difficult to interpret the magnitude and direction from the raw form of vector analysis
Hybrid change detection	A combination of different change detection techniques	Not necessary to collect training data at less changeable areas	No specific limitation

Various change detection techniques have been used for seagrass dynamics assessment. However, the performance metrics of the classification are frequently reported rather than change detection techniques in the published research papers. An arithmetic approach enables presentation of a spatial change in area and distribution of seagrass corresponding to the years with classification accuracy ranged from 63% to 80% (Yang & Yang, 2009; Lauer & Aswani, 2010; Misbari & Hashim, 2016). Other research papers involve image pixel analysis in various years by clustering pixels into classes (Dekker et al., 2003; Shapiro & Rohmann, 2006; Nobi & Thangaradjou, 2012; Kim et al., 2015; Chen et al., 2016) and then applying a visual interpretation or arithmetic technique (mapping accuracy ranged between 67% and 91%). These steps aim to reduce the impact of atmospheric, water column and acquisition date on the process of image change detection. Two time points, namely initial and final phase, are defined by the users and act as the inputs for change detection. Then, a differential map is created and indicates a positive, negative or no change of classified objects. Object based image analysis (OBIA) with geographic rules, for example in terms of defining seagrass patches, combined with field survey,

has also been used to trace seagrass dynamics from a very high resolution (VHR) image (Nguyen et al., 2014; Roelfsema et al., 2014). These authors argue for a higher accuracy of an OBIA classified map (accuracy 68% - 83%) compared to the pixel-based approach (35%). Additionally, contextual editing is another option to better adapt for change detection in a geographical information system (GIS) environment. The classified pixels are converted into shapefile format and then analyzed with simple or advanced spatial analysis technique (Wabnitz et al., 2008; Carter et al., 2011; Nguyen et al., 2014). The overall accuracy, however is widely varied (46% - 91%) among research papers, indicating the challenges for seagrass habitat mapping in different conditions. In the case of monthly or annual detection, time series analysis is a feasible selection to identify the trend and discover the peak point of this variation. This approach, however, is sensor and climate dependent, and as a result, few studies have been conducted during recent years (Lyons et al., 2013; Traganos & Reinartz, 2018). Fundamentally, this is a pixel-based technique where pixel value presents the area, coverage or aboveground biomass of seagrass. Then, these values are spatially extracted and transferred into a time series chart to denote the variation of mentioned parameters. Further statistical analysis from classified maps with sufficient accuracy (85% - 100% as reported by Traganos & Reinartz. (2017) contributes to the certainty in change detection as well as uncovering driving factors in the study site.

The hybrid option is applied when separate techniques fail to satisfy the requirement of change detection. Similarly, a combination of pixel-based and object-based classification techniques increased the overall accuracy from 75% to 88% for land cover change in Mexico (Aguirre-Gutiérrez et al., 2012) and Donaxi et al. (2012) developed an urban change detection framework using object-based classification, data filtering, and multivariate alteration detection (MAD) transformation. The latter paper reported approximately 100% correctness of change detection

(Doxani et al., 2012). Bruzzone and Bovolo (2013) described a change detection framework for very high resolution multi-spectral imagery (Bruzzone & Bovolo, 2013) which was based on a taxonomy of radiometric change and context-sensitive decision approach. The results showed an improvement of the proposed framework compared to pixel-based and parcel-based change detection. The overall accuracy increased to 93.91% and was higher than pixel-based (90.86%) and parcel-based (91.56%) approaches. In addition, a time series analysis may contribute informative data to the last step of change detection to depict a clearer picture of the dynamics (Lyons et al., 2013).

Despite the improvement of sensor resolution, image processing, and classification frameworks, the field survey still plays an essential role for both seagrass mapping and change detection (Hossain et al., 2015; Ferwerda et al., 2007). Ground truth data provides not only seagrass distribution but other ecological characteristics and enables an effective empirical modeling for mapping and post-classification assessment. Furthermore, a uniform approach for image selection, image pre-processing and classification will ensure the accuracy of output maps which is essential for change detection comparison.

In summary, the Landsat family and Sentinel-2 are the most appropriate satellite images for long-term and large-scale monitoring. These sensors are promising due to their stability, and long-life operation, as well as their similarity in spectral resolution. However, Landsat-7 suffered a scan line corrector hardware failure (SLC-off) since 2003 and this leads to a significant loss of pixel information. As a result, an interval of 10 years monitoring should be investigated to fill this gap, otherwise, a gap-filling procedure must be considered. In a range of seagrass ecosystem, Landsat-8 (launched in 2013) and Sentinel-2 (launched in 2015) imagery, which are both high in temporal and spatial resolutions, together with ML models, are expected as the next rationale

approaches for mapping and change detection of seagrass dynamics. The classification accuracy for published remote sensing outputs on seagrass cover (which are not always reported) varies by study site, seagrass cover, satellite imagery selected, and classification models with the ranges of 70% - 96% and 35% - 95% for long-term and short-term change assessment, respectively. This situation motivates the use of long-term operated satellite sensors to reduce the uncertainty in image processing, and investigation of novel classification approaches to improve the certainty in detection maps.

2.4. Blue carbon mapping of seagrass ecosystems

2.4.1. Blue carbon – A novel initiation for seagrass ecosystem conservation

Anthropogenic carbon dioxide (CO_2) emission is one of the most serious environmental problems in the early 21st century; this leads to greenhouse gas increase and indirectly results in global warming. Approximately one-third of anthropogenic carbon dioxide is currently absorbed by the oceans which reinforces the importance of the ocean ecosystem for carbon sequestration at a global scale (Rhein et al., 2013). Notably, mangrove, salt marsh and seagrass meadows contribute a 50-fold (Mcleod et al., 2011) higher carbon sequestration rate (g C m^{-2} yr^{-1}) than other terrestrial ecosystems, including tropical or temperate forests (Pendleton et al., 2012; Duarte et al., 2013). As an autotrophic ecosystem, one hectare of seagrass sequestrates more than 600 Mg CO_2 with a large proportion stored in the top 0.5 - 1 m of the soil layer (Fourqurean et al., 2012). Despite representing a small area (approximately 0.1% of ocean surface), roughly 20% of total buried carbon has been stored by seagrass (Duarte et al., 2013). Conversely, the loss of this ecosystem may lead to the emission of a huge amount of CO_2 to the atmosphere. Under the pressure of coastal ecosystem degradation and CO_2 emission, the blue carbon initiative was

born to fortify the conservation of coastal resources and contribute to climate change mitigation (Duarte et al., 2013).

Blue carbon includes autochthonous and allochthonous carbon sequestered by an ecosystem (Mitra & Zaman, 2015). Autochthonous carbon is a product of *in situ* photosynthesis and is stored in the plant tissue and in organisms using those tissues, whilst allochthonous carbon originates from external sources. The downstream transfer of organic matter from terrestrial ecosystems significantly contributes to the allochthonous carbon in coastal areas. Blue carbon of mangroves, salt marshes, and seagrass meadows is calculated as total stored carbon in the soil, aboveground biomass (leaves, stem, and branch), below-ground biomass (root, rhizome), and non-living biomass (litter, dead wood). These ecosystems are acknowledged as high primary productivity ecosystems (Bouillon et al., 2007; Howard et al., 2014; Alongi et al., 2016) and maintain a considerable burial rate of organic carbon (Watanabe & Kuwae, 2015). Burial rate, which is a measurement of the amount of sequestered carbon in a specific time period, is a critical metric of mitigation of climate change. This number is well estimated for mangrove forest and salt marsh (Schile et al., 2017) but not well replicated for the seagrass ecosystem.

Seagrass species, seagrass meadow structure and the source of carbon are the key factors that impact on carbon stocks and sequestration. At a meadow scale, carbon sequestration of seagrass varies due to the diversity of seagrass species, area and age of the meadow, and geographical features of the coastal zones (Gazeau et al., 2005; Bouillon et al., 2007; Duarte et al., 2013; Tokoro et al., 2014; Zarate-Barrera & Maldonado, 2015; Deyanova et al., 2017; Schile et al., 2017). Oreska et al. (2017) found that stored carbon is proportional to sediment grain size and edge proximity at a meadow scale. At a plot scale, shoot density and seagrass age significantly correlated to carbon stocking. In other words, a dense, continuous and long-lived meadow tends

to accumulate more carbon than patchy, younger and smaller ones (Oreska et al., 2017). Notably, the loss of seagrass aboveground biomass may lead to a rapid reduction of stocked carbon in seagrass meadows (Duarte et al., 2013; Macreadie et al., 2014; Campbell et al., 2015; Alongi et al., 2016; Oreska et al., 2017), result in a negative carbon balance, raises the demand for accurate mapping of seagrass aboveground biomass changes (Deyanova et al., 2017). Restored seagrass meadows are also proven to have high potency for carbon sequestration (Duarte et al., 2013). The latter results demonstrated that an expensive investment in seagrass restoration can be effectively recouped through the carbon tax offset for CO_2 capturing by recovered seagrass canopies.

In addition, sequestered carbon of seagrass meadow also receives contributions from various sources, involving atmosphere-ocean CO_2 exchange, allochthonous carbon, sediment, and the water column. Whilst the water column contributes an effective sequestration in the open ocean, burial of carbon in sediment is determined as the most important mechanism for shallow ecosystems (Tokoro et al., 2014). In this context, a meadow structure of seagrass supports a large volume of CO_2 exchange by suspended sediment trapping and prevention of resuspension. Low oxygen concentration and decomposition rate in sediments of seagrass meadow preserves inactive carbon for a long period (Iacono et al., 2008). The top 10 cm of the soil layer tends to store the highest carbon concentration with large contributions from seagrass roots and rhizomes.

Despite the importance of the blue carbon initiative (Fourqurean et al., 2012), its study still remains at the development phase with several knowledge gaps including: spatial distribution, and carbon sequestration in different ecosystems as well as external driving factors (Howard et al., 2014). In the case of seagrass ecosystems, there is a high demand for more blue carbon assessment due to known sensitivity to both internal and external stressors. In 2015, a large

survey was conducted in Indonesia and resulted in blue carbon estimation of the species *Enhalus acoroides, Thalassia hemprichii, Halodule uninervis, Cymodocea rotundata* for the entire country (Alongi et al., 2016). On average, the total blue carbon was calculated as 119 Mg C ha^{-1}. Below-ground biomass carbon was significantly greater than aboveground biomass carbon. Duarte et al. (2013) summarized a short list of carbon storage assessment research in the Mediterranean and Australia (Duarte et al., 2013). Generally, the carbon burial rate of species of the genus *Zostera* (*Zostera marina, Zostera noltii*) is not as high as the rate for *Posidonia oceania* and *Cymodocea nodosa*. Organic carbon content in seagrass sediment was also measured for *Enhalus acoroides, Thalassodendron ciliatum, Thalassia hemprichii* and *Cymodeocea* spp. in the coastal zones of Tanzania and Mozambique (Gullström et al., 2006). On average, the content ranged from 213.4 - 730.5 Mg C ha^{-1} and was much higher than unvegetated areas in similar study sites. Among seagrass species, sediment associated with the very large seagrass *Enhalus acoroides* had the highest content of stored organic carbon (700 Mg C ha^{-1}). Sediment organic carbon negatively correlated with sediment density but positively with belowground and aboveground seagrass biomass. In Kenya, seagrass blue carbon was estimated in Gazi Bay for *Thalassodendron ciliatum, Thalassia hemprichii, Enhalus acoroides, Syringodium isoetifolium* seagrass species (Githaiga et al., 2017). There was no overall correlation found between sediment organic carbon and aboveground organic carbon. Despite this, the relationship was species dependent. Currently, roughly 67% of seagrass' carbon sequestration studies focus on tropical and subtropical regions, with a particular shortage of information in the Southern hemisphere, and North and West Pacific coastal areas (Duarte et al., 2013).

2.4.2. Blue carbon mapping from remote sensing

Blue carbon mapping from satellite data is an attractive topic due to the advantages of low cost, large-scale mapping and reliable monitoring. Recently, a majority of remote sensing work has concentrated on mangrove and salt marsh ecosystems by utilizing optical or SAR data (Avitabile et al., 2011; Wijaya et al., 2013; Candra et al., 2016), drone mapping (AGEDI, 2015), ShoreZone database (Cook, 2013). For seagrass ecosystems, blue carbon mapping using remote sensing is less extensive. Mapping can be effective where sufficient data on soil, belowground and aboveground stocking carbon is available, but this is not a simple task, since these data are not always available for several regions. Additionally, current remote sensing sensors are not able to detect below the sediment surface. For this reason, the task devolves to seagrass aboveground biomass mapping and development of an empirical relationship with total stored carbon (Roelfsema et al., 2014; Koedsin et al., 2016).

Aboveground biomass detection has achieved some success. For example, spectroscopic aerial mapping system with on-board navigation (SAMSON) – a hyperspectral satellite imagery – was applied to retrieve a seagrass aboveground biomass through leaf area index for *Thalassia testudinum, Halodule wrightii, Syringodium filiforme, Ruppia maritima, and Halophila engelmannii* (Hill et al., 2014). Kim et al. (2016) developed a significant correlation between aboveground biomass of mudflat rush (*Bolboschoenus planiculmis*) and the combination of spectral band 3, 4 and 7 of Landsat-8 OLI (Kim et al., 2016); this model presented a good correlation between observed and predicted aboveground biomass with a Pearson correlation coefficient of 0.84. Similarly, Worldview-2, Ikonos and Quickbird-2 were combined with object-based classification to retrieve aboveground biomass mapping of seagrass in the Eastern Banks and Moreton Bay (Australia) (Roelfsema et al., 2014). Biomass was modeled as a function of

percentage cover and confirmed by ground truth data. The determination coefficient (R^2) of AGB regressions ranged from 0.26 to 0.82 for different seagrass species. High spatial resolution of satellite images provides more detail of seagrass species distribution, meadow texture and therefore, releases a higher accuracy mapping when applying an object-based approach. In Thailand, Worldview-2 imagery was utilized for seagrass mapping, retrieving an estimation of spatial distribution, species, coverage, and aboveground biomass. Linear relationships between coverage (percentage cover) and aboveground biomass were developed for *Enhalus acoroides* (R^2 = 0.94), *Halophila ovalis* (R^2 = 0.8), and *Thalassia hemprichii* (R^2 = 0.67). The biomass map was modeled directly from reflectance bands of Worldview-2 with a strong correlation range from 0.68 (for all seagrass species) to 0.99 (for *Enhalus acroides* seagrass species) (Koedsin et al., 2016). Pu et al. (2017) conducted an ISODATA (unsupervised classification technique) clustering prior to combining truth data and supervised classification techniques (MLC and SVM) to retrieve the map of submerged aquatic vegetation (SAV) coverage (Pu & Bell, 2017). The continuous, patchy, and no SAV classes were mapped from Landsat TM and Ikonos with a higher accuracy of Ikonos mapping. Notably, Landsat TM presented a clear discrimination of large SAV (continuous or patchy meadows). The overall accuracy and Kappa coefficient of SAV mapping from Ikonos is not much higher than Landsat TM.

There are, however, few attempts to connect AGB to blue carbon mapping by developing correlations with below-ground carbon. Among those that do exist, the database of shore zone imaging and mapping (http://www.shorezone.org) has been used for carbon stock assessment (Short et al., 2014). These authors also argued a serious data gap of seagrass blue carbon habitat maps as well as their spatial extent and variation. Their approach is very promising; however, it requires a high density of ground truth data as well as a synchronized feature of GIS system for

visualization and decision making. In Malaysia, Landsat image was processed in 2004, 2009 and 2013 to review the dynamics of aboveground biomass as well as the aboveground blue carbon stock of seagrass (Misbari & Hashim, 2016). Seagrass aboveground biomass declined during the period of 2004 - 2013 and reduced the carbon stock to 18.77 kg C $pixel^{-1}$ (2013) from 90.68 kg C $pixel^{-1}$ (2004). The results presented a high potential for mapping for carbon sequestration dynamics from the medium of remote sensing, particularly if the link to below-ground biomass can be resolved.

In summary, blue carbon mapping from remote sensing is promising. Nevertheless, in addition to all the challenges previously discussed related to estimation from remotely sensed data, the lack of *in situ* carbon stock data, the diversity of optical water types, high labor and laboratory cost, as well as the temporal changes of seagrass aboveground biomass remain the obstacles for accurate mapping and monitoring of seagrass AGB and aboveground blue carbon. This thesis proposes the use of free satellite data which helps to reduce the research budget, novel ML models which might benefit a higher accuracy of AGB retrieval, and advanced feature selection techniques which decrease the number of input variables as well as the requisite computation power. The successful estimation of seagrass aboveground biomass will provide a baseline for further assessment of aboveground stocking of blue carbon by seagrass globally, and ultimately will include total areal blue carbon.

2.5. Dynamic web-based application for seagrass data visualization and collaboration

In recent years, the rapid development of information technology has strengthened a wider application of web-GIS platforms in natural resources mapping and management, spatial planning, and social interaction (Veenendaal et al., 2017). A web-GIS based management system confers several benefits: (i) a near real time and clear visualization of geospatial database and

thematic maps; (ii) an instant access to geospatial database and data collaboration; (iii) provision of a platform for social interaction and collaboration which stimulates the sharing of GIS data among the stakeholders and improves community awareness (Fu, 2018). For this reason, the researchers have paid close attention to the application of web-GIS, not only to effectively visualize the geospatial database but to connect people for resource conservation through social interaction (Andris, 2016; Sui & Goodchild, 2011).

A geospatial database is a group of remote sensing, global positioning system (GPS) and geographical information system (GIS) data, inter-connected and managed using a database management system. These datasets are managed and visualized within a GIS server and can be used to visualize specific maps at various scales and accessed from anywhere. In the case of an open system, the user is allowed to contribute their formatted data to the database. The use of geospatial data enables an overview of classified habitats regarding geographic features, social and environmental factors. In addition, time series analysis from a spatial decision support system will suggest driving parameters which are behind the changes in the habitats (Karnatak et al., 2007). As a function of the database system, the relational feature ensures a connection between separate datasets and allows a huge number of queries from remote users. This kind of system is known as a relational database management system (RDBMS). PostgreSQL (https://www.postgresql.org/) + PostGIS (https://postgis.net/), ArcGIS Enterprise (https://enterprise.arcgis.com/en/), Oracle Spatial (http://www.oracle.com/technetwork/database/options/spatialandgraph/overview/index.html), MySQL (https://www.mysql.com/), TerraLib (http://www.dpi.inpe.br/terralib5/codedocs_5.0.1/) or SpatiaLite (https://www.gaia-gis.it/fossil/libspatialite/index) are popular foundations for database building and management. In order to render a specific database into accepted formats

as web pages, a GIS map server is required to connect, overlap and analyze data prior to presenting the results to the end users. Recently, GeoServer, UMN Mapserver, ArcGIS Server, Mapguide, Degree, ERDAS APOLLO Server, Intergraph GeoWeb Server, QGIS server are feasible platforms to build up a web-based system. Both database management and the web server system are released under a commercial or open source license that diversifies the options and satisfies different demands across disciplines. In order to reduce the cost and take advantage of open sources technologies, the following web-based mapping applications are listed as the priority (Table 2.6).

Table 2.6. Open sources web-based mapping application

	Name	Web address
1	MapGuide project	https://mapguide.osgeo.org/
2	GeoTools	http://www.geotools.org/
3	GLG Map Server	http://www.genlogic.com/free_map_server.html
4	Map Server	http://mapserver.org/
5	Google Earth and Google My Map	https://www.google.com/earth/
6	Degree	https://www.deegree.org/
7	QGIS Server	http://qgis.org
8	GeoServer	http://geoserver.org/
9	Lizmap	https://www.lizmap.com/en/

The application should satisfy crucial requirements of security, database connection, effective management, user-friendliness multi-platform installation, fast loading and international web browser capability.

In terms of blue carbon, this ideal has been successfully conducted in the USA with ShoreZone imaging and mapping (http://www.shorezone.org) (Cook, 2013), internationally within Mapping

Ocean Wealth (http://maps.oceanwealth.org/#) and the Blue Carbon Mapping Tool (http://bluecarbontoolkit.ae/en/layout) in UAE. As web-based systems, these applications allow end users a remote access for database visualization and query. Locally, the ShoreZone and UAE mapping tools include coastal vegetation distribution and blue carbon in their database whilst the Mapping Ocean Wealth concentrates on coral reefs and their economic value globally. Currently, there is not an overview of system structure and operation due to the unavailability of the source codes and technical documents. In detail, their core functions allow a spatial query at a specific area (UAE blue carbon mapping tool), visualization of coral reef values (mapping ocean wealth) and more advanced function for mapping with ShoreZone.

For a more general purpose, there are several web-GIS services available for natural resource or soil carbon management. The United States fish and wildlife service has released a version of wetlands mapper (https://www.fws.gov/wetlands/) using a geodatabase in the ArcSDE environment (Stout et al., 2007). This web-based inventory allows remote access for wetland types and surface water habitat query (involving estuarine, marine deep water, freshwater wetland, freshwater pond, lake and riverine), measurement of the area as well as the facility to export a wetland map in PDF format. In addition, a huge database permits the users to download wetland distribution and metadata datasets by states or watershed. Historical distribution is another option, however, it is not available for all regions. More broadly, the Global Earth Observation System (GEOSS) portal (http://www.geoportal.org/) and Ocean Data Viewer (http://data.unep-wcmc.org) provide visualization of a biodiversity dataset globally. Ocean Data Viewer focuses on the global distribution of marine ecosystems, biodiversity index (including risk index) as well as providing available data for downloading. On a more ambitious scale, the GEOSS portal aims to collect all database from around the world, involving metadata and

geospatial information. The user makes a query with the keyword; however, it is time-consuming to filter the results from a big database. On the other hand, several records exist as raw information and should be updated in the next few years. The SeagrassSpotter (https://seagrassspotter.org/) supports a simple web-GIS system for seagrass image visualization globally, and a morphology-based classification for seagrass species involved.

Soil carbon database is another collaborative endeavor between USGS Powell center and the Max Planck institute for biogeochemistry (https://powellcenter-soilcarbon.github.io/soilcarbon/). This work aims to create a database of soil carbon that allows users to view, make plots and contribute formatted data. The project's data has been uploaded to GitHub, however, most information is not distributed widely at this time. Supporting $REDD_+$ activities, Forest Carbon Database (http://carbonstock.cifor.org/user/HomeMap) has released an open access web-based mapping application (Kurnianto & Murdiyarso, 2010). The application supports basic functions of visualization in map and table view. The user is motivated to contribute carbon stocking data for any location. Nevertheless, current data is only available in Indonesia within limited data points. In another promising approach, Global Risk Assessment Services (GRAS) (http://gst-prod.gras-system.org/webui/index.html#/worldmap/show) provides the database of total biomass carbon, total organic carbon, peatland, and aboveground biomass of woody vegetation in 28 members of the EU and 13 other countries worldwide which covers carbon stocking, biodiversity, land use change, social index, and fires. The spatial maps allow the user to view one layer of specific data in listed countries. However, neither blue carbon nor coastal vegetation cover is involved in the GRAS database.

Undoubtedly, these approaches take advantage of collaboration, peer exchange, monitoring cost reduction and unification of data formats. However, none of them concurrently (1) provides a

database of seagrass dynamics; (2) entirely uses open source applications for web-GIS building; (3) is easy for data visualization and extraction, simple in configuration and deployment, and social-media-emphasized with features of data sharing and collaboration; and (4) has integrated seagrass blue carbon into the database. The development of a web-GIS toolset, which is dedicated to the seagrass ecosystem, is very necessary to not only effectively protect extant seagrass but also improve public awareness of the need for seagrass conservation, fosters a win-win mechanism in the future.

The review has revealed that web-based applications at various scales, can be used not only for terrestrial ecosystems but present high potential for the visualization of coastal resource change. There are, however, only a modest number of dedicated database management and web-GIS systems for seagrass in specific regions. In New Zealand, a full feature and open source web-GIS system is not available for seagrass at the moment, which is a strong motivation to build a simple, low cost but solid web-GIS system for the visualization and collaborative building of a seagrass dynamics database.

2.6. Knowledge gap and research question

The literature review has provided an in-depth overview to identify the knowledge gaps and research questions for the mapping of temporal spatial distribution of seagrass' biophysical parameters and web-GIS based applications (Table 2.7).

Table 2.7. Identified knowledge gap and proposed research question

Knowledge gap	Research question
Seagrass mapping	
1. Large-scale and accurate mapping of *Z. muelleri* has not been undertaken in New Zealand. 2. Limited application of ML algorithms for (sparse) seagrass maping. 3. The potential of mapping using the SAR image has not been examined for the seagrass ecosystem.	1. Is Sentinel - 2 imagery suitable for large-scale seagrass mapping in New Zealand? 2. Traditional classification approach (i.e: Maximum likelihood) and ML algorithms - which one results in more accurate mapping for seagrass meadows? 3. Can we use SAR images for seagrass mapping in the intertidal zones?
Seagrass change detection	
1. Large scale and long-term change detection for seagrass (*Zostera muelleri*) dynamics using space borne remotely sensed data have not been undertaken in New Zealand; 2. Long-term monitoring from an historical remote sensing data framework has not been proposed for investigating *Zostera muelleri* dynamics in New Zealand	1. Is it applicable and reliable to detect the large-scale and long-term change of the seagrass *Zostera muelleri* in New Zealand from Landsat image? 2. Can we use ML models to map seagrass changes, and which one will be the best model integrating to Landsat imagery? 3. Is Landsat imagery sufficient to plan a long-term monitoring of *Zostera muelleri* in New Zealand?
Seagrass AGB mapping	
1. Accuracy of seagrass AGB estimation from space-born satellite image is very low for the species *Z. muelleri*; 2. ML application has not been tested for seagrass AGB mapping from space-borne satellite images.	1. Is it reasonable and reliable to either apply multi-spectral/SAR data or using both multi-spectral and SAR data for *Z. muelleri* AGB estimation? 2. With what accuracy do the ML models derive *Z. muelleri* AGB estimation?
Web-GIS based application	
1. A geospatial-database of seagrass dynamics has not been built for the species *Z. muelleri* in New Zealand; 2. An open-source, light-weight, and simple web-GIS interface for seagrass dynamics is not available in New Zealand.	1. Is it solid and rational to integrate open source GIS applications for building a simple but light-weight, elegant web-GIS interface for database visualization and collaboration in the study site?

The thesis will involve 4 research chapters which aim to fill the gaps in the literature and answer the research questions in Table 2.7. Chapter 3 describes the application of various novel ML models, in comparison to the traditional approach, to mapping the spatial distribution of dense and sparse seagrass meadows from multi-spectral Sentinel-2 imagery. Chapter 4 describes the testing of a combination of ML models and Landsat imagery for a large-scale and temporal detection of seagrass change which is expected to improve classification accuracy, and confidence in the long-term assessment of the seagrass ecosystem. In chapter 5, the potential application of both Sentinel-1 and Sentinel-2 images is examined for seagrass spatial distribution and AGB mapping using the novel ML models and metaheuristic optimization for feature selection. Sentinel-1 was initially selected for the mapping of seagrass meadows whilst various combinations using Sentinel-1 and Sentinel-2 bands were examined for seagrass AGB mapping. While chapter 3 and 4 contribute advanced methods for the mapping of spatial and temporal distribution, chapter 5 attempts to test the potency of both novel remotely sensed data (Sentinel-1 image which has the advantages of being cloud and weather impactless) and retrieval models (ML models and metaheuristic optimization) for assessing seagrass distribution and AGB. Chapter 6 builds an experimental web-GIS interface to visualize the geodatabase of seagrass dynamics and provides a free access hub to the relevant stakeholders using open-source GIS application.

Chapter 3 A comparative assessment of ensemble-based machine learning and Maximum Likelihood methods for mapping seagrass using Sentinel-2 imagery in Tauranga Harbor, New Zealand

The contents of this chapter have been published in the Remote Sensing Journal (https://www.mdpi.com/2072-4292/12/3/355/htm). A copy of the published paper is bound into Appendix 2.

Abstract: Seagrass has been acknowledged as a productive blue carbon ecosystem in significant decline across much of the world. A first step towards conservation is mapping and monitoring of extant seagrass meadows. Several methods are currently in use but mapping the resource from satellite images using machine learning is not widely applied, despite its successful use in various comparable applications. This research aims to develop a novel approach for seagrass monitoring, using state-of-the-art machine learning with data from Sentinel-2 imagery. We used Tauranga Harbor, New Zealand, for which extensive ground truth data are available, as a validation site to compare ensemble machine learning methods, involving Random Forests (RF), Rotation Forests (RoF), and Canonical Correlation Forests (CCF) with the more traditional Maximum Likelihood Classifier (MLC) technique. Using a group of validation metrics, including F_1, precision, recall, accuracy, and McNemar test, our result indicated that machine learning techniques outperformed MLC with RoF the best performer (F_1 scores ranging from 0.75 - 0.91 for sparse and dense seagrass meadows, respectively). Our study is the first comparison of various ensemble-based methods for seagrass mapping that we are aware of, and promises an effective approach to enhance the accuracy of seagrass monitoring.

Keywords: seagrass, Sentinel-2, Random Forest, Rotation Forest, Canonical Correlation Forest, Maximum Likelihood, Tauranga, machine learning, remote sensing

3.1. Introduction

Alongside mangrove and salt marsh, seagrass has been evaluated as an effective coastal ecosystem for blue carbon storage (Duarte & Krause-Jensen, 2017; Gullström et al., 2018; Oreska et al., 2017). However, ongoing degradation of seagrass meadows (Waycott et al., 2009) is leading to a requirement for accurate mapping and monitoring methods to facilitate the MRV (Monitoring, Reporting, and Verification) approach necessary for broad scale evaluation of their contribution to blue carbon reservoirs (Herold & Skutsch, 2011). In the last decade, satellite imagery has been used extensively in developing seagrass mapping techniques, employing various classification algorithms with or without parallel traditional field surveys (Dat Pham et al., 2019). Among them, Sentinel-2 imagery is becoming more popular for seagrass mapping. Operated by European Space Agency (ESA) since 2015, this sensor supports a high quality image at the spatial resolutions of between 10 and 60 m (ESA, 2015). Sentinel-2 data has been distributed free-of-charge at the top-of-atmosphere corrected level (level 1C), for ρ_{Blue}, ρ_{Green}, ρ_{Red}, and ρ_{NIR} bands at 10 m resolution, and provides a very good resource for intertidal and subtidal ecosystem mapping. Using these data to derive ecosystem spatial properties requires classification algorithms, and overfitting, inaccurate edge detection of different substrata remain as limitations of traditional classification methods (Hossain et al., 2015; Winters et al., 2017; Dat Pham et al., 2019; Gumusay et al., 2019). For seagrass mapping, the problems of misclassification usually relate to the impact of deep water on pixel values or the mixture of substrata within a seagrass meadow (Wicaksono & Lazuardi, 2018). To overcome this problem, very high resolution (VHR) imagery and a variety of classification approaches can be considered (Poursanidis et al., 2018, 2019). Most frequently, probability-theory based models, such as Maximum Likelihood Classifier (MLC), have been applied for seagrass classification (Hossain et al., 2015; Dat Pham et al., 2019). This approach,

however, requires conditions that are difficult to satisfy in the marine environment, including a normal distribution of probabilities, equal co-variance, and large amounts of validation input data (Asmala, 2012; Richards, 2013). In addition, the utilization of the linear or quadratic discrimination functions of a MLC may not work when the boundaries of classes are not well defined (Richards, 2013).

In recent years, machine learning (ML) has emerged as a novel approach for seagrass mapping and monitoring (Dat Pham et al., 2019). Machine learning has the benefits of rapid learning, accommodation of non-linearity (Holloway & Mengersen, 2018) and the availability of an increasing number of new, open source algorithms (Liu, 2017). In the field of seagrass mapping and monitoring, however, the application of machine learning is still in its infancy (Dat Pham et al., 2019). Examples used to date include weighted majority voting using Quickbird image (Mohamed et al., 2018); Logistic Model Trees (LMT), AdaBoost, Random Forest (RF), and Artificial Neural Networks (ANN) using digital images (Bonin-Font et al., 2016)**;** Support Vector Machine (SVM) using Sentinel-2 images (Poursanidis et al., 2019; Traganos & Reinartz, 2017)**;** and Decision Tree (DT) using aerial photographs (Pe'eri et al., 2016). In these examples, when used with high spatial resolution images (< 1 m), machine learning models achieved an accuracy of 92 - 100%. Decision tree models using aerial photographs, however, achieved a lower accuracy of 66% for seagrass meadows when plant cover was below 60% (Pe'eri et al., 2016). These mixed results support the exploration of novel machine learning approaches, particularly for improving low coverage seagrass mapping.

Among various DT ensemble machine learning algorithms, Rotation Forests (RoF) and Canonical Correlation Forests (CCF) algorithms are now emerging as reliable techniques for land cover mapping (Colkesen & Kavzoglu, 2017), landslide mapping using multi-spectral (Sahin et al.,

2018) or hyper-spectral (Moughal, 2013) imagery, and rapid building mapping using multi-source data (Adriano et al., 2019). Through bootstrap sampling, combining multiple independent base classifiers, and applying statistical analysis (principle component analysis (PCA) in RoF model) (Probst et al., 2019; Xiu et al., 2017), these learning algorithms are well-known for reducing the variance and overfitting of the classification results, resulting in a better detection of multi-classes boundaries (Bagnall et al., 2018; Belgiu & Drăguţ, 2016; Feng et al., 2018). In addition, the CCF model does not require optimizing of hyperparameters (Rainforth & Wood, 2015) which make this model simpler to apply for mapping tasks. To our knowledge, these techniques have not been used for seagrass mapping, yet they potentially offer benefits in the classification of low coverage through enhanced recognition of edge boundaries. Therefore, our goal in this study is to compare the use of three ML algorithms, RF, RoF and CCF to the more traditional MLC approach for mapping the aboveground distribution of seagrass communities at low and high coverage, using Sentinel-2 data.

Our target was Tauranga Harbor, New Zealand, for which ground truth data were available, and where a mosaic of dense, sparse, and no seagrass present. We discuss here the difference in performance of selected models for seagrass detection at two densities. Our results are expected to contribute alternative solutions for the mapping and monitoring of seagrass at various regions in the world, and assist conservation of this important blue carbon ecosystem.

3.2. Materials and Methods

3.2.1. Study site

Tauranga Harbor, North Island, New Zealand was selected as our study site (Figure 3.1). The site supports a single seagrass species, *Zostera muelleri*, distributed in the intertidal parts of the harbor (Park, 1999). *Z. muelleri* tolerates at a wide range of salinity (10 - 30 practical salinity unit

(psu)), however a salinity at 12 psu is optimal to produce highest shoot density (Collier et al., 2014). *Z. muelleri* is a small plant when growing intertidally, as in the Tauranga Harbor, with leaves 5 - 30 cm in length and 0.1 - 0.4 cm in width. It has a maximum growth rate in austral summer (December to March) with optimal temperature ranging from 27 - 33 ^{o}C (Collier et al., 2011, 2017; York et al., 2013). Biomass declines gradually over winter to reach a minimum in early spring (October) (Turner, 2007). Since satellite image based mapping uses surface reflectance as input data for the classification, our mapping addresses only the aboveground part of seagrass meadows, and may thus underestimate colonized area if applied in late winter when main leaves have senesced. Flowering and seed production in *Z. muelleri* is rare in New Zealand, and lateral spread is slow and by vegetative mechanisms (Ramage & Schiel, 1998; Schwarz & Turner, 2006).

In Tauranga Harbor, the tide regime is semi-diurnal, with a range of 0.2 - 2.1 m. The size of the harbor means that tide timings are different in various areas in the harbor (Reeve et al., 2018). Examination of harbor bathymetry and seagrass distribution, which is almost exclusively intertidal in the harbor, indicates that seagrass was occupying a depth range of 0.0 - 1.5 m at Sentinel-2 acquisition time. Across the years 2018 - 2019, the mean air temperature ranged from 2.5 ^{o}C in winter to 31.6 ^{o}C in summer (data from the New Zealand MetService: https://www.metservice.com/towns-cities/locations/tauranga/past-weather) whereas mean sea temperature ranged from 14 ^{o}C in winter to 23 ^{o}C in summer (data from the New Zealand sea temperature: https://www.seatemperature.org/australia-pacific/new-zealand/tauranga.htm).

Previous mapping, using aerial photography and manual classifications for the years 1959, 1996, and 2011 (Park, 2011), suggested a loss of ~50% of seagrass area, making a strong case for on-

going monitoring of seagrass cover. In addition, a wide range of seagrass density and substratum make Tauranga Harbor a suitable location for testing novel classification algorithms (Figure 3.2).

Figure 3.1. Study site in Tauranga Harbor using a ρ_{Red}-ρ_{Green}-ρ_{Blue} combination of Sentinel-2 scene (1 May 2019)

Figure 3.2. Dense (*a*) and sparse (*b*) seagrass meadows in Tauranga Harbor (photos taken by N.T.H)

3.2.2. Field survey

A seagrass mapping survey was undertaken from 1 - 7 April 2019 (Figure 3.1) in the intertidal areas of the habour. At low tide, the boundary of seagrass meadows was delimited using a Garmin Etrex 30 global positioning system (GPS) with an accuracy of ± 2 m. Other substrata recorded during the field survey, were bare sand and muddy sand. Macroalgae were neither detected from our field survey nor mentioned in previous mapping reports (Park, 1999, 2011).

Ground truth points (GTPs) were recorded by following the boundary between seagrass meadows and unvegetated areas. In addition, we also recorded the internal boundary of seagrass and bare substrate if they coexisted in the same meadow. The frequency with which GTPs were recorded was related to the 10 m pixel size of Sentinel-2, and varied according to seagrass density. The number of GTPs was 2 - 5 GTPs per pixel in the case of patchy meadows, and decreased to one GTP per pixel or one GTP per 2 - 3 pixels for continuous, dense meadows. GTPs were not collected for seagrass meadows smaller than 100 square meters (fitting with the pixel size of Sentinel-2 imagery), and some meadows were not able to be accessed due to logistic constraints.

Seagrass class boundaries were determined visually, with “dense” and “sparse” boundaries recorded. A meadow was identified as “dense” if the coverage was larger than 80% (Figure 3.2

(*a*)), and "sparse" if the coverage was less than 80% (Figure 3.2 (*b*)). A total of 4315 GTPs were recorded, with 2751 and 1564 for sparse and dense meadows, respectively; 237 GTPs were recorded for other substrata.

3.2.3. Satellite data acquisition and image pre-processing

A Sentinel-2 scene acquired on 1 May 2019 was selected and downloaded from the GLOVIS website (https://glovis.usgs.gov) (Table 3.1). Sentinel-2 scene was pre-processed at level 1C (atmospheric correction at the top of atmosphere), and in the projection of WGS-84 UTM 60S. Sentinel pixels were classified into non-seagrass, sparse and dense seagrass classes according to our field observations. Field and remote sampling were closely synchronous, and we considered the field data to provide a sufficiently accurate representation of seagrass spatial distribution to develop the models.

Table 3.1. Sentinel-2 data acquisitions used for seagrass mapping in this study

Date of acquisition	Time of acquisition*	Spatial resolution (m)	Cloud coverage (%)	First low tide	Second low tide
1 May 2019	10:16 AM	10	0	10:33 AM	22:52 PM

*: Time is local time of New Zealand zone

Atmospheric correction

Atmospheric correction was executed in a Python™ environment using the dark spectrum fitting method in ACOLITE (RBINS, 2018). This fast and free-to-download tool has been adapted for aquatic application and presented a reliable atmospheric correction for Landsat and Sentinel-2 (Vanhellemont, 2019). This process converts pixel values from top of atmospheric (at level 1C) to surface reflectance for water pixels.

Table 3.2. Selected parameters for atmospheric correction using ACOLITE

Parameters	Values
Ancillary data	
Gas transmittance	True
Ozone concentration (cm^{-1})	0.3
Water vapour concentration ($g\ cm^{-2}$)	1.5
Pressure	Normal pressure
Masking	
Level 2 water masking (nm)	1600
Negative reflectance masking	True
Cirrus masking	True
Other parameters	
Sky correction	True
Dark spectrum fitting	Fixed
Sun glint correction	False
Output parameter	
Surface reflectance for water pixel (ρ_w)	ρ_{w443} ρ_{w560} ρ_{w665}

The coefficients and options of selected parameters are presented in Table 3.2. For our study site, sun glint was not observed in the acquired scene. Therefore, the parameter of sun glint correction was set to False. SWIR, a spectral band of Sentinel-2 with the wavelength close to 1600 nm, was used to mask land and cloud pixels from water pixels. Corrected surface reflectance for water pixels of blue (ρ_{w443}), green (ρ_{w560}), and red (ρ_{w665}) bands were used for the next step of water column correction.

Water column correction

Satellite image acquisition did not coincide with low tide in Tauranga Harbor (Table 3.1), and most, but not all seagrass were inundated at the time of acquisition. We therefore next applied a water column correction. Previous studies suggested that visible wavelengths are the most sensitive to seagrass meadows (Green et al., 1996; Ha et al., 2012; Garcia et al., 2015) and penetrate well into water (Green et al., 1996). Conversely, several studies indicated that the near infrared band is rapidly absorbed in the underwater environment (Green & Edwards, 2000; Traganos & Reinartz, 2017), and therefore may contribute noise to the image, leading to a low accuracy of underwater habitat detection (Ha et al., 2012; Chen et al., 2018). As a result, we decided on the use of only visible spectra in the current study without the ρ_{NIR} band. After the water column correction step, water pixels in the ρ_{Blue} (458 - 523 nm), ρ_{Green} (543 - 578 nm), and ρ_{Red} (650 - 680 nm) spectral bands were selected as input data for the evaluation of the model's performance. Water column correction was conducted using bottom reflectance index (BRI), as proposed by Sagawa et al (2010). Instead of absolute values of bottom reflectance, this approach creates different indexes for various bottom types which are used for the step of classification. The index is calculated using Equation (3-1):

$$\mathrm{BRI}_i = \frac{\mathrm{SR}_i}{e^{-ki*g*z}} \tag{3-1}$$

in which:

BRI_i*:* Bottom reflectance index of band *i*

SR_i: Surface reflectance for water pixel of band *i* (ρ_{wi})

k_i: attenuation coefficient of solar radiance in water column (m^{-1}) of band *i*

g: a geometric factor accounting for the path length through the water

z: water depth (m)

Values of SR and k of blue, green, and red bands were retrieved from the atmospheric correction step. Water depth (z) was extracted from bathymetry published by the national institute of water and atmospheric research (NIWA). g was calculated using Equation (3-2) (Lyzenga et al., 2006):

$$g = \frac{1}{sec(SolarZenithAngle) + sec(SatelliteNadirAngle)} \quad (3\text{-}2)$$

in which:

$$sec(SolarZenithAngle) = \frac{1}{cos(SolarZenithAngle)}$$

$$sec(SatelliteNadirAngle) = \frac{1}{cos(SatelliteNadirAngle)}$$

g factor was calculated as 0.0245 for Sentinel-2 scene at study site.

3.2.4. Image classification with ML ensemble-based and MLC methods

Selection of maximum likelihood, random forest, rotation forest, and canonical correlation forest classifiers

MLC is the most popular classification method in remote sensing (Dat Pham et al., 2019) and is based on probability theory. This model requires a normal distribution, equal covariance, and sufficient number of training samples (Hogland et al., 2013) to maintain a reliable result. Class mean vector and covariance matrices are used to minimize class distance and maximize the probability of a feature belonging to selected class, using quadratic or linear discrimination functions. This method, however, may overfit posterior probabilities and result in inaccurate classification when the assumptions are violated.

RF (Breiman, 2001) builds a forest of decision trees from a bootstrap sampling of training data points, and uses only a selection of all the samples for each decision tree. For each subset, a decision tree is built using two thirds of the total samples for training and one third for validation of the RF model. Majority voting is applied to arrange labels to given classes. The RF model constructs and combines a large number of base-decision trees using the classification and regression tree (CART) algorithm. Known as a robust and consistent model, RF is the most popular ensemble based model for classification problems, and, in particular, for seagrass mapping (Dat Pham et al., 2019).

Rodriguez et al., (2006) presented RoF – a feature extraction based method (Rodriguez et al., 2006). RoF is expected to provide an alternative selection for both regression and classification tasks (Bagnall et al., 2018). The training data created by a bootstrap sampling is split into *K* subsets and then principal component analysis (PCA) is applied for each subset. All the principal components are retained and a number of decision trees are built from these transformed datasets. RoF can perform better than RF with a smaller number of trees, and therefore reduces the time for running the model. Instead of estimating the average value from all the trees in RF, a confidence value is calculated to assign a label to a given class with the highest value of the confidence.

CCF creates a number of canonical correlation trees (CCTs) using canonical correlation analysis (CCA) to maximize the correlation between the input data and the selected labels. The authors (Rainforth & Wood, 2015) confirmed the robustness of the model and also introduced the projection bootstrapping which uses of all of the data and improves prediction accuracy over RF models. CCF is different from RF and RoF, lacking the tuning hyper-parameter, but offers similar

accuracy with a smaller number of trees (Rainforth & Wood, 2015) and therefore, may consume less computation time for the training phase.

Training and testing dataset

Differences between collected date of GTPs and acquisition date of satellite image existed and, in the variable marine environment, this may be inevitable (Koedsin et al., 2016; Kovacs et al., 2018; Meyer & Pu, 2012; Tsujimoto et al., 2016). In our study site, the difference between the field survey (1 - 7 April 2019) and acquired date of Sentinel-2 (1 May 2019) is approximately 23 days which was acceptable comparing to various literatures (Meyer & Pu, 2012; Tsujimoto et al., 2016; Wicaksono & Lazuardi, 2018; Kovacs et al., 2018). In addition, April and May are in the austral autumn. Therefore, the GTPs are reliable to support the selection of image pixel for training and testing dataset. Following the GTPs, the regions of interest (pixels) of three classes: dense seagrass, sparse seagrass, and non-seagrass, were selected and randomly split into 60% for training and 40% for testing phases, and used as a unique input for all selected models (Table 3.3).

Table 3.3. Number of pixels for training and testing at various acquisition dates

Sentinel acquisition date	Number of pixels	
	60% for training	40% for testing
1 May 2019	8586	5724

Performing of MLC, RF, RoF and CCF models

The optimization and performance of the RF, RoF, and MLC algorithms for image classification were conducted in a Python™ environment, using Jupyter notebook as the code editor. RF and RoF codes were sourced from Scikit-learn (Pedregosa et al., 2011), GitHub (Joshua, 2016), and MLC from mlpy (Albanese et al., 2012; Davide, 2012). For the RF and the RoF models, optimi-

zation of the hyper-parameters used a grid search with fivefold cross-validation (Bagnall et al., 2018; Probst et al., 2019). The optimization and performance of the CCF model was conducted in the Matlab environment, using the source code of Rainforth and Wood (Rainforth, 2018; Rainforth & Wood, 2015). The CCF models were performed with 10, 30, 50, 100, 200, and 500 trees, and an optimal number was selected based on the lowest misclassification rate, highest Kappa coefficient, and an acceptable computation time.

The results of the CCF model were exported to a CSV format for model comparisons in Python™. All source codes were open access, and the codes developed in this study will be submitted to GitHub (https://github.com). The image processing and classification were performed using a desktop computer with four 3.8 GHz physical cores and 16 Gb RAM.

Evaluation criteria

Equations (3-3) - (3-7), involving accuracy, Kappa coefficient, precision, recall, and F_1 were used to compare the performance of selected models. High precision means that the classifier is able to detect seagrass pixels precisely, whilst high recall means that the classifier is able to find all possible pixels of seagrass.

$$accuracy(y, y_{pred}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} 1\left(y_{predi} = y_i\right) \tag{3-3}$$

in which:

y_{pred}: predicted value

y: corresponding true value

$$Kappa = \frac{p_o - p_e}{1 - p_e} \tag{3-4}$$

in which:

p_o: observed agreement ratio

p_e: expected agreement

$$\text{Precision} = \frac{tp}{tp + fp} \tag{3-5}$$

$$\text{Recall} = \frac{tp}{tp + fn} \tag{3-6}$$

$$F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \tag{3-7}$$

in which:

tp: true positive

fp: false positive

fn: false negative

In addition, the non-parametric McNemar test was used to compare statistically the accuracy of selected models in this research. The test was executed in a Python™ environment using the mlxtend library (Raschka, 2018). The chi-square value (χ^2) is calculated from Equation (3-8) with Edward's continuity correction.

$$\chi^2 = \frac{(|fn - fp| - 1)^2}{(fn + fp)} \tag{3-8}$$

in which:

fn: false negative

fp: false positive

3.3. Results

Pixels were selected from each ground-truth class for evaluation of automated classification. The main challenge to the automated classification emerged as the problem of the mixing of sparse seagrass and bare sand in the same meadow. In particular, the substratum in very shallow water belonging to the non-seagrass class could be confused with sparse seagrass.

3.3.1. Hyper-parameters tuning for RF, RoF, and optimizing number of trees for CCF models

The hyper-parameters were tuned and consistently maintained for RF and RoF running during the training and testing phases (Table 3.4).

Table 3.4. Hyper-parameters selected for use in RF and RoF models

	RF	RoF
Bootstrap	True	True
Max depth	20	30
Max feature	Auto	Auto
Min sample leaf	1	3
Min sample split	9	7
Number of tree	100	100
Number of subset		3

For the CCF model, lowest misclassification rates, and highest Kappa coefficient values were recorded at 200 trees (Figures 3.3 (*a*), 3.3 (*d*)). As a result, we selected 200 trees as an optimal choice for CCF, which gave a computation time for training and testing runs, which increased linearly with the number of trees, of 27 and 4.5 s respectively (Figures 3.3 (*b*), 3.3 (*c*)).

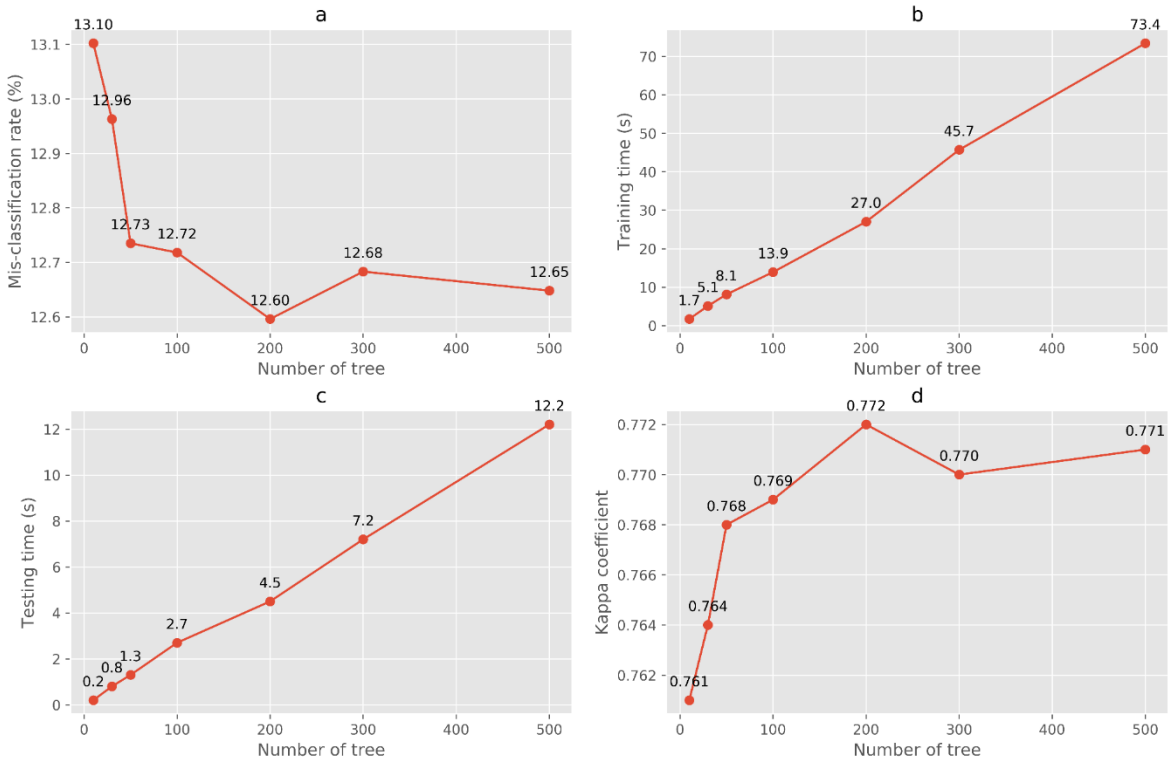


Figure 3.3. Number of tree, mean values of mis-classification rate, computation time (training and testing time is the computation time measured for training and testing phases, respectively), and Kappa coefficient using data described in Table 3.3

3.3.2. Comparing the performance of RF, RoF, CCF, and MLC models for seagrass mapping

The ML methods consistently outperformed the MLC for all evaluation metrics (Table 3.5), and the McNemar test indicated these effects as significant (Table 3.6). In particular, the precision of ML models are similarly high, and greater than MLC methods for dense and sparse seagrass, whilst very high recall was observed for both classes using the MLC model. F_1 values of MLC were lower than the ML methods for both dense and sparse seagrass. Among the ML methods, RoF outperformed CCF and RF models (Figure 3.4 and Table 3.5), and this difference was also statistically significant (Table 3.6). RoF values show highest values for precision and F_1 across all three ML methods for both dense and sparse seagrass classes (Table 3.5). The ML models are less effective at discriminating sparse than dense seagrass.

All ensemble-based ML models were able to detect dense seagrass meadows with very high F_1 scores ranging from 0.90 - 0.91 whereas the MLC model produced a F_1 score of 0.75 (Table 3.5). The performance of these ML models is consistent with a balance of high precision and recall. A good performance for dense seagrass class is conceivably due to a spectral separation of this class from sparse and non-seagrass classes. As can be seen from Table 3.5, the RoF model improved the accuracy of the classification, in terms of F_1 score, by 1% compared to the CCF, and RF, but by nearly 17% over the MLC model for dense seagrass class. Corresponding improvements were 3% (CCF, RF) and 33% (MLC) for the sparse seagrass class.

In addition, the findings also demonstrate how seagrass meadows made the classification more challenging due to the mixing between seagrass and bare sand in the same meadow. The RoF model still produced the highest F_1 score (0.75), followed by the CCF and RF (0.73) models (Table 3.5). Conversely, the MLC model yielded a F_1 score of only 0.50, which was significantly lower than the three ML models. Regarding computation time, the CCF model requires more time to run whilst the MLC model executes very fast for both training and testing phases (Table 3.5).

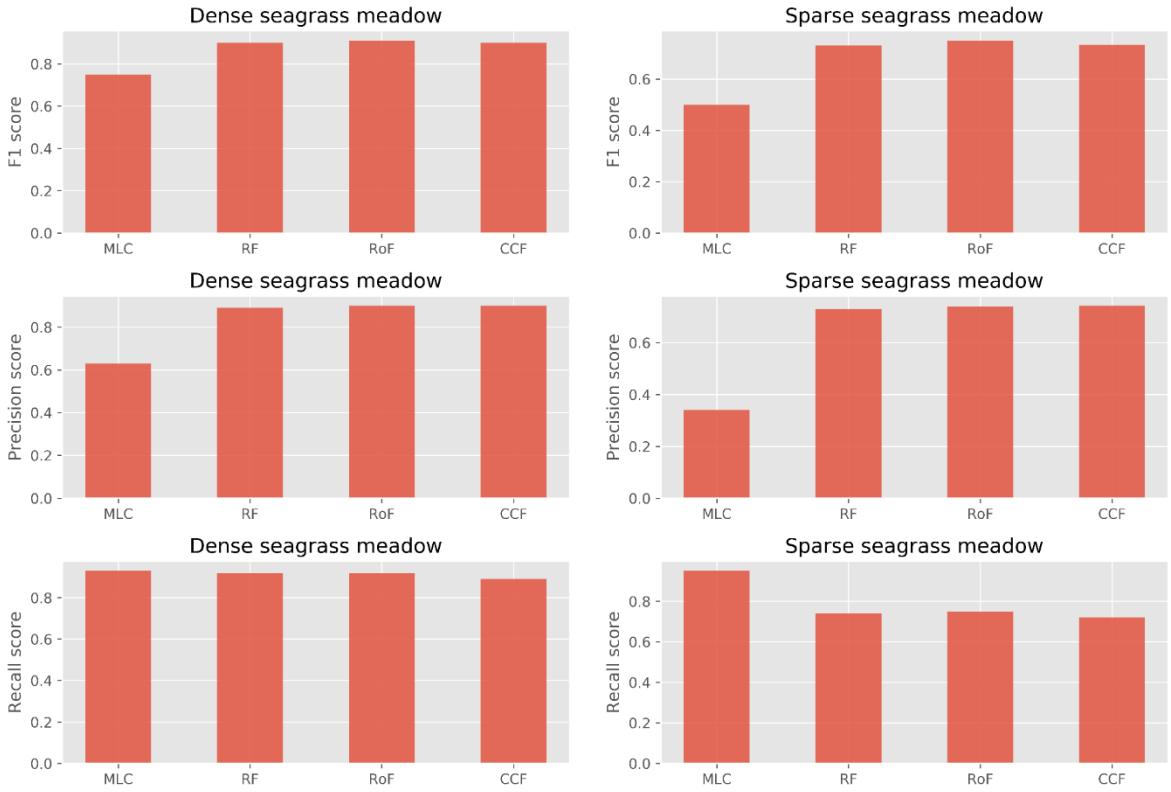


Figure 3.4. A comparison of F_1, precision, and recall scores for Sentinel-2 scene (1 May 2019) using ensemble-based and traditional approaches

Table 3.5. Accuracy, precision, recall, and F_1 of model performing for selected Setinel-2 scene

Model	Accuracy	Precision DS*	Precision SS*	Recall DS	Recall SS	F_1 DS	F_1 SS	Training time (s)	Prediction time (s)
RoF	**0.88**	**0.90****	**0.74**	0.92	0.75	**0.91**	**0.75**	3.54	0.37
CCF	0.87	**0.90**	**0.74**	0.89	0.72	0.90	0.73	27.06	4.57
RF	0.87	0.89	0.73	0.92	0.74	0.90	0.73	0.07	0.00
MLC	0.51	0.63	0.34	**0.93**	**0.95**	0.75	0.50	0.00	0.01

*: DS: dense seagrass and SS: sparse seagrass classes

**: Bold values indicate the best performance

Table 3.6. McNemar test comparing the performance of selected models in prediction of seagrass class

	χ^2	*p*-value
	Scene 1 May 2019	
RoF – RF	9.03	0.00
RoF – MLC	1,667.29	0.00
RoF – CCF	2,069.37	0.00
RF – MLC	1,599.46	0.00
RF – CCF	1,995.93	0.00
CCF – MLC	42.51	0.00

As the RoF model yielded the highest performance assessment for seagrass mapping in this research, we employed this model to create the classified maps (Figure 3.5). The distribution of seagrasss meadows were consistently in the middle and southern part of the harbor which may suggest optimal sites for blue carbon assessment as well as potential targets for long-term conservation of seagrass in Tauranga Harbor. Seagrass area was approximately estimated as 1,027.59 ha in May 2019.

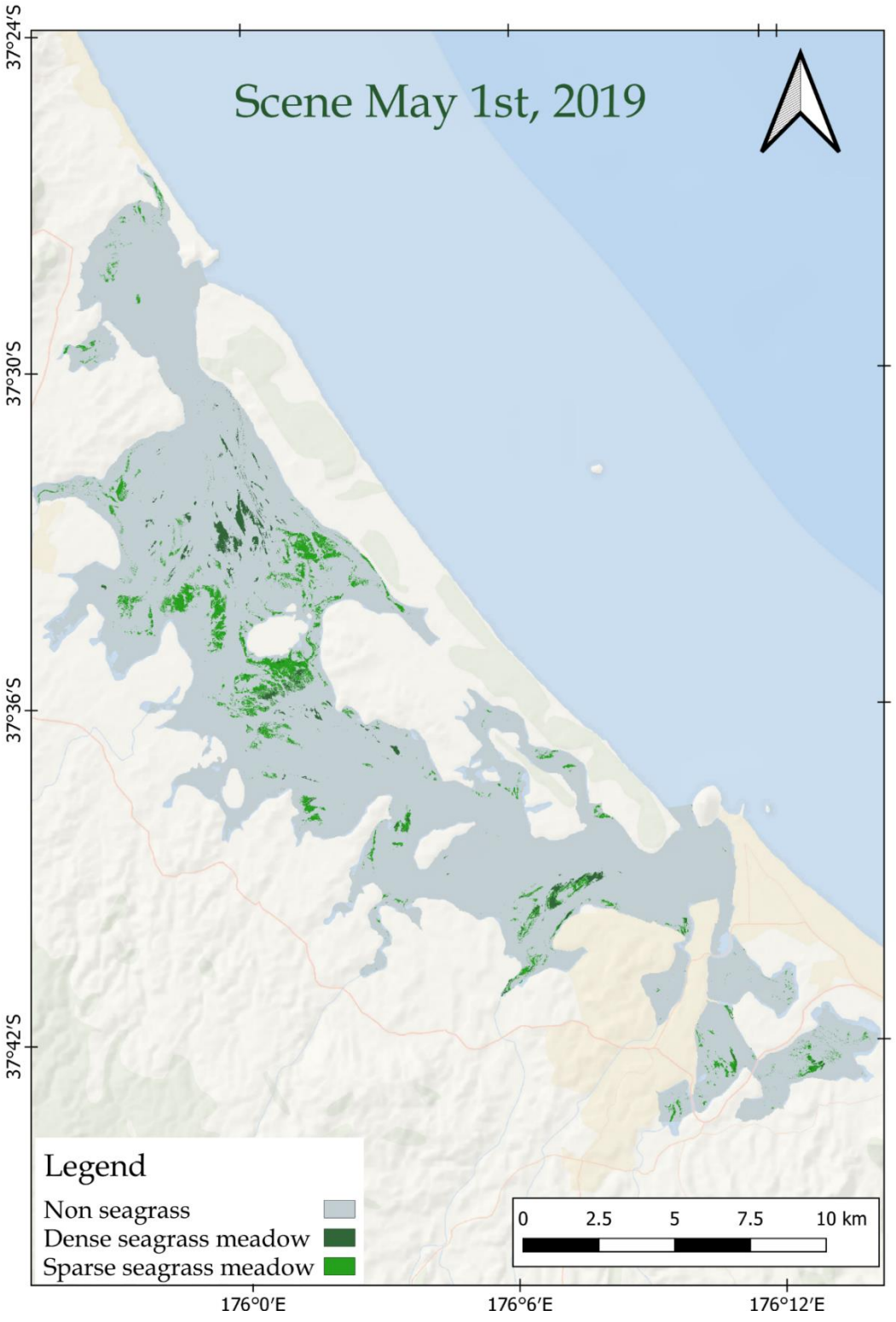


Figure 3.5. Seagrass map for Tauranga Harbor, 1 May 2019, derived using the RoF model applied to Sentinel-2 imagery

3.4. Discussion

As far as we are aware, this research is the first attempt to compare the performance of RF, RoF, CCF and MLC methods for seagrass mapping with a full radiometric correction of the images. Desirable characteristics for seagrass mapping are both high precision and recall. To give a final coherence score and harmonize the values of precision and recall, F_1 score is usually preferred to evaluate a model's performance. The research presented here suggests that ML models detect dense seagrass meadows well, and out-perform the traditional MLC approach.

Of the machine learning ensemble approaches used, CCF and RF models performed less well than the RoF model, contradicting a superior performance of CCF in other studies (Rainforth & Wood, 2015). CCF produced lower recall whilst RF created lower precision for both dense and sparse seagrass classes. For sparse seagrass meadows, CCF detected more precisely than the RF model. In addition, MLC produced very high recall but low precision scores, for both seagrass classes, leading to lower F_1 score and accuracy than ensemble based models. Overall, our results show that RF, RoF, and CCF are good performers with a balance of high precision and recall scores, whilst very low precision scores of dense and sparse seagrass classes, ranging from 0.34 - 0.63, were found for MLC. These results confirm the robustness and consistency of machine learning ensemble based methods in comparison to MLC. We hypothesis that the poor performance of MLC may be because of the need of input data to satisfy built-in assumptions, described above, which are difficult to sustain in the spatially heterogeneous marine environment.

Of the methods tested here, only the RF technique has previously been applied to seagrass mapping, using very high spatial resolution imagery. In that case, high precision (0.947) and recall (0.968) values were determined mapping *Posidonia oceanica* from digital airborne images, though no comparison to other methods was attempted (Bonin-Font et al., 2016). In another

seagrass study, overall accuracy only reached 82% using the RF algorithm applied to RapidEye imagery (Traganos & Reinartz, 2018). Considering the size of seagrass meadows and the mix of substrate in Tauranga Harbor, the measured scores in our results are reliable for both dense and sparse seagrass mapping using medium spatial resolution of Sentinel-2 data (10 m pixel size). In other studies that allow comparison of RF, RoF, and CCF models, CCF slightly outperformed RF and RoF models for land cover mapping (Colkesen & Kavzoglu, 2017), RoF outperformed RF and CCF for mangrove mapping (Pham et al., 2019) whilst a similar performance of RoF and CCF models was noted previously for land slide mapping (Sahin et al., 2018).

In addition to performance advantages, RF and RoF are easy to execute in Python™, whilst CCF is confined to the Matlab environment. The open source operating environment and diverse Python™ libraries provide multiple solutions for seagrass mapping and monitoring, and enhance the capacity to develop novel algorithms for various tasks in marine science (Lemenkova, 2019). Several libraries in the Python™ environment support a built-in framework for classification problems with a long list of state-of-the-art machine learning algorithms (Raschka & Mirjalili, 2017). This ease-of-use approach allows a person with a minimum programming skill to make the classification more reliable with machine learning. Recently, cloud computing divers the executable environments for the big Earth observation data, and coastal resources mapping, in particular. A cloud computation system is able to deal with a massive remote sensing data, parallel processing of satellite image using multiple data centers, and enables to respond to on-demand information requests at the country and global scales (Yan et al., 2018; Yao et al., 2019). The using of open source machine leaning algorithms in Python™ and cloud system, therefore promises the big scale and reliable mapping in the future.

Our results validated the performance of RF, RoF, and CCF models for seagrass mapping and suggest the RoF technique is a promising novel approach to further seagrass monitoring at various sites in the world. The current study does have some limitations. The mismatch between the number of ground truth points (GTPs) and Sentinel-2 pixel size (10 × 10 m) may raise a degree of uncertainty in classification, particularly for sparse seagrass. However, we consider that despite this mismatch, a sufficient and representative number of GTPs for each class was collected during the field survey (as presented in the subsection 3.2.2) to have confidence in classification. Related issues that might explain the low values of precision and recall for sparse seagrass meadows are issues of mixed pixel, whereby small seagrass patches or dispersed clumps within a pixel challenge the classification process. To our knowledge, these effects are not easy to compensate for in the case of low to very low coverage seagrass, using Sentinel-2 imagery. Thus, the use of very high spatial resolution sensors such as Worldview (~ 0.3 - 0.4 m) (Koedsin et al., 2016; Kovacs et al., 2018) or Pleiades-1 (0.5 m) is currently being investigated for future studies for seagrass mapping. Moreover, with the development of computer vision and pattern recognition, deep learning approaches with a Convolutional Neural Network (CNN) such as U-Net or Recurrent Neural Network (RNN) for semantic segmented imagery applied sub-pixel techniques should be encouraged for future studies (Dat Pham et al., 2019).

3.5. Conclusions

We tested the performance of ML ensemble-based and MLC methods for seagrass mapping from Sentinel-2 data. Using Tauranga Harbor as a validation site, our comparison indicated that all ML-based approaches significantly outperformed MLC. MLC failed to detect sparse seagrass meadows, with a low F_1 score of 0.50. We noted a better performance of RoF compared to RF and CCF models with a highest F_1 score of 0.91 and 0.75 for dense and sparse seagrass classes.

Our results attest to the reliable application of RoF model for the mapping and monitoring of seagrass in shallow water, using Sentinel-2 imagery. Despite a lower accuracy for sparse than dense seagrass meadow classification, CCF model shows potential for the mapping of seagrass and merits further testing at various scales and in various case studies. Regarding MLC, this model is still an applicable candidate for dense seagrass meadows, however may be not applicable for sparse to very sparse seagrass meadows mapping.

Chapter 4 Detecting multi-decadal changes in seagrass cover in Tauranga Harbor, New Zealand, using Landsat imagery and boosting ensemble classification techniques

The contents of this chapter have been published in the ISPRS International Journal of Geo-Information (https://www.mdpi.com/2220-9964/10/6/371). A copy of the published paper is bound into Appendix 4.

Abstract: Seagrass provides a wide range of essential ecosystem services, supports climate change mitigation, and contributes to blue carbon sequestration. This resource, however, is undergoing significant declines across the globe, and there is an urgent need to develop change detection techniques appropriate to the scale of loss and applicable to the complex coastal marine environment. Our work aimed to develop remote-sensing-based techniques for detection of changes between 1990 and 2019 in the area of seagrass meadows in Tauranga Harbour, New Zealand. Four state-of-the-art machine-learning models, Random Forest (RF), Support Vector Machine (SVM), Extreme Gradient Boost (XGB), and CatBoost (CB), were evaluated for classification of seagrass cover (presence/absence) in a Landsat 8 image from 2019, using near-concurrent Ground-Truth Points (GTPs). We then used the most accurate one of these models, CB, with historic Landsat imagery supported by classified aerial photographs for an estimation of change in cover over time. The CB model produced the highest accuracies (precision, recall, F_1 scores of 0.94, 0.96, and 0.95 respectively). We were able to use Landsat imagery to document the trajectory and spatial distribution of an approximately 50% reduction in seagrass area from 2237 ha to 1184 ha between the years 1990–2019. Our illustration of change detection of seagrass in Tauranga Harbour suggests that machine-learning techniques, coupled with historic satellite imagery, offers potential for evaluation of historic as well as ongoing seagrass dynamics.

Keywords: seagrass mapping; Tauranga Harbour; change detection; Landsat; Random Forest; Support Vector Machine; Extreme Gradient Boost; CatBoost; machine learning

4.1. Introduction

Seagrass provides a number of valuable ecosystem services in coastal areas, including primary production, biogenic habitat production, water filtering, wave energy attenuation, and sediment trapping (Nordlund et al., 2016; Nordlund et al., 2018). In recent years, blue carbon, including seagrass meadows, has been acknowledged as an important service for climate change mitigation because of its value in the sequestration of carbon (Fourqurean et al., 2012; Gullström et al., 2018). Seagrass meadows, however, have declined and degraded across various regions in the world, a change largely attributed to anthropogenic effects (Marbà et al., 2015; Orth et al., 2006; Waycott et al., 2009).

The destruction of seagrass leads to the loss of various ecosystem services (Cullen-Unsworth & Unsworth, 2013; Marbà et al., 2015), threatens the stability (Waycott et al., 2009) and long-term livelihood of the fisherman in the coastal areas (Bujang et al., 2018; Hejnowicz et al., 2015). Therefore, an accurate and rapid technique to inventory this resource is in high demand (Orth et al., 2006; Unsworth et al., 2018; Dat Pham et al., 2019), to contribute baseline data for the evaluation of coastal ecosystem dynamics, establishment of marine protected areas, and functional zoning fitting to the local conditions. Where this can include a historic perspective, it can provide a comprehensive understanding of system change.

Several attempts at mapping and monitoring seagrass meadows using different satellite sensors and approaches have been reported (Dat Pham et al., 2019). Change of seagrass cover has been assessed using RapidEye (Traganos & Reinartz, 2018), Indian satellite image (IRS LISS IV) (Paulose et al., 2013), WorldView-2, Ikonos, Quickbird-2 (Roelfsema et al., 2015), and Landsat (Chen et al., 2016; Hossain et al., 2015; Amone-Mabuto et al., 2017; Phinn et al., 2018) in various parts of the globe including the Mediterranean, the USA, Australia and Malaysia. The

temporal range of these attempts is constrained by the various platform launch dates, and typically range from 5 to ~ 25 years. Few efforts have attempted a longer-term change detection (30–40 years) of seagrass, and accuracy assessment has not frequently been reported for such long-term change detection. The reasons for this may relate to a deficiency of ground-truth data against which to evaluate older satellite scenes, and a need for imagery for the development of robust models for the classification of seagrass meadows in variably submerged conditions to be captured at optimal times to allow traditional classification procedures to be applied.

In recent years, machine learning (ML) has been emerging as an effective approach in various classification tasks, including for seagrass mapping (Dat Pham et al., 2019; Ha et al., 2020). ML provides improvements over the traditional Earth observation (EO) data classification approaches, to deal better with the challenges of mixed habitat, coarse spatial resolution of satellite imagery, water column and atmospheric interference in coastal habitats (Ha et al., 2020; McCarthy & Halls, 2014; Sousa et al., 2016). Advantages of ML models are their use of non-parametric approaches, requiring no assumptions of normal distribution of input data, effective use of noisy data, and capability for multiple feature extraction (Ahmad, 2019; Camps-Valls, 2009; Lary et al., 2016; Maxwell et al., 2018). The application of ML techniques to multitemporal satellite data, gathered from different satellite platforms, may therefore improve the overall accuracy of the classification result and enhance the reliability of seagrass change assessment. A range of different ensemble-based supervised classification techniques, such as boosting and bagging approaches (Camps-Valls, 2009; Sousa et al., 2016; Lary et al., 2016; Huettmann, 2018) have been considered and tested in the literature for this type of task (Bühlmann, 2012; Machova et al., 2006). The most important differences between the bagging and boosting methods come from the approaches to creation of training and testing datasets, and

how the bagging and boosting methods deal with weak learners during the learning process. Despite the potential for improved classification accuracy in suboptimal datasets, these approaches have not yet been fully implemented for seagrass change detection (Dat Pham et al., 2019). We are aware of only a single study, using Random Forest (RF) classification, for mapping the change of seagrass cover (Traganos & Reinartz, 2018). In the case study reported by these authors, the performance of the model was unstable and the accuracy varied among acquired scenes. Here we test the performance of a range of ML models, both boosting and bagging methods, with a time-series of satellite images, to compare their performances for assessment of seagrass cover and long-term change in Tauranga Harbour. Our goal is to improve the accuracy of tools for seagrass mapping and change detection.

Landsat time-series data were selected for the current study as the longest available time series and as freely available satellite remotely sensed resources. Landsat has operated since 1972 and provides continuous, homogeneous input data up to the most current Landsat 8 operational land imager (OLI) in orbit (Northrop, 2015). The Landsat multitemporal data has been used previously for several long-term change detection tasks (Zhu, 2017; Pham et al., 2019) with the combination of long-term acquisition, medium spatial resolution, and the high quality of atmosphere-corrected products cited as important attributes. The spatial resolution has been retained as 30 m through eight generations (Landsat 1–Landsat 8); however, the radiometric resolution has been improved from 8 bit to 12 bit, leading to a better recognition of surface objects (Ihlen, 2019). In addition, Landsat imagery includes blue, green, and red wavebands, which are the most appropriate for underwater resource mapping (Green et al., 1996; Ha et al., 2012; Garcia et al., 2015), but have not yet been evaluated for long-term seagrass change detection (Pham et al., 2019). Thus, our work attempts to fill a gap in the current literature by

assessing the performance of historic Landsat imagery, coupled with various machine-learning boosting and bagging models implemented in an open-source environment, in mapping changes in seagrass extent in a tidally inundated environment.

We employed two well-known models, i.e., Support Vector Machine (SVM), Random Forests (RF); and two novel techniques, Extreme Gradient Boosting (XGB), and CatBoost (CB)) for the classification of seagrass meadows in Tauranga Harbour from Landsat imagery, and for detecting change across 29 years. The results demonstrated that the novel classification method CB was successful in describing the dynamics of change in seagrass in the study site as well as contributing baseline data for further assessment of change.

4.2. Materials and Methods

4.2.1. Study site

We selected Tauranga Harbour (North Island, New Zealand) as the study site (Figure 1), due to its large size (201 km^2 in surface area (Park, 2011)), variation in water depth (from 0 m when exposed to 20 m in deep channels (Reeve et al., 2018)), widely distributed but patchy seagrass cover and the availability of historic ground-truth information. The tidal regime is semidiurnal, with a range of 0.2–2.1 m, and the estuary has an average water residence time of 3–8 days (Tay et al., 2013). *Zostera muelleri* is the only species of seagrass, occurring primarily in the intertidal parts of the harbor (Park, 2011; Ha et al., 2020). The growth rate of *Z. muelleri* is optimal at 12 practical salinity units (psu) (Collier et al., 2014) and 27–33 °C (York et al., 2013; Collier et al., 2017). It attains its highest biomass in the austral summer and declines gradually over the winter, reaching a minimum cover in early spring (Turner, 2007). Flowering and seed production of *Z. muelleri* is rare in New Zealand, reproduction is primarily vegetative and patch dynamics are correspondingly slow (Ramage & Schiel, 1998; Schwarz & Turner, 2006). Seagrass is primarily

intertidal in the estuary and, based on bathymetry and tidal predictions (Reeve et al., 2018) at the time of the Landsat image acquisition, water depths ranged between 0.0–1.5 m in the locations where seagrass was present.

In recent decades, Tauranga Harbour has been increasingly influenced by agricultural activities in the northern part (between 37.44° S and 37.54° S) and urban development in the southern part (between 37.62° S and 37.72° S). Episodic high loadings of sediment have been recorded and have resulted in the accumulation of sediment and high turbidity over the autumn and winter seasons (Hicks et al., 2019; Hicks, 2019). Changes in the sedimentary environment have been implicated in negative impacts on the growth of seagrass (Cabaço et al., 2008; Saunders et al., 2017), though other factors may also be involved. Available maps of seagrass in 1959, 1996, and 2011 derived from manual classification of aerial photography provided a resource for model validation (Park, 2011).

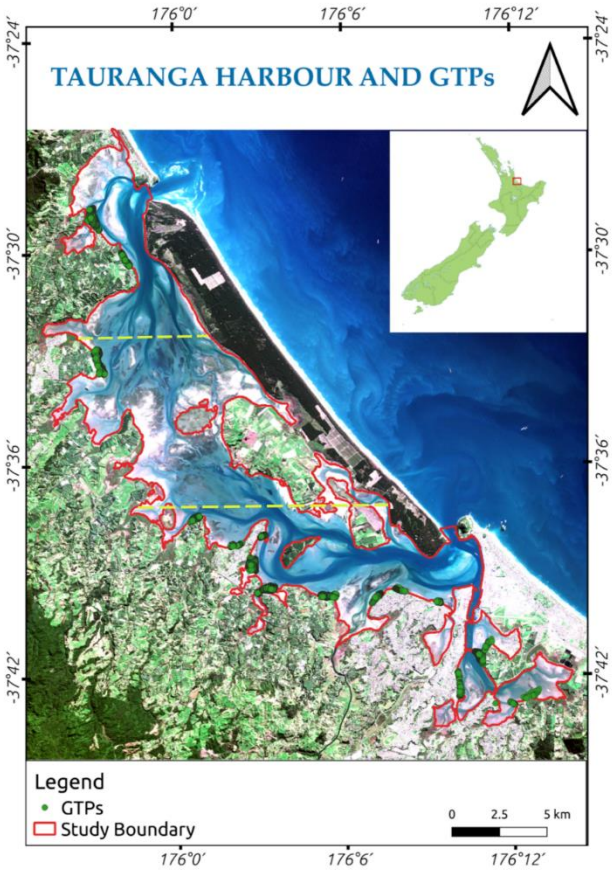


Figure 4.1. Tauranga Harbor – A study site (ρ_{Red} - ρ_{Geen} - ρ_{Blue} composited Landsat image, date on 23 May 2019). GTPs, collected on 1 - 7 April 2019, are indicated by green circles (yellow lines indicate the boundaries of the northern, central, and southern harbor).

4.2.2. Satellite image acquisition

Landsat images were downloaded from the GLOVIS website (USGS, 2019) for the years 1990, 2001, 2011, 2014, and 2019 (Table 1) at process level 1 (pixel value in digital number), and in the projection of WGS-84 UTM 60S. Landsat images were selected based on: (1) the acquired time of the Landsat image that coincided as closely as possible to low tide at the study site; (2) the image that had the lowest coverage of cloud; (3) whether there existed a similar acquisition month among the scenes. In practice, we selected scenes that ranged 1–2 months around March (Table 2).

Table 4.1. Landsat data acquisitions used for seagrass mapping and change detection.

Date of Acquisition (MM/DD/YYY)	**Landsat Generation**	**Time of Acquisition [a]**	**Spatial Resolution (m)**	**Cloud Coverage (%)**	**First Low Tide [b]**	**Second Low Tide [b]**
4 April 1990	Landsat 4 TM	10:16 AM	30	2	02:49 AM	15:09 PM
10 March 2001	Landsat 7 ETM+	10:16 AM	30	0	08:14 AM	20:35 PM
17 February 2011	Landsat 5 TM	10:15 AM	30	2	06:33 AM	18:57 PM
6 March 2014	Landsat 8 OLI	10:15 AM	30	0	11:41 AM [c]	
23 May 2019	Landsat 8 OLI	10:15 AM	30	0	04:14 AM	16:29 PM

[a]: Local time of New Zealand zone. [b]: Tide data was retrieved from the National Institute of Water and Atmospheric Research (NIWA). [c]: Only one low tide at the study site.

Table 4.2. Available aerial, Google Earth images corresponding to historic Landsat images acquisition.

Landsat Image Acquisition	Nearest Aerial Image Acquisition	Aerial Image Spatial Resolution (m)	Google Earth Image (Year of Acquisition)
April 1990	February 1991 March 1992	0.23	December 1990
March 2001	February 2003	0.23	December 2001
February 2011	February 2011	0.25	
March 2014	March 2014	0.125	

4.2.3. Field survey data

A field survey was undertaken from 1–7 April 2019 (Figure 1) in the intertidal areas of the harbor. At low tide, the boundary of seagrass meadows was delimited using a Global Positioning System (GPS) Garmin Etrex 30 with an accuracy of ±2 m. Other substrata recorded during the field survey were bare sand and muddy sand. Macroalgae were neither detected from our field survey nor mentioned in previous mapping reports (Park, 1999, 2011).

Ground-Truth Points (GTPs), which were the base points to make the regions of interest (ROIs) for given classes, were recorded by following the boundary between seagrass meadows and non-vegetated areas. A total of 4315 GTPs were recorded for seagrass distribution, and 237 GTPs for other substrata in the harbor.

4.2.4. Ground-truth historical scenes

Before 2019, no GTPs from field surveying were available, therefore we used aerial and Google Earth images (Table 2) and published documents (Park, 1999, 2011) to identify regions of inter-

ests (ROIs), within which we were able to determine seagrass presence/absence with sufficient confidence to develop the models and to evaluate the accuracy of the hindcast seagrass maps. High-resolution aerial imagery exists from the years 2011 and 2014, and cloud-free, near-low-tide Landsat scenes, from February 2011 and March 2014, could be found that coincided with these. However, for the Landsat scenes in 1990 and 2001, aerial images were only available with a gap of 1–2 years. These included aerial images in 1991–1992 (monochrome and colour) and 2003 (colour). We found Google Earth images (identified as Landsat/Copernicus images in the Google Earth application) for both December 1990 and December 2001, which were in the austral summer and were close to the acquisition time of the Landsat scenes in April 1990 (austral autumn) and March 2001 (austral summer). Due to concerns over circularity of use of Landsat data, we used both Google Earth and aerial images to select the ROIs for Landsat scenes in 1990 and 2001, ensuring that ROIs were only used where both sources showed seagrass present. We considered that the slow dynamics of seagrass patches in Tauranga Harbour (Ramage & Schiel, 1998; Schwarz & Turner, 2006) made this approach robust.

4.2.5. Development of seagrass maps and detection of change

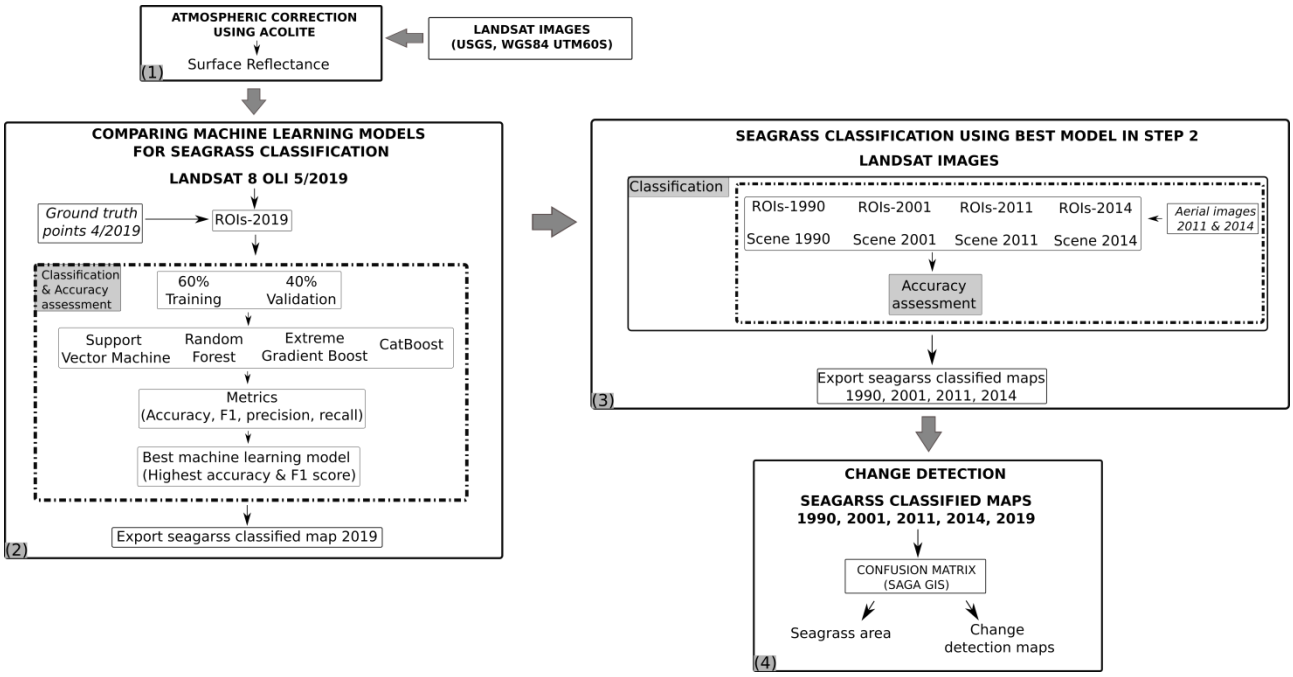


Figure 4.2. Flowchart of image processing and change detection using Landsat images in Tauranga Harbor

Our method of seagrass change detection using Landsat images involved four steps (Figure 2): (1) atmospheric correction, necessary to convert the pixel values from digital number to water surface reflectance; (2) selection of the best ML technique by comparing the accuracies of classification models for 2019 data; (3) application of the selected ML model (from step 2) for seagrass mapping to Landsat images from 1990, 2001, 2011, and 2014; (4) identifying the changes of area and spatial distribution. Due to the deficiency of field data in the past, a binary classification (seagrass and non-seagrass) was adopted to deliver the most consistent change detection.

Atmospheric correction

An atmospheric correction for all Landsat scenes was conducted using ACOLITE, in the Python™ environment (Table A1, Appendix A) (RBINS, 2018). The original pixel values in physical digital number were converted to surface reflectance. Atmospheric corrected surface reflec-

tance for pixels (limited by the study boundary, Figure 1) for the ρ_{Blue} (ρ_{w482}), the ρ_{Geen} (ρ_{w561}), and the ρ_{Red} (ρ_{w654}) bands were retained for all Landsat scenes in the years 1990, 2001, 2011, 2014, and 2019 for further processing steps. In the years 2014 and 2019, when Landsat 8 images were available, the coastal aerosol band (ρ_{w443}) was used, together with the ρ_{Blue}, ρ_{Geen}, and ρ_{Red} bands. The selected bands were used as independent variables in ML model prediction of the presence/absence of seagrass.

Due to inconsistency between the tidal status and the acquisition time of Landsat images, our study site was considered to contain both exposed and submerged areas. Therefore, the near infrared (ρ_{NIR}) band, which attenuates rapidly in water, was not used in the analysis. A water column correction was not employed for water pixels in Tauranga Harbour, since the water depth and water optical characteristics (i.e., attenuation coefficient of the solar radiance in the water column) were unavailable for the historic scenes (1990, 2001, 2011).

Application of Machine learning algorithms

Hyper-parameter tuning for selected machine learning models

Machine-learning models comprise several hyperparameters (i.e., the parameters that control the learning process during the implementation of ML models), which often need to be optimized (i.e., by the process of tuning) to find the best combination to achieve best classification performance. The hyper-parameters of the RF, the SVM, the XGB, and the CB models were tuned using a grid search with threefold cross-validation in the scikit-learn library (Pedregosa et al., 2011). The hyperparameters for each of the models were maintained during the training and the testing phases (Table A2, Appendix A).

Theoretical background of the machine learning algorithms used

Random Forest

Random Forests (RF) (Breiman, 2001) is perhaps the most popular machine-learning model for both classification and regression problems in remote sensing (Belgiu & Drăguţ, 2016). It is an ensemble bagging method, which uses a bootstrap sampling approach to build the training and the testing data and a voting method to select the most accurate decision from a large group of input decision tress. The RF model is a nonparametric method that is insensitive to the data's distribution, reducing the overfitting. The RF technique supplies various hyperparameters for tuning; however, the large number of parameters in the model results in slow optimization.

Support Vector Machine

Support Vector Machine (SVM) (Mountrakis et al., 2011; Khemchandani et al., 2015) supports linear, poly-nominal, and radial basis function (RBF) kernels and can be adapted to various linear or non-linear data types. It has relatively few tunable hyperparameters but performance speed is still relatively slow when dealing with a large dataset. The SVM model uses a hyperplane to find the separation space among the classes with the most typical rules being: (i) better segregation of data; (ii) maximization of the distance between the closest data points and the hyperplane. Despite an accurate prediction and robustness to outliers, the SVM technique is not effective on overlaid classes or noisy datasets.

Extreme Gradient Boost

Extreme Gradient Boost (XGB) (Chen & Guestrin, 2016) is different from Gradient Boosting as it uses a more regularized model, which reduces over-fitting and results in a higher prediction accuracy. In the regularized gradient boosting mode, a selection of L_1 or L_2 regularization can be made to adapt the model to suit input data. Similar to other boosting models, the XGB technique supports various hyperparameters that are tuned using a grid search or genetic algorithm (GA) (Georganos et al., 2018).

CatBoost

CatBoost (CB) was introduced in 2018 (Prokhorenkova et al., 2019) for classification, regression, and ranking tasks. It can handle both category and numerical data types. Using ordered boosting on decision trees, a permutation-driven derivation from classic boosting, the CB yields a fast and reliable performance, even with a small dataset. The model itself produces robust predictive results with default hyperparameters, reducing the requirement of tuning, and its novel gradient boosting scheme results in less overfitting.

Comparison of ML algorithms for seagrass mapping using Landsat image taken in 2019

Four ML models, SVM, RF, XGB, CB, were compared for seagrass mapping using the Landsat image from May 2019 and near-synchronous GTPs collected in April 2019 to identify the regions of interest (hereafter referred to as ROIs-2019) known to either seagrass or non-seagrass classes. The 1-month gap between the acquisition date of the Landsat image and the field survey date is acceptable due to the stable condition of the weather (i.e., no extreme weather phenomena) (NIWA, 2020), and seagrass dynamics are slow in the study site (Ramage & Schiel, 1998; Schwarz & Turner, 2006). A dataset of pixel reflectance values was extracted from ROIs-2019 and its corresponding Landsat image (dataset DS5, Table A3, Appendix A), split randomly into 60% for the training and 40% for the testing of selected ML models. The best model was selected as the model with highest accuracy and F_1 score.

Seagrass mapping using Landsat images in 1990, 2001, 2011, and 2014

The best ML model identified using the 2019 data was applied for mapping of seagrass using Landsat images from 1990, 2001, 2011, and 2014 (see Table 1 for date acquisition and spatial resolution of satellite images). The hyper-parameters developed using the 2019 data were retained for subsequent analysis, while the year-specific model was developed using ROIs

containing seagrass and non-seagrass classes from the relevant year. For the years 2011 and 2014, we created these ROIs using aerial imagery (BOPRC, 2018) (hereafter referred to as ROIs-2011 and ROIs-2014). For 1990 and 2001, we used Google Earth images cross-referenced with the aerial images acquired between 1991 and 1992 (for creating ROIs-1990) and 2003 (for creating ROIs-2001). A dataset of pixel reflectance values was extracted from corresponding Landsat images (dataset DS1, DS2, DS3, and DS4, Table A3, Appendix A) for ROIs-1990, ROIs-2001, ROIs-2011, and ROIs-2014. Datasets were split randomly into 60% for training of the classification and 40% for the accuracy assessment for 1990, 2001, 2011, and 2014.

Change detection

Change detection was conducted using the standard confusion matrix tool in the SAGA GIS (Conrad et al., 2015). The confusion matrix analyzed the changes of the pairs of classified maps (years 1990–2011 and 1990–2019), reporting in the map as seagrass loss (seagrass to non-seagrass), seagrass recovery (non-seagrass to seagrass), and unchanging seagrass.

4.2.6. Evaluation criteria

We employed standard metrics for the evaluation of the classification skill: accuracy, Kappa coefficient (κ), Kendall's tau coefficient (τ), precision, recall, and F_1 (Equations (1)–(6)). These were applied independently to the five datasets listed in Table A3, to yield the skill of the initial model based on GTPs from DS5 (2019), and to check its performance when applied to the historic Landsat data in DS1 (1990), DS2 (2001), DS3 (2011), and DS4 (2014). Kendall's tau coefficient was calculated using the SciPy library (Virtanen et al., 2020).

$$accuracy(y, y_{pred}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} 1(y_{predi} = y_i) \quad (4\text{-}1)$$

in which:

y_{pred}: predicted value

y: corresponding true value

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{4-2}$$

in which:

$p_o$is the observed agreement

$p_e$is the expected agreement

$$\tau = \frac{P - Q}{\sqrt{(P + Q + T) \times (P + Q + U)}} \tag{4-3}$$

in which:

P: the number of concordant pair

Q: the number of discordant pair

U: the number of ties in predicted value

T: the number of ties in true value

$$\text{Precision} = \frac{tp}{tp + fp} \tag{4-4}$$

$$\text{Recall} = \frac{tp}{tp + fn} \tag{4-5}$$

$$F_1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (4\text{-}6)$$

in which:

tp: true positive

fp: false positive

fn: false negative

In addition, the non-parametric McNemar test was used to compare statistically significant difference of the overall accuracy of the selected models in this research. The test was executed in a Python™ environment using the mlxtend library (Raschka, 2018). The chi-square value (χ^2) was calculated from Equation (4-5) with Edward's continuity correction.

$$\chi^2 = \frac{(|fn - fp| - 1)^2}{(fn + fp)} \quad (4\text{-}7)$$

in which:

fn: false negative

fp: false positive

4.3. Results

4.3.1. Performance of the RF, SVM, XGB, and CB models using Landsat image and GTPs for 2019 data

Of the four machine-learning models applied to the 2019 data, the CB model outperformed all others, with the F_1 score, κ and τ coefficients reaching 0.95 (Table 4.3), and 0.92 (Table A4, Supplemental Materials), respectively. The difference between models was statistically significant (McNemar's test, Table 4.4) with the exception of the XGB and RF models. The CB model required a longer computation time (3.71 s) than the RF model (0.33 s), the XGB model (0.15 s), and the SVM model (0.04 s). The RF and XGB techniques showed an equivalent performance (Table 3) with F_1 score of 0.93, while the SVM model underperformed the other models with a F_1 score of 0.91.

All models tested were able to classify seagrass from other bottom types in the harbor with a precision exceeding 0.89, but the highest precision was again from the CB model. Despite a similar F_1 score, the XGB model gained a higher precision than the RF technique.

Table 4.3. Model performance for seagrass detection in Tauranga Harbor for the 2019 dataset

Model	Accuracy	Precision	Recall	F_1	Training Time (s)	Testing time (s)
RF	0.96	0.92	0.95	0.93	0.33	0.02
CB	**0.97**	**0.94**	**0.96**	**0.95**	3.71	0.006
XGB	0.96	0.93	0.94	0.93	0.15	0.004
SVM	0.94	0.89	0.92	0.91	0.04	0.02

Bold values indicate the best performance of the model.

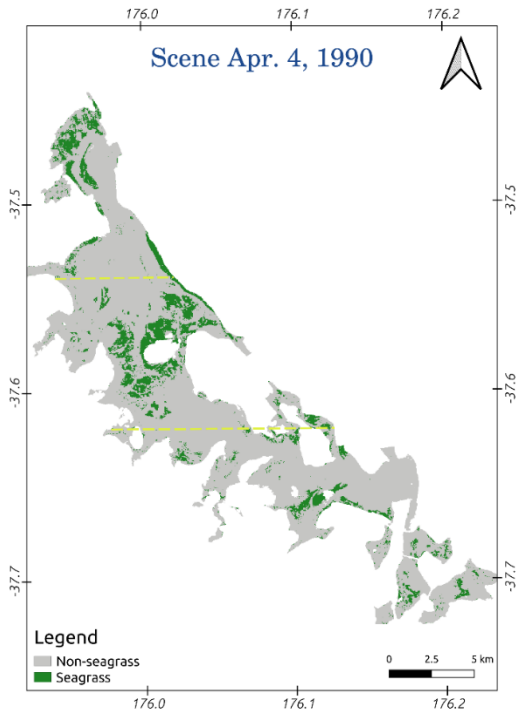
Table 4.4. Model performance comparison using the McNemar's test

	χ^2	*p*-value
CB–RF	5.88	0.01
CB–SVM	19.11	0.00
CB–XGB	4.50	0.03
XGB–RF	0.00	1.00
XGB–SVM	8.20	0.00
RF–SVM	9.25	0.00

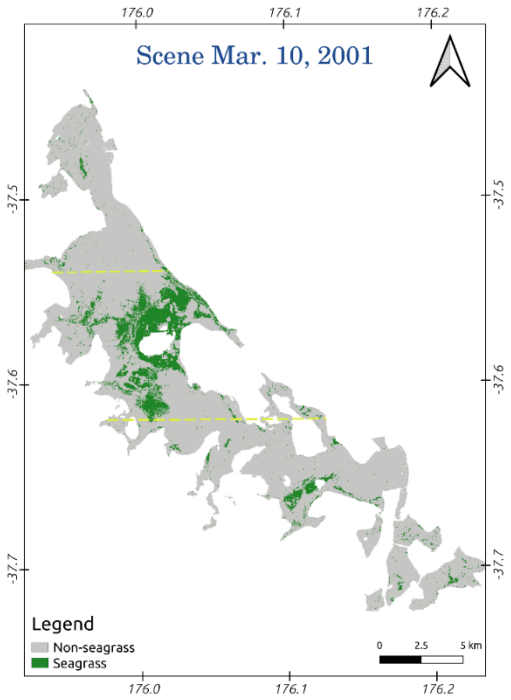
p-value < 0.05 indicates a significant difference between two models.

4.3.2. Seagrass change detection from 1990 – 2019

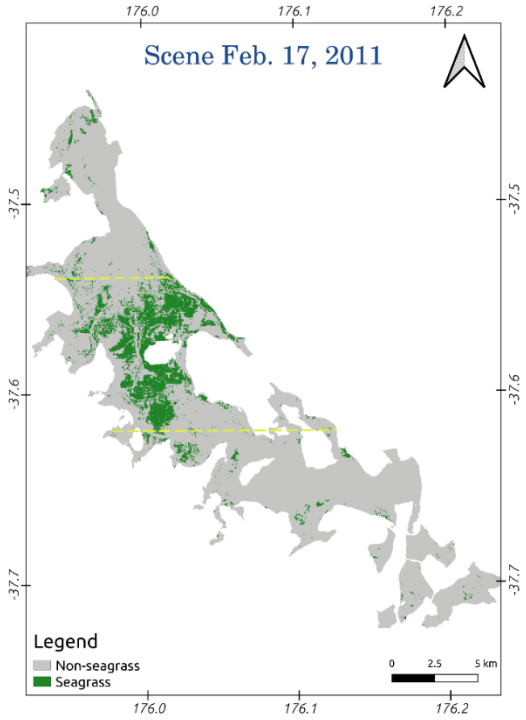
The CB technique was then used to make classification maps for the years 1990, 2001, 2011, and 2014 (Figure 4.3). Our results indicated a performance across all metrics that was equivalent to that in the 2019 case, with accuracy and F_1 scores over 95% for the binary classification of seagrass and nonseagrass (Table 4.5 and Table A4, Appendix A).


Scene Apr. 4, 1990
176.0
176.1
176.2
-37.5
-37.6
-37.7
Legend
Non-seagrass
Seagrass
0 2.5 5 km

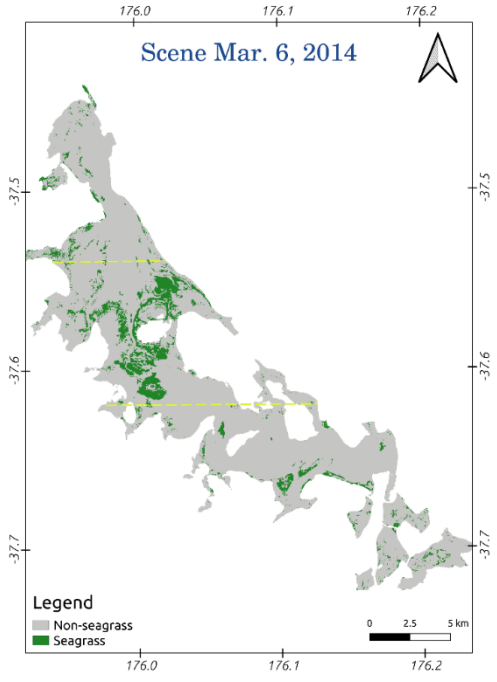
(a)


Scene Mar. 10, 2001
176.0
176.1
176.2
-37.5
-37.6
-37.7
Legend
Non-seagrass
Seagrass
0 2.5 5 km

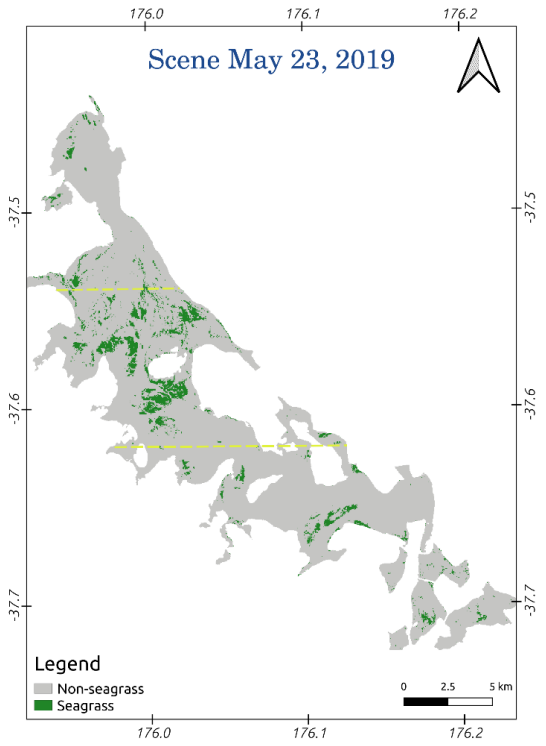
(b)


Scene Feb. 17, 2011
176.0
176.1
176.2
-37.5
-37.6
-37.7
Legend
Non-seagrass
Seagrass
0 2.5 5 km

(c)


Scene Mar. 6, 2014
176.0
176.1
176.2
-37.5
-37.6
-37.7
Legend
Non-seagrass
Seagrass
0 2.5 5 km

(d)



(e)

Figure 4.3. Seagrass distribution in the years 1990, 2001, 2011, 2014, and 2019 (a–e) using the CB model (yellow lines indicate the boundaries of the northern, central, and southern harbor).

Table 4.5. Accuracy assessment of classified map using Landsat images in 1990, 2001, 2011 and 2014

Date acquisition	Accuracy	Precision	Recall	F_1
4 April 1990	0.97	0.98	0.98	0.98
10 Mar. 2001	0.96	0.95	0.96	0.96
17 Feb. 2011	0.97	0.98	0.96	0.97
6 Mar. 2014	0.96	0.96	0.96	0.96

The time series shows that the seagrass meadow area decreased from 2237 ha in 1990 to 1184 ha in 2019, though not monotonically (Figures 4.3 and 4.4). A downward trend from 1990 (2237 ha) to 2001 (2035 ha), was followed by a recovery in 2011 (to 2380 ha), followed by a second decline to 1184 ha in 2019 (Figure 4.4a). Different trends, though all with an overall decline to 2019, were discovered in the northern (Figure 4.4b), the central (Figure 4.4c), and the southern (Figure 4.4d) harbor. Seagrass attained the largest area in the central harbor, where it reached the peak of 1985 ha in 2011; however, it declined to 776 ha in 2019. In the northern harbor, seagrass was very abundant in 1990 with 549 ha, but strongly decreased to only 92 ha in 2001. This number increased to 242 ha in 2014 before suffered a second decline to 148 ha in 2019. Seagrass loss was also recorded in the southern harbor, at a slower rate of degradation, dropping from 576 ha in 1990 to 222 ha in 2011, and around 260 ha in 2019. Across the entire harbor, the recovery in 2011 was due to a large increase of seagrass areas from 2001 in the northern and the central harbors.

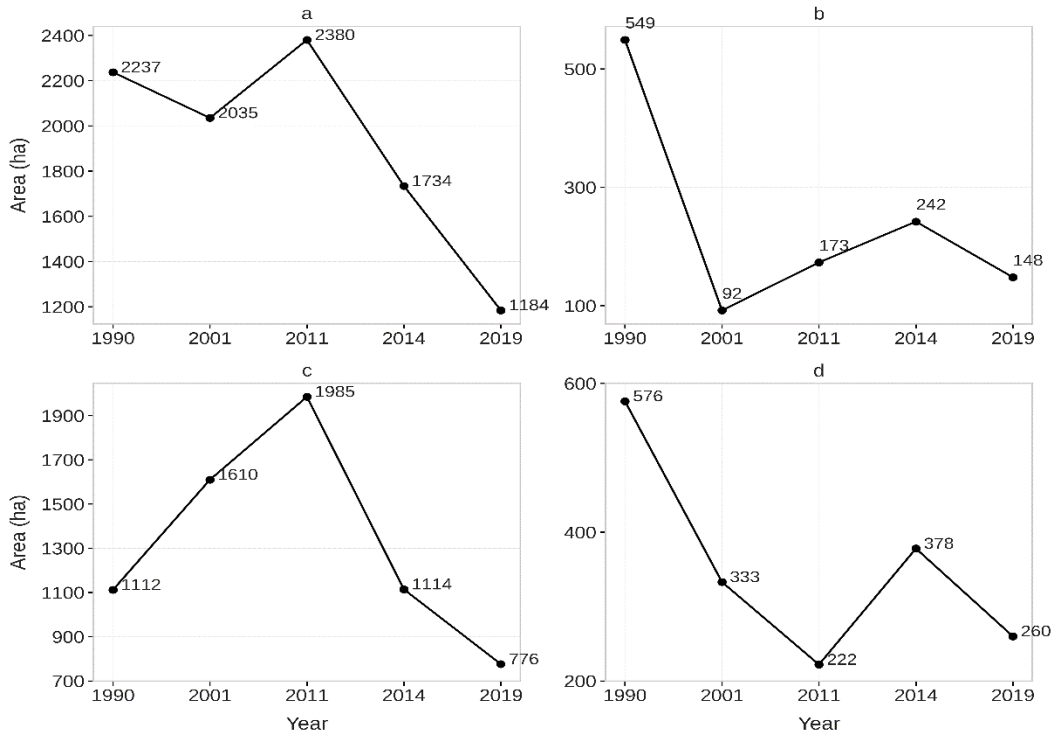
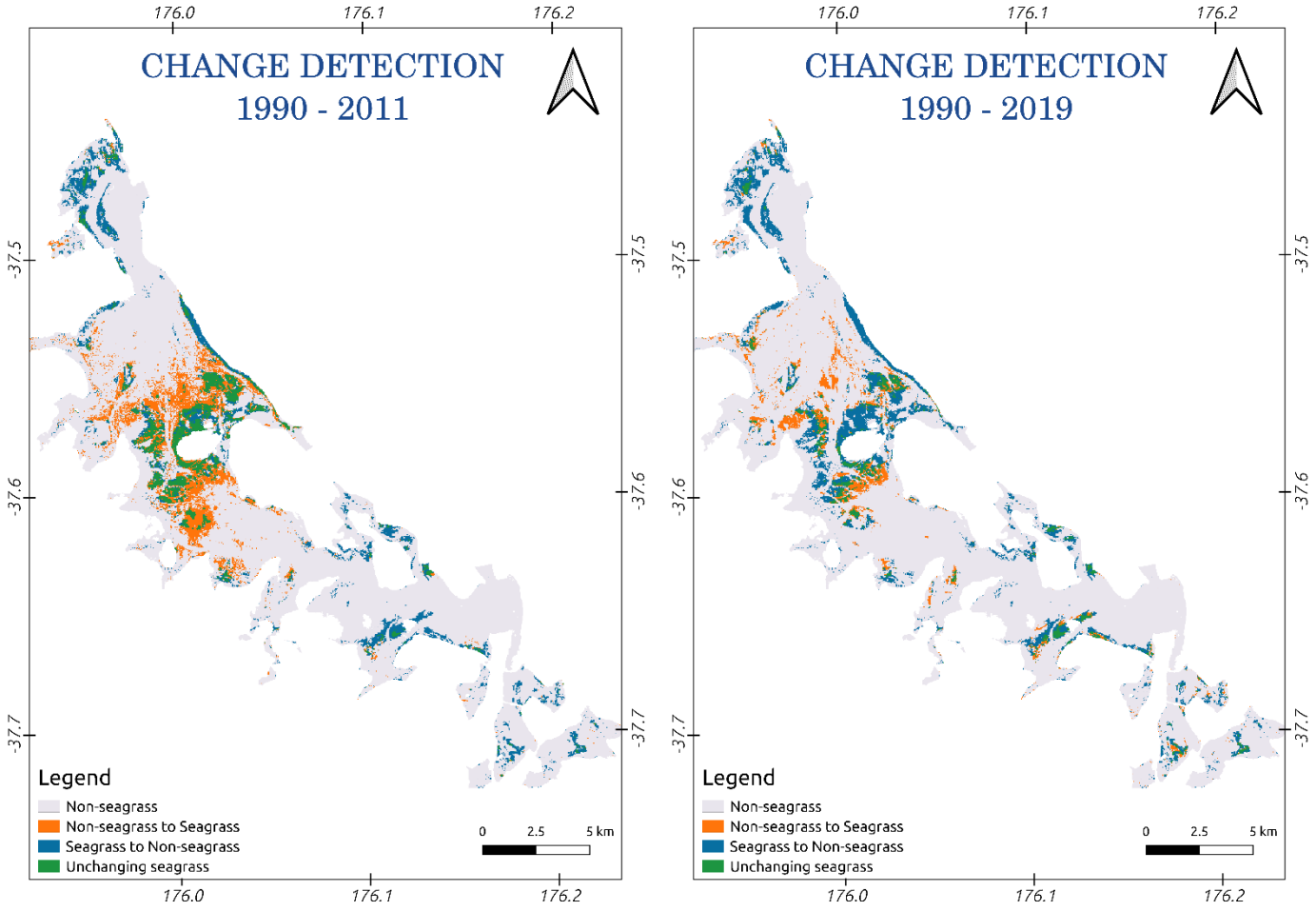


Figure 4.4. Seagrass area in Tauranga Harbour from 1990–2019 derived from Landsat imagery with the variation in: (a) entire harbor; (b) northern harbor; (c) central harbor; (d) southern harbor.

The distribution of seagrass has also changed over time. In 1990, the meadows were similarly abundant in the northern, central, and southern harbors. Declines to 2001 mostly reflected losses from the northern and southern meadows, while the central meadows remained and were responsible for most of the expansion between 2001 and 2011 (Figures 4.3 and 4.5). After 2011, there was no detectable recovery of the northern or the southern meadows, and the renewed overall decline was due to degradation of the central meadows, declining in area and becoming patchier by 2019 (Figures 4.3 and 4.5).



(a) (b)

Figure 4.5. Seagrass change detection between 1990-2011 (a) and 1990-2019 (b).

4.4. Discussion

In this investigation, we have demonstrated the use of machine-learning approaches successfully to classify seagrass in Landsat images of Tauranga Harbour, and to use this classification to detect changes in seagrass cover over a period of 29 years. Due to the relative paucity of field validation data from most of the time series of this analysis, we only tested a binary classification (seagrass and nonseagrass classes), but the four machine-learning models, RF, XGB, SVM, and CB, were all capable of detecting seagrass from other bottom types with high precision and recall

scores. Previously, the RF and the SVM models have been tested for seagrass classification (Traganos & Reinartz, 2017) and the RF for seagrass change detection in the Mediterranean (Traganos & Reinartz, 2018). These previous attempts have produced accuracies from 76–98% for *Posidonia oceania* and 32–62% for *Cymodocea nodosa* using higher resolution RapidEye imagery (Traganos & Reinartz, 2017, 2018), both lower than were achieved in this study using the CB technique. *P. oceania* and *C. nodosa* are structurally similar to *Z. muelleri* and would seem likely to offer a similar target. This suggests that the use of the state-of-the-art ML models with optimized hyper-parameters is an important factor contributing to the high-precision classification of seagrass presence/absence. Both the XGB and the CB techniques have been proven as potential candidates for a range of classification (Georganos et al., 2018; Sun et al., 2019; Pham, Yokoya, et al., 2020; Pham, Le, et al., 2020), and regression (Lou et al., 2016; Maier & Keller, 2018; Lee et al., 2019) problems but have not previously been applied to seagrasses, or to any other semisubmerged targets, so it is not clear if this is a general performance advantage in this type of application.

Other advantages over previous studies may, however, exist in Tauranga Harbour. Specifically, *Z. muelleri* occurs as monospecific meadows, without a substantial presence of macroalgae, which can degrade classifications (Park, 2011; Ha et al., 2020), and where the reflectance value of seagrass is considerably different from the other common bottom types (sand, muddy sand, deep water). In addition, we were able to use cloud-free Landsat scenes, with atmospheric correction using ACOLITE, which has been designed for the aquatic application of Landsat imagery, which likely reduced the uncertainty of atmospheric impact and derived a higher quality of corrected surface reflectance (Vanhellemont, 2019).

In this study, the two boosting techniques (XGB, CB) and the one bagging (RF) outperformed the more traditional SVM methods. The SVM model does not work well with noisy data, where unclear margins exist between classes (Karamizadeh et al., 2014). Such fuzzy margins were observed at the study site at the overlap between seagrass and nonseagrass (sand, muddy sand) classes, where the distinction between present and absent was gradual. This likely resulted in the relatively poor performance of the SVM model. The boosting techniques XGB (0.93) and CB (0.94) show slightly higher precision than the bagging RF (0.92), which might have resulted from the advancement in decision-tree growth of the boosting techniques. Unlike the RF model, which builds the independent decision tree from the bootstrapped samples, the boosting XGB and CB models sequentially grow new trees using the residual information of previous trees, which allows the new learner to solve the errors of the previous tree by minimizing the residual of the next model fitting. For a final prediction of a classification task, the bagging RF takes a majority vote from all decision trees while a weighted majority vote is adapted to the boosting techniques, such as XGB and CB, and potentially results in a higher precision of a class prediction.

Given the classification skill metrics, the CB is the best candidate for the mapping and change detection of seagrass in the study site. The CB is also amongst the latest emerging algorithms developed in the computer vision and pattern recognition fields (released in 2018); is easy to tune with fewer hyperparameters than the RF and the XGB techniques; and is using symmetric trees, which potentially results in faster optimization and prediction (Prokhorenkova et al., 2019). The CB model differs from the boosting algorithm family by using ordered boosting on a random permutation of given dataset, which prevent the prediction shift and alleviate the overfitting in model prediction. The outperformance of the CB over other ML models has been reported for

mangrove total carbon estimation (Pham et al., 2020), various testing datasets (Prokhorenkova et al., 2019), and forest aboveground biomass (Luo et al., 2021), which confirm the reliability and the capability of the CB implementation for seagrass mapping in our study. Our accurate long-term (29-year) change detection of seagrass meadows using the CB machine-learning model in Tauranga Harbour is a significant advance in the classification and monitoring of seagrass ecosystem using multispectral, remotely sensed data.

Our analysis has confirmed a general declining trend of seagrass cover in Tauranga Harbour reported previously (Park, 2011) using aerial photography. In absolute terms, Park (2011) reported 2744 ha in March 2011, close to our estimate of 2380 ha at that time. Also, like Park, our analysis was able to resolve areas within the estuary where the greatest loss has occurred between assessments. We specifically noted that the seagrass loss was initially focused in the northern and southern parts of the harbor. High flux of sediment was recorded into the northern part, due to agricultural intensification, and the southern part, due to urban development, particularly after 2011 (Hicks, 2019) and may explain the long-term decline of seagrass in those areas. The potential impact of agricultural and urban developments in the northern and the southern parts is supported by the observation that recovery was only observed in the central part of the harbor (Figure 4.3, year 2011, and Figure 4.5). Another potential factor contributing to long-term loss of seagrass is the grazing of black swans, which has previously been linked to variations in seagrass cover in the southern harbor (Park, 2011; Dos Santos et al., 2012). Further analysis is required to develop a detailed explanation on the dynamics of seagrass meadows in Tauranga Harbour.

Here, we advocate the use of novel and advanced ML models, in combination with multitemporal Landsat images to obtain a long-term, historic series of observations on seagrass dy-

namics that will continue to be supported into the future through ongoing developments of the Landsat series. The proposed method potentially provides a low-cost, high-precision classification tool that can be extended to other estuaries with similar target conditions. While aerial photography and very high spatial resolution (VHR) satellite images have higher spatial resolution than Landsat images, they come at a high cost, and spatial coverage can be limited. Currently, Landsat is the most suitable satellite image resource for any long-term change detection due to its long time in service. A 30 m spatial resolution was found suitable to support a binary classification of seagrass in this study, and accuracy was unaffected by the small changes in spectral information that have accompanied the incremental changes in Landsat optical sensors. The most recent generation Landsat 8, with an improvement in radiometric resolution up to 16 bit (in the level 1 product), compared to the 8 bit in previous generations, and the addition of a coastal aerosol band, has good potential for accurate detection of the dynamics of seagrass. An improvement of spectral and radiometric resolution in Landsat 9 (scheduled for launching in 2021) is expected to provide continuity into the future monitoring of seagrass (USGS, 2019). For a short-term observation of seagrass change, our proposed methods for seagrass classification are also potentially applicable to a wide range of VHR images (Quickbird, Ikonos, Unmanned Aerial Vehicle (UAV)) with consideration of the trade-off among the spatial coverage of the study site, the spatial resolution of the image, and the available budget.

The open-source approach is another significant advantage of our proposed methodology. The Python environment provides an excellent option for the end users to apply the novel machine-learning algorithms and remote-sensing data processing platform to support accurate mapping and estimation of the blue carbon budget of seagrass ecosystem (Macreadie et al., 2019). Most commercial software only provides a limited number of processing and classification algo-

rithms, with few, older ML options (e.g., SVM) and has a high license cost. Our proposed methods are more flexible, free of charge, and offer a high efficiency for mapping the dynamics of seagrass meadows in the complex coastal marine environment.

Despite a successful application of the CB model for seagrass classification and change detection, this research still comes with limitations. Since we used a supervised classification technique, both classification and validation require an independent assessment of seagrass cover in at least part of the remote image, to provide the ROIs that allow the training and validation steps. In addition, the seasonal growth of seagrass in temperate waters, and its intertidal habit, raise the uncertainty of change detection between various time points unless imagery is available at the same time, and under similar tidal conditions. The offset between Landsat, the time of image acquisition, and tidal regime (Table 4.1) is unavoidable in the study site; however, we consider that it is unlikely to significantly impact on classification accuracy. In Tauranga Harbour, seagrass meadows are distributed in the intertidal regions at a water depth ranging from 0 m (exposed) to a maximum of 1.5 m (at high tide) (Park, 1999, 2011). The ρ_{Blue}, ρ_{Green}, and ρ_{Red} bands have nominal maximum penetration depths of 15, 10, 5 m respectively (Green et al., 1996), and while moderate, but variable, coastal turbidity in the harbor will increase attenuation rate, the maximum immersion depth of 1.5 m suggests that the spectral bands reflectance signatures are highly likely to have been impacted by seagrass. Average vertical attenuation rate of the downwelling radiation within the 400–700 nm band in Tauranga Harbour is 0.40 m^{-1} (range 0.16–0.98 m^{-1}) (Cussioli et al., 2019) and these authors found that 65% of incident radiation reached the estuary floor at 1.2 m depth. Again, this suggests that water clarity is sufficient to ensure that, even at maximum water depth, seagrass will contribute to the reflectance spectrum detected by the satel-

lite. As with all satellite-based remote sensing, a cloud-free view is required, which constrains use of this technology.

To compensate for the limitation, we attempted to select all Landsat images acquired in the growing season of seagrass in Tauranga Harbour (austral summer and autumn) and at low tide, but this further constrains the availability of verifiable Landsat imagery for seagrass cover estimation. Further research focusing on expanding the novel approach used in the current study for long-term change detection of seagrass meadows is underway.

4.5. Conclusion

In this research, we used the novel machine-learning model CatBoost (CB) and other well-known ML models (RF, SVM, and XGB) for seagrass cover classification (present/absent), using Landsat satellite imagery, in Tauranga Harbour, New Zealand. Our results showed a high level of accuracy for all approaches, but the CB model outperformed the other selected models, with precision, recall, and F_1 scores of 0.94, 0.96, and 0.95 respectively.

We then applied the CB technique to multispectral Landsat data for the detection of change in seagrass cover over a 29-year period between 1990 and 2019 in Tauranga Harbour. The change detection analysis determined an overarching declining trend of seagrass cover in Tauranga Harbour with approximately 50% loss over the 29 years period (from 2237 ha in 1990 to 1184 ha in 2019); these results concurred with a study using aerial imaging. Seagrass was lost in the far northern and southern areas of the harbor during the first part of this time, then more gradually from the central region. This analysis of change using Landsat images combined with the CB model demonstrates the value of historic satellite imagery and machine-learning for accurate documentation of the change over time in this difficult-to-quantify coastal vegetation.

4.6. Chapter supplementary material

Table A1. Selected parameters for atmospheric correction using ACOLITE.

Parameter	Value
Ancillary data	
Gas transmittance	True
Ozone concentration (cm^{-1})	0.3
Water vapor concentration ($g\ cm^{-2}$)	1.5
Pressure	Normal pressure
Masking	
Negative reflectance masking	True
Cirrus masking	True
Other parameters	
Sky correction	True
Dark spectrum fitting	Fixed
Sun glint correction	False
Output parameter	
Surface reflectance for water pixel ()	ρ_{w443} ρ_{w482} ρ_{w561} ρ_{w654}

Table A2. The tuned hyperparameters of the RF, the SVM, the XGB, and the CB models.

Random Forest		Extreme Gradient Boost	
Bootstrap	True	Booster	GbTree
Max. depth	8	Gamma	1
Max. features	Auto	Learning rate	0.2
Min. sample leaf	1	Max. depth	5
Min. sample split	3	Min. child weight	3
Number of trees	100	Number of trees	100
Support Vector Machine		CatBoost	
Kernel	RBF	Depth	7
C	100	Iteration	200

		(Number of trees)	
Gamma	1000	Learning rate	0.2
		L2 leaf reg	1

Table A3. Number of pixels for the training and the testing sets at various acquisition dates.

Dataset	Landsat Acquisition Date	Number of Pixels	
		60% for Training	40% For testing
DS1	4 April 1990	2171	1448
DS2	10 March 2001	3000	2001
DS3	17 February 2011	2618	1746
DS4	6 March 2014	2544	1696
DS5	23 May 2019	1830	1221

Table A4. Kappa and Kendall's tau coefficients of the classification.

Model	κ	τ	p-Value of τ
Data DS5, date 23 May 2019			
RF	0.90	0.90	0.00
SVC	0.87	0.87	0.00
CB	0.92	0.92	0.00
XGB	0.90	0.90	0.00
Data DS1, date 4 April 1990			
CB	0.95	0.95	0.00
Data DS2, date 10 March 2001			
CB	0.92	0.92	0.00
Data DS2, date 17 February 2011			
CB	0.94	0.94	0.00
Data DS4, date 6 March 2014			
CB	0.93	0.93	0.00

Table A5. List of acronyms and abbreviations.

Acronym/Abbreviation	Meaning	Explanation

ACOLITE	Atmospheric correction for operational land imager (OLI) 'lite' toolbox	A Python language-based application for atmospheric correction of satellite imagery
	Accuracy	An agreement degree between the classified values and the ground-truth values in a classification task
CB	CatBoost	A machine-learning algorithm
XGB	Extreme Gradient Boost	A machine-learning algorithm
F_1	F_1	A harmonic measurement of precision and recall scores in the prediction of a machine-learning model
GPS	Global Positioning System	A satellite-based system providing positioning services
GTPs	Ground-Truth Points	GTPs are the boundary points of any given classes in the study site, defined by GPS
K	Kappa coefficient	A statistical index measuring the accuracy (agreement between predictions and ground-truthed values) of the classification. A higher Kappa coefficient denotes a more accurate classification
τ	Kendall's tau coefficient	A nonparametric measurement to evaluate the classification's accuracy. A higher Kendall's tau coefficient denotes a more accurate classification
ML	Machine learning	An artificial intelligence (AI) approach that builds an application/algorithm for a specific output by learning from data
NIR	Near infrared	The near infrared region in the electromagnetic spectrum

	Precision	A score to measure the success of the prediction of a machine-learning model. A higher precision denotes a more accurate prediction
RF	Random Forest	A machine-learning algorithm
	Recall	A score to measure the success of the prediction of a machine-learning model. A higher recall denotes a more accurate prediction
RBF	Radial basis function	A function used in the Support Vector Machine model, together with linear and polynomial functions
ROI	Region of interest	A bounded region used in image classification where the pixels contain a given class
SVM	Support Vector Machine	A machine-learning algorithm
UAV	Unmanned aerial vehicle	An aircraft without a human pilot
GLOVIS	USGS Global Visualization Viewer	A web-based system for satellite image visualization and downloading
UTM	Universal Transverse Mercator	A map projection
VHR	Very high spatial resolution	Indicating satellite images that have spatial resolution from centimeters to a few meters
WGS	World Geodetic System	A standard coordinate system used in cartography.

Chapter 5 The use of radar and optical satellite imagery combined with advanced machine learning and metaheuristic optimization techniques to detect and quantify above ground biomass of intertidal seagrass in a New Zealand estuary

The contents of this chapter have been published in the International Journal of Remote Sensing (http://dx.doi.org/10.1080/01431161.2021.1899335). A copy of the published paper is bound into Appendix 6.

Abstract: Seagrass provides numerous valuable ecosystem services across a wide range of climatic regions. However, in terms of area and habitat, this resource is in decline globally and there is an urgent need for accurate mapping of extant meadows and biomass to support sustainable seagrass blue carbon conservation and management. This study develops a novel method for a binary mapping of seagrass distribution and estimating seagrass aboveground biomass (AGB) by applying a suite of advanced machine learning (ML) algorithms combined with and without a metaheuristic optimization approach (Particle Swarm Optimization - PSO) to various combinations of multispectral (Sentinel-2) and Synthetic Aperture Radar (Sentinel-1) remote sensing data. Our results reveal that the Sentinel-1 data has potential for the binary mapping of seagrass meadows using an extreme gradient boosting (XGB) model (scores of precision (P) = 0.82, recall (R) = 0.90, and F_1 = 0.86) but is less effective at estimating AGB. The optimal method for estimation of AGB used both Sentinel-1 and Sentinel-2 imagery, the XGB model and PSO optimization (coefficient of determination (R^2) = 0.75, root mean squared error (RMSE) = 0.35, Akaike information criteria (AIC) = 24.80, Bayesian information criteria (BIC) = 44.70). Our findings contribute novel and advanced methods for seagrass detection and improvement of AGB estimation, which are fast and reliable, use open source data and software and should be easily applicable to intertidal zones across many regions of the world.

Keywords: seagrass, aboveground biomass, metaheuristic, Particle Swarm Optimization, Extreme Gradient Boost, Sentinel-2

5.1. Introduction

Seagrasses are a group of marine angiosperms, which survive in a wide range of environmental conditions and are widely distributed in coastal zones. Several valuable ecosystem services are recognized for this resource (Nordlund et al., 2016), including wave attenuation (Fonseca & Cahalan, 1992; Reidenbach & Thomas, 2018), pollution attenuation (Short & Short, 1984), nursery and breeding habitat (Heck Hay et al., 2003) and, most recently, carbon sequestration (Bedulli et al., 2020). Recognition of the importance of seagrass systems as a blue carbon resource has resulted in initiatives to foster a long-term strategy of reversing the current decline in seagrass area/biomass (Waycott et al., 2009), and carbon sequestration (Githaiga et al., 2019; Aoki et al., 2020) world-wide as a nature-based solution in averting climate change (Macreadie et al., 2019). To facilitate this, a precise method for mapping the distribution and biomass in seagrass ecosystems, under different environmental conditions (Unsworth et al., 2018), is needed to support its inclusion in the carbon credits market (Macreadie et al., 2019; Ricart et al., 2020).

Recently, mapping seagrass distribution and aboveground biomass (AGB) has been investigated either using intensive field surveys or data-driven models developed from Earth observation (EO) data and statistical techniques (Roelfsema et al., 2015; Sani et al., 2019). The former approach is highly accurate, using *in situ* measurements for each sampling plot. However, it is labor intensive, is time-consuming, and cannot be readily upscaled (Unsworth et al., 2018). In contrast, the use of optical remote sensing data for mapping seagrass ecosystems allows low cost, large area spatial mapping, and fast performance (Hossain et al., 2015; Dat Pham et al., 2019; Veettil et al., 2020; Ha et al., 2020). EO-based techniques, however, have unresolved challenges of coarse spatial resolution, insufficient accuracy for retrieval models, cloud cover constraints, and the effect of immersion on retrieval algorithms in intertidal environments.

For seagrass distribution mapping, researchers have developed various techniques for image pre-processing and classification (Hossain et al., 2015; Dat Pham et al., 2019). Recent studies have used the multispectral Land Remote-Sensing Satellite (System, Landsat) 5 Thematic Map (TM), Landsat-8 Operational Land Imager (OLI) (Misbari & Hashim, 2016), Landsat-5 TM, and Hyperion (Pu & Bell, 2013), or very high spatial resolution (VHR) optical data, including Quickbird, Ikonos, Worldview-2 (Roelfsema et al., 2014; Koedsin et al., 2016) for the retrievals of seagrass AGB using linear regression models. The coefficient of determination (R^2) varied between studies, with higher values, ranging from 0.34 to 0.68, observed when using VHR images.

Despite the obvious advantages of using spectral bands for seagrass (species) detection, these EO techniques are constrained by the need for scene to be captured at suitable time during daylight and free of cloud. The high cost of VHR images adds to the constraints of analyzing large areas. Therefore, the development of novel approaches using other EO data that escape some of these constraints (*i.e* synthetic aperture radar (SAR) images that can be obtained regardless of time or weather conditions during overpass) may better suit the requirement of rapid assessment of seagrass change globally. To the date of this study, we found no application of SAR data for seagrass meadows mapping of presence-absence, cover, or AGB.

The current situation regarding the broad scale estimation of the contribution of seagrass to blue carbon storage, is that optical satellite data can be used for seagrass detection and estimation of AGB, however with a variable and generally low precision, and with the constraint of requiring an absence of cloud cover, suitable ambient lighting and low tide to coincide with satellite overpass. In a previous study that used Landsat imagery to address 40 year trends in seagrass cover in Tauranga Harbor (New Zealand), we found this to seriously constrain usable observations. To address this limitation we elected to assess the potential application of SAR remote sensing to

map seagrass distribution and to estimate seagrass biomass, due to its freedom from the need for cloud-free observing conditions with the sun in a suitable position. We also investigate the use of fusions of multi-spectral and SAR remotely sensed images, to improve accuracy of models for biophysical parameters estimation of seagrass ecosystems in the coastal areas.

Given a list of available satellite resources, we selected Sentinel-1 (S-1) and Sentinel-2 (S-2), which were launched in the years 2014 and 2015, are operated by the European Space Agency (ESA), and provide freely accessible remote sensing data (ESA-S1, 2020; ESA-S2, 2020). S-1 provides dual-polarimetric SAR imagery, acquired day and night at C-band (dual-polarization vertical transmitting and horizontal receiving (VH) and vertical transmitting and vertical receiving (VV)) in any weather conditions, at a wide swath of 250 km and a spatial resolution of 20 m (level 1 production). A favourable S-1 image therefore only requires that the overpass is at low tide. S-2 provides 12 multispectral bands ranging from 400 to 2400 nm and spatial resolutions ranging from 10 to 60 m. For coastal environments, the blue, and the green wavelengths are expected to have a good penetration through the water environment, whilst the red and the three red-edge bands of S-2 produce a range of vegetation and soil indices, which might be useful for the estimation of biophysical parameters such as AGB retrievals (Pham, Le et al., 2020; Pham, Yokoya et al., 2020). S-1 and S-2 have repeating cycles of 12 and 5 days, respectively. In this work, we hypothesize that the S-1 C-band sensor may be suited to estimate the meadow structure despite the low canopy height of seagrass ecosystems (El Hajj et al., 2018; Howell et al., 2018) whereas the S-2 optical sensor can be useful to extract spectral information that will assist in the estimation of seagrass AGB.

To extract information from these Earth observing sensors we elected to use machine learning (ML) models, which have been developing rapidly in recent years. The current literature reveals

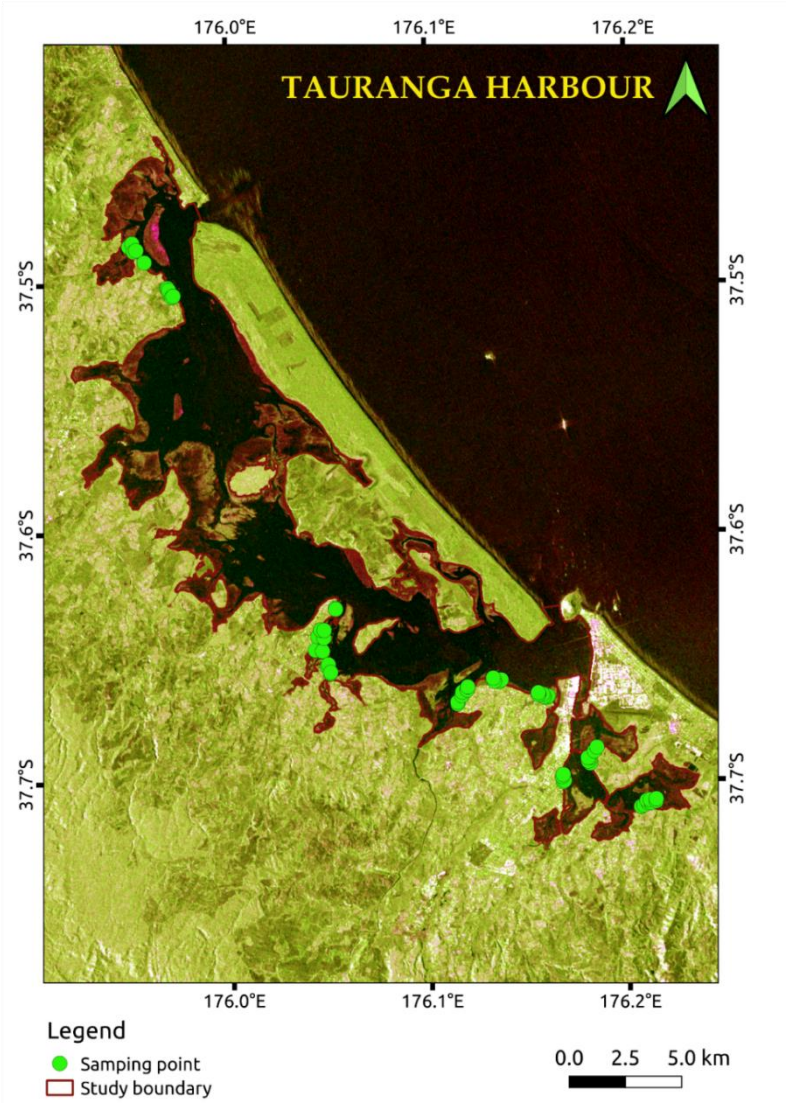
a limited application of ML models for the retrieval of biophysical parameters of seagrass ecosystem, but successful implementation in various domains (Huettmann, 2018; Thessen, 2016). Currently, the ensemble decision tree-based approaches of bagging and boosting algorithms that have been applied in numerous field of studies include the Random Forest (RF) (Belgiu & Drăguţ, 2016), Rotation Forest (RoF) (Ha et al., 2020; Rodriguez et al., 2006), Canonical Correlation Forest (CCF) (Rainforth & Wood, 2015), Support Vector Machine (SVM) (Mountrakis et al., 2011), Gradient Boost (GB) (Natekin & Knoll, 2013), Extreme Gradient Boost (XGB) (Chen & Guestrin, 2016), CatBoost (CB) (Prokhorenkova et al., 2019), and Light Gradient Boosting Machine (LGBM) (Ke et al., 2017). In general, they produce better predictive performance, faster speed and often outperform parametric models but, to our knowledge, have seldom been used for seagrass mapping and have never been used to retrieve seagrass AGB. Therefore, in this work we attempted to develop novel approaches for first the binary (presence/absence) mapping of seagrass meadows using S-1 images. For retrieving seagrass AGB, we investigated the use of various combinations of S-1 and S-2 images and a range of advanced ML techniques. Our novel, integrated approach is expected to strengthen the utilization of multisource remote sensing data in accurately detecting the distribution and predicting the AGB of extant seagrass meadows and to support their integration into the blue carbon strategy to deal with climate change impacts.

5.2. Materials and Methods

5.2.1. Study site

Tauranga Harbor (New Zealand) (Figure 5.1 (*a*) - (*b*)) was the study site; at this site, seagrass meadows are patchy and distributed mainly in the center and southern parts of the harbor (Ha et al., 2020). The tide regime is semi-diurnal with a range from 0.20 to 2.10 m. The only species of seagrass in the harbor, as in the rest of New Zealand, is *Zostera muelleri*, which is distributed in

the inter-tidal zone of the harbor (Park, 1999). Compared to other seagrasses, *Z. muelleri* is small (0.10 to 0.40 cm in leaf width), with optimal growth at 27 to 30 °C (Collier et al., 2017; York et al., 2013) and 12 practical salinity units (psu) (Collier et al., 2014). *Z. muelleri* biomass shows a seasonal rhythm (Turner, 2007), with the highest biomass in the austral summer (December to March) and declining during the austral winter to minimum biomass in early September.



(*a*)

(*b*)

Figure 5.1. Sentinel imagery of Tauranga Harbor (New Zealand) and ground truth points used in the investigation. (*a*) S-1 image (acquisition date 31 March 2020) used for seagrass binary mapping. (*b*) S-2 image (pseudo color using a composition of bands ρ_{Red} - ρ_{Green} - ρ_{Blue}, acquisition date 5 April 2020). The red line indicates the mask used in analysis of remote imagery

5.2.2. Field survey

Seagrass collection

The field survey was conducted during the austral summer (March 2020) season in the inter-tidal areas of the harbor, when biomass would be approaching the seasonal maximum. 57 sampling

plots (size 10 m × 10 m) across the harbor were selected (Figure 5.1 (*a*) – (*b*)). The geographic position of each plot was recorded using a global positioning system (GPS) Extrex-30 with an accuracy of ± 2 m and used as the ground truth points (GTPs) for the seagrass binary mapping using the S-1 image.

A 0.5 m × 0.5 m square quadrat and a plastic core were used to collect seagrass for the measurement of the AGB as dry weight. The plastic core was used (Howard et al., 2014; Lyons et al., 2015) within each plot to take three cores, each 15 cm in diameter and 40 cm in depth in representative locations (Figure 5.2 (*a*) – (*b*)). The cores were inserted to a depth of 20 cm and the seagrass samples were washed free of sediment at the field sites, then kept at – 10 °C until analysis in the laboratory.

Sample processing and dry AGB measurement

In the laboratory, the seagrass sample was divided into aboveground and below-ground parts. The aboveground part consisted of seagrass shoots whilst the below-ground part comprised the rhizome and the root. We observed no epiphytes on the leaves of the seagrass samples. The aboveground part of seagrass was dried at 60 °C for 48 hours, cooled in a desiccator and weighed to ± 1 mg. An average AGB, normalized to the unit of g dry weight m^{-2} (g DW m^{-2}), for each site was used in the production of the AGB distribution map.

(*a*)

(*b*)

Figure 5.2. Seagrass samples collection in Tauranga Harbor using (*a*) quadrat and (*b*) plastic core. Photos were taken by N.T.H in March 2020

5.2.3. Development of models

Satellite image acquisition

S-1 and S-2 data were downloaded from the Copernicus open access hub (ESA 2020) and the United States geological survey (USGS) global visualization viewer (USGS-Glovis) (USGS 2020), respectively in the projection of World Geodetic System (WGS84) (S-1) and WGS84-Universal Transverse Mercator (UTM), zone 60 South (60S) (S-2) (Table 5.1). The images were acquired to have the acquisition time closest to each other and the low tide in Tauranga Harbor, and as close as possible to the field data collection dates (4 to 25 March 2020). Despite the close coincidence of low tide and overpass, comparison with bathymetry suggests that both submerged and exposed intertidal areas in the harbor were present at the time of satellite image acquisition (Reeve et al., 2018).

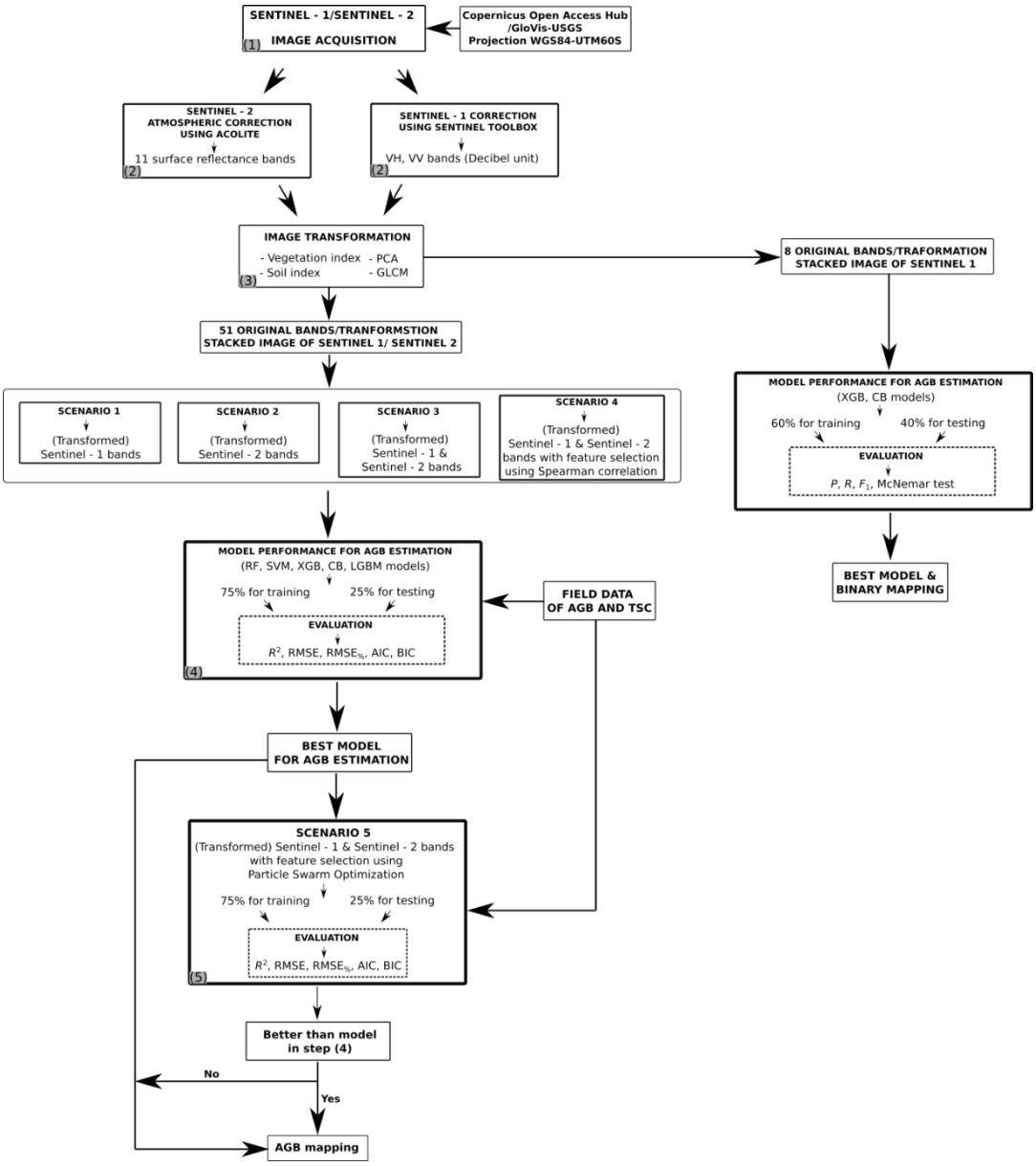


Figure 5.3. Flow chart of image processing, retrieval modeling (steps (1) to (5)) for seagrass AGB estimation and binary mapping

Table 5.1. S-1 and S-2 data acquisition details

Sensor	Processing level	Band used	Date of acquisition	Time of acquisition	Corresponding low tide
S-1	Ground range detected (GRD) level 1 sensor	Dual-polarimetric VH and VV	31 March 2020	19:07	19:02

	mode: interferometric wide swath (IW)				
S-2	Level 1C	Band 1 to 8A vegetation (8A-V), 11, 12	5 April 2020	10:15	10:17

Sentinel-1 preprocessing and image transformation

The preprocessing of the S-1 image consisted of multiple steps to convert the pixel values from the raw intensity signal to the backscattering coefficients (σ^0) in decibel units (dB), which were used for seagrass detection and retrieval of AGB. The S-1 C-band intensities of the VH and the VV polarizations were processed to the normalized radar σ^0 as suggested by (Filipponi, 2019). This involved the following steps: (1) Correct the orbit file; (2) Thermal noise removal; (3) Border noise removal; (4) Radiometric calibration; (5) Speckle filtering; (6) Range Doppler terrain correction: before (7) conversion of the pixel values to the normalized radar backscattering coefficient (σ^0) using Equation (5-1).

$$\sigma^0 = 10 \times \log_{10}(\mathrm{DN}^2) \quad (5\text{-}1)$$

in which DN is the digital number of the intensity image.

The band transformation for S-1 VH and VV polarizations was conducted as follows: (1) band combinations were derived as VV / VH, VH / VV, VV − VH, VH − VV, (VV + VH) / 2; (2) applying the gray level co-occurrence matrix (GLCM) feature extraction to create twenty new bands (VH contrast (VH_cons), VH dissimilarity (VH_diss), VH homogeneity (VH_homo), VH

asm (VH_asm), VH energy (VH_energy), VH max (VH_max), VH entropy (VH_entropy), VH mean (VH_mean), VH variance (VH_variance), VH correlation (VH_corr), VV contrast (VV_cons), VV dissimilarity (VV_diss), VV homogeneity (VV_homo), VV asm (VV_asm), VV energy (VV_energy), VV max (VV_max), VV entropy (VV_entropy), VV mean (VV_mean), VV variance (VV_variance), VV correlation (VV_corr)); (3) extracting the first principal component (PCA1) from principal component analyses (PCA) of firstly the seven bands developed in step 1 (VV, VH, VV / VH, VH / VV, VV − VH, VH − VV, (VV + VH) / 2)) and secondly the 20 GLCM bands developed in step (2), resulting in the PCA1_7_band and PCA1_GLCM. The GLCM and PCA features were both extracted using the Sentinel application platform (SNAP) toolbox (ESA, 2020). A total of 29 bands (2 dual polarizations and 27 transformed bands) were involved as the input variables/features from the S-1 analysis for use in the estimations of seagrass AGB whilst a total of 8 bands (2 dual polarizations and 6 transformation bands (VV / VH, VH / VV, VV − VH, VH − VV, (VV + VH) / 2), PCA1_7_band) were used for the binary mapping in this study (see Table S5, the supplemental material) for the abbreviations and full names of the transformations).

All the steps were accomplished in the SNAP application using the SAR toolset (ESA, 2020). Finally, we resampled to a spatial resolution of 10 m and co-registered the projection to WGS84 UTM 60S to match S-2 imagery.

Sentinel-2 image atmospheric correction and transformation

The level 1C S-2 is downloaded as the top of atmosphere (TOA) processing level and was converted to surface reflectance using the atmospheric correction for operational land imager (OLI) “lite” toolbox (ACOLITE) (Vanhellemont, 2016) with the dark spectrum algorithm (Vanhellemont, 2019) (Table 5.S1, the supplemental material).

For the retrieval of AGB, the literature indicates the usefulness of different vegetation indices (VIs) and soil radiometric indices (SIs) (Xue & Su, 2017; Pham, Yokoya et al., 2020). Therefore, we computed a number of band transformations to generate different VIs and SIs (Table 5.2). A total of 22 bands (11 multispectral bands, 11 bands of VIs and SIs) were computed as the input variables for the retrievals of AGB from the S-2 scene (see Table 5.S5, the supplemental material for the abbreviations and full names of the transformations).

Seagrass mapping using Sentinel-1

Our previous work on seagrass binary mapping using Landsat images in Tauranga Harbor indicated that XGB and CB outperformed other ML models (Thang et al. in review, chapter 4). Therefore, we evaluated the classification skills of XGB and CB models for the presence/absence mapping of seagrass meadows using only S-1 images here. Using the GTPs from the field survey (March 2020), two classes of regions of interest (ROIs) (seagrass and non-seagrass) were created with a total of 5435 pixels selected from S-1 images (Table 5.3). The dataset was randomly split into 60% for the training and 40% for the testing of the selected ML models. The best model with highest scores of precision (P) and F1 was used to produce the binary map of seagrass meadows in Tauranga Harbor.

Table 5.2. Vegetation and soil index transformed from S-2 bands. Band wavelengths of S-2: ρ_{Blue} (460 nm), ρ_{Green} (560 nm), ρ_{Red} (665 nm), Red-edge 1 (RE_1 (704 nm)), Red-edge 2 (RE_2 (740 nm)), Red-edge 3 (RE_3 (783 nm)), Near-infrared (ρ_{NIR}) (833 nm)

Vegetation index	Acronym	S-2 band wavelength used	Reference
Ratio vegetation index	RVI	$\frac{\rho_{\text{NIR}}}{\rho_{\text{Red}}}$	(Tucker, 1979)
Normalized difference vegetation index	NDVI	$\frac{\rho_{\text{NIR}} - \rho_{\text{Red}}}{\rho_{\text{NIR}} + \rho_{\text{Red}}}$	(Rouse et al., 1974)
Green normalized difference vegetation index	GNDVI	$\frac{\rho_{\text{NIR}} - \rho_{\text{Green}}}{\rho_{\text{NIR}} + \rho_{\text{Green}}}$	(Gitelson et al., 1996)
Enhanced vegetation index 2	EVI2	$2.5 \times \frac{\rho_{\text{NIR}} - \rho_{\text{Red}}}{\rho_{\text{NIR}} + 2.4 \times \rho_{\text{Red}} + 1}$	(Jiang et al., 2008)
Normalized difference index using bands 4 & 5 of S-2	NDI45	$\frac{\text{RE}_1 - \rho_{\text{Red}}}{\text{RE}_1 + \rho_{\text{Red}}}$	(Delegido et al., 2011)
Soil-adjusted vegetation index	SAVI	$(1 + L) \times \frac{\rho_{\text{NIR}} - \rho_{\text{Red}}}{\rho_{\text{NIR}} + \rho_{\text{Red}} + L}$ L = 0.5 in most conditions	(Huete, 1988)
Inverted red-edge chlorophyll index	IRECI	$\frac{\text{RE}_3 - \rho_{\text{Red}}}{\frac{\text{RE}_1}{\text{RE}_2}}$	(Frampton et al., 2013)
Modified chlorophyll absorption in reflectance index	MCARI	$[(\text{RE}_1 - \rho_{\text{Red}}) - 0.2 \times (\text{RE}_1 - \rho_{\text{Green}})] \times (\text{RE}_1 - \rho_{\text{Red}})$	(Daughtry et al., 2000)

Soil Index	Acronym	S-2 Band Wavelengths Used	Reference
Brightness index	BI	$\sqrt{\frac{\rho_{\text{Blue}}{}^2 + \rho_{\text{Green}}{}^2 + \rho_{\text{Red}}{}^2}{3}}$	(Escadafal, 1989)
Redness index	RI	$\frac{\rho_{\text{Red}}{}^2}{\rho_{\text{Blue}} \times \rho_{\text{Green}}{}^3}$	(Mathieu et al., 1998)
Colour index	CI	$\frac{\rho_{\text{Red}} - \rho_{\text{Green}}}{\rho_{\text{Red}} + \rho_{\text{Green}}}$	(Mathieu et al., 1998)

Table 5.3. Number of input bands and field data observations used for various scenarios

		For AGB estimation		For binary mapping	
	Sensor	No. of input band	No. of observation	No. of input band	No. of training/testing pixel
	S1 + SAR transformation			8	3261/2174
Scenario 1	S-1 + SAR transformation	29	57		
Scenario 2	S-2 + VIs + SIs	22	57		
Scenario 3	S-1 + SAR transformation + S-2 + VIs + SIs	51	57		
Scenario 4	S-1 + SAR transformation + S-2 + VIs + SIs and Spearman feature selection	47	57		
Scenario 5	S-1 + SAR transformation + S-2 + VIs + SIs and PSO feature selection	26	57		

Seagrass AGB estimation

We examined all available integrations of S-1 and S-2 data using five scenarios, and in each the performance of various ML models for AGB estimation was evaluated (Table 5.3). Scenarios 1

and 2 used only the original bands and transformations of S-1 and S-2 images respectively. Scenario 3 used both the original bands and transformations of S-1 and S-2 images and Scenario 4 used a band subset derived from a Spearman correlation analysis between each input band and AGB. In this case, input bands with a Spearman coefficient > 0.10 were selected to include in scenario 4. In scenario 5, the best retrieval model from previous scenarios was integrated with a Particle Swarm Optimization (PSO) procedure designed to select the best combination of bands (see section 2.3.7).

Retrieval of AGB using ML models

Random Forest (RF)

To date, the RF algorithm (Breiman, 2001) would be the most popular ML model applied for a wide range of classification and regression problems (Belgiu & Drăguţ, 2016; Pal, 2005). The RF model is reliable, with bootstrap sampling and ensemble bagging trees used to derive the most accurate decision from an ensemble of the weak learners. During the learning, approximately 2/3 of the samples (in-bag) is used for the training phase and 1/3 of the samples (out-of-bag) is used for the testing phase. Similar to other tree-based and boosting ML models, the RF model estimates the importance of each variable as predictors in the model. The most notable hyper-parameters of the RF model involve the maximum depth, minimum sample leaf, minimum sample split, maximum features, and number of trees that can be tuned to fit to a specific dataset.

Support Vector Machine (SVM)

SVM is a well-known algorithm for supervised learning applied for various non-linear problems (Gholami & Fakhari, 2017; Mountrakis et al., 2011). Given a classification or regression problem, the SVM algorithm defines the hyperplanes with the support vectors, the closest points

to the hyperplanes, and the margin, the gap between the classified classes. Compared to other ML models, the SVM model consists of fewer hyper-parameters of the kernel function (linear, poly-nominal or radial basis), gamma (γ) controlling the overfitting, and regularization parameter (C). The SVM model is able to deal with non-linear data, however not efficiently when the data is noisy or comprises overlapped data classes (Fehr et al., 2008; Mountrakis et al., 2011).

Extreme Gradient Boost (XGB)

Developed by Chen & Guestrin (2016), the XGB algorithm shares the same theory of boosting technique with the gradient boosting family, however it uses a more regularized model to produce scalable, and accurate prediction. The XGB model computes the second-order gradients of the loss function which provide essential information on minimizing the loss function. The introduction of L_1 and L_2 regularization models improves the generalization, and therefore reduces the overfitting. Parallelization, out-of-core computation and cache optimization are further advantages of the XGB, which reduces the training time when dealing with a big dataset. In addition, the XGB algorithm can be wrapped inside various optimization algorithms to tune the hyper-parameters for specific problems. The most important hyper-parameters involve the booster, maximum depth, minimum child weight, number of trees and learning rate.

Light Gradient Boosting Machine (LGBM)

The LGBM algorithm is designed for big data processing, but is still applicable for small-scale datasets (Ke et al., 2017). The LGBM technique inherits the advantages of decision tree-based and gradient boosting algorithms, in addition to a fast training speed with parallel computation, lower memory usage, and is capable of large-scale data processing. The LGBM grows the decision tree using the leaf-wise mode rather than the level-wise mode used in other tree-based ML models, which results in higher prediction accuracy, however, it may also lead to excessive

tree complexity and overfitting. The most important hyper-parameters of the LGBM model include the number of trees, number of leaves, maximum depth, which may reduce the model overfitting, and the gradient boosting model such as dropouts meet multiple additive regression tree (DART), gradient-based one-side sampling (GOSS), and gradient boosting decision tree (GBDT). On the other hand, the LGBM algorithm performs well with high dimensional data (Li et al., 2018), which might be appropriate to the multiple bands data in this study.

CatBoost (CB)

Of the ML algorithms used in this study, the CB algorithm is the most recent addition to ML techniques, that have been used for an extensive range of classification, regression, and ranking tasks (Prokhorenkova et al., 2019). The CB algorithm is able to handle both category and numerical data. The CB model introduced ordered boosting, which was modified from the standard gradient boosting algorithm and helps to avoid "shifting" in prediction due to target leakage. During the algorithm's learning, the CB algorithm uses the oblivious decision tree instead of the asymmetric tree used in the XGB and the LGBM algorithms. The CB technique supports graphics processing unit (GPU) processing, significantly reduces training time and is easy to apply for a user-friendly application programming interface (API). The CB algorithm is robust as it uses only a few defaults hyper-parameters such as depth, number of trees, and learning rate.

Hyper-parameter tuning for selected ML models

In this study, we employed a five-fold cross-validation (CV) using the GridSearch in the Scikit-learn library (Pedregosa et al., 2011a) to find the best combination of each model's hyper-parameters (See Table 5.S2, the supplemental material for seagrass binary mapping and Table 5.S3 (*a*) - (*d*), the supplemental material for seagrass AGB estimation).

Particle Swarm Optimization (PSO) application for feature selection

Introduction to PSO

PSO is a powerful metaheuristic search algorithm for optimization of solutions to non-linear functions introduced by Kennedy & Eberhart (1995). Implementing a swarm-based stochastic optimization, the PSO uses the knowledge gained by each of the swarm of particles to affect the behavior of other particles in a search space of n dimensions. Each particle is controlled by three parameters inertia weight (w), cognitive (c_1) and social (c_2) components. The particles then implement the searches in n dimensional space, which correspond to the input features, returning the best positions in the search space with the minimum fitness during the optimization of feature selection (Sengupta et al., 2018; Seixas Gomes de Almeida & Coppo Leite, 2019). Compared to the Genetic algorithm (GA), PSO shares the idea of updating the particle positions after the iterations, however it requires a less complex model structure and is thus capable of being implemented with fewer hyper-parameters (Lambert-Torres et al., 2009).

PSO implementation

The PSO algorithm was separately employed for estimating seagrass AGB using the PySwarms library (Tisimst, 2020) in the Python™ environment. The cost function in the PSO was defined using the best retrieval model for retrieving AGB in step 4 (Figure 5.3). At the end of the optimization, the positions corresponding to the lowest root mean squared error (RMSE) was selected and tested for the model performance in scenario 5 (step 5, Figure 5.3).

PSO parameter tuning for AGB estimation

PSO consists of different hyper-parameters, which should be tuned to derive the best optimization. In this work, we used GridSearch in the PySwarms library (Tisimst, 2020) to find

the best combination of number of particles, w, c_1, and c_2 parameters (Table 5.S4, the supplemental material). The search algorithm was fixed to global search.

Evaluation criteria

To evaluate the model performance for the binary classification, we used the accuracy (A), precision (P), recall (R), and F_1 scores to measure the performance (Equations (5-2) to (5-5)) whilst the McNemar test was used to check the statistical difference from the two ML models (the chi-square (χ^2) and p-value were produced using the Python library mlxtend (Raschka, 2018)).

For seagrass AGB retrieval, we used the following metrics to evaluate the model performance: the coefficient of determination (R^2), root mean square error (RMSE), and RMSE percent of mean ($RMSE_{\%}$) (Equations (5-6) to (5-8)). Additionally, the Akaike Information Criteria (AIC) (Akaike, 1974) and the Bayesian Information Criteria (BIC) (Schwarz, 1978) were used to test the statistical difference among the selected models (Equations (5-9) to (5-10)). Lower values of AIC and BIC indicate a better performance of the model (Vrieze, 2012).

$$A(y, \hat{y}) = \frac{1}{n_{\text{samples}}} \sum_{i=0}^{n_{\text{samples}}-1} (\hat{y_i} = y_i) \tag{5-2}$$

in which:

$\hat{y_i}$: predicted value

y_i : corresponding true value

n_{samples} : is the total number of validation samples

$$P = \frac{(\text{TP})}{(\text{TP}) + (\text{FP})} \tag{5-3}$$

$$R = \frac{(\mathrm{TP})}{(\mathrm{TP}) + (\mathrm{FN})} \tag{5-4}$$

$$F_1 = 2 \times \frac{P \times R}{R + P} \tag{5-5}$$

in which:

TP : true positive

FP : false positive

FN : false negative

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{5-6}$$

in which:

$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \varepsilon_i^2$

ε : the error term

n : is the total number of validation samples

$$\mathrm{RMSE}(y, \hat{y}) = \sqrt{\frac{1}{n_{\mathrm{samples}}} \sum_{i=0}^{n_{\mathrm{samples}}-1} (y_i - \hat{y}_i)^2} \tag{5-7}$$

$$\mathrm{RMSE}_{\%}(y,\hat{y}) = \sqrt{\frac{1}{n_{\mathrm{samples}}}\sum_{i=0}^{n_{\mathrm{samples}}-1}\left(\frac{y_i-\hat{y}_i}{y_i}\right)^2 \times 100} \quad (5\text{-}8)$$

in which:

$\hat{y}_i$: predicted value of the *i* samples

y_i : corresponding true value of the *i* samples

n_{samples} : is the total number of validation samples

$$\mathrm{AIC} = n \times \log\left(\frac{\mathrm{RSS}}{n}\right) + 2 \times K \quad (5\text{-}9)$$

$$\mathrm{BIC} = n \times \log\left(\frac{\mathrm{RSS}}{n}\right) + K \times \log(n) \quad (5\text{-}10)$$

in which:

RSS : residuals sum of squares

K : number of parameters (including intercept)

n : number of observations

Seagrass AGB mapping

The best performing model in step (4) or step (5) (Figure 5.3) was used to generate the distribution maps of the seagrass AGB parameters. The binary seagrass map produced in this study was used to mask out the non-seagrass area in the AGB map.

5.3. Results

5.3.1. Seagrass binary mapping

The results indicate a high accuracy for seagrass binary mapping using the S-1 image data. The XGB obtained a higher metrics of P (0.82) and F_1 (0.86) comparing to the CB model (Table 5.4) and the accuracy is significantly different from two ML models (McNemar test, Table 5.4).

Table 5.4. Model performance for seagrass binary mapping

	A	P	R	F_1
XGB	0.92	0.82	0.90	0.86
CB	0.90	0.78	0.91	0.84
	McNemar test			
	$\chi 2$ value	p-value		
XGB-CB	10.89	0.00		

The feature importance determines the contribution of the input variables to the seagrass binary mapping, in which the band VH is the most influential variable (62%), following by the PCA_1_7_band (17.2%), VV / VH, VV, and VV − VH, VH / VV (Figure 5.4). The transformations of VH − VV and (VV + VH) / 2 have no impact on the seagrass binary mapping in this study.

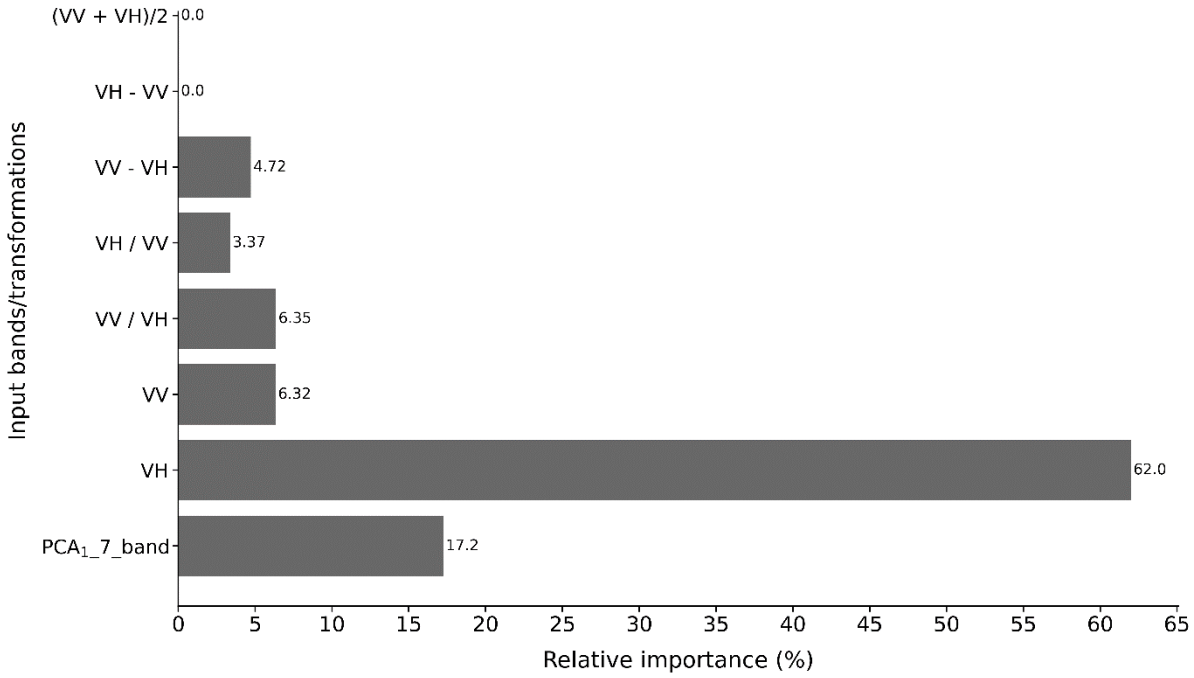


Figure 5.4. Influence of input variables for seagrass binary mapping

Using the XGB model, we produced a seagrass binary map for Tauranga Harbor (Figure 5.5). The map indicates an estimated total area of approximately 1,100 ha, with continuous meadows in all parts of the harbor, and these are particularly abundant in the middle section.

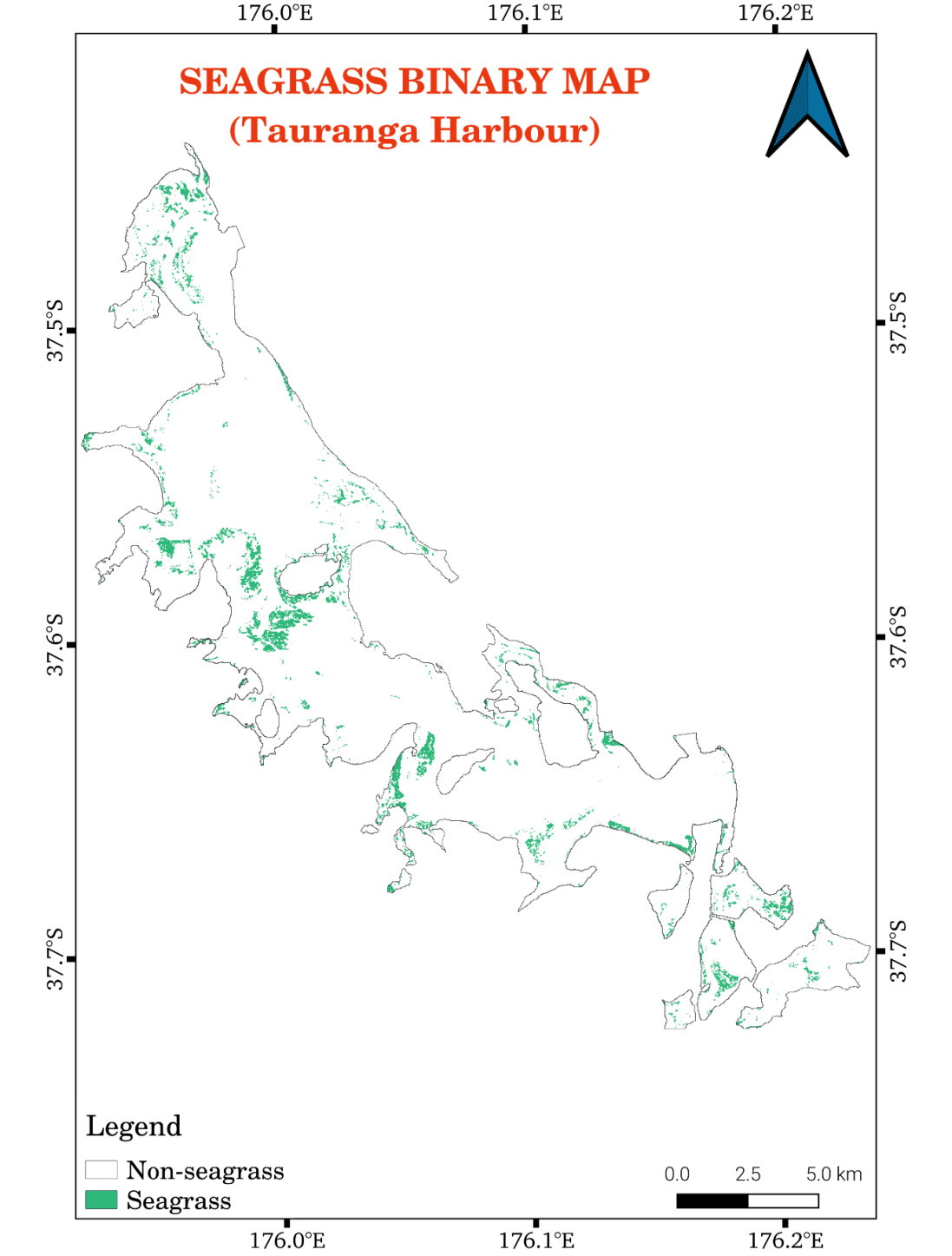


Figure 5.5. Seagrass binary distribution map derived from XGB model using S-1 image data

5.3.2. Seagrass AGB estimation

During the field survey, the seagrass AGB varied over an order of magnitude from 23 to 237 g DW m^{-2}, with a mean of 69 g DW m^{-2}, standard deviation of 42 g DW m^{-2} and 99% of the field data ranged between 23 to 178 g DW m^{-2}. The result indicates a wide variation of seagrass AGB from plots collected in the harbor.

In scenario 1, when only S-1 bands were considered, the LGBM model produced the highest R^2 of 0.49 whilst the XGB model was the better model, however at a lower R^2 of 0.47 in scenario 2 (Table 5.5). The accuracy was improved for AGB estimation when combining S-1 and S-2 bands in scenario 3, with the LGBM model yielding an R^2 of 0.54 and the lowest values of AIC and BIC (Table 5.5).

Table 5.5. Comparison of the ML models' performances for AGB retrieval in four scenarios[5]

		Model				
		RF	SVM	XGB	CB	LGBM
Scenario 1	R^2	0.34	-0.04	0.23	0.19	**0.49**
	RMSE (g)	0.55	0.70	0.60	0.62	**0.49**
	$RMSE_{\%}$ (%)	63.57	82.70	69.07	74.27	**50.02**
	AIC	40.40	47.50	42.90	43.70	**36.70**
	BIC	61.00	68.00	63.40	64.20	**57.20**
Scenario 2	R^2	0.21	0.00	**0.47**	0.22	0.26
	RMSE (g)	0.48	0.54	**0.39**	0.47	0.46
	$RMSE_{\%}$ (%)	68.11	71.88	**59.78**	68.61	60.75
	AIC	22.00	25.70	**16.20**	21.90	21.10
	BIC	37.60	41.30	**31.80**	37.50	36.70
Scenario 3	R^2	0.24	0.01	0.12	0.31	**0.54**
	RMSE (g)	0.50	0.57	0.54	0.48	**0.39**
	$RMSE_{\%}$ (%)	79.64	88.00	79.85	73.01	**51.61**

*5*The RMSE was calculated differently due to the random sample splitting from each scenarios. Therefore, the RMSE was used only to compare the model's performance inside the scenario whilst the R^2 was used to compare the model among the scenarios. Bold values indicate the best performance of the model.

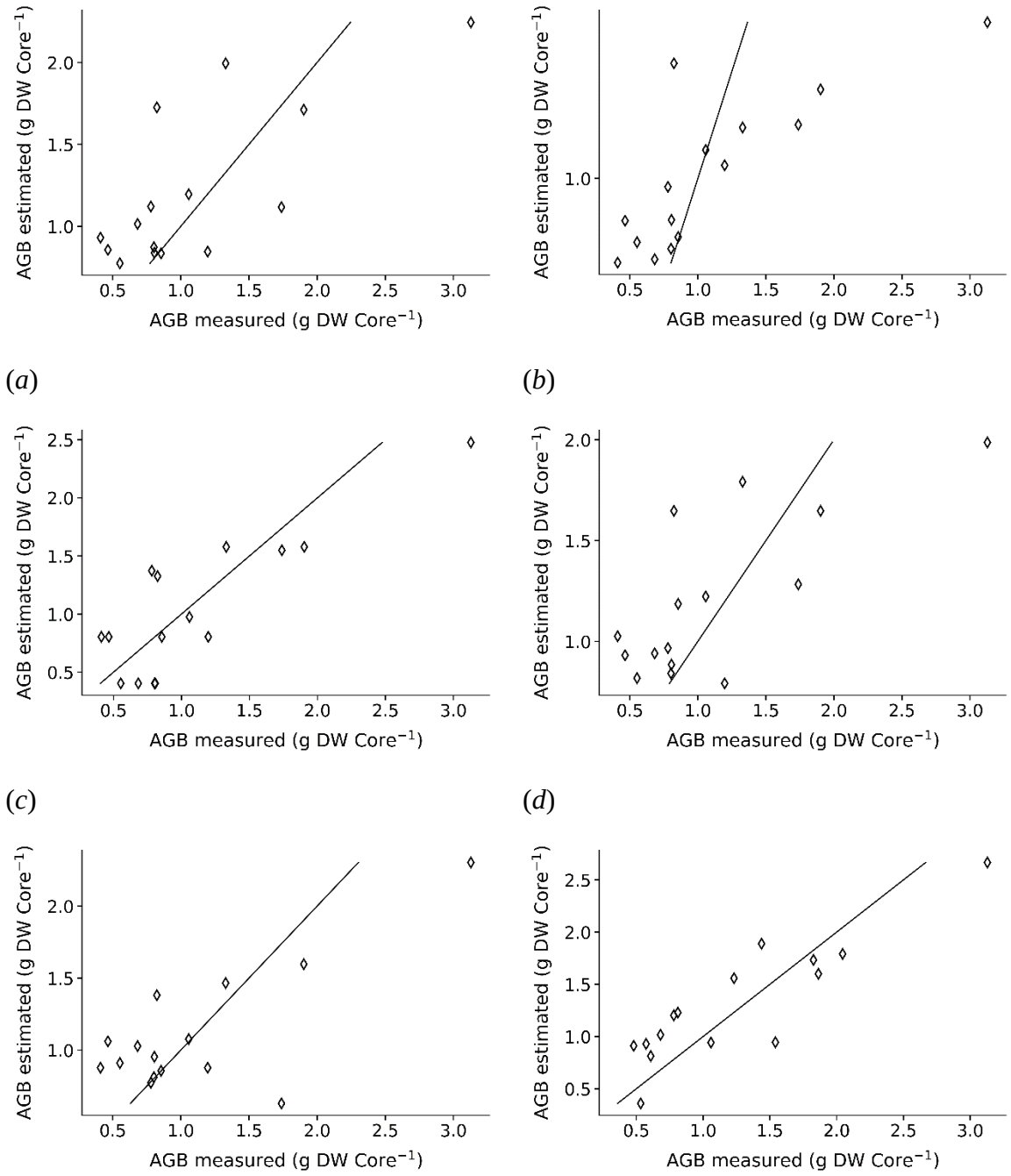
	AIC	82.20	85.60	83.80	80.10	**74.20**
	BIC	118.30	121.70	119.90	116.80	**110.30**
Scenario 4	R^2	0.52	0.32	**0.70**	0.49	0.53
	RMSE (g)	0.46	0.56	**0.37**	0.48	0.46
	$RMSE_{\%}$ (%)	55.87	45.60	**47.40**	59.56	55.89
	AIC	71.30	76.70	**64.40**	72.30	71.10
	BIC	104.50	109.90	**97.60**	105.60	104.40

In scenario 4, we tested the model performance with the feature selection using only bands that yielded a Spearman correlation coefficient > 0.10, which reduced the selected bands to 47. Using these, the XGB model derived an enhanced prediction of AGB at R^2 of 0.70 (Table 5.5). R^2, RMSE, AIC, BIC confirmed that XGB was superior to all other ML models. The model performances were visualized (Figure 5.6), and this confirmed that the XGB was a good fit with residuals smaller and more evenly spread than other models.

For Scenario 5, the best performing model from scenarios 1 to 4 (XGB model) was used as the cost function in the PSO algorithm, and 26 bands were selected from optical and SAR data fusion. The XGB model with PSO for feature selection (XGB-PSO) outperformed the Spearman correlation analysis used in Scenario 4, with a higher R^2 (0.75), and lower AIC (24.80), BIC (44.70), $RMSE_{\%}$ (41.69%). Overall, the XGB-PSO was the best model for seagrass AGB retrieval in Tauranga Harbor, with the highest R^2, lowest values of AIC, BIC, $RMSE_{\%}$ and a good agreement between estimated and measured AGB values (Figure 5.6 (*f*)). The bands that emerged as of greatest significance in the various models confirm that similar numbers of bands computed from SAR and multispectral data contribute to the final XGB-PSO model (Figure 5.7).

For seagrass AGB estimation, the most influential bands involve ρ_{Red}, IRECI of S-2 and VH_cons, VH_diss, VH_entropy, VV_corr of S-1 images. Among selected vegetation indices using the PSO algorithm, several new VIs computed from S-2 data such as MCARI, IRECI,

EVI2 play a role in estimating seagrass AGB in the study area. On the other hand, other common VIs (NDVI, SAVI) are less important when retrieving AGB.



(*e*) (*f*)

Figure 5.6. Scatter plot of AGB estimation use S-1 and S-2 bands with feature selection from Spearman correlation (*a*) - (*e*) and PSO (*f*): (*a*) RF, (*b*) SVM, (*c*): XGB, (*d*): CB, (*e*) LGBM, (*f*): XGB-PSO

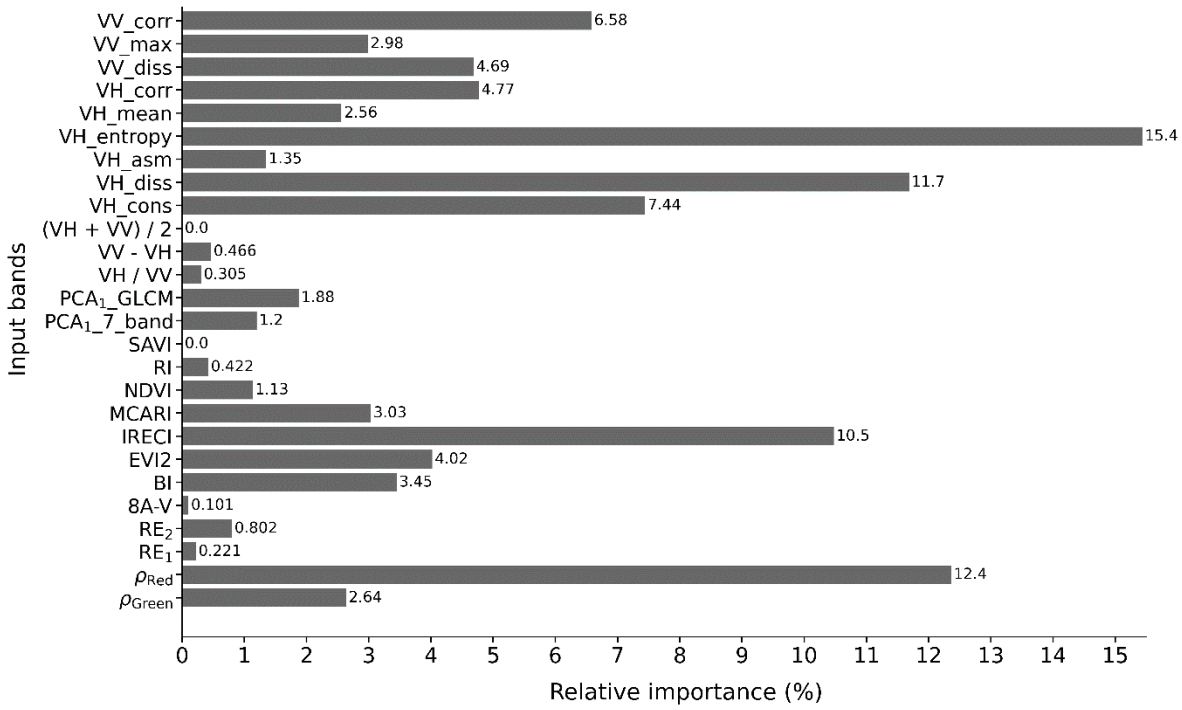


Figure 5.7. Variable importance for AGB estimation in Tauranga Harbor

Seagrass AGB distribution map

We created an AGB distribution map for Tauranga Harbor using the XGB-PSO model (Figure 5.8). Predicted seagrass AGB ranged from 15 to 201 g DW m^{-2} and almost all seagrass meadows were estimated to have AGB from 61 to 139 g DW m^{-2}. We predict highest seagrass AGB values in the areas in the center and south of the harbor.

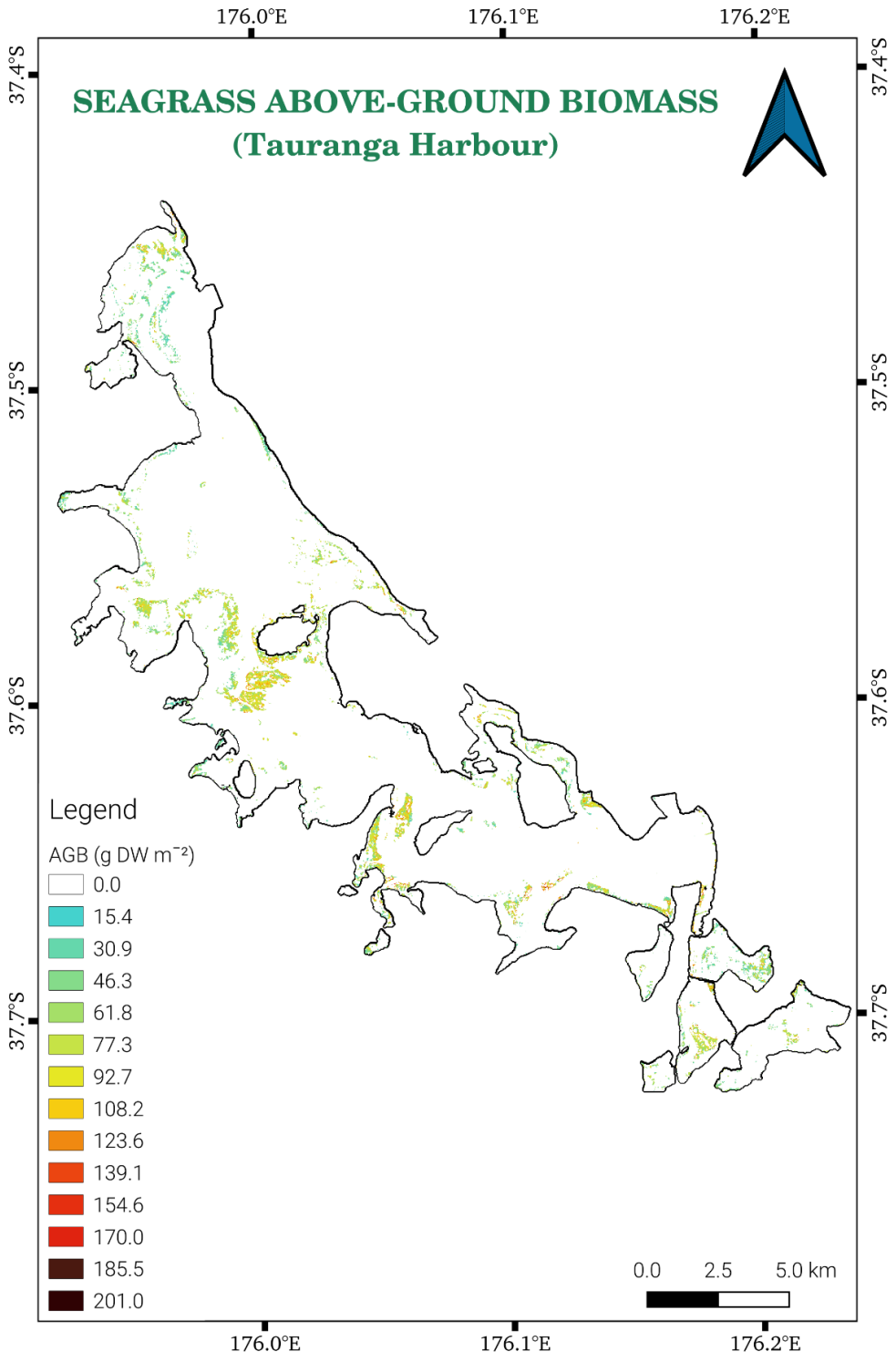


Figure 5.8. Seagrass AGB map for Tauranga Harbor created from the XGB-PSO model

5.4. Discussion

In this research, we evaluated novel approaches for estimation of seagrass properties using EO resources. We attempted to map the seagrass distribution using only SAR data, to offset constraints imposed on optical sensors by viewing conditions. We then went on to assess whether AGB could be best estimated using SAR, multispectral or a fusion of these remote sensing data, when combined with advanced ML models. To our knowledge, no other studies have applied EO data fusion techniques for such applications. In addition, we demonstrate how an advanced feature selection technique, using PSO metaheuristic optimization, improved the reliability of the retrievals to yield spatial AGB maps at 10 m resolution.

We were capable of mapping the distribution of intertidal seagrass meadows (present-absent) using 8 input variables derived from S-1 data (2 original bands and 6 transformations) with high accuracy (F_1 score 0.86). Our mapping is the first report on the use of SAR data for seagrass detection. The finding of the high contribution of the VH to the detection of vegetation is in agreement with the literature, where this polarization conveys information on stem scattering (Shang et al., 2020; Xu et al., 2019). Our results indicate that complex relationships across the SAR signals exists, with transformations and the PCA of all bands contributing 33% to the XGB model performance.

For seagrass AGB, the spatial map indicates high biomass values in the middle and southern parts of the harbor which coincides with the distribution of healthy seagrass meadows in a previous map (Ha et al., 2020). Compared to the task of binary/trinary classification, the quantification of AGB is more of a challenge due to the limited amount of training/testing data and the difference in the input data ranges (*i.e* the limited number of output classes for classification comparing to a wide range of output variables for regression against biomass). The

accuracy of the regression is, therefore usually lower than the binary/trinary classification task as presented in this study.

Substantial performance differences were observed with the use of S-1, S-2 or both, and between the various ML techniques applied. Best performance consistently was associated with scenarios using both the S-1 and S-2 images, and overall the XGB-PSO was the most effective ML model. Vorster et al. (2020) discussed the variation of $RMSE_{\%}$ in the forest biomass estimation, in which satellite image based prediction models contribute 25% to 75% to the total uncertainty of the estimation (Vorster et al., 2020). Salum et al. (2020) reported $RMSE_{\%}$ of 20.66% to 43.81% for mangrove AGB estimation (R^2 ranged from 0.90 to 0.97) (Salum et al., 2020) whilst Li et al. (2020) indicated the values of $RMSE_{\%}$ of 38% to 54% for forest AGB retrieval (R^2 ranged from 0.38 to 0.75) (Li et al., 2020). The $RMSE_{\%}$ of XGB-PSO model (41%), therefore is rationale and acceptable when dealing with the small size and low AGB of the seagrass *Z. muelleri* in this study.

Feature selection resulted in higher accuracy of AGB estimation, which exceeded the accuracy of a simple linear regression model for predicting AGB for seagrasses from VHR optical images (Roelfsema et al., 2014; Koedsin et al., 2016). In other words, the high performance of the XGB model indicates that high to very high resolution images might not be the only solution to return high accuracy seagrass AGB estimation. Naidoo et al. (2019) and Byrd et al. (2018) applied similar scenarios with and without the fusion of S-1 and S-2 images for predicting biomass of emergent wetland plants, and retrieved AGB at an R^2 ranging from 0.36 to 0.63, slightly less than those achieved in this study (R^2 = 0.75). Given the targets of the other studies had a higher biomass (5 to 3000 g m^{-2} in the wetland), and likely presented a more favorable radar target than

emerged seagrass, our proposed models are promising and could be expanded to include not only the seagrass but also other blue carbon ecosystems in the coastal areas.

The integration of the metaheuristic optimization with remote sensing data fusion is another new approach to the estimation of seagrass biophysical parameters from remote sensing data. Our results show the value of including advanced feature selection tools to gain higher accuracy of seagrass AGB estimation. Compared to the results in scenarios 3 (no feature selection) and 4 (feature selection using Spearman correlation analysis), the model performance was improved substantially by the addition of the PSO (scenario 5). The PSO is fast proving a powerful tool in ecological models. Through the harmonization of the control parameters of w, c_1, and c_2 inside the swarm, the next best location is updated among the particles via each iteration. In addition, the memory function (*i.e* the personal best feature in the algorithm) provides another solution for the particles to revert to the best way when they explore a "wrong way" in the searching space. Two features of the PSO - the iterative updating of the next best solution within swarm space and the ability to revert when the algorithm explores a "wrong way" in the searching space - usually result in a faster convergence of the swarm and better exploration of the target variable (seagrass AGB in this study) (Bansal et al., 2009; Elbeltagi et al., 2005; Panda & Padhy, 2008; Xiaohui et al., 2020). Since the effective application of the PSO for seagrass AGB estimation is first reported here, it is expected that our results will pave the way for advanced methods to be applied to other seagrass species in other international locations (Akbari et al., 2020; Sengupta et al., 2018; Wang et al., 2018).

We observed the most important variables for retrieval of seagrass AGB from S-2 are the single color bands ρ_{Green} and ρ_{Red}, the vegetation indices EVI2, IRECI, MCARI, the soil index BI and the GLCM bands VH_cons, VH_diss, VH_entropy, VV_corr of S-1 image (Figure 5.7). The

ρ_{Green} and ρ_{Red} bands are likely sensitive to the pigment information, whilst the VIs and the SIs provide useful information on the density of the seagrass. These indices were derived to provide information on vegetation from reflectance data, and have been shown to be effective in other situations (Morcillo-Pallarés et al., 2019; Mulder et al., 2011; Xue & Su, 2017). The GLCM bands, on the other hand, have not previously been used for seagrass, but here appear to provide quantitative information on the shape and texture of the meadows, and make a significant contribution to the final model. Whether the SAR data are truly addressing the texture of seagrass, or some other component of the seagrass ecosystem such as cockles which typically co-occur in Tauranga Harbor (Morrison et al., 2014), is not clear. Our results are, however, consistent with recent studies for mangrove AGB retrieval in the tropics which suggested that S-1 is appropriate for quantification of low biomass vegetation (Pham, Le et al., 2020; Pham, Yokoya et al., 2020).

The method described here provides a novel direction using data fusion and state-of-the-art machine learning and optimized techniques in detecting the distribution and retrieving seagrass AGB in the coastal area at 10 m spatial resolution, and may be useful for other blue carbon ecosystem analyses. All satellite and analytical products used in this work are open access, and present opportunities for low-cost analyses where suitable data to supervise classification and regression can be obtained. Our work could provide a baseline for blue carbon credit markets, toward a nature-based solution to climate change bb current study, however is spatially narrowed to the intertidal zones where the seagrass meadows can be exposed at low tide and all the bands of S-1/S-2 are available for further processing. In addition, the data saturation for AGB estimation was reached at approximately 170 g DW m^{-2} which might be considered as the limitation when using S-1 data.

Future research might be to expand the proposed methods with more field surveying data on various seagrass species across different coastal regions at a larger-scale using different SAR sensors such as advanced land observing satellite 2 (ALOS-2) phased array type L band synthetic aperture radar (PALSAR-2). The L band SAR has a longer wavelength, penetrates deeper than the C band of S-1 image, and is potentially higher in data saturation (Pham, Yokoya et al., 2020). In addition, various metaheuristic optimization algorithms (Harris Hawks Optimization (HHO) (Heidari et al., 2019), Firefly Algorithm (FFA) (Pan et al., 2019)) could be investigated for better feature selection to improve the accuracy of the retrieval models.

5.5. Conclusion

Our work pioneers the use of SAR remote sensing data (S-1 imagery) for mapping the distribution of intertidal seagrasses, and the fusion of optical remote sensing data (S-2 imagery) with advanced machine learning and metaheuristic optimization techniques (PSO) for estimating AGB of the small seagrass species *Z. muelleri* in the intertidal zone in Tauranga Harbor.

S-1 provides useful remote SAR image data for the binary mapping of seagrass with high accuracy (A 0.92 and F_1 0.86 when combined with the XGB model). For seagrass AGB, both optical and SAR sensing provided information to drive models and data fusion provided the best retrieval accuracy. In addition, feature selection improved the model accuracy for AGB estimation with a superior performance of the metaheuristic optimization PSO using the XGB model over the Spearman correlation.

Our results contribute a novel advance in the use of multispectral and SAR remote sensing data fusion and state-of-the-art ML models together with metaheuristic optimization, for both seagrass detection and seagrass AGB estimation in the intertidal zones that may be expandable to various seagrass species globally.

5.6. Chapter supplemental material

Table 5.S1. Selected parameters for atmospheric correction using ACOLITE

Parameter	Value
Ancillary data	
Gas transmittance	True
Ozone concentration (cm^{-1})	0.30
Water vapour concentration ($g\ cm^{-2}$)	1.50
Pressure	Normal pressure
Masking	
Negative reflectance masking	True
Cirrus masking	True
Other parameters	
Sky correction	True
Dark spectrum fitting	Fixed
Sun glint correction	False

Table 5.S2. List of tuned hyper-parameters for seagrass binary mapping

CatBoost		Extreme Gradient Boost	
Depth	5	Boosting type	GBTREE
Learning rate	0.01	Gamma	2
L2 leaf regression	3	Learning rate	0.20
Number of trees	30	Max depth	5
		Min child weight	3
		Number of trees	50

Table 5.S3a. List of tuned hyper-parameters for AGB estimation in Scenario 1

Random Forest		Extreme Gradient Boost	
Bootstrap	True	Boosting type	GBTREE
Max depth	1	Gamma	3
Max features	2	Learning rate	0.10
Min sample leaf	4	Max depth	3
Min sample split	3	Min child weight	7
Number of trees	100	Number of trees	30
CatBoost		Light Gradient Boosting Machine	
Depth	10	Boosting type	DART
Learning rate	1	Learning rate	0.10
L2 leaf regression	4	Max depth	-1
Number of trees	10	Number of leaves	10
		Number of trees	200
Support Vector Machine			
Kernel	RBF		
C	10		
Epsilon	0.01		
Gamma	1000		

Table 5.S3b. List of tuned hyper-parameters for AGB estimation in Scenario 2

Random Forest		Extreme Gradient Boost	
Bootstrap	True	Boosting type	GBTREE
Max depth	3	Gamma	1
Max features	2	Learning rate	0.10
Min sample leaf	4	Max depth	3
Min sample split	9	Min child weight	7
Number of trees	100	Number of trees	50
CatBoost		Light Gradient Boosting Machine	
Depth	10	Boosting type	GBDT
Learning rate	0.30	Learning rate	0.50
L2 leaf regression	4	Max depth	-1
Number of trees	10	Number of leaves	10
		Number of trees	10
Support Vector Machine			
Kernel	RBF		
C	1000		
Epsilon	0.50		
Gamma	0.00		

Table 5.S3c. List of tuned hyper-parameters for AGB estimation in Scenario 3

Random Forest		Extreme Gradient Boost	
Bootstrap	True	Boosting type	GBTREE
Max depth	3	Gamma	3
Max features	2	Learning rate	0.20
Min sample leaf	4	Max depth	3
Min sample split	3	Min child weight	7
Number of trees	50	Number of trees	30
CatBoost		Light Gradient Boosting Machine	
Depth	10	Boosting type	DART
Learning rate	0.20	Learning rate	0.30
L2 leaf regression	4	Max depth	-1
Number of trees	100	Number of leaves	10
		Number of trees	50
Support Vector Machine			
Kernel	RBF		
C	1000		
Epsilon	0.50		
Gamma	0.00		

Table 5.S3d. List of tuned hyper-parameters for AGB estimation in Scenario 4

Random Forest		Extreme Gradient Boost	
Bootstrap	True	Boosting type	GBTREE
Max depth	3	Gamma	1
Max features	8	Learning rate	0.51
Min sample leaf	1	Max depth	3
Min sample split	2	Min child weight	6
Number of trees	30	Number of trees	50
CatBoost		Light Gradient Boosting Machine	
Depth	2	Boosting type	GBDT
Learning rate	0.185	Learning rate	0.92
L2 leaf regression	4	Max depth	-1
Number of trees	30	Number of leaves	10
		Number of trees	100
Support Vector Machine			
Kernel	RBF		
C	100000		
Epsilon	0.20		
Gamma	0.00		

Table 5.S4. PSO tuned hyper-parameters for AGB estimation

Estimation	Search algorithm	Number of particles	w	c_1	c_2
AGB	Global search	30	0	0.50	0.50

Table 5.S5. Abbreviation of Sentinel-1 and Sentinel-2 bands used in the study

Abbreviation	Full name	Abbreviation	Full name
Sentinel-1		Sentinel-2	
VV	Dual-polarization vertical transmitting and vertical receiving	8A-V	8A vegetation
VH	Dual-polarization vertical transmitting and horizontal receiving	RE	Red edge
VH_cons:	VH contrast	NIR	Near infrared
VH_diss:	VH dissimilarity	BI	Brightness index
VH_homo:	VH homogeneity	CI	Color index
VH_asm:	VH asm	EVI2	Enhanced vegetation index 2
VH_energy:	VH energy	GNDVI	Green normalized difference vegetation index
VH_max:	VH max	IRECI	Inverted red edge chlorophyll index
VH_entropy:	VH entropy	MCARI	Modified chlorophyll absorption in reflectance index
VH_mean	VH mean	NDI45	Normalized difference index derived from band 4, 5
VH_variance:	VH variance	NDVI	Normalized difference vegetation index
VH_corr:	VH correlation	RI	Red index
VV_cons:	VV contrast	RVI	Ratio vegetation index

VV_diss:	VV dissimilarity	SAVI	Soil adjusted vegetation index
VV_homo:	VV homogeneity		
VV_asm:	VV asm		
VV_energy:	VV energy		
VV_max:	VV max		
VV_entropy:	VV entropy		
VV_mean	VV mean		
VV_variance:	VV variance		
VV_corr:	VV correlation		

Chapter 6 A novel and open source web-GIS approach for seagrass data visualization and collaboration

Abstract: Web-GIS provides an effective platform to visualize scientific data in the form of a readily interpretable thematic map, which is capable of benefiting a broad community. Using this approach, web-GIS fosters geospatial data sharing, collaboration, and strengthens community-based conservation efforts. In the current study, we developed a web-GIS application based on an integrated framework using open-source assets, including QGIS, QGIS server, Lizmap web client, PostGIS, and PostgreSQL to successfully visualize a database of seagrass dynamics in Tauranga Harbor, New Zealand. The name of this application is Blue Carbon Web (BCW). BCW is designed to support diverse functions, be secure, with data loading, and to facilitate collaboration. Our work is the first application of open source web-GIS for seagrass ecosystems in New Zealand, provides an elegant and lightweight interface, which supports the improvement of community awareness in conservation of this blue carbon ecosystem in New Zealand.

Keywords: seagrass, web-GIS, QGIS, PostgreSQL, PostGIS, Lizmap

6.1. Introduction

Seagrasses are a polyphyletic group of angiosperm species which have adapted to marine life, and are widely distributed across the climatic regions. Recent studies reveal various important ecosystem services provided by seagrass, including filtering water, acting as a nursery and breeding ground, coastal erosion attenuation, and carbon sequestration (Nordlund et al., 2016, 2018). A large amount of research effort has been dedicated to accurate mapping, change detection and long-term policy engagement, toward sustainable conservation of seagrass meadows across the world (Cullen-Unsworth & Unsworth, 2013; Nordlund et al., 2018; Unsworth et al., 2018; Dat Pham et al., 2019; Ha et al., 2020).

Despite the demonstration by researchers of valuable ecological functions and the development of tools for updating temporal distribution and estimating carbon sequestration capability (Nordlund et al., 2016; Dat Pham et al., 2019; Macreadie et al., 2019), seagrass meadows have been degraded in both area and habitat quality globally (Waycott et al., 2009; de los Santos et al., 2019). Challenges facing seagrass conservation include failures in transplantation (Suykerbuyk et al., 2016; Paulo et al., 2019), impacts of intensive aquaculture and unsustainable fisheries (Herbeck et al., 2014; Ferriss et al., 2019), extreme weather phenomena (Fraser et al., 2014; Oprandi et al., 2020), water pollution and sediment accumulation (Fraser & Kendrick, 2017), and importantly, a failure in effective social communication (Orth et al., 2006; Ramesh et al., 2018; Unsworth et al., 2018). The current state of seagrass dynamics is widely investigated by a small community of scientists and a limited number of the managers (Hossain et al., 2015; Dat Pham et al., 2019), whilst a broader community, including fishermen, local managers, and tourist operators, are still out of the box (Unsworth et al., 2018). The reasons for this may come from a limited access to the information generated in research databases, poor communication via social

media, and a lack of interest from local managers. To deal with these issues, social media have been identified as high impact, essential tools that provide instant, easy access, and simple data visualization for end users (Di Minin et al., 2015; Wu et al., 2018; Toivonen et al., 2019). In this context, a web-GIS approach will be the most popular practical solution with a web-based interface providing an intuitive interaction, readily updated information and ease of use for multiple purposes (Rouse et al., 2007; Sui & Goodchild, 2011; Werts et al., 2012). Such an approach requires a geographic information system (GIS), web server, web client, GIS server, and database management application as essential system components.

Recently, a limited number of web-GIS based applications have been developed that are specifically designed for seagrass ecosystems or integrate seagrass dynamics in their database. A detailed review of the literature was presented in Section 2.5, Chapter 2 in this thesis. Of the existing web-GIS systems, seagrass dynamics data was only integrated into the global database of the Ocean Data Viewer (http://data.unep-wcmc.org), Mapping Ocean Wealth (http://maps.oceanwealth.org/#), simplified as the image data web-GIS (https://seagrassspotter.org/), or as a local database of in the Blue Carbon Mapping Tool (http://bluecarbontoolkit.ae/en/layout) and seagrass extent developed by the Bay of Plenty Regional Council (https://maps.boprc.govt.nz/datasets/seagrass-extents-bay-of-plenty). These web-GIS systems take advantage of global visualization, data sharing, and support various ecosystems. However, we have observed a few limitations from the listed web-GIS systems, including a failure to update seagrass data, slow performance, discrete illustrations of spatial dynamics as well as lack of a mapping interface. In addition, many come at high cost, due to the problems of a big database, web client and use of commercial software. Currently there is no web-GIS application showing integrated seagrass blue carbon resources.

The recent literature review motivated us to build an open geospatial database of seagrass dynamics for Tauranga Harbor, New Zealand. To make this readily applicable and transferable, our goal was to base this around open access platforms. The database is structured in the format of PostgeSQL - a secure and advanced open source database management system (PGDG, 2020), mapping using QGIS - the best-known GIS open source application (Brunsdon & Singleton, 2015; Coetzee et al., 2020), connection to the QGIS server via Lizmap web client (Douchin, 2020) - an emerging web client and natively integrated to the QGIS/QGIS server, to create an elegant and lightweight web-GIS interface (hereafter namely Blue Carbon Web or BCW). Our application is expected to provide an effective communication web gate to a broader community, increasing the flow of information on seagrass ecosystem conservation, and engaging the willingness of the general population to value and protect the extant seagrass meadows in New Zealand.

6.2. Methodology

6.2.1. Visualization site

Tauranga Harbor, New Zealand (Figure 6.1) is our visualization site. A single species, *Zostera muelleri*, has colonized the inter-tidal zones of the harbor, where it forms one of the largest areas of seagrass in New Zealand (Park, 2011). Our previous work has indicated a significant change in area of seagrass meadows in the harbor over the last 29 years (1990 - 2019) (Ha et al., in revision, Chapter 4). In addition, our fieldwork has provided good observation data of the aboveground, and belowground dry biomass, and the total carbon of seagrass meadows at 57 random plots (Ha et al., 2020) that support the development of maps of cover, biomass and change. The available data provides an excellent case study to develop an experimental open source web-GIS application.

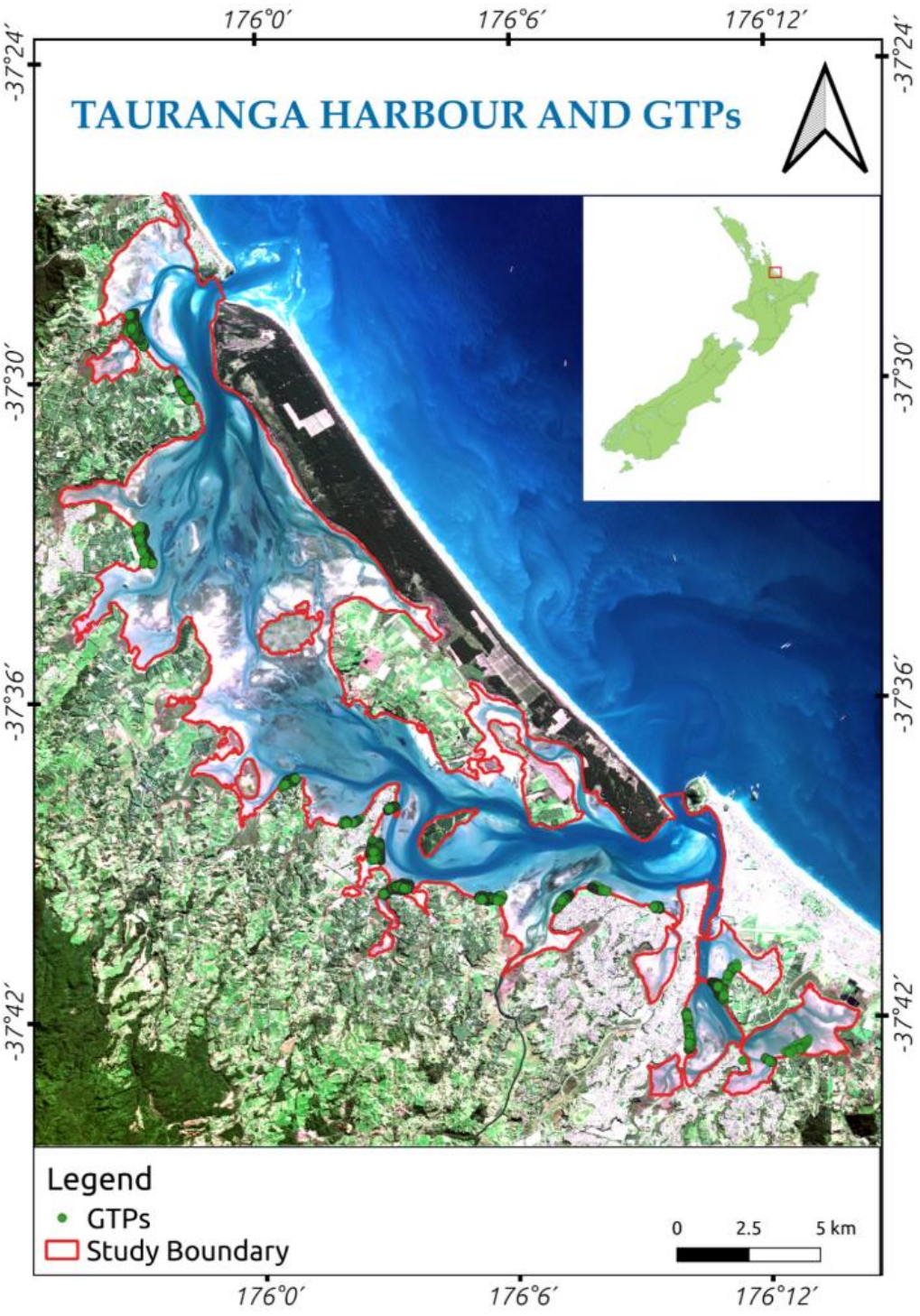


Figure 6.1. Tauranga Harbor – A web-GIS visualization site (pseudo color using ρ_{Red} - ρ_{Green} -ρ_{Blue} composition of Sentinel-2 imagery). The green dots indicate the location of ground-truth point collection and blue carbon assessment

6.2.2. Seagrass changes, dry biomass and total soil carbon data

The seagrass changes (Table 6.1), dry biomass and total soil carbon (Table 6.2) data are taken from our previous works in the years 2019 and 2020. The data of dry biomass and total soil carbon are formatted in GeoTIFF raster whilst GPS data of field survey are saved in Shapefile format. In total, seven (7) GeoTIFF files, and two (2) Shapefile are structured and stored in an PostgreSQL database with a file size approximately 78 megabyte (Mb).

Table 6.1. Seagrass changes data in Tauranga Harbor

Month year	Area (ha)	Format	Projection	Mapping source	Thematic map
April 1990	2237	GeoTIFF	WGS-84, UTM-60S	Landsat-4 TM	Raster map of seagrass distribution
Mar. 2001	2035			Landsat-7 ETM	
Feb. 2011	2380			Landsat-5 TM	
Mar. 2014	1734			Landsat-8 OLI	
May 2019	1184			Landsat-8 OLI	

Table 6.2. Dry biomass and total soil carbon data in Tauranga Harbor

Month/Year	Amount of plots	Format	Projection	Mapping source	Thematic map
Sep. 2019 & Mar. 2020	57	Shapefile	WGS-84, UTM-60S		Map of point-based field survey
		Shapefile	WGS-84, UTM-60S		Map of study site boundary
Mar. 2020		GeoTIFF	WGS-84, UTM-60S	Sentinel-1 & Sentinel-2	Raster map of dry biomass and total soil carbon

6.2.3. Methodology

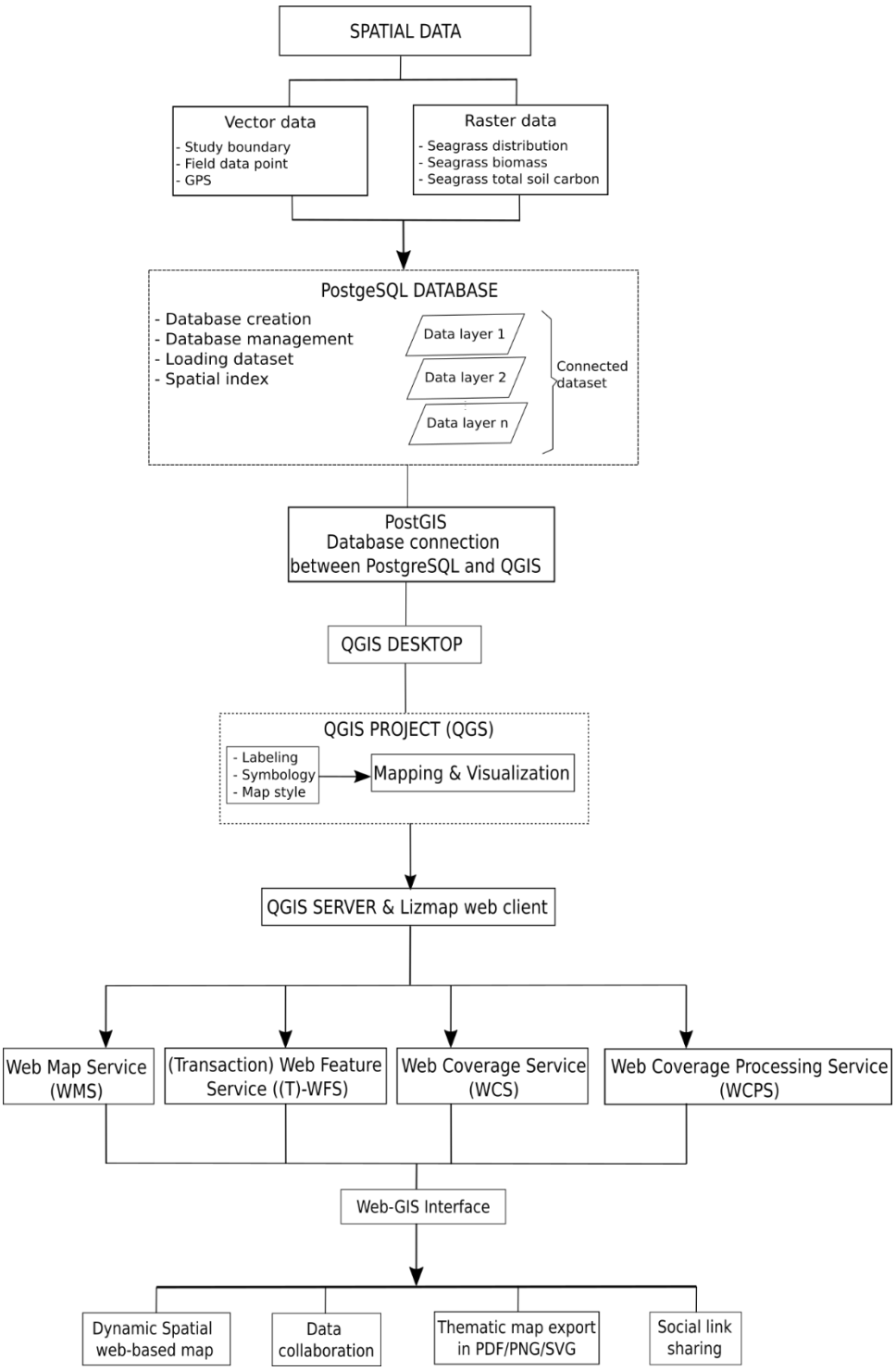


Figure 6.2. Flowchart of geospatial database and web-GIS building

Several components are required to build a web-GIS interface (Figure 6.2). The application of PostgreSQL is used to build and manage the geospatial database, PostGIS is used to connect the geospatial database between PostgreSQL and QGIS, QGIS desktop is used to do the data visualization, decoration in a map format whilst QGIS server and Lizmap client help to convey the styled map to the WWW via a web-GIS interface (the BCW application in this study).

Install and configure web-GIS components

In this study, the web-GIS components (Table 6.3) are installed and configured under the Linux environment (Lubuntu distribution, version 20.04).

Table 6.3. Application information used in the installation

Application	Version	Operating system	Source code	Link to download
QGIS server	3.16	Linux	Open source	https://qgis.org/en/site/forusers/alldownloads.html#debian-ubuntu
QGIS	3.16	Linux	Open source	https://qgis.org/en/site/forusers/download.html
Lizmap web client/ QGIS plugin	3.2	Linux	Open source	https://github.com/3liz/lizmap-web-client
PostgreSQL	12	Linux	Open source	https://www.postgresql.org/download/
PostGIS	3.0	Linux	Open source	https://postgis.net/install/

Installation instructions for the application of QGIS server, Lizmap plugin for QGIS, Lizmap web client, PostgreSQL, PostGIS and QGIS are presented in the supplemental material of the chapter.

Geospatial database building

PostgreSQL was used to build, manage the geospatial database using both command line and a user interface PgAdmin4 (Table 6.4). To make the web-GIS accessible and ease data collaboration, the popular Shapefile (shp) and GeoTIFF (tif) formats were used to stored vector and raster dataset in the geospatial database (Table 6.4).

Table 6.4. Functional level and data format of geospatial database

Function	Tool
Create server	PgAdmin4
Create database	PgAdmin4
Delete database	PgAdmin4
Add vector data	Database manager in QGIS
Add raster data	Command line

Data format	Data type
Shapefile	Point
Shapefile	Polygon
GeoTIFF	64bit real

Web-GIS performance evaluation

As a web-based platform, the BCW application should be fast in page loading, friendly in user interface, and easy for searching by the user. Considering the criteria mentioned, the four web services (Table 6.5) were implemented to test the BCW performance from the scores of (i) performance, (ii) page speed, (iii) load time which measure the speed of the web page, (iv) search engine optimization which quantifies the visibility of the web page to the user, (v) accessibility, and (vi) best practice which evaluates the friendly design of the web user interface. The Google web development tool (GWDT) supports a direct assessment for the local-host web whilst the GTmetrix, Dareboost, and Uptrends only work with the online website. For this reason, we used the ngrok application (https://www.ngrok.com/) to create a communication between the local folder of BCW and the WWW. An online link will be created and the performance of BCW can be tested using Gtmetrix, Dareboost, and Uptrents services.

Table 6.5. Web services for BCW performance evaluation

Web services	Measured score	Web link
GTmetrix	Page speed, load time	https://gtmetrix.com/
Dareboost	Load time	https://www.dareboost.com/en
Uptrends	Google page speed, load time	https://www.uptrends.com/tools/website-speed-test
Google web development tool (GWDT)	Performance, accessibility, best practice, SEO	https://developers.google.com/web/tools

6.3. Results

6.3.1. Seagrass distribution and blue carbon geospatial database

In this study, a geospatial database *postgres* was created, then connected to the QGIS application using the external extension PostGIS. Two vector (in Shapefile format) and seven raster (GeoTIFF format) layers were added into the database (Table 6.6) with a size of approximately 78 Mb. All the vector and raster dataset are managed using PgAdmin4 (Figure 6.3).

Table 6.6. Vector and raster dataset in the geospatial database

Name	Format	Amount	Size (Mb)
Sampling point	Shapefile (point)	1	0.32
Tauranga boundary	Shapefile (polygon)	1	0.10
Seagrass distribution map	GeoTIFF	5	11.1
Seagrass biomass map	GeoTIFF	1	33.3
Seagrass total carbon map	GeoTIFF	1	33.3
Total (Mb)			78.12

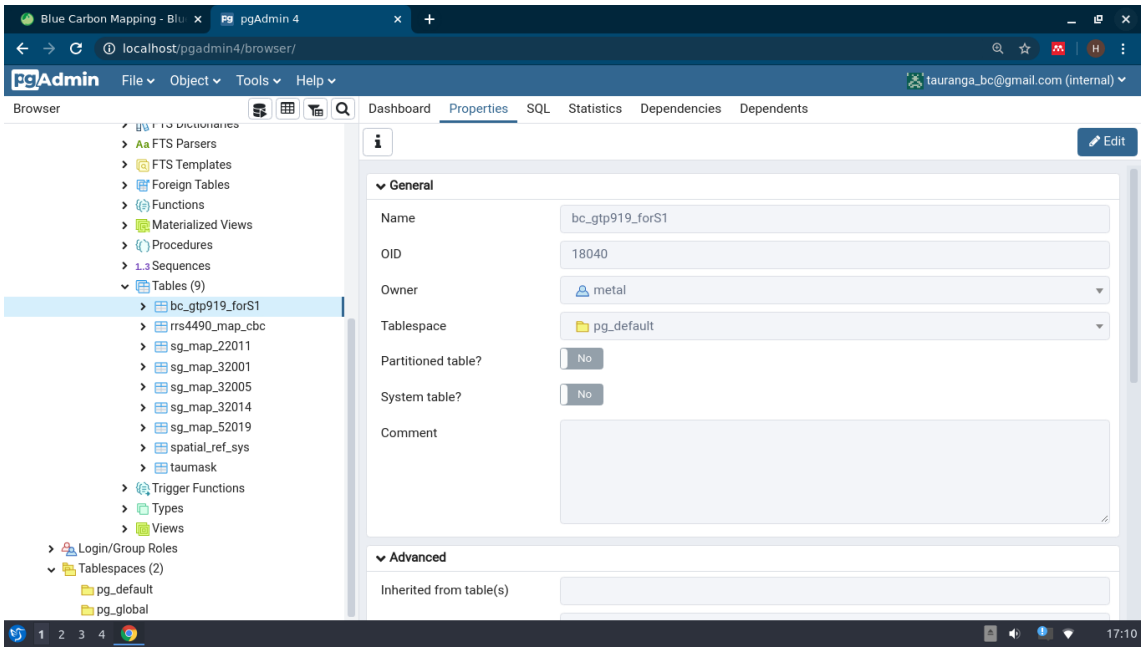


Figure 6.3. Data table in the interface of PgAdmin4

6.3.2. BCW performance

The performance of the local-host BCW was evaluated using the web services of GWDT, GTmetrix, Dareboost, and Uptrents. The analysis results (Table 6.7) indicate a fast and stable performance (scores of performance, page speed, load time), good search engine optimization (SEO) (score of SEO), and friendly user interface (scores of accessibility and best practice) that are essential to keep the visitors staying longer at the website. The GWDT service gave the BCW score 100/100 for performance whilst the Uptrents and GTmetrix score the page speed 86 - 92/100. Importantly, the load time ranges from 2.4 - 3.4 s which were smaller than the standard 4 s for a website loading, indicating an acceptable loading time for the BCW. We also noted the high scores of SEO (89/100) and best practice (85/100), implying a friendly, trustworthy, and safe web interface. In other words, the BCW bring a comfortable experience to the visitors, that will motivate them to spend more time to view the content of the website. At the time of page

loading, only a limited number of functions were loaded and this might explain the modest score of accessibility (61/100).

Table 6.7. BCW performance evaluation from GWDT, GTmetrix, Dareboost, and Uptrents

Service	Metric			
	Performance	Accessibility	Best practice	SEO
GWDT	100/100	61/100	85/100	89/100
	Google page speed score	Load time (s)		
Uptrents	92/100	2.4/4		
Dareboost	Overall score	Load time (s)		
	0.78	2.5/4		
	Page speed score	Load time (s)		
GTmetrix	86/100	3.4/4		

6.3.3. BCW functions

Data visualization

BCW provides a variety of visualization functions for both vector and raster data, including mapping, plotting with several options of base maps (Table 6.8 and Figure 6.4).

Table 6.8. BCW visualization functions

Functions	Explanation
Mapping	Present the maps of seagrass distribution, biomass and total carbon
Zoom in/ out Pan map Overview map Distance and area measurement Different base maps (Google terrain, Google satellite, open street map) Layer management (show/hide, expand/collapse)	

Functions	Explanation
Plotting	Plot the quantitative data (biomass, total carbon)
Histogram plot Box plot Scatter plot Download plot as image Zoom in/ out Auto scale	
Attribute table	View, manage and download data table
Show/hide list of tables Show/hide several tables in tab format Select/unselect rows Select number of entries to show Export and download table data	
Time series	Present various maps in a time series
Window pop-up	Present pop-up windows to view all relevant information

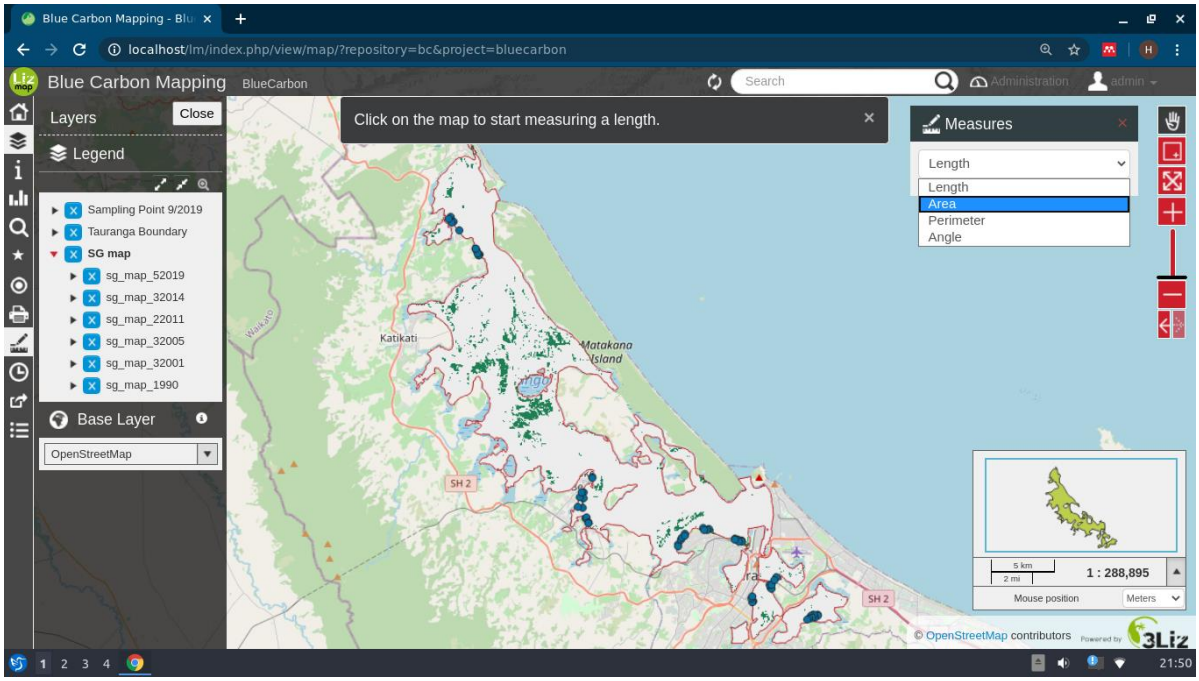
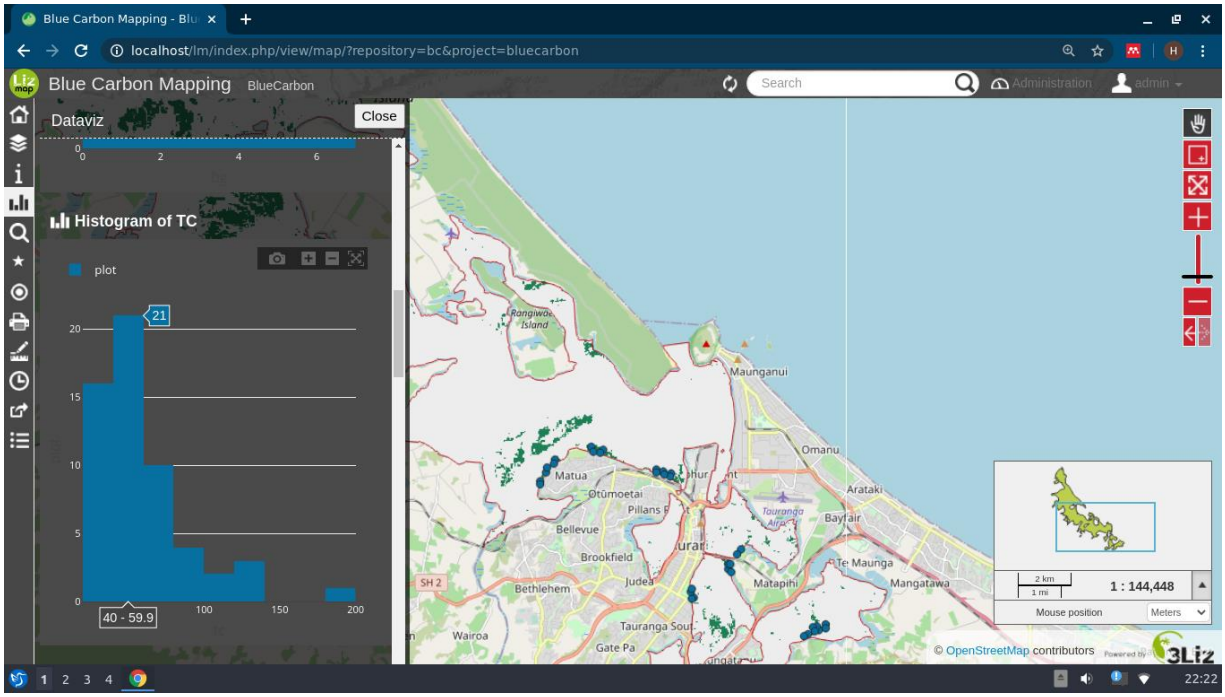
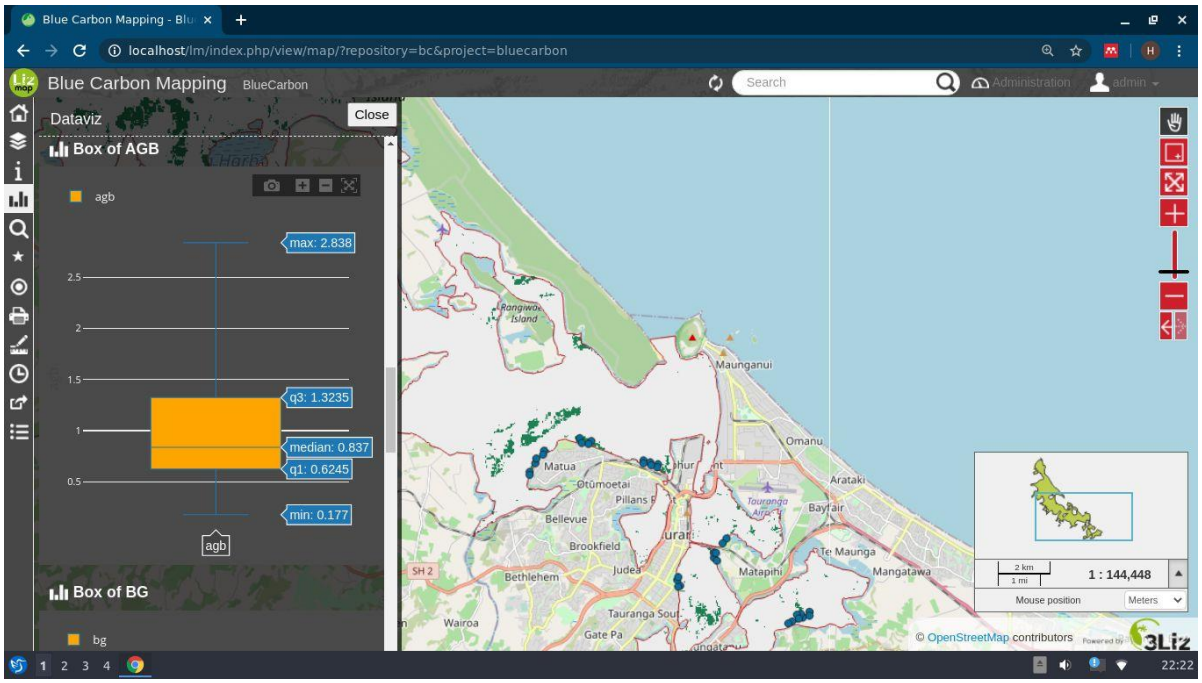


Figure 6.4. An overview of BCW web-GIS application

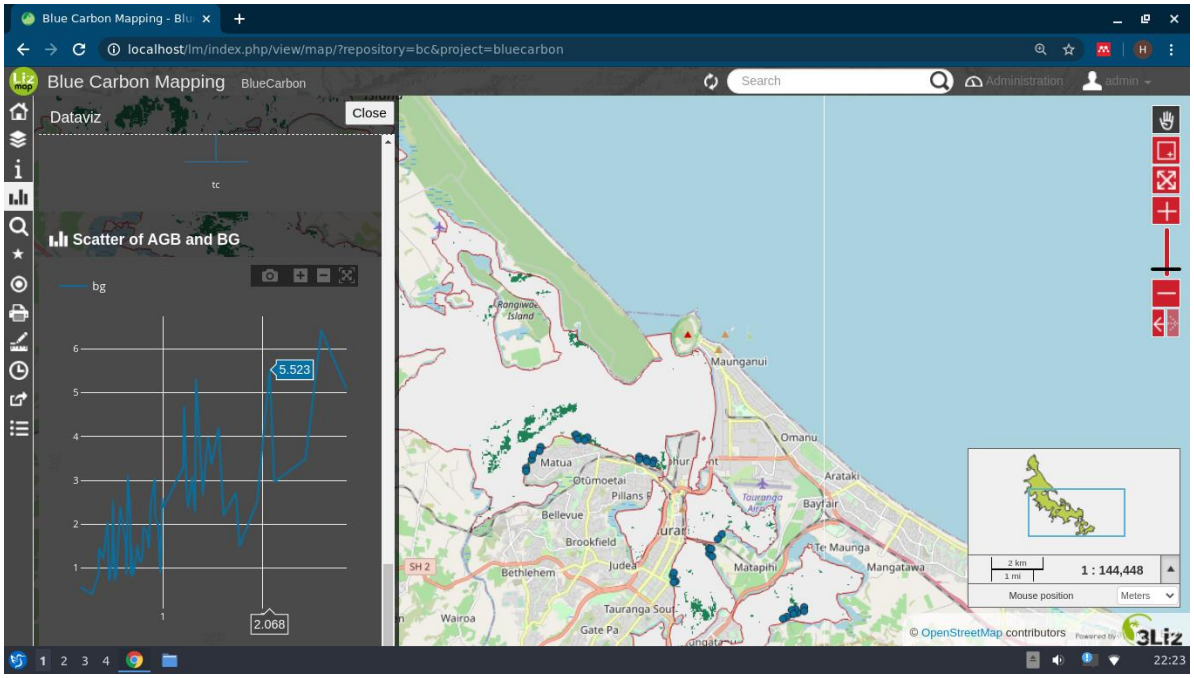
Using mapping, the user is able to view the distribution at various supported time points, the biomass and the total soil carbon of seagrass meadows in the year 2020. The layer control panel supports a convenience view to show or hide any layers that provides an overview of seagrass distribution, change detection, and the spatial variation of biomass and total soil carbon in the harbor. The measurement toolset provides the means to measure length, area, and angle of spatial objects. In addition, BCW supports three basic plots for data visualization, involving histogram, box and scatter plots (Figure 6.5). Plotting with parameters defined by the user will be supported in the future.



(*a*)



(*b*)



(*c*)

Figure 6.5. Plotting functions (*a*) histogram, (*b*) box, (*c*) scatter plots in the BCW

Another useful function is to view, manage and export tables from vector data. The data table is accessible through the table manager to show or hide any tables, select the desired rows or export all the tables to the formats of GML and GeoJSON (Figure 6.6).

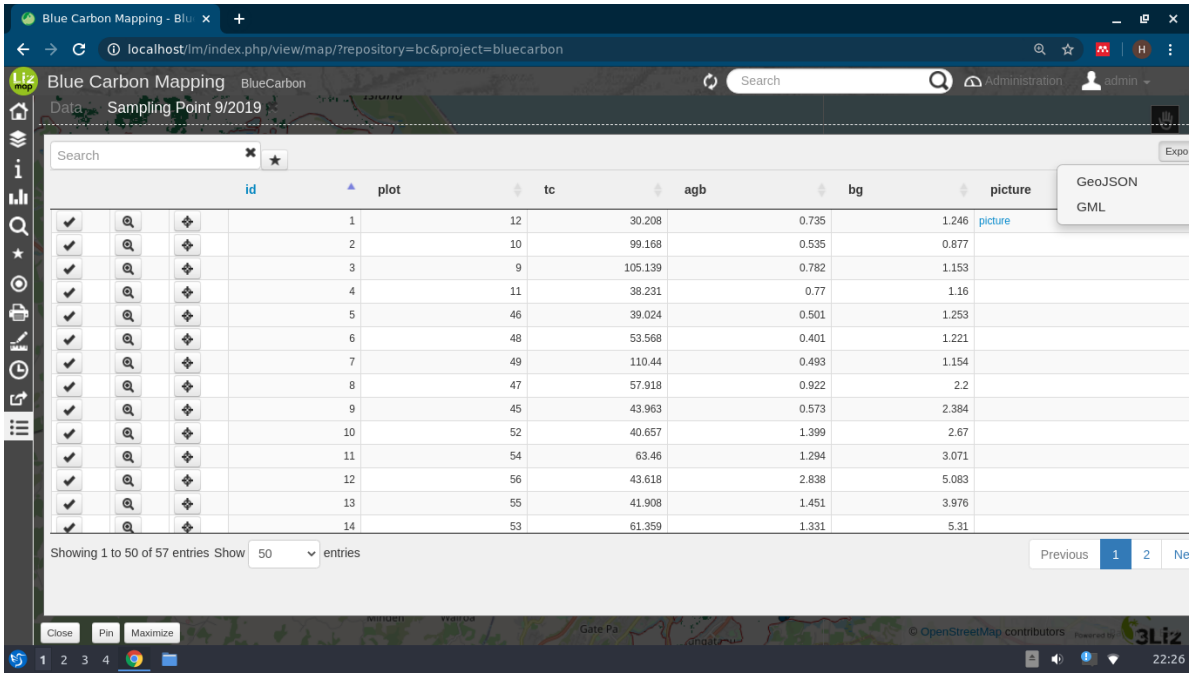


Figure 6.6. Table management and GeoJSON/GML export

Moreover, the user can open a pop-up window to view the detail information at any data point (sampling plot). This function is very helpful to check all relevant information regarding the sampling plots with an attached photograph of seagrass meadows at the time of sampling (Figure 6.7).

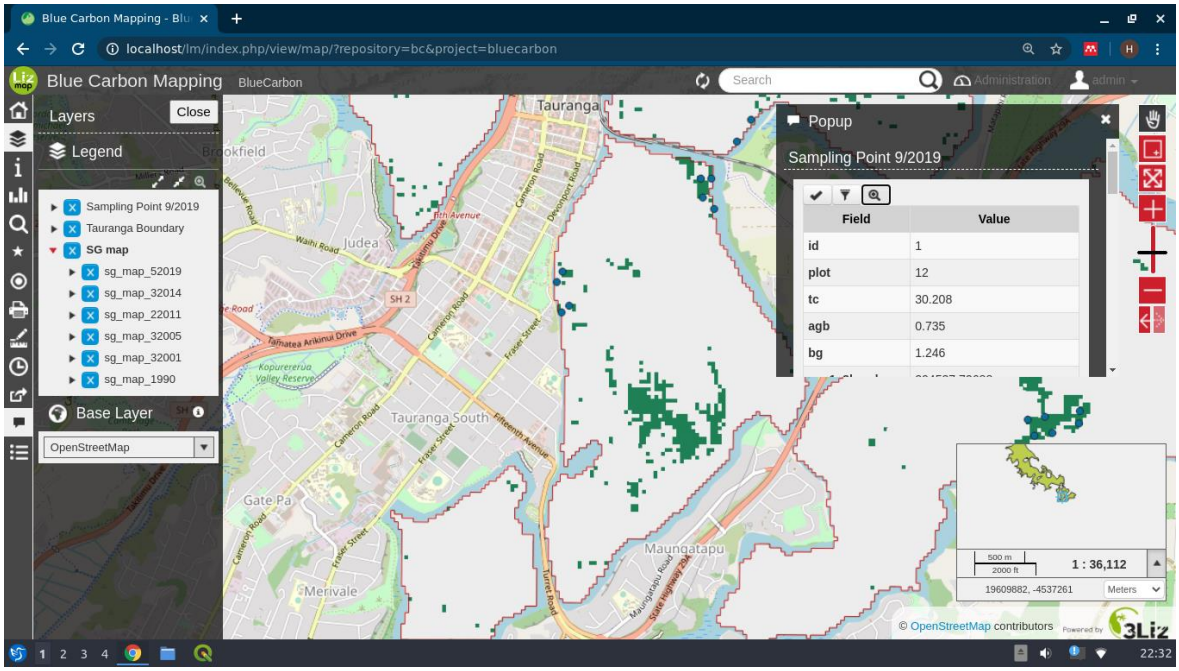


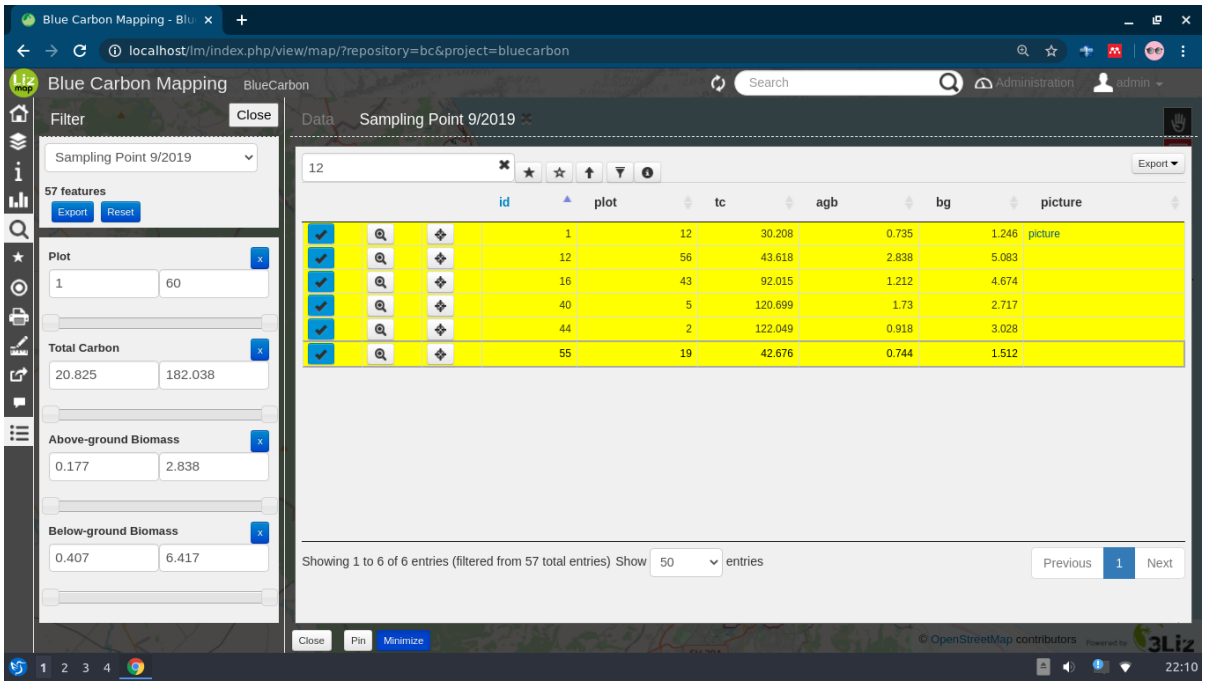
Figure 6.7. Pop-up window for data point in BCW

Finally, a time series animation is another function, which supports a continuous and attractive presentation of the maps at various time points.

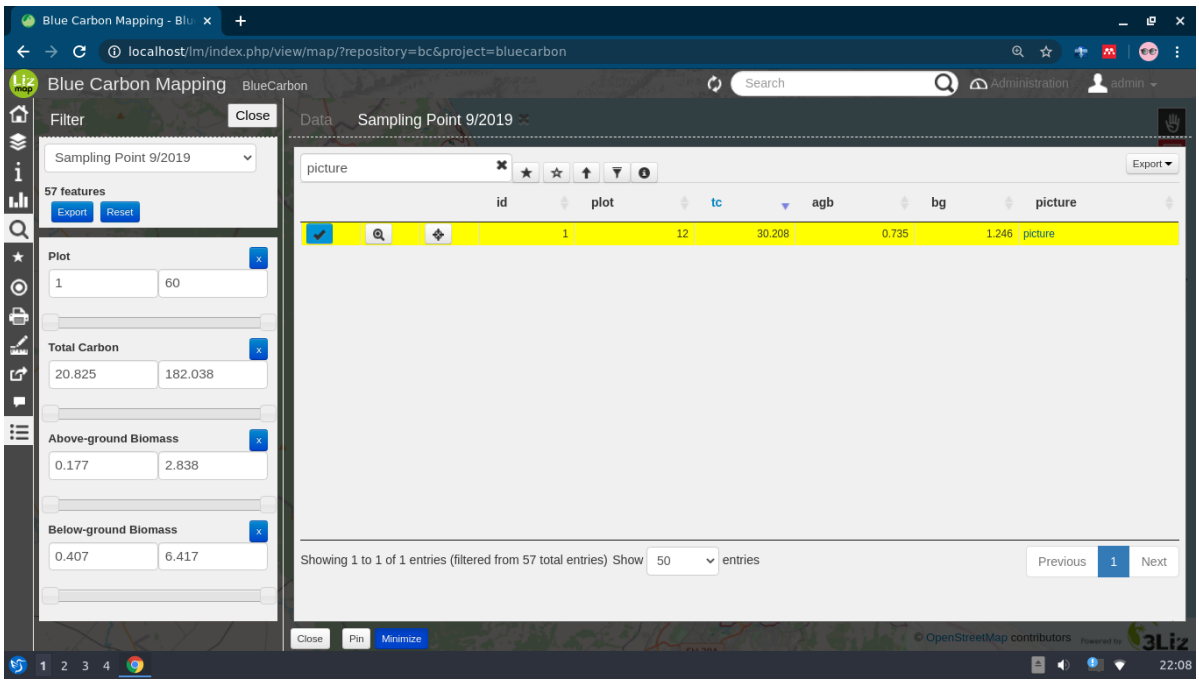
Data query and filtering

Data query

Recently, BCW supports a query protocol using the search box in an attribute table (Figure 6.8). This search box enables the user to find the desired information in both format of text and number.



(*a*)



(*b*)

Figure 6.8. Attribute query using number and text in BCW using (*a*) number and (*b*) text query

For a large dataset table, this function is helpful to export a subset of required data instead of the entire dataset. The query is fast, however it only supports a single query without the usage of SQL syntax.

Data filtering

BCW provides a filter for various data fields for imported vector data. It is easy to define a range of different parameters (data fields) in the data table to subset the sampling points from the mapping window (Figure 6.9).

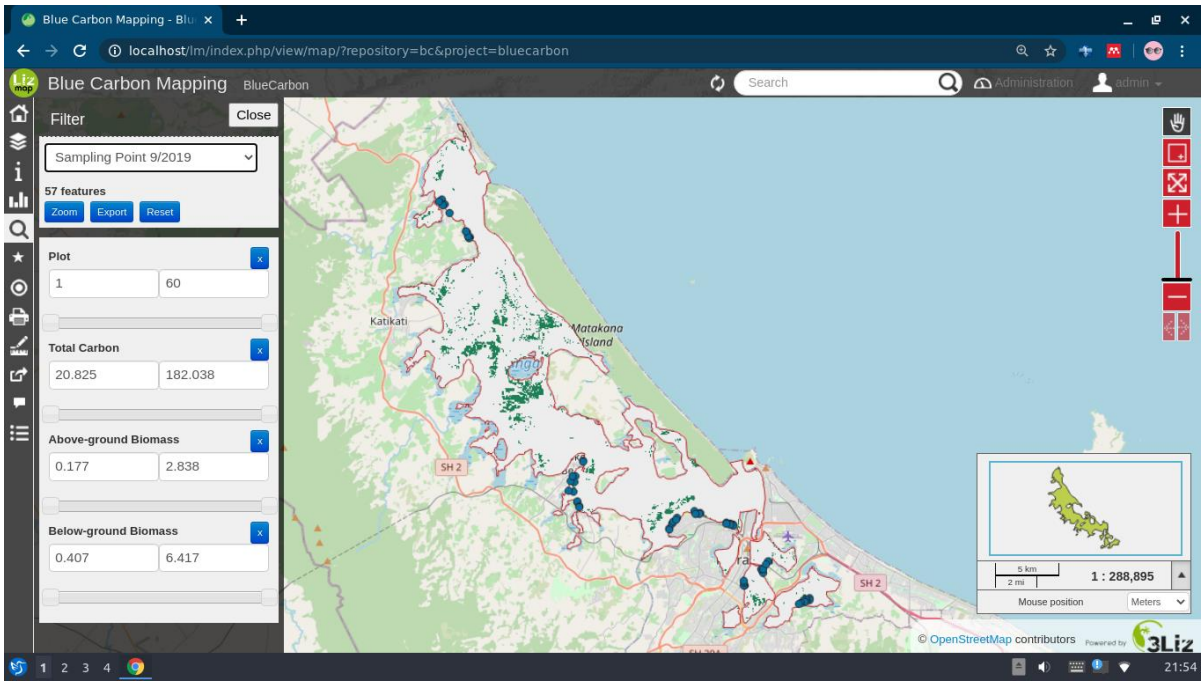


Figure 6.9. Data filter using the filter tool in BCW

All the vector files containing data tables are accessible from a drop list. In addition, BCW provides an export function to open document spreadsheet (ODS) format, which enable the export of the entire data table. This function is effective when the user would like to filter the sampling points for a specific ranges of sampling plots, total soil carbon, aboveground, and belowground biomass.

Data printing

BCW provides a simple solution to print out the desired maps (Table 6.9). The user is able to change the title of the map, set the desired scale, image format and image resolution. The printing function is friendly, supporting a wide range of map scale from the entire region (scale 1: 500,000) to the smallest area (scale 1: 2,000) containing one or two sampling plots. For an easier measurement to work with data on the map, the coordinate reference system (CRS) was set to the world geodetic system (WGS) 84 universal transverse mercator (UTM) 60S (Figure 6.10).

Table 6.9. Printing functions in BCW

Function	Explanation
Scale printing 1:2,000 - 1:500,000	Print the maps at selected scales
Dot per inch (DPI) resolution 100 - 300	The image resolution
Title	Change the title
Format export PDF, JPG, SVG, PNG	The image format to export
Base map Open street map	Print the map with a base map for better visualization

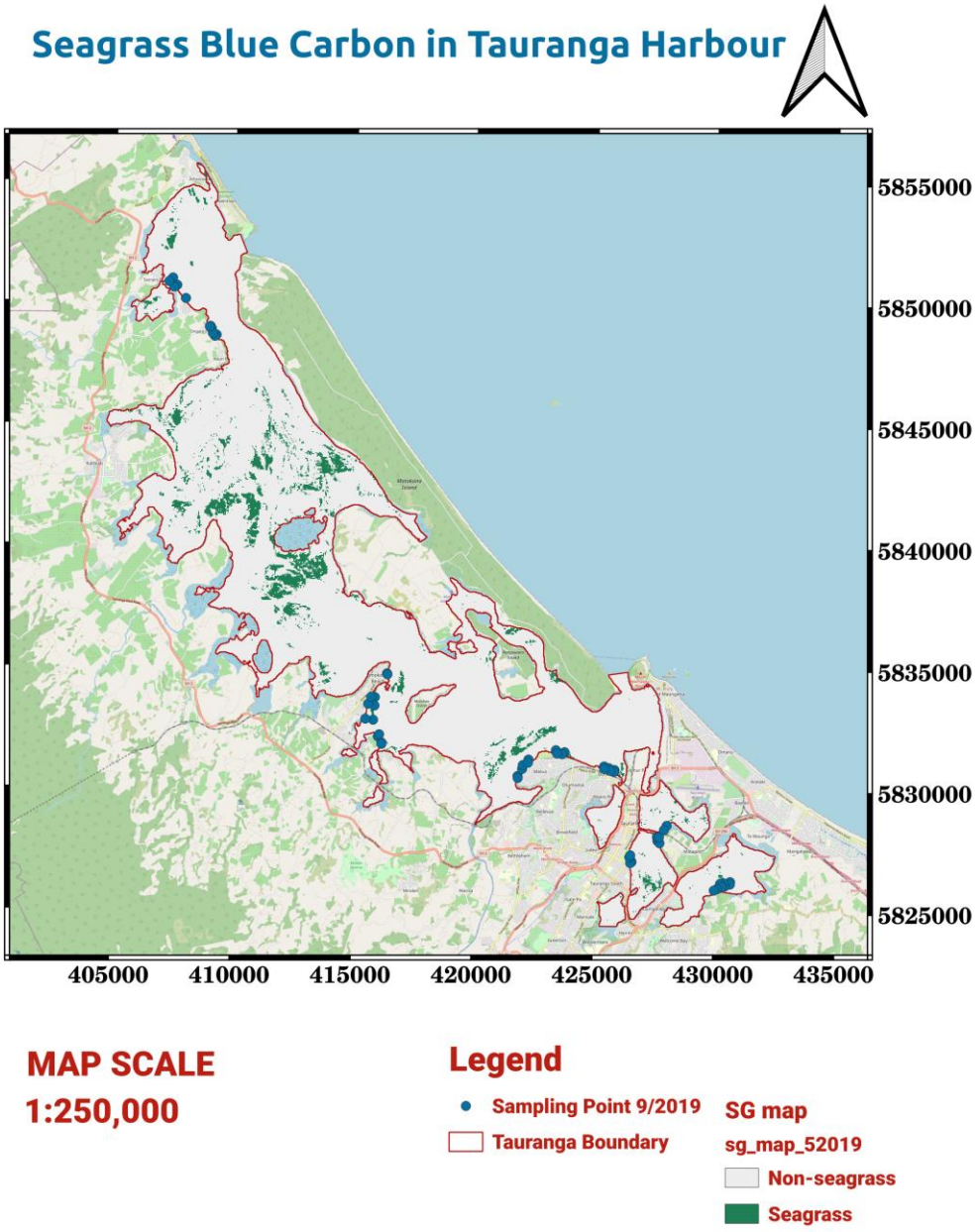


Figure 6.10. An exported map using the printing function in BCW

6.3.4. Geospatial data sharing and collaboration

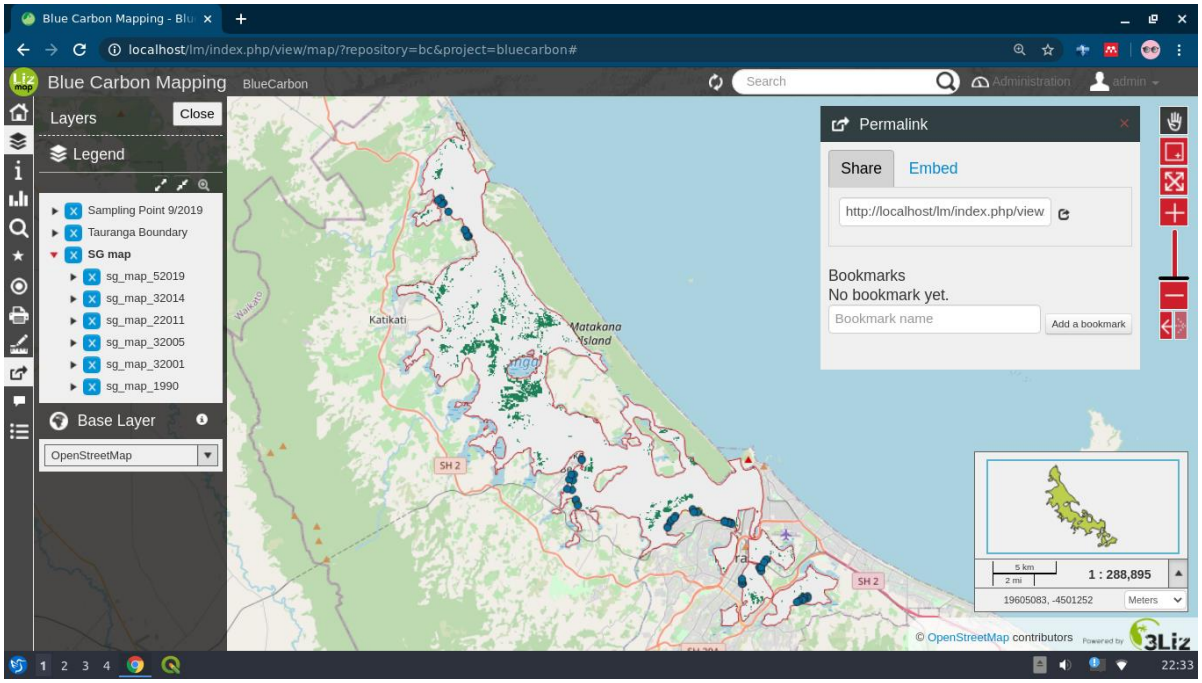
Collaboration is important and is facilitated by the data being updated and shared very quickly via social networks. When accessing the BCW application, the user is able to download and export the original data table and thematic maps from our repository (Table 6.10).

Table 6.10. Downloadable and exportable format from BCW

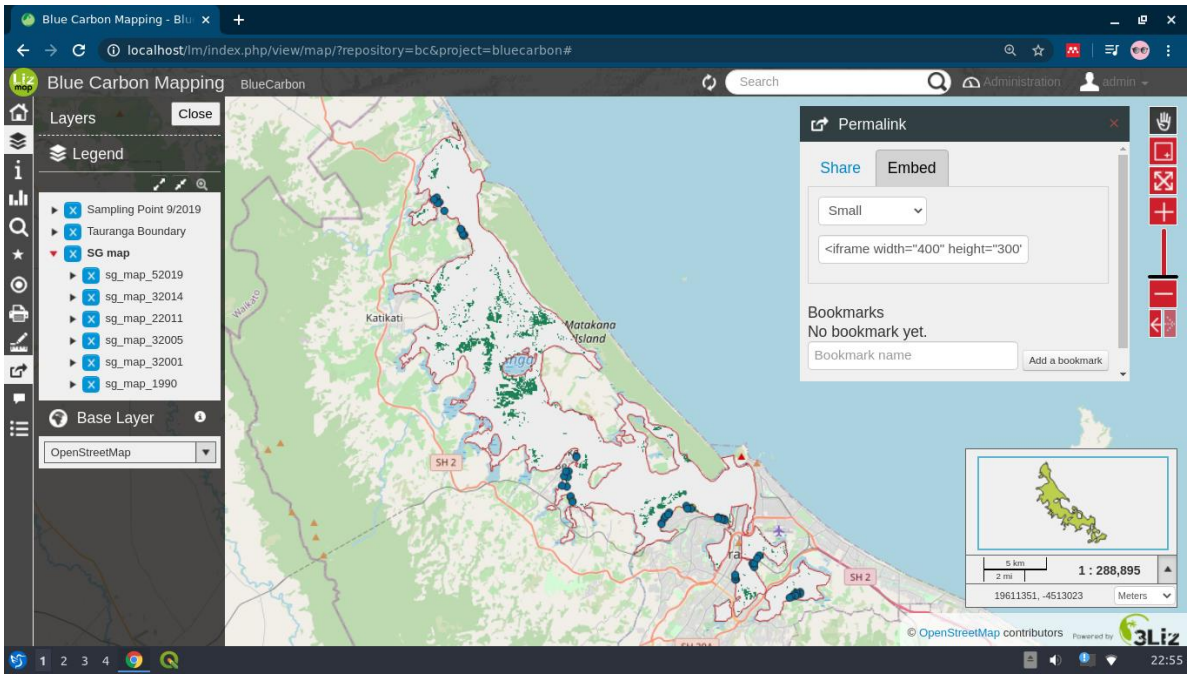
Format	
ODS	Export and download
GeoJSON	Export and download
GML 2 and GML3	Export and download
Image (PNG/JPG/SVG)	Download
PDF	Download

On the other hand, where the user wishes to submit a new dataset or support any update to existing tables in the geospatial database of BCW, this is also facilitated. An automatic upload and management of external data is not currently available. However, it is possible for external users, who have been granted appropriate privileges, to add new data to a specific data table. Data exchange via the administrator's email in the format of shapefile and GeoTIFF for the vector and raster files is also possible, and the administrator can integrate the new data in the existed geospatial database and publish to the BCW.

BCW provides the option of creation of permanent links using a sharable link or a HTML code to various web platforms (Figure 6.11). This function allows the user to instantly share the desired content to other partners or embed into different social platforms to improve social interaction and facilitate information updates.



(*a*)



(*b*)

Figure 6.11. Web link sharing (*a*) and HTML code (*b*) embedding in BCW

6.4. Discussion

This study describes an integrated and simple approach to stimulate data sharing and collaboration and present scientific data to a broader community in a readily assimilable format. The study aims to improve the awareness of the importance of seagrass conversation for not only in Tauranga Harbor, New Zealand but is applicable across various regions in the world. Several web-GIS applications have been developed for various types of data ranging from soil (https://powellcenter-soilcarbon.github.io/soilcarbon/), marine resources (http://maps.oceanwealth.org/#) to forest conversation (http://carbonstock.cifor.org/user/HomeMap), however only a few existing applications include seagrass data or have been designed entirely for blue carbon ecosystems.

Comparing to other web-GIS application, the BCW has the advantage of being based on open source software, is lightweight, easy to install and configure, and can adapt to various levels of data, from a few megabytes of vector and raster files to gigabytes of complex data structure with hundreds of data files (for example: hierarchical data format (HDF), network common data form (NetCDF)). Our open geodatabase provides a platform for continuous and long-term assessment of seagrass changes in the harbor, seagrass aboveground and belowground biomass, and seagrass total soil carbon which can be scaled up to the entire country, motivating community-based interactions with data. Given the variety of designed functions, the BCW is unique for the seagrass ecosystem with an opening geodatabase and open source installed applications, is a well-designed interface and is good in data quality with additional data of seagrass dynamics and blue carbon ((Ha et al., 2020), Ha et al., in revision (Chapter 4)). Importantly, the website has a friendly interface, simple sharing and printing functions which optimize the user's experiment and assist the spreading of information within the community. All the maps are accessible via

direct layer management which enables show/hide for a specific layer, the essential buttons for map viewing are located on the right side of the window and leave the center part for the map to maximize the viewing screen. The simple but full features BCW interface brings a feeling of a GIS desktop software to the end user. Moreover, the printing and embedding options provide an efficient way to share beautiful maps or selected data to other users, stimulating interest in seagrass protection as well as raising awareness of seagrass conversation in the wider community.

Since our BCW application used the state-of-the-art open source GIS (QGIS, QGIS server) and database manager (PostgreSQL, PostGIS) together with a compelling emerging web client (Lizmap), the cost is only for maintaining the world wide web (WWW) server and collecting more field data in the future. In addition, this web-GIS system is capable of scaling up to any desired management levels, from a small local area to large-scale regions in New Zealand. Importantly, our web-GIS system is easily integrated to various open source web-GIS projects in the world using the QGIS application for mapping/ map server and PostgreSQL/PostGIS for database management/ connection that would strengthen a joint collaboration on blue carbon ecosystems mapping, investigation, and data sharing globally.

Despite of a successful performance of BCW, the application may come with some unavoidable limitations. First, the BCW is running at a localhost level and is not entirely published through the WWW network due to the renting fee constrain of the web-GIS server. This drawback might be solved with a donation from the community to maintain an annually renewable web-GIS server. Second, despite well designed functions and web interface, Lizmap web client still lacks the capacity to handle advance data queries using the SQL syntax, and instant connection to various social networks. However, Lizmap web client is a community driven project and similar

to PostgreSQL, QGIS, thus progress is controlled and development is secured. For this reason, it is expected that more functions will be included and will make the web-GIS application better at handling real time and complex data.

6.5. Conclusion

Collection of data on seagrass ecosystem services and the impact of loss of this habitat is important and necessary to improve our understanding of the resources' contribution to the strategy of climate change mitigation. However, the need to share these data with non-scientists is becoming more urgent as it is necessary to improve popular awareness of protection and conservation of extant seagrass meadows in the world. In the era of social networks, web-GIS is emerging as an effective toolset for geospatial data visualization, query, instant sharing, and strong collaboration in the community.

In this study, we have proposed an integrated solution for publishing geospatial data via a protocol of web-GIS. The open source applications, which are used PostgreSQL, QGIS, QGIS server, PostGIS, and Lizmap web client are integrated in the Linux environment to create a localhost web-GIS application, BCW (Blue Carbon Web), which aims at publishing all seagrass data in Tauranga Harbor, New Zealand. Using PostgreSQL as the database manager, we have created the seagrass geospatial database with the data of seagrass historic distribution, sampling plots, dry aboveground and belowground biomass, and total carbon acquired during field work in the austral winter 2019 and summer 2020. The external extension PostGIS was used to connect the database to the mapping application QGIS and then published all formatted data to the web via Lizmap web client.

The localhost BCW provides diverse functions, ranging from thematic map visualization, data query and filtering, data sharing and collaboration, to the quick printing of various GeoTIFF

images. BCW is simple and elegant in the web interface, function richness, lightweight, yet still secured for long-term development. Currently, BCW is sufficient as the web-GIS to visualize geospatial data, and permit instant sharing of the recent conditions of extant seagrass in Tauranga Habour. The beautiful maps and simple interaction will support a fast and reliable platform to spread the data to various partners, improving popular awareness and make the conservation of this resource more efficient in the future.

6.6. Chapter supplemental material

Installation instructions for the application of QGIS server, Lizmap plugin for QGIS, Lizmap web client, PostgreSQL, and QGIS. The codes and installation are implemented in the Linux environment (Lubuntu 20.04 distribution).

QGIS server

Instructions to install and configure the QGIS server:

```
# Step 1: Install Apache2
sudo apt install apache2
sudo /etc/init.d/apache2 restart
# or use this line if the line above not work
sudo service apache2 restart
# Step 2: Install QGIS server
sudo apt install qgis-server libapache2-mod-fcgid
# Step 3: Activate the server services
sudo a2enmod fcgid
sudo a2enconf serve-cgi-bin
# Step 4: Restart apache2 service
sudo service apache2 restart
# Step 5: Add the following code
ScriptAlias /cgi-bin/ /usr/lib/cgi-bin/
<Directory "/usr/lib/cgi-bin/">
Options ExecCGI FollowSymLinks
Require all granted
AddHandler fcgid-script .fcgi
</Directory>
# into /etc/apache2/sites-available/000-default.conf and save the changes.
# Replace gedit with any text editor:
sudo gedit /etc/apache2/sites-available/000-default.conf
```

After adding the lines, the configuration file will be like this:

```
<VirtualHost *:80>
    # The ServerName directive sets the request scheme, hostname and port that
    # the server uses to identify itself. This is used when creating
    # redirection URLs. In the context of virtual hosts, the ServerName
    # specifies what hostname must appear in the request's Host: header to
    # match this virtual host. For the default virtual host (this file) this
    # value is not decisive as it is used as a last resort host regardless.
    # However, you must set it for any further virtual host explicitly.
    #ServerName www.example.com
    ScriptAlias /cgi-bin/ /usr/lib/cgi-bin/
    <Directory "/usr/lib/cgi-bin/">
    Options ExecCGI FollowSymLinks
    Require all granted
    AddHandler fcgid-script .fcgi
    </Directory>

    ServerAdmin webmaster@localhost
    DocumentRoot /var/www/html

    # Available loglevels: trace8, ..., trace1, debug, info, notice, warn,
    # error, crit, alert, emerg.
    # It is also possible to configure the loglevel for particular
    # modules, e.g.
    #LogLevel info ssl:warn

    ErrorLog ${APACHE_LOG_DIR}/error.log
    CustomLog ${APACHE_LOG_DIR}/access.log combined

    # For most configuration files from conf-available/, which are
    # enabled or disabled at a global level, it is possible to
    # include a line for only one particular virtual host. For example the
    # following line enables the CGI configuration for this host only
    # after it has been globally disabled with "a2disconf".
```

#Include conf-available/serve-cgi-bin.conf

</VirtualHost>

Then, we restart and test the server to ensure the server working.

Step 6: Restart the server

sudo service apache2 restart

Step 7: Test the 'GetCapabilities' of the web server by run this line in the web browser:

http://localhost/cgi-bin/qgis_mapserv.fcgi?SERVICE=WMS&VERSION=1.3.0&REQUEST=GetCapabilities

The following figure (Figure 6.S1) indicated that the QGIS server was successfully installed and configured.

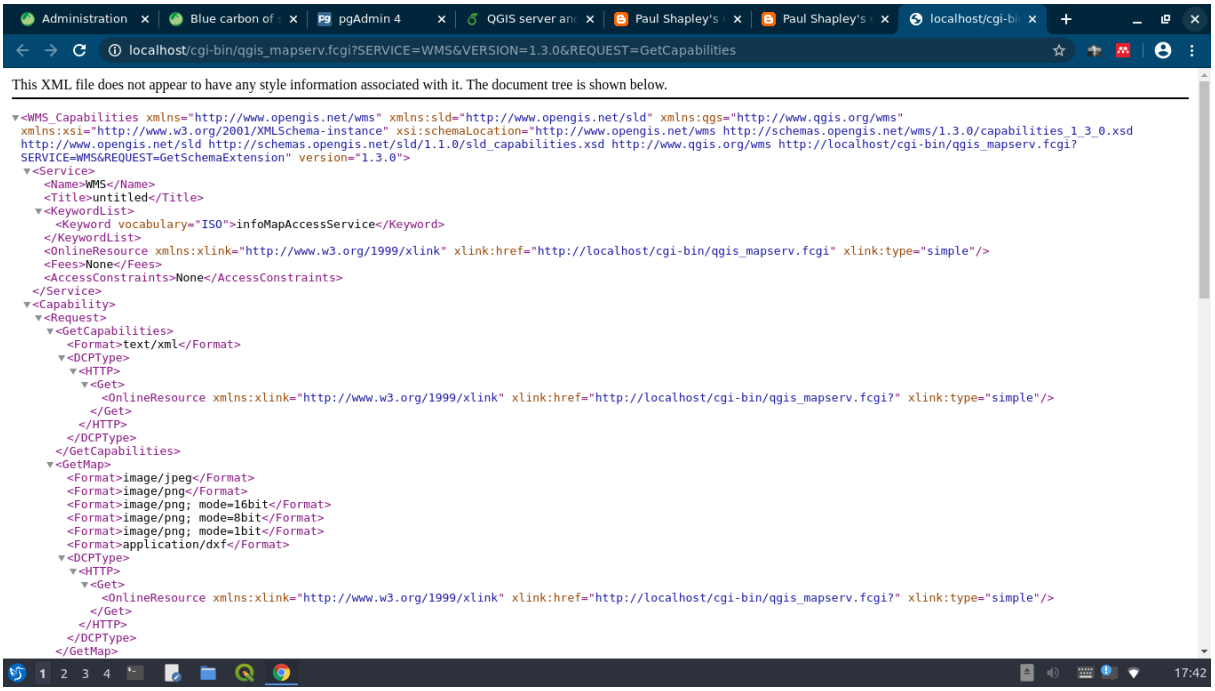


Figure 6.S1. Test the installation and configuration of QGIS server

Add the ”project” folder in ”/usr/lib/cgi-bin/project” and create the symlinks of the ’wms-metadata.xml’ and the ’qgis-mapserv.fcgi’ as well as link the ’projectname.qgs’ file (the QGS file created by QGIS and saved in “/home/user/qgis/project”) to the “/usr/lib/cgi-bin/project” folder. The following lines will do the tasks.

```
# Step 8:
cd /usr/lib/cgi-bin/project
sudo ln -s ../qgis_mapserv.fcgi .
sudo ln -s ../wms_metadata.xml .
# The letters after “sudo” are considered as one (1) line:
# Step 9:
sudo ln -s /home/user/project/projectname.qgs /usr/lib/cgi-bin/project/projectname.qgs
sudo ln -s /home/paul/project/projecname.qgs.cfg /usr/lib/cgi-bin/project/projectname.qgs.cfg
# Step 10: Restart Apache2 server
sudo service apache2 restart
```

Lizmap QGIS plugin and Lizmap web client

Lizmap QGIS plugin

In QGIS, Lizmap is designed as a plugin, which aims to connect the data in QGIS to the QGIS server and display in the web browser. The Lizmap plugin is installed using the plugin manager in QGIS. Figure 6.S2 indicates a successful installation of the plugin.

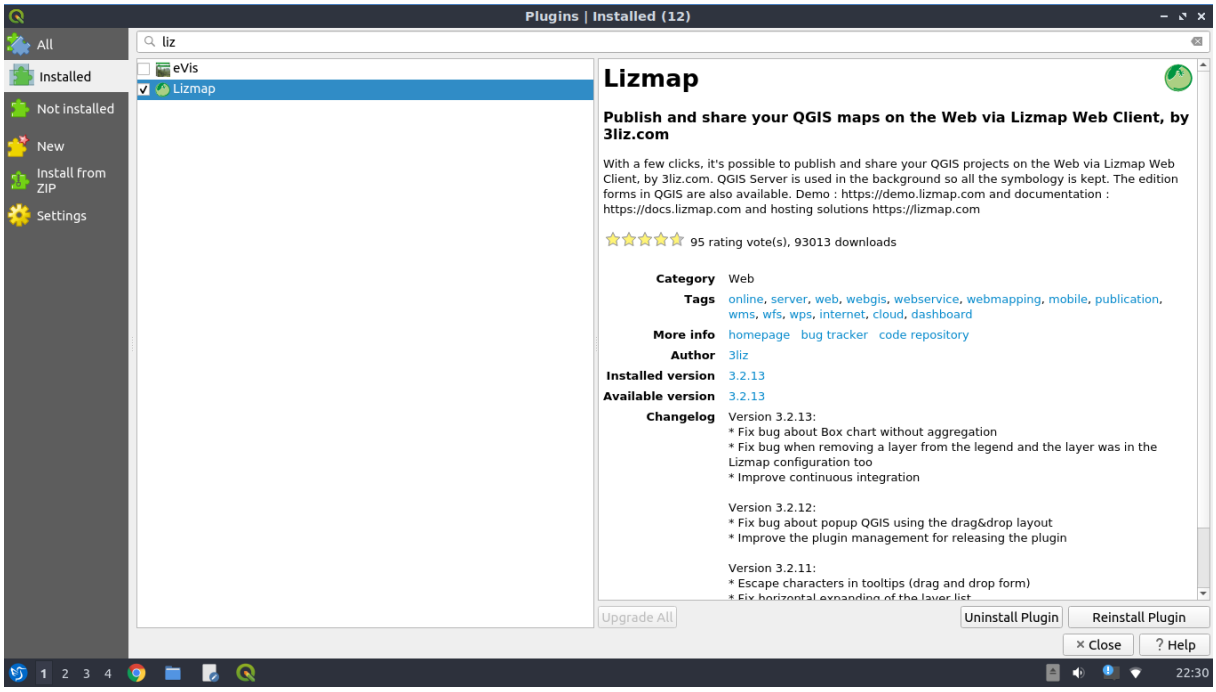


Figure 6.S2. Lizmap plugin installation in QGIS plugin manager

Lizmap web client

The following codes will download and run the installation of Lizmap web client.

Step 1: Go to a specific folder. The downloaded files will stay in this folder:

cd /home/user/.../folder

Step 2: Download the source code of Lizmap web client from master branch:

git clone https://github.com/3liz/lizmap-web-client

Step 3: Copy the downloaded client into '/var/www':

sudo cp lizmap-web-client /var/www

Step 4: Link the lizmap client into the apache2 server 'html' directory to make the service is available for web browser:

sudo ln -s /var/www/lizmap-web-client/lizmap/www/ /var/www/html/lm

cd /var/www/lizmap-web-client

sudo lizmap/install/set_rights.sh www-data www-data

cd /var/www/lizmap-web-client/lizmap/var/config

sudo cp lizmapConfig.ini.php.dist lizmapConfig.ini.php

sudo cp localconfig.ini.php.dist localconfig.ini.php

sudo cp profiles.ini.php.dist profiles.ini.php

cd ../../..

Step 5: Install required python php packages (the command after “sudo” is understood as one (1) line):

sudo apt install xauth htop curl libapache2-mod-php7.4 php7.4-cgi php7.4-gd php7.4-sqlite3 php7.4-xml php7.4-curl php7.4-xmlrpc php7.4-pgsql python-simplejson python-software-properties

Step 6: Install the lizmap web client

sudo php lizmap/install/installer.php

The output should be like this:

sudo php lizmap/install/installer.php

Installation start..

[notice] Installation starts for the entry point index

All modules dependencies are ok

Module jelix installed

Module jauthdb_admin installed

Module master_admin installed

Module jacl2 installed

Module jacl2db installed

Module jacl2db_admin installed

Module jauth installed

Module jauthdb installed

Module admin installed

Module lizmap installed

Module view installed

Module proj4php installed

All modules are installed or upgraded for the entry point index

[notice] Installation starts for the entry point admin

All modules dependencies are ok

Module jelix installed

Module jauthdb_admin installed

Module master_admin installed

Module jacl2 installed

Module jacl2db installed

Module jacl2db_admin installed

Module jauth installed

Module jauthdb installed

Module admin installed

Module lizmap installed

Module view installed

Module proj4php installed

All modules are installed or upgraded for the entry point admin

[notice] Installation starts for the entry point script

All modules dependencies are ok

Module jelix installed

Module jauthdb_admin installed

Module master_admin installed

Module jacl2 installed

Module jacl2db installed

Module jacl2db_admin installed

Module jauth installed

Module jauthdb installed

Module admin installed

Module lizmap installed

Module view installed

Module proj4php installed

All modules are installed or upgraded for the entry point script

Installation ended.

Step 7: Restart the server:

sudo service apache2 restart

Step 8: Set the rights for user to the folder:

cd /var/www/lizmap-web-client

sudo chown :www-data temp/ lizmap/var/ lizmap/www -R

sudo chmod 775 temp/ lizmap/var/ lizmap/www -R

Step 9: Restart the server:

sudo service apache2 restart

Step 10: Test the working of Lizmap web client by typing this line in a web browser:

http://localhost/lm

During the installation, an error regarding the vendor and autoload.php may appear. The solution is to install the package composer and then deploy this package into lizmap.

Step 1: Install composer:

sudo apt install composer

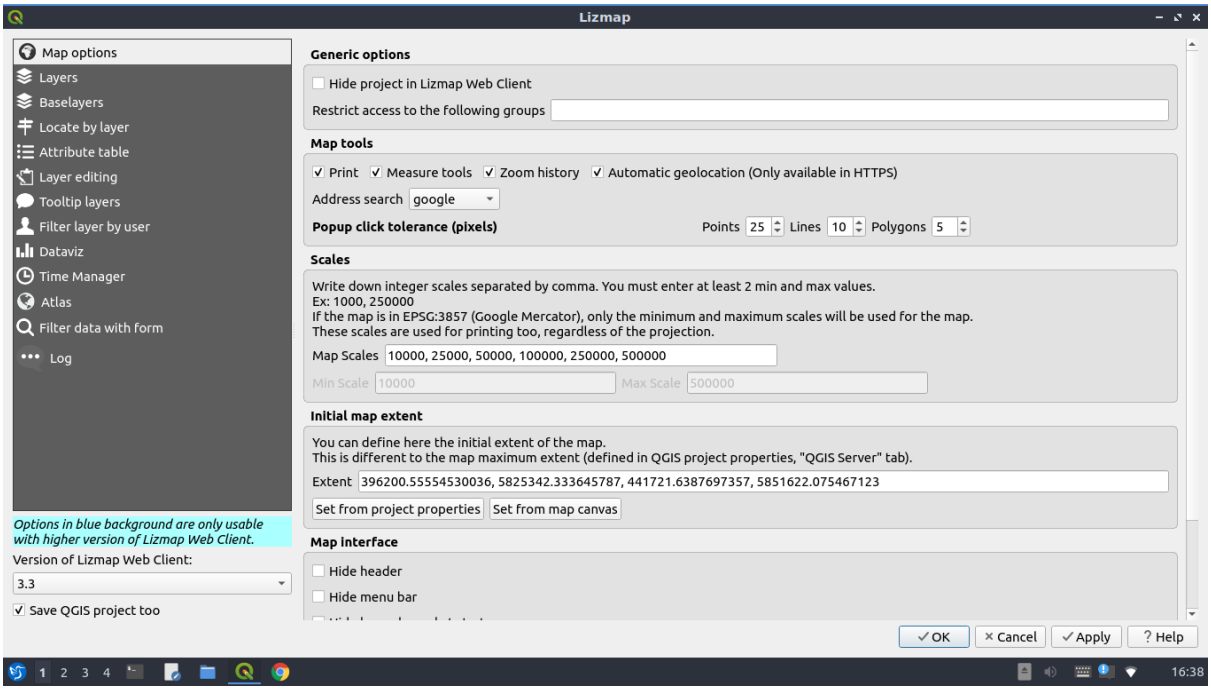
Step 2: deploy composer into lizmap:

cd /path_to_lizmap/lizmap-web-client-x.x/lizmap

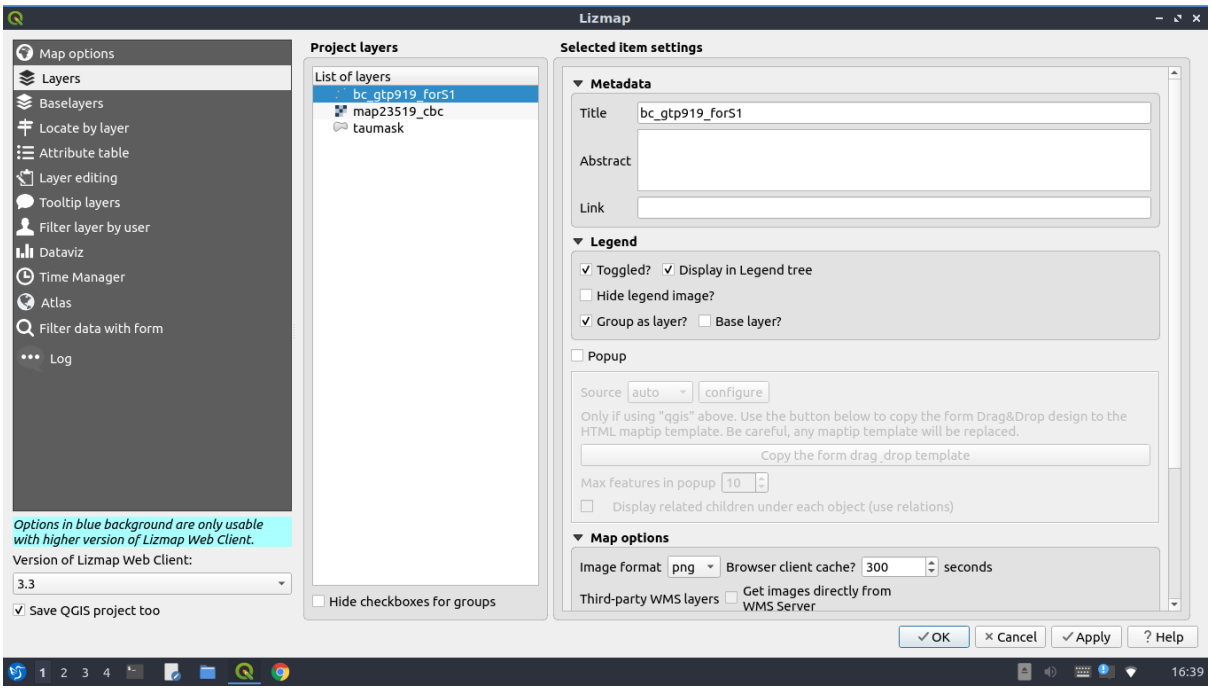
sudo composer install

Configure lizmap web client

The lizmap web client was configured in the administrator page (Table 6.S1) and QGIS plugin (Figure 6.S3).



(*a*)



(*b*)

Figure 6.S3. Lizmap QGIS plugin configuration with options in Map (*a*) and Layers (*b*)

Table 6.S1. Configuration parameters in Lizmap administration page

Parameters	
Interface	Set as default
Email	Set as default
Projects	Leave as blank
Cache	Set as default
QGIS server	
QGIS server version	>= 3.0
Other parameters	Set as default
System	Set as default

Finally, we create a new repository to store the QGIS project's information. This can be done from the dashboard of the Lizmap administrator page inside the Maps Management option (Table 6.S2).

Table 6.S2. Configuration parameters in Lizmap map management

Data configuration	Explanation
Label	Set a name
Local folder path	Path to QGIS project (.QGS format)
Allow theme/javascript code for this repository	Yes
Rights and granted groups	Set to the admins right

PostgreSQL

Install PostgreSQL

```
# Step 1: Import the GPG key using for PostgreSQL packages:
sudo apt-get install wget ca-certificateswget --quiet -O - https://www.postgresql.org/media/keys/ACCC4CF8.asc | sudo apt-key add -
# Step 2: Add the repository to the system:
sudo sh -c 'echo "deb http://apt.postgresql.org/pub/repos/apt/ `lsb_release -cs`-pgdg main" >> /etc/apt/sources.list.d/pgdg.list'
# Step 3: Install PostgreSQL:
sudo apt-get update
sudo apt-get install postgresql postgresql-contrib
```

At this step, PostgreSQL is installed with a default user name and database as *postgres* and *postgres*. To manage the database, user privilege, and extension, the application PgAdmin4 is installed and used as the front-end manager.

```
# Step 1: Install PgAdmin4:
sudo apt-get install pgadmin4 pgadmin4-apache2
# Step 2: Access PgAdmin4:
http://localhost/pgadmin4
```

In addition, the extensions *postgis* and *postgis_raster* are installed using PgAdmin4 to add the data in the format of Shapefile and raster into PostgreSQL database. In this study, we used both default user (*postgres*) and database (*postgres*) without a password attached. This setting aims to make the demonstration faster and ease to access from a level of local host for a demonstration purpose.

Configure PostgreSQL

We used PgAdmin4 to configure all required parameters (Table 6.S3) to active the database and connect to QGIS application.

Table 6.S3. Parameters configuration in PgAdmin4

Parameter		Explanation
Server	Tauranga	The server to manage all user and database
Database	Postgres	The database
User	Postgres	The user to do all the tasks regarding management database
Extension	Plpgsql	To manage basic task in PostgreSQL
	Postgis	To import and manage vector data
	postgis_raster	To import and manage raster data

PostGIS

Since we installed PostgreSQL version 12, the following code was used to install PostGIS:

```
# Step 1: Install PostGIS:
sudo apt install postgis postgresql-12-postgis-3
```

Connect PostgreSQL database to QGIS

To make the thematic maps in QGIS, vector and raster data are required to be imported from a PostgreSQL database. Technically, the connection is established using a PostgreSQL's external extension, PostGIS. By connecting the PostgreSQL database to QGIS via PostGIS using parameters in Table 6.S4, we are able to import, export, and create thematic maps inside the QGIS application.

The following codes are used to add the raster data to the geospatial database:

```
# add raster data in PostgreSQL
raster2pgsql -s 32760 /path/to/raster_data/raster.tif | psql postgres
# where
# postgres is the name of the geospatial database
# -s 32760 is the coordinator reference system of the raster datase
```

Table 6.S4. PostGIS parameters connection to QGIS

Parameter		Explanation
Name	Postgres	The user is with full privilege of the database
Host	localhost	The localhost
Port	5432	Default port to access to the database
Database	Postgres	The name of the database
SSL mode	Disable	No SSL mode
Authentication	No authentication	No user/password requirement

QGIS

There is no specific codes for QGIS installation. Instead, the user is encouraged to use the official guidance from the QGIS website for a Linux distribution (https://qgis.org/en/site/forusers/alldownloads.html#debian-ubuntu).

Chapter 7 Synthesis and future research

7.1. Seagrass dynamics and conservation from space – the state-of-the art challenges

In the last decade, coastal zones significantly contribute to the development of coastal and ocean economies, with the intensification of agriculture, tourism, urbanization, and aquaculture (He et al., 2014; Wenhai et al., 2019). The development of economic activities has significantly benefited the community, however they have led to several unsolved challenges of water pollution, over population, aquatic resource depletion, and coastal habitat degradation (He et al., 2014; de Alencar et al., 2020). In particular, many reports have provided evidence of seagrass degradation in terms of area and habitat quality (Waycott et al., 2009; de los Santos et al., 2019). This has in turn led to a significant reduction in ecosystem services of provision of breeding and nursery ground, coastal zone protection, and carbon sequestration (Marbà et al., 2015; Githaiga et al., 2019), and might result in an unsustainable social-economic impact in various regions of the world (Ferrol-Schulte et al., 2013; Cahaya, 2015). Several efforts to address this issue have been considered and are encapsulated in the initiative of Coastal Blue Carbon, which aims to protect, conserve, and move toward a sustainable maintenance of coastal ecosystem services (Hejnowicz et al., 2015). To do this, the strategy of monitoring, reporting, and verification (MRV) has been advocated (Herold & Skutsch, 2011) to allow updating of the most recent situation and to facilitate a fast and reliable response policy. The first part of this strategy, monitoring, includes mapping and assessment of ecosystem changes using space-born satellite images. This has been to some extent successful but has left unsolved challenges that require further studies, some of which are outlined below, and which form the motivation of this thesis.

Mapping the spatial and temporal distribution of seagrass meadows using remotely sensed imagery differs from mapping of terrestrial ecosystems. This is due to the differences between

the optical properties of the aquatic and terrestrial environments, the phenology of the vegetation, and the anthropogenic stressors specific to aquatic systems, particularly fine sediment pollution, that impact the ability to observe underwater meadows. This leads to the need for additional steps in image processing for atmospheric and water column correction, and the use of advanced methods for image classification and retrieval models that are better suited to the more complex task of remote estimation of vegetation in an aquatic, often tidal, environment (Hossain et al., 2015; Dat Pham et al., 2019).

The impact of the water column and the mixed substrate usually result in misclassification and low confidence in seagrass mapping when using traditional methods. Several attempts to enhance accuracy using very high resolution (VHR) image (Poursanidis et al., 2018), increasing ground truth points for parametric classification methods (Koedsin et al., 2016) or application of a narrow range of machine learning (ML) models have been made (Bonin-Font et al., 2016; Mohamed et al., 2018; Poursanidis et al., 2019), however the accuracy thus obtained varies significantly varied depending on the water depth and the cover of seagrass meadow. In addition, spaceborn VHR images only cover a small area at very high cost, leading to a limited number of VHR based mapping applications worldwide (Dat Pham et al., 2019).

Broad-scale seagrass change assessment, on the other hand, requires a long-term operated satellite platform combined with an advanced classification technique to deal with the variation in environment conditions at various historical scales. It is still not an easy task, and the current research literature contains only a modest number of studies dedicated to long-term change detection (30 - 40 years) (Dat Pham et al., 2019). Attempts at temporal mapping of seagrass meadows also reveals a fundamental drawback of multi-spectral imagery, its requirement of cloud free acquisition, which significantly reduces the number of available satellite images

(Chapter 4). Moreover, the use of classical learning models (*e.g* parametric model maximum likelihood classifier (MLC)) during the classification process has not delivered high accuracy and low uncertainty in the change evaluation (Chapter 3). MLC performs well when working with healthy seagrass meadows in stable water environments, however, underperforms in classification accuracy when dealing with degraded (scattered and low coverage) meadows (Ha et al., 2020).

When estimating aboveground biomass (AGB), traditional linear models result in a very wide variation of R^2 values (0.26 - 0.99) for seagrass AGB estimation (Roelfsema et al., 2014; Koedsin et al., 2016), depending on the seagrass species and the coverage. VHR imagery was successfully applied for AGB retrieval with large size seagrass species (*e.g. Enhalus acoroides*) (Koedsin et al., 2016), however only achived a very low accuracy (R^2 0.26) when dealing with the smaller species (*e.g. Zostera muelleri*) (Roelfsema et al., 2014). Improving seagrass AGB estimation accuracy is very necessary since the AGB data contributes to the precise evaluation of seagrass blue carbon globally (Howard et al., 2014; Susi et al., 2019) and might lead to significant changes in conservation policy (Chapter 5).

In recent years, web based social media that allowed widespread access to current research information was recognized as an effective toolset to fill the gap of social interaction in conservation science (Toivonen et al., 2019) and seagrass conservation in particular. Several attempts have been made to bring scientific data to a wider spectrum of audiences via various social platforms (Di Minin et al., 2015; Wu et al., 2018). Nonetheless, this movement is spreading only slowly, in comparison with the rapid change of many seagrass ecosystems at the moment (Orth et al., 2006; Unsworth & Coles, 2014; Unsworth et al., 2018). Developing a low cost, effective social web based application, would greatly conservation awareness and support a

community based management of seagrass resource in the future (Toivonen et al., 2019) (Chapter 6).

7.2. Novel approaches for accurate mapping of seagrass dynamics and seagrass conservation

This thesis describes the development of novel methods to solve the problems of accurate mapping of seagrass dynamics and proposes a social web-based approach to foster community-based conservation of seagrass ecosystems.

In Chapter 3, the performance of a range of ML models was assessed to provide a preferred solution for the mapping of dense (> 80% cover) and sparse (< 80% cover) seagrass meadows from S-2 imagery. A wide range of decision tree based ML models (*i.e* Random Forest (RF), Rotation Forest (RoF), and Canonical Correlation Forest (CCF)) were tested in comparison to the traditional approach (MLC), resulted in a high precision mapping (F_1 0.91 for dense and F_1 0.75 for sparse meadows, respectively) using the non-parametric RoF model, and much improved to the parametric MLC (F_1 0.75 for dense and F_1 0.50 for sparse meadows).

The successful implementation of the open source ML models for seagrass mapping significantly benefits the research community. Firstly, the rapidly increasing number of ML models available helps to diversify the model selection, and gives better fitting with the mapping of seagrass meadows in degraded condition. Secondly, ML models can be effectively trained/undergo self-learning using various spatial and temporal datasets and potentially resulting in higher accuracy of the classification/ regression (Ha et al., 2020; Pham, Le et al., 2020; Pham, Yokoya et al., 2020). In addition, the open source license of ML algorithms supports the opportunities for developers to integrate the algorithms into different programming languages (*e.g* Python programming language), to build simple but effective toolsets for non-programming researchers.

The integration with free-of-charge medium resolution satellite images (*e.g* the multi-spectral Sentinel-2 (S-2) imagery) will improve the spatial resolution of mapped seagrass and diversify the options in seagrass mapping.

It is, however, noted here that this approach is challenged by the presence of mixed pixels due to conflicts between image pixel size (10 m for S-2) and meadow patch size (1 - 50 m), such that seagrass and sand were mixed in many image pixels (*i.e* sub-pixel mixture). Because of this issue, we have to accept a lower threshold of mapping accuracy (up to 91% of F_1 score using RoF model in this study) using S-2 imagery. On the other hand, greater accuracy could be expected through application of commercial VHR images for the mapping using the most recent generations of Quickbird, Ikonos, or Worldview, highlighting potential considerations of conflicts between spatial coverage, mapping details, mapping accuracy, and available budgets.

The ML, tree based models were useful in high precision mapping of seagrass meadows at different coverages. To extend the application of ML models for seagrass change assessment, two novel models, Extreme Gradient Boost (XGB) and CatBoost (CB), were compared to the more traditional RF and Support Vector Machine (SVM) models to map the temporal changes (29 years) of seagrass meadows from the long-life operated Landsat imagery (Chapter 4). The CB model outperformed other models, and delivered high and consistent scores of F_1 (0.96 - 0.98) during the historical mapping. Here, the results contribute a solid mapping framework for the task of long-term change assessment using the free-of-charge Landsat image which has been secured in operation since 1972 (Roy et al., 2014) and advanced ML models which potentially improve the mapping accuracy and certainty in seagrass change evaluation.

The practical implementation of multi-temporal mapping illustrates the usefulness of the multi-spectral satellite sensors in assisting long-term change assessment. However, the requirement of

cloud free acquisition emerged as a vital drawback and significantly constrained the number of images for various study sites. To overcome this challenge and for the first time, the Sentinel-1 (S-1) synthetic aperture radar (SAR) imagery was tested for spatial mapping of seagrass meadow in the intertidal zone of the harbor. In addition, the S-1 image was combined with S-2 to validate the hypothesis of improving accuracy of seagrass AGB estimation. With the support of ML XGB model, seagrass distribution was successfully mapped using only S-1 image with a surprisingly high level of confidence (F_1 0.92). The implication here is that change detection can be supported by combining use of S-1 data to fill gaps in the records of optical sensors due to cloud cover. In addition, this thesis shows that S-1 has potential for enhancing seagrass AGB retrieval when combined with optical (S-2) imagery. Despite the small size of the local seagrass species (*Z. muelleri*), the integration of S-1, S-2 images together with the ML model and metaheuristic optimization provided a high accuracy of AGB retrieval (R^2 0.75), which was higher than that achieved using a linear parametric model with VHR imagery (R^2 0.26) (Roelfsema et al., 2014). The proposed methods for seagrass mapping from S-1 and seagrass AGB retrieval using the fusion of both S-1 and S-2 images are reliable and applicable, therefore will diversify the mapping strategy for not only seagrass, but for the blue carbon ecosystems worldwide. It will however be necessary to validate the performance of S-1 imagery for AGB estimation of large size and high biomass seagrass species.

In Chapter 6, the thesis developed an open source, low cost web-GIS application (*i.e* the Blue Carbon Web or BCW) as the social bridge between the scientist and the community and foster a community-based conservation effort for the seagrass ecosystem. The BCW is a lightweight, open hub where people are able to view the spatial data of seagrass dynamics, extract high dimensional information, produce elegant thematic maps for further usage, and easily share

seagrass data via social media. The BCW avoids commercial GIS platforms, thus freeing funds for data collection for ongoing ground-truth validation and laboratory analysis. The only cost attached to the BCW application is the renting fee for a web GIS hosting.

7.3. Future research

The successful implementation of the proposed methods helped to overcome some of the current challenges in accurate mapping of seagrass dynamics and the need to awake community awareness of seagrass conservation. However, a variety of research topics should be considered in the future to further progress these issues in order to better deal with rapid change in the blue carbon ecosystem. In general, the challenges continue to lie in the areas of enhancing precision and accuracy across species and density. This involves both enhanced processing of imagery, but also better ways to develop training and validation datasets for supervised classification systems, including for historic images. Specific issues for future research include:

– Improve the classification accuracy for sparse – very sparse meadows using multi-spectral S-1/S-2 data (10 m in spatial resolution) in integration with the state-of-the-art algorithms for atmospheric correction, water column correction, and classification.

– Validate the ML model performance for various seagrass species and regions to attain the best fitting models.

– Develop the novel accuracy assessment in change detection without using historical validation data.

– Verify the proposed mapping methods and retrieval models for various seagrass species in different estuary/lagoon conditions, using S-1, S-2, and Landsat images.

– Verify the application of state-of-the-art meta-heuristic optimization algorithms (Harris Hawk Optimization, Sparrow Search Algorithm, Firefly Algorithm, Particle Swamp

Optimization, etc.) for feature selection assessment in the retrieval of seagrass biophysical parameters.

– Develop the meta-heuristic optimization Python toolset as the independent library or wrapped component inside the scikit-learn library which would highly benefit the research community in the future.

References

Adriano, B., Xia, J., Baier, G., Yokoya, N., & Koshimura, S. (2019). Multi-source data fusion based on ensemble learning for rapid building damage mapping during the 2018 sulawesi earthquake and tsunami in Palu, Indonesia. *Remote Sensing*, *11*(7), 886. https://doi.org/10.3390/rs11070886

AGEDI. (2015). *The coastal ecosystem mapping and media viability project* (pp. 44–44). https://gridarendal-website-live.s3.amazonaws.com/production/documents/:s_document/325/original/The_Coastal_Ecosystem_Mapping_and_Media_Viability_Project_Final-web.pdf?1490611734

Aguirre-Gutiérrez, J., Seijmonsbergen, A. C., & Duivenvoorden, J. F. (2012). Optimizing land cover classification accuracy for change detection, a combined pixel-based and object-based approach in a mountainous area in Mexico. *Applied Geography*, *34*, 29–37. https://doi.org/10.1016/j.apgeog.2011.10.010

Ahmad, H. (2019). Machine learning applications in oceanography. *Aquatic Research*, 161–169. https://doi.org/10.3153/AR19014

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Akbari, E., Darvishi Boloorani, A., Neysani Samany, N., Hamzeh, S., Soufizadeh, S., & Pignatti, S. (2020). Crop mapping using random forest and particle swarm optimization based on multi-temporal Sentinel-2. *Remote Sensing*, *12*(9), 1449. https://doi.org/10.3390/rs12091449

Albanese, D., Visintainer, R., Merler, S., Riccadonna, S., Jurman, G., & Furlanello, C. (2012). mlpy: Machine Learning Python. *ArXiv:1202.6548 [Cs, Stat]*. http://arxiv.org/abs/1202.6548

Alongi, D. M., Murdiyarso, D., Fourqurean, J. W., Kauffman, J. B., Hutahaean, A., Crooks, S., Lovelock, C. E., Howard, J., Herr, D., Fortes, M., Pidgeon, E., & Wagey, T. (2016). Indonesia's blue carbon: A globally significant and vulnerable sink for seagrass and mangrove carbon. *Wetlands Ecology and Management*, *24*(1), 3–13. https://doi.org/10.1007/s11273-015-9446-y

Amone-Mabuto, M., Bandeira, S., & da Silva, A. (2017). *Long-term changes in seagrass coverage and potential links to climate-related factors: The case of Inhambane Bay, Southern Mozambique*. *16*(2), 13–25.

Amran, M. A. (2010). Estimation of seagrass coverage by Depth Invariant Indices on Quickbird imagery. *Journal of BIOTROPIA*, *17*(1). http://dx.doi.org/10.11598/btb.2010.17.1.43

Andris, C. (2016). Integrating social network data into GISystems. *International Journal of Geographical Information Science*, *30*(10), 2009–2031. https://doi.org/10.1080/13658816.2016.1153103

Aoki, L. R., McGlathery, K. J., Wiberg, P. L., & Al-Haj, A. (2020). Depth affects seagrass restoration success and resilience to marine heat wave disturbance. *Estuaries and Coasts*, *43*(2), 316–328. https://doi.org/10.1007/s12237-019-00685-0

Asmala, A. (2012). *Analysis of maximum likelihood classification multispectral data*.

Avitabile, V., Herold, M., Henry, M., & Schmullius, C. (2011). Mapping biomass with remote sensing: A comparison of methods for the case study of Uganda. *Carbon Balance and Management*, *6*(1), 7–7. https://doi.org/10.1186/1750-0680-6-7

Bagnall, A., Bostrom, A., Cawley, G., Flynn, M., Large, J., & Lines, J. (2018). Is rotation forest the best classifier for problems with continuous features? *ArXiv:1809.06705 [Cs, Stat]*. http://arxiv.org/abs/1809.06705

Bansal, S., Gupta, D., Panchal, V. K., & Kumar, S. (2009). Swarm intelligence inspired classifiers in comparison with fuzzy and rough classifiers: a remote sensing approach. In S. Ranka, S. Aluru, R. Buyya, Y.-C. Chung, S. Dua, A. Grama, S. K. S. Gupta, R. Kumar, & V. V. Phoha (Eds.), *Contemporary Computing* (pp. 284–294). Springer. https://doi.org/10.1007/978-3-642-03547-0_27

Baumstark, R., Duffey, R., & Pu, R. (2016). Mapping seagrass and colonized hard bottom in Springs Coast, Florida using WorldView-2 satellite imagery. *Estuarine, Coastal and Shelf Science*, *181*, 83–92. https://doi.org/10.1016/j.ecss.2016.08.019

Bedulli, C., Lavery, P. S., Harvey, M., Duarte, C. M., & Serrano, O. (2020). Contribution of seagrass blue carbon toward carbon neutral policies in a touristic and environmentally-friendly island. *Frontiers in Marine Science*, *7*, 1. https://doi.org/10.3389/fmars.2020.00001

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS Journal of Photogrammetry and Remote Sensing*, *114*, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Blomberg, E., Ulander, L. M. H., Tebaldini, S., & Ferro-Famil, L. (2021). Evaluating P-Band TomoSAR for Biomass Retrieval in Boreal Forest. *IEEE Transactions on Geoscience and Remote Sensing*, *59*(5), 3793–3804. https://doi.org/10.1109/TGRS.2020.3020775

Bonin-Font, F., Campos, M. M., & Codina, G. O. (2016). Towards visual detection, mapping and quantification of *Posidonia Oceanica* using a lightweight AUV. *IFAC-PapersOnLine*, *49*(23), 500–505. https://doi.org/10.1016/j.ifacol.2016.10.485

BOPRC. (2018). *Seagrass extent—Bay of Plenty*. http://gis.boplass.govt.nz/arcgis/rest/services/imagery

Bouillon, S., Dehairs, F., Velimirov, B., Abril, G., & Borges, A. V. (2007). Dynamics of organic and inorganic carbon across contiguous mangrove and seagrass systems (Gazi Bay, Kenya). *Journal of Geophysical Research*, *112*(G2), G02018–G02018. https://doi.org/10.1029/2006JG000325

Breiman, L. (2001). Random forest. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Brown, C. J., Smith, S. J., Lawton, P., & Anderson, J. T. (2011). Benthic habitat mapping: A review of progress towards improved understanding of the spatial ecology of the seafloor using acoustic techniques. *Estuarine, Coastal and Shelf Science*, *92*(3), 502–520. https://doi.org/10.1016/j.ecss.2011.02.007

Brunsdon, C., & Singleton, A. (2015). *Geocomputation: A practical primer*. SAGE Publications, Inc. https://doi.org/10.4135/9781473916432

Bruzzone, L., & Bovolo, F. (2013). A novel framework for the design of change-detection systems for very-high-resolution remote sensing images. *Proceedings of the IEEE*, *101*(3), 609–630. https://doi.org/10.1109/JPROC.2012.2197169

Bühlmann, P. (2012). Bagging, boosting and ensemble methods. In J. E. Gentle, W. K. Härdle, & Y. Mori (Eds.), *Handbook of Computational Statistics* (pp. 985–1022). Springer Berlin

Heidelberg. http://link.springer.com/10.1007/978-3-642-21551-3_33 https://www.econstor.eu/bitstream/10419/22204/1/31_pb.pdf

Bujang, J. S., Zakaria, M. H., & Short, F. T. (2018). Seagrass in Malaysia: Issues and challenges ahead. In C. M. Finlayson, G. R. Milton, R. C. Prentice, & N. C. Davidson (Eds.), *The Wetland Book* (pp. 1875–1883). Springer Netherlands. https://doi.org/10.1007/978-94-007-4001-3_268

Byrd, K. B., Ballanti, L., Thomas, N., Nguyen, D., Holmquist, J. R., Simard, M., & Windham-Myers, L. (2018). A remote sensing-based model of tidal marsh aboveground carbon stocks for the conterminous United States. *ISPRS Journal of Photogrammetry and Remote Sensing*, *139*, 255–271. https://doi.org/10.1016/j.isprsjprs.2018.03.019

Cabaço, S., Santos, R., & Duarte, C. M. (2008). The impact of sediment burial and erosion on seagrasses: A review. *Estuarine, Coastal and Shelf Science*, *79*(3), 354–366. https://doi.org/10.1016/j.ecss.2008.04.021

Cahaya, A. (2015). Fishermen community in the coastal area: a note from Indonesian poor family. *Procedia Economics and Finance*, *26*, 29–33. https://doi.org/10.1016/S2212-5671(15)00801-1

Campbell, J. E., Lacey, E. A., Decker, R. A., Crooks, S., & Fourqurean, J. W. (2015). Carbon storage in seagrass beds of Abu Dhabi, United Arab Emirates. *Estuaries and Coasts*, *38*(1), 242–251. https://doi.org/10.1007/s12237-014-9802-9

Camps-Valls, G. (2009). Machine learning in remote sensing data processing. *2009 IEEE International Workshop on Machine Learning for Signal Processing*, 1–6. https://doi.org/10.1109/MLSP.2009.5306233

Candra, E. D., Hartono, & Wicaksono, P. (2016). Above ground carbon stock estimates of mangrove forest using Worldview-2 imagery in Teluk Benoa, Bali. *IOP Conference Series: Earth and Environmental Science*, *47*(1), 12014–12014.

Carter, G. A., Lucas, K. L., Biber, P. D., Criss, G. A., & Blossom, G. A. (2011). Historical changes in seagrass coverage on the Mississippi barrier islands, northern Gulf of Mexico, determined from vertical aerial imagery (1940–2007). *Geocarto International*, *26*(8), 663–673. https://doi.org/10.1080/10106049.2011.620634

Casal, G., Kutser, T., Domínguez-Gómez, J. A., Sánchez-Carnero, N., & Freire, J. (2013). Assessment of the hyperspectral sensor CASI-2 for macroalgal discrimination on the Ría de Vigo coast (NW Spain) using field spectroscopy and modelled spectral libraries. *Continental Shelf Research*, *55*, 129–140. https://doi.org/10.1016/j.csr.2013.01.010

Chavez, P. S. (1988). An improved dark-object subtraction technique for atmospheric scattering correction of multispectral data. *Remote Sensing of Environment*, *24*(3), 459–479. https://doi.org/10.1016/0034-4257(88)90019-3

Chen, C.-F., Lau, V.-K., Chang, N.-B., Son, N.-T., Tong, P.-H.-S., & Chiang, S.-H. (2016). Multi-temporal change detection of seagrass beds using integrated Landsat TM/ETM+/OLI imageries in Cam Ranh Bay, Vietnam. *Ecological Informatics*, *35*, 43–54. https://doi.org/10.1016/j.ecoinf.2016.07.005

Chen, Q., Yu, R., Hao, Y., Wu, L., Zhang, W., Zhang, Q., & Bu, X. (2018). A new method for mapping aquatic vegetation especially underwater vegetation in lake Ulansuhai using GF-1 satellite data. *Remote Sensing*, *10*(8), 1279. https://doi.org/10.3390/rs10081279

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, 785–794. https://doi.org/10.1145/2939672.2939785

Coetzee, S., Ivánová, I., Mitasova, H., & Brovelli, M. A. (2020). Open geospatial software and data: A review of the current state and a perspective into the future. *ISPRS International Journal of Geo-Information*, *9*(2), 90. https://doi.org/10.3390/ijgi9020090

Cole, M. van den B. and A. (2014). *Ecosystem goods and services in marine protected areas (MPAs)* (pp. 100–100). http://www.doc.govt.nz/Documents/science-and-technical/sfc326entire.pdf

Colkesen, I., & Kavzoglu, T. (2017). Ensemble-based canonical correlation forest (CCF) for land use and land cover classification using Sentinel-2 and Landsat OLI imagery. *Remote Sensing Letters*, *8*(11), 1082–1091. https://doi.org/10.1080/2150704X.2017.1354262

Collier, C. J., Ow, Y. X., Langlois, L., Uthicke, S., Johansson, C. L., O'Brien, K. R., Hrebien, V., & Adams, M. P. (2017). Optimum temperatures for net primary productivity of three tropical seagrass species. *Frontiers in Plant Science*, *8*. https://doi.org/10.3389/fpls.2017.01446

Collier, C. J., Uthicke, S., & Waycott, M. (2011). Thermal tolerance of two seagrass species at contrasting light levels: Implications for future distribution in the Great Barrier Reef. *Limnology and Oceanography*, *56*(6), 2200–2210. https://doi.org/10.4319/lo.2011.56.6.2200

Collier, C. J., Villacorta-Rath, C., van Dijk, K., Takahashi, M., & Waycott, M. (2014). Seagrass proliferation precedes mortality during hypo-salinity events: A stress-induced

morphometric response. *PLoS ONE*, *9*(4), e94014. https://doi.org/10.1371/journal.pone.0094014

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., & Böhner, J. (2015). System for automated geoscientific analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, *8*(7), 1991–2007. https://doi.org/10.5194/gmd-8-1991-2015

Cook, S. (2013). *Mapping blue carbon in south slough national estuarine research reserve using ShoreZone* (pp. 19–19). http://www.shorezone.org/documents/Mapping%20Blue%20Carbon%20-%20Sarah%20Cook,%20CORI.pdf

Cullen-Unsworth, L., & Unsworth, R. (2013). Seagrass meadows, ecosystem services, and sustainability. *Environment: Science and Policy for Sustainable Development*, *55*(3), 14–28. https://doi.org/10.1080/00139157.2013.785864

Dat Pham, T., Xia, J., Thang Ha, N., Tien Bui, D., Nhu Le, N., & Tekeuchi, W. (2019). A review of remote sensing approaches for monitoring blue carbon ecosystems: Mangroves, sea grasses and salt marshes during 2010–2018. In *Sensors (Switzerland)* (Vol. 19). MDPI AG. https://doi.org/10.3390/s19081933

Daughtry, C. S. T., Walthall, C. L., Kim, M. S., De Colstoun, E. B., & McMurtrey, J. E. (2000). Estimating corn leaf chlorophyll concentration from leaf and canopy reflectance. *Remote Sensing of Environment*, *74*(2), 229–239. https://doi.org/10.1016/S0034-4257(00)00113-9

Davide, A. (2012). Nonlinear methods for classification: maximum likelihood classifier. *MLPY's Document*. http://mlpy.sourceforge.net/docs/3.5/nonlin_class.html#maximum-likelihood-classifier

de los Santos, C. B., Krause-Jensen, D., Alcoverro, T., Marbà, N., Duarte, C. M., van Katwijk, M. M., Pérez, M., Romero, J., Sánchez-Lizaso, J. L., Roca, G., Jankowska, E., Pérez-Lloréns, J. L., Fournier, J., Montefalcone, M., Pergent, G., Ruiz, J. M., Cabaço, S., Cook, K., Wilkes, R. J., … Santos, R. (2019). Recent trend reversal for declining European seagrass meadows. *Nature Communications*, *10*(1), 3356. https://doi.org/10.1038/s41467-019-11340-4

Deilmai, B. R., Ahmad, B. B., & Zabihi, H. (2014). Comparison of two classification methods (MLC and SVM) to extract land use and land cover in Johor Malaysia. *IOP Conference Series: Earth and Environmental Science*, *20*, 012052–012052. https://doi.org/10.1088/1755-1315/20/1/012052

Dekker, A., Brando, V., Anstee, J., Fyfe, S., Malthus, T., & Karpouzli, E. (2006). Remote sensing of seagrass ecosystems: use of spaceborne and airborne sensors. In S*eagrasses: Biology, Ecology and Conservation* (pp. 347–359). Springer Netherlands. https://doi.org/10.1007/978-1-4020-2983-7_15

Dekker, A. G., Anstee, J. M., & Brando, V. E. (2003). *Seagrass change assessment using satellite data for wallis lake, NSW* (pp. 64–64). http://www.clw.csiro.au/publications/technical2003/tr13-03.pdf

Dekker, A. G., Brando, V. E., & Anstee, J. M. (2005). Retrospective seagrass change detection in a shallow coastal tidal Australian lake. *Remote Sensing of Environment*, *97*(4), 415–433. https://doi.org/10.1016/j.rse.2005.02.017

Delegido, J., Verrelst, J., Alonso, L., & Moreno, J. (2011). Evaluation of Sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. *Sensors*, *11*(7), 7063–7081. https://doi.org/10.3390/s110707063

Devi, G. K., Ganasri, B. P., & Dwarakish, G. S. (2015). Applications of remote sensing in satellite oceanography: A Review. *Aquatic Procedia*, *4*, 579–584. https://doi.org/10.1016/j.aqpro.2015.02.075

Deyanova, D., Gullström, M., Lyimo, L. D., Dahl, M., Hamisi, M. I., Mtolera, M. S. P., & Björk, M. (2017). Contribution of seagrass plants to CO_2 capture in a tropical seagrass meadow under experimental disturbance. *PLOS ONE*, *12*(7), e0181386–e0181386. https://doi.org/10.1371/journal.pone.0181386

Di Minin, E., Tenkanen, H., & Toivonen, T. (2015). Prospects and challenges for social media data in conservation science. *Frontiers in Environmental Science*, *3*. https://doi.org/10.3389/fenvs.2015.00063

DOC. (2011). *Coastal marine habitats and marine protected areas in the New Zealand Territorial Sea: A broad scale gap analysis* (pp. 50–50). http://www.doc.govt.nz/Documents/conservation/marine-and-coastal/marine-protected-areas/coastal-marine-habitats-marine-protected-areas.pdf

Douchin, M. (2020, September 2). *Lizmap—Documentation* [Document]. https://docs.lizmap.com/current/en/index.html

Doxani, G., Karantzalos, K., & Strati, M. T.-. (2012). Monitoring urban changes based on scale-space filtering and object-oriented classification. *International Journal of Applied Earth Observation and Geoinformation*, *15*, 38–48. https://doi.org/10.1016/j.jag.2011.07.002

Duarte, C. M., Kennedy, H., Marbà, N., & Hendriks, I. (2013). Assessing the capacity of seagrass meadows for carbon burial: Current limitations and future strategies. *Ocean & Coastal Management*, *83*, 32–38. https://doi.org/10.1016/j.ocecoaman.2011.09.001

Duarte, C. M., & Krause-Jensen, D. (2017). Export from seagrass meadows contributes to marine carbon sequestration. *Frontiers in Marine Science*, *4*. https://doi.org/10.3389/fmars.2017.00013

El Hajj, M., Baghdadi, N., Bazzi, H., & Zribi, M. (2018). Penetration analysis of SAR signals in the C and L bands for wheat, maize, and grasslands. *Remote Sensing*, *11*(1), 31. https://doi.org/10.3390/rs11010031

Elbeltagi, E., Hegazy, T., & Grierson, D. (2005). Comparison among five evolutionary-based optimization algorithms. *Advanced Engineering Informatics*, *19*(1), 43–53. https://doi.org/10.1016/j.aei.2005.01.004

Elhefnawy, M., & Sri Sumantyo, J. T. (2016). A review on designing antenna arrays for long range synthetic aperture radar. 2016 International Workshop on Recent Advances in Robotics and Sensor Technology for Humanitarian Demining and Counter-IEDs (RST), 1–5. https://doi.org/10.1109/RST.2016.7869857

ESA. (2015). *Sentinel—2 user handbook* (pp. 64–64).

ESA. (2020). *SNAP v7.0*. http://step.esa.int

ESA. (2021a). *Sentinel-1 SAR*. https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar/overview

ESA. (2021b). *Sentinel-2 MSI*. https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-2-msi/overview

ESA. (2021c). *THEOS*. https://earth.esa.int/web/eoportal/satellite-missions/t/theos

ESA-S1. (2020). *Sentinel-1 SAR*. https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar

ESA-S2. (2020). *Sentinel-2 MSI*. https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi

Escadafal, R. (1989). Remote sensing of arid soil surface color with Landsat thematic mapper. *Advances in Space Research*, *9*(1), 159–163. https://doi.org/10.1016/0273-1177(89)90481-X

Fauzan, M. A., Kumara, I. S. W., Yogyantoro, R. N., Suwardana, S. W., Fadhilah, N., Nurmalasari, I., Apriyani, S., & Wicaksono, P. (2017). Assessing the capability of Sentinel-2a data for mapping seagrass percent cover in Jerowaru, East Lombok. *Indonesian Journal of Geography*, *49*(2), 195–195. https://doi.org/10.22146/ijg.28407

Fehr, J., Arreola, K. Z., & Burkhardt, H. (2008). Fast support vector machine classification of very large datasets. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme, & R. Decker (Eds.), *Data Analysis, Machine Learning and Applications* (pp. 11–18). Springer Berlin Heidelberg. http://link.springer.com/10.1007/978-3-540-78246-9_2 https://lmb.informatik.uni-freiburg.de/papers/download/fe_za_bu_GFKL07.pdf

Feng, W., Sui, H., Tu, J., Huang, W., Xu, C., & Sun, K. (2018). A novel change detection approach for multi-temporal high-resolution remote sensing images based on rotation forest and coarse-to-fine uncertainty analyses. *Remote Sensing*, *10*(7), 1015. https://doi.org/10.3390/rs10071015

Ferriss, B. E., Conway-Cranos, L. L., Sanderson, B. L., & Hoberecht, L. (2019). Bivalve aquaculture and eelgrass: A global meta-analysis. *Aquaculture*, *498*, 254–262. https://doi.org/10.1016/j.aquaculture.2018.08.046

Ferrol-Schulte, D., Wolff, M., Ferse, S., & Glaser, M. (2013). Sustainable livelihoods approach in tropical coastal and marine social–ecological systems: A review. *Marine Policy*, *42*, 253–258. https://doi.org/10.1016/j.marpol.2013.03.007

Ferwerda, J. G., de Leeuw, J., Atzberger, C., & Vekerdy, Z. (2007). Satellite-based monitoring of tropical seagrass vegetation: Current techniques and future developments. *Hydrobiologia*, *591*(1), 59–71. https://doi.org/10.1007/s10750-007-0784-5

Filipponi, F. (2019). Sentinel-1 GRD preprocessing workflow. *Proceedings*, *18*(1), 11. https://doi.org/10.3390/ECRS-3-06201

Fonseca, M. S., & Cahalan, J. A. (1992). A preliminary evaluation of wave attenuation by four species of seagrass. *Estuarine, Coastal and Shelf Science*, *35*(6), 565–576. https://doi.org/10.1016/S0272-7714(05)80039-3

Fourqurean, J. W., Duarte, C. M., Kennedy, H., Marbà, N., Holmer, M., Mateo, M. A., Apostolaki, E. T., Kendrick, G. A., Krause-Jensen, D., McGlathery, K. J., & Serrano, O. (2012). Seagrass ecosystems as a globally significant carbon stock. *Nature Geoscience*, *5*(7), 505–509. https://doi.org/10.1038/ngeo1477

Frampton, W. J., Dash, J., Watmough, G., & Milton, E. J. (2013). Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation. *ISPRS Journal of Photogrammetry and Remote Sensing*, *82*, 83–92. https://doi.org/10.1016/j.isprsjprs.2013.04.007

Fraser, M. W., & Kendrick, G. A. (2017). Belowground stressors and long-term seagrass declines in a historically degraded seagrass ecosystem after improved water quality. *Scientific Reports*, *7*(1), 14469. https://doi.org/10.1038/s41598-017-14044-1

Fraser, M. W., Kendrick, G. A., Statton, J., Hovey, R. K., Zavala-Perez, A., & Walker, D. I. (2014). Extreme climate events lower resilience of foundation seagrass at edge of biogeographical range. *Journal of Ecology*, *102*(6), 1528–1536. https://doi.org/10.1111/1365-2745.12300

Fu, P. (2018). *Getting to know Web GIS* (Third edition). Esri Press.

Garcia, R., Hedley, J., Tin, H., & Fearns, P. (2015). A method to analyze the potential of optical remote sensing for benthic habitat mapping. *Remote Sensing*, *7*(10), 13157–13189. https://doi.org/10.3390/rs71013157

Gazeau, F., Duarte, C. M., Gattuso, J.-P., Barrón, C., Navarro, N., Ruiz, S., Prairie, Y. T., Calleja, M., Delille, B., Frankignoulle, M., & Borges, A. V. (2005). Whole-system metabolism and CO_2 fluxes in a Mediterranean Bay dominated by seagrass beds (Palma Bay, NW Mediterranean). *Biogeosciences*, *2*(1), 43–60. https://doi.org/10.5194/bg-2-43-2005

Georganos, S., Grippa, T., Vanhuysse, S., Lennert, M., Shimoni, M., & Wolff, E. (2018). Very high resolution object-based land use–land cover urban classification using extreme gradient boosting. *IEEE Geoscience and Remote Sensing Letters*, *15*(4), 607–611. https://doi.org/10.1109/LGRS.2018.2803259

Gholami, R., & Fakhari, N. (2017). Support vector machine: principles, parameters, and applications. In *Handbook of Neural Computation* (pp. 515–535). Elsevier. https://linkinghub.elsevier.com/retrieve/pii/B9780128113189000272

Gholizadeh, M., Melesse, A., & Reddi, L. (2016). A comprehensive review on water quality parameters estimation using remote sensing techniques. *Sensors*, *16*(8), 1298–1298. https://doi.org/10.3390/s16081298

Gitelson, A. A., Kaufman, Y. J., & Merzlyak, M. N. (1996). Use of a green channel in remote sensing of global vegetation from EOS- MODIS. *Remote Sensing of Environment*, *58*(3), 289–298. https://doi.org/10.1016/S0034-4257(96)00072-7

Githaiga, M. N., Frouws, A. M., Kairo, J. G., & Huxham, M. (2019). Seagrass removal leads to rapid changes in fauna and loss of carbon. *Frontiers in Ecology and Evolution*, *7*, 62. https://doi.org/10.3389/fevo.2019.00062

Githaiga, M. N., Kairo, J. G., Gilpin, L., & Huxham, M. (2017). Carbon storage in the seagrass meadows of Gazi Bay, Kenya. *PLOS ONE*, *12*(5), e0177001–e0177001. https://doi.org/10.1371/journal.pone.0177001

Gómez, C., White, J. C., & Wulder, M. A. (2016). Optical remotely sensed time series data for land cover classification: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *116*, 55–72. https://doi.org/10.1016/j.isprsjprs.2016.03.008

Green, E. P., Mumby, P. J., Edwards, A. J., & Clark, C. D. (1996). A review of remote sensing for the assessment and management of tropical coastal resources. *Coastal Management*, *24*(1), 1–40. https://doi.org/10.1080/08920759609362279

Green, E. P., & Short, F. T. (2003). *World atlas of seagrasses*. University of California Press.

Green, Edmund P., & Edwards, A. J. (Eds.). (2000). *Remote sensing handbook for tropical coastal management*. Unesco Pub.

Gullström, M., Lundén, B., Bodin, M., Kangwe, J., Öhman, M. C., Mtolera, M. S. P., & Björk, M. (2006). Assessment of changes in the seagrass-dominated submerged vegetation of tropical Chwaka Bay (Zanzibar) using satellite remote sensing. *Estuarine, Coastal and Shelf Science*, *67*(3), 399–408. https://doi.org/10.1016/j.ecss.2005.11.020

Gullström, M., Lyimo, L. D., Dahl, M., Samuelsson, G. S., Eggertsen, M., Anderberg, E., Rasmusson, L. M., Linderholm, H. W., Knudby, A., Bandeira, S., Nordlund, L. M., & Björk, M. (2018). Blue carbon storage in tropical seagrass meadows relates to carbonate stock dynamics, plant–sediment processes, and landscape context: Insights from the Western Indian ocean. *Ecosystems*, *21*(3), 551–566. https://doi.org/10.1007/s10021-017-0170-8

Gumusay, M. U., Bakirman, T., Tuney Kizilkaya, I., & Aykut, N. O. (2019). A review of seagrass detection, mapping and monitoring applications using acoustic systems. *European Journal of Remote Sensing*, *52*(1), 1–29. https://doi.org/10.1080/22797254.2018.1544838

Ha, N. T., Manley-Harris, M., Pham, T. D., & Hawes, I. (2020). A comparative assessment of ensemble-based machine learning and maximum likelihood methods for mapping seagrass using Sentinel-2 imagery in Tauranga Harbor, New Zealand. *Remote Sensing*, *12*(3), 355. https://doi.org/10.3390/rs12030355

Ha, N.-T., Manley-Harris, M., Pham, T. D., & Hawes, I. (2021). The use of radar and optical satellite imagery combined with advanced machine learning and metaheuristic optimization techniques to detect and quantify above ground biomass of intertidal seagrass in a New Zealand estuary. *International Journal of Remote Sensing*, *42*(12), 4712–4738. https://doi.org/10.1080/01431161.2021.1899335

Ha, N. T., Yoshino, K., & Hoang Son, T. P. (2012). *Seagrass mapping using ALOS AVNIR-2 data in Lap An Lagoon, Thua Thien Hue, Viet Nam* (R. J. Frouin, N. Ebuchi, D. Pan, & T. Saino, Eds.; pp. 85250S-85250S). SPIE. https://doi.org/10.1117/12.977188

He, Q., Bertness, M. D., Bruno, J. F., Li, B., Chen, G., Coverdale, T. C., Altieri, A. H., Bai, J., Sun, T., Pennings, S. C., Liu, J., Ehrlich, P. R., & Cui, B. (2014). Economic development and coastal ecosystem change in China. *Scientific Reports*, *4*(1), 5995. https://doi.org/10.1038/srep05995

Heck Hay, K., Hays, G., & Orth, R. (2003). Critical evaluation of the nursery role hypothesis for seagrass meadows. *Marine Ecology Progress Series*, *253*, 123–136. https://doi.org/10.3354/meps253123

Hedley, J. D., Russell, B. J., Randolph, K., Pérez-Castro, M. Á., Vásquez-Elizondo, R. M., Enríquez, S., & Dierssen, H. M. (2017). Remote sensing of seagrass leaf area index and species: the capability of a model inversion method assessed by sensitivity analysis and hyperspectral data of Florida Bay. *Frontiers in Marine Science*, *4*. https://doi.org/10.3389/fmars.2017.00362

Heidari, A. A., Mirjalili, S., Faris, H., Aljarah, I., Mafarja, M., & Chen, H. (2019). Harris hawks optimization: Algorithm and applications. *Future Generation Computer Systems*, *97*, 849–872. https://doi.org/10.1016/j.future.2019.02.028

Heiden, U., Keim, S., Schrader, S., Foerster, S., Kuester, T., Elger, K., & G., R. (2017). *EnMAP Campaign Portal – Flight campaigns of the "EnMAP Data Exploitation and Ap plication Development Program"* (pp. 14–14). https://doi.org/10.2312/enmap.2017.001

Hejnowicz, A. P., Kennedy, H., Rudd, M. A., & Huxham, M. R. (2015). Harnessing the climate mitigation, conservation and poverty alleviation potential of seagrasses: Prospects for developing blue carbon initiatives and payment for ecosystem service programmes. *Frontiers in Marine Science*, *2*. https://doi.org/10.3389/fmars.2015.00032

Herbeck, L. S., Sollich, M., Unger, D., Holmer, M., & Jennerjahn, T. C. (2014). Impact of pond aquaculture effluents on seagrass performance in NE Hainan, tropical China. *Marine Pollution Bulletin*, *85*(1), 190–203. https://doi.org/10.1016/j.marpolbul.2014.05.050

Herold, M., & Skutsch, M. (2011). Monitoring, reporting and verification for national REDD + programmes: Two proposals. *Environmental Research Letters*, *6*(1), 014002. https://doi.org/10.1088/1748-9326/6/1/014002

Hicks, M. (2019). *Review and analysis of suspended sediment monitoring in the Tauranga Moana Catchment* (No. 2019183CH; NIWA Client Report 2019183CH Prepared for Bay of Plenty Regional Council, p. 57). https://atlas.boprc.govt.nz/api/v1/edms/document/A3297536/content

Hicks, M., Semadeni-Davies, A., Haddadchi, A., Shankar, U., & Plew, D. (2019). *Updated sediment load estimator for New Zealand* (No. 2018341CH; NIWA Client Report 2018341CH Prepared for Ministry for the Environment, p. 190). https://www.mfe.govt.nz/sites/default/files/media/Fresh%20water/updated-sediment-load-estimator-for-nz.pdf

Hill, V. J., Zimmerman, R. C., Bissett, W. P., Dierssen, H., & Kohler, D. D. R. (2014). Evaluating light availability, seagrass biomass, and productivity using hyperspectral airborne remote sensing in Saint Joseph's Bay, Florida. *Estuaries and Coasts*, *37*(6), 1467–1489. https://doi.org/10.1007/s12237-013-9764-3

Hogland, J., Billor, N., & Anderson, N. (2013). Comparison of standard maximum likelihood classification and polytomous logistic regression used in remote sensing. *European Journal of Remote Sensing*, *46*(1), 623–640. https://doi.org/10.5721/EuJRS20134637

Holloway, J., & Mengersen, K. (2018). Statistical machine learning methods and remote sensing for sustainable development goals: A review. *Remote Sensing*, *10*(9), 1365. https://doi.org/10.3390/rs10091365

Hossain, M. S., Bujang, J. S., Zakaria, M. H., & Hashim, M. (2015). The application of remote sensing to seagrass ecosystems: An overview and future research prospects. *International Journal of Remote Sensing*, *36*(1), 61–114. https://doi.org/10.1080/01431161.2014.990649

Howard, J., Hoyt, S., Isensee, K., Pidgeon, E., & Telszewski, M. (2014). *Coastal Blue Carbon: Methods for assessing carbon stocks and emissions factors in mangroves, tidal salt marshes, and seagrass meadows* (M. Telszewski, Ed.). Intergovernmental Oceanographic Commission of UNESCO, International Union for Conservation of Nature. http://thebluecarboninitiative.org/manual/

Howell, S. E. L., Komarov, A. S., Dabboor, M., Montpetit, B., Brady, M., Scharien, R. K., Mahmud, M. S., Nandan, V., Geldsetzer, T., & Yackel, J. J. (2018). Comparing L- and C-band synthetic aperture radar estimates of sea ice motion over different ice regimes. *Remote Sensing of Environment*, *204*, 380–391. https://doi.org/10.1016/j.rse.2017.10.017

Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, *25*(3), 295–309. https://doi.org/10.1016/0034-4257(88)90106-X

Huettmann, F. (2018). Boosting, Bagging and Ensembles in the Real World: An Overview, some Explanations and a Practical Synthesis for Holistic Global Wildlife Conservation Applications Based on Machine Learning with Decision Trees. In *Machine Learning for Ecology and Sustainable Natural Resource Management* (pp. 63–83). Springer International Publishing. https://doi.org/10.1007/978-3-319-96978-7_3

Ihlen, V. (2019). *Landsat 8 (L8) data users handbook*. USGS. https://prd-wret.s3-us-west-2.amazonaws.com/assets/palladium/production/atoms/files/LSDS-1574_L8_Data_Users_Handbook-v5.0.pdf

ISRO. (2021). *The saga of Indian remote sensing satellite system*. https://www.isro.gov.in/saga-of-indian-remote-sensing-satellite-system

Jia, J., Wang, Y., Chen, J., Guo, R., Shu, R., & Wang, J. (2020). Status and application of advanced airborne hyperspectral imaging technology: A review. In *Infrared Physics and Technology* (Vol. 104). Elsevier B.V. https://doi.org/10.1016/j.infrared.2019.103115

Jiang, Z., Huete, A. R., Didan, K., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment*, *112*(10), 3833–3845. https://doi.org/10.1016/j.rse.2008.06.006

Joshi, N., Baumann, M., Ehammer, A., Fensholt, R., Grogan, K., Hostert, P., Jepsen, M., Kuemmerle, T., Meyfroidt, P., Mitchard, E., Reiche, J., Ryan, C., & Waske, B. (2016). A review of the application of optical and radar remote sensing data fusion to land use mapping and monitoring. *Remote Sensing*, *8*(1), 70–70. https://doi.org/10.3390/rs8010070

Joshua, L. (2016). *Rotation forest*. https://github.com/joshloyal/RotationForest

Karan, S. K., & Samadder, S. R. (2016). Accuracy of land use change detection using support vector machine and maximum likelihood techniques for open-cast coal mining areas. *Environmental Monitoring and Assessment*, *188*(8), 486–486. https://doi.org/10.1007/s10661-016-5494-x

Karamizadeh, S., Abdullah, S. M., Halimi, M., Shayan, J., & Rajabi, M. javad. (2014). Advantage and drawback of support vector machine functionality. *2014 International*

Conference on Computer, Communications, and Control Technology (I4CT), 63–65. https://doi.org/10.1109/I4CT.2014.6914146

Karnatak, H. C., Saran, S., Bhatia, K., & Roy, P. S. (2007). Multicriteria spatial decision analysis in web GIS environment. *GeoInformatica*, *11*(4), 407–429. https://doi.org/10.1007/s10707-006-0014-8

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, *2017-Decem*, 3147–3155. https://www.microsoft.com/en-us/research/publication/lightgbm-a-highly-efficient-gradient-boosting-decision-tree/

Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks*, *4*, 1942–1948 vol.4. https://doi.org/10.1109/ICNN.1995.488968

Kerr, V. (2009). *Marine habitat map of Northland: Mangawhai to Ahipara* (pp. 33–33). http://www.doc.govt.nz/Documents/conservation/marine-and-coastal/northland-marine-habitat-map-report-1.pdf

Kim, J.-Y., Im, R.-Y., Do, Y., Kim, G.-Y., & Joo, G.-J. (2016). Above-ground biomass estimation of tuberous bulrush (*Bolboschoenus planiculmis*) in mudflats using remotely sensed multispectral image. *Ocean Science Journal*, *51*(1), 151–158. https://doi.org/10.1007/s12601-016-0013-0

Kim, K., Choi, J.-K., Ryu, J.-H., Jeong, H. J., Lee, K., Park, M. G., & Kim, K. Y. (2015). Observation of typhoon-induced seagrass die-off using remote sensing. *Estuarine, Coastal and Shelf Science*, *154*, 111–121. https://doi.org/10.1016/j.ecss.2014.12.036

Koedsin, W., Intararuang, W., Ritchie, R., & Huete, A. (2016). An integrated field and remote sensing method for mapping seagrass species, cover, and biomass in Southern Thailand. *Remote Sensing*, *8*(4), 292. https://doi.org/10.3390/rs8040292

Konar, B., & Iken, K. (2018). The use of unmanned aerial vehicle imagery in intertidal monitoring. *Deep Sea Research Part II: Topical Studies in Oceanography*, *147*, 79–86. https://doi.org/10.1016/j.dsr2.2017.04.010

Kovacs, E., Roelfsema, C., Lyons, M., Zhao, S., & Phinn, S. (2018). Seagrass habitat mapping: How do Landsat 8 OLI, Sentinel-2, ZY-3A, and Worldview-3 perform? *Remote Sensing Letters*, *9*(7), 686–695. https://doi.org/10.1080/2150704X.2018.1468101

Kuenzer, C., Bluemel, A., Gebhardt, S., Quoc, T. V., & Dech, S. (2011). Remote sensing of mangrove ecosystems: A Review. *Remote Sensing*, *3*(12), 878–928. https://doi.org/10.3390/rs3050878

Kurnianto, S., & Murdiyarso, D. (2010). *Forest carbon database: A web-based carbon stock data repository and exchange system*. Center for International Forestry Research (CIFOR). https://doi.org/10.17528/cifor/003234

Lambert-Torres, G., Martins, H. G., Coutinho, M. P., Salomon, C. P., Matsunaga, F. M., & Carminati, R. A. (2009, December). Comparison between PSO and GA in system restoration solution. *2009 15th International Conference on Intelligent System Applications to Power Systems, ISAP '09*. https://doi.org/10.1109/ISAP.2009.5352885

Lary, D. J., Alavi, A. H., Gandomi, A. H., & Walker, A. L. (2016). Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, *7*(1), 3–10. https://doi.org/10.1016/j.gsf.2015.07.003

Lauer, M., & Aswani, S. (2010). Indigenous knowledge and long-term ecological change: detection, interpretation, and responses to changing ecological conditions in pacific island communities. *Environmental Management*, *45*(5), 985–997. https://doi.org/10.1007/s00267-010-9471-9

Lee, Y., Han, D., Ahn, M.-H., Im, J., & Lee, S. J. (2019). Retrieval of total precipitable water from Himawari-8 AHI data: a comparison of random forest, extreme gradient boosting, and deep neural network. *Remote Sensing*, *11*(15), 1741. https://doi.org/10.3390/rs11151741

Lee, Zhongping, Carder, K. L., Mobley, C. D., Steward, R. G., & Patch, J. S. (1999). Hyperspectral remote sensing for shallow waters: Deriving bottom depths and water properties by optimization. *Applied Optics*, *38*(18), 3831. https://doi.org/10.1364/ao.38.003831

Lee, ZhongPing, & Hu, C. (2006). Global distribution of Case-1 waters: An analysis from SeaWiFS measurements. *Remote Sensing of Environment*, *101*(2), 270–276. https://doi.org/10.1016/j.rse.2005.11.008

Lemenkova, P. (2019). Processing oceanographic data by python libraries Numpy, Scipy, and Pandas. *Aquatic Research*, 73–91. https://doi.org/10.3153/AR19009

Li, C., Wang, J., Wang, L., Hu, L., & Gong, P. (2014). Comparison of classification algorithms and training sample sizes in urban land classification with Landsat thematic mapper imagery. *Remote Sensing*, *6*(2), 964–983. https://doi.org/10.3390/rs6020964

Li, F., Zhang, L., Chen, B., Gao, D., Cheng, Y., Zhang, X., Yang, Y., Gao, K., Huang, Z., & Peng, J. (2018). A light gradient boosting machine for remaining useful life estimation of

aircraft engines. *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, 2018-November*, 3562–3567. https://doi.org/10.1109/ITSC.2018.8569801

Li, Y., Li, M., Li, C., & Liu, Z. (2020). Forest aboveground biomass estimation using Landsat 8 and Sentinel-1A data with machine learning algorithms. *Scientific Reports*, *10*(1), 9952. https://doi.org/10.1038/s41598-020-67024-3

Liu, H., & Zhou, Q. (2004). Accuracy analysis of remote sensing change detection by rule-based rationality evaluation with post-classification comparison. *International Journal of Remote Sensing*, *25*(5), 1037–1050. https://doi.org/10.1080/0143116031000150004

Liu, Y. (2017). *Python machine learning by example: Easy-to-follow examples that get you up and running with machine learning*. Packt Publishing.

Lo Iacono, C., Mateo, M. A., Gràcia, E., Guasch, L., Carbonell, R., Serrano, L., Serrano, O., & Dañobeitia, J. (2008). Very high-resolution seismo-acoustic imaging of seagrass meadows (Mediterranean Sea): Implications for carbon sink estimates. *Geophysical Research Letters*, *35*(18), L18601–L18601. https://doi.org/10.1029/2008GL034773

Luo, M., Wang, Y., Xie, Y., Zhou, L., Qiao, J., Qiu, S., & Sun, Y. (2021). Combination of Feature Selection and CatBoost for Prediction: The First Application to the Estimation of Aboveground Biomass. *Forests*, *12*(2), 216. https://doi.org/10.3390/f12020216

Lyons, M. B., Roelfsema, C. M., & Phinn, S. R. (2013). Towards understanding temporal and spatial dynamics of seagrass landscapes using time-series remote sensing. *Estuarine, Coastal and Shelf Science*, *120*, 42–53. https://doi.org/10.1016/j.ecss.2013.01.015

Lyons, M, Roelfsema, C., Kovacs, E., Samper-Villarreal, J., Saunders, M., Maxwell, P., & Phinn, S. (2015). Rapid monitoring of seagrass biomass using a simple linear modelling

approach, in the field and from space. *Marine Ecology Progress Series*, *530*, 1–14. https://doi.org/10.3354/meps11321

Lyons, Mitchell, Phinn, S., & Roelfsema, C. (2011). Integrating Quickbird Multi-Spectral Satellite and Field Data: Mapping Bathymetry, Seagrass Cover, Seagrass Species and Change in Moreton Bay, Australia in 2004 and 2007. *Remote Sensing*, *3*(1), 42–64. https://doi.org/10.3390/rs3010042

Lyzenga, D. R. (1978). Passive remote sensing techniques for mapping water depth and bottom features. *Applied Optics*, *17*(3), 379–379. https://doi.org/10.1364/AO.17.000379

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. https://doi.org/10.1109/TGRS.2006.872909

Machova, K., Puszta, M., Barcak, F., & Bednar, P. (2006). A comparison of the bagging and the boosting methods using the decision trees classifiers. *Computer Science and Information Systems*, *3*(2), 57–72. https://doi.org/10.2298/CSIS0602057M

Macreadie, Peter I., Anton, A., Raven, J. A., Beaumont, N., Connolly, R. M., Friess, D. A., Kelleway, J. J., Kennedy, H., Kuwae, T., Lavery, P. S., Lovelock, C. E., Smale, D. A., Apostolaki, E. T., Atwood, T. B., Baldock, J., Bianchi, T. S., Chmura, G. L., Eyre, B. D., Fourqurean, J. W., … Duarte, C. M. (2019). The future of blue carbon science. *Nature Communications*, *10*(1), 3998. https://doi.org/10.1038/s41467-019-11693-w

Macreadie, P.I., Baird, M. E., Trevathan-Tackett, S. M., Larkum, A. W. D., & Ralph, P. J. (2014). Quantifying and modelling the carbon sequestration capacity of seagrass meadows – A critical assessment. *Marine Pollution Bulletin*, *83*(2), 430–439. https://doi.org/10.1016/j.marpolbul.2013.07.038

Maier, P. M., & Keller, S. (2018). Machine learning regression on hyperspectral data to estimate multiple water parameters. *2018 9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 1–5. https://doi.org/10.1109/WHISPERS.2018.8747010

Manessa, M., Kanno, A., Sekine, M., Ampou, E., Widagti, N., & As-syakur, Abd. (2014). Shallow-water benthic identification using multispectral satellite imagery: investigation on the effects of improving noise correction method and spectral cover. *Remote Sensing*, *6*(5), 4454–4472. https://doi.org/10.3390/rs6054454

Manuputty, A., Gaol, J. L., Agus, S. B., & Nurjaya, I. W. (2017). The utilization of depth invariant index and principle component analysis for mapping seagrass ecosystem of Kotok Island and Karang Bongkok, Indonesia. *IOP Conference Series: Earth and Environmental Science*, *54*(1), 12083–12083.

Marbà, N., Arias-Ortiz, A., Masqué, P., Kendrick, G. A., Mazarrasa, I., Bastyan, G. R., Garcia-Orellana, J., & Duarte, C. M. (2015). Impact of seagrass loss and subsequent revegetation on carbon sequestration and stocks. *Journal of Ecology*, *103*(2), 296–302. https://doi.org/10.1111/1365-2745.12370

Maritorena, S., Morel, A., & Gentili, B. (1994). Diffuse reflectance of oceanic shallow waters: Influence of water depth and bottom albedo. *Limnology and Oceanography*, *39*(7), 1689–1703. https://doi.org/10.4319/lo.1994.39.7.1689

Matheson, F., & Wadhwa, S. (2012). *Seagrass in Porirua Harbor Preliminary assessment of restoration potential Prepared for Greater Wellington Regional Council* (p. 35). http://www.gw.govt.nz/assets/Seagrass-restoration-in-Porirua-Harbor.pdf

Mathieu, R., Pouget, M., Cervelle, B., & Escadafal, R. (1998). Relationships between satellite-based radiometric indices simulated using laboratory reflectance data and typic soil color of an arid environment. *Remote Sensing of Environment*, *66*(1), 17–28. https://doi.org/10.1016/S0034-4257(98)00030-3

Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. *International Journal of Remote Sensing*, *39*(9), 2784–2817. https://doi.org/10.1080/01431161.2018.1433343

McCarthy, M. J., & Halls, J. N. (2014). Habitat mapping and change assessment of coastal environments: an examination of WorldView-2, QuickBird, and IKONOS satellite imagery and airborne lidar for mapping barrier island habitats. *ISPRS International Journal of Geo-Information*, *3*(1), 297–325. https://doi.org/10.3390/ijgi3010297

Mcleod, E., Chmura, G. L., Bouillon, S., Salm, R., Björk, M., Duarte, C. M., Lovelock, C. E., Schlesinger, W. H., & Silliman, B. R. (2011). A blueprint for blue carbon: Toward an improved understanding of the role of vegetated coastal habitats in sequestering CO 2. *Frontiers in Ecology and the Environment*, *9*(10), 552–560. https://doi.org/10.1890/110004

Meyer, C. A., & Pu, R. (2012). Seagrass resource assessment using remote sensing methods in St. Joseph Sound and Clearwater Harbor, Florida, USA. *Environmental Monitoring and Assessment*, *184*(2), 1131–1143. https://doi.org/10.1007/s10661-011-2028-4

Meyer, R. P. and S. B. and K. H. L. and C. (2010). Mapping detailed seagrass habitats using satellite imagery. *2010 IEEE International Geoscience and Remote Sensing Symposium*, 4–4. https://doi.org/10.1109/IGARSS.2010.5651884

Misbari, S., & Hashim, M. (2016). Change detection of submerged seagrass biomass in shallow coastal water. *Remote Sensing*, *8*(3), 200. https://doi.org/10.3390/rs8030200

Mitra, A., & Zaman, S. (2015). Blue carbon assessment. In *Blue Carbon Reservoir of the Blue Planet* (pp. 93–111). Springer India. https://doi.org/10.1007/978-81-322-2107-4_3

Mohamed, H., Nadaoka, K., & Nakamura, T. (2018). Assessment of machine learning algorithms for automatic benthic cover monitoring and mapping using towed underwater video camera and high-resolution satellite images. *Remote Sensing*, *10*(5), 773. https://doi.org/10.3390/rs10050773

Morcillo-Pallarés, P., Rivera-Caicedo, J. P., Belda, S., De Grave, C., Burriel, H., Moreno, J., & Verrelst, J. (2019). Quantifying the robustness of vegetation indices through global sensitivity analysis of homogeneous and forest leaf-canopy radiative transfer models. *Remote Sensing*, *11*(20), 2418. https://doi.org/10.3390/rs11202418

Morrison, M., Lowe, M. L., Grant, C. M., Smith, P. J., Carbines, G. D., Reed, J., Bury, S., Brown, J., New Zealand, & Ministry for Primary Industries. (2014). *Seagrass meadows as biodiversity and productivity hotspots*. http://fs.fish.govt.nz/Doc/23703/AEBR_137_2701_ZBD2004-08%20Objective%201-4;%20Milestones%208,9,10,17,18,23,24,%2025,28,29%20and%2030.pdf.ashx

Moughal, T. A. (2013). Hyperspectral image classification using support vector machine. *Journal of Physics: Conference Series*, *439*, 012042–012042. https://doi.org/10.1088/1742-6596/439/1/012042

Mountrakis, G., Im, J., & Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(3), 247–259. https://doi.org/10.1016/j.isprsjprs.2010.11.001

Mtwana Nordlund, L., Koch, E. W., Barbier, E. B., & Creed, J. C. (2016). Seagrass ecosystem services and their variability across genera and geographical regions. *Plos One*, *11*(10), e0163091–e0163091. https://doi.org/10.1371/journal.pone.0163091

Mulder, V. L., de Bruin, S., Schaepman, M. E., & Mayr, T. R. (2011). The use of remote sensing in soil and terrain mapping—A review. *Geoderma*, *162*(1–2), 1–19. https://doi.org/10.1016/j.geoderma.2010.12.018

Mumby, P. J., Clark, C. D., Green, E. P., & Edwards, A. J. (1998). Benefits of water column correction and contextual editing for mapping coral reefs. *International Journal of Remote Sensing*, *19*(1), 203–210. https://doi.org/10.1080/014311698216521

Mustapha, M. R., Lim, H. S., & Mat Jafri, M. Z. (2010). Comparison of Neural Network and Maximum Likelihood Approaches in Image Classification. *Journal of Applied Sciences*, *10*(22), 2847–2854. https://doi.org/10.3923/jas.2010.2847.2854

Naidoo, L., van Deventer, H., Ramoelo, A., Mathieu, R., Nondlazi, B., & Gangat, R. (2019). Estimating above ground biomass as an indicator of carbon storage in vegetated wetlands of the grassland biome of South Africa. *International Journal of Applied Earth Observation and Geoinformation*, *78*, 118–129. https://doi.org/10.1016/j.jag.2019.01.021

Natekin, A., & Knoll, A. (2013). Gradient boosting machines, a tutorial. *Frontiers in Neurorobotics*, *7*(DEC), 21. https://doi.org/10.3389/fnbot.2013.00021

Navarro, J. A., Algeet, N., Fernández-Landa, A., Esteban, J., Rodríguez-Noriega, P., & Guillén-Climent, M. L. (2019). Integration of UAV, Sentinel-1, and Sentinel-2 Data for mangrove plantation aboveground biomass monitoring in Senegal. *Remote Sensing*, *11*(1), 77. https://doi.org/10.3390/rs11010077

Nguyen, D. T. H., Tripathi, N. K., Gallardo, W. G., & Tipdecho, T. (2014). Coastal and marine ecological changes and fish cage culture development in Phu Quoc, Vietnam (2001 to 2011). *Geocarto International*, *29*(5), 486–506. https://doi.org/10.1080/10106049.2013.798358

Nitze, I., Schulthess, U., & Asche, H. (2012). Comparison of machine learning algorithms random forest, artificial neural network and support vector machine to maximum likelihood for supervised crop type classification. *Proceedings of the 4th GEOBIA*, 35–40. https://pdfs.semanticscholar.org/a10e/79e21be020daab308d0fb5aafe3b3efa5adf.pdf

NIWA. (2020). *The national climate database*. https://cliflo.niwa.co.nz/

Nobi, E. P., & Thangaradjou, T. (2012). Evaluation of the spatial changes in seagrass cover in the lagoons of Lakshadweep islands, India, using IRS LISS III satellite images. *Geocarto International*, *27*(8), 647–660. https://doi.org/10.1080/10106049.2012.665501

Nordlund, L. M., Jackson, E. L., Nakaoka, M., Samper-Villarreal, J., Beca-Carretero, P., & Creed, J. C. (2018). Seagrass ecosystem services – What's next? *Marine Pollution Bulletin*, *134*, 145–151. https://doi.org/10.1016/j.marpolbul.2017.09.014

Nordlund, L. M., Koch, E. W., Barbier, E. B., & Creed, J. C. (2016). Seagrass ecosystem services and their variability across genera and geographical regions. *Plos One*, *11*(10), e0163091–e0163091. https://doi.org/10.1371/journal.pone.0163091

Northrop, A. (2015). *IDEAS – LANDSAT products description document* (IDEAS-VEG-SRV-REP-1320; p. 68). USGS. https://earth.esa.int/documents/10174/679851/LANDSAT_Products_Description_Document.pdf

Nurdin, N., Amri, K., Djalil, Abd. R., Jaya, I., Aris, A., & Akbar AS, M. (2014). *Geospatial dynamic of seagrass in outer zone, Spermonde Archipelago, Indonesia using Landsat data from 1972-2013* (R. J. Frouin, D. Pan, & H. Murakami, Eds.; pp. 92610N-92610N). https://doi.org/10.1117/12.2062898

Oprandi, A., Mucerino, L., De Leo, F., Bianchi, C. N., Morri, C., Azzola, A., Benelli, F., Besio, G., Ferrari, M., & Montefalcone, M. (2020). Effects of a severe storm on seagrass meadows. *Science of The Total Environment*, *748*, 141373. https://doi.org/10.1016/j.scitotenv.2020.141373

Oreska, M. P. J., McGlathery, K. J., & Porter, J. H. (2017). Seagrass blue carbon spatial patterns at the meadow-scale. *PLOS ONE*, *12*(4), e0176630–e0176630. https://doi.org/10.1371/journal.pone.0176630

Orth, R. J., Carruthers, T. J. B., Dennison, W. C., Duarte, C. M., Fourqurean, J. W., Heck, K. L., Hughes, A. R., Kendrick, G. A., Kenworthy, W. J., Olyarnik, S., Short, F. T., Waycott, M., & Williams, S. L. (2006). A Global Crisis for Seagrass Ecosystems. *BioScience*, *56*(12), 987–987. https://doi.org/10.1641/0006-3568(2006)56[987:AGCFSE]2.0.CO;2

Otukei, J. R., & Blaschke, T. (2010). Land cover change assessment using decision trees, support vector machines and maximum likelihood classification algorithms. *International Journal of Applied Earth Observation and Geoinformation*, *12*, S27–S31. https://doi.org/10.1016/j.jag.2009.11.002

P. de Alencar, N. M., Le Tissier, M., Paterson, S. K., & Newton, A. (2020). Circles of Coastal Sustainability: A Framework for Coastal Management. *Sustainability*, *12*(12), 4886. https://doi.org/10.3390/su12124886

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, *26*(1), 217–222. https://doi.org/10.1080/01431160412331269698

Pan, X., Xue, L., & Li, R. (2019). A new and efficient firefly algorithm for numerical optimization problems. *Neural Computing and Applications*, *31*(5), 1445–1453. https://doi.org/10.1007/s00521-018-3449-6

Pan, Z., Glennie, C., Fernandez-Diaz, J. C., & Starek, M. (2016). Comparison of bathymetry and seagrass mapping with hyperspectral imagery and airborne bathymetric lidar in a shallow estuarine environment. *International Journal of Remote Sensing*, *37*(3), 516–536. https://doi.org/10.1080/01431161.2015.1131869

Panda, S., & Padhy, N. P. (2008). Comparison of particle swarm optimization and genetic algorithm for FACTS-based controller design. *Applied Soft Computing*, *8*(4), 1418–1427. https://doi.org/10.1016/j.asoc.2007.10.009

Park, S. G. (1999). *Changes in abundance of seagrass (Zostera spp.) in Tauranga Harbor from 1959—96* (Environment Report No. 99/30; p. 19). https://cdn.boprc.govt.nz/media/362713/changes-in-abundance-of-seagrass-zostera-spp-in-tauranga-harbor-from-1959-96.pdf

Park, S. G. (2011). *Extent of seagrass in the Bay of Plenty in 2011* (No. 1179–9471; Environmental Publication, p. 52). Bay of Plenty Reginal Council. https://cdn.boprc.govt.nz/media/509090/extent-of-seagrass-in-the-bay-of-plenty-in-2011.pdf

Pasqualini, V., Pergent-Martini, C., Pergent, G., Agreil, M., Skoufas, G., Sourbes, L., & Tsirika, A. (2005). Use of SPOT 5 for mapping seagrasses: An application to Posidonia oceanica. *Remote Sensing of Environment*, *94*(1), 39–45. https://doi.org/10.1016/j.rse.2004.09.010

Paulo, D., Cunha, A. H., Boavida, J., Serrão, E. A., Gonçalves, E. J., & Fonseca, M. (2019). Open coast seagrass restoration. Can we do it? Large scale seagrass transplants. *Frontiers in Marine Science*, *6*(MAR), 52. https://doi.org/10.3389/fmars.2019.00052

Paulose, N. E., Dilipan, E., & Thangaradjou, T. (2013). Integrating Indian remote sensing multi-spectral satellite and field data to estimate seagrass cover change in the Andaman and Nicobar Islands, India. *Ocean Science Journal*, *48*(2), 173–181. https://doi.org/10.1007/s12601-013-0014-1

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pe'eri, S., Morrison, J. R., Short, F., Mathieson, A., & Lippmann, T. (2016). Eelgrass and macroalgal mapping to develop nutrient criteria in New Hampshire's estuaries using hyperspectral imagery. *Journal of Coastal Research*, *76*, 209–218. https://doi.org/10.2112/SI76-018

Pendleton, L., Donato, D. C., Murray, B. C., Crooks, S., Jenkins, W. A., Sifleet, S., Craft, C., Fourqurean, J. W., Kauffman, J. B., Marbà, N., Megonigal, P., Pidgeon, E., Herr, D., Gordon, D., & Baldera, A. (2012). Estimating global "blue carbon" emissions from conversion and degradation of vegetated coastal ecosystems. *PLoS ONE*, *7*(9), e43542–e43542. https://doi.org/10.1371/journal.pone.0043542

PGDG. (2020, August 26). *PostgreSQL: Documentation: 12: PostgreSQL 12.4 Documentation*. PostgreSQL 12.4 Documentation. https://www.postgresql.org/docs/12/index.html

Pham, T. D., Le, N. N., Ha, N. T., Nguyen, L. V., Xia, J., Yokoya, N., To, T. T., Trinh, H. X., Kieu, L. Q., & Takeuchi, W. (2020). Estimating mangrove above-ground biomass using extreme gradient boosting decision trees algorithm with fused Sentinel-2 and ALOS-2 PALSAR-2 data in Can Gio biosphere reserve, Vietnam. *Remote Sensing*, *12*(5), 777. https://doi.org/10.3390/rs12050777

Pham, T. D., Xia, J., Baier, G., Le, N. N., & Yokoya, N. (2019). Mangrove species mapping using Sentinel-1 and Sentinel-2 data in North Vietnam. *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 6102–6105. https://doi.org/10.1109/IGARSS.2019.8898987

Pham, T. D., Yokoya, N., Nguyen, T. T. T., Le, N. N., Ha, N. T., Xia, J., Takeuchi, W., & Pham, T. D. (2020). Improvement of mangrove soil carbon stocks estimation in North Vietnam using Sentinel-2 data and machine learning approach. *GIScience & Remote Sensing*, *0*(0), 1–20. https://doi.org/10.1080/15481603.2020.1857623

Pham, T. D., Yokoya, N., Xia, J., Ha, N. T., Le, N. N., Nguyen, T. T. T., Dao, T. H., Vu, T. T. P., Pham, T. D., & Takeuchi, W. (2020). Comparison of machine learning methods for estimating mangrove above-ground biomass using multiple source remote sensing data in the Red River delta biosphere reserve, Vietnam. *Remote Sensing*, *12*(8), 1334. https://doi.org/10.3390/rs12081334

Phinn, S. R., Kovacs, E. M., Roelfsema, C. M., Canto, R. F., Collier, C. J., & McKenzie, L. J. (2018). Assessing the potential for satellite image monitoring of seagrass thermal dynamics: For inter- and shallow sub-tidal seagrasses in the inshore Great Barrier reef World Heritage Area, Australia. *International Journal of Digital Earth*, *11*(8), 803–824. https://doi.org/10.1080/17538947.2017.1359343

Phinn, S., Roelfsema, C., Dekker, A., Brando, V., & Anstee, J. (2008). Mapping seagrass species, cover and biomass in shallow waters: An assessment of satellite multi-spectral and airborne hyper-spectral imaging systems in Moreton Bay (Australia). *Remote Sensing of Environment*, *112*(8), 3413–3425. https://doi.org/10.1016/j.rse.2007.09.017

Poursanidis, D., Topouzelis, K., & Chrysoulakis, N. (2018). Mapping coastal marine habitats and delineating the deep limits of the Neptune's seagrass meadows using very high resolution Earth observation data. *International Journal of Remote Sensing*, 1–18. https://doi.org/10.1080/01431161.2018.1490974

Poursanidis, D., Traganos, D., Reinartz, P., & Chrysoulakis, N. (2019). On the use of Sentinel-2 for coastal habitat mapping and satellite-derived bathymetry estimation using downscaled coastal aerosol band. *International Journal of Applied Earth Observation and Geoinformation*, *80*, 58–70. https://doi.org/10.1016/j.jag.2019.03.012

Probst, P., Wright, M., & Boulesteix, A.-L. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *9*(3), e1301. https://doi.org/10.1002/widm.1301

Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2019). CatBoost: Unbiased boosting with categorical features. *ArXiv:1706.09516 [Cs]*. http://arxiv.org/abs/1706.09516 https://arxiv.org/pdf/1706.09516.pdf https://arxiv.org/abs/1706.09516

Pu, R., & Bell, S. (2013). A protocol for improving mapping and assessing of seagrass abundance along the West Central Coast of Florida using Landsat TM and EO-1 ALI/Hyperion images. *ISPRS Journal of Photogrammetry and Remote Sensing*, *83*, 116–129. https://doi.org/10.1016/j.isprsjprs.2013.06.008

Pu, R., & Bell, S. (2017). Mapping seagrass coverage and spatial patterns with high spatial resolution IKONOS imagery. *International Journal of Applied Earth Observation and Geoinformation*, *54*, 145–158. https://doi.org/10.1016/j.jag.2016.09.011

Pu, R., Bell, S., Meyer, C., Baggett, L., & Zhao, Y. (2012). Mapping and assessing seagrass along the western coast of Florida using Landsat TM and EO-1 ALI/Hyperion imagery. *Estuarine, Coastal and Shelf Science*, *115*, 234–245. https://doi.org/10.1016/j.ecss.2012.09.006

Purnawan, S., Aulia, F., & Khairuman, T. (2016). Using Landsat-8 imagery data on mapping of the seagrass distribution in Matahari Island, Pulau Banyak district, Aceh province, Indonesia. *Aceh International Journal of Science and Technology*, *5*.

Rainforth, T. (2018). *Canonical correlation forests*. https://github.com/twgr/ccfs

Rainforth, T., & Wood, F. (2015). Canonical correlation forests. *ArXiv:1507.05444 [Cs, Stat]*. http://arxiv.org/abs/1507.05444

Ramage, D. L., & Schiel, D. R. (1998). Reproduction in the seagrass *Zostera novazelandica* on intertidal platforms in southern New Zealand. *Marine Biology*, *130*(3), 479–489. https://doi.org/10.1007/s002270050268

Ramesh, R., Banerjee, K., Paneerselvam, A., Raghuraman, R., Purvaja, R., & Lakshmi, A. (2018). Importance of seagrass management for effective mitigation of climate change. In *Coastal Management: Global Challenges and Innovations* (pp. 283–299). Elsevier. https://doi.org/10.1016/B978-0-12-810473-6.00015-7

Raschka, S. (2018). MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. *Journal of Open Source Software*, *3*(24), 638. https://doi.org/10.21105/joss.00638

Raschka, S., & Mirjalili, V. (2017). *Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow* (Second edition, fourth release,[fully revised and updated]). Packt Publishing.

RBINS. (2018). *ACOLITE Python user manual* (pp. 10–10).

Reeve, G., Stephens, S., & Wadhwa, A. (2018). *Tauranga Harbor inundation modelling* (NIWA Client Report 2018269HN to Bay of Plenty Regional Council No. 2018269HN; p. 107). NIWA. https://atlas.boprc.govt.nz/api/v1/edms/document/A3338785/content

Reidenbach, M. A., & Thomas, E. L. (2018). Influence of the seagrass, *Zostera marina*, on wave attenuation and bed shear stress within a shallow coastal bay. *Frontiers in Marine Science*, *5*(OCT), 397. https://doi.org/10.3389/fmars.2018.00397

Rhein, M., Rintoul, S. R., Aoki, S., Campos, E., Chambers, D., Feely, R. A., Gulev, S., Johnson, G. C., Josey, S. A., Kostianoy, A., Mauritzen, C., Roemmich, D., Talley, L. D., & Wan, F. (2013). *Observations: Ocean*. Cambridge University Press. https://www.ipcc.ch/report/ar5/wg1/

Ricart, A. M., York, P. H., Bryant, C. V., Rasheed, M. A., Ierodiaconou, D., & Macreadie, P. I. (2020). High variability of blue carbon storage in seagrass meadows at the estuary scale. *Scientific Reports*, *10*(1), 1–12. https://doi.org/10.1038/s41598-020-62639-y

Richards, J. A. (2013). Supervised classification techniques. In *Remote Sensing Digital Image Analysis*. Springer.

Robertson, B., & Stevens, L. (2016). *Manawatu-Wanganui Estuaries. Habitat Mapping, Vulnerability Assessment and Monitoring Recommendations Related to Issues of Eutrophication and Sedimentation* (pp. 120–120). http://www.envirolink.govt.nz/assets/Envirolink/1624-HZLC127-Manawatu-Wanganui-

Estuaries-Habitat-Mapping-Vulnerability-Assessment-and-Monitoring-Recommendations-Related-to-Issues-of-Eutrophication-and-Sedimentation.pdf

Rodriguez, J. J., Kuncheva, L. I., & Alonso, C. J. (2006). Rotation forest: A new classifier ensemble method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*(10), 1619–1630. https://doi.org/10.1109/TPAMI.2006.211

Roelfsema, C., Lyons, M., Dunbabin, M., Kovacs, E. M., & Phinn, S. (2015). Integrating field survey data with satellite image data to improve shallow water seagrass maps: The role of AUV and snorkeller surveys? *Remote Sensing Letters*, *6*(2), 135–144. https://doi.org/10.1080/2150704X.2015.1013643

Roelfsema, C. M., Lyons, M., Kovacs, E. M., Maxwell, P., Saunders, M. I., Samper-Villarreal, J., & Phinn, S. R. (2014). Multi-temporal mapping of seagrass cover, species and biomass: A semi-automated object based image analysis approach. *Remote Sensing of Environment*, *150*, 172–187. https://doi.org/10.1016/j.rse.2014.05.001

Rouse, J. W., Jr., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *NASA. Goddard Space Flight Center 3d ERTS-1 Symp., Vol. 1, Sect. A*. https://ntrs.nasa.gov/citations/19740022614

Rouse, L. J., Bergeron, S. J., & Harris, T. M. (2007). Participating in the geospatial web: collaborative mapping, social networks and participatory GIS. In A. Scharl & K. Tochtermann (Eds.), *The Geospatial Web: How Geobrowsers, Social Software and the Web 2.0 are Shaping the Network Society* (pp. 153–158). Springer. https://doi.org/10.1007/978-1-84628-827-2_14

Roy, D. P., Wulder, M. A., Loveland, T. R., C.E., W., Allen, R. G., Anderson, M. C., Helder, D., Irons, J. R., Johnson, D. M., Kennedy, R., Scambos, T. A., Schaaf, C. B., Schott, J. R.,

Sheng, Y., Vermote, E. F., Belward, A. S., Bindschadler, R., Cohen, W. B., Gao, F., … Zhu, Z. (2014). Landsat-8: Science and product vision for terrestrial global change research. *Remote Sensing of Environment*, *145*, 154–172. https://doi.org/10.1016/j.rse.2014.02.001

Sagawa, T., Boisnier, E., Komatsu, T., Mustapha, K. B., Hattour, A., Kosaka, N., & Miyazaki, S. (2010). Using bottom surface reflectance to map coastal marine areas: A new application method for Lyzenga's model. *International Journal of Remote Sensing*, *31*(12), 3051–3064. https://doi.org/10.1080/01431160903154341

Sahin, E. K., Colkesen, I., & Kavzoglu, T. (2018). A comparative assessment of canonical correlation forest, random forest, rotation forest and logistic regression methods for landslide susceptibility mapping. *Geocarto International*, *33*, 1–23. https://doi.org/10.1080/10106049.2018.1516248

Salinas, C., Duarte, C. M., Lavery, P. S., Masque, P., Arias-Ortiz, A., Leon, J. X., Callaghan, D., Kendrick, G. A., & Serrano, O. (2020). Seagrass losses since mid-20th century fuelled CO2 emissions from soil carbon stocks. *Global Change Biology*, *26*(9), 4772–4784. https://doi.org/10.1111/gcb.15204

Salum, R. B., Souza-Filho, P. W. M., Simard, M., Silva, C. A., Fernandes, M. E. B., Cougo, M. F., do Nascimento, W., & Rogers, K. (2020). Improving mangrove above-ground biomass estimates using LiDAR. *Estuarine, Coastal and Shelf Science*, *236*, 106585. https://doi.org/10.1016/j.ecss.2020.106585

Sani, D. A., Hashim, M., & Hossain, M. S. (2019). Recent advancement on estimation of blue carbon biomass using satellite-based approach. *International Journal of Remote Sensing*, *40*(20), 7679–7715. https://doi.org/10.1080/01431161.2019.1601289

Satellite Imaging Corp. (2021a). *ALOS satellite sensor*. https://www.satimagingcorp.com/satellite-sensors/other-satellite-sensors/alos/

Satellite Imaging Corp. (2021b). *ASTER satellite sensor*. https://www.satimagingcorp.com/satellite-sensors/other-satellite-sensors/aster/

Satellite Imaging Corp. (2021c). *IKONOS satellite imagery*. https://www.satimagingcorp.com/satellite-sensors/ikonos/

Satellite Imaging Corp. (2021d). *KOMPSAT-3A satellite sensor*. https://www.satimagingcorp.com/satellite-sensors/kompsat-3a/

Satellite Imaging Corp. (2021e). *QuickBird satellite sensor*. https://www.satimagingcorp.com/satellite-sensors/quickbird/

Satellite Imaging Corp. (2021f). *SPOT-7 satellite sensor*. https://www.satimagingcorp.com/satellite-sensors/spot-7/

Satellite Imaging Corp. (2021g). *WorldView-4 Satellite Sensor*. https://www.satimagingcorp.com/satellite-sensors/geoeye-2/

Saunders, M. I., Atkinson, S., Klein, C. J., Weber, T., & Possingham, H. P. (2017). Increased sediment loads cause non-linear decreases in seagrass suitable habitat extent. *Plos One*, *12*(11), e0187284. https://doi.org/10.1371/journal.pone.0187284

Schile, L. M., Kauffman, J. B., Crooks, S., Fourqurean, J. W., Glavan, J., & Megonigal, J. P. (2017). Limits on carbon sequestration in arid blue carbon ecosystems. *Ecological Applications*, *27*(3), 859–874. https://doi.org/10.1002/eap.1489

Schwarz, A.-M., Morrison, M., Hawes, I., & Halliday, J. (2004a). *Physical and biological characteristics of a rare marine habitat: Sub-tidal seagrass beds of offshore islands* (p. 30). Department of Conservation.

Schwarz, A.-M., Morrison, M., Hawes, I., & Halliday, J. (2004b). *Physical and biological characteristics of a rare marine habitat: Sub-tidal seagrass beds of offshore islands* (pp. 30–30). https://www.doc.govt.nz/documents/science-and-technical/sfc269.pdf

Schwarz, A.-M., & Turner, S. (2006). *Management and conservation of seagrass in New Zealand: An introduction* (pp. 90–90). www.doc.govt.nz/documents/science-and-technical/sfc264.pdf

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464. https://doi.org/10.1214/aos/1176344136

Seixas Gomes de Almeida, B., & Coppo Leite, V. (2019). Particle swarm optimization: a powerful technique for solving engineering problems. In *Swarm Intelligence—Recent Advances, New Perspectives and Applications*. IntechOpen. https://doi.org/10.5772/intechopen.89633

Sengupta, S., Basak, S., & Peters, R. (2018). Particle Swarm Optimization: A survey of historical and recent developments with hybridization perspectives. *Machine Learning and Knowledge Extraction*, *1*(1), 157–191. https://doi.org/10.3390/make1010010

Shang, J., Liu, J., Poncos, V., Geng, X., Qian, B., Chen, Q., Dong, T., Macdonald, D., Martin, T., Kovacs, J., & Walters, D. (2020). Detection of crop seeding and harvest through analysis of time-series Sentinel-1 interferometric SAR data. *Remote Sensing*, *12*(10), 1551. https://doi.org/10.3390/rs12101551

Shapiro, A., & Rohmann, S. (2006). Mapping changes in submerged aquatic vegetation using Landsat imagery and benthic habitat data: Coral reef ecosystem monitoring in Vieques Sound between 1985 and 2000. *Bulletin of Marine Science, 79*, 375–388.

Short, F. T., & Short, C. A. (1984). The seagrass filter: Purification of estuarine and coastal waters. In *The Estuary As a Filter* (pp. 395–413). Elsevier. https://doi.org/10.1016/b978-0-12-405070-9.50024-4

Short, F., Torio, D., Hessing-Lewis, M., Reshitnyk, L., Denouden, T., McInnes, W., & Prentice, C. (2014). *Blue carbon seagrass mapping in Canada and the United States: British Columbia, Washington and Oregon* (pp. 8–8). http://www3.cec.org/islandora/es/item/11725-blue-carbon-seagrass-mapping-in-canada-and-united-states-british-columbia-en.pdf

Silva, G. C. M., Souza, F. E. S., & Marinho-Soriano, E. (2017). Application of ALOS AVNIR-2 for the detection of seaweed and seagrass beds on the northeast of Brazil. *International Journal of Remote Sensing*, *38*(3), 662–678. https://doi.org/10.1080/01431161.2016.1268738

Sousa, L. P., Sousa, A. I., Alves, F. L., & Lillebø, A. I. (2016). Ecosystem services provided by a complex coastal region: Challenges of classification and mapping. *Scientific Reports*, *6*(1), 22782–22782. https://doi.org/10.1038/srep22782

Stout, D. J., Kodis, M., & Wilen, B. O. (2007). *Wetlands layer—national spatial data infrastructure: a phased approach to completion and modernization* (pp. 12–12). https://www.fws.gov/wetlands/documents/Wetlands-Layer-National-Spatial-Data-Infrastructure-A-Phased-Approach-to-Completion-and-Modernization.pdf

Su, L., & Huang, Y. (2019). Seagrass resource assessment using WorldView-2 imagery in the Redfish Bay, Texas. *Journal of Marine Science and Engineering*, *7*(4), 98. https://doi.org/10.3390/jmse7040098

Sui, D., & Goodchild, M. (2011). The convergence of GIS and social media: Challenges for GIScience. *International Journal of Geographical Information Science*, *25*(11), 1737–1748. https://doi.org/10.1080/13658816.2011.604636

Sun, F., Wang, R., Wan, B., Su, Y., Guo, Q., Huang, Y., & Wu, X. (2019). Efficiency of extreme gradient boosting for imbalanced land cover classification using an extended margin and disagreement performance. *ISPRS International Journal of Geo-Information*, *8*(7), 315. https://doi.org/10.3390/ijgi8070315

Susi, R., Udhi, E. H., Kathryn, M., Bayu, P., Hanif, B. P., A'an, J. W., & Mat, V. (2019). Blue carbon in seagrass ecosystem: Guideline for the Assessment of Carbon Stock and Sequestration in Southeast Asia. In *Elements*. UGM PRESS. https://scholarbank.nus.edu.sg/handle/10635/171783

Suykerbuyk, W., Govers, L. L., Bouma, T. J., Giesen, W. B. J. T., Jong, D. J. de, Voort, R. van de, Giesen, K., Giesen, P. T., & Katwijk, M. M. van. (2016). Unpredictability in seagrass restoration: Analysing the role of positive feedback and environmental stress on Zostera noltii transplants. *Journal of Applied Ecology*, *53*(3), 774–784. https://doi.org/10.1111/1365-2664.12614

Tamondong, A. M., Blanco, A. C., Fortes, M. D., & Nadaoka, K. (2013). Mapping of seagrass and other benthic habitats in Bolinao, Pangasinan using Worldview-2 satellite image. *2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS*, 1579–1582. https://doi.org/10.1109/IGARSS.2013.6723091

Tara, J. A., Mark, M., Alison, M., Malcolm, C., Roberta, D., Nelson, W., Tracey, D., Gordon, D., Read, G., Kettles, H., Morrisey, D., Wood, A., Anderson, O., Smith, A. M., Page, M., Paul-Burke, K., Schnabel, K., & Wadhwa., S. (2019). *Review of New Zealand's key*

biogenic habitats (p. 190). https://www.mfe.govt.nz/sites/default/files/media/Marine/NZ-biogenic-habitat-review.pdf

Tay, H., Bryan, K., de Lange, W., & Pilditch, C. (2013). The hydrodynamics of the southern basin of Tauranga Harbor. *New Zealand Journal of Marine and Freshwater Research*, *47*(2), 249–274. https://doi.org/10.1080/00288330.2013.778300

Tewkesbury, A. P., Comber, A. J., Tate, N. J., Lamb, A., & Fisher, P. F. (2015). A critical synthesis of remotely sensed optical image change detection techniques. *Remote Sensing of Environment*, *160*, 1–14. https://doi.org/10.1016/j.rse.2015.01.006

Thessen, A. E. (2016). Adoption of machine learning techniques in ecology and earth science. *One Ecosystem*, *1*, e8621. https://doi.org/10.3897/oneeco.1.e8621

Tisimst. (2020). *Pyswarm*. https://github.com/tisimst/pyswarm

Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järv, O., Tenkanen, H., & Di Minin, E. (2019). Social media data for conservation science: A methodological overview. *Biological Conservation*, *233*, 298–315. https://doi.org/10.1016/j.biocon.2019.01.023

Tokoro, T., Hosokawa, S., Miyoshi, E., Tada, K., Watanabe, K., Montani, S., Kayanne, H., & Kuwae, T. (2014). Net uptake of atmospheric CO_2 by coastal submerged aquatic vegetation. *Global Change Biology*, *20*(6), 1873–1884. https://doi.org/10.1111/gcb.12543

Tomar, D., & Agarwal, S. (2015). Twin support vector machine: A review from 2007 to 2014. *Egyptian Informatics Journal*, *16*(1), 55–69. https://doi.org/10.1016/j.eij.2014.12.003

Torres-Pulliza, D., Wilson, J. R., Darmawan, A., Campbell, S. J., & Andréfouët, S. (2013). Ecoregional scale seagrass mapping: A tool to support resilient MPA network design in

the Coral Triangle. *Ocean & Coastal Management*, *80*, 55–64. https://doi.org/10.1016/j.ocecoaman.2013.04.005

Traganos, D., & Reinartz, P. (2017). Mapping Mediterranean seagrasses with Sentinel-2 imagery. *Marine Pollution Bulletin*. https://doi.org/10.1016/j.marpolbul.2017.06.075

Traganos, D., & Reinartz, P. (2018). Interannual change detection of Mediterranean seagrasses using Rapideye image time series. *Frontiers in Plant Science*, *9*. https://doi.org/10.3389/fpls.2018.00096

Tsouros, D. C., Bibi, S., & Sarigiannidis, P. G. (2019). A review on UAV-based applications for precision agriculture. *Information*, *10*(11), 349. https://doi.org/10.3390/info10110349

Tsujimoto, R., Terauchi, G., Sasaki, H., Sakamoto, S. X., Sawayama, S., Sasa, S., Yagi, H., & Komatsu, T. (2016). Damage to seagrass and seaweed beds in Matsushima Bay, Japan, caused by the huge tsunami of the Great East Japan Earthquake on 11 March 2011. *International Journal of Remote Sensing*, *37*(24), 5843–5863. https://doi.org/10.1080/01431161.2016.1249300

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*(2), 127–150. https://doi.org/10.1016/0034-4257(79)90013-0

Turner, S. J. (2007). Growth and productivity of intertidal *Zostera capricorni* in New Zealand estuaries. *New Zealand Journal of Marine and Freshwater Research*, *41*(1), 77–90. https://doi.org/10.1080/00288330709509897

UNEP-WCMC, S. F. (2018). Global distribution of seagrasses (version 6.0). Sixth update to the data layer used in Green and Short (2003). *Cambridge (UK): UNEP World Conservation Monitoring Centre*. https://data.unep-wcmc.org/datasets/7

Unsworth, R. K. F., & Coles, R. G. (2014). Seagrass meadows in a globally changing environment. *Marine Pollution Bulletin*, *83*(2), 383–386. https://doi.org/10.1016/j.marpolbul.2014.02.026

Unsworth, R. K. F., McKenzie, L. J., Collier, C. J., Cullen-Unsworth, L. C., Duarte, C. M., Eklöf, J. S., Jarvis, J. C., Jones, B. L., & Nordlund, L. M. (2018). Global challenges for seagrass conservation. *Ambio*. https://doi.org/10.1007/s13280-018-1115-y

Urbański, J., Mazur, A., & Janas, U. (2009). Object-oriented classification of QuickBird data for mapping seagrass spatial structure. *Oceanological and Hydrobiological Studies*, *38*(1). https://doi.org/10.2478/v10009-009-0013-9

USGS. (2018). *Landsat 7 data users handbook* (LSDS-1927; p. 154). https://landsat.usgs.gov/landsat-7-data-users-handbook

USGS. (2019). *Landsat 9*. https://www.usgs.gov/land-resources/nli/landsat/landsat-9

USGS. (2021a). *EO1-ALI*. https://www.usgs.gov/centers/eros/science/usgs-eros-archive-earth-observing-one-eo-1-ali

USGS. (2021b). *EO1-Hyperion*. https://www.usgs.gov/centers/eros/science/usgs-eros-archive-earth-observing-one-eo-1-hyperion

USGS. (2021c). *Landsat satellite*. https://www.usgs.gov/core-science-systems/nli/landsat/landsat-satellite-missions

Ustuner, M. (2015). Application of support vector machines for land use classification using high-resolution RapidEye images: A sensitivity analysis. *European Journal of Remote Sensing*, 403–403. https://doi.org/10.5721/EuJRS20154823

Valle, M., Palà, V., Lafon, V., Dehouck, A., Garmendia, J. M., Borja, Á., & Chust, G. (2015). Mapping estuarine habitats using airborne hyperspectral imagery, with special focus on

seagrass meadows. *Estuarine, Coastal and Shelf Science*, *164*, 433–442. https://doi.org/10.1016/j.ecss.2015.07.034

Vanhellemont, Q. (2019). Adaptation of the dark spectrum fitting atmospheric correction for aquatic applications of the Landsat and Sentinel-2 archives. *Remote Sensing of Environment*, *225*, 175–192. https://doi.org/10.1016/j.rse.2019.03.010

Vanhellemont, R. K. (2016). ACOLITE For Sentinel-2: Aquatic applications of MSI imagery. *Proceedings of the 2016 ESA Living Planet Symposium*, 8. https://odnature.naturalsciences.be/downloads/publications/vanhellemontruddick_esa_lps2016_coastalapplications_final_header.pdf

Veenendaal, B., Brovelli, M. A., & Li, S. (2017). Review of web mapping: eras, trends and directions. *ISPRS International Journal of Geo-Information*, *6*(10), 317. https://doi.org/10.3390/ijgi6100317

Veettil, B. K., Ward, R. D., Lima, M. D. A. C., Stankovic, M., Hoai, P. N., & Quang, N. X. (2020). Opportunities for seagrass research derived from remote sensing: A review of current methods. In *Ecological Indicators* (Vol. 117). Elsevier B.V. https://doi.org/10.1016/j.ecolind.2020.106560

Vela, A., Pasqualini, V., Leoni, V., Djelouli, A., Langar, H., Pergent, G., Pergent-Martini, C., Ferrat, L., Ridha, M., & Djabou, H. (2008). Use of SPOT 5 and IKONOS imagery for mapping biocenoses in a Tunisian Coastal Lagoon (Mediterranean Sea). *Estuarine, Coastal and Shelf Science*, *79*(4), 591–598. https://doi.org/10.1016/j.ecss.2008.05.014

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E.,

… SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Vorster, A. G., Evangelista, P. H., Stovall, A. E. L., & Ex, S. (2020). Variability and uncertainty in forest biomass estimates from the tree to landscape scale: The role of allometric equations. *Carbon Balance and Management*, *15*(1), 8. https://doi.org/10.1186/s13021-020-00143-6

Vrieze, S. I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods*, *17*(2), 228–243. https://doi.org/10.1037/a0027127

Wabnitz, C. C., Andréfouët, S., Torres-Pulliza, D., Müller-Karger, F. E., & Kramer, P. A. (2008). Regional-scale seagrass habitat mapping in the Wider Caribbean region using Landsat sensors: Applications to conservation and ecology. *Remote Sensing of Environment*, *112*(8), 3455–3467. https://doi.org/10.1016/j.rse.2008.01.020

Wang, D., Tan, D., & Liu, L. (2018). Particle swarm optimization algorithm: An overview. *Soft Computing*, *22*(2), 387–408. https://doi.org/10.1007/s00500-016-2474-6

Wang, L., Sousa, W. P., Gong, P., & Biging, G. S. (2004). Comparison of IKONOS and QuickBird images for mapping mangrove species on the Caribbean coast of Panama. *Remote Sensing of Environment*, *91*(3–4), 432–440. https://doi.org/10.1016/j.rse.2004.04.005

Watanabe, K., & Kuwae, T. (2015). How organic carbon derived from multiple sources contributes to carbon sequestration processes in a shallow coastal system? *Global Change Biology*, *21*(7), 2612–2623. https://doi.org/10.1111/gcb.12924

Watkins, R., L. (2015). *A Methodology for Classification of Benthic Features using WorldView-2 Imagery*. 29–29.

Waycott, M., Duarte, C. M., Carruthers, T. J. B., Orth, R. J., Dennison, W. C., Olyarnik, S., Calladine, A., Fourqurean, J. W., Heck, K. L., Hughes, A. R., Kendrick, G. A., Kenworthy, W. J., Short, F. T., & Williams, S. L. (2009). Accelerating loss of seagrasses across the globe threatens coastal ecosystems. *Proceedings of the National Academy of Sciences*, *106*(30), 12377–12381. https://doi.org/10.1073/pnas.0905620106

Wenhai, L., Cusack, C., Baker, M., Tao, W., Mingbao, C., Paige, K., Xiaofan, Z., Levin, L., Escobar, E., Amon, D., Yue, Y., Reitz, A., Neves, A. A. S., O'Rourke, E., Mannarini, G., Pearlman, J., Tinker, J., Horsburgh, K. J., Lehodey, P., … Yufeng, Y. (2019). Successful blue economy examples with an emphasis on international perspectives. *Frontiers in Marine Science*, *6*. https://doi.org/10.3389/fmars.2019.00261

Werts, J. D., Mikhailova, E. A., Post, C. J., & Sharp, J. L. (2012). An integrated WebGIS framework for volunteered geographic information and social media in soil and water conservation. *Environmental Management*, *49*(4), 816–832. https://doi.org/10.1007/s00267-012-9818-5

Wettle, M. B., Vittorio E. (2006). *SAMBUCA - Semi-analytical model for bathymetry, un-mixing and concentration assessment*. https://doi.org/10.4225/08/5866a187b7a3c

Wicaksono, P. (2016). Improving the accuracy of multispectral-based benthic habitats mapping using image rotations: The application of principle component analysis and independent component analysis. *European Journal of Remote Sensing*, *49*(1), 433–463. https://doi.org/10.5721/EuJRS20164924

Wicaksono, P., & Lazuardi, W. (2018). Assessment of PlanetScope images for benthic habitat and seagrass species mapping in a complex optically shallow water environment. *International Journal of Remote Sensing*, *39*(17), 5739–5765. https://doi.org/10.1080/01431161.2018.1506951

Wijaya, A., Sasmito, S. D., Purbopuspito, J., & Murdiyarso, D. (2013). *Calibration of global above ground biomass estimate using multi-source remote sensing data*. http://www.cifor.org/nc/online-library/browse/view-publication/publication/5180.html%3Cbr /%3E

Wilton, A. D., Schönberger, I., Boardman, K. F., Breitwieser, I., Cochrane, M., Dawson, M. I., de Lange, P. J., de Pauw, B., Fife, A. J., Ford, K. A., Gibb, E. S., Glenny, D. S., Heenan, P. B., Korver, M. A., Novis, P. M. ;, Redmond, D. N., Smissen, R. D., & Tawiri, K. (2016). *Checklist of the New Zealand flora—seed plants (2016)—Datasets—Manaaki Whenua—Landcare Research DataStore* (p. 385). https://doi.org/10.7931/P1PP42

Winters, G., Edelist, D., Shem-Tov, R., Beer, S., & Rilov, G. (2017). A low cost field-survey method for mapping seagrasses and their potential threats: An example from the northern Gulf of Aqaba, Red Sea: Mapping seagrasses and their potential threats in the Gulf of Aqaba. *Aquatic Conservation: Marine and Freshwater Ecosystems*, *27*(2), 324–339. https://doi.org/10.1002/aqc.2688

Wolter, P. T., Johnston, C. A., & Niemi, G. J. (2005). Mapping submergent aquatic vegetation in the US Great Lakes using Quickbird satellite data. *International Journal of Remote Sensing*, *26*(23), 5255–5274. https://doi.org/10.1080/01431160500219208

Wotton, K. G. M. and D. M. (2009). *Conservation and the delivery of ecosystem services: A literature review* (pp. 81–81). http://www.doc.govt.nz/Documents/science-and-technical/sfc295entire.pdf

Wu, T.-F., Lin, C.-J., & Weng, R. C. (2004). Probability estimates for multi-class classification by pairwise coupling. *J. Mach. Learn. Res.*, *5*, 975–1005.

Wu, Y., Xie, L., Huang, S.-L., Li, P., Yuan, Z., & Liu, W. (2018). Using social media to strengthen public awareness of wildlife conservation. *Ocean & Coastal Management*, *153*, 76–83. https://doi.org/10.1016/j.ocecoaman.2017.12.010

Xiaohui, D., Huapeng, L., Yong, L., Ji, Y., & Shuqing, Z. (2020). Comparison of swarm intelligence algorithms for optimized band selection of hyperspectral remote sensing image. *Open Geosciences*, *12*(1), 425–442. https://doi.org/10.1515/geo-2020-0155

Xie, Y., Sha, Z., & Yu, M. (2008). Remote sensing imagery in vegetation mapping: A review. *Journal of Plant Ecology*, *1*(1), 9–23. https://doi.org/10.1093/jpe/rtm005

Xiu, Y., Liu, W., & Yang, W. (2017). An improved rotation forest for multi-feature remote-sensing imagery classification. *Remote Sensing*, *9*(11), 1205. https://doi.org/10.3390/rs9111205

Xu, L., Zhang, H., Wang, C., Zhang, B., & Liu, M. (2019). Crop classification based on temporal information using Sentinel-1 SAR time-series data. *Remote Sensing*, *11*(1), 53. https://doi.org/10.3390/rs11010053

Xue, J., & Su, B. (2017). Significant remote sensing vegetation indices: A review of developments and applications. *Journal of Sensors*, *2017*. https://doi.org/10.1155/2017/1353691

Yan, J., Ma, Y., Wang, L., Choo, K.-K. R., & Jie, W. (2018). A cloud-based remote sensing data production system. *Future Generation Computer Systems*, *86*, 1154–1166. https://doi.org/10.1016/j.future.2017.02.044

Yang, B., Hawthorne, T. L., Hessing-Lewis, M., Duffy, E. J., Reshitnyk, L. Y., Feinman, M., & Searson, H. (2020). Developing an introductory UAV/drone mapping training program for seagrass monitoring and research. *Drones*, *4*(4), 70. https://doi.org/10.3390/drones4040070

Yang, C., Yang, D., Cao, W., Zhao, J., Wang, G., Sun, Z., Xu, Z., & Ravi Kumar, M. S. (2010). Analysis of seagrass reflectivity by using a water column correction algorithm. *International Journal of Remote Sensing*, *31*(17), 4595–4608. https://doi.org/10.1080/01431161.2010.485138

Yang, D., & Yang, C. (2009). Detection of seagrass distribution changes from 1991 to 2006 in Xincun Bay, Hainan, with satellite remote sensing. *Sensors*, *9*(2), 830–844. https://doi.org/10.3390/s90200830

Yao, H., Qin, R., & Chen, X. (2019). Unmanned Aerial Vehicle for remote sensing applications - A review. *Remote Sensing*, *11*(12), 1443. https://doi.org/10.3390/rs11121443

York, P. H., Gruber, R. K., Hill, R., Ralph, P. J., Booth, D. J., & Macreadie, P. I. (2013). Physiological and morphological responses of the temperate seagrass *Zostera muelleri* to multiple stressors: Investigating the interactive effects of light and temperature. *Plos One*, *8*(10), e76377. https://doi.org/10.1371/journal.pone.0076377

Zarate-Barrera, T. G., & Maldonado, J. H. (2015). Valuing blue carbon: carbon sequestration benefits provided by the marine protected areas in Colombia. *PLOS ONE*, *10*(5), e0126627–e0126627. https://doi.org/10.1371/journal.pone.0126627

Zhu, Z. (2017). Change detection using Landsat time series: A review of frequencies, preprocessing, algorithms, and applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, *130*, 370–384. https://doi.org/10.1016/j.isprsjprs.2017.06.013

Zoffoli, M., Frouin, R., & Kampel, M. (2014). Water column correction for coral reef studies by remote sensing. *Sensors*, *14*(9), 16881–16931. https://doi.org/10.3390/s140916881

Appendices

Appendix 1. Co-authorship form of research chapter 3

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

Postgraduate Studies Office
Student and Academic Services Division
Wahanga Ratonga Matauranga Akonga
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 4439
Website: http://www.waikato.ac.nz/sasd/postgraduate/

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 3: A Comparative Assessment of Ensemble-Based Machine Learning and Maximum Likelihood Methods for Mapping Seagrass Using Sentinel-2 Imagery in Tauranga Harbor, New Zealand

Nature of contribution by PhD candidate	Conceptualization, methodology, software, validation, resources, writing-original draft preparation, write-review and editing.
Extent of contribution by PhD candidate (%)	75

CO-AUTHORS

Name	Nature of Contribution
Merilyn Manley-Harris	Resources, writing-review and editing.
Tien Dat Pham	Validation, writing-review and editing.
Ian Hawes	Conceptualization, resources, writing-review and editing, supervision.

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and

Name	Signature	Date
Merilyn Manley-Harris		21.1.21
Tien Dat Pham		21.01.2021
Ian Hawes		24.2.2021

July 2015

Appendix 2. Published paper in the Remote Sensing Journal (chapter 3)

remote sensing

MDPI

Article

A Comparative Assessment of Ensemble-Based Machine Learning and Maximum Likelihood Methods for Mapping Seagrass Using Sentinel-2 Imagery in Tauranga Harbor, New Zealand

Nam Thang Ha [1,2,*], Merilyn Manley-Harris [1], Tien Dat Pham [3] and Ian Hawes [1]

[1] Environmental Research Institute, School of Science, University of Waikato, Hamilton 3260, New Zealand; manleyha@waikato.ac.nz (M.M.-H.); ian.hawes@waikato.ac.nz (I.H.)
[2] Faculty of Fisheries, University of Agriculture and Forestry, Hue University, Hue 530000, Vietnam
[3] Center for Agricultural Research and Ecological Studies (CARES), Vietnam National University of Agriculture (VNUA), Trau Quy, Gia Lam, Hanoi 10000, Vietnam; tiendat@vnua.edu.vn
* Correspondence: hanamthang@huaf.edu.vn

Received: 22 December 2019; Accepted: 20 January 2020; Published: 21 January 2020

check for updates

Abstract: Seagrass has been acknowledged as a productive blue carbon ecosystem that is in significant decline across much of the world. A first step toward conservation is the mapping and monitoring of extant seagrass meadows. Several methods are currently in use, but mapping the resource from satellite images using machine learning is not widely applied, despite its successful use in various comparable applications. This research aimed to develop a novel approach for seagrass monitoring using state-of-the-art machine learning with data from Sentinel–2 imagery. We used Tauranga Harbor, New Zealand as a validation site for which extensive ground truth data are available to compare ensemble machine learning methods involving random forests (RF), rotation forests (RoF), and canonical correlation forests (CCF) with the more traditional maximum likelihood classifier (MLC) technique. Using a group of validation metrics including F1, precision, recall, accuracy, and the McNemar test, our results indicated that machine learning techniques outperformed the MLC with RoF as the best performer (F1 scores ranging from 0.75–0.91 for sparse and dense seagrass meadows, respectively). Our study is the first comparison of various ensemble-based methods for seagrass mapping of which we are aware, and promises to be an effective approach to enhance the accuracy of seagrass monitoring.

Keywords: seagrass; Sentinel-2; random forest; rotation forest; canonical correlation forest; maximum likelihood; Tauranga; machine learning; remote sensing

1. Introduction

Together with mangrove and salt marsh, seagrass has been evaluated as an effective coastal ecosystem for blue carbon storage [1–3]. However, ongoing degradation of seagrass meadows [4] is leading to a requirement for accurate mapping and monitoring methods to facilitate the MRV (Monitoring, Reporting, and Verification) approach necessary for broad scale evaluation of their contribution to blue carbon reservoirs [5]. In the last decade, satellite imagery has been used extensively in developing seagrass mapping techniques by employing various classification algorithms with or without parallel traditional field surveys [6]. Among them, Sentinel-2 imagery is becoming more popular for seagrass mapping. Operated by the European Space Agency since 2015, this sensor supports a high quality image at spatial resolutions between 10 and 60 m [7]. Sentinel-2 data have been distributed free-of-charge at the top-of-atmosphere corrected level (level 1C) for blue, green, red,

and near infrared (NIR) bands at 10 m resolution, and provides a very good resource for intertidal and subtidal ecosystem mapping. Using these data to derive ecosystem spatial properties requires classification algorithms and overfitting, and the inaccurate edge detection of different substrata remains a limitation of traditional classification methods [6,8–10]. For seagrass mapping, the problems of misclassification usually relate to the impact of deep water on pixel values or the mixture of substrata within a seagrass meadow [11]. To overcome this problem, very high resolution (VHR) imagery and a variety of classification approaches can be considered [12,13]. Most frequently, probability-theory based models such as the maximum likelihood classifier (MLC) have been applied for seagrass classification [6,8]. This approach, however, requires conditions that are difficult to satisfy in the marine environment including a normal distribution of probabilities, equal co-variance, and large amounts of validation input data [14,15]. In addition, the utilization of the linear or quadratic discrimination functions of a MLC may not work when the boundaries of classes are not well defined [15].

In recent years, machine learning (ML) has emerged as a novel approach for seagrass mapping and monitoring [6]. Machine learning has the benefits of rapid learning, accommodation of non-linearity [16], and the availability of an increasing number of new, open source algorithms [17]. In the field of seagrass mapping and monitoring, however, the application of machine learning is still in its infancy [6]. Examples used to date include weighted majority voting using Quickbird images [18]; logistic model trees (LMT), AdaBoost, random forest (RF), and artificial neural networks (ANN) using digital images [19]; support vector machine (SVM) using Sentinel-2 images [13,20]; and decision trees (DTs) using aerial photographs [21]. In these examples, when used with high spatial resolution images (<1 m), machine learning models achieved an accuracy of 92–100%. Decision tree models using aerial photographs, however, achieved a lower accuracy of 66% for seagrass meadows when the plant cover was below 60% [21]. These mixed results support the exploration of novel machine learning approaches, particularly for improving low coverage seagrass mapping.

Among the various DT ensemble machine learning algorithms, rotation forest (RoF) and canonical correlation forest (CCF) algorithms are now emerging as reliable techniques for land cover mapping [22], landslide mapping using multi-spectral [23] or hyper-spectral [24] imagery, and rapid building mapping using multi-source data [25]. Using bootstrap sampling, combining multiple independent base classifiers, and applying statistical analysis (principal component analysis in the RoF model) [26,27], these learning algorithms are well-known for reducing the variance and overfitting of the classification results, resulting in a better detection of multi-class boundaries [28–30]. In addition, the CCF model does not require the optimization of hyperparameters [31], which makes this model simpler to apply for mapping tasks. To our knowledge, these techniques have not been used for seagrass mapping, however, they potentially offer benefits in the classification of low coverage through enhanced recognition of edge boundaries. Therefore, our goal in this study was to compare the use of three ML algorithms, RF, RoF, and CCF, to the more traditional MLC approach for mapping the aboveground distribution of seagrass communities at low and high coverage using Sentinel-2 data.

Our target was Tauranga Harbor, New Zealand, for which ground truth data were available, and which offers a mosaic of dense, sparse, and zero seagrass coverage. We discuss here the difference in performance of the selected models for seagrass detection at two densities. Our results are expected to contribute alternative solutions for the mapping and monitoring of seagrass at various regions in the world, and assist in the conservation of this important blue carbon ecosystem.

2. Materials and Methods

2.1. Study Site

Tauranga Harbor, North Island, New Zealand was selected as our study site (Figure 1). The site supports a single seagrass species, *Zostera muelleri*, distributed in the intertidal parts of the harbor [32]. *Z. muelleri* tolerates a wide range of salinity (10–30 practical salinity unit (psu)), however, a salinity at 12 psu is optimal to produce the highest shoot density [33]. *Z. muelleri* is a small plant when growing

intertidally as in the Tauranga Harbor, with leaves 5–30 cm in length and 0.1–0.4 cm in width. It has a maximum growth rate in the austral summer (December to March) with optimal temperature ranging from 27–33 °C [34–36]. Biomass declines gradually over winter to reach a minimum in early spring (October) [37]. Since satellite image based mapping uses surface reflectance as input data for the classification, our mapping only addressed the above-ground part of seagrass meadows, and may thus underestimate the colonized area if applied in late winter when the aboveground parts have senesced. Flowering and seed production in *Z. muelleri* is rare in New Zealand, and lateral spread is slow and by vegetative mechanisms [38,39].

Figure 1. Study site in Tauranga Harbor using a Red-Green-Blue (RGB) combination of Sentinel-2 scene (5/1/2019).

In Tauranga Harbor, the tide regime is semi-diurnal, with a range of 0.2–2.1 m. The size of the harbor means that tide timings are different in various areas in the harbor [40]. Examination of harbor bathymetry and seagrass distribution, which is almost exclusively intertidal in the harbor, indicates that seagrass was occupying a depth range of 0.0–1.5 m at the Sentinel-2 acquisition time. Across the years 2018–2019, the mean air temperature ranged from 2.5 °C in winter to 31.6 °C in summer (data from the New Zealand Meteorological Service [41] whereas mean sea temperature ranged from 14 °C in winter to 23 °C in summer [42] .

Previous mapping using aerial photography and manual classifications for the years 1959, 1996, and 2011 [43] suggested a loss of ~50% of seagrass area, making a strong case for the ongoing monitoring of seagrass cover. In addition, a wide range of seagrass density and substratum makes Tauranga Harbor a suitable location for testing novel classification algorithms (Figure 2).

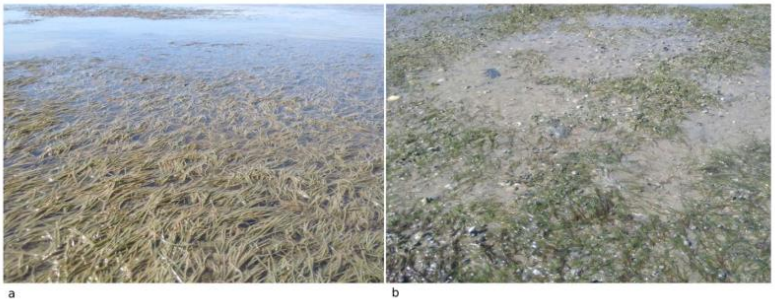


Figure 2. Dense (**a**) and sparse (**b**) seagrass meadows in Tauranga Harbor (photographs taken by Nam Thang Ha.

2.2. Field Survey

A seagrass mapping survey was undertaken between 1 and 7, April, 2019 (Figure 1) in the intertidal areas of the harbor. At low tide, the boundary of seagrass meadows was delimited using a Garmin Etrex 30 global positioning system (GPS) with an accuracy of ±2 m. Other substrata recorded during the field survey were bare sand and muddy sand. Macroalgae were neither detected from our field survey nor mentioned in previous mapping reports [32,43].

Ground truth points (GTPs) were recorded by following the boundary between seagrass meadows and unvegetated areas. In addition, the internal boundaries of seagrass and bare substrate were recorded if they coexisted in the same meadow. The frequency with which GTPs were recorded was related to the 10 m pixel size of Sentinel-2, and varied according to seagrass density. The number of GTPs was 2–5 GTPs per pixel in the case of patchy meadows, and decreased to one GTP per pixel or one GTP per 2–3 pixels for continuous, dense meadows. GTPs were not collected for seagrass meadows smaller than 100 square meters (corresponding to the pixel size of Sentinel-2 imagery), and some meadows could not be accessed due to logistic constraints.

Seagrass class boundaries were determined visually, with "dense" and "sparse" boundaries recorded. A meadow was identified as "dense" if the coverage was greater than 80% (Figure 2a), and "sparse" if the coverage was less than 80% (Figure 2b). A total of 4315 GTPs were recorded, with 2751 and 1564 for sparse and dense meadows, respectively; 237 GTPs were recorded for other substrata.

2.3. Satellite Data Acquisition and Image Pre-Processing

A Sentinel-2 scene acquired on May 1, 2019 was selected and downloaded from the GLOVIS website [44] (Table 1). The Sentinel-2 scene was pre-processed at level 1C (atmospheric correction at the top of atmosphere), and in the projection of WGS-84 UTM 60S. Sentinel pixels were classified into non-seagrass, sparse, and dense seagrass classes according to our field observations. Field and remote

sampling were closely synchronous, and we considered the field data to provide a sufficiently accurate representation of seagrass spatial distribution to develop the models.

Table 1. Sentinel-2 data acquisitions used for seagrass mapping in this study.

Date of Acquisition	Time of Acquisition [a]	Spatial Resolution (m)	Cloud Coverage (%)	First Low Tide	Second Low Tide
5/1/2019	10:16 AM	10	0	10:33 AM	22:52 PM

[a] Local New Zealand summer time (NZDT).

2.3.1. Atmospheric Correction

Atmospheric correction was executed in a Python™ environment using the dark spectrum fitting method in ACOLITE [45]. This fast and free-to-download tool has been adapted for aquatic application and presents a reliable atmospheric correction for Landsat and Sentinel-2 [46]. This process converts pixel values from the top of atmosphere (at level 1C) to surface reflectance for water pixels.

The values and options of selected parameters are presented in Table 2. For our study site, sun glint was not observed in the acquired scene. Therefore, the parameter of sun glint correction was set to False. Short wave infrared spectral range (SWIR), a spectral band of Sentinel-2 with the wavelength close to 1600 nm, was used to mask land and cloud pixels from water pixels. Corrected surface reflectance for water pixels of blue (ρ_{w443}), green (ρ_{w560}), and red (ρ_{w665}) bands were used for the next step of water column correction.

Table 2. Selected parameters for atmospheric correction using ACOLITE.

Parameters	Values
Ancillary data	
Gas transmittance	True
Ozone concentration (cm^{-1})	0.3
Water vapor concentration (g/cm^2)	1.5
Pressure	Normal pressure
Masking	
Level 2 water masking (nm)	1600
Negative reflectance masking	True
Cirrus masking	True
Other parameters	
Sky correction	True
Dark spectrum fitting	Fixed
Sun glint correction	False
Output parameter	
Surface reflectance for water pixel (ρ_w)	ρ_{w443} ρ_{w560} ρ_{w665}

2.3.2. Water Column Correction

Satellite image acquisition did not coincide with low tide in Tauranga Harbor (Table 1), and most, but not all, seagrass areas were inundated at the time of acquisition. We therefore applied a water column correction. Previous studies suggested that visible wavelengths are the most sensitive to seagrass meadows [47–49] and penetrate well into water [47]. Several studies have indicated that the NIR band is rapidly absorbed in an underwater environment [20,50], and may therefore contribute noise to the image, leading to a low accuracy of underwater habitat detection [48,51]. As a result, we decided upon the use of only visible spectra in the current study without the NIR band. After the water column correction step, water pixels in the blue (458–523 nm), green (543–578 nm), and red (650–680 nm) spectral bands were selected as input data for the evaluation of the model's performance.

Water column correction was conducted using bottom reflectance index (BRI), as proposed by Sagawa et al. (2010) [52]. Instead of absolute values of bottom reflectance, this approach creates different indexes for various bottom types, which are used for the step of classification. The index is calculated using Equation (1):

$$BRI_i = \frac{SR_i}{e^{-ki*g*z}} \quad (1)$$

where BRI_i is the bottom reflectance index of band i; SR_i is the surface reflectance for water pixel of band i (ρ_{wi}); k_i is the attenuation coefficient of solar radiance in water column (m^{-1}) of band i; g is a geometric factor accounting for the path length through the water; and z is the water depth (m).

Values of SR and k of blue, green, and red bands were retrieved from the atmospheric correction step. Water depth (z) was extracted from the bathymetry published by the National Institute of Water and Atmospheric Research (NIWA). g was calculated using Equation (2) [53]:

$$g = \frac{1}{sec(SolarZenithAngle) + sec(SatelliteNadirAngle)}, \quad (2)$$

where

$$sec(SolarZenithAngle) = \frac{1}{cos(SolarZenithAngle)}$$

$$sec(SatelliteNadirAngle) = \frac{1}{cos(SatelliteNadirAngle)}$$

and g was calculated as 0.0245 for the Sentinel-2 scene at the study site.

2.4. Image Classification with Machine Learning Ensemble-Based and Maximum Likelihood Methods

2.4.1. Selection of Maximum Likelihood, Random Forest, Rotation Forest, and Canonical Correlation Forest Classifiers

MLC is the most popular classification method in remote sensing [6] and is based on probability theory. This model requires a normal distribution, equal covariance, and a sufficient number of training samples [54] to maintain a reliable result. Class mean vector and covariance matrices are used to minimize the class distance and maximize the probability of a feature belonging to the selected class by using quadratic or linear discrimination functions. This method, however, may overfit posterior probabilities and result in inaccurate classification when the assumptions are violated.

RF [55] builds a forest of decision trees from a bootstrap sampling of training data points, and uses only a selection of all the samples for each decision tree. For each subset, a decision tree is built using two thirds of the total samples for training and one third for validation of the RF model. Majority voting is applied to arrange labels to given classes. The RF model constructs and combines a large number of base-decision trees using the classification and regression tree (CART) algorithm. Known as a robust and consistent model, RF is the most popular ensemble based model for classification problems, and, in particular, for seagrass mapping [6].

In 2006, Rodriguez et al. (2006) presented RoF, a feature extraction based method [56]. RoF is expected to provide an alternative selection for both regression and classification tasks [28]. The training data created by a bootstrap sampling is split into K subsets and then principal component analysis (PCA) is applied for each subset. All the principal components are retained and a number of decision trees are built from these transformed datasets. RoF can perform better than RF with a smaller number of trees, and therefore reduces the time for running the model. Instead of estimating the average value from all the trees in RF, a confidence value is calculated to assign a label to a given class with the highest value of confidence.

CCF creates a number of canonical correlation trees (CCTs) using canonical correlation analysis (CCA) to maximize the correlation between the input data and the selected labels. The authors in [31] confirmed the robustness of the model and also introduced projection bootstrapping, which uses all of

the data and improves the prediction accuracy over RF models. CCF is different from RF and RoF, as it lacks the tuning hyper-parameter, but offers similar accuracy with a smaller number of trees [31] and therefore, may consume less computation time for the training phase.

2.4.2. Training and Testing Dataset

Differences exist between the collection date of the GTPs and the acquisition date of the satellite image and, in the variable marine environment, this may be inevitable [57–60]. The sampling dates in April and May both fell in the austral fall. The difference between the field survey (1–7, April 2019) and the acquisition date of the Sentinel-2 image (1 May, 2019) was approximately 23 days, which was considered acceptable when compared with various reports in the literature [11,58–60]. Therefore, the GTPs were reliable to support the selection of image pixels for the training and testing datasets. Following the GTPs, the regions of interest (pixels) of three classes (dense seagrass, sparse seagrass, and non-seagrass) were selected and randomly split into 60% for training and 40% for the testing phases, and used as a unique input for all selected models (Table 3).

Table 3. Number of pixels for training and testing at various acquisition dates.

Sentinel Acquisition Date	Number of Pixels	
	60% for training	40% for testing
5/1/2019	8586	5724

2.4.3. Use of Maximum Likelihood, Random Forest, Rotation Forest, and Canonical Correlation Forest Models

The optimization and performance of the RF, RoF, and MLC algorithms for image classification were conducted in a Python™ environment using a Jupyter notebook as the code editor. RF and RoF codes were sourced from Scikit-learn [61], GitHub [62], and MLC from mlpy [63,64]. For the RF and RoF models, optimization of the hyper-parameters used a grid search with fivefold cross-validation [26,28]. The optimization and performance of the CCF model was conducted in the MATLAB environment using the source code of Rainforth and Wood [31,65]. The CCF models were performed with 10, 30, 50, 100, 200, and 500 trees, and an optimal number was selected based on the lowest misclassification rate, highest Kappa coefficient, and an acceptable computation time.

The results of the CCF model were exported to a CSV format for model comparisons in Python™. All source codes were open access, and the codes developed in this study will be submitted to GitHub. Image processing and classification were performed using a desktop computer with four 3.8 GHz physical cores and 16 Gb RAM.

2.4.4. Evaluation Criteria

Equations (3)–(7) involving accuracy, Kappa coefficient, precision, recall, and F1 were used to compare the performance of the selected models.

$$accuracy\left(y, y_{pred}\right) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} 1\left(y_{predi} = y_i\right) \quad (3)$$

where y_{pred} is the predicted value and y is the corresponding true value

$$Kappa = \frac{p_o - p_e}{1 - p_e} \quad (4)$$

where p_o is the observed agreement ratio and p_e is the expected agreement

$$Precision = \frac{tp}{tp + fp} \tag{5}$$

$$Recall = \frac{tp}{tp + fn} \tag{6}$$

$$F1 = \frac{2 * precision * recall}{precision + recall} \tag{7}$$

where tp is the true positive; fp is the false positive; and fn is the false negative.

In addition, the non-parametric McNemar test was used to statistically compare the accuracy of the selected models in this research. The test was executed in a Python™ environment using the mlxtend library [66]. The chi-square value (χ^2) was calculated from Equation (8) with Edward's continuity correction.

$$\chi^2 = \frac{\left(|fn - fp| - 1\right)^2}{(fn + fp)} \tag{8}$$

where fn is the false negative and fp is the false positive.

3. Results

Pixels were selected from each class, which had been verified by ground truth samplings, for the evaluation of automated classification. The main challenge to the automated classification that emerged relate to the mixing of sparse seagrass and bare sand in the same meadow. In particular, the substratum in very shallow water belonging to the non-seagrass class could be confused with sparse seagrass.

3.1. Hyper-Parameter Tuning for Random Forest, Rotation Forest, and Optimizing the Number of Trees for Canonical Correlation Forest Models

The hyper-parameters were tuned and consistently maintained for RF and RoF running during the training and testing phases (Table 4).

Table 4. Hyper-parameters selected for use in the RF and RoF models.

	RF	**RoF**
Bootstrap	True	True
Max depth	20	30
Max feature	Auto	Auto
Min sample leaf	1	3
Min sample split	9	7
Number of tree	100	100
Number of subset		3

For the CCF model, the lowest misclassification rates, and highest Kappa coefficient values were recorded at 200 trees (Figure 3a,d). As a result, we selected 200 trees as an optimal choice for CCF, which gave a computation time for the training and testing runs of 27 and 4.5 s, respectively; these times increased linearly with the number of trees (Figure 3b,c).

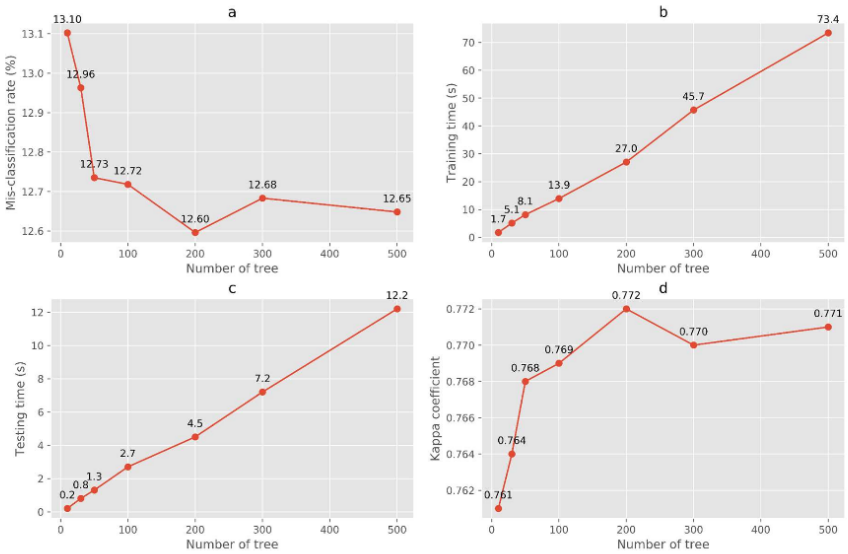


Figure 3. Number of trees optimizing for CCF model: mis-classification rate (**a**), computation time (**b** and **c**, the training and testing times are the computation time measured for the training and testing phases, respectively), and Kappa coefficient (**d**) using data described in Table 3.

3.2. Comparing the Performance of Random Forest, Rotation Forest, Canonical Correlation Forest, and Maximum Likelihood Models for Seagrass Mapping

The ML methods consistently outperformed the MLC model for all evaluation metrics (Table 5), and the McNemar test indicated these effects were significant (Table 6). In particular, the precision values of the ML models were similarly high, and greater than that obtained by the MLC method for dense and sparse seagrass, whilst very high recall was observed for both classes using the MLC model. F1 values of the MLC were lower than the ML methods for both dense and sparse seagrass. Among the ML methods, RoF outperformed the CCF and RF models (Figure 4 and Table 5), and this difference was also statistically significant (Table 6). The RoF model showed the highest values for precision and F1 across all three ML methods for both dense and sparse seagrass classes (Table 5).

Table 5. Accuracy, precision, recall, and F1 of the model performing for the selected Sentinel-2 scene; best performance is indicated by the characters in bold.

Model	Accuracy	Precision DS [a]	Precision SS [a]	Recall DS	Recall SS	F1 DS	F1 SS	Training time (s)	Prediction time (s)
RoF	**0.88**	**0.90**	**0.74**	0.92	0.75	**0.91**	**0.75**	3.54	0.37
CCF	0.87	**0.90**	**0.74**	0.89	0.72	0.90	0.73	27.06	4.57
RF	0.87	0.89	0.73	0.92	0.74	0.90	0.73	0.07	0.00
MLC	0.51	0.63	0.34	**0.93**	**0.95**	0.75	0.50	0.00	0.01

[a] DS: dense seagrass and SS: sparse seagrass classes.

Table 6. McNemar test comparing the performance of selected models in prediction of seagrass class.

	χ^2	*p* Value
	Scene 05/01/2019	
RoF–RF	9.03	0.00
RoF–MLC	1667.29	0.00
RoF–CCF	2069.37	0.00
RF–MLC	1599.46	0.00
RF–CCF	1995.93	0.00
CCF–MLC	42.51	0.00

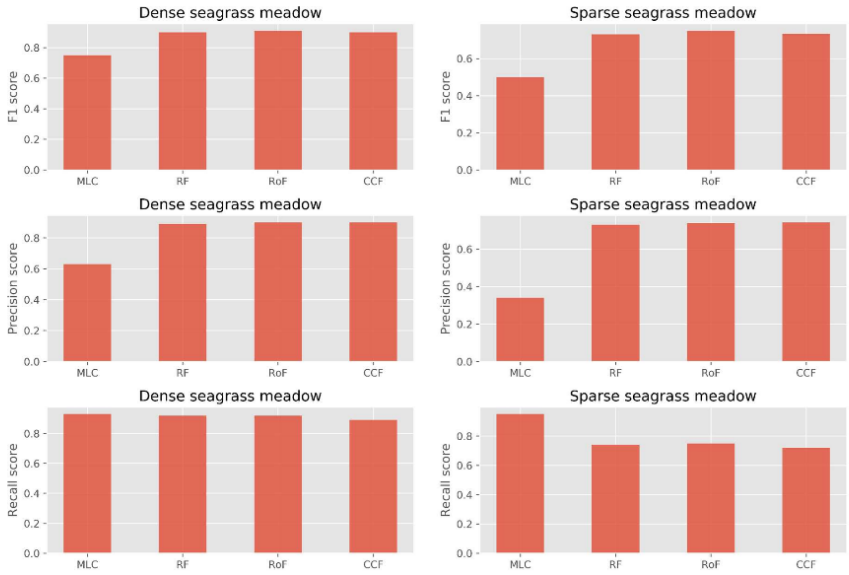


Figure 4. A comparison of the F1, precision, and recall scores for the Sentinel-2 scene (5/1/2019) using ensemble-based and traditional approaches.

All ensemble-based ML models were able to detect dense seagrass meadows with very high F1 scores ranging from 0.90–0.91 whereas the MLC model produced a F1 score of 0.75 (Table 5). The performance of these ML models was consistent with a balance of high precision and recall. The good performance for the dense seagrass class was conceivably due to a spectral separation of this class from the sparse and non-seagrass classes. As can be seen from Table 5, the RoF model improved the accuracy of the classification in terms of F1 score by 1% compared to the CCF and RF, but by nearly 17% over the MLC model for the dense seagrass class. Corresponding improvements were 3% (CCF, RF) and 33% (MLC) for the sparse seagrass class.

As indicated above, the classification was more challenging when a mixture of seagrass and bare sand were present in the same meadow. The RoF model still produced the highest F1 score (0.75), followed by the CCF and RF (0.73) models (Table 5). Conversely, the MLC model yielded a F1 score of only 0.50, which was significantly lower than the three ML models. With respect to computation time, the CCF model requires more time to run whilst the MLC model executes both the training and testing phases very quickly (Table 5).

As the RoF model yielded the highest performance assessment for seagrass mapping in this research, we employed this model to create the classified maps (Figure 5). The distribution of seagrass

meadows was consistent in the middle and southern part of the harbor, which may suggest optimal sites for blue carbon assessment as well as potential targets for the long-term conservation of seagrass in Tauranga Harbor. The seagrass area was estimated as approximately 1027.59 ha in May 2019.

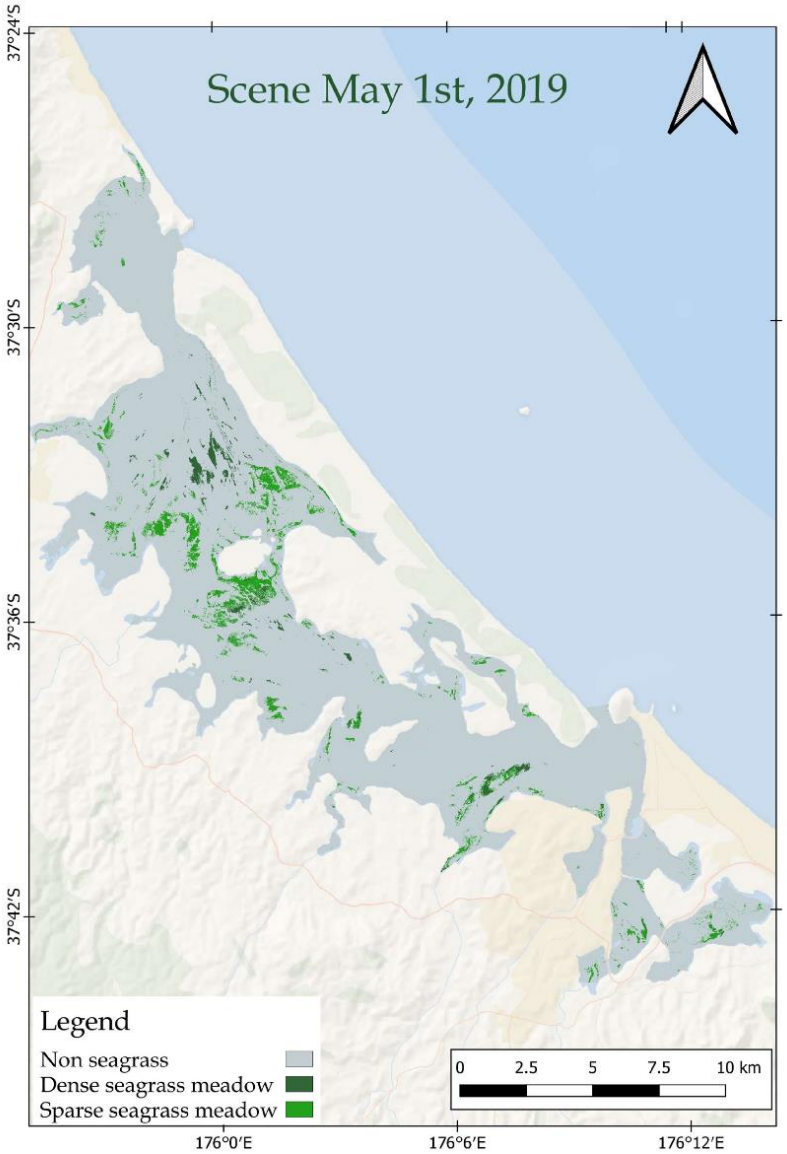


Figure 5. Seagrass map for Tauranga Harbor, 1 May 2019, derived using the RoF model applied to Sentinel-2 imagery.

4. Discussion

As far as we are aware, this research is the first attempt to compare the performance of the RF, RoF, CCF, and MLC methods for seagrass mapping with a full radiometric correction of the image. Desirable characteristics for seagrass mapping are both high precision and recall. High precision means that the classifier is able to detect the seagrass pixels precisely, whilst high recall means that the classifier is able to find all possible pixels of seagrass. To give a final coherence score and harmonize the values of precision and recall, the F1 score is usually preferred to evaluate a model's performance. The research presented here suggests that ML models detect dense seagrass meadows well, and outperform the traditional MLC approach.

Of the machine learning ensemble approaches used, the CCF and RF models performed less well than the RoF model, contradicting a superior performance of CCF in other studies [31]. CCF produced lower recall whilst RF created a lower precision for both the dense and sparse seagrass classes. For sparse seagrass meadows, CCF detected more precisely than the RF model. In addition, MLC produced very high recall, but low precision scores for both seagrass classes, leading to a lower F1 score and accuracy than the ensemble based models. Overall, our results show that RF, RoF, and CCF are good performers with a balance of high precision and recall scores, whilst very low precision scores of dense and sparse seagrass classes ranging from 0.34–0.63 were found for MLC. These results confirm the robustness and consistency of machine learning ensemble based methods in comparison to the MLC. We hypothesize that the poor performance of MLC may be because of the need for input data to satisfy the built-in assumptions, described above, which are difficult to sustain in a spatially heterogeneous marine environment.

Of the methods tested here, only the RF technique has previously been applied to seagrass mapping using very high spatial resolution imagery. In that case, high precision (0.947) and recall (0.968) values were determined mapping *Posidonia oceanica* from digital airborne images, though no comparison to other methods was attempted [19]. In another seagrass study, the overall accuracy only reached 82% using the RF algorithm applied to RapidEye imagery [67]. Considering the size of the seagrass meadows and the mix of substrate in Tauranga Harbor, the measured scores in our results were reliable for both dense and sparse seagrass mapping using medium spatial resolution of Sentinel-2 data (10 m pixel size). In other studies that allowed for a comparison of RF, RoF, and CCF models, CCF slightly outperformed the RF and RoF models for land cover mapping [22], RoF outperformed RF and CCF for mangrove mapping [68] whilst a similar performance of RoF and CCF models was noted for landslide mapping [23].

In addition to the performance advantages, RF and RoF are easy to execute in Python™, whilst CCF is confined to the MATLAB environment. The open source operating environment and diverse Python™ libraries provide multiple solutions for seagrass mapping and monitoring, and enhance the capacity to develop novel algorithms for various tasks in marine science [69]. Several libraries in the Python™ environment support a built-in framework for classification problems with a long list of state-of-the-art machine learning algorithms [70]. This ease-of-use approach allows a person with minimum programming skills to make a classification more reliable with machine learning. Recently, cloud computing broadens the executable environments for the big Earth observation data, especially in coastal resource mapping. A cloud computation system is able to deal with massive amounts of remote sensing datasets, parallel processing of satellite image using multiple data centers, and can respond to real time monitoring on country and global scales [71,72]. The use of open source machine learning algorithms in Python™ and a cloud system, therefore, promises large scale and more reliable mapping in the future.

Our results have validated the performance of the RF, RoF, and CCF models for seagrass mapping and suggest that the RoF technique is a promising novel approach to further seagrass monitoring at various sites around the world. However, the current study does have some limitations. The mismatch between the number of ground truth points (GTPs) and Sentinel-2 pixel size (10 × 10 m) may raise a degree of uncertainty in classification, particularly for sparse seagrass. However, we considered that

despite this mismatch, a sufficient and representative number of GTPs for each class was collected during the field survey (as presented in Section 2.2) to have confidence in the classification. Related issues that might explain the low values of precision and recall for sparse seagrass meadows are issues of mixed pixels, whereby small seagrass patches or dispersed clumps within a pixel challenge the classification process. To our knowledge, these effects are not easy to compensate for in the case of low to very low coverage seagrass using Sentinel-2 imagery. Thus, the use of very high spatial resolution sensors such as WorldView (~0.3–0.4 meters) [57,58] or Pleiades-1 (0.5 meters) is currently being investigated for future studies for seagrass mapping. Moreover, with the development of computer vision and pattern recognition, deep learning approaches using a variety of algorithms such as convolutional neural networks (CNNs or recurrent neural networks (RNNs) for semantic segmented imagery applied sub-pixel techniques should be encouraged for future studies [6].

5. Conclusions

We tested the performance of ML ensemble-based and MLC methods for seagrass mapping from Sentinel-2 data. Using Tauranga Harbor as a validation site, our comparison indicated that all ML-based approaches significantly outperformed MLC. MLC failed to detect sparse seagrass meadows, with a low F1 score of 0.50. We noted a better performance of RoF compared to the RF and CCF models with the highest F1 scores of 0.91 and 0.75 for dense and sparse seagrass classes, respectively.

Our results attest to the reliable application of the RoF model for the mapping and monitoring of seagrass in shallow water using Sentinel-2 imagery. Despite a lower accuracy for sparse than dense seagrass meadow classification, the CCF model shows potential for the mapping of seagrass and merits further testing at various scales and in various case studies. Regarding MLC, this model is still an applicable candidate for dense seagrass meadows, however, it may not be applicable for the mapping of sparse to very sparse seagrass meadows.

Author Contributions: All authors have read and agree to the published version of the manuscript. Conceptualization, N.T.H. and I.H.; methodology, N.T.H.; software, N.T.H.; validation, N.T.H. and T.D.P; resources, N.T.H, I.H, M.M.-H.; writing-original draft preparation, N.T.H.; writing-review and editing, N.T.H., T.D.P., M.M.-H., and I.H.; supervision, I.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: Bathymetry data were supplied by the National Institute of Water and Atmospheric Research (NIWA): Reeve G., Stephens S.A., Wadhwa S. 2018. Tauranga Harbor inundation modeling. NIWA Client Report 2018269HN to Bay of Plenty Regional Council, December 2018, p. 107. Field surveys were completed with a support from Christine E.C. Gunfield and Matthew J. Finnigan in Marine Field Station, Tauranga, New Zealand.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gullström, M.; Lyimo, L.D.; Dahl, M.; Samuelsson, G.S.; Eggertsen, M.; Anderberg, E.; Rasmusson, L.M.; Linderholm, H.W.; Knudby, A.; Bandeira, S.; et al. Blue carbon storage in tropical seagrass meadows relates to carbonate stock dynamics, plant–sediment processes, and landscape context: Insights from the Western Indian ocean. *Ecosystems* **2018**, *21*, 551–566.
2. Oreska, M.P.J.; McGlathery, K.J.; Porter, J.H. Seagrass blue carbon spatial patterns at the meadow-scale. *PLoS ONE* **2017**, *12*, e0176630. [CrossRef] [PubMed]
3. Duarte, C.M.; Krause-Jensen, D. Export from seagrass meadows contributes to marine carbon sequestration. *Front. Mar. Sci.* **2017**, *4*. [CrossRef]
4. Waycott, M.; Duarte, C.M.; Carruthers, T.J.B.; Orth, R.J.; Dennison, W.C.; Olyarnik, S.; Calladine, A.; Fourqurean, J.W.; Heck, K.L.; Hughes, A.R.; et al. Accelerating loss of seagrasses across the globe threatens coastal ecosystems. *Proc. Natl. Acad. Sci.* **2009**, *106*, 12377–12381. [CrossRef]
5. Herold, M.; Skutsch, M. Monitoring, reporting and verification for national REDD + programmes: Two proposals. *Environ. Res. Lett.* **2011**, *6*, 014002. [CrossRef]

6. Pham, T.D.; Xia, J.; Ha, N.T.; Bui, D.T.; Le, N.N.; Tekeuchi, W. A review of remote sensing approaches for monitoring blue carbon ecosystems: Mangroves, seagrassesand salt marshes during 2010–2018. *Sensors* **2019**, *19*, 1933. [CrossRef]
7. *ESA Sentinel—2 User Handbook*; ESA: Paris, France, 2015; p. 64.
8. Hossain, M.S.; Bujang, J.S.; Zakaria, M.H.; Hashim, M. The application of remote sensing to seagrass ecosystems: An overview and future research prospects. *Int. J. Remote Sens.* **2015**, *36*, 61–114. [CrossRef]
9. Winters, G.; Edelist, D.; Shem-Tov, R.; Beer, S.; Rilov, G. A low cost field-survey method for mapping seagrasses and their potential threats: An example from the northern Gulf of Aqaba, Red Sea: Mapping seagrasses and their potential threats in the Gulf of Aqaba. *Aquat. Conserv. Mar. Freshw. Ecosyst.* **2017**, *27*, 324–339. [CrossRef]
10. Gumusay, M.U.; Bakirman, T.; Tuney Kizilkaya, I.; Aykut, N.O. A review of seagrass detection, mapping and monitoring applications using acoustic systems. *Eur. J. Remote Sens.* **2019**, *52*, 1–29. [CrossRef]
11. Wicaksono, P.; Lazuardi, W. Assessment of PlanetScope images for benthic habitat and seagrass species mapping in a complex optically shallow water environment. *Int. J. Remote Sens.* **2018**, *39*, 5739–5765. [CrossRef]
12. Poursanidis, D.; Topouzelis, K.; Chrysoulakis, N. Mapping coastal marine habitats and delineating the deep limits of the Neptune's seagrass meadows using very high resolution Earth observation data. *Int. J. Remote Sens.* **2018**, 1–18. [CrossRef]
13. Poursanidis, D.; Traganos, D.; Reinartz, P.; Chrysoulakis, N. On the use of Sentinel-2 for coastal habitat mapping and satellite-derived bathymetry estimation using downscaled coastal aerosol band. *Int. J. Appl. Earth Obs. Geoinformation* **2019**, *80*, 58–70. [CrossRef]
14. Asmala, A. Analysis of Maximum Likelihood Classification on Multispectral Data. *Appl. Math. Sci.* **2012**.
15. Richards, J.A. Supervised Classification Techniques. In *Remote Sensing Digital Image Analysis*; Springer: Berlin/Heidelberg, Germany, 2013; ISBN 978-3-642-30062-2.
16. Holloway, J.; Mengersen, K. Statistical machine learning methods and remote sensing for sustainable development goals: A review. *Remote Sens.* **2018**, *10*, 1365. [CrossRef]
17. Liu, Y. *Python Machine Learning by Example: Easy-to-follow Examples that Get You up and Running with Machine Learning*; Packt Publishing: Birmingham, UK; Mumbai, India, 2017; ISBN 978-1-78355-311-2.
18. Mohamed, H.; Nadaoka, K.; Nakamura, T. Assessment of machine learning algorithms for automatic benthic cover monitoring and mapping using towed underwater video camera and high-resolution satellite images. *Remote Sens.* **2018**, *10*, 773. [CrossRef]
19. Bonin-Font, F.; Campos, M.M.; Codina, G.O. Towards visual detection, mapping and quantification of Posidonia Oceanica using a lightweight AUV. *IFAC-Pap.* **2016**, *49*, 500–505. [CrossRef]
20. Traganos, D.; Reinartz, P. Mapping Mediterranean seagrasses with Sentinel-2 imagery. *Mar. Pollut. Bull.* **2017**. [CrossRef] [PubMed]
21. Pe'eri, S.; Morrison, J.R.; Short, F.; Mathieson, A.; Lippmann, T. Eelgrass and macroalgal mapping to develop nutrient criteria in New Hampshire's estuaries using hyperspectral imagery. *J. Coast. Res.* **2016**, *76*, 209–218. [CrossRef]
22. Colkesen, I.; Kavzoglu, T. Ensemble-based canonical correlation forest (CCF) for land use and land cover classification using sentinel-2 and Landsat OLI imagery. *Remote Sens. Lett.* **2017**, *8*, 1082–1091. [CrossRef]
23. Sahin, E.K.; Colkesen, I.; Kavzoglu, T. A comparative assessment of canonical correlation forest, random forest, rotation forest and logistic regression methods for landslide susceptibility mapping. *Geocarto Int.* **2018**, *33*, 1–23. [CrossRef]
24. Moughal, T.A. Hyperspectral image classification using Support Vector Machine. *J. Phys. Conf. Ser.* **2013**, *439*, 012042. [CrossRef]
25. Adriano, B.; Xia, J.; Baier, G.; Yokoya, N.; Koshimura, S. Multi-source data fusion based on ensemble learning for rapid building damage mapping during the 2018 sulawesi earthquake and tsunami in Palu, Indonesia. *Remote Sens.* **2019**, *11*, 886. [CrossRef]
26. Probst, P.; Wright, M.; Boulesteix, A.-L. Hyperparameters and tuning strategies for random forest. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2019**, *9*, e1301. [CrossRef]
27. Xiu, Y.; Liu, W.; Yang, W. An improved rotation forest for multi-feature remote-sensing imagery classification. *Remote Sens.* **2017**, *9*, 1205. [CrossRef]

28. Bagnall, A.; Bostrom, A.; Cawley, G.; Flynn, M.; Large, J.; Lines, J. Is rotation forest the best classifier for problems with continuous features? *arXiv* **2018**, arXiv:1809.06705.
29. Belgiu, M.; Drăguţ, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
30. Feng, W.; Sui, H.; Tu, J.; Huang, W.; Xu, C.; Sun, K. A Novel Change Detection Approach for Multi-Temporal High-Resolution Remote Sensing Images Based on Rotation Forest and Coarse-to-Fine Uncertainty Analyses. *Remote Sens.* **2018**, *10*, 1015. [CrossRef]
31. Rainforth, T.; Wood, F. Canonical Correlation Forests. *arXiv* **2015**, arXiv:150705444.
32. Park, S.G. *Changes in abundance of seagrass (Zostera spp.) in Tauranga Harbour from 1959–96*; Environmental Report 99/30; Environment BOP: Whakatane, New Zealand, 1999; p. 19.
33. Collier, C.J.; Villacorta-Rath, C.; van Dijk, K.; Takahashi, M.; Waycott, M. Seagrass proliferation precedes mortality during hypo-salinity events: A stress-induced morphometric response. *PLoS ONE* **2014**, *9*, e94014. [CrossRef]
34. Collier, C.J.; Ow, Y.X.; Langlois, L.; Uthicke, S.; Johansson, C.L.; O'Brien, K.R.; Hrebien, V.; Adams, M.P. Optimum Temperatures for Net Primary Productivity of Three Tropical Seagrass Species. *Front. Plant Sci.* **2017**, *8*. [CrossRef]
35. York, P.H.; Gruber, R.K.; Hill, R.; Ralph, P.J.; Booth, D.J.; Macreadie, P.I. Physiological and Morphological Responses of the Temperate Seagrass Zostera muelleri to Multiple Stressors: Investigating the Interactive Effects of Light and Temperature. *PLoS ONE* **2013**, *8*, e76377. [CrossRef] [PubMed]
36. Collier, C.J.; Uthicke, S.; Waycott, M. Thermal tolerance of two seagrass species at contrasting light levels: Implications for future distribution in the Great Barrier Reef. *Limnol. Oceanogr.* **2011**, *56*, 2200–2210. [CrossRef]
37. Turner, S.J. Growth and productivity of intertidal Zostera capricorni in New Zealand estuaries. *N. Z. J. Mar. Freshw. Res.* **2007**, *41*, 77–90. [CrossRef]
38. Ramage, D.L.; Schiel, D.R. Reproduction in the seagrass Zostera novazelandica on intertidal platforms in southern New Zealand. *Mar. Biol.* **1998**, *130*, 479–489. [CrossRef]
39. Schwarz, A.-M.; Turner, S. *Management and Conservation of Seagrass in New Zealand: An Introduction*; Wellington, New Zeland, 2006; p. 90.
40. Reeve, G.; Stephens, S.; Wadhwa, A. *Tauranga Harbour Inundation Modelling*; NIWA: Tauranga, New Zealand, 2018; p. 107.
41. Past Weather for Tauranga Airport. Available online: https://www.metservice.com/towns-cities/locations/tauranga/past-weather (accessed on 5 January 2020).
42. Tauranga Sea Temperature. Available online: https://www.seatemperature.org/australia-pacific/new-zealand/tauranga.htm (accessed on 5 January 2020).
43. Park, S. *Extent of Seagrass in the Bay of Plenty in 2011*; Environmental publication; Bay of Plenty Reginal Council: Whakatane, New Zealand, 2011.
44. Glovis. Available online: https://glovis.usgs.gov (accessed on 12 October 2019).
45. RBINS Acolite Atmospheric Correction Processor. Available online: https://odnature.naturalsciences.be/remsem/software-and-data/acolite (accessed on 1 October 2018).
46. Vanhellemont, Q. Adaptation of the dark spectrum fitting atmospheric correction for aquatic applications of the Landsat and Sentinel-2 archives. *Remote Sens. Environ.* **2019**, *225*, 175–192. [CrossRef]
47. Green, E.P.; Mumby, P.J.; Edwards, A.J.; Clark, C.D. A review of remote sensing for the assessment and management of tropical coastal resources. *Coast. Manag.* **1996**, *24*, 1–40. [CrossRef]
48. Thang, H.N.; Yoshino, K.; Hoang Son, T.P. *Seagrass mapping using ALOS AVNIR-2 data in Lap An Lagoon, Thua Thien Hue, Vietnam*; Frouin, R.J., Ebuchi, N., Pan, D., Saino, T., Eds.; SPIE: Kyoto, Japan, 2012; p. 85250S.
49. Garcia, R.; Hedley, J.; Tin, H.; Fearns, P. A method to analyze the potential of optical remote sensing for benthic habitat mapping. *Remote Sens.* **2015**, *7*, 13157–13189. [CrossRef]
50. *Remote Sensing Handbook for Tropical Coastal Management*; Green, E.P.; Edwards, A.J. (Eds.) Coastal management sourcebooks; Unesco Pub: Paris, France, 2000; ISBN 978-92-3-103736-8.
51. Chen, Q.; Yu, R.; Hao, Y.; Wu, L.; Zhang, W.; Zhang, Q.; Bu, X. A new method for mapping aquatic vegetation especially underwater vegetation in lake Ulansuhai using GF-1 satellite data. *Remote Sens.* **2018**, *10*, 1279. [CrossRef]

52. Sagawa, T.; Boisnier, E.; Komatsu, T.; Mustapha, K.B.; Hattour, A.; Kosaka, N.; Miyazaki, S. Using bottom surface reflectance to map coastal marine areas: A new application method for Lyzenga's model. *Int. J. Remote Sens.* **2010**, *31*, 3051–3064. [CrossRef]
53. Lyzenga, D.R.; Malinas, N.P.; Tanis, F.J. Multispectral bathymetry using a simple physically based algorithm. *IEEE Trans. Geosci. Remote Sens.* **2006**, *44*, 2251–2259. [CrossRef]
54. Hogland, J.; Billor, N.; Anderson, N. Comparison of standard maximum likelihood classification and polytomous logistic regression used in remote sensing. *Eur. J. Remote Sens.* **2013**, *46*, 623–640. [CrossRef]
55. Breiman, L. Random Forest. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
56. Rodriguez, J.J.; Kuncheva, L.I.; Alonso, C.J. Rotation Forest: A New Classifier Ensemble Method. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 1619–1630. [CrossRef] [PubMed]
57. Koedsin, W.; Intararuang, W.; Ritchie, R.; Huete, A. An Integrated Field and Remote Sensing Method for Mapping Seagrass Species, Cover, and Biomass in Southern Thailand. *Remote Sens.* **2016**, *8*, 292. [CrossRef]
58. Kovacs, E.; Roelfsema, C.; Lyons, M.; Zhao, S.; Phinn, S. Seagrass habitat mapping: How do Landsat 8 OLI, Sentinel-2, ZY-3A, and Worldview-3 perform? *Remote Sens. Lett.* **2018**, *9*, 686–695. [CrossRef]
59. Meyer, C.A.; Pu, R. Seagrass resource assessment using remote sensing methods in St. Joseph Sound and Clearwater Harbor, Florida, USA. *Environ. Monit. Assess.* **2012**, *184*, 1131–1143. [CrossRef]
60. Tsujimoto, R.; Terauchi, G.; Sasaki, H.; Sakamoto, S.X.; Sawayama, S.; Sasa, S.; Yagi, H.; Komatsu, T. Damage to seagrass and seaweed beds in Matsushima Bay, Japan, caused by the huge tsunami of the Great East Japan Earthquake on 11 March 2011. *Int. J. Remote Sens.* **2016**, *37*, 5843–5863. [CrossRef]
61. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
62. Joshua, L. Rotation Forest 2016. Available online: https://github.com/joshloyal/RotationForest (accessed on 13 February 2019).
63. Albanese, D.; Visintainer, R.; Merler, S.; Riccadonna, S.; Jurman, G.; Furlanello, C. Mlpy: Machine Learning Python. *arXiv* **2012**, arXiv:12026548.
64. Davide, A. Non Linear Methods for Classification: Maximum Likelihood Classifier. Available online: http://mlpy.sourceforge.net/docs/3.5/nonlin_class.html#maximum-likelihood-classifier (accessed on 15 February 2019).
65. Rainforth, T. Canonical Correlation Forests 2018. Available online: https://github.com/twgr/ccfs (accessed on 17 February 2019).
66. Raschka, S. MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. *J. Open Source Softw.* **2018**, *3*, 638. [CrossRef]
67. Traganos, D.; Reinartz, P. Interannual Change Detection of Mediterranean Seagrasses Using RapidEye Image Time Series. *Front. Plant Sci.* **2018**, *9*. [CrossRef] [PubMed]
68. Pham, T.D.; Xia, J.; Baier, G.; Le, N.N.; Yokoya, N. Mangrove Species Mapping Using Sentinel-1 and Sentinel-2 Data in North Vietnam. In Proceedings of the IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium; IEEE: Yokohama, Japan, 2019; pp. 6102–6105.
69. Lemenkova, P. Processing oceanographic data by python libraries Numpy, Scipy, and Pandas. *Aquat. Res.* **2019**, 73–91. [CrossRef]
70. Raschka, S.; Mirjalili, V. *Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow*, 2nd ed.; Packt Publishing: Birmingham, UK; Mumbai, India, 2017; ISBN 978-1-78712-593-3.
71. Yan, J.; Ma, Y.; Wang, L.; Choo, K.-K.R.; Jie, W. A cloud-based remote sensing data production system. *Future Gener. Comput. Syst.* **2018**, *86*, 1154–1166. [CrossRef]
72. Yao, X.; Li, G.; Xia, J.; Ben, J.; Cao, Q.; Zhao, L.; Ma, Y.; Zhang, L.; Zhu, D. Enabling the Big Earth Observation Data via Cloud Computing and DGGS: Opportunities and Challenges. *Remote Sens.* **2019**, *12*, 62. [CrossRef]

Appendix 3. Co-authorship form of research chapter 4

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

Postgraduate Studies Office
Student and Academic Services Division
Wahanga Ratonga Matauranga Akonga
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 4439
Website: http://www.waikato.ac.nz/sasd/postgraduate/

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 4: Detecting multi-decadal changes in seagrass cover in Tauranga Harbour, New Zealand, using Landsat imagery and boosting ensemble classification techniques

Nature of contribution by PhD candidate	Conceptualization, methodology, software, validation, resources, writing-original draft preparation, write-review and editing.
Extent of contribution by PhD candidate (%)	80

CO-AUTHORS

Name	Nature of Contribution
Merilyn Manley-Harris	Resources, writing-review and editing.
Tien Dat Pham	Validation, writing-review and editing.
Ian Hawes	Conceptualization, resources, writing-review and editing, supervision.

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and

Name	Signature	Date
Merilyn Manley-Harris		21.1.21
Tien Dat Pham		21.01.2021
Ian Hawes		24.02.2021

July 2015

Appendix 4. Published paper in the ISPRS International Journal of Geo-Information (chapter 4)

International Journal of Geo-Information

MDPI

Article

Detecting Multi-Decadal Changes in Seagrass Cover in Tauranga Harbour, New Zealand, Using Landsat Imagery and Boosting Ensemble Classification Techniques

Nam-Thang Ha [1,2], Merilyn Manley-Harris [1], Tien-Dat Pham [3,*] and Ian Hawes [1]

[1] Environmental Research Institute, School of Science, University of Waikato, Hamilton 3260, New Zealand; hanamthang@hueuni.edu.vn (N.-T.H.); manleyha@waikato.ac.nz (M.M.-H.); ian.hawes@waikato.ac.nz (I.H.)
[2] Faculty of Fisheries, University of Agriculture and Forestry, Hue University, Hue 530000, Vietnam
[3] Department of Biological Sciences, Florida International University (FIU), Miami, FL 33199, USA
* Correspondence: dat6784@gmail.com

Abstract: Seagrass provides a wide range of essential ecosystem services, supports climate change mitigation, and contributes to blue carbon sequestration. This resource, however, is undergoing significant declines across the globe, and there is an urgent need to develop change detection techniques appropriate to the scale of loss and applicable to the complex coastal marine environment. Our work aimed to develop remote-sensing-based techniques for detection of changes between 1990 and 2019 in the area of seagrass meadows in Tauranga Harbour, New Zealand. Four state-of-the-art machine-learning models, Random Forest (RF), Support Vector Machine (SVM), Extreme Gradient Boost (XGB), and CatBoost (CB), were evaluated for classification of seagrass cover (presence/absence) in a Landsat 8 image from 2019, using near-concurrent Ground-Truth Points (GTPs). We then used the most accurate one of these models, CB, with historic Landsat imagery supported by classified aerial photographs for an estimation of change in cover over time. The CB model produced the highest accuracies (precision, recall, F_1 scores of 0.94, 0.96, and 0.95 respectively). We were able to use Landsat imagery to document the trajectory and spatial distribution of an approximately 50% reduction in seagrass area from 2237 ha to 1184 ha between the years 1990–2019. Our illustration of change detection of seagrass in Tauranga Harbour suggests that machine-learning techniques, coupled with historic satellite imagery, offers potential for evaluation of historic as well as ongoing seagrass dynamics.

Keywords: seagrass mapping; Tauranga Harbour; change detection; landsat; random forest; support vector machine; extreme gradient boost; CatBoost; machine learning

check for updates

Citation: Ha, N.-T.; Manley-Harris, M.; Pham, T.-D.; Hawes, I. Detecting Multi-Decadal Changes in Seagrass Cover in Tauranga Harbour, New Zealand, Using Landsat Imagery and Boosting Ensemble Classification Techniques. *ISPRS Int. J. Geo-Inf.* **2021**, *10*, 371. https://doi.org/10.3390/ijgi10060371

Academic Editors: George P. Petropoulos and Wolfgang Kainz

Received: 15 April 2021
Accepted: 26 May 2021
Published: 31 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Seagrass provides a number of valuable ecosystem services in coastal areas, including primary production, biogenic habitat production, water filtering, wave energy attenuation, and sediment trapping [1,2]. In recent years, blue carbon, including seagrass meadows, has been acknowledged as an important service for climate change mitigation because of its value in the sequestration of carbon [3,4]. Seagrass meadows, however, have declined and degraded across most regions in the world, a change largely attributed to anthropogenic effects [5–7].

The destruction of seagrass leads to the loss of various ecosystem services [7,8] and threatens the stability [6] and long-term livelihood of the fisherman in coastal areas [9,10]. Therefore, an accurate and rapid technique to inventory this resource is in high demand [5,11,12], to contribute baseline data for the evaluation of coastal ecosystem dynamics, establishment of marine protected areas, and functional zoning fitting to the local conditions. Where this can include a historic perspective, it can provide a comprehensive understanding of system change.

Several attempts at mapping and monitoring seagrass meadows using different satellite sensors and approaches have been reported [12]. Change of seagrass cover has been assessed using RapidEye [13], Indian satellite image (IRS LISS IV) [14], WorldView-2, IKONOS, Quickbird-2 [15], and Landsat [16–19] in various parts of the globe including the Mediterranean, the USA, Australia, and Malaysia. The temporal range of these attempts is constrained by the various platform launch dates, and typically range from 5 to around 25 years. Few efforts have attempted a longer-term change detection (30–40 years) of seagrass, and accuracy assessment has not frequently been reported for such long-term change detection. The reasons for this may relate to a deficiency of ground-truth data against which to evaluate older satellite scenes, and a need for imagery for the development of robust models for the classification of seagrass meadows in variably submerged conditions to be captured at optimal times to allow traditional classification procedures to be applied.

In recent years, machine learning (ML) has been emerging as an effective approach in various classification tasks, including for seagrass mapping [12,20]. ML provides improvements over the traditional Earth observation (EO) data classification approaches, to deal better with the challenges of mixed habitat, coarse spatial resolution of satellite imagery, and water column and atmospheric interference in coastal habitats [20–22]. Advantages of ML models are their use of non-parametric approaches, requiring no assumptions of normal distribution of input data, effective use of noisy data, and capability for multiple feature extraction [23–26]. The application of ML techniques to multitemporal satellite data, gathered from different satellite platforms, may therefore improve the overall accuracy of the classification result and enhance the reliability of seagrass change assessment. A range of different ensemble-based supervised classification techniques, such as boosting and bagging approaches [21–24], have been considered and tested in the literature for this type of task [27,28]. The most important differences between the bagging and boosting methods come from the approaches to the creation of training and testing datasets, and how the bagging and boosting methods deal with weak learners during the learning process [29,30]. Despite the potential for improved classification accuracy in suboptimal datasets, these approaches have not yet been fully implemented for seagrass change detection [12]. We are aware of only a single study, using Random Forest (RF) classification, for mapping the change of seagrass cover [13]. In the case study reported by these authors, the performance of the model was unstable and the accuracy varied among acquired scenes. Here we test the performance of a range of ML models, both boosting and bagging methods, with a time-series of satellite images, to compare their performances for assessment of seagrass cover and long-term change in Tauranga Harbour. Our goal is to improve the accuracy of tools for seagrass mapping and change detection.

Landsat time-series data were selected for the current study as the longest available time series and as freely available satellite remotely sensed resources. Landsat has operated since 1972 and provides continuous, homogeneous input data up to the most current Landsat 8 operational land imager (OLI) in orbit [31]. The Landsat multitemporal data has been used previously for several long-term change detection tasks [12,32] with the combination of long-term acquisition, medium spatial resolution, and the high quality of atmosphere-corrected products cited as important attributes. The spatial resolution has been retained as 30 m through eight generations (Landsat 1–Landsat 8); however, the radiometric resolution has been improved from 8 bit to 12 bit, leading to a better recognition of surface objects [33]. In addition, Landsat imagery includes blue, green, and red wavebands, which are the most appropriate for underwater resource mapping [34–36], but have not yet been evaluated for long-term seagrass change detection [12]. Thus, our work attempts to fill a gap in the current literature by assessing the performance of historic Landsat imagery, coupled with various machine-learning boosting and bagging models implemented in an open-source environment, in mapping changes in seagrass extent in a tidally inundated environment.

We employed two well-known models, i.e., Support Vector Machine (SVM), Random Forests (RF); and two novel techniques, Extreme Gradient Boosting (XGB), and CatBoost

(CB) for the classification of seagrass meadows in Tauranga Harbour from Landsat imagery, and for detecting change across 29 years. The results demonstrated that the novel classification method CB was successful in describing the dynamics of change in seagrass in the study site as well as contributing baseline data for further assessment of change.

2. Materials and Methods

2.1. Study Site

We selected Tauranga Harbour (North Island, New Zealand) as the study site (Figure 1), due to its large size (201 km^2 in surface area [37]), variation in water depth (from 0 m when exposed to 20 m in deep channels [38]), widely distributed but patchy seagrass cover and the availability of historic ground-truth information. The tidal regime is semidiurnal, with a range of 0.2–2.1 m, and the estuary has an average water residence time of 3–8 days [39]. *Zostera muelleri* is the only species of seagrass, occurring primarily in the intertidal parts of the harbor [20,37]. The growth rate of *Z. muelleri* is optimal at 12 practical salinity units (psu) [40] and 27–33 °C [41,42]. It attains its highest biomass in the austral summer and declines gradually over the winter, reaching a minimum cover in early spring [43]. Flowering and seed production of *Z. muelleri* is rare in New Zealand, reproduction is primarily vegetative and patch dynamics are correspondingly slow [44,45]. Seagrass is primarily intertidal in the estuary and, based on bathymetry and tidal predictions [38] at the time of the Landsat image acquisition, water depths ranged between 0.0–1.5 m in the locations where seagrass was present.

In recent decades, Tauranga Harbour has been increasingly influenced by agricultural activities in the northern part (between 37.44° S and 37.54° S) and urban development in the southern part (between 37.62° S and 37.72° S). Episodic high loadings of sediment have been recorded and have resulted in the accumulation of sediment and high turbidity over the autumn and winter seasons [46,47]. Changes in the sedimentary environment have been implicated in negative impacts on the growth of seagrass [48,49], though other factors may also be involved. Available maps of seagrass in 1959, 1996, and 2011 derived from manual classification of aerial photography provided a resource for model validation [37].

2.2. Satellite Image Acquisition

Landsat images were downloaded from the GLOVIS website [50] for the years 1990, 2001, 2011, 2014, and 2019 (Table 1) at process level 1 (pixel value in digital number), and in the projection of WGS-84 UTM 60S. Landsat images were selected based on: (1) the acquired time of the Landsat image that coincided as closely as possible to low tide at the study site; (2) the image that had the lowest coverage of cloud; (3) whether there existed a similar acquisition month among the scenes. In practice, we selected scenes that ranged 1–2 months around March (Table 2).

Table 1. Landsat data acquisitions used for seagrass mapping and change detection.

Date of Acquisition (MM/DD/YYY)	Landsat Generation	Time of Acquisition [a]	Spatial Resolution (m)	Cloud Coverage (%)	First Low Tide [b]	Second Low Tide [b]
4 April 1990	Landsat 4 TM	10:16 a.m.	30	2	02:49 a.m.	15:09 p.m.
10 March 2001	Landsat 7 ETM+	10:16 a.m.	30	0	08:14 a.m.	20:35 p.m.
17 February 2011	Landsat 5 TM	10:15 a.m.	30	2	06:33 a.m.	18:57 p.m.
6 March 2014	Landsat 8 OLI	10:15 a.m.	30	0	11:41 a.m. [c]	
23 May 2019	Landsat 8 OLI	10:15 a.m.	30	0	04:14 a.m.	16:29 p.m.

[a]: Local time of New Zealand zone. [b]: Tide data was retrieved from the National Institute of Water and Atmospheric Research (NIWA). [c]: Only one low tide at the study site.

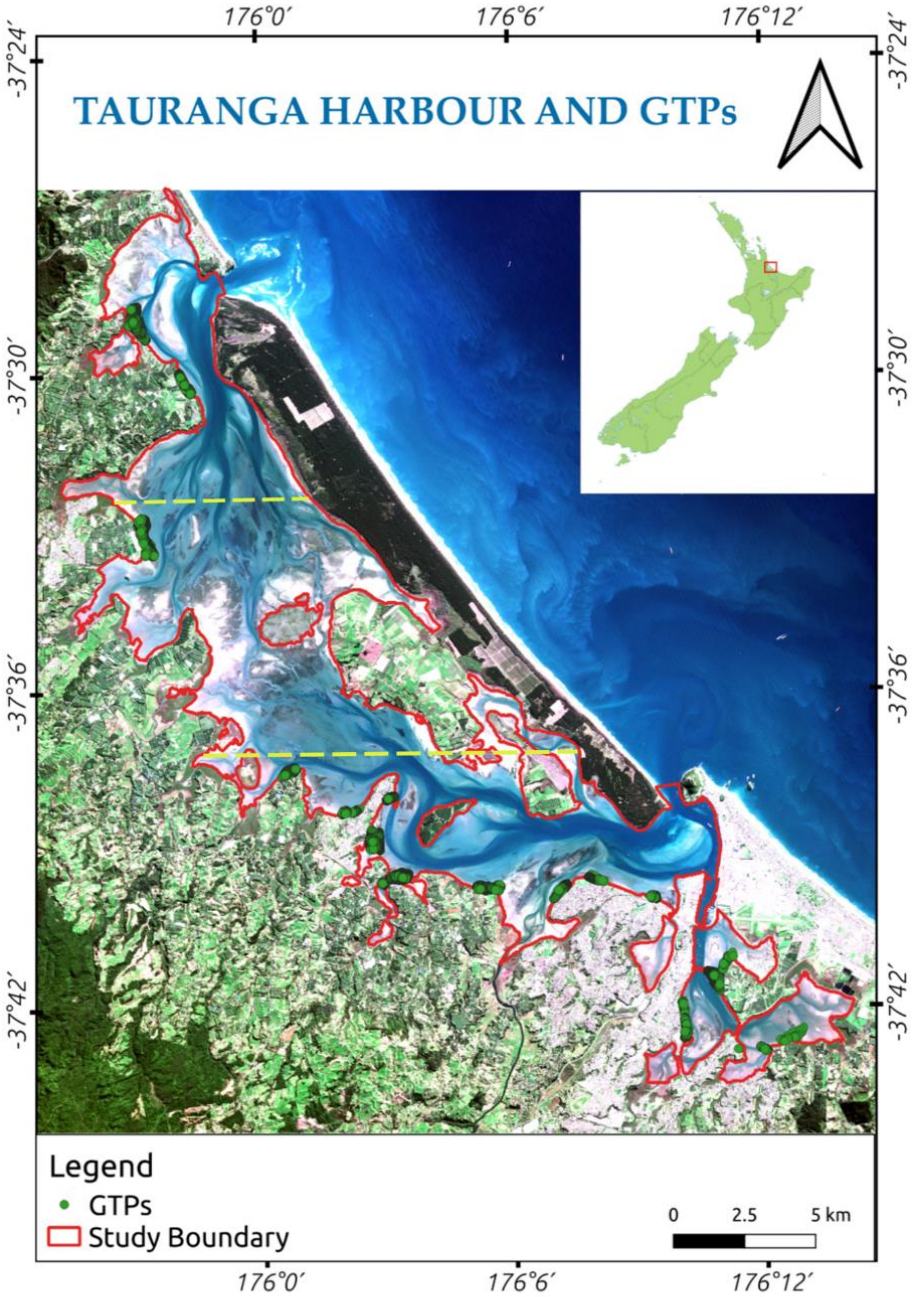


Figure 1. Tauranga Harbour—our study site (ρ_{Red}—ρ_{Geen}—ρ_{Blue} composited Landsat image, date on 23 May 2019). Ground-Truth Points, collected on 1–7 April 2019, are indicated by green circles (yellow lines indicate the boundaries of the northern, central, and southern harbor).

Table 2. Available aerial, Google Earth images corresponding to historic Landsat images acquisition.

Landsat Image Acquisition	Nearest Aerial Image Acquisition	Aerial Image Spatial Resolution (m)	Google Earth Image (Year of Acquisition)
April 1990	February 1991 March 1992	0.23	December 1990
March 2001	February 2003	0.23	December 2001
February 2011	February 2011	0.25	
March 2014	March 2014	0.125	

2.3. Field Survey Data

A field survey was undertaken from 1–7 April 2019 (Figure 1) in the intertidal areas of the harbor. At low tide, the boundary of seagrass meadows was delimited using a Global Positioning System (GPS) Garmin Etrex 30 with an accuracy of ±2 m. Other substrata recorded during the field survey were bare sand and muddy sand. Macroalgae were neither detected from our field survey nor mentioned in previous mapping reports [37,51].

Ground-Truth Points (GTPs), which were the base points to make the regions of interest (ROIs) for given classes, were recorded by following the boundary between seagrass meadows and non-vegetated areas. A total of 4315 GTPs were recorded for seagrass distribution, and 237 GTPs for other substrata in the harbor.

2.4. Ground-Truth Historical Scenes

Before 2019, no GTPs from field surveying were available, therefore we used aerial and Google Earth images (Table 2) and published documents [37,51] to identify regions of interests (ROIs), within which we were able to determine seagrass presence/absence with sufficient confidence to develop the models and to evaluate the accuracy of the hindcast seagrass maps. High-resolution aerial imagery exists from the years 2011 and 2014, and cloud-free, near-low-tide Landsat scenes, from February 2011 and March 2014, could be found that coincided with these. However, for the Landsat scenes in 1990 and 2001, aerial images were only available with a gap of 1–2 years. These included aerial images in 1991–1992 (monochrome and colour) and 2003 (colour). We found Google Earth images (identified as Landsat/Copernicus images in the Google Earth application) for both December 1990 and December 2001, which were in the austral summer and were close to the acquisition time of the Landsat scenes in April 1990 (austral autumn) and March 2001 (austral summer). Due to concerns over circularity of use of Landsat data, we used both Google Earth and aerial images to select the ROIs for Landsat scenes in 1990 and 2001, ensuring that ROIs were only used where both sources showed seagrass present. We considered that the slow dynamics of seagrass patches in Tauranga Harbour [44,45] made this approach robust.

2.5. Development of Seagrass Maps and Detection of Change

Our method of seagrass change detection using Landsat images involved four steps (Figure 2): (1) atmospheric correction, necessary to convert the pixel values from digital number to surface reflectance; (2) selection of the best ML technique by comparing the accuracies of classification models for 2019 data; (3) application of the selected ML model (from step 2) for seagrass mapping to Landsat images from 1990, 2001, 2011, and 2014; (4) identifying the changes of area and spatial distribution. Due to the deficiency of field data in the past, a binary classification (seagrass and non-seagrass) was adopted to deliver the most consistent change detection.

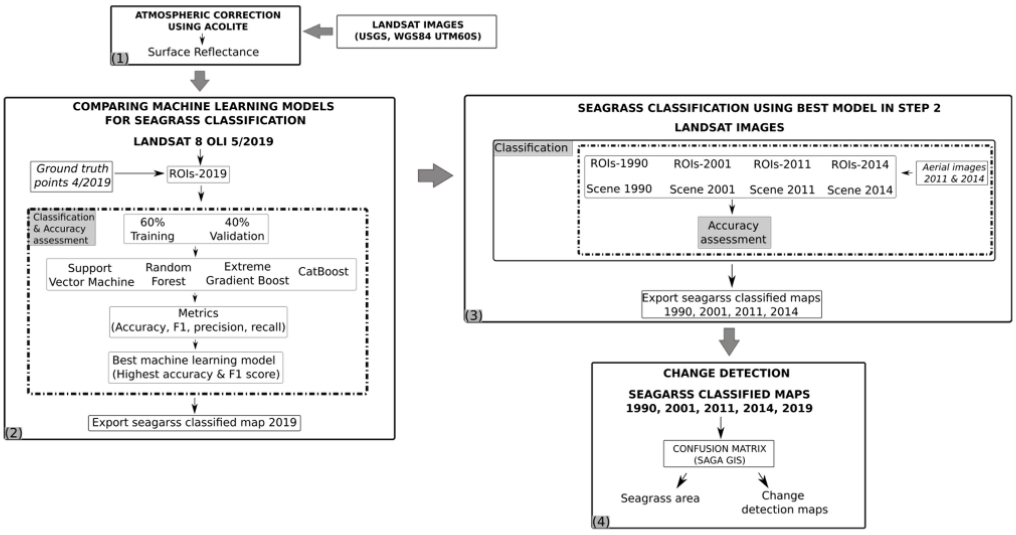


Figure 2. Flowchart of image processing and change detection using Landsat images in Tauranga Harbour.

2.5.1. Atmospheric Correction

An atmospheric correction for all Landsat scenes was conducted using ACOLITE, in the Python™ environment (Table A1, Appendix A) [52]. The original pixel values in physical digital number were converted to surface reflectance. Atmospheric corrected surface reflectance for pixels (limited by the study boundary, Figure 1) for the ρ_{Blue} (ρ_{w482}), the ρ_{Geen} (ρ_{w561}), and the ρ_{Red} (ρ_{w654}) bands were retained for all Landsat scenes in the years 1990, 2001, 2011, 2014, and 2019 for further processing steps. In the years 2014 and 2019, when Landsat 8 images were available, the coastal aerosol band (ρ_{w443}) was used, together with the ρ_{Blue}, ρ_{Geen}, and ρ_{Red} bands. The selected bands were used as independent variables in ML model prediction of the presence/absence of seagrass.

Due to inconsistency between the tidal status and the acquisition time of Landsat images, our study site was considered to contain both exposed and submerged areas. Therefore, the near infrared (ρ_{NIR}) band, which attenuates rapidly in water, was not used in the analysis. A water column correction was not employed for water pixels in Tauranga Harbour, since the water depth and water optical characteristics (i.e., attenuation coefficient of the solar radiance in the water column) were unavailable for the historic scenes (1990, 2001, 2011).

2.5.2. Application of Machine-Learning Algorithms

Hyper-Parameter Tuning for Selected Machine-Learning Models

Machine-learning models comprise several hyperparameters (i.e., the parameters that control the learning process during the implementation of ML models), which often need to be optimized (i.e., by the process of tuning) to find the best combination to achieve best classification performance. The hyper-parameters of the RF, the SVM, the XGB, and the CB models were tuned using a grid search with threefold cross-validation in the scikit-learn library [53]. The hyperparameters for each of the models were maintained during the training and the testing phases (Table A2, Appendix A).

Theoretical Background of the Machine-Learning Algorithms Used Random Forests

Random Forests (RF) [54] is perhaps the most popular machine-learning model for both classification and regression problems in remote sensing [55]. It is an ensemble

bagging method, which uses a bootstrap sampling approach to build the training and the testing data and a voting method to select the most accurate decision from a large group of input decision tress. The RF model is a nonparametric method that is insensitive to the data's distribution, reducing the overfitting. The RF technique supplies various hyperparameters for tuning; however, the large number of parameters in the model results in slow optimization.

Support Vector Machine

Support Vector Machine (SVM) [56,57] supports linear, poly-nominal, and radial basis function (RBF) kernels and can be adapted to various linear or non-linear data types. It has relatively few tunable hyperparameters but performance speed is still relatively slow when dealing with a large dataset. The SVM model uses a hyperplane to find the separation space among the classes with the most typical rules being: (i) better segregation of data; (ii) maximization of the distance between the closest data points and the hyperplane. Despite an accurate prediction and robustness to outliers, the SVM technique is not effective on overlaid classes or noisy datasets.

Extreme Gradient Boost

Extreme Gradient Boost (XGB) [17] is different from Gradient Boosting as it uses a more regularized model, which reduces over-fitting and results in a higher prediction accuracy. In the regularized gradient boosting mode, a selection of L_1 or L_2 regularization can be made to adapt the model to suit input data. Similar to other boosting models, the XGB technique supports various hyperparameters that are tuned using a grid search or genetic algorithm (GA) [58].

CatBoost

CatBoost (CB) was introduced in 2018 [59] for classification, regression, and ranking tasks. It can handle both category and numerical data types. Using ordered boosting on decision trees, a permutation-driven derivation from classic boosting, the CB yields a fast and reliable performance, even with a small dataset. The model itself produces robust predictive results with default hyperparameters, reducing the requirement of tuning, and its novel gradient boosting scheme results in less overfitting.

2.5.3. Comparison of ML Algorithms for Seagrass Mapping Using the Landsat Image Taken in 2019

Four ML models, SVM, RF, XGB, CB, were compared for seagrass mapping using the Landsat image from May 2019 and near-synchronous GTPs collected in April 2019 to identify the regions of interest (hereafter referred to as ROIs-2019) known to either seagrass or non-seagrass classes. The 1-month gap between the acquisition date of the Landsat image and the field survey date is acceptable due to the stable condition of the weather (i.e., no extreme weather phenomena) [60], and seagrass dynamics are slow in the study site [44,45]. A dataset of pixel reflectance values was extracted from ROIs-2019 and its corresponding Landsat image (dataset DS5, Table A3, Appendix A), split randomly into 60% for the training and 40% for the testing of selected ML models. The best model was selected as the model with highest accuracy and F_1 score.

2.5.4. Seagrass Mapping Using Landsat Images in 1990, 2001, 2011, and 2014

The best ML model identified using the 2019 data was applied for mapping of seagrass using Landsat images from 1990, 2001, 2011, and 2014 (see Table 1 for date acquisition and spatial resolution of satellite images). The hyper-parameters developed using the 2019 data were retained for subsequent analysis, while the year-specific model was developed using ROIs containing seagrass and non-seagrass classes from the relevant year. For the years 2011 and 2014, we created these ROIs using aerial imagery [61] (hereafter referred to as ROIs-2011 and ROIs-2014). For 1990 and 2001, we used Google Earth images cross-

referenced with the aerial images acquired between 1991 and 1992 (for creating ROIs-1990) and 2003 (for creating ROIs-2001). A dataset of pixel reflectance values was extracted from corresponding Landsat images (dataset DS1, DS2, DS3, and DS4, Table A3, Appendix A) for ROIs-1990, ROIs-2001, ROIs-2011, and ROIs-2014. Datasets were split randomly into 60% for training of the classification and 40% for the accuracy assessment for 1990, 2001, 2011, and 2014.

2.5.5. Change Detection

Change detection was conducted using the standard confusion matrix tool in the SAGA GIS [62]. The confusion matrix analyzed the changes of the pairs of classified maps (years 1990–2011 and 1990–2019), reporting in the map as seagrass loss (seagrass to non-seagrass), seagrass recovery (non-seagrass to seagrass), and unchanging seagrass.

2.6. Evaluation Criteria

We employed standard metrics for the evaluation of the classification skill: accuracy, Kappa coefficient (κ), Kendall's tau coefficient (τ), precision, recall, and F_1 (Equations (1)–(6)). These were applied independently to the five datasets listed in Table A3, to yield the skill of the initial model based on GTPs from DS5 (2019), and to check its performance when applied to the historic Landsat data in DS1 (1990), DS2 (2001), DS3 (2011), and DS4 (2014). Kendall's tau coefficient was calculated using the SciPy library [63].

$$accuracy\left(y, y_{pred}\right) = \frac{1}{\mathrm{n}_{samples}} \sum_{i=0}^{\mathrm{n}_{samples}-1} 1\left(y_{predi} = y_i\right) \tag{1}$$

in which:

y_{pred}: predicted value
y: corresponding true value

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{2}$$

in which:

p_o is the observed agreement
p_e is the expected agreement

$$\tau = \frac{P - Q}{\sqrt{(P + Q + T) \times (P + Q + U)}} \tag{3}$$

in which:

P: the number of concordant pair
Q: the number of discordant pair
U: the number of ties in predicted value
T: the number of ties in true value

$$Precision = \frac{tp}{tp + fp} \tag{4}$$

$$Recall = \frac{tp}{tp + fn} \tag{5}$$

$$F_1 = \frac{2 \times precision \times recall}{precision + recall} \tag{6}$$

in which:

tp: true positive
fp: false positive

fn: false negative

In addition, the nonparametric McNemar test was used to assess the statistical significance of the differences of the overall accuracy of the selected models in this research. The test was executed in a Python™ environment using the mlxtend library [64]. The chi-square value (χ^2) was calculated from Equation (7) with Edward's continuity correction.

$$\chi^2 = \frac{(|fn - fp| - 1)^2}{(fn + fp)} \quad (7)$$

in which:

fn: false negative
fp: false positive

3. Results

3.1. Performance of the RF, SVM, XGB, and CB Models Using Landsat Image and GTPs for 2019 Data

Of the four machine-learning models applied to the 2019 data, the CB model outperformed all others, with the F_1 score, κ and τ coefficients reaching 0.95 (Table 3), and 0.92 (Table A4, Appendix A), respectively. The difference between models was statistically significant (McNemar's test, Table 4) with the exception of the XGB and RF models. The CB model required a longer computation time (3.71 s) than the RF model (0.33 s), the XGB model (0.15 s), and the SVM model (0.04 s). The RF and XGB techniques showed an equivalent performance (Table 3) with F_1 score of 0.93, while the SVM model underperformed the other models with a F_1 score of 0.91.

Table 3. Model performance for seagrass detection in Tauranga Harbour for the 2019 dataset.

Model	Accuracy	Precision	Recall	F_1	Training Time (s)	Testing Time (s)
RF	0.96	0.92	0.95	0.93	0.33	0.02
CB	**0.97**	**0.94**	**0.96**	**0.95**	3.71	0.006
XGB	0.96	0.93	0.94	0.93	0.15	0.004
SVM	0.94	0.89	0.92	0.91	0.04	0.02

Bold values indicate the best performance of the model.

Table 4. Model performance comparison using McNemar's test.

	χ^2	*p*-Value
CB–RF	5.88	0.01
CB–SVM	19.11	0.00
CB–XGB	4.50	0.03
XGB–RF	0.00	1.00
XGB–SVM	8.20	0.00
RF–SVM	9.25	0.00

p-value < 0.05 indicates a significant difference between two models.

All models tested were able to classify seagrass from other bottom types in the harbor with a precision exceeding 0.89, but the highest precision was again from the CB model. Despite a similar F_1 score, the XGB model gained a higher precision than the RF technique.

3.2. Seagrass Change Detection from 1990–2019

The CB technique was then used to make classification maps for the years 1990, 2001, 2011, and 2014 (Figure 3). Our results indicated a performance across all metrics that was equivalent to that in the 2019 case, with accuracy and F_1 scores over 95% for the binary classification of seagrass and nonseagrass (Tables 5 and A4, Appendix A).

Table 5. Accuracy assessment of the classified map using Landsat images in 1990, 2001, 2011, and 2014.

Date Acquisition	Accuracy	Precision	Recall	F_1
4 April 1990	0.97	0.98	0.98	0.98
10 March 2001	0.96	0.95	0.96	0.96
17 February 2011	0.97	0.98	0.96	0.97
6 March 2014	0.96	0.96	0.96	0.96

The time series shows that the seagrass meadow area decreased from 2237 ha in 1990 to 1184 ha in 2019, though not monotonically (Figures 3 and 4). A downward trend from 1990 (2237 ha) to 2001 (2035 ha), was followed by a recovery in 2011 (to 2380 ha), followed by a second decline to 1184 ha in 2019 (Figure 4a). Different trends, though all with an overall decline to 2019, were discovered in the northern (Figure 4b), the central (Figure 4c), and the southern (Figure 4d) harbor. Seagrass attained the largest area in the central harbor, where it reached the peak of 1985 ha in 2011; however, it declined to 776 ha in 2019. In the northern harbor, seagrass was very abundant in 1990 with 549 ha, but strongly decreased to only 92 ha in 2001. This number increased to 242 ha in 2014 before suffered a second decline to 148 ha in 2019. Seagrass loss was also recorded in the southern harbor, at a slower rate of degradation, dropping from 576 ha in 1990 to 222 ha in 2011, and around 260 ha in 2019. Across the entire harbor, the recovery in 2011 was due to a large increase of seagrass areas from 2001 in the northern and the central harbors.

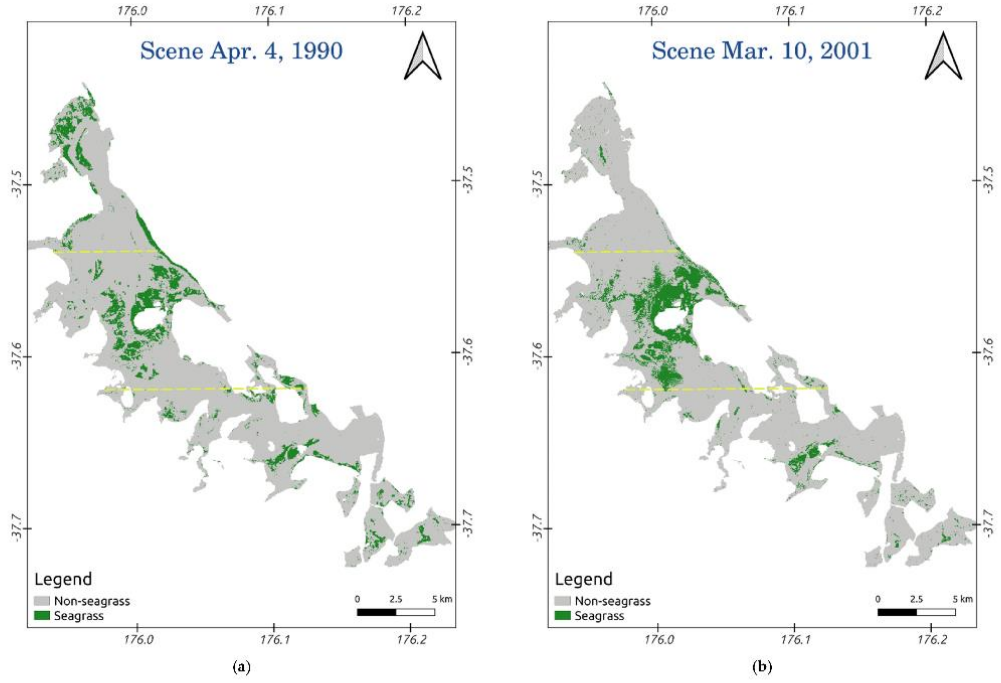


Figure 3. *Cont.*

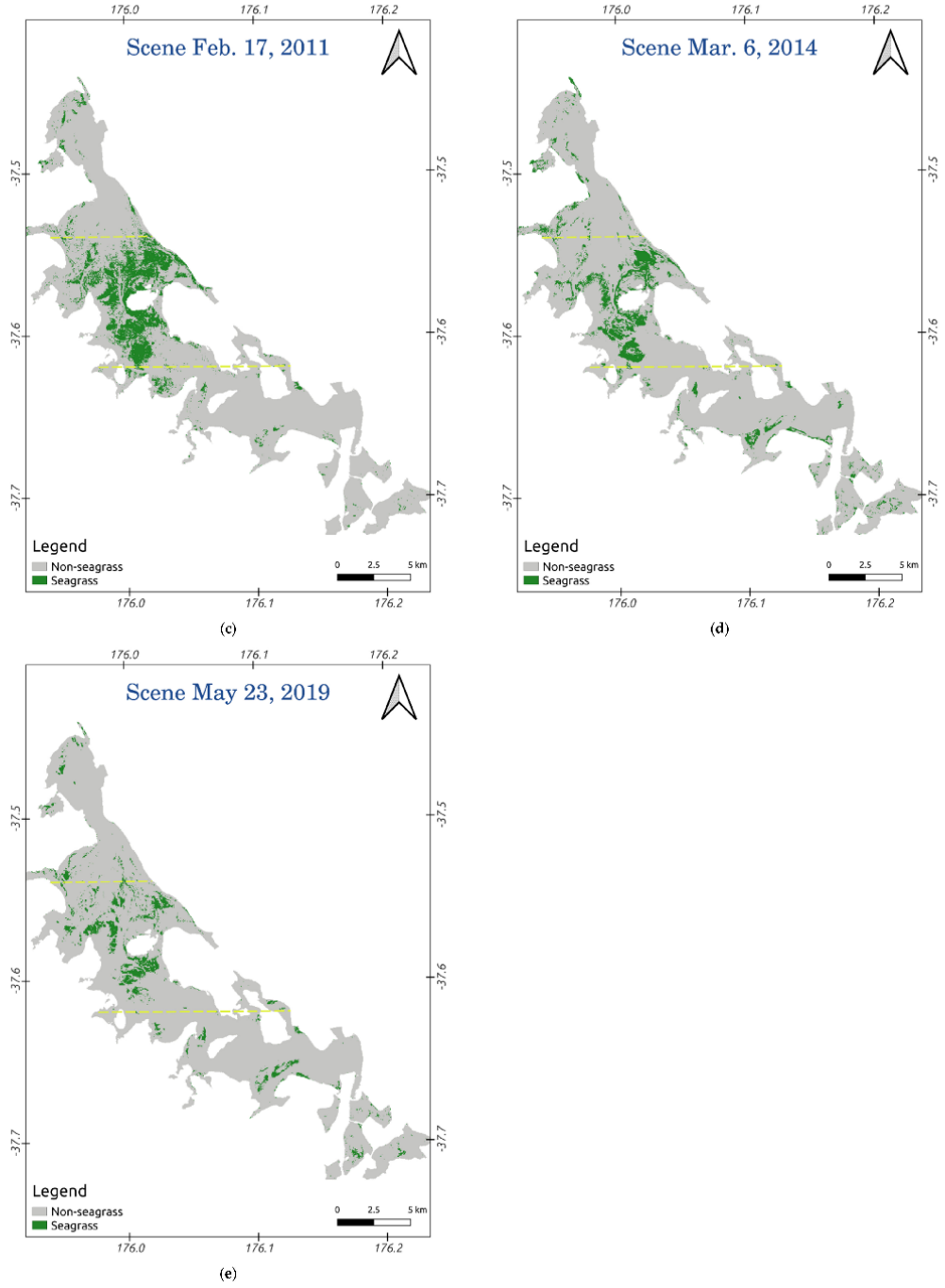


Figure 3. Seagrass distribution in the years 1990, 2001, 2011, 2014, and 2019 (**a**–**e**) using the CB model (yellow lines indicate the boundaries of the northern, central, and southern harbor).

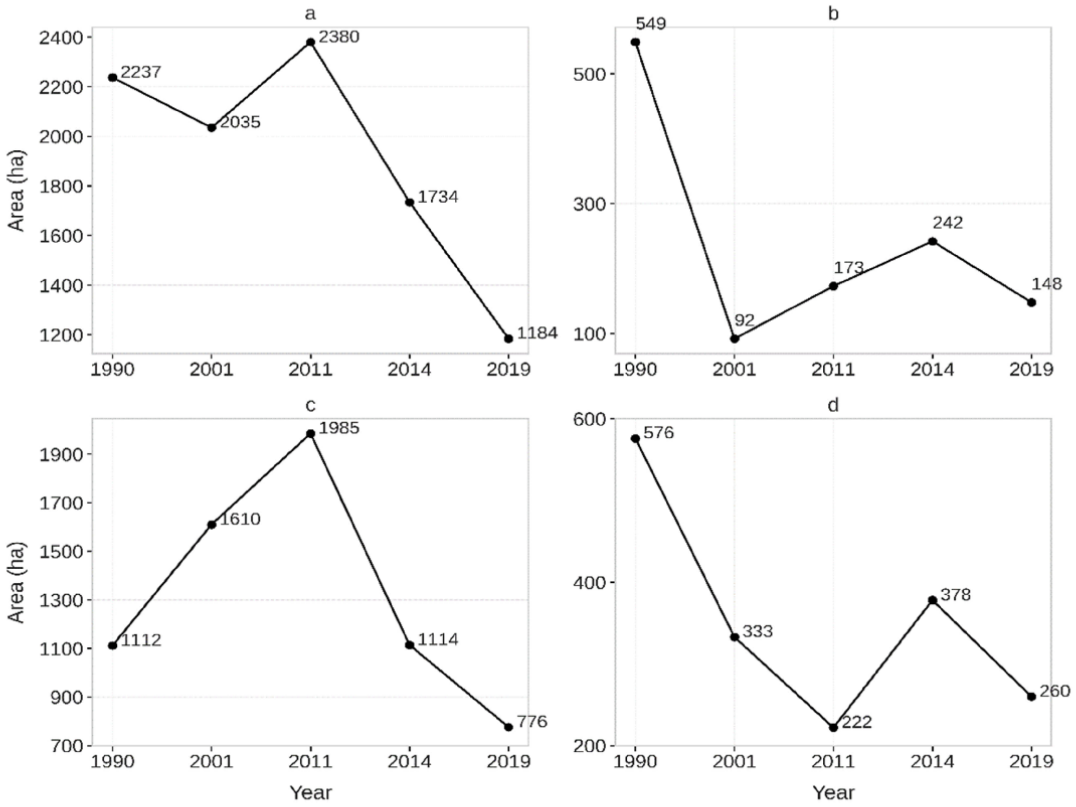


Figure 4. Seagrass area in Tauranga Harbour from 1990–2019 derived from Landsat imagery with the variation in: (**a**) entire harbor; (**b**) northern harbor; (**c**) central harbor; (**d**) southern harbor.

The distribution of seagrass has also changed over time. In 1990, the meadows were similarly abundant in the northern, central, and southern harbors. Declines to 2001 mostly reflected losses from the northern and southern meadows, while the central meadows remained and were responsible for most of the expansion between 2001 and 2011 (Figures 3 and 5). After 2011, there was no detectable recovery of the northern or the southern meadows, and the renewed overall decline was due to degradation of the central meadows, declining in area and becoming patchier by 2019 (Figures 3 and 5).

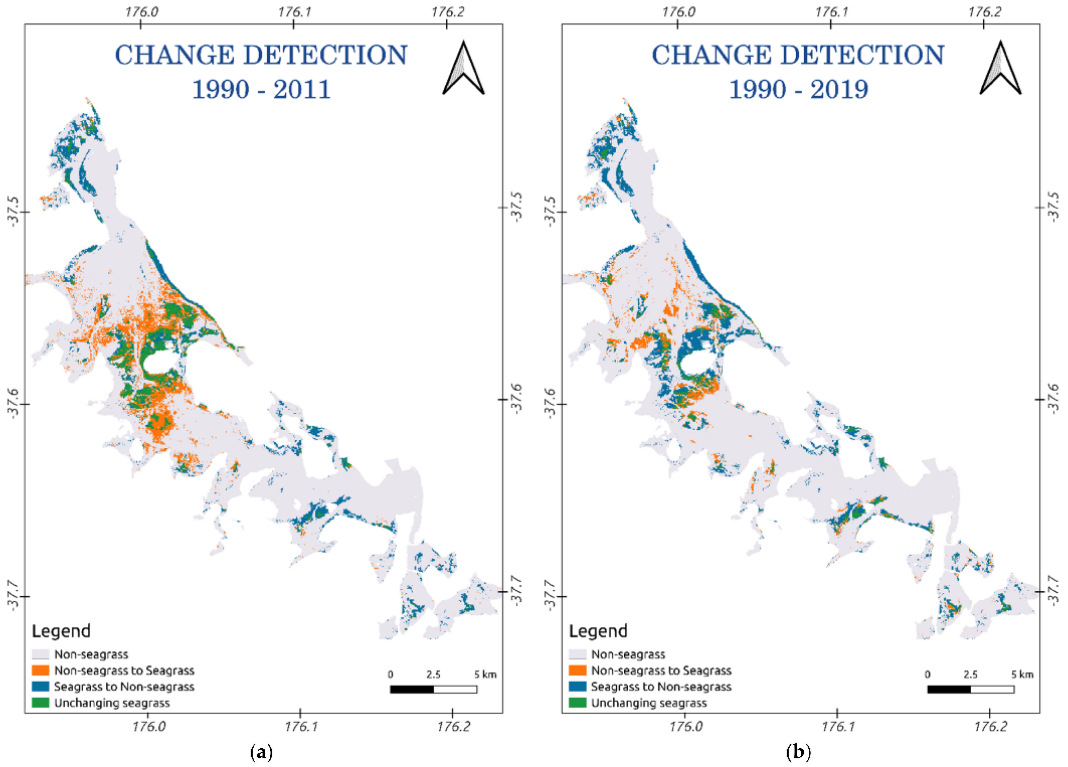


Figure 5. Seagrass change detection between 1990–2011 (**a**) and 1990–2019 (**b**).

4. Discussion

In this investigation, we have demonstrated the use of machine-learning approaches successfully to classify seagrass in Landsat images of Tauranga Harbour, and to use this classification to detect changes in seagrass cover over a period of 29 years. Due to the relative paucity of field validation data from most of the time series of this analysis, we only tested a binary classification (seagrass and non-seagrass classes), but the four machine-learning models, RF, XGB, SVM, and CB, were all capable of detecting seagrass from other bottom types with high precision and recall scores. Previously, the RF and the SVM models have been tested for seagrass classification [65] and the RF for seagrass change detection in the Mediterranean [13]. These previous attempts have produced accuracies from 76–98% for *Posidonia oceania* [13,65] and 32–62% for *Cymodocea nodosa* using higher resolution RapidEye imagery [13], both lower than were achieved in this study using the CB technique. *P. oceania* and *C. nodosa* are structurally similar to *Z. muelleri* and would seem likely to offer a similar target. This suggests that the use of the state-of-the-art ML models with optimized hyper-parameters is an important factor contributing to the high-precision classification of seagrass presence/absence. Both the XGB and the CB techniques have been proven as potential candidates for a range of classification [58,66–68], and regression [69–71] problems but have not previously been applied to seagrasses, or to any other semi-submerged targets, so it is not clear if this is a general performance advantage in this type of application.

Other advantages over previous studies may, however, exist in Tauranga Harbour. Specifically, *Z. muelleri* occurs as monospecific meadows, without a substantial presence of

macroalgae, which can degrade classifications [20,37], and where the reflectance value of seagrass is considerably different from the other common bottom types (sand, muddy sand, deep water). In addition, we were able to use cloud-free Landsat scenes, with atmospheric correction using ACOLITE, which has been designed for the aquatic application of Landsat imagery, which likely reduced the uncertainty of atmospheric impact and derived a higher quality of corrected surface reflectance [72].

In this study, the two boosting techniques (XGB, CB) and the one bagging (RF) outperformed the more traditional SVM methods. The SVM model does not work well with noisy data, where unclear margins exist between classes [73]. Such fuzzy margins were observed at the study site at the overlap between seagrass and non-seagrass (sand, muddy sand) classes, where the distinction between present and absent was gradual. This likely resulted in the relatively poor performance of the SVM model. The boosting techniques XGB (0.93) and CB (0.94) show slightly higher precision than the bagging RF (0.92), which might have resulted from the advancement in decision-tree growth of the boosting techniques. Unlike the RF model, which builds the independent decision tree from the bootstrapped samples, the boosting XGB and CB models sequentially grow new trees using the residual information of previous trees, which allows the new learner to solve the errors of the previous tree by minimizing the residual of the next model fitting. For a final prediction of a classification task, the bagging RF takes a majority vote from all decision trees while a weighted majority vote is adapted to the boosting techniques, such as XGB and CB, and potentially results in a higher precision of a class prediction.

Given the classification skill metrics, the CB is the best candidate for the mapping and change detection of seagrass in the study site. The CB is also amongst the latest emerging algorithms developed in the computer vision and pattern recognition fields (released in 2018); is easy to tune with fewer hyperparameters than the RF and the XGB techniques; and is using symmetric trees, which potentially results in faster optimization and prediction [59]. The CB model differs from the boosting algorithm family by using ordered boosting on a random permutation of given dataset, which prevent the prediction shift and alleviate the overfitting in model prediction. The outperformance of the CB over other ML models has been reported for mangrove total carbon estimation [74], various testing datasets [59], and forest aboveground biomass [75], which confirm the reliability and the capability of the CB implementation for seagrass mapping in our study. Our accurate long-term (29-year) change detection of seagrass meadows using the CB machine-learning model in Tauranga Harbour is a significant advance in the classification and monitoring of seagrass ecosystem using multispectral, remotely sensed data.

Our analysis has confirmed a general declining trend of seagrass cover in Tauranga Harbour reported previously [37] using aerial photography. In absolute terms, Park (2011) reported 2744 ha in March 2011, close to our estimate of 2380 ha at that time. Also, like Park, our analysis was able to resolve areas within the estuary where the greatest loss has occurred between assessments. We specifically noted that the seagrass loss was initially focused in the northern and southern parts of the harbor. High flux of sediment was recorded into the northern part, due to agricultural intensification, and the southern part, due to urban development, particularly after 2011 [47] and may explain the long-term decline of seagrass in those areas. The potential impact of agricultural and urban developments in the northern and the southern parts is supported by the observation that recovery was only observed in the central part of the harbor (Figure 3, year 2011, and Figure 5). Another potential factor contributing to long-term loss of seagrass is the grazing of black swans, which has previously been linked to variations in seagrass cover in the southern harbor [37,76]. Further analysis is required to develop a detailed explanation on the dynamics of seagrass meadows in Tauranga Harbour.

Here, we advocate the use of novel and advanced ML models, in combination with multitemporal Landsat images to obtain a long-term, historic series of observations on seagrass dynamics that will continue to be supported into the future through ongoing developments of the Landsat series. The proposed method potentially provides a low-cost,

high-precision classification tool that can be extended to other estuaries with similar target conditions. While aerial photography and very high spatial resolution (VHR) satellite images have higher spatial resolution than Landsat images, they come at a high cost, and spatial coverage can be limited. Currently, Landsat is the most suitable satellite image resource for any long-term change detection due to its long time in service. A 30 m spatial resolution was found suitable to support a binary classification of seagrass in this study, and accuracy was unaffected by the small changes in spectral information that have accompanied the incremental changes in Landsat optical sensors. The most recent generation Landsat 8, with an improvement in radiometric resolution up to 16 bit (in the level 1 product), compared to the 8 bit in previous generations, and the addition of a coastal aerosol band, has good potential for accurate detection of the dynamics of seagrass. An improvement of spectral and radiometric resolution in Landsat 9 (scheduled for launching in 2021) is expected to provide continuity into the future monitoring of seagrass [77]. For a short-term observation of seagrass change, our proposed methods for seagrass classification are also potentially applicable to a wide range of VHR images (Quickbird, Ikonos, Unmanned Aerial Vehicle (UAV)) with consideration of the trade-off among the spatial coverage of the study site, the spatial resolution of the image, and the available budget.

The open-source approach is another significant advantage of our proposed methodology. The Python environment provides an excellent option for the end users to apply the novel machine-learning algorithms and remote-sensing data processing platform to support accurate mapping and estimation of the blue carbon budget of seagrass ecosystem [78]. Most commercial software only provides a limited number of processing and classification algorithms, with few, older ML options (e.g., SVM) and has a high license cost. Our proposed methods are more flexible, free of charge, and offer a high efficiency for mapping the dynamics of seagrass meadows in the complex coastal marine environment.

Despite a successful application of the CB model for seagrass classification and change detection, this research still comes with limitations. Since we used a supervised classification technique, both classification and validation require an independent assessment of seagrass cover in at least part of the remote image, to provide the ROIs that allow the training and validation steps. In addition, the seasonal growth of seagrass in temperate waters, and its intertidal habit, raise the uncertainty of change detection between various time points unless imagery is available at the same time, and under similar tidal conditions. The offset between Landsat, the time of image acquisition, and tidal regime (Table 1) is unavoidable in the study site; however, we consider that it is unlikely to significantly impact on classification accuracy. In Tauranga Harbour, seagrass meadows are distributed in the intertidal regions at a water depth ranging from 0 m (exposed) to a maximum of 1.5 m (at high tide) [37,51]. The ρ_{Blue}, ρ_{Green}, and ρ_{Red} bands have nominal maximum penetration depths of 15, 10, 5 m respectively [34], and while moderate, but variable, coastal turbidity in the harbor will increase attenuation rate, the maximum immersion depth of 1.5 m suggests that the spectral bands reflectance signatures are highly likely to have been impacted by seagrass. Average vertical attenuation rate of the downwelling radiation within the 400–700 nm band in Tauranga Harbour is 0.40 m^{-1} (range 0.16–0.98 m^{-1}) [79] and these authors found that 65% of incident radiation reached the estuary floor at 1.2 m depth. Again, this suggests that water clarity is sufficient to ensure that, even at maximum water depth, seagrass will contribute to the reflectance spectrum detected by the satellite. As with all satellite-based remote sensing, a cloud-free view is required, which constrains use of this technology.

To compensate for the limitation, we attempted to select all Landsat images acquired in the growing season of seagrass in Tauranga Harbour (austral summer and autumn) and at low tide, but this further constrains the availability of verifiable Landsat imagery for seagrass cover estimation. Further research focusing on expanding the novel approach used in the current study for long-term change detection of seagrass meadows is underway.

5. Conclusions

In this research, we used the novel machine-learning model CatBoost (CB) and other well-known ML models (RF, SVM, and XGB) for seagrass cover classification (present/absent), using Landsat satellite imagery, in Tauranga Harbour, New Zealand. Our results showed a high level of accuracy for all approaches, but the CB model outperformed the other selected models, with precision, recall, and F_1 scores of 0.94, 0.96, and 0.95 respectively.

We then applied the CB technique to multispectral Landsat data for the detection of change in seagrass cover over a 29-year period between 1990 and 2019 in Tauranga Harbour. The change detection analysis determined an overarching declining trend of seagrass cover in Tauranga Harbour with approximately 50% loss over the 29 years period (from 2237 ha in 1990 to 1184 ha in 2019); these results concurred with a study using aerial imaging. Seagrass was lost in the far northern and southern areas of the harbor during the first part of this time, then more gradually from the central region. This analysis of change using Landsat images combined with the CB model demonstrates the value of historic satellite imagery and machine-learning for accurate documentation of the change over time in this difficult-to-quantify coastal vegetation.

Author Contributions: Conceptualization, Nam-Thang Ha and Ian Hawes; methodology, Nam-Thang Ha; software, Nam-Thang Ha; validation, Nam-Thang Ha and Tien-Dat Pham; resources, Nam-Thang Ha, Ian Hawes, Merilyn Manley-Harris; writing—original draft preparation, Nam-Thang Ha; writing—review and editing, Nam-Thang Ha, Tien-Dat Pham, Merilyn Manley-Harris, and Ian Hawes; supervision, Ian Hawes. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data/material are available on request from the authors: The data and material that support the findings of this study are available from the corresponding author, Nam Thang Ha, upon reasonable request.

Acknowledgments: We are grateful for the support from Christine E. C. Gunfield and Matthew J. Finnigan from the Marine Field Station, Tauranga, New Zealand for the field survey.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Selected parameters for atmospheric correction using ACOLITE.

Parameter	Value
Ancillary data	
Gas transmittance	True
Ozone concentration (cm^{-1})	0.3
Water vapor concentration (g cm^{-2})	1.5
Pressure	Normal pressure
Masking	
Negative reflectance masking	True
Cirrus masking	True
Other parameters	
Sky correction	True
Dark spectrum fitting	Fixed
Sun glint correction	False
Output parameter	
Surface reflectance for water pixel (ρ_w)	ρ_{w443} ρ_{w482} ρ_{w561} ρ_{w654}

Table A2. The tuned hyperparameters of the RF, the SVM, the XGB, and the CB models.

Random Forest		Extreme Gradient Boost	
Bootstrap	**True**	**Booster**	**GbTree**
Max. depth	8	Gamma	1
Max. features	Auto	Learning rate	0.2
Min. sample leaf	1	Max. depth	5
Min. sample split	3	Min. child weight	3
Number of trees	100	Number of trees	100
Support Vector Machine		CatBoost	
Kernel	RBF	Depth	7
C	100	Iteration (Number of trees)	200
Gamma	1000	Learning rate	0.2
		L2 leaf reg	1

Table A3. Number of pixels for the training and the testing sets at various acquisition dates.

Dataset	Landsat Acquisition Date	Number of Pixels	
		60% for Training	**40% For Testing**
DS1	4 April 1990	2171	1448
DS2	10 March 2001	3000	2001
DS3	17 February 2011	2618	1746
DS4	6 March 2014	2544	1696
DS5	23 May 2019	1830	1221

Table A4. Kappa and Kendall's tau coefficients of the classification.

Model	κ	τ	p-Value of τ
Data DS5, date 23 May 2019			
RF	0.90	0.90	0.00
SVC	0.87	0.87	0.00
CB	0.92	0.92	0.00
XGB	0.90	0.90	0.00
Data DS1, date 4 April 1990			
CB	0.95	0.95	0.00
Data DS2, date 10 March 2001			
CB	0.92	0.92	0.00
Data DS2, date 17 February 2011			
CB	0.94	0.94	0.00
Data DS4, date 6 March 2014			
CB	0.93	0.93	0.00

Table A5. List of acronyms and abbreviations.

Acronym/Abbreviation	Meaning	Explanation
ACOLITE	Atmospheric correction for operational land imager (OLI) 'lite' toolbox	A Python language-based application for atmospheric correction of satellite imagery
	Accuracy	An agreement degree between the classified values and the ground-truth values in a classification task
CB	CatBoost	A machine-learning algorithm
XGB	Extreme Gradient Boost	A machine-learning algorithm
F_1	F_1	A harmonic measurement of precision and recall scores in the prediction of a machine-learning model

Table A5. *Cont.*

Acronym/Abbreviation	Meaning	Explanation
GPS	Global Positioning System	A satellite-based system providing positioning services
GTPs	Ground-Truth Points	GTPs are the boundary points of any given classes in the study site, defined by GPS
K	Kappa coefficient	A statistical index measuring the accuracy (agreement between predictions and ground-truthed values) of the classification. A higher Kappa coefficient denotes a more accurate classification
τ	Kendall's tau coefficient	A nonparametric measurement to evaluate the classification's accuracy. A higher Kendall's tau coefficient denotes a more accurate classification
ML	Machine learning	An artificial intelligence (AI) approach that builds an application/algorithm for a specific output by learning from data
NIR	Near infrared	The near infrared region in the electromagnetic spectrum
	Precision	A score to measure the success of the prediction of a machine-learning model. A higher precision denotes a more accurate prediction
RF	Random Forest	A machine-learning algorithm
	Recall	A score to measure the success of the prediction of a machine-learning model. A higher recall denotes a more accurate prediction
RBF	Radial basis function	A function used in the Support Vector Machine model, together with linear and polynomial functions
ROI	Region of interest	A bounded region used in image classification where the pixels contain a given class
SVM	Support Vector Machine	A machine-learning algorithm
UAV	Unmanned aerial vehicle	An aircraft without a human pilot
GLOVIS	USGS Global Visualization Viewer	A web-based system for satellite image visualization and downloading
UTM	Universal Transverse Mercator	A map projection
VHR	Very high spatial resolution	Indicating satellite images that have spatial resolution from centimeters to a few meters
WGS	World Geodetic System	A standard coordinate system used in cartography.

References

1. Nordlund, L.M.; Jackson, E.L.; Nakaoka, M.; Samper-Villarreal, J.; Beca-Carretero, P.; Creed, J.C. Seagrass Ecosystem Services—What's Next? *Mar. Pollut. Bull.* **2018**, *134*, 145–151. [CrossRef]
2. Nordlund, L.M.; Koch, E.W.; Barbier, E.B.; Creed, J.C. Seagrass Ecosystem Services and Their Variability across Genera and Geographical Regions. *PLoS ONE* **2016**, *11*, e0163091. [CrossRef] [PubMed]
3. Fourqurean, J.W.; Duarte, C.M.; Kennedy, H.; Marbà, N.; Holmer, M.; Mateo, M.A.; Apostolaki, E.T.; Kendrick, G.A.; Krause-Jensen, D.; McGlathery, K.J.; et al. Seagrass Ecosystems as a Globally Significant Carbon Stock. *Nat. Geosci.* **2012**, *5*, 505–509. [CrossRef]
4. Gullström, M.; Lyimo, L.D.; Dahl, M.; Samuelsson, G.S.; Eggertsen, M.; Anderberg, E.; Rasmusson, L.M.; Linderholm, H.W.; Knudby, A.; Bandeira, S.; et al. Blue Carbon Storage in Tropical Seagrass Meadows Relates to Carbonate Stock Dynamics, Plant–Sediment Processes, and Landscape Context: Insights from the Western Indian Ocean. *Ecosystems* **2018**, *21*, 551–566. [CrossRef]
5. Orth, R.J.; Carruthers, T.J.B.; Dennison, W.C.; Duarte, C.M.; Fourqurean, J.W.; Heck, K.L.; Hughes, A.R.; Kendrick, G.A.; Kenworthy, W.J.; Olyarnik, S.; et al. A Global Crisis for Seagrass Ecosystems. *BioScience* **2006**, *56*, 987. [CrossRef]

6. Waycott, M.; Duarte, C.M.; Carruthers, T.J.B.; Orth, R.J.; Dennison, W.C.; Olyarnik, S.; Calladine, A.; Fourqurean, J.W.; Heck, K.L.; Hughes, A.R.; et al. Accelerating Loss of Seagrasses across the Globe Threatens Coastal Ecosystems. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 12377–12381. [CrossRef] [PubMed]
7. Marbà, N.; Arias-Ortiz, A.; Masqué, P.; Kendrick, G.A.; Mazarrasa, I.; Bastyan, G.R.; Garcia-Orellana, J.; Duarte, C.M. Impact of Seagrass Loss and Subsequent Revegetation on Carbon Sequestration and Stocks. *J. Ecol.* **2015**, *103*, 296–302. [CrossRef]
8. Cullen-Unsworth, L.; Unsworth, R. Seagrass Meadows, Ecosystem Services, and Sustainability. *Environ. Sci. Policy Sustain. Dev.* **2013**, *55*, 14–28. [CrossRef]
9. Hejnowicz, A.P.; Kennedy, H.; Rudd, M.A.; Huxham, M.R. Harnessing the Climate Mitigation, Conservation and Poverty Alleviation Potential of Seagrasses: Prospects for Developing Blue Carbon Initiatives and Payment for Ecosystem Service Programmes. *Front. Mar. Sci.* **2015**, *2*. [CrossRef]
10. Bujang, J.S.; Zakaria, M.H.; Short, F.T. Seagrass in Malaysia: Issues and Challenges Ahead. In *The Wetland Book*; Finlayson, C.M., Milton, G.R., Prentice, R.C., Davidson, N.C., Eds.; Springer: Dordrecht, The Netherlands, 2018; pp. 1875–1883, ISBN 978-94-007-4000-6.
11. Unsworth, R.K.F.; McKenzie, L.J.; Collier, C.J.; Cullen-Unsworth, L.C.; Duarte, C.M.; Eklöf, J.S.; Jarvis, J.C.; Jones, B.L.; Nordlund, L.M. Global Challenges for Seagrass Conservation. *Ambio* **2018**. [CrossRef]
12. Pham, D.; Yokoya, N.; Bui, D.; Yoshino, K.; Friess, D. Remote Sensing Approaches for Monitoring Mangrove Species, Structure, and Biomass: Opportunities and Challenges. *Remote Sens.* **2019**, *11*, 230. [CrossRef]
13. Traganos, D.; Reinartz, P. Interannual Change Detection of Mediterranean Seagrasses Using RapidEye Image Time Series. *Front. Plant Sci.* **2018**, *9*. [CrossRef] [PubMed]
14. Paulose, N.E.; Dilipan, E.; Thangaradjou, T. Integrating Indian Remote Sensing Multi-Spectral Satellite and Field Data to Estimate Seagrass Cover Change in the Andaman and Nicobar Islands, India. *Ocean Sci. J.* **2013**, *48*, 173–181. [CrossRef]
15. Roelfsema, C.; Lyons, M.; Dunbabin, M.; Kovacs, E.M.; Phinn, S. Integrating Field Survey Data with Satellite Image Data to Improve Shallow Water Seagrass Maps: The Role of AUV and Snorkeller Surveys? *Remote Sens. Lett.* **2015**, *6*, 135–144. [CrossRef]
16. Hossain, M.S.; Bujang, J.S.; Zakaria, M.H.; Hashim, M. The Application of Remote Sensing to Seagrass Ecosystems: An Overview and Future Research Prospects. *Int. J. Remote Sens.* **2015**, *36*, 61–114. [CrossRef]
17. Chen, C.-F.; Lau, V.-K.; Chang, N.-B.; Son, N.-T.; Tong, P.-H.-S.; Chiang, S.-H. Multi-Temporal Change Detection of Seagrass Beds Using Integrated Landsat TM/ETM+/OLI Imageries in Cam Ranh Bay, Vietnam. *Ecol. Inform.* **2016**, *35*, 43–54. [CrossRef]
18. Amone-Mabuto, M.; Bandeira, S.; da Silva, A. Long-Term Changes in Seagrass Coverage and Potential Links to Climate-Related Factors: The Case of Inhambane Bay, Southern Mozambique. *West. Indian Ocean J. Mar. Sci.* **2017**, *16*, 13–25.
19. Phinn, S.R.; Kovacs, E.M.; Roelfsema, C.M.; Canto, R.F.; Collier, C.J.; McKenzie, L.J. Assessing the Potential for Satellite Image Monitoring of Seagrass Thermal Dynamics: For Inter- and Shallow Sub-Tidal Seagrasses in the Inshore Great Barrier Reef World Heritage Area, Australia. *Int. J. Digit. Earth* **2018**, *11*, 803–824. [CrossRef]
20. Ha, N.T.; Manley-Harris, M.; Pham, T.D.; Hawes, I. A Comparative Assessment of Ensemble-Based Machine Learning and Maximum Likelihood Methods for Mapping Seagrass Using Sentinel-2 Imagery in Tauranga Harbor, New Zealand. *Remote Sens.* **2020**, *12*, 355. [CrossRef]
21. McCarthy, M.J.; Halls, J.N. Habitat Mapping and Change Assessment of Coastal Environments: An Examination of WorldView-2, QuickBird, and IKONOS Satellite Imagery and Airborne LiDAR for Mapping Barrier Island Habitats. *ISPRS Int. J. Geo-Inf.* **2014**, *3*, 297–325. [CrossRef]
22. Sousa, L.P.; Sousa, A.I.; Alves, F.L.; Lillebø, A.I. Ecosystem Services Provided by a Complex Coastal Region: Challenges of Classification and Mapping. *Sci. Rep.* **2016**, *6*, 22782. [CrossRef]
23. Camps-Valls, G. Machine Learning in Remote Sensing Data Processing. In Proceedings of the 2009 IEEE International Workshop on Machine Learning for Signal Processing, Grenoble, France, 1–4 September 2009; pp. 1–6.
24. Lary, D.J.; Alavi, A.H.; Gandomi, A.H.; Walker, A.L. Machine Learning in Geosciences and Remote Sensing. *Geosci. Front.* **2016**, *7*, 3–10. [CrossRef]
25. Maxwell, A.E.; Warner, T.A.; Fang, F. Implementation of Machine-Learning Classification in Remote Sensing: An Applied Review. *Int. J. Remote Sens.* **2018**, *39*, 2784–2817. [CrossRef]
26. Ahmad, H. Machine learning applications in oceanography. *Aquat. Res.* **2019**, 161–169. [CrossRef]
27. Machova, K.; Puszta, M.; Barcak, F.; Bednar, P. A Comparison of the Bagging and the Boosting Methods Using the Decision Trees Classifiers. *Comput. Sci. Inf. Syst.* **2006**, *3*, 57–72. [CrossRef]
28. Bühlmann, P. Bagging, Boosting and Ensemble Methods. In *Handbook of Computational Statistics*; Gentle, J.E., Härdle, W.K., Mori, Y., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 985–1022. ISBN 978-3-642-21550-6.
29. Huettmann, F. Boosting, Bagging and Ensembles in the Real World: An Overview, some Explanations and a Practical Synthesis for Holistic Global Wildlife Conservation Applications Based on Machine Learning with Decision Trees. In *Machine Learning for Ecology and Sustainable Natural Resource Management*; Springer International Publishing: Berlin/Heidelberg, Germany, 2018; pp. 63–83.
30. Yaman, E.; Subasi, A. Comparison of Bagging and Boosting Ensemble Machine Learning Methods for Automated EMG Signal Classification. Available online: https://www.hindawi.com/journals/bmri/2019/9152506/ (accessed on 2 December 2020).
31. Northrop, A. *IDEAS—LANDSAT Products Description Document*; USGS: Reston, VA, USA, 2015; p. 68.

32. Zhu, Z. Change Detection Using Landsat Time Series: A Review of Frequencies, Preprocessing, Algorithms, and Applications. *ISPRS J. Photogramm. Remote Sens.* **2017**, *130*, 370–384. [CrossRef]
33. Ihlen, V. *Landsat 8 (L8) Data Users Handbook*; USGS: Reston, VA, USA, 2019.
34. Green, E.P.; Mumby, P.J.; Edwards, A.J.; Clark, C.D. A Review of Remote Sensing for the Assessment and Management of Tropical Coastal Resources. *Coast. Manag.* **1996**, *24*, 1–40. [CrossRef]
35. Ha, N.T.; Yoshino, K.; Hoang Son, T.P. *Seagrass Mapping Using ALOS AVNIR-2 Data in Lap An Lagoon, Thua Thien Hue, Viet Nam*; Frouin, R.J., Ebuchi, N., Pan, D., Saino, T., Eds.; SPIE: Kyoto, Japan, 2012; p. 85250S.
36. Garcia, R.; Hedley, J.; Tin, H.; Fearns, P. A Method to Analyze the Potential of Optical Remote Sensing for Benthic Habitat Mapping. *Remote Sens.* **2015**, *7*, 13157–13189. [CrossRef]
37. Park, S.G. *Extent of Seagrass in the Bay of Plenty in 2011*; Environmental Publication; Bay of Plenty Reginal Council: Whakatane, New Zealand, 2011; p. 52.
38. Reeve, G.; Stephens, S.; Wadhwa, A. *Tauranga Harbour Inundation Modelling*; NIWA: Tauranga, New Zealand, 2018; p. 107.
39. Tay, H.; Bryan, K.; de Lange, W.; Pilditch, C. The Hydrodynamics of the Southern Basin of Tauranga Harbour. *N. Z. J. Mar. Freshw. Res.* **2013**, *47*, 249–274. [CrossRef]
40. Collier, C.J.; Villacorta-Rath, C.; van Dijk, K.; Takahashi, M.; Waycott, M. Seagrass Proliferation Precedes Mortality during Hypo-Salinity Events: A Stress-Induced Morphometric Response. *PLoS ONE* **2014**, *9*, e94014. [CrossRef] [PubMed]
41. York, P.H.; Gruber, R.K.; Hill, R.; Ralph, P.J.; Booth, D.J.; Macreadie, P.I. Physiological and Morphological Responses of the Temperate Seagrass Zostera Muelleri to Multiple Stressors: Investigating the Interactive Effects of Light and Temperature. *PLoS ONE* **2013**, *8*, e76377. [CrossRef] [PubMed]
42. Collier, C.J.; Ow, Y.X.; Langlois, L.; Uthicke, S.; Johansson, C.L.; O'Brien, K.R.; Hrebien, V.; Adams, M.P. Optimum Temperatures for Net Primary Productivity of Three Tropical Seagrass Species. *Front. Plant Sci.* **2017**, *8*. [CrossRef] [PubMed]
43. Turner, S.J. Growth and Productivity of Intertidal Zostera Capricorni in New Zealand Estuaries. *N. Z. J. Mar. Freshw. Res.* **2007**, *41*, 77–90. [CrossRef]
44. Ramage, D.L.; Schiel, D.R. Reproduction in the Seagrass Zostera Novazelandica on Intertidal Platforms in Southern New Zealand. *Mar. Biol.* **1998**, *130*, 479–489. [CrossRef]
45. Schwarz, A.-M.; Turner, S. *Management and Conservation of Seagrass in New Zealand: An Introduction*; Science & Technical Publishing: Wellington, New Zealand, 2006; p. 90.
46. Hicks, M.; Semadeni-Davies, A.; Haddadchi, A.; Shankar, U.; Plew, D. *Updated Sediment Load Estimator for New Zealand*; NIWA Client Report 2018341CH prepared for Ministry for the Environment; National Institute of Water and Atmospheric Research Ltd.: Christchurch, New Zealand, 2019; p. 190.
47. Hicks, M. *Review and Analysis of Suspended Sediment Monitoring in the Tauranga Moana Catchment*; NIWA Client Report 2019183CH prepared for Bay of Plenty Regional Council; NIWA: Auckland, New Zealand, 2019; p. 57.
48. Cabaço, S.; Santos, R.; Duarte, C.M. The Impact of Sediment Burial and Erosion on Seagrasses: A Review. *Estuar. Coast. Shelf Sci.* **2008**, *79*, 354–366. [CrossRef]
49. Saunders, M.I.; Atkinson, S.; Klein, C.J.; Weber, T.; Possingham, H.P. Increased Sediment Loads Cause Non-Linear Decreases in Seagrass Suitable Habitat Extent. *PLoS ONE* **2017**, *12*, e0187284. [CrossRef]
50. USGS GloVis. Available online: https://glovis.usgs.gov/ (accessed on 2 February 2019).
51. Park, S.G. *Changes in Abundance of Seagrass (Zostera spp.) in Tauranga Harbour from 1959–1996*; Environment BOP: Whakatane, New Zealand, 1999; p. 19.
52. RBINS. *ACOLITE Python User Manual*; RBINS: Brussels, Belgium, 2018; p. 10.
53. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
54. Breiman, L. Random Forest. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
55. Belgiu, M.; Drăguţ, L. Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS J. Photogramm. Remote Sens.* **2016**, *114*, 24–31. [CrossRef]
56. Mountrakis, G.; Im, J.; Ogole, C. Support Vector Machines in Remote Sensing: A Review. *ISPRS J. Photogramm. Remote Sens.* **2011**, *66*, 247–259. [CrossRef]
57. Tomar, D.; Agarwal, S. Twin Support Vector Machine: A Review from 2007 to 2014. *Egypt. Inform. J.* **2015**, *16*, 55–69. [CrossRef]
58. Georganos, S.; Grippa, T.; Vanhuysse, S.; Lennert, M.; Shimoni, M.; Wolff, E. Very High Resolution Object-Based Land Use–Land Cover Urban Classification Using Extreme Gradient Boosting. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 607–611. [CrossRef]
59. Prokhorenkova, L.; Gusev, G.; Vorobev, A.; Dorogush, A.V.; Gulin, A. CatBoost: Unbiased Boosting with Categorical Features. *arXiv* **2019**, arXiv:170609516.
60. NIWA. The National Climate Database. Available online: https://cliflo.niwa.co.nz/ (accessed on 20 July 2020).
61. BOPRC Seagrass Extent—Bay of Plenty. Available online: http://gis.boplass.govt.nz/arcgis/rest/services/imagery (accessed on 20 October 2019).
62. Conrad, O.; Bechtel, B.; Bock, M.; Dietrich, H.; Fischer, E.; Gerlitz, L.; Wehberg, J.; Wichmann, V.; Böhner, J. System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geosci. Model Dev.* **2015**, *8*, 1991–2007. [CrossRef]

63. Virtanen, P.; Gommers, R.; Oliphant, T.E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nat. Methods* **2020**, *17*, 261–272. [CrossRef] [PubMed]
64. Raschka, S. MLxtend: Providing Machine Learning and Data Science Utilities and Extensions to Python's Scientific Computing Stack. *J. Open Source Softw.* **2018**, *3*, 638. [CrossRef]
65. Traganos, D.; Reinartz, P. Mapping Mediterranean Seagrasses with Sentinel-2 Imagery. *Mar. Pollut. Bull.* **2017**. [CrossRef]
66. Sun, F.; Wang, R.; Wan, B.; Su, Y.; Guo, Q.; Huang, Y.; Wu, X. Efficiency of Extreme Gradient Boosting for Imbalanced Land Cover Classification Using an Extended Margin and Disagreement Performance. *ISPRS Int. J. Geo-Inf.* **2019**, *8*, 315. [CrossRef]
67. Pham, T.D.; Le, N.N.; Ha, N.T.; Nguyen, L.V.; Xia, J.; Yokoya, N.; To, T.T.; Trinh, H.X.; Kieu, L.Q.; Takeuchi, W. Estimating Mangrove Above-Ground Biomass Using Extreme Gradient Boosting Decision Trees Algorithm with Fused Sentinel-2 and ALOS-2 PALSAR-2 Data in Can Gio Biosphere Reserve, Vietnam. *Remote Sens.* **2020**, *12*, 777. [CrossRef]
68. Pham, T.D.; Yokoya, N.; Xia, J.; Ha, N.T.; Le, N.N.; Nguyen, T.T.T.; Dao, T.H.; Vu, T.T.P.; Pham, T.D.; Takeuchi, W. Comparison of Machine Learning Methods for Estimating Mangrove Above-Ground Biomass Using Multiple Source Remote Sensing Data in the Red River Delta Biosphere Reserve, Vietnam. *Remote Sens.* **2020**, *12*, 1334. [CrossRef]
69. Lou, I.; Xie, Z.; Ung, W.K.; Mok, K.M. Freshwater Algal Bloom Prediction by Extreme Learning Machine in Macau Storage Reservoirs. *Neural Comput. Appl.* **2016**, *27*, 19–26. [CrossRef]
70. Maier, P.M.; Keller, S. Machine Learning Regression on Hyperspectral Data to Estimate Multiple Water Parameters. In Proceedings of the 2018 9th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Amsterdam, The Netherlands, 23–26 September 2018; IEEE: Amsterdam, The Netherlands, September 2018; pp. 1–5.
71. Lee, Y.; Han, D.; Ahn, M.-H.; Im, J.; Lee, S.J. Retrieval of Total Precipitable Water from Himawari-8 AHI Data: A Comparison of Random Forest, Extreme Gradient Boosting, and Deep Neural Network. *Remote Sens.* **2019**, *11*, 1741. [CrossRef]
72. Vanhellemont, Q. Adaptation of the Dark Spectrum Fitting Atmospheric Correction for Aquatic Applications of the Landsat and Sentinel-2 Archives. *Remote Sens. Environ.* **2019**, *225*, 175–192. [CrossRef]
73. Karamizadeh, S.; Abdullah, S.M.; Halimi, M.; Shayan, J.; Rajabi, M.j. Advantage and Drawback of Support Vector Machine Functionality. In Proceedings of the 2014 International Conference on Computer, Communications, and Control Technology (I4CT), Langkawi, Malaysia, 2–4 September 2014; IEEE: Langkawi, Malaysia, 2014; pp. 63–65.
74. Pham, T.D.; Yokoya, N.; Nguyen, T.T.T.; Le, N.N.; Ha, N.T.; Xia, J.; Takeuchi, W.; Pham, T.D. Improvement of Mangrove Soil Carbon Stocks Estimation in North Vietnam Using Sentinel-2 Data and Machine Learning Approach. *GISci. Remote Sens.* **2020**, *58*, 1–20. [CrossRef]
75. Luo, M.; Wang, Y.; Xie, Y.; Zhou, L.; Qiao, J.; Qiu, S.; Sun, Y. Combination of Feature Selection and CatBoost for Prediction: The First Application to the Estimation of Aboveground Biomass. *Forests* **2021**, *12*, 216. [CrossRef]
76. Dos Santos, V.M.; Matheson, F.E.; Pilditch, C.A.; Elger, A. Is Black Swan Grazing a Threat to Seagrass? Indications from an Observational Study in New Zealand. *Aquat. Bot.* **2012**, *100*, 41–50. [CrossRef]
77. USGS Landsat 9. Available online: Https://Www.Usgs.Gov/Land-Resources/Nli/Landsat/Landsat-9 (accessed on 20 November 2019).
78. Macreadie, P.I.; Anton, A.; Raven, J.A.; Beaumont, N.; Connolly, R.M.; Friess, D.A.; Kelleway, J.J.; Kennedy, H.; Kuwae, T.; Lavery, P.S.; et al. The Future of Blue Carbon Science. *Nat. Commun.* **2019**, *10*, 3998. [CrossRef] [PubMed]
79. Cussioli, M.C.; Bryan, K.R.; Pilditch, C.A.; de Lange, W.P.; Bischof, K. Light Penetration in a Temperate Meso-Tidal Lagoon: Implications for Seagrass Growth and Dredging in Tauranga Harbour, New Zealand. *Ocean Coast. Manag.* **2019**, *174*, 25–37. [CrossRef]

Appendix 5. Co-authorship form of research chapter 5

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

Postgraduate Studies Office
Student and Academic Services Division
Wahanga Ratonga Matauranga Akonga
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 4439
Website: http://www.waikato.ac.nz/sasd/postgraduate/

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work.

Chapter 5: The use of radar and optical satellite imagery combined with advanced machine learning and metaheuristic optimization techniques to detect and quantify above ground biomass of intertidal seagrass in a New Zealand estuary

Nature of contribution by PhD candidate	Conceptualization, methodology, software, validation, resources, writing-original draft preparation, write-review and editing.
Extent of contribution by PhD candidate (%)	75

CO-AUTHORS

Name	Nature of Contribution
Merilyn Manley-Harris	Resources, writing-review and editing.
Tien Dat Pham	Validation, writing-review and editing.
Ian Hawes	Conceptualization, resources, writing-review and editing, supervision.

Certification by Co-Authors

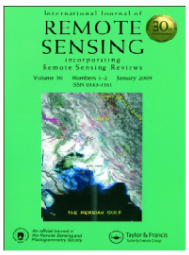
The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and

Name	Signature	Date
Merilyn Manley-Harris		21.1.21
Tien Dat Pham		21.01.2021
Ian Hawes		24.02.2021

July 2015

Appendix 6. Published paper in the International Journal of Remote Sensing (chapter 5)



International Journal of Remote Sensing

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tres20

The use of radar and optical satellite imagery combined with advanced machine learning and metaheuristic optimization techniques to detect and quantify above ground biomass of intertidal seagrass in a New Zealand estuary

Nam Thang Ha, Merilyn Manley-Harris, Tien Dat Pham & Ian Hawes

To cite this article: Nam Thang Ha, Merilyn Manley-Harris, Tien Dat Pham & Ian Hawes (2021) The use of radar and optical satellite imagery combined with advanced machine learning and metaheuristic optimization techniques to detect and quantify above ground biomass of intertidal seagrass in a New Zealand estuary, International Journal of Remote Sensing, 42:12, 4716-4742, DOI: 10.1080/01431161.2021.1899335

To link to this article: https://doi.org/10.1080/01431161.2021.1899335

Published online: 23 Mar 2021.

INTERNATIONAL JOURNAL OF REMOTE SENSING
2021, VOL. 42, NO. 12, 4716–4742
https://doi.org/10.1080/01431161.2021.1899335

The use of radar and optical satellite imagery combined with advanced machine learning and metaheuristic optimization techniques to detect and quantify above ground biomass of intertidal seagrass in a New Zealand estuary

Nam Thang Ha[a,b], Merilyn Manley-Harris[a], Tien Dat Pham[c,d] and Ian Hawes[a]

[a]Environmental Research Institute, School of Science, University of Waikato, Hamilton, New Zealand; [b]Faculty of Fisheries, University of Agriculture and Forestry, Hue University, Hue, Vietnam; [c]Department of Biological Sciences, Florida International University (FIU), Miami, USA; [d]Institute of Environment, Florida International University (FIU), Miami, USA

ABSTRACT

Seagrass provides numerous valuable ecosystem services across a wide range of climatic regions. However, in terms of area and habitat, this resource is in decline globally and there is an urgent need for accurate mapping of extant meadows and biomass to support sustainable seagrass blue carbon conservation and management. This study develops a novel method for a binary mapping of seagrass distribution and estimating seagrass above-ground biomass (AGB) by applying a suite of advanced machine learning (ML) algorithms combined with and without a metaheuristic optimization approach (particle swarm optimization – PSO) to various combinations of multispectral (Sentinel-2) and synthetic aperture radar (Sentinel-1) remote sensing data. Our results reveal that the Sentinel-1 data has potential for the binary mapping of seagrass meadows using an extreme gradient boosting (XGB) model (scores of precision (P) = 0.82, recall (R) = 0.90, and F_1 = 0.86) but is less effective at estimating AGB. The optimal method for estimation of AGB used both Sentinel-1 and Sentinel-2 imagery, the XGB model, and PSO optimization (coefficient of determination (R^2) = 0.75, root mean squared error (RMSE) = 0.35, Akaike information criteria (AIC) = 24.80, Bayesian information criteria (BIC) = 44.70). Our findings contribute novel and advanced methods for seagrass detection and improvement of AGB estimation, which are fast and reliable, use open-source data and software and should be easily applicable to intertidal zones across many regions of the world.

ARTICLE HISTORY
Received 13 November 2020
Accepted 4 February 2021

1. Introduction

Seagrasses are a group of marine angiosperms, which survive in a wide range of environmental conditions and are widely distributed in coastal zones. Several valuable ecosystem services are recognized for this resource (Mtwana Nordlund et al. 2016), including wave attenuation (Fonseca and Cahalan 1992; Reidenbach and Thomas 2018), pollution

CONTACT Nam Thang Ha hanamthang@huaf.edu.vn Environmental Research Institute, School of Science, University of Waikato, Hamilton, New Zealand.

attenuation (Short and Short 1984), nursery and breeding habitat (Heck Hay, Hays, and Orth 2003) and, most recently, carbon sequestration (Bedulli et al. 2020). Recognition of the importance of seagrass systems as a blue carbon resource has resulted in initiatives to foster a long-term strategy of reversing the current decline in seagrass area/biomass (Waycott et al. 2009), and carbon sequestration (Aoki et al. 2020; Githaiga et al. 2019) world-wide as a nature-based solution in averting climate change (Macreadie et al. 2019). To facilitate this, a precise method for mapping the distribution and biomass in seagrass ecosystems, under different environmental conditions (Unsworth et al. 2018), is needed to support its inclusion in the carbon credits market (Macreadie et al. 2019; Ricart et al. 2020).

Recently, mapping seagrass distribution and above-ground biomass (AGB) has been investigated either using intensive field surveys or data-driven models developed from earth observation (EO) data and statistical techniques (Roelfsema et al. 2015; Sani, Hashim, and Hossain 2019). The former approach is highly accurate, using *in-situ* measurements for each sampling plot. However, it is labour intensive, is time-consuming, and cannot be readily upscaled (Unsworth et al. 2018). In contrast, the use of optical remote sensing data for mapping seagrass ecosystems allows low cost, large area spatial mapping, and fast performance (Hossain et al. 2015; Pham et al. 2019; Veettil et al. 2020; Ha et al. 2020). EO-based techniques, however, have unresolved challenges of coarse spatial resolution, insufficient accuracy for retrieval models, cloud cover constraints, and the effect of immersion on retrieval algorithms in intertidal environments.

For seagrass distribution mapping, researchers have developed various techniques for image preprocessing and classification (Hossain et al. 2015; Pham et al. 2019). Recent studies have used the multispectral Land Remote-Sensing Satellite (System, Landsat) 5 Thematic Map (TM), Landsat 8 Operational Land Imager (OLI) (Misbari and Hashim 2016), Landsat 5 TM, and Hyperion (Pu and Bell 2013), or very high spatial resolution (VHR) optical data, including QUICKBIRD, IKONOS, Worldview-2 (Koedsin et al. 2016; Roelfsema et al. 2014) for the retrievals of seagrass AGB using linear regression models. The coefficient of determination (R^2) varied between studies, with higher values, ranging from 0.34 to 0.68, observed when using VHR images.

Despite the obvious advantages of using spectral bands for seagrass (species) detection, these EO techniques are constrained by the need for scene to be captured at suitable time during daylight and free of cloud. The high cost of VHR images adds to the constraints of analysing large areas. Therefore, the development of novel approaches using other EO data that escape some of these constraints (*i.e*: synthetic aperture radar (SAR) images that can be obtained regardless of time or weather conditions during overpass) may better suit the requirement of rapid assessment of seagrass change globally. To the date of this study, we found no application of SAR data for seagrass meadows mapping of presence-absence, cover, or AGB.

The current situation regarding the broad-scale estimation of the contribution of seagrass to blue carbon storage, is that optical satellite data can be used for seagrass detection and estimation of AGB, however with a variable and generally low precision, and with the constraint of requiring the absence of cloud cover, suitable ambient lighting, and low tide to coincide with satellite overpass. In a previous study that used Landsat imagery to address 40 year trends in seagrass cover in Tauranga Harbour (New Zealand), we found this to seriously constrain usable observations. To address this limitation we elected to assess the potential application of SAR remote

sensing to map seagrass distribution and to estimate seagrass biomass, due to its freedom from the need for cloud-free observing conditions with the sun in a suitable position. We also investigate the use of fusions of multi-spectral and SAR remotely sensed images, to improve accuracy of models for biophysical parameters estimation of seagrass ecosystems in the coastal areas.

Given a list of available satellite resources, we selected Sentinel-1 (S-1) and Sentinel-2 (S-2), which were launched in the years 2014 and 2015, are operated by the European Space Agency (ESA), and provide freely accessible remote sensing data (ESA-S1 2020; ESA-S2 2020). S-1 provides dual-polarimetric SAR imagery, acquired day and night at C-band (dual-polarization vertical transmitting and horizontal receiving (VH) and vertical transmitting and vertical receiving (VV)) in any weather conditions, at a wide swath of 250 km and a spatial resolution of 20 m (level 1 production). A favourable S-1 image therefore only requires that the overpass is at low tide. S-2 provides 12 multispectral bands ranging from 400 to 2400 nm and spatial resolutions ranging from 10 to 60 m. For coastal environments, the blue, and the green wavelengths are expected to have a good penetration through the water environment, whilst the red and the three red-edge bands of S-2 produce a range of vegetation and soil indices, which might be useful for the estimation of biophysical parameters such as AGB retrievals (Pham et al. 2020a; 2020b; 2020c). S-1 and S-2 have repeating cycles of 12 and 5 days, respectively. In this work, we hypothesize that the S-1 C-band sensor may be suited to estimate the meadow structure despite the low canopy height of seagrass ecosystems (El Hajj et al. 2018; Howell et al. 2018) whereas the S-2 optical sensor can be useful to extract spectral information that will assist in the estimation of seagrass AGB.

To extract information from these earth-observing sensors we elected to use machine learning (ML) models, which have been developing rapidly in recent years. The current literature reveals a limited application of ML models for the retrieval of biophysical parameters of seagrass ecosystem, but successful implementation in various domains (Huettmann 2018; Thessen 2016). Currently, the ensemble decision tree-based approaches of bagging and boosting algorithms that have been applied in numerous field of studies include the Random Forest (RF) (Belgiu and Drăguţ 2016), Rotation Forest (RoF) (Rodriguez, Kuncheva, and Alonso 2006; Ha et al. 2020), Canonical Correlation Forest (CCF) (Rainforth and Wood 2015), Support Vector Machine (SVM) ((Mountrakis et al. 2011), Gradient Boost (GB) (Natekin and Knoll 2013), Extreme Gradient Boost (XGB) (Chen and Guestrin 2016), CatBoost (CB) (Prokhorenkova et al. 2019), and Light Gradient Boosting Machine (LGBM) (Ke et al. 2017). In general, they produce better predictive performance, faster speed, and often outperform parametric models but, to our knowledge, have seldom been used for seagrass mapping and have never been used to retrieve seagrass AGB. Therefore, in this work we attempted to develop novel approaches for first the binary (presence/absence) mapping of seagrass meadows using S-1 images. For retrieving seagrass AGB, we investigated the use of various combinations of S-1 and S-2 images and a range of advanced ML techniques. Our novel, integrated approach is expected to strengthen the utilization of multisource remote sensing data in accurately detecting the distribution and predicting the AGB of extant seagrass meadows and to support their integration into the blue carbon strategy to deal with climate change impacts.

2. Materials and Methods

2.1. Study site

Tauranga Harbour (New Zealand) (Figures 1*(a)* – *(b)*) was the study site; at this site, seagrass meadows are patchy and distributed mainly in the centre and southern parts of the harbour (Ha et al. 2020). The tide regime is semi-diurnal with a range from 0.20 to 2.10 m. The only species of seagrass in the harbour, as in the rest of New Zealand, is *Zostera muelleri*, which is distributed in the inter-tidal zone of the harbour (Park 1999). Compared to other seagrasses, *Z. muelleri* is small (0.10 to 0.40 cm in leaf width), with optimal growth at 27°C to 30°C (Collier et al. 2017; York et al. 2013) and 12 practical salinity units (psu) (Collier et al. 2014). *Z. muelleri* biomass shows a seasonal rhythm (Turner 2007), with the highest biomass in the austral summer (December to March) and declining during the austral winter to minimum biomass in early September.

2.2. Field survey

2.2.1. Seagrass collection

The field survey was conducted during the austral summer (March 2020) season in the inter-tidal areas of the harbour, when biomass would be approaching the seasonal maximum. Fifty-seven sampling plots (size 10 m × 10 m) across the harbour were selected (Figures 1*(a)* – *(b)*). The geographic position of each plot was recorded using a Global

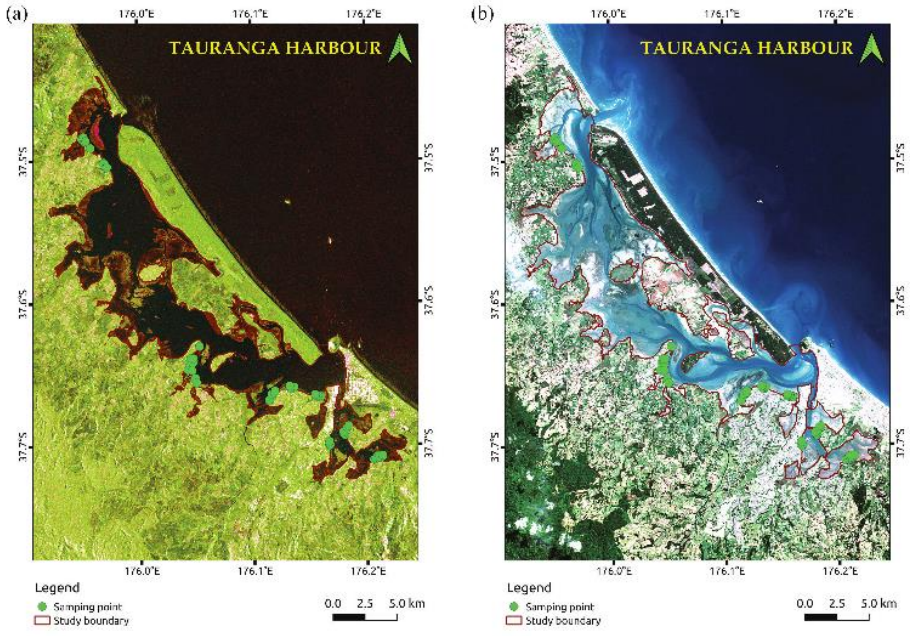


Figure 1. Sentinel imagery of Tauranga Harbour (New Zealand) and ground truth points used in the investigation: (*a*) S-1 image (acquisition date 31 March 2020) used for seagrass binary mapping, (*b*) S-2 image (pseudo colour using a composition of bands ρ_{Red} – ρ_{Green} – ρ_{Blue}, acquisition date 5 April 2020). The red line indicates the mask used in analysis of remote imagery.

Positioning System (GPS) Extrex-30 with an accuracy of ± 2 m and used as the ground truth points (GTPs) for the seagrass binary mapping using the S-1 image.

A 0.5 m × 0.5 m square quadrat and a plastic core were used to collect seagrass for the measurement of the AGB as dry weight. The plastic core was used within each plot to take three cores, each 15 cm in diameter and 40 cm in depth in representative locations (Figures 2*(a)* – *(b)*). The cores were inserted to a depth of 20 cm and the seagrass samples were washed free of sediment at the field sites, then kept at – 10°C until analysis in the laboratory.

2.2.2. Sample processing and dry AGB measurement

In the laboratory, the seagrass sample was divided into above-ground and below-ground parts. The above-ground part consisted of seagrass shoots whilst the below-ground part comprised the rhizome and the root. We observed no epiphytes on the leaves of the seagrass samples. The above-ground part of seagrass was dried at 60°C for 48 hours, cooled in a desiccator, and weighed to ± 1 mg. An average AGB, normalized to the unit of g dry weight m^{-2} (g DW m^{-2}), for each site was used in the production of the AGB distribution map.

2.3. Development of models

2.3.1. Satellite image acquisition

S-1 and S-2 data were downloaded from the Copernicus open access hub (ESA 2020b) and the United States geological survey (USGS) global visualization viewer (USGS-Glovis) (USGS 2020), respectively, in the projection of World Geodetic System (WGS84) (S-1) and WGS84-Universal Transverse Mercator (UTM), zone 60 South (60S) (S-2) (Table 1). The images were acquired to have the acquisition time closest to each other and the low tide in Tauranga Harbour, and as close as possible to the field data collection dates (4 to 25 March 2020). Despite the close coincidence of low tide and overpass, comparison with bathymetry suggests that both submerged and exposed intertidal areas in the harbour

(a) (b)

Figure 2. Seagrass samples collection in Tauranga Harbour using (*a*) quadrat and (*b*) plastic core. Photos were taken by N.T.H in March 2020.

Table 1. S-1 and S-2 data acquisition details.

Sensor	Processing level	Band used	Date of acquisition	Time of acquisition	Corresponding low tide
S-1	Ground range detected (GRD) level 1 sensor mode: interferometric wide swath (IW)	Dual-polarimetric VH and VV	31 March 2020	19:07	19:02
S-2	Level 1 C	Band 1 to 8A vegetation (8A-V), 11, 12	5 April 2020	10:15	10:17

were present at the time of satellite image acquisition (Reeve, Stephens, and Wadhwa 2018).

2.3.2. Sentinel-1 preprocessing and image transformation

The preprocessing of the S-1 image consisted of multiple steps to convert the pixel values from the raw intensity signal to the backscattering coefficients (σ^0) in decibel units (dB), which were used for seagrass detection and retrieval of AGB. The S-1 C-band intensities of the VH and the VV polarizations were processed to the normalized radar σ^0 as suggested by (Filipponi 2019). This involves the following steps: (1) Correct the orbit file; (2) Thermal noise removal; (3) Border noise removal; (4) Radiometric calibration; (5) Speckle filtering; (6) Range Doppler terrain correction: before (7) conversion of the pixel values to the normalized radar backscattering coefficient (σ^0) using Equation (1) in which DN is the digital number of the intensity image.

$$\sigma^0 = 10 \times \log_{10}(\mathrm{DN}^2) \tag{1}$$

The band transformation was conducted for S-1 VH and VV polarizations as follows: (1) band combinations were derived as VV/VH, VH/VV, VV− VH, VH− VV, (VV + VH)/2; (2) applying the grey level co-occurrence matrix (GLCM) feature extraction to create 20 new bands (VH contrast (VH_cons), VH dissimilarity (VH_diss), VH homogeneity (VH_homo), VH asm (VH_asm), VH energy (VH_energy), VH max (VH_max), VH entropy (VH_entropy), VH mean (VH_mean), VH variance (VH_variance), VH correlation (VH_corr), VV contrast (VV_cons), VV dissimilarity (VV_diss), VV homogeneity (VV_homo), VV asm (VV_asm), VV energy (VV_energy), VV max (VV_max), VV entropy (VV_entropy), VV mean (VV_mean), VV variance (VV_variance), VV correlation (VV_corr)); (3) extracting the first principal component (PCA_1) from principal component analyses (PCA) of firstly the seven bands developed in step 1 (VV, VH, VV/VH, VH/VV, VV− VH, VH− VV, (VV + VH)/2)) and secondly the 20 GLCM bands developed in step (2), resulting in the PCA_1_7_band and PCA_1_GLCM. The GLCM and PCA features were both extracted using the Sentinel application platform (SNAP) toolbox (ESA 2020a). A total of 29 bands (2 dual polarizations and 27 transformed bands) were involved as the input variables/features from the S-1 analysis for use in the estimations of seagrass AGB whilst a total of 8 bands (2 dual polarizations and 6 transformation bands (VV/VH, VH/VV, VV− VH, VH− VV, (VV + VH)/2), PCA_1_7_band) were used for the binary mapping in this study (see Table S5, the supplemental material) for the abbreviations and full names of the transformations).

All the steps were accomplished in the SNAP application using the SAR toolset (ESA 2020a). Finally, we resampled to a spatial resolution of 10 m and co-registered the projection to WGS84 UTM 60S to match S-2 imagery.

2.3.3. Sentinel-2 image atmospheric correction and transformation

The level 1 C S-2 is downloaded as the top of atmosphere (TOA) processing level and was converted to surface reflectance using the atmospheric correction for operational land imager (OLI) 'lite' toolbox (ACOLITE) (Vanhellemont 2016) with the dark spectrum algorithm (Vanhellemont 2019) (Table S1, the supplemental material).

For the retrieval of AGB, the literature indicates the usefulness of different vegetation indices (VIs) and soil radiometric indices (SIs) (Pham et al. 2020a; Xue and Baofeng 2017). Therefore, we computed a number of band transformations to generate different VIs and SIs (Table 2). A total of 22 bands (11 multispectral bands, 11 bands of VIs and SIs) were computed as the input variables for the retrievals of AGB from the S-2 scene (see Table S5, the supplemental material for the abbreviations and full names of the transformations).

2.3.4. Seagrass mapping using Sentinel-1

Our previous work on seagrass binary mapping using Landsat images in Tauranga Harbour indicated that XGB and CB outperformed other ML models (Thang et al. in review). Therefore, we evaluated the classification skills of XGB and CB models for the presence/absence mapping of seagrass meadows using only S-1 images here. Using the GTPs from the field survey (March 2020), two classes of regions of interest (ROIs) (seagrass

Table 2. Vegetation and soil index transformed from S-2 bands. Band wavelengths of S-2: ρ_{Blue} (460 nm), ρ_{Green} (560 nm), ρ_{Red} (665 nm), Red-edge 1 (RE_1 (704 nm)), Red-edge 2 (RE_2 (740 nm)), Red-edge 3 (RE_3 (783 nm)), Near-infrared (ρ_{NIR}) (833 nm).

Vegetation index	Acronym	S-2 band wavelength used	Reference
Ratio vegetation index	RVI	$\frac{\rho_{NIR}}{\rho_{Red}}$	(Tucker 1979)
Normalized difference vegetation index	NDVI	$\frac{\rho_{NIR}-\rho_{Red}}{\rho_{NIR}+\rho_{Red}}$	(Rouse et al. 1974)
Green normalized difference vegetation index	GNDVI	$\frac{\rho_{NIR}-\rho_{Green}}{\rho_{NIR}+\rho_{Green}}$	(Gitelson, Kaufman, and Merzlyak 1996)
Enhanced vegetation index 2	EVI2	$2.5 \times \frac{\rho_{NIR}-\rho_{Red}}{\rho_{NIR}+2.4\times\rho_{Red}+1}$	(Jiang et al. 2008)
Normalized difference index using bands 4 & 5 of S-2	NDI45	$\frac{RE_1-\rho_{Red}}{RE_1+\rho_{Red}}$	(Delegido et al. 2011)
Soil-adjusted vegetation index	SAVI	$(1+L)\times\frac{\rho_{NIR}-\rho_{Red}}{\rho_{NIR}+\rho_{Red}+L}$ $L = 0.5$ in most conditions	(Huete 1988)
Inverted red-edge chlorophyll index	IRECI	$\frac{RE_3-\rho_{Red}}{RE_1/RE_2}$	(Frampton et al. 2013)
Modified chlorophyll absorption in reflectance index	MCARI	$[(RE_1-\rho_{Red})-0.2\times(RE_1-\rho_{Green})]\times(RE_1-\rho_{Red})$	(Daughtry 2000)
Soil Index	Acronym	S-2 Band Wavelengths Used	Reference
Brightness index	BI	$\sqrt{\frac{\rho_{Blue}^2+\rho_{Green}^2+\rho_{Red}^2}{3}}$	(Escadafal 1989)
Redness index	RI	$\frac{\rho_{Red}^2}{\rho_{Blue}\times\rho_{Green}^3}$	(Mathieu et al. 1998)
Colour index	CI	$\frac{\rho_{Red}-\rho_{Green}}{\rho_{Red}+\rho_{Green}}$	(Mathieu et al. 1998)

Table 3. Number of input bands and field data observations used for various scenarios.

		For AGB estimation		For binary mapping	
	Sensor	No. of input band	No. of observation	No. of input band	No. of training/testing pixel
	S1 + SAR transformation			8	3261/2174
Scenario 1	S-1 + SAR transformation	29	57		
Scenario 2	S-2 + VIs + SIs	22	57		
Scenario 3	S-1 + SAR transformation + S-2 + VIs + SIs	51	57		
Scenario 4	S-1 + SAR transformation + S-2 + VIs + SIs and Spearman feature selection	47	57		
Scenario 5	S-1 + SAR transformation + S-2 + VIs + SIs and PSO feature selection	26	57		

and non-seagrass) were created with a total of 5435 pixels selected from S-1 images (Table 3). The dataset was randomly split into 60% for the training and 40% for the testing of the selected ML models. The best model with highest scores of precision (P) and F_1 was used to produce the binary map of seagrass meadows in Tauranga Harbour.

2.3.5. Seagrass AGB estimation

We examined all available integrations of S-1 and S-2 data using five scenarios, and in each the performance of various ML models for AGB estimation was evaluated (Table 3). Scenarios 1 and 2 used only the original bands and transformations of S-1 and S-2 images, respectively. Scenario 3 used both the original bands and transformations of S-1 and S-2 images and Scenario 4 used a band subset derived from a Spearman correlation analysis between each input band and AGB. In this case, input bands with a Spearman coefficient > 0.10 were selected to include in scenario 4. In scenario 5, the best retrieval model from previous scenarios was integrated with a Particle Swarm Optimization (PSO) procedure designed to select the best combination of bands (see section 2.3.7).

2.3.6. Retrieval of AGB using ML models

2.3.6.1. ML models used in the study.

2.3.6.1.1. Random Forest (RF). To date, the RF algorithm (Breiman 2001) would be the most popular ML model applied for a wide range of classification and regression problems (Belgiu and Drăguț 2016; Pal 2005). The RF model is reliable, with bootstrap sampling and ensemble bagging trees used to derive the most accurate decision from an ensemble of the weak learners. During the learning, approximately 2/3 of the samples (in-bag) is used for the training phase and 1/3 of the samples (out-of-bag) is used for the testing phase. Similar to other tree-based and boosting ML models, the RF model estimates the importance of each variable as predictors in the model. The most notable hyper-parameters of the RF model involve the maximum depth, minimum sample leaf, minimum sample split, maximum features, and number of trees that can be tuned to fit to a specific dataset.

2.3.6.1.2. Support Vector Machine (SVM). SVM is a well-known algorithm for supervised learning applied for various non-linear problems (Gholami and Fakhari 2017; Mountrakis et al. 2011). Given a classification or regression problem, the SVM algorithm defines the hyperplanes with the support vectors, the closest points to the hyperplanes, and the margin, the gap between the classified classes. Compared to other ML models, the SVM model consists of fewer hyper-parameters of the kernel function (linear, poly-nominal or radial basis), gamma (γ) controlling the overfitting, and regularization parameter (C). The SVM model is able to deal with non-linear data; however, not efficiently when the data is noisy or comprises overlapped data classes (Fehr, Arreola, and Burkhardt 2008; Mountrakis et al. 2011).

2.3.6.1.3. Extreme Gradient Boost (XGB). Developed by (Chen and Guestrin 2016), the XGB algorithm shares the same theory of boosting technique with the gradient-boosting family; however it uses a more regularized model to produce scalable, and accurate prediction. The XGB model computes the second-order gradients of the loss function which provide essential information on minimizing the loss function. The introduction of L_1 and L_2 regularization models improves the generalization, and therefore reduces the

overfitting. Parallelization, out-of-core computation, and cache optimization are further advantages of the XGB, which reduces the training time when dealing with a big dataset. In addition, the XGB algorithm can be wrapped inside various optimization algorithms to tune the hyper-parameters for specific problems. The most important hyper-parameters involve the booster, maximum depth, minimum child weight, number of trees and learning rate.

2.3.6.1.4. Light Gradient Boosting Machine (LGBM). The LGBM algorithm is designed for big data processing, but is still applicable for small-scale datasets (Ke et al. 2017). The LGBM technique inherits the advantages of decision tree-based and gradient boosting algorithms, in addition to a fast training speed with parallel computation, lower memory usage, and is capable of large-scale data processing. The LGBM grows the decision tree using the leaf-wise mode rather than the level-wise mode used in other tree-based ML models, which results in higher prediction accuracy; however, it may also lead to excessive tree complexity and overfitting. The most important hyper-parameters of the LGBM model include the number of trees, number of leaves, maximum depth, which may reduce the model overfitting, and the gradient boosting model such as Dropouts meet Multiple Additive Regression Tree (DART), Gradient-based One-Side Sampling (GOSS), and Gradient Boosting Decision Tree (GBDT). On the other hand, the LGBM algorithm performs well with high dimensional data (Li et al. 2018), which might be appropriate to the multiple bands data in this study.

2.3.6.1.5. CatBoost (CB). Of the ML algorithms used in this study, the CB algorithm is the most recent addition to ML techniques, that have been used for an extensive range of classification, regression, and ranking tasks (Prokhorenkova et al. 2019). The CB algorithm is able to handle both category and numerical data. The CB model introduced ordered boosting, which was modified from the standard gradient boosting algorithm and helps to avoid 'shifting' in prediction due to target leakage. During the algorithm's learning, the CB algorithm uses the oblivious decision tree instead of the asymmetric tree used in the XGB and the LGBM algorithms. The CB technique supports graphics processing unit (GPU) processing, significantly reduces training time, and is easy to apply for a user-friendly application programming interface (API). The CB algorithm is robust as it uses only a few default hyper-parameters such as depth, number of trees, and learning rate.

2.3.6.1.6. Hyper-parameter tuning for selected ML models. In this study, we employed a five-fold cross-validation (CV) using the GridSearch in the Scikit-learn library (Pedregosa et al. 2011) to find the best combination of each model's hyper-parameters (See Table S2, the supplemental material for seagrass binary mapping and Table S3 (*a*) – (*d*), the supplemental material for seagrass AGB estimation).

2.3.7. Particle Swarm Optimization (PSO) application for feature selection

2.3.7.1. Introduction to PSO.

PSO is a powerful metaheuristic search algorithm for optimization of solutions to non-linear functions introduced by (Kennedy and Eberhart 1995). Implementing a swarm-based stochastic optimization, the PSO uses the knowledge gained by each of the swarm of particles to affect the behaviour of other particles in a search space of n dimensions. Each particle is controlled by three parameters inertia weight (w), cognitive (c_1) and social (c_2) components. The particles then implement the searches in n dimensional space,

which correspond to the input features, returning the best positions in the search space with the minimum fitness during the optimization of feature selection (De Almeida, Bruno, and Leite 2019; Sengupta, Basak, and Peters 2018). Compared to the Genetic algorithm (GA), PSO shares the idea of updating the particle positions after the iterations; however, it requires a less complex model structure and is thus capable of being implemented with fewer hyper-parameters (Lambert-Torres et al. 2009).

2.3.7.2. PSO implementation. The PSO algorithm was separately employed for estimating seagrass AGB using the PySwarms library (Tisimst 2020) in the Python™ environment. The cost function in the PSO was defined using the best retrieval model for retrieving AGB in step 4. At the end of the optimization, the positions corresponding to the lowest root mean squared error (RMSE) was selected and tested for the model performance in scenario 5 (step 5).

2.3.7.3. PSO parameter tuning for AGB estimation. PSO consists of different hyper-parameters, which should be tuned to derive the best optimization. In this work, we used GridSearch in the PySwarms library (Tisimst 2020) to find the best combination of number of particles, w, c_1, and c_2 parameters (Table S4, the supplemental material). The search algorithm was fixed to global search.

2.3.8. Evaluation criteria

To evaluate the model performance for the binary classification, we used the accuracy (A), precision (P), recall (R), and F_1 scores to measure the performance (Equations (2) to (5)) whilst the McNemar test was used to check the statistical difference from the two ML models (the chi-square (χ^2) and p-value were produced using the Python library mlxtend (Raschka 2018)).

For seagrass AGB retrieval, we used the following metrics to evaluate the model performance: the coefficient of determination (R^2), root mean square error (RMSE), and RMSE percent of mean ($RMSE_{\%}$) (Equations (6) to (8)). Additionally, the Akaike Information Criteria (AIC) (Akaike 1974) and the Bayesian Information Criteria (BIC) (Schwarz 1978) were used to test the statistical difference among the selected models (Equations (9) to (10)). Lower values of AIC and BIC indicate a better performance of the model (Vrieze 2012).

$$A(y, \hat{y}) = \frac{1}{n_{samples}} \sum_{i=0}^{n_{samples}-1} (\hat{y}_i = y_i) \tag{2}$$

in which:

$\hat{y}_i$: predicted value

y_i: corresponding true value

$n_{samples}$: is the total number of validation samples

$$P = \frac{(TP)}{(TP) + (FP)} \tag{3}$$

$$R = \frac{(TP)}{(TP) + (FN)} \tag{4}$$

$$F_1 = 2 \times \frac{P \times R}{R + P} \tag{5}$$

in which:

TP: true positive

FP: false positive

FN: false negative

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{6}$$

in which:

and $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \varepsilon_i^2$

ε: the error term

n: is the total number of validation samples

$$\mathrm{RMSE}(y, \hat{y}) = \sqrt{\frac{1}{n_{\mathrm{samples}}} \sum_{i=0}^{n_{\mathrm{samples}}-1} (y_i - \hat{y}_i)^2} \tag{7}$$

$$\mathrm{RMSE}_{\%}(y, \hat{y}) = \sqrt{\frac{1}{n_{\mathrm{samples}}} \sum_{i=0}^{n_{\mathrm{samples}}-1} \left(\frac{y_i - \hat{y}_i}{y_i}\right)^2 \times 100} \tag{8}$$

in which:

$\hat{y}_i$: predicted value of the i samples

y_i: corresponding true value of the i samples

n_{samples}: is the total number of validation samples

$$\mathrm{AIC} = n \times \log\left(\frac{RSS}{n}\right) + 2 \times K \tag{9}$$

$$\mathrm{BIC} = n \times \log\left(\frac{RSS}{n}\right) + K \times \log(n) \tag{10}$$

in which:

RSS: residuals sum of squares

K: number of parameters (including intercept)

n: number of observations

2.3.9. Seagrass AGB mapping

The best performing model in step (4) or step (5) (Figure 3) was used to generate the distribution maps of the seagrass AGB parameters. The binary seagrass map produced in this study was used to mask out the non-seagrass area in the AGB map.

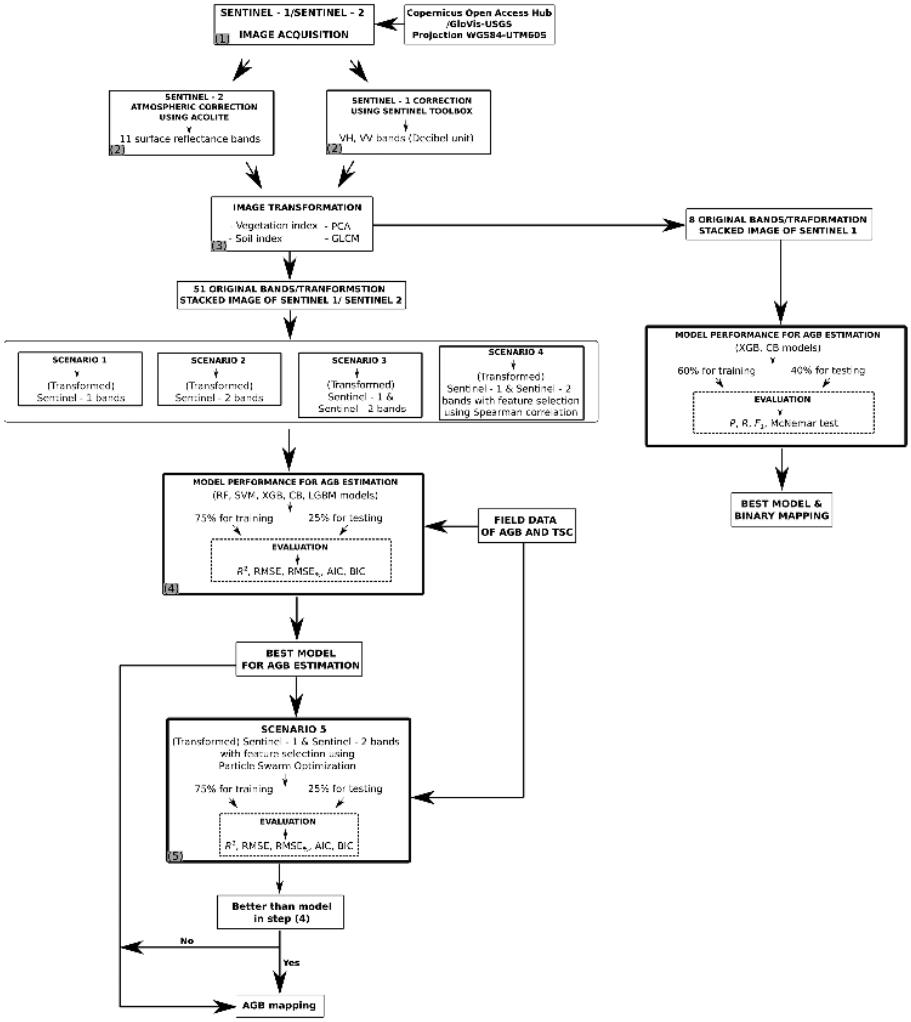


Figure 3. Flow chart of image processing, retrieval modelling (steps (1) to (5)) for seagrass AGB estimation and binary mapping.

3. Results

3.1. Seagrass binary mapping

The results indicate a high accuracy for seagrass binary mapping using the S-1 image data. The XGB obtained a higher metrics of P (0.82) and F_1 (0.86) comparing to the CB model (Table 4) and the accuracy is significantly difference from two ML models (McNemar test, Table 4).

Table 4. Model performance for seagrass binary mapping.

	A	*P*	*R*	F_1
XGB	0.92	0.82	0.90	0.86
CB	0.90	0.78	0.91	0.84
	McNemar test			
	χ^2 value	*p*-value		
XGB-CB	10.89	0.00		

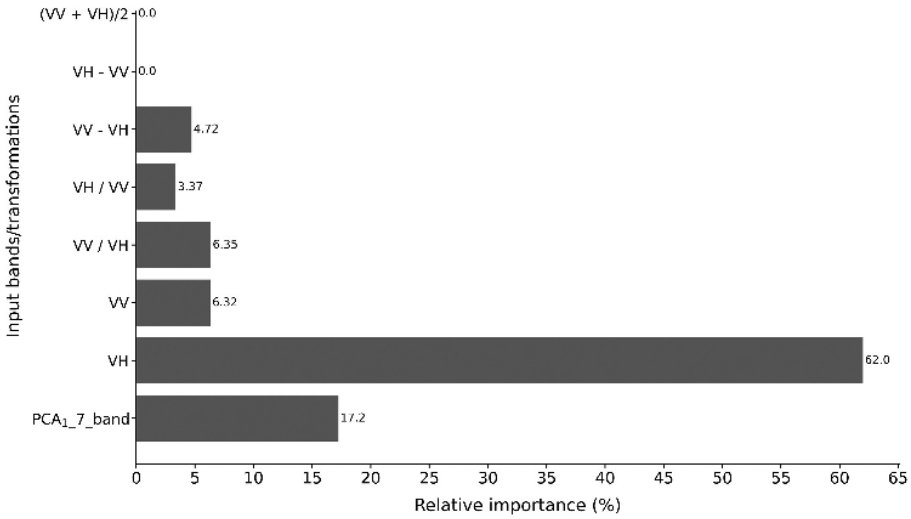


Figure 4. Influence of input variables for seagrass binary mapping.

The feature importance determines the contribution of the input variables to the seagrass binary mapping, in which the band VH is the most influential variable (62%), following by the PCA_1_7_band (17.2%), VV/VH, VV, and VV– VH, VH/VV (Figure 4). The transformations of VH– VV and (VV + VH)/2 have no impact on the seagrass binary mapping in this study.

Using the XGB model, we produced a seagrass binary map for Tauranga Harbour (Figure 5). The map indicates an estimated total area of approximately 1,100 ha, with continuous meadows in all parts of the harbour, and these are particularly abundant in the middle section.

3.2. Seagrass AGB estimation

During the field survey, the seagrass AGB varied over an order of magnitude from 23 to 237 g DW m^{-2}, with a mean of 69 g DW m^{-2}, standard deviation of 42 g DW m^{-2} and 99% of the field data ranged between 23 to 178 g DW m^{-2}. The result indicates a wide variation of seagrass AGB from plots collected in the harbour.

In scenario 1, when only S-1 bands were considered, the LGBM model produced the highest R^2 of 0.49 whilst the XGB model was the better model, however at a lower R^2 of 0.47 in scenario 2 (Table 5). The accuracy was improved for AGB estimation when combining S-1 and S-2 bands in

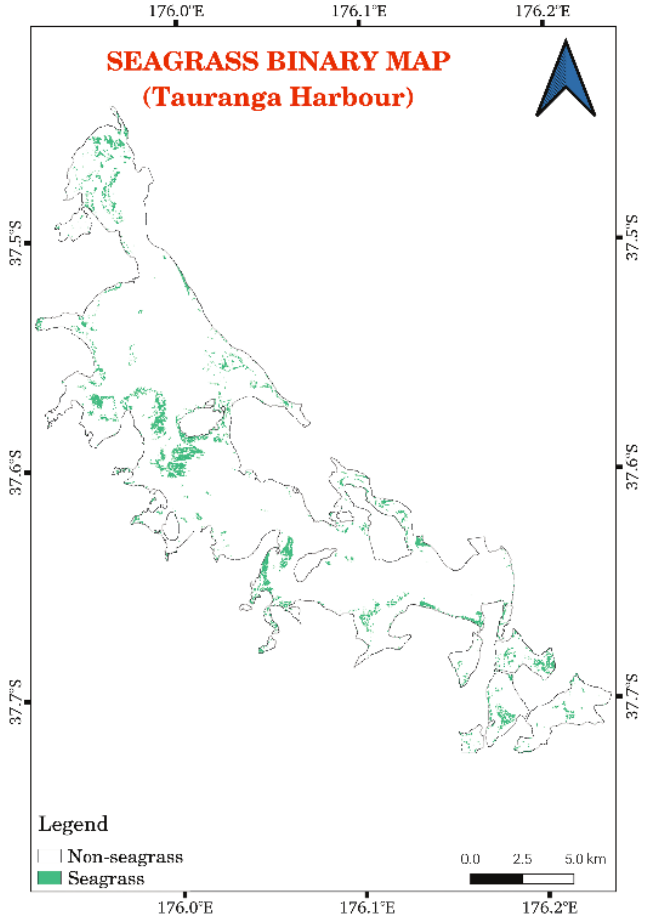


Figure 5. Seagrass binary distribution map derived from XGB model using S-1 image data.

scenario 3, with the LGBM model yielding an R^2 of 0.54 and the lowest values of AIC and BIC (Table 5).

In scenario 4, we tested the model performance with the feature selection using only bands that yielded a Spearman correlation coefficient > 0.10, which reduced the selected bands to 47. Using these, the XGB model derived an enhanced prediction of AGB at R^2 of 0.70 (Table 5). R^2, RMSE, AIC, BIC confirmed that XGB was superior to all other ML models. The model performances were visualized (Figure 6), and this confirmed that the XGB was a good fit with residuals smaller and more evenly spread than other models.

For Scenario 5, the best performing model from scenarios 1 to 4 (XGB model) was used as the cost function in the PSO algorithm, and 26 bands were selected from optical and SAR data fusion. The XGB model with PSO for feature selection (XGB-PSO) outperformed the Spearman correlation analysis used in Scenario 4, with a higher R^2 (0.75), and lower AIC (24.80), BIC (44.70), $RMSE_{\%}$ (41.69%). Overall, the XGB-PSO was the best model for seagrass AGB retrieval in Tauranga Harbour, with the highest R^2, lowest values of AIC, BIC, $RMSE_{\%}$ and a good agreement between

Table 5. Comparison of the ML models' performances for AGB retrieval in four scenarios[1].

SC	Scenario 1					Scenario 2					Scenario 3					Scenario 4				
Model	R^2	RMSE (g)	$RMSE_{\%}$ (%)	AIC	BIC	R^2	RMSE (g)	$RMSE_{\%}$ (%)	AIC	BIC	R^2	RMSE (g)	$RMSE_{\%}$ (%)	AIC	BIC	R^2	RMSE (g)	$RMSE_{\%}$ (%)	AIC	BIC
RF	0.34	0.55	63.57	40.40	61.00	0.21	0.48	68.11	22.00	37.60	0.24	0.50	79.64	82.20	118.30	0.52	0.46	55.87	71.30	104.50
SVM	0.04	0.70	82.70	47.50	68.00	0.00	0.54	71.88	25.70	41.30	0.01	0.57	88.00	85.60	121.70	0.32	0.56	45.60	76.70	109.90
XGB	0.23	0.60	69.07	42.90	63.40	**0.47**	**0.39**	**59.78**	**16.20**	**31.80**	0.12	0.54	79.85	83.80	119.90	**0.70**	**0.37**	**47.40**	**64.40**	**97.60**
CB	0.19	0.62	74.27	43.70	64.20	0.22	0.47	68.61	21.90	37.50	0.31	0.48	73.01	80.10	116.80	0.49	0.48	59.56	72.30	105.60
LGBM	**0.49**	**0.49**	**50.02**	**36.70**	**57.20**	0.26	0.46	60.75	21.10	36.70	**0.54**	**0.39**	**51.61**	**74.20**	**110.30**	0.53	0.46	55.89	71.10	104.40

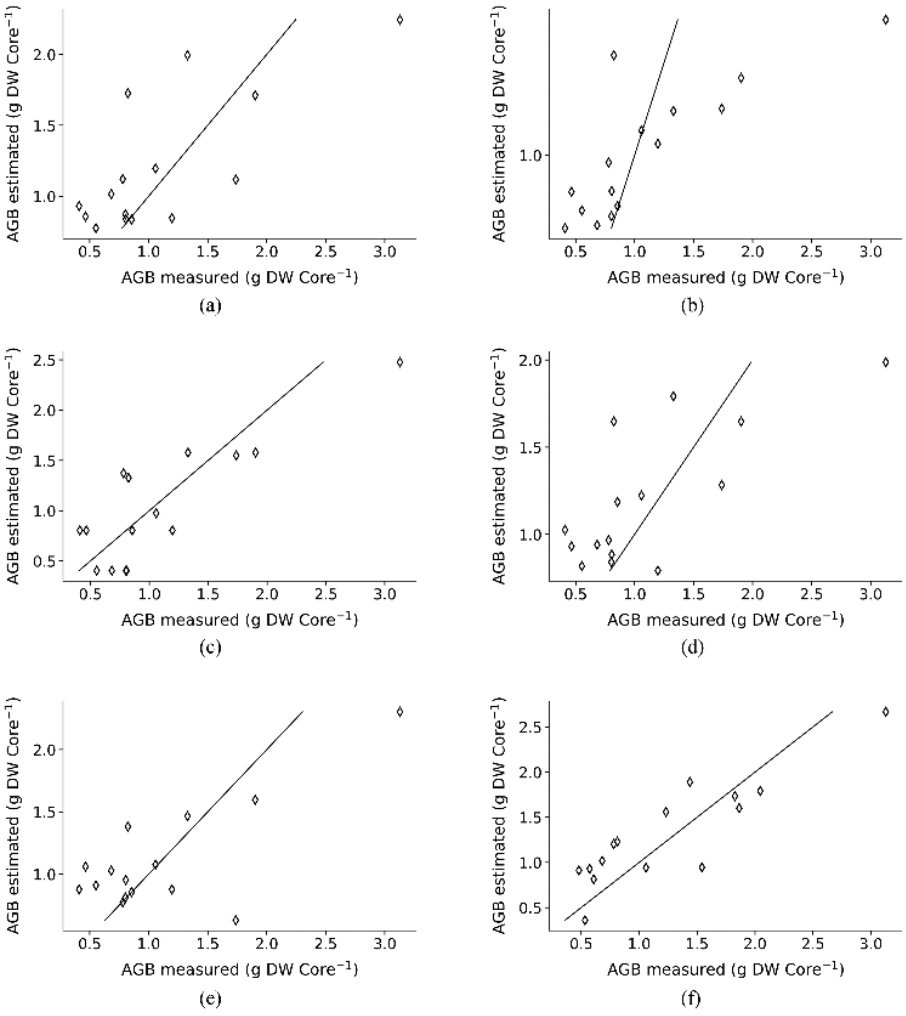


Figure 6. Scatter plot of AGB estimation use S-1 and S-2 bands with feature selection from Spearman correlation (a) – (e) and PSO (f): (a) RF, (b) SVM, (c): XGB, (d): CB, (e) LGBM, (f): XGB-PSO.

estimated and measured AGB values (Figure 6(f)). *The bands that emerged as of greatest significance in the various models confirm that similar numbers of bands computed from SAR and multispectral data contribute to the final XGB-PSO model* (Figure 7).

For seagrass AGB estimation, the most influential bands involve ρ_{Red}, IRECI of S-2 and VH_cons, VH_diss, VH_entropy, VV_corr of S-1 images. Among selected vegetation indices using the PSO algorithm, several new VIs computed from S-2 data such as MCARI, IRECI, EVI2 play a role in estimating seagrass AGB in the study area. On the other hand, other common VIs (NDVI, SAVI) are less important when retrieving AGB.

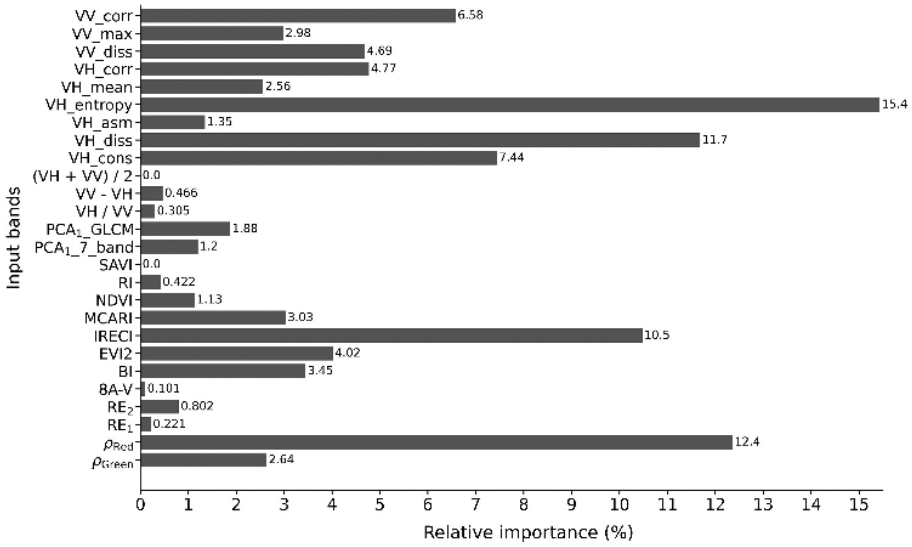


Figure 7. Variable importance for AGB estimation in Tauranga Harbour.

3.3. Seagrass AGB distribution map

We created an AGB distribution map for Tauranga Harbour using the XGB-PSO model (Figure 8). Predicted seagrass AGB ranged from 15 to 201 g DW m^{-2} and almost all seagrass meadows were estimated to have AGB from 61 to 139 g DW m^{-2}. We predict highest seagrass AGB values in the areas in the centre and south of the harbour.

4. Discussion

In this research, we evaluated novel approaches for estimation of seagrass properties using EO resources. We attempted to map the seagrass distribution using only SAR data, to offset constraints imposed on optical sensors by viewing conditions. We then went on to assess whether AGB could be best estimated using SAR, multispectral or a fusion of these remote sensing data, when combined with advanced ML models. To our knowledge, no other studies have applied EO data fusion techniques for such applications. In addition, we demonstrate how an advanced feature selection technique, using PSO metaheuristic optimization, improved the reliability of the retrievals to yield spatial AGB maps at 10 m resolution.

We were capable of mapping the distribution of intertidal seagrass meadows (present-absent) using eight input variables derived from S-1 data (two original bands and six transformations) with high accuracy (F_1 score 0.86). Our mapping is the first report on the use of SAR data for seagrass detection. The finding of the high contribution of the VH to the detection of vegetation is in agreement with the literature, where this polarization conveys information on stem scattering (Shang et al. 2020; Xu et al. 2019). Our results indicate that complex relationships across the SAR signals exist, with transformations and the PCA of all bands contributing 33% to the XGB model performance.

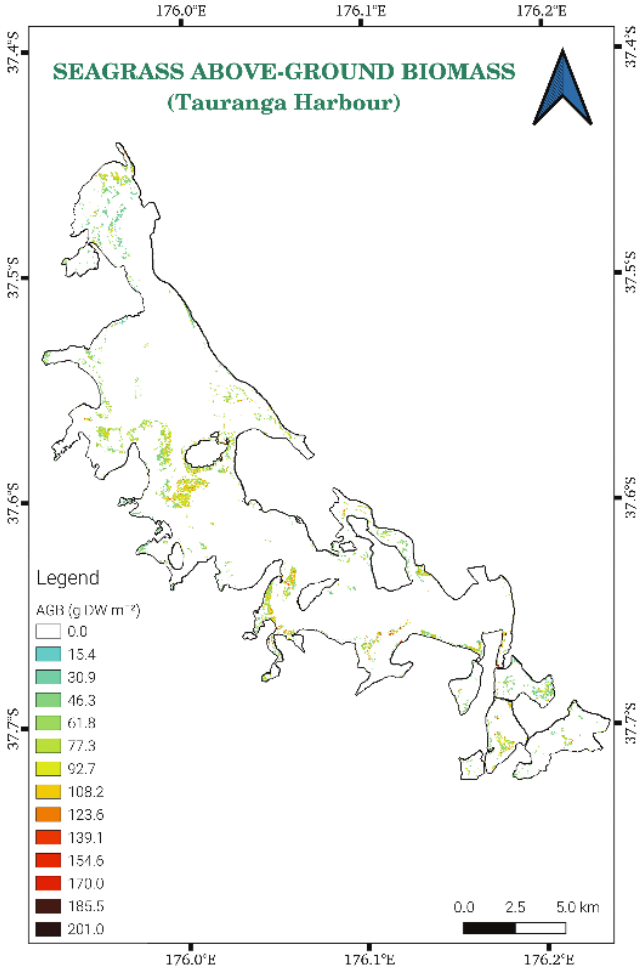


Figure 8. Seagrass AGB map for Tauranga Harbour created from the XGB-PSO model.

For seagrass AGB, the spatial map indicates high biomass values in the middle and southern parts of the harbour which coincides with the distribution of healthy seagrass meadows in a previous map (Ha et al. 2020). Compared to the task of binary/trinary classification, the quantification of AGB is more of a challenge due to the limited amount of training/testing data and the difference in the input data ranges (*i.e*: the limited number of output classes for classification compared to a wide range of output variables for regression against biomass). The accuracy of the regression is, therefore usually lower than the binary/trinary classification task as presented in this study.

Substantial performance differences were observed with the use of S-1, S-2 or both, and between the various ML techniques applied. Best performance consistently was associated with scenarios using both the S-1 and S-2 images, and overall the XGB-PSO was the most effective ML model. Vorster et al. (2020) discussed the variation of $RMSE_{\%}$ in the forest biomass estimation, in which satellite image-based prediction models

contribute 25% to 75% to the total uncertainty of the estimation (Vorster et al. 2020). Salum et al. (2020) reported $RMSE_{\%}$ of 20.66% to 43.81% for mangrove AGB estimation (R^2 ranged from 0.90 to 0.97) (Salum et al. 2020) whilst Li et al. (2020) indicated the values of $RMSE_{\%}$ of 38% to 54% for forest AGB retrieval (R^2 ranged from 0.38 to 0.75) (Li et al. 2020). The $RMSE_{\%}$ of XGB-PSO model (41%), therefore is rationale and acceptable when dealing with the small size and low AGB of the seagrass *Z. muelleri* in this study.

Feature selection resulted in higher accuracy of AGB estimation, which exceeded the accuracy of a simple linear regression model for predicting AGB for seagrasses from VHR optical images (Koedsin et al. 2016; Roelfsema et al. 2014). In other words, the high performance of the XGB model indicates that high to very high-resolution images might not be the only solution to return high accuracy seagrass AGB estimation. (Naidoo et al. 2019) and (Byrd et al. 2018) applied similar scenarios with and without the fusion of S-1 and S-2 images for predicting biomass of emergent wetland plants, and retrieved AGB at an R^2 ranging from 0.36 to 0.63, slightly less than those achieved in this study (R^2 = 0.75). Given the targets of the other studies had a higher biomass (5 to 3000 g m^{-2} in the wetland), and likely presented a more favourable radar target than emerged seagrass, our proposed models are promising and could be expanded to include not only the seagrass but also other Blue Carbon ecosystems in the coastal areas.

The integration of the metaheuristic optimization with remote sensing data fusion is another new approach to the estimation of seagrass biophysical parameters from remote sensing data. Our results show the value of including advanced feature selection tools to gain higher accuracy of seagrass AGB estimation. Compared to the results in scenarios 3 (no feature selection) and 4 (feature selection using Spearman correlation analysis), the model performance was improved substantially by the addition of the PSO (scenario 5). The PSO is fast proving a powerful tool in ecological models. Through the harmonization of the control parameters of w, c_1, and c_2 inside the swarm, the next best location is updated among the particles via each iteration. In addition, the memory function (*i.e*: the personal best feature in the algorithm) provides another solution for the particles to revert to the best way when they explore a 'wrong way' in the searching space. Two features of the PSO – the iterative updating of the next best solution within swarm space and the ability to revert when the algorithm explores a 'wrong way' in the searching space – usually result in a faster convergence of the swarm and better exploration of the target variable (seagrass AGB in this study) (Bansal et al. 2009; Elbeltagi, Hegazy, and Grierson 2005; Panda and Padhy 2008; Xiaohui et al. 2020). Since the effective application of the PSO for seagrass AGB estimation is first reported here, it is expected that our results will pave the way for advanced methods to be applied to other seagrass species in other international locations (Akbari et al. 2020; Sengupta, Basak, and Peters 2018; Wang, Tan, and Liu 2018).

We observed the most important variables for retrieval of seagrass AGB from S-2 are the single colour bands ρ_{Green} and ρ_{Red}, the vegetation indices EVI2, IRECI, MCARI, the soil index BI, and the GLCM bands VH_cons, VH_diss, VH_entropy, VV_corr of S-1 image (Figure 7). The ρ_{Green} and ρ_{Red} bands are likely sensitive to the pigment information, whilst the VIs and the SIs provide useful information on the density of the seagrass. These indices were derived to provide information on vegetation from reflectance data, and have been shown to be effective in other situations (Morcillo-Pallarés et al. 2019; Mulder et al. 2011; Xue and Baofeng 2017). The GLCM bands, on the other hand, have not previously been used for seagrass, but here appear to provide quantitative information

on the shape and texture of the meadows, and make a significant contribution to the final model. Whether the SAR data are truly addressing the texture of seagrass, or some other component of the seagrass ecosystem such as cockles which typically co-occur in Tauranga Harbour (Morrison et al. 2014), is not clear. Our results are, however, consistent with recent studies for mangrove AGB retrieval in the tropics which suggested that S-1 is appropriate for quantification of low biomass vegetation (Pham et al. 2020a, 2020c).

The method described here provides a novel direction using data fusion and state-of-the-art machine learning and optimized techniques in detecting the distribution and retrieving seagrass AGB in the coastal area at 10 m spatial resolution, and may be useful for other Blue Carbon ecosystem analyses. All satellite and analytical products used in this work are open access and present opportunities for low-cost analyses where suitable data to supervise classification and regression can be obtained. Our work could provide a baseline for blue carbon credit markets, towards a nature-based solution to climate change mitigation (Lovelock and Duarte 2019; Macreadie et al. 2019). The current study, however, is spatially narrowed to the intertidal zones where the seagrass meadows can be exposed at low tide and all the bands of S-1/S-2 are available for further processing. In addition, the data saturation for AGB estimation was reached at approximately 170 g DW m^{-2} which might be considered as the limitation when using S-1 data.

Future research might be to expand the proposed methods with more field surveying data on various seagrass species across different coastal regions at a larger-scale using different SAR sensors such as advanced land observing satellite 2 (ALOS-2) phased array type L band synthetic aperture radar (PALSAR-2). The L band SAR has a longer wavelength, penetrates deeper than the C band of S-1 image, and is potentially higher in data saturation (Pham et al. 2020a). In addition, various metaheuristic optimization algorithms (Harris Hawks Optimization (HHO) (Heidari et al. 2019), Firefly algorithm (FFA) (Pan, Xue, and Ruixiang 2019)) could be investigated for better feature selection to improve the accuracy of the retrieval models.

5. Conclusion

Our work pioneers the use of SAR remote sensing data (S-1 imagery) for mapping the distribution of intertidal seagrasses, and the fusion of optical remote sensing data (S-2 imagery) with advanced machine learning and metaheuristic optimization techniques (PSO) for estimating AGB of the small seagrass species *Z. muelleri* in the intertidal zone in Tauranga Harbour.

S-1 provides useful remote SAR image data for the binary mapping of seagrass with high accuracy (A 0.92 and F_1 0.86 when combined with the XGB model). For seagrass AGB, both optical and SAR sensing provided information to drive models and data fusion provided the best retrieval accuracy. In addition, feature selection improved the model accuracy for AGB estimation with a superior performance of the metaheuristic optimization PSO using the XGB model over the Spearman correlation.

Our results contribute a novel advance in the use of multispectral and SAR remote sensing data fusion and state-of-the-art ML models together with metaheuristic optimization, for both seagrass detection and seagrass AGB estimation in the intertidal zones that may be expandable to various seagrass species globally.

Note

1. The RMSE was calculated differently due to the random sample splitting from each scenarios. Therefore, the RMSE was used only to compare the model's performance inside the scenario whilst the R^2 was used to compare the model among the scenarios. Bold values indicate the best performance of the model.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Akaike, H. 1974. "A New Look at the Statistical Model Identification." *IEEE Transactions on Automatic Control* 19 (6): 716–723. doi:10.1109/TAC.1974.1100705.

Akbari, E., A. D. Boloorani, N. N. Samany, S. Hamzeh, S. Soufizadeh, and S. Pignatti. 2020. "Crop Mapping Using Random Forest and Particle Swarm Optimization Based on Multi-Temporal Sentinel-2." *Remote Sensing* 12 (9): 1449. Multidisciplinary Digital Publishing Institute. doi10.3390/rs12091449.

Aoki, L. R., K. J. McGlathery, P. L. Wiberg, and A. Al-Haj. 2020. "Depth Affects Seagrass Restoration Success and Resilience to Marine Heat Wave Disturbance." *Estuaries and Coasts* 43 (2): 316–328. Springer. doi10.1007/s12237-019-00685-0.

Bansal, S., D. Gupta, V. K. Panchal, and S. Kumar. 2009. "Swarm Intelligence Inspired Classifiers in Comparison with Fuzzy and Rough Classifiers: A Remote Sensing Approach."In *Contemporary Computing*, edited by S. Ranka, S. Aluru, R. Buyya, Y.-C. Chung, S. Dua, S. K. Ananth Grama, S. Gupta, R. Kumar, and V. V. Phoha, 284–294. Communications in Computer and Information Science. Berlin, Heidelberg: Springer.

Bedulli, C., P. S. Lavery, M. Harvey, C. M. Duarte, and O. Serrano. 2020. "Contribution of Seagrass Blue Carbon toward Carbon Neutral Policies in a Touristic and Environmentally-Friendly Island." *Frontiers in Marine Science* 7 (January): 1. doi:10.3389/fmars.2020.00001.

Belgiu, M., and L. Drăguţ. 2016. "Random Forest in Remote Sensing: A Review of Applications and Future Directions." *ISPRS Journal of Photogrammetry and Remote Sensing* 114 (April): 24–31. doi:10.1016/j.isprsjprs.2016.01.011.

Breiman, L. 2001. "Random Forest." *Machine Learning* 45 (1): 5–32. doi:10.1023/A:1010933404324.

Byrd, K. B., L. Ballanti, N. Thomas, D. Dung Nguyen, J. R. Holmquist, M. Simard, and L. Windham-Myers. 2018. "A Remote Sensing-Based Model of Tidal Marsh Aboveground Carbon Stocks for the Conterminous United States." *ISPRS Journal of Photogrammetry and Remote Sensing* 139 (May): 255–271. doi:10.1016/j.isprsjprs.2018.03.019.

Chen, T., and C. Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, 785–794. San Francisco, California, USA: ACM Press. doi:10.1145/2939672.2939785.

Collier, C. J., C. Villacorta-Rath, K.-J. Van Dijk, M. Takahashi, and M. Waycott. 2014. "Seagrass Proliferation Precedes Mortality during Hypo-Salinity Events: A Stress-Induced Morphometric Response." Edited by J. F. Valentine. *PLoS ONE* 9 (4): e94014. doi:10.1371/journal.pone.0094014.

Collier, C. J., Y. X. Ow, L. Langlois, S. Uthicke, C. L. Johansson, K. R. O'Brien, V. Hrebien, and M. P. Adams. 2017. "Optimum Temperatures for Net Primary Productivity of Three Tropical Seagrass Species." *Frontiers in Plant Science* 8 (August). doi:10.3389/fpls.2017.01446.

Daughtry, C. S. T. 2000. "Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance." *Remote Sensing of Environment* 74 (2): 229–239. Elsevier Science Inc. doi10.1016/S0034-4257(00)00113-9.

De Almeida, S. G., Bruno, and V. C. Leite. 2019. "Particle Swarm Optimization: A Powerful Technique for Solving Engineering Problems." In *Swarm Intelligence - Recent Advances, New Perspectives and Applications*. IntechOpen. doi:10.5772/intechopen.89633.

Delegido, J., J. Verrelst, L. Alonso, and J. Moreno. 2011. "Evaluation of Sentinel-2 Red-Edge Bands for Empirical Estimation of Green LAI and Chlorophyll Content." *Sensors* 11 (7): 7063–7081. Multidisciplinary Digital Publishing Institute (MDPI). doi10.3390/s110707063.

El Hajj, M., N. B. Baghdadi, H. Bazzi, and M. Zribi. 2018. "Penetration Analysis of SAR Signals in the C and L Bands for Wheat, Maize, and Grasslands." *Remote Sensing* 11 (1): 31. MDPI AG. doi10.3390/rs11010031.

Elbeltagi, E., T. Hegazy, and D. Grierson. 2005. "Comparison among Five Evolutionary-Based Optimization Algorithms." *Advanced Engineering Informatics* 19 (1): 43–53. doi:10.1016/j.aei.2005.01.004.

ESA. 2020a. "SNAP V7.0." http://step.esa.int

ESA. 2020b. "Copernicus Open Access Hub." Accessed February 2. https://scihub.copernicus.eu/dhus/#/home

ESA-S1. 2020. "User Guides - Sentinel-1 SAR." https://sentinel.esa.int/web/sentinel/user-guides/sentinel-1-sar

ESA-S2. 2020. "User Guides - Sentinel-2 MSI." https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi

Escadafal, R. 1989. "Remote Sensing of Arid Soil Surface Color with Landsat Thematic Mapper." *Advances in Space Research* 9 (1): 159–163. Pergamon. doi10.1016/0273-1177(89)90481-X.

Fehr, J., K. Z. Arreola, and H. Burkhardt. 2008. "Fast Support Vector Machine Classification of Very Large Datasets." *Data Analysis, Machine Learning and Applications*. edited by C. Preisach, H. Burkhardt, L. Schmidt-Thieme, and R. Decker, 11–18. Berlin, Heidelberg: Springer Berlin Heidelberg. http://link.springer.com/10.1007/978-3-540-78246-9_2https://lmb.informatik.uni-freiburg.de/papers/download/fe_za_bu_GFKL07.pdf

Filipponi, F. 2019. "Sentinel-1 GRD Preprocessing Workflow." *Proceedings* 18 (1): 11. doi:10.3390/ECRS-3-06201.

Fonseca, M. S., and J. A. Cahalan. 1992. "A Preliminary Evaluation of Wave Attenuation by Four Species of Seagrass." *Estuarine, Coastal and Shelf Science* 35 (6): 565–576. Academic Press. doi10.1016/S0272-7714(05)80039-3.

Frampton, W. J., J. Dash, G. Watmough, and E. J. Milton. 2013. "Evaluating the Capabilities of Sentinel-2 for Quantitative Estimation of Biophysical Variables in Vegetation." *ISPRS Journal of Photogrammetry and Remote Sensing* 82 (August): 83–92. Elsevier B.V. doi10.1016/j.isprsjprs.2013.04.007.

Gholami, R., and N. Fakhari. 2017. "Support Vector Machine: Principles, Parameters, and Applications." In *Handbook of Neural Computation*, 515–535. Elsevier. https://linkinghub.elsevier.com/retrieve/pii/B9780128113189000272

Gitelson, A. A., Y. J. Kaufman, and M. N. Merzlyak. 1996. "Use of a Green Channel in Remote Sensing of Global Vegetation from EOS-MODIS." *Remote Sensing of Environment* 58 (3): 289–298. Elsevier. doi10.1016/S0034-4257(96)00072-7.

Githaiga, M. N., A. M. Frouws, J. G. Kairo, and M. Huxham. 2019. "Seagrass Removal Leads to Rapid Changes in Fauna and Loss of Carbon." *Frontiers in Ecology and Evolution* 7 (March): 62. doi:10.3389/fevo.2019.00062.

Ha, N. T., M. Manley-Harris, T. D. Pham, and I. Hawes. 2020. "A Comparative Assessment of Ensemble-Based Machine Learning and Maximum Likelihood Methods for Mapping Seagrass Using Sentinel-2 Imagery in Tauranga Harbor, New Zealand." *Remote Sensing* 12 (3): 355. doi:10.3390/rs12030355.

Heck Hay, K. L., G. Hays, and R. J. Orth. 2003. "Critical Evaluation of the Nursery Role Hypothesis for Seagrass Meadows." *Marine Ecology Progress Series* 253 (May): 123–136. Inter-Research. doi10.3354/meps253123.

Heidari, A. A., S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen. 2019. "Harris Hawks Optimization: Algorithm and Applications." *Future Generation Computer Systems* 97 (August): 849–872. doi:10.1016/j.future.2019.02.028.

Hossain, M. S., J. S. Bujang, M. H. Zakaria, and M. Hashim. 2015. "The Application of Remote Sensing to Seagrass Ecosystems: An Overview and Future Research Prospects." *International Journal of Remote Sensing* 36 (1): 61–114. doi:10.1080/01431161.2014.990649.

Howell, S. E. L., A. S. Komarov, M. Dabboor, B. Montpetit, M. Michael Brady, R. K. Scharien, M. S. Mahmud, V. Nandan, A. T. Geldsetzer, and J. J. Yackel. 2018. "Comparing L- and C-Band Synthetic Aperture Radar Estimates of Sea Ice Motion over Different Ice Regimes." *Remote Sensing of Environment* 204 (January): 380–391. Elsevier Inc. doi10.1016/j.rse.2017.10.017.

Huete, A. R. 1988. "A Soil-Adjusted Vegetation Index (SAVI)." *Remote Sensing of Environment* 25 (3): 295–309. Elsevier. doi10.1016/0034-4257(88)90106-X.

Huettmann, F. 2018. "Boosting, Bagging and Ensembles in the Real World: An Overview, Some Explanations and a Practical Synthesis for Holistic Global Wildlife Conservation Applications Based on Machine Learning with Decision Trees." In *Machine Learning for Ecology and Sustainable Natural Resource Management*, 63–83. Springer, Cham; Springer International Publishing. doi:10.1007/978-3-319-96978-7_3.

Jiang, Z., A. R. Huete, K. Didan, and T. Miura. 2008. "Development of a Two-Band Enhanced Vegetation Index without a Blue Band." *Remote Sensing of Environment* 112 (10): 3833–3845. Elsevier. doi10.1016/j.rse.2008.06.006.

Ke, G., Q. Meng, T. Finley, T. Wang, W. Chen, M. Weidong, Y. Qiwei, and T. Y. Liu. 2017. "LightGBM: A Highly Efficient Gradient Boosting Decision Tree." In *Advances in Neural Information Processing Systems*, 2017-Decem: Vancouver, Canada. 3147–3155. https://www.microsoft.com/en-us/research/publication/lightgbm-a-highly-efficient-gradient-boosting-decision-tree/

Kennedy, J., and R. Eberhart. 1995. "Particle Swarm Optimization." In *Proceedings of ICNN'95 - International Conference on Neural Networks*, Perth, Western Australia, Australia, 4:1942–1948 doi:10.1109/ICNN.1995.488968.

Koedsin, W., W. Intararuang, R. Ritchie, and A. Huete. 2016. "An Integrated Field and Remote Sensing Method for Mapping Seagrass Species, Cover, and Biomass in Southern Thailand." *Remote Sensing* 8 (4): 292. doi:10.3390/rs8040292.

Lambert-Torres, G., H. G. Martins, M. P. Coutinho, C. P. Salomon, F. M. Matsunaga, and R. A. Carminati. 2009. "Comparison between PSO and GA in System Restoration Solution." In *2009 15th International Conference on Intelligent System Applications to Power Systems, ISAP '09*. Curitiba, Brazil. doi:10.1109/ISAP.2009.5352885.

Li, F., L. Zhang, B. Chen, D. Gao, Y. Cheng, X. Zhang, Y. Yang, K. Gao, Z. Huang, and J. Peng. 2018. "A Light Gradient Boosting Machine for Remainning Useful Life Estimation of Aircraft Engines." In *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, 2018-November:3562–3567. Institute of Electrical and Electronics Engineers, Maui, Hawaii, USA. doi:10.1109/ITSC.2018.8569801.

Li, Y., M. Li, C. Li, and Z. Liu. 2020. "Forest Aboveground Biomass Estimation Using Landsat 8 and Sentinel-1A Data with Machine Learning Algorithms." *Scientific Reports* 10 (1): 9952. Nature Publishing Group. doi10.1038/s41598-020-67024-3.

Lovelock, C. E., and C. M. Duarte. 2019. "Dimensions of Blue Carbon and Emerging Perspectives." *Biology Letters* 15 (3): 20180781. Royal Society Publishing. doi10.1098/rsbl.2018.0781.

Macreadie, P. I., J. A. Andrea Anton, J. A. Raven, N. Beaumont, R. M. Connolly, D. A. Friess, J. J. Kelleway, et al. 2019. "The Future of Blue Carbon Science." *Nature Communications* 10 (1): 3998. doi:10.1038/s41467-019-11693-w.

Mathieu, R., M. Pouget, B. Cervelle, and R. Escadafal. 1998. "Relationships between Satellite-Based Radiometric Indices Simulated Using Laboratory Reflectance Data and Typic Soil Color of an Arid Environment." *Remote Sensing of Environment* 66 (1): 17–28. Elsevier Science Inc. doi10.1016/S0034-4257(98)00030-3.

Misbari, S., and M. Hashim. 2016. "Change Detection of Submerged Seagrass Biomass in Shallow Coastal Water." *Remote Sensing* 8 (3): 200. doi:10.3390/rs8030200.

Morcillo-Pallarés, P., J. P. Rivera-Caicedo, S. Belda, C. De Grave, H. Burriel, J. Moreno, and J. Verrelst. 2019. "Quantifying the Robustness of Vegetation Indices through Global Sensitivity Analysis of Homogeneous and Forest Leaf-Canopy Radiative Transfer Models." *Remote Sensing* 11 (20): 2418. Multidisciplinary Digital Publishing Institute. doi10.3390/rs11202418.

Morrison, M., M. L. Lowe, C. M. Grant, P. J. Smith, G. D. Carbines, J. Reed, S. Bury, and J. Brown, "New Zealand, and Ministry for Primary Industries." 2014. *Seagrass Meadows as Biodiversity*

and Productivity Hotspots. http://fs.fish.govt.nz/Doc/23703/AEBR_137_2701_ZBD2004-08%20Objective%201-4;%20Milestones%208,9,10,17,18,23,24,%2025,28,29%20and%2030.pdf.ashx

Mountrakis, G., I. Jungho, and C. Ogole. 2011. "Support Vector Machines in Remote Sensing: A Review." *ISPRS Journal of Photogrammetry and Remote Sensing* 66 (3): 247–259. doi:10.1016/j.isprsjprs.2010.11.001.

Mtwana Nordlund, L., E. W. Koch, E. B. Barbier, and J. C. Creed. 2016. "Seagrass Ecosystem Services and Their Variability across Genera and Geographical Regions." Edited by K. O. Reinhart. *Plos One* 11 (10): e0163091–e0163091. doi:10.1371/journal.pone.0163091.

Mulder, V. L., S. De Bruin, M. E. Schaepman, and T. R. Mayr. 2011. "The Use of Remote Sensing in Soil and Terrain Mapping — A Review." *Geoderma* 162 (1–2): 1–19. doi:10.1016/j.geoderma.2010.12.018.

Naidoo, L., H. Van Deventer, A. Ramoelo, R. Mathieu, B. Nondlazi, and R. Gangat. 2019. "Estimating above Ground Biomass as an Indicator of Carbon Storage in Vegetated Wetlands of the Grassland Biome of South Africa." *International Journal of Applied Earth Observation and Geoinformation* 78 (June): 118–129. doi:10.1016/j.jag.2019.01.021.

Natekin, A., and A. Knoll. 2013. "Gradient Boosting Machines, a Tutorial." *Frontiers in Neurorobotics* 7 (DEC): 21. doi:10.3389/fnbot.2013.00021.

Pal, M. 2005. "Random Forest Classifier for Remote Sensing Classification." *International Journal of Remote Sensing* 26 (1): 217–222. doi:10.1080/01431160412331269698.

Pan, X., L. Xue, and L. Ruixiang. 2019. "A New and Efficient Firefly Algorithm for Numerical Optimization Problems." *Neural Computing & Applications* 31 (5): 1445–1453. doi:10.1007/s00521-018-3449-6.

Panda, S., and N. P. Padhy. 2008. "Comparison of Particle Swarm Optimization and Genetic Algorithm for FACTS-Based Controller Design." *Applied Soft Computing, Soft Computing for Dynamic Data Mining* 8 (4): 1418–1427. doi:10.1016/j.asoc.2007.10.009.

Park, S. G. 1999. "*Changes in Abundance of Seagrass (Zostera* Spp.) *In Tauranga Harbour from 1959-96*." Environment Report 99/30. Whakatane, New Zealand. https://cdn.boprc.govt.nz/media/362713/changes-in-abundance-of-seagrass-zostera-spp-in-tauranga-harbour-from-1959-96.pdf

Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel et al. 2011. "Scikit-Learn: Machine Learning in Python." *Journal of Machine Learning Research* 12:2825–2830.

Pham, T. D., N. Yokoya, T. T. T. Nguyen, N. N. Le, N. T. Ha, J. Xia, W. Takeuchi, and T. D. Pham. 2020b. "Improvement of Mangrove Soil Carbon Stocks Estimation in North Vietnam Using Sentinel-2 Data and Machine Learning Approach." GIScience & Remote Sensing. 1–20. Taylor & Francis. doi10.1080/15481603.2020

Pham, T. D., J. Xia, N. T. Ha, D. T. Bui, N. N. Le, and W. Tekeuchi. 2019. "A Review of Remote Sensing Approaches for Monitoring Blue Carbon Ecosystems: Mangroves, Seagrassesand Salt Marshes during 2010–2018." *Sensors* 19 (8): 1933. Multidisciplinary Digital Publishing Institute. doi10.3390/s19081933.

Pham, T. D., N. Yokoya, J. Xia, N. T. Ha, N. N. Le, T. T. T. Nguyen, T. H. Dao, T. T. P. Vu, T. D. Pham, and W. Takeuchi. 2020c. "Comparison of Machine Learning Methods for Estimating Mangrove Above-Ground Biomass Using Multiple Source Remote Sensing Data in the Red River Delta Biosphere Reserve, Vietnam." *Remote Sensing* 12 (8): 1334. doi:10.3390/rs12081334.

Pham, T. D., N. Yokoya, T. T. T. Nguyen, N. N. Le, N. T. Ha, J. Xia, W. Takeuchi, and T. D. Pham. 2020b. "Improvement of Mangrove Soil Carbon Stocks Estimation in North Vietnam Using Sentinel-2 Data and Machine Learning Approach." *GIScience & Remote Sensing*. 1–20. Taylor & Francis. doi10.1080/15481603.2020.1857623.

Pham, T. D., N. N. Le, N. T. Ha, L. V. Nguyen, J. Xia, N. Yokoya, T. T. To, H. X. Trinh, L. Q. Kieu, and W. Takeuchi. 2020a. "Estimating Mangrove Above-Ground Biomass Using Extreme Gradient Boosting Decision Trees Algorithm with Fused Sentinel-2 and ALOS-2 PALSAR-2 Data in Can Gio Biosphere Reserve, Vietnam." *Remote Sensing* 12 (5): 777. doi:10.3390/rs12050777.

Pham, Tien Dat, Naoto Yokoya, Junshi Xia, Nam Thang Ha, Nga Nhu Le, Thi Thu Trang Nguyen, Thi Huong Dao, Thuy Thi Phuong Vu, Tien Duc Pham, and Wataru Takeuchi. 2020c. "Comparison of

Machine Learning Methods for Estimating Mangrove Above-Ground Biomass Using Multiple Source Remote Sensing Data in the Red River Delta Biosphere Reserve, Vietnam." Remote Sensing 12 (8). MDPI AG: 1334. doi:10.3390/RS12081334

Prokhorenkova, L., G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin. 2019. "CatBoost: Unbiased Boosting with Categorical Features." *ArXiv:1706.09516 [Cs]*, January. http://arxiv.org/abs/1706.09516

Pu, R., and S. Bell. 2013. "A Protocol for Improving Mapping and Assessing of Seagrass Abundance along the West Central Coast of Florida Using Landsat TM and EO-1 ALI/Hyperion Images." *ISPRS Journal of Photogrammetry and Remote Sensing* 83: 116–129. doi:10.1016/j.isprsjprs.2013.06.008.

Rainforth, T., and F. Wood. 2015. "Canonical Correlation Forests." *ArXiv:1507.05444 [Cs, Stat]*, July. http://arxiv.org/abs/1507.05444

Raschka, S. 2018. "MLxtend: Providing Machine Learning and Data Science Utilities and Extensions to Python's Scientific Computing Stack." *Journal of Open Source Software* 3 (24): 638. doi:10.21105/joss.00638.

Reeve, G., S. A. Stephens, and A. Wadhwa. 2018. "Tauranga Harbour Inundation Modelling." NIWA Client Report 2018269HN to Bay of Plenty Regional Council 2018269HN. Tauranga, New Zealand: NIWA. https://atlas.boprc.govt.nz/api/v1/edms/document/A3338785/content

Reidenbach, M. A., and E. L. Thomas. 2018. "Influence of the Seagrass, Zostera Marina, on Wave Attenuation and Bed Shear Stress within a Shallow Coastal Bay." *Frontiers in Marine Science* 5 (OCT): 397. Frontiers Media S.A. doi10.3389/fmars.2018.00397.

Ricart, A. M., P. H. York, C. V. Bryant, M. A. Rasheed, D. Ierodiaconou, and P. I. Macreadie. 2020. "High Variability of Blue Carbon Storage in Seagrass Meadows at the Estuary Scale." *Scientific Reports* 10 (1): 1–12. Nature Research. doi10.1038/s41598-020-62639-y.

Rodriguez, J. J., L. I. Kuncheva, and C. J. Alonso. 2006. "Rotation Forest: A New Classifier Ensemble Method." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (10): 1619–1630. doi:10.1109/TPAMI.2006.211.

Roelfsema, C., M. Lyons, M. Dunbabin, E. M. Kovacs, and S. Phinn. 2015. "Integrating Field Survey Data with Satellite Image Data to Improve Shallow Water Seagrass Maps: The Role of AUV and Snorkeller Surveys?" *Remote Sensing Letters* 6 (2): 135–144. doi:10.1080/2150704X.2015.1013643.

Roelfsema, C. M., M. Lyons, E. M. Kovacs, P. Maxwell, M. I. Saunders, J. Samper-Villarreal, and S. R. Phinn. 2014. "Multi-Temporal Mapping of Seagrass Cover, Species and Biomass: A Semi-Automated Object Based Image Analysis Approach." *Remote Sensing of Environment* 150: 172–187. doi:10.1016/j.rse.2014.05.001.

Rouse, J. W., Jr., R. H. Haas, J. A. Schell, and D. W. Deering. 1974. "Monitoring Vegetation Systems in the Great Plains with ERTS." In *NASA. Goddard Space Flight Center 3d ERTS-1 Symp., Vol. 1, Sect. A*. https://ntrs.nasa.gov/citations/19740022614

Salum, R. B., P. W. M. Souza-Filho, M. Simard, C. A. Silva, M. E. B. Fernandes, M. F. Cougo, W. Do Nascimento, and K. Rogers. 2020. "Improving Mangrove Above-Ground Biomass Estimates Using LiDAR." *Estuarine, Coastal and Shelf Science* 236 (May): 106585. doi:10.1016/j.ecss.2020.106585.

Sani, D. A., M. Hashim, and M. S. Hossain. 2019. "Recent Advancement on Estimation of Blue Carbon Biomass Using Satellite-Based Approach." *International Journal of Remote Sensing* 40 (20): 7679–7715. doi:10.1080/01431161.2019.1601289.

Schwarz, G. 1978. "Estimating the Dimension of a Model." *The Annals of Statistics* 6 (2): 461–464. doi:10.1214/aos/1176344136.

Sengupta, S., S. Basak, and R. Peters. 2018. "Particle Swarm Optimization: A Survey of Historical and Recent Developments with Hybridization Perspectives." *Machine Learning and Knowledge Extraction* 1 (1): 157–191. doi:10.3390/make1010010.

Shang, J., J. Liu, V. Poncos, X. Geng, B. Qian, Q. Chen, T. Dong, et al. 2020. "Detection of Crop Seeding and Harvest through Analysis of Time-Series Sentinel-1 Interferometric SAR Data." *Remote Sensing* Multidisciplinary Digital Publishing Institute 1210:1551. doi: 10.3390/rs12101551.

Short, F. T., and C. A. Short. 1984. "The Seagrass Filter: Purification of Estuarine and Coastal Waters" *The Estuary as a Filter* 395–413. Elsevier. doi: 10.1016/b978-0-12-405070-9.50024-4.

Thessen, A. E. 2016. "Adoption of Machine Learning Techniques in Ecology and Earth Science." *One Ecosystem* 1: e8621. doi:10.3897/oneeco.1.e8621.

Tisimst. 2020. "Pyswarm." https://github.com/tisimst/pyswarm

Tucker, C. J. 1979. "Red and Photographic Infrared Linear Combinations for Monitoring Vegetation." *Remote Sensing of Environment* 8 (2): 127–150. Elsevier. doi10.1016/0034-4257(79)90013-0.

Turner, S. J. 2007. "Growth and Productivity of Intertidal Zostera Capricorni in New Zealand Estuaries." *New Zealand Journal of Marine and Freshwater Research* 41 (1): 77–90. doi:10.1080/00288330709509897.

Unsworth, R. K. F., L. J. McKenzie, C. J. Collier, L. C. Cullen-Unsworth, C. M. Duarte, J. S. Eklöf, J. C. Jarvis, B. L. Jones, and L. M. Nordlund. 2018. "Global Challenges for Seagrass Conservation." *Ambio*. November. doi: 10.1007/s13280-018-1115-y.

USGS. 2020. "GloVis - Home." Accessed February 2. https://glovis.usgs.gov/

Vanhellemont, Q. 2019. "Adaptation of the Dark Spectrum Fitting Atmospheric Correction for Aquatic Applications of the Landsat and Sentinel-2 Archives." *Remote Sensing of Environment* 225 (May): 175–192. doi:10.1016/j.rse.2019.03.010.

Vanhellemont, R. K. 2016. "ACOLITE for Sentinel-2: Aquatic Applications of MSI Imagery." In *Proceedings of the 2016 ESA Living Planet Symposium*, 8. Prague, Czech Republic: ESA Special Publication. https://odnature.naturalsciences.be/downloads/publications/vanhellemontruddick_esa_lps2016_coastalapplications_final_header.pdf

Veettil, B. K., R. D. Ward, M. D. A. C. Lima, M. Stankovic, P. N. Hoai, and N. X. Quang. 2020. "Opportunities for Seagrass Research Derived from Remote Sensing: A Review of Current Methods." *Ecological Indicators* 117. Elsevier B.V. doi:10.1016/j.ecolind.2020.106560.

Vorster, A. G., P. H. Evangelista, A. E. L. Stovall, and E. Seth. 2020. "Variability and Uncertainty in Forest Biomass Estimates from the Tree to Landscape Scale: The Role of Allometric Equations." *Carbon Balance and Management* 15 (1): 8. doi:10.1186/s13021-020-00143-6.

Vrieze, S. I. 2012. "Model Selection and Psychological Theory: A Discussion of the Differences between the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC)." *Psychological Methods* 17 (2): 228–243. doi:10.1037/a0027127.

Wang, D., D. Tan, and L. Liu. 2018. "Particle Swarm Optimization Algorithm: An Overview." *Soft Computing* 22 (2): 387–408. doi:10.1007/s00500-016-2474-6.

Waycott, M., C. M. Duarte, T. J. B. Carruthers, R. J. Orth, W. C. Dennison, S. Olyarnik, A. Calladine, et al. 2009. "Accelerating Loss of Seagrasses across the Globe Threatens Coastal Ecosystems." *Proceedings of the National Academy of Sciences* 106 (30): 12377–12381. doi:10.1073/pnas.0905620106.

Xiaohui, D., L. Huapeng, L. Yong, J. Yang, and Z. Shuqing. 2020. "Comparison of Swarm Intelligence Algorithms for Optimized Band Selection of Hyperspectral Remote Sensing Image." *Open Geosciences* 12 (1): 425–442. De Gruyter. doi10.1515/geo-2020-0155.

Xu, L., H. Zhang, C. Wang, B. Zhang, and M. Liu. 2019. "Crop Classification Based on Temporal Information Using Sentinel-1 SAR Time-Series Data." *Remote Sensing* 11 (1): 53. Multidisciplinary Digital Publishing Institute. doi10.3390/rs11010053.

Xue, J., and S. Baofeng. 2017. "Significant Remote Sensing Vegetation Indices: A Review of Developments and Applications." *Journal of Sensors* 2017. doi:10.1155/2017/1353691.

York, P. H., R. K. Gruber, R. Hill, P. J. Ralph, D. J. Booth, and P. I. Macreadie. 2013. "Physiological and Morphological Responses of the Temperate Seagrass Zostera Muelleri to Multiple Stressors: Investigating the Interactive Effects of Light and Temperature." Edited by M. Chapman. *Plos One* 8 (10): e76377. doi:10.1371/journal.pone.0076377.

Appendix 7. Co-authorship form of research chapter 6

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

Co-Authorship Form

Postgraduate Studies Office
Student and Academic Services Division
Wahanga Ratonga Matauranga Akonga
The University of Waikato
Private Bag 3105
Hamilton 3240, New Zealand
Phone +64 7 838 4439
Website: http://www.waikato.ac.nz/sasd/postgraduate/

This form is to accompany the submission of any PhD that contains research reported in published or unpublished co-authored work. **Please include one copy of this form for each co-authored work.** Completed forms should be included in your appendices for all the copies of your thesis submitted for examination and library deposit (including digital deposit).

Please indicate the chapter/section/pages of this thesis that are extracted from a co-authored work and give the title and publication details or details of submission of the co-authored work. Chapter 6: A novel and open source web-GIS approach for seagrass da-ta visualization and collaboration	
Nature of contribution by PhD candidate	Conceptualization, methodology, software, formal analysis, writing-original draft preparation, resources, write-review and editing.
Extent of contribution by PhD candidate (%)	85

CO-AUTHORS

Name	Nature of Contribution
Merilyn Manley-Harris	Resources, writing-review and editing.
Ian Hawes	Conceptualization, resources, writing-review and editing, supervision.

Certification by Co-Authors

The undersigned hereby certify that:

- the above statement correctly reflects the nature and extent of the PhD candidate's contribution to this work, and the nature of the contribution of each of the co-authors; and

Name	Signature	Date
Merilyn Manley-Harris		26.2.21.
Ian Hawes	Ian Hawes	1 March 2021

July 2015

Appendix 8. The contribution of machine learning (ML) hyper-parameters to model performance

	Shareable hyper-parameters among ML model	Contribution to ML model performance
	Depth (or max depth)	Higher values might increase the complexity of the decision tree and result in overfitting during the prediction.
	Learning rate	Lower values result in longer training time. Decrease the learning rate in case of detected overfitting.
	Number of trees	The number of decision trees in the model. Higher number of trees might produce better accuracy prediction.
	Boosting type	Define the boosting strategy during the learning of the model. Selection of boosting methods might have different impact on the prediction accuracy and are user-dependent experience.
ML model	Unique hyper-parameters of ML model	Contribution to ML model performance
CatBoost	L_2 leaf regression	Different values might have different impact on the performance of the cost

		function in the model.
Extreme Gradient Boost	Gamma	Define the conservative feature of the model. Value depends on the used loss function.
	Min child weight	Control the overfitting during the learning and prediction of ML model. Higher values improve the global learning (i.e the model does not learn only from specific observations), however too high values might lead to the underfitting.
Random Forest & Rotation Forest	Boostrap	Define the sampling strategy during the training/ testing process. Should be set to "True" in the configuration of the model.
	Max features	Impact on the best splitting of nodes in a decision tree. High value might result in overfitting during the prediction. Max features should be the square root of available number of features in the dataset.
	Min sample leaf	Similar to "min sample split", this hyperparameter control the growth of decision tree and should be tuned with specific input data.

	Min sample split	Too low value might lead to an overgrowth of decision tree and overfitting in the prediction.
	Number of subset	Maximize the gained information through a Principle Component Analysis (PCA) from a given dataset. Higher value might lead to longer training time and overfitting. A 3-subset is recommended as the starting point to train the model.
Light Gradient Boosting Machine	Number of leaves	Maximum number of leaves that weak learners have. Higher value might increase the training and prediction accuracy, however too high value might lead to the overfitting.
Support Vector Machine	Kernel	Define the discrimination space in the classification. Radial Basis Function (RBF) is preferred to the linear and poly-nominal kernels.
	C	Control the soft margin in the model. Low/ high value of C might lead to a decision boundary with different margins, and impact on the prediction accuracy.

Epsilon	Higher value of epsilon, lower number of support vectors is selected to construct the model performance. Lower value of epsilon is recommended in almost cases.
Gamma	Control the impact distance on the training points. High to very high values might lead to overfitting in the prediction.