





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



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).

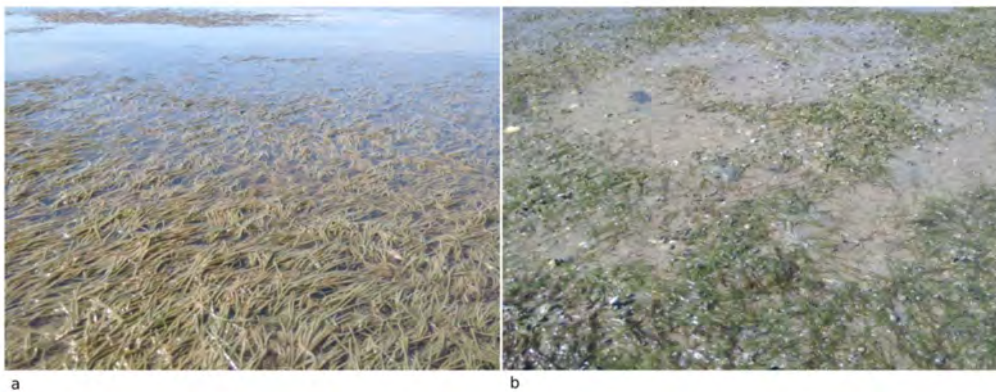


Figure 2. Dense (a) and sparse (b) seagrass meadows in Tauranga Harbor (photographs taken by Nam Thang Ha).

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
<i>Ancillary data</i>	
Gas transmittance	True
Ozone concentration (cm^{-1})	0.3
Water vapor concentration (g/cm^2)	1.5
Pressure	Normal pressure
<i>Masking</i>	
Level 2 water masking (nm)	1600
Negative reflectance masking	True
Cirrus masking	True
<i>Other parameters</i>	
Sky correction	True
Dark spectrum fitting	Fixed
Sun glint correction	False
<i>Output parameter</i>	
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):

$$BRI_i = \frac{SR_i}{e^{-k_i * g * z}} \quad (1)$$

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]:

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

where

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

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

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	
5/1/2019	60% for training 8586	40% for testing 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.

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

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

$$\text{Precision} = \frac{tp}{tp + fp} \quad (5)$$

$$\text{Recall} = \frac{tp}{tp + fn} \quad (6)$$

$$F1 = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \quad (7)$$

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.

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

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).

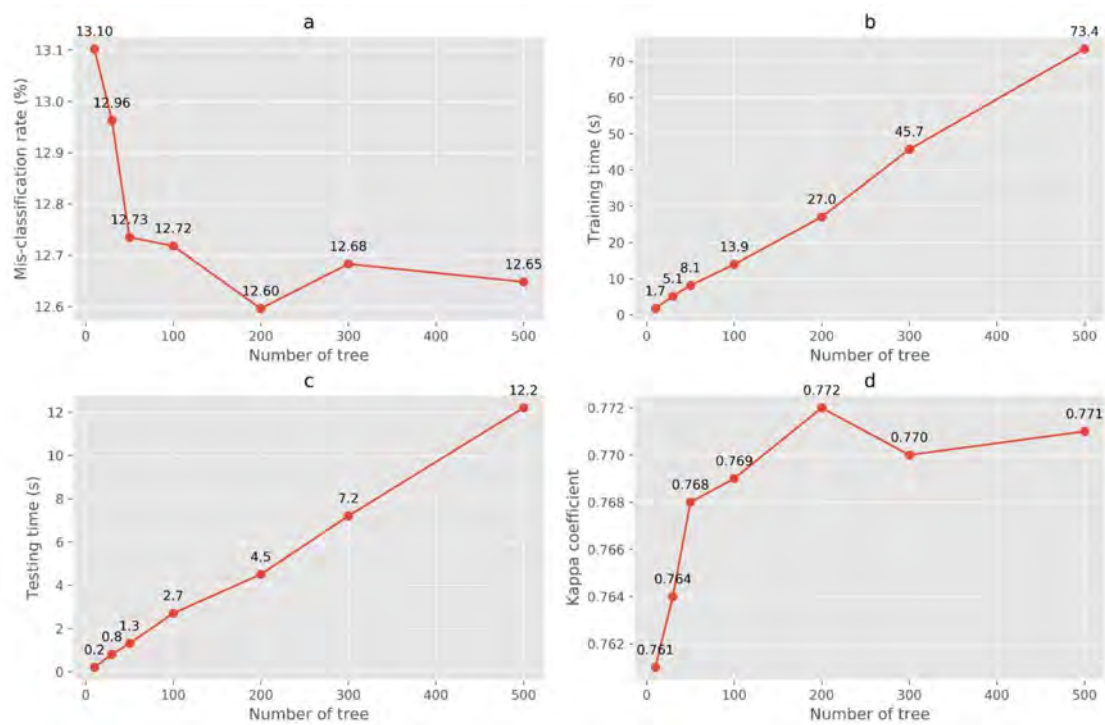


Figure 3. Number of trees optimizing for CCF model: mis-classification rate (a), computation time (b and c, the training and testing times are the computation time measured for the training and testing phases, respectively), and Kappa coefficient (d) using data described in Table 3.

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).

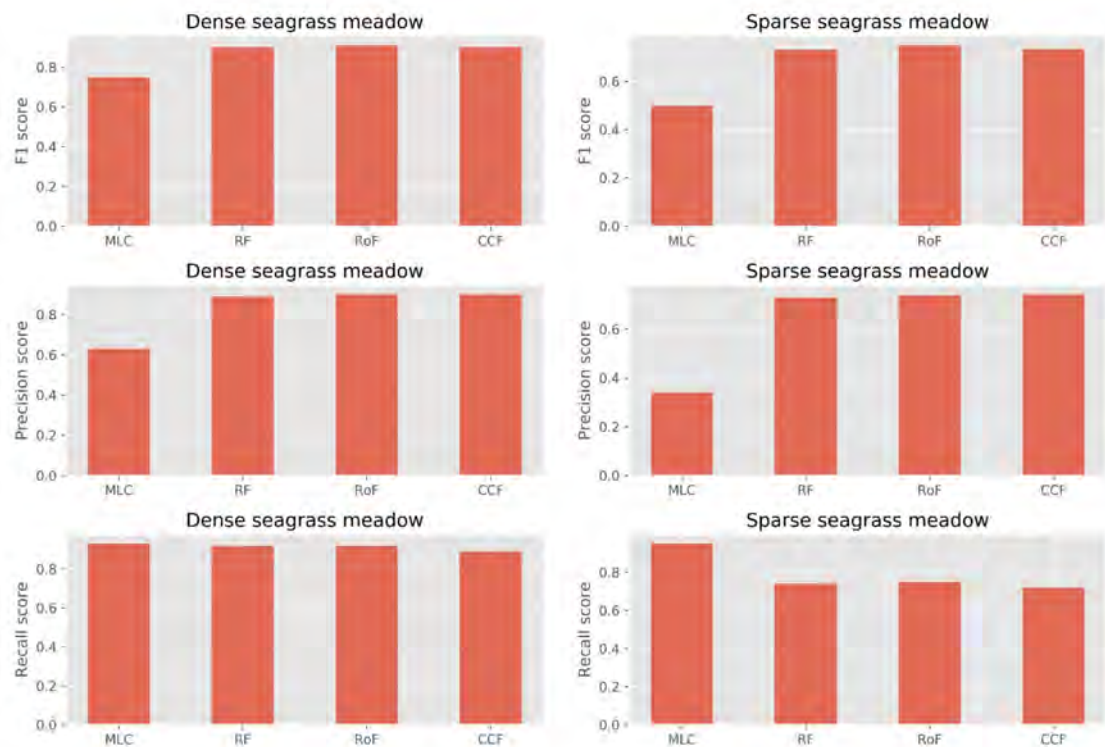
Table 5. Accuracy, precision, recall, and F1 of the model performing for the selected Sentinel-2 scene; best performance is indicated by the characters in bold.

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	<i>p</i> 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.

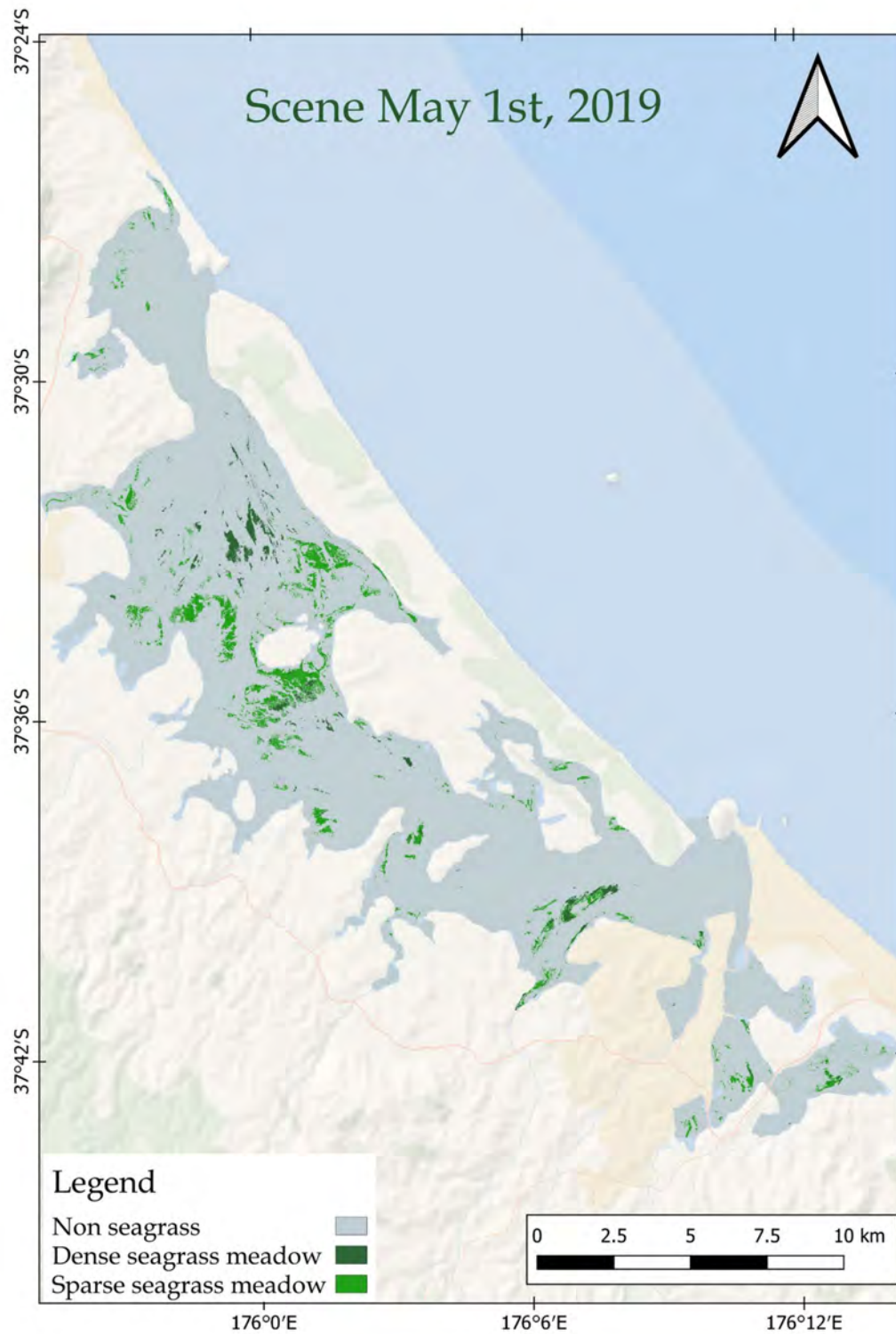


Figure 5. Seagrass map for Tauranga Harbor, 1 May 2019, derived using the RoF model applied to Sentinel-2 imagery.

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, seagrasses and 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.; Drăguț, 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 Zealand, 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 Regional 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.; Fearn, 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.; Kunecheva, 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]

