



UNIVERSITY OF LEEDS

This is a repository copy of *Comparisons of regression and machine learning methods for estimating mangrove above-ground biomass using multiple remote sensing data in the red River Estuaries of Vietnam*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/185739/>

Version: Accepted Version

Article:

Quang, NH, Quinn, CH orcid.org/0000-0002-2085-0446, Carrie, R et al. (4 more authors) (2022) Comparisons of regression and machine learning methods for estimating mangrove above-ground biomass using multiple remote sensing data in the red River Estuaries of Vietnam. *Remote Sensing Applications: Society and Environment*, 26. 100725. ISSN 2352-9385

<https://doi.org/10.1016/j.rsase.2022.100725>

© 2022, Elsevier. This manuscript version is made available under the CC-BY-NC-ND 4.0 license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Comparisons of Regression and Machine Learning Methods for Estimating Mangrove Above-Ground Biomass Using Multiple Remote Sensing Data in the Red River Estuaries of Vietnam

Nguyen Hong Quang^{a,*}, Claire H. Quinn^b, Rachael Carrie^b, Lindsay C. Stringer^c, Le Thi Van Hue^d, Christopher R. Hackney^e, Dao Van Tan^f

^a Vietnam National Space Center (VNSC), Vietnam Academy of Science and Technology (VAST), 18 Hoang Quoc Viet, Hanoi 100000, Vietnam

^b Sustainability Research Institute, School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK; C.H.Quinn@leeds.ac.uk (C.Q.); R.H.Carrie@leeds.ac.uk (R.C.)

^c Department of Environment and Geography, Wentworth Way, University of York, Heslington, York, YO10 5NG, UK; lindsay.stringer@york.ac.uk (L.S.)

^d Central Institute for Natural Resources and Environmental Studies, Vietnam National University (VNU), No 19 Le Thanh Tong, Hoan Kiem, Ha Noi 100000, Vietnam; huele@cres.edu.vn (L.T.V.H.)

^e School of Geography, Politics and Sociology, Newcastle University, UK, NUI 7RU; christopher.hackney@ncl.ac.uk (C.H.)

^f Faculty of Biology, Hanoi National University of Education (HNUE), 136 Xuan Thuy, Cau Giay, Ha Noi 100000, Vietnam; tandv@hnue.edu.vn (D.V.T.)

* Correspondence: nhquang@vnsc.org.vn (N.H. Quang)

Abstract

Currently, remote sensing platforms provide state-of-the-art data for multiple purposes including applications related to coastal wetlands. Mangrove above-ground biomass (MAGB) together with its extent is considered well correlated with the habitats' environmental and economic values. Above-ground biomass can be estimated by models that integrate remote sensing, field data and statistical information. However, it remains difficult to decide which model and which data offer the best performance for any one study location. Hence, this study aims to assess the spatial change in MAGB over a 45-year period and investigate different approaches to quantify this change through linear and multi linear regression models. Specifically, we test a non-linear model (Multivariate Adaptive Regression Splines; MARS), and non-parametric machine learning models, to predict MAGB using vegetation indices and biophysical variables derived from optical remote sensing data from Sentinel-2, Landsat-8, SPOT-7 and synthetic aperture radar remote sensing data from ALOS-2. The multi linear regression (MLR) and the MARS models were trained by field measured MAGB data to a good level of accuracy ($R^2 = 0.80$ and $RMSE = 5.56 \text{ Mg ha}^{-1}$ for MLR and $R^2 = 0.89$, $RMSE = 5.42 \text{ Mg ha}^{-1}$ for MARS). These models were subsequently applied to Landsat 2, 5 and 8 time-series images to assess changes in MAGB values and mangrove forest extent over the period 1975 to 2020. To ensure accurate training data for the models, we conducted field work to measure MAGB in 24 plots measured in May 2019. Findings showed that the MARS model generated MAGB values with higher accuracy than linear regression and multi linear regression models. Uses of vegetation indices (Normalized Differenced Vegetation Index, Soil-adjusted Vegetation Index, Green-Normalized Differenced Vegetation Index, Simple Ratio, and Red-edge Simple Ratio) generated MAGB values with accuracy slightly higher than using biophysical variables (Leaf area index, Fraction of Absorbed Radiation, Fractional vegetation cover, and Leaf chlorophyll content). Sentinel-2 and Landsat 8 were effective data sources for MAGB estimates, while SPOT-7 and ALOS-2 produced acceptable MAGB accuracy. Modelling the Landsat time series found an increase in both MAGB values and forest extent over the 1975-2020 period. The MARS model, Sentinel-2, Landsat 8 and vegetation indices are the recommended models and data to use

to measure MAGB and could be used to understand changes in MAGB and forest extent at national and regional scales.

Keywords: regression model; MARS; vegetation indices; biophysical variables; Thuy Truong commune

1. Introduction

45 Mangrove forests are one of the most diverse ecosystems on Earth and deliver numerous provisioning, regulating, cultural and supporting services that benefit coastal and inland communities (Menéndez et al., 2020, Worthington et al., 2020, Iqbal, 2020). Recognizing these benefits, strategies for mangrove preservation and growth have been developed by the Vietnamese government to reduce global climate change impacts and secure the livelihoods of coastal communities, and have been implemented in the five northern coastal provinces since the 1980s (VEA, 2016).
50 One key strategy has been forest plantation and restoration, which has attracted domestic investment and foreign donors (such as Denmark and Japan). Alongside the conversion of some mangrove areas to aquaculture, rice and salt farming, this strategy has led to fluctuations in mangrove forest extent in the Red River Delta (RRD) over the ensuing decades (Hoa, 2016, Quang et al., 2020).

Above-ground biomass (AGB), accounts for the living biomass above the soil. This includes stems, bark, branches, foliage, and seeds, and constitutes the main carbon pool in trees with a diameter at breast height (DBH) greater than
55 10 cm (Spawn et al., 2020). Mangrove above-ground biomass (MAGB), together with its spatial extent, is well correlated with environmental and economic values (Baloloy et al., 2018). In tropical regions, mangroves and seagrass are one of the most carbon-rich ecosystems, storing up to 1023 t C ha⁻¹ (Laffoley and Grimsditch, 2009), therefore MAGB is important to estimating the above-ground carbon pool (Clark and Clark, 2000), for monitoring forest
60 changes and in creating management protocols.

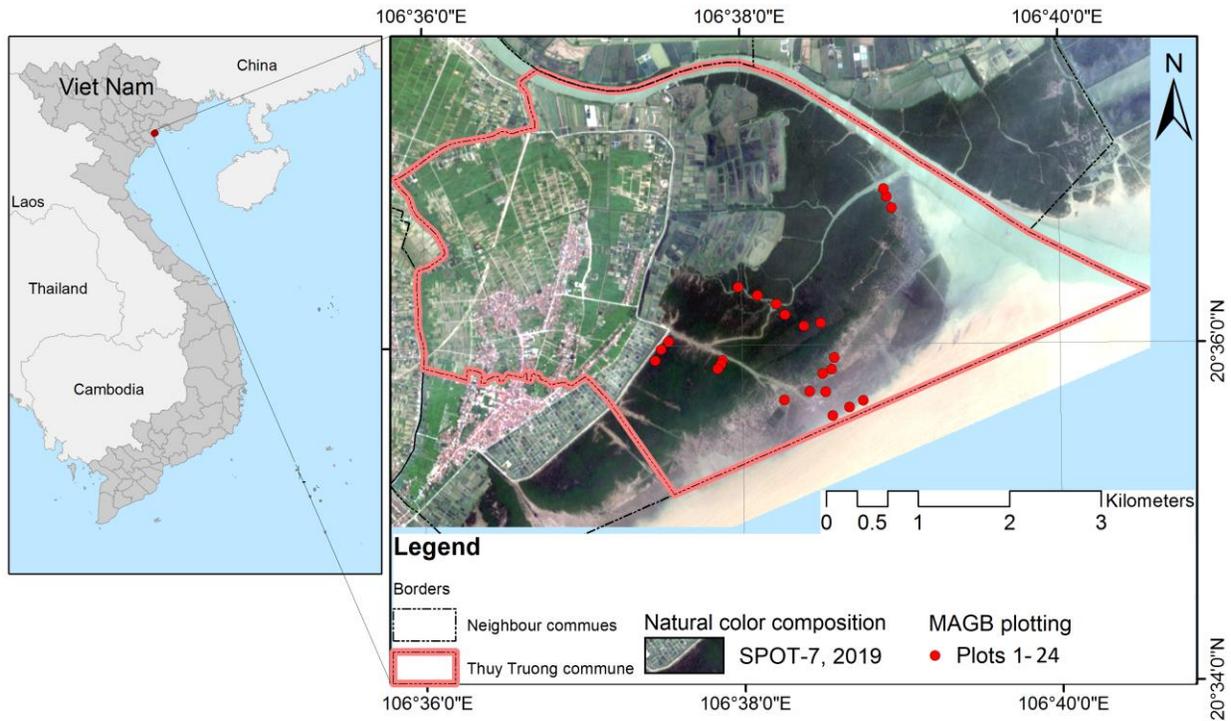
With the advantages of observation at large spatial scales and across multiple decades, remote sensing data (e.g. hyperspectral, multispectral, radar, and light detection and ranging (LiDAR)) are a valuable resource for mapping mangrove extent, distribution and change (Green et al., 1998, Zhang et al., 2012), and for extracting vegetation biophysical parameters such as chlorophyll content (Gitelson et al., 2005, Zhang et al., 2012). Among the available
65 optical remote sensing data, SPOT and Landsat images are most commonly used and users can choose between many different image processing techniques. Suitable data and techniques depend on the objective of the study. At large spatial scales satellite sensors are more appropriate than airborne and unmanned aerial vehicles (UAVs), and for better spatial coverage space and airborne data should be used rather than in situ point measurement (Congalton, 1991). Remote sensing-based methods have demonstrated their effectiveness for assessing MAGB for large forests, and
70 compared to traditional above ground biomass estimates, remote sensing offers a non-destructive more robust approach, enabling continuous, spatially explicit biomass assessment (Herold and Johns, 2007) at low cost (Pham et al., 2019, Castillo et al., 2017). Despite these advantages, remote sensing techniques can be limited because a single data pixel will contain information for a number of different individual trees. The complexity of mangrove communities and the low spatial resolution of remote sensed images can therefore result in lower accuracy in
75 estimating MAGB (Pham et al., 2019).

Various MAGB models using remote sensing data have been developed to improve MAGB estimation accuracy (Argamosa et al., 2018), with more modern approaches generating higher accuracy (Pham et al., 2019). However, MAGB is a highly heterogeneous parameter that varies across different regions, species, ages, forest types (natural and plantation) and climates, hence it is difficult to find a single “best” model to predict MAGB for all types of mangrove forests. Therefore, the aim of this study is to assess the abilities of different approaches to estimate MAGB in a region of the Red River Delta in Vietnam, to discover which model and what data will produce the most accurate MAGB values. Approaches to estimate MAGB include parametric regression models (Rosillo-Calle et al., 2015, Avtar et al., 2017) with frequent use of vegetation indices and biophysical variables (Baloloy et al., 2018) and non-parametric machine learning methods (Jachowski et al., 2013). We tested and compared simple but effective linear regression, multi linear regression, and a non-linear regression model (Multivariate Adaptive Regression Splines; MARS), and recent common machine learning models. We have used three different machine learning algorithms: Artificial Neural Network (ANN), random forest (RF) and Support Vector Machines (SVM) with data from different optical images from Sentinel-2, Landsat-8, SPOT-7 and SAR remote sensing data from ALOS-2. Finally, a Landsat time-series was used to spatially assess change in MAGB and mangrove forest extent over the last 45 years.

2. Materials and Methods

2.1. Study area

Here, we focus on the mangrove forests of the Red River Delta, Vietnam. Specifically, this research focuses on Thuy Truong commune located in Thai Thuy district in the north-east of Thai Binh province at the mouth of the Thai Binh River (central coordinates: 106°38'00E and 20°36'00N) (Fig. 1). Thuy Truong commune covers 9.3 km² and is home to approximately ten thousand people (Quang et al., 2020). The main livelihoods are based on agriculture (rice), aquaculture and the harvesting of clams, fish and crabs from nearby mangrove forests (Pham, 2007). The mangrove forest in this region plays an important role in providing habitat for birds, snakes and other animals, protecting local citizens from strong storms and waves (Mazda et al., 1997), and as a source of food for consumption or sale (Hoa, 2016). As a result, the forest is protected by the government, and by local citizens. The area has benefited from forest plantation supported by the Danish and Japanese Red Cross programs, both of which ended in 2006. Hence, although the general picture of total mangrove extent in Vietnam as a whole is one of decline (Powell et al., 2011), the Thuy Truong mangrove forest has increased in extent by about 3.5 km seaward (Quang et al., 2020). The forest comprises three main mangrove species: *Sonneratia caseolaris*, *Kandelia obovata* and *Aegiceras corniculatum* (Quang et al., 2020, Loan et al., 2020). The study area has four distinct seasons; spring from February to May, summer from June to August, autumn from September to November, and winter from December to January, and the climate of the region is influenced by the South-East Asian tropical monsoon. As the maximum and minimum monthly average temperatures are approximately 28°C in July and 16°C in January, the mangrove is not seasonally deciduous, staying green all year round. Hence, the mangrove leaf biomass does not display much seasonal variation.



110 **Fig. 1.** Location of Thuy Truong commune on the SPOT-7 natural color composition image, the red circles denote 24 plots in which field mangrove above ground biomass (MAGB) measurements were taken in May 2019.

2.2. Data collection

2.2.1. Remote sensing data

115 For the purposes of establishing MAGB models, we collected different remote sensing data with sensing dates as close as possible to the dates of the field measurements ([Table 1](#)) to minimize MAGB differences due to seasonal and/or growth effects. These data were Sentinel-2, Landsat-8 and SPOT-7 images and ALOS-2/PALSAR-2 (AL2). The Sentinel-2 scene (optical) was at level 1-C obtained freely from Copernicus Open Access Hub of the European Space Agency (ESA). The Landsat-8 OLI was obtained from the United States Geological Survey (USGS) at level 1, and

120 SPOT-7 data were purchased from the Airbus group. The SAR remote sensing image of ALOS-2 was provided by the Japan Aerospace Exploration Agency (JAXA) at level 1.5 (L15) and used for estimating MAGB, and for comparison with the results from the optical images using identical methods of regression and machine learning. A time series of Landsat images for each 5-year period from 1975 to 2020 with minimal cloud cover (in October and November) were acquired to analyze MAGB changes over time. Metadata is summarized within [Table 1](#).

125 **Table 1**

Remote sensing data covering the Thuy Truong commune, Thai Binh province; where B and BQA stand for band and band quality, respectively; MSS is Multispectral Scanner Sensor; TM stands for Thematic Mapper; OLI is Operational Land Imager; Mul and Pan are short for multispectral and panchromatic bands, respectively; GPL is geometric processing level; RPL is radiometric processing level; and GRD is ground-range detected. V and H are vertical and horizontal, respectively, and coupled letters of VH and VV indicate SAR cross-polarizations.

Sensors	Product level	Bands	Spatial resolution (m)	Acquisition date
Sentinel-2	Multispectral image-L1C	B2-B8, B11, B12	10 (B2-4, B8) 20 (B5-7, B11,12)	2019/06/25
Landsat-8 OLI	L1TP	1–11, BQA	30	2019/05/18
SPOT-7	GPL: Sensor Basic	RPL: Band 0–3 (Mul) Pan	6 (Mul) 1.5 (Pan)	2019/05/17
ALOS-2	L15 (level 1.5, terrain correction)	HH, and VH	3.125	2019/07/31 (Ascending)
Landsat-X	(2) L1TP	(2) 4–6	(2) 60m	1975/04/20 (2MSS)
Time series (X=2, 5 and 8)	(5) L1TP	(5) 1–7, BQA	(5) 30m	1988/11/04 (5TM)
	(5) L1TP	(5) 1–7, BQA	(5) 30m	1993/11/02 (5TM)
	(5) L1TP	(5) 1–7, BQA	(5) 30m	1998/10/15 (5TM)
	(5) L1TP	(5) 1–7, BQA	(5) 30m	2003/10/10 (5TM)
	(5) L1TP	(5) 1–7, BQA	(5) 30m	2008/11/11 (5TM)
	(8) L1TP	(8) 1–11, BQA	(8) 30m	2013/10/08 (8OLI)
	(8) L1TP	(8) 1–11, BQA	(8) 30m	2018/10/06 (8OLI)
	(8) L1TP	(8) 1–11, BQA	(8) 30m	2020/05/20 (8OLI)

2.2.2. Field measurements for estimating mangrove above-ground biomass

Field measurements of MAGB were undertaken in May 2019 with clear skies and low cloud cover. Areas characterized by different mangrove species were measured in different plots, resulting in 24 distinct measurement plots. All field measurement activities were undertaken in 24 10m × 10m plots for three mangrove types of *S. caseolaris* plots for *K. obovata* and *A. corniculatum*. Plot borders were demarked using ropes. A GPS (Garmin Montana 680) with an integrated 5M camera was used to position the four corners of each plot. Plot locations are shown in Fig. 1 with the points corresponding to the center of each plot. The GPS was used for measuring the coordinates of the plots which subsequently enables to overlay the extraction of remote sensing-based MAGB values with plot-based data for comparison.

The above-ground biomass (AGB) values for each mangrove species were estimated using established allometric equations (Table 2), using parameters of tree height (H), diameter at breast height (D_{BH}), and stem diameter at the

position of 10% of the tree height ($D_{0.1}$) (Hoque et al., 2011). The total height and diameter at 1.3 meters above the ground (D_{BH}) of *S. caseolaris* individuals and the tree height and diameter at 10% of the tree height ($D_{0.1}$) of *K. obovata* and *A. corniculatum* were measured.

To calibrate the allometric equations for predicting biomass, a sample of measured (diameter and height) trees were cut down at ground level. After harvesting, the stems, branches and leaves of each individual were separated and immediately weighed to record fresh biomass. 200 – 1000 g of stems or branches and 100 g of leaves were sampled from each cut tree and weighed before drying in an oven at 105°C until a constant weight to establish the relationship between dry and fresh weight. For *S. caseolaris*, the cut trees (n=12) had a height and D_{BH} that ranged from 0.6 – 6.0 m and 2.5 – 25 cm, respectively. *K. obovata* (n=28) and *A. corniculatum* (n=15) individuals had heights and D_{10} that ranged from 0.8 – 4.6 m and 3.5 – 12 cm, respectively. The coefficient of determination between $D_{0.1}^2H$ or D_{BH}^2H and individual tree biomass (R^2) values greater than 0.9 and square error (SE) values less than 2.3 kg presented the allometric equations predicted accurate MAGB values.

Table 2

Allometric equations developed for each mangrove species in the study site ($P < 0.001$)

Mangrove Species	AGB allometric equations	Parameter	R^2	SE (kg)
<i>Sonneratia caseolaris</i>	Biomass (kg) = $0.2123 \times (D_{BH}^2 H)^{0.7083}$	D_{BH} (cm), H (m)	0.982	0.251
<i>Aegiceras corniculatum</i>	Biomass (kg) = $0.0463 \times (D_{0.1}^2 H)^{0.761}$	$D_{0.1}$ (cm), H (m)	0.887	0.182
<i>Kandelia obovata</i>	Biomass (kg) = $0.0513 \times (D_{0.1}^2 H)^{0.8416}$	$D_{0.1}$ (cm), H (m)	0.928	0.246

Note: AGB is the above-ground biomass, D_{BH} is stem diameter at breast height (1.3m), $D_{0.1}$ is stem diameter at the position of 10% of the height, and H is the height of the tree.

2.3. Methodology

To estimate and map MAGB we developed a flowchart of the methodology (Fig. 2) which consisted of the following steps; (i) pre-processing of optical images from Sentinel-2, Landsat-8 and SPOT-7 and SAR images of ALOS-2/PALSAR-2 (AL2) before they were all co-registered and projected in the singular coordinate system of the Universal Transverse Mercator World Geodetic System 1984 (UTM-WGS 84) zone 48 North (section 2.3.1); (ii) calculation of vegetation indices from optical remote sensing bands (section 2.3.2); (iii) calculation of biophysical variables also from optical remote sensing bands (section 2.3.3); (iv) setting up three MAGB estimation and modelling approaches (section 2.3.4); (v) processing the Landsat-X time series with a pre-processing step (section 2.3.5); and (vi) mapping of MAGB (section 2.3.6). The best linear regression (LR) (and variable), multiple linear regression (MLR) and multivariate adaptive regression spline (MARS) models were used to generate time-series MAGB maps.

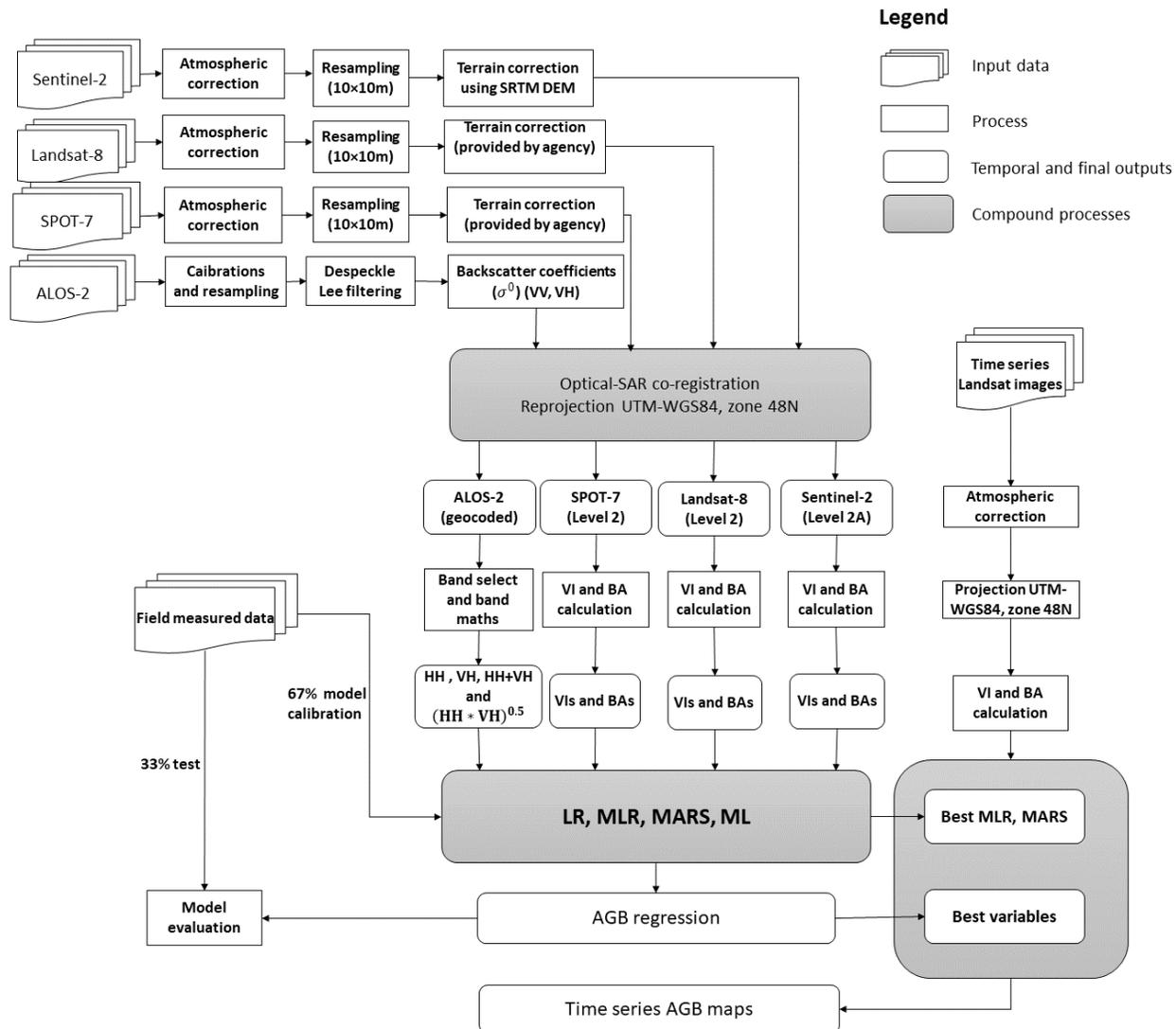


Fig. 2. Work flow for mangrove above-ground biomass (MAGB) estimation, where: V and H are vertical and horizontal, respectively, and coupled letters HH and VH indicate polarization configurations for a SAR system, SRTM DEM stands for Shuttle Radar Topography Mission Digital Elevation Model, VI and BA are vegetation indices and biophysical variable respectively, MLR stands for Multiple linear regression, MARS is the Multivariate Adaptive Regression Splines model, ML is machine learning and BF is basic function, and UTM-WGS84, zone 48N is the Universal Transverse Mercator World Geodetic System 1984, zone 48 North.

2.3.1. Image pre-processing

To combine multi-source remote sensing images, all the data were pre-processed including radiometric calibration into top-of-atmosphere (ToA), resampling in order to homogenize the spatial resolution of all images to 10m (using cubic convolution method, and an output pixel size of 10m), terrain correction for the Sentinel-2 images using the Shuttle Radar Topography Mission Digital Elevation Model (0.1degree resolution). The ESA Sen2Cor toolbox was used for the Sentinel-2 image atmospheric corrections. The Fast Line-of-sight Atmospheric Analysis of Hypercubes

180 (FLAASH) tool was applied for the Landsat-X image atmospheric corrections. Next, the surface reflectance of all the pre-processed optical images were used for the vegetation index and biophysical variable calculations. The AL2 SAR image at level 1.5, was calibrated to reduce image radiometric bias, filled using the Lee filtering method (Ali et al., 2015) to reduce image speckles, and converted to the backscattering coefficient (σ^0 in decibel (dB)) using Equation (1) where CF is the calibration factor and set at -83 dB:

$$185 \quad \sigma_{dB}^0 = 10 \times \log_{10} \sigma^0 + CF \quad (1)$$

$$\sigma^0 = \frac{DN^2}{A_{dn}^2} \times \frac{1}{G_{eap}^2} \times \left(\frac{R}{R_{ref}}\right)^3 \times \sin(\alpha) \quad (2)$$

Where $\frac{1}{G_{eap}^2}$ is the elevation antenna pattern (EAP) correction (2-way), $\left(\frac{R}{R_{ref}}\right)^3$ indicates the range spreading loss (RSL) correction, A_{dn}^2 is the product final scaling from internal SLC to final SLC or GRD, α is the local incidence angle and DN^2 is the average product intensity and has a value of 22142.71.

All optical and SAR images were then projected in the unique Universal Transverse Mercator World Geodetic System 1984 zone 48 northern hemisphere (UTM-WGS 84, zone 48N) to enable overlay.

2.3.2. Calculation of vegetation indices

Numerous vegetation indices can be derived from remote sensing data by transformation of multispectral information. Since the Normalized Difference Vegetation Index (NDVI), Soil-adjusted Vegetation Index (SAVI), Green NDVI (GNDVI), Simple Ratio (SR) and Red-edge Simple Ratio (SRre) have been most frequently applied to mangrove data (Green et al., 1998), we selected them for our model inputs. The indices were calculated using the formulas presented in Table 3.

Table 3

200 Equations for the vegetation indices used in the biomass models

Vegetation index	Formula	Reference
Normalized Difference Vegetation Index (NDVI)	$(NIR-R) / (NIR + R)$	(Rousel et al., 1974)
Green NDVI (GNDVI)	$(NIR-G) / (NIR + G)$	(Gitelson et al., 1996)
Soil Adjusted vegetation Index (SAVI)	$\frac{(NIR-R)}{(NIR+R+L)} \times (1 + L)$	(Huete, 1988)
Simple ratio (SR)	NIR/R	(Jordan, 1969)
Red-edge simple ratio (SRre)*	$NIR/R-edge$	(Gitelson and Merzlyak, 1994)

Note: NIR is the near infrared band (~833 nm), R is red band (~665 nm), G is green band (~560 nm), L is the correction factor, for the higher vegetation density of mangrove forests L values of 0.75 are recommended (Huete, 1988), * is not applicable for Landsat 8 and SPOT-7.

2.3.3. Calculation of biophysical variables

Biophysical variables are useful for biomass estimation as they describe the spatial distribution of vegetation state and dynamics (Widlowski et al., 2004) and are central to the use of multispectral remotely-sensed images to understand the correlation between the spectral reflectance of mangrove forests and their biophysical parameters (Pham et al., 2019). We selected Leaf Area Index (LAI), fraction of absorbed photosynthetically active radiation (fAPAR), fractional vegetation cover (FVC), and leaf chlorophyll content (Cla) because they are the most sensitive parameters for MAGB estimation (Bilgili et al., 2010). These were calculated using the built-in Biophysical processor within SNAP Toolbox (Kganyago et al., 2020) supporting Sentinel-2 and Landsat 8 images. Since SPOT images have no biophysical variables included in their products, we calculated the LAI, FVC and Cla using the following equations (Ali et al., 2015);

$$LAI_{NDVI} = \frac{-\log(1-FVC_{NDVI})}{k(\theta)} \quad (3)$$

Where LAI_{NDVI} is NDVI-derived leaf area index, FVC_{NDVI} is the fractional vegetation cover based on NDVI and $k(\theta)$ indicates the light extinction coefficient for a given solar zenith angle (range between 0.4 and 0.65) and set at value of 0.5. The FVC_{NDVI} in equation 3 is solved by equation 4.

$$FVC_{NDVI} = \frac{NDVI - NDVI_s}{NDVI_v - NDVI_s} \quad (4)$$

Where $NDVI_s$ is NDVI for bare soil and $NDVI_v$ is NDVI for vegetated land.

Chlorophyll-a green (Cl_{aGreen}) was obtained using the Green Chlorophyll Index model (CIGREEN) developed by Gitelson et al. (2003) for comparison with the Sentinel-2.

$$Cl_{aGreen} = (NIR - Green) - 1 \quad (5)$$

Where Green is the green band (0.530 μ m – 0.590 μ m) of the SPOT-7 and Landsat 8 image.

The SPOT-7 fAPAR was computed from the leaf area index (LAI) using equation (6) suggested by Yuan et al. (2015):

$$fAPAR = 0.2058 * \ln(LAI) + 0.5378 \quad (6)$$

2.3.4. MAGB models

- *Linear and multiple linear regression models*

A linear regression is a simple but effective method to estimate MAGB, not only of mangrove forest but other types of vegetation, based on linear correlation between a single remote sensing band layer or other indices and the target variable (MAGB etc.). Relationships between a response variable y , and a single explanatory variable x , given a set of data that includes observations for both of these variables for a particular sample, is estimated by a simple linear regression (Tranmer and Elliot, 2008). For example, the above ground biomass of mangrove forest (MAGB) as a response variable can be predicted from examination of the results of the calculation of vegetation indices or by using a single reflectance band from remote sensed images - the explanatory variable (V_i).

$$MAGB = \beta_0 + \beta_1 * V_i + e_i \quad (7)$$

Where β_0 is the intercept also called the constant. β_1 is the slope of the line – this is how much the value of MAGB increases and e_i is the error term for the explanatory variable (V_i).

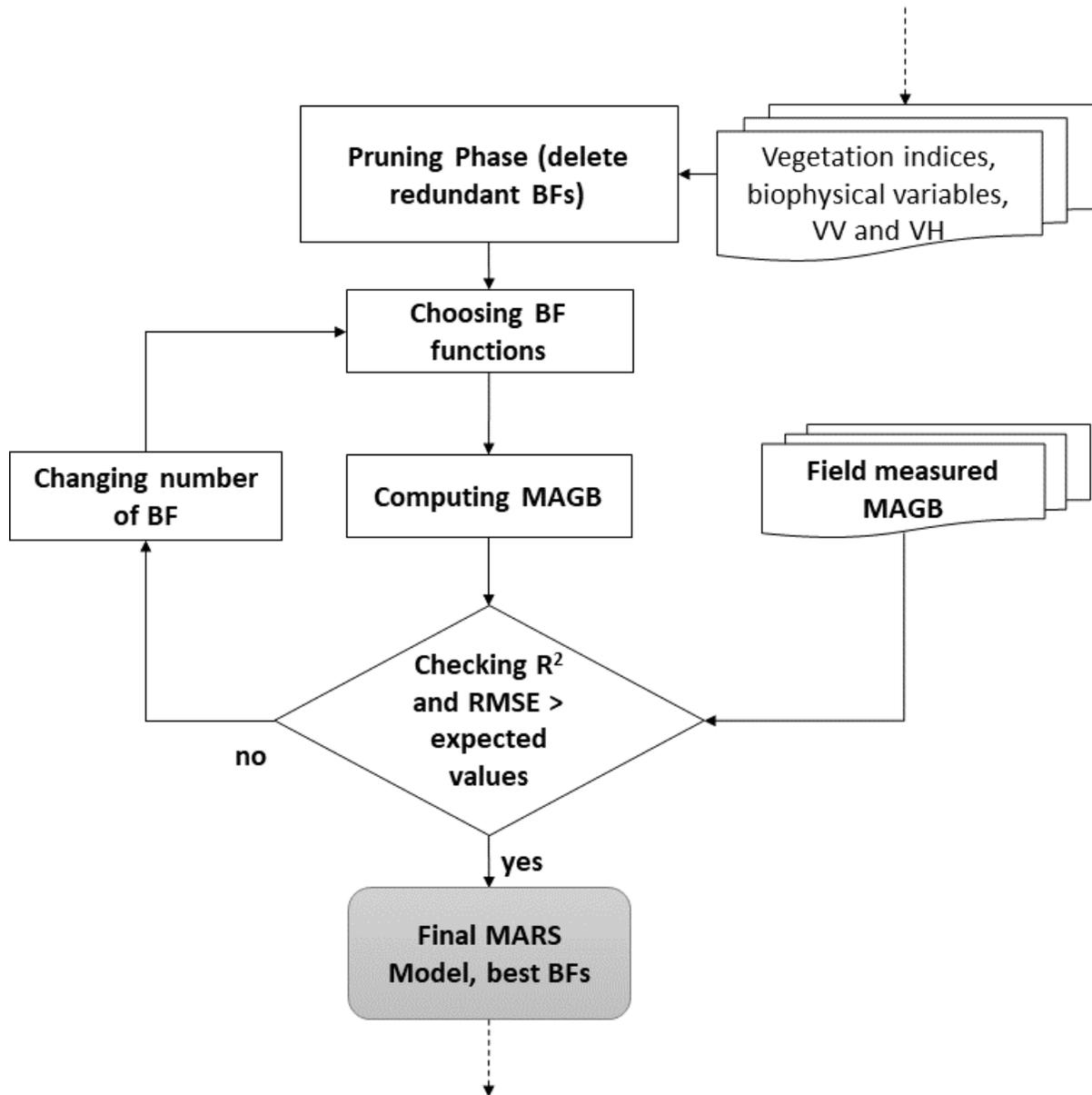
More than one explanatory variable (V_{ni}) is included in multiple linear regression (MLR) (Tranmer and Elliot, 2008, Marill, 2004) to estimate a single dependent outcome variable $MAGB_i$ and each linear regression has its own slope ($\beta_{1,2,...n}$).

$$MAGB = \beta_0 + \beta_1 * V_{1i} + \beta_2 * V_{2i} + \dots + \beta_n * V_{ni} + e_i \quad (8)$$

In this study, we grouped explanatory variables (V_{ni}) into three groups of vegetation indices and biophysical variables for Sentinel-2, Landsat-8 and SPOT-7 and two dual polarizations HH and VH and ratios HH+VH and $(HH \times VH)^{0.5}$ of the ALOS-2 backscatters by compartments (Soares and Schaeffer-Novelli, 2005). The MLR models were developed after hundreds of loops of slope (β_1) and intercept or constant (β_0) adjustments from the initial models. This procedure was repeated for the three groups of the target variables ($MAGB$). The initial models were established from linear regression models (LRM) of individual explanatory variables, and the Pearson's correlation coefficient (PCC) of each LRM was used as its weight for the adjustments (increment and reduction). After each increment and reduction, RMSEs were calculated and compared to the previous values. Loops were stopped when the RMSE values approached minimum values.

- *Multivariate adaptive regression splines model*

Multivariate Adaptive Regression Splines (MARS) is a regression and data mining technique developed by Friedman (Friedman, 1991). Basic functions in modeling the predictor and response variables are used in this method. The generated basic functions were employed as the new set of predictor variables to generate the final model. A forward algorithm, which selects all possible basic functions and their corresponding knots (constants of the basic functions) is used in the initial step of MARS. Next, the backward algorithm discards basis functions which do not contribute significantly to the accuracy of the fit. In the final step, the model selects the remaining basic functions (Bilgili et al., 2010). We applied the MARS model to estimate $MAGB$ in this study (Fig. 3) as an extension of the overall study work flow presented in Fig. 2. The vegetation indices (VI), biophysical variables (BV) and ALOS-2 polarizations (VV, VH) calculated from previous steps were pruned to eliminate redundant basic functions (BFs) before they were selected to estimate $MAGB$. BFs were chosen to compute $MAGB$ and the condition that the coefficient of determination (R^2) and RMSE values were greater than the expected ratios from field measured $MAGB$ were checked. This process continued until the final model results consisted of the best matched calculated $MAGB$ with field measured $MAGB$ and a set of BFs based on the highest R^2 and lowest RMSE values. In theory, single remote sensing bands and mixed parameters could be used in one model. However, in order to allow comparisons, we grouped VIs, BVs and SAR data for the MARS runs separately. This also shortened the time taken for model runs and reduced the complexity of the model.



265

Fig. 3. Sub-work flow for mangrove above-ground biomass (MAGB) estimation-MARS model; where: R^2 is coefficient of determination, RMSE stands for root-mean-square error, BF is basic function and MARS is the Multivariate Adaptive Regression Splines model.

- *Machine learning approaches*

270

Non-parametric algorithms in machine learning have been increasingly applied because they increase the robustness of image supervised classification methods (Maxwell et al., 2018, Ou et al., 2019), model complex class signatures, and accept a variety of training data, and are routinely found to have higher accuracies than the maximum likelihood method (Yu et al., 2014). The artificial neural network (ANN), random forest (RF), and support vector machine (SVM) methods were selected from among the many available because they are considered robust and suitable for MAGB

275

estimations (Pham et al., 2019, Pham et al., 2020b, Hu et al., 2020, Pham et al., 2020a).

Machine learning methods have been well explained and documented (Quang et al., 2020, Jachowski et al., 2013, Hu et al., 2020, Pham et al., 2020b, Pham et al., 2019), however they can be applied differently. In this study, we have applied these three methods using combinations of all the VIs and BVs from each optical image and a combination of SAR polarizations (HH, VH) and HH+VH and (HH×VH)^{0.5} inputting into the machine learning classifiers. As they are all supervised classifiers, training data are required, but field collected training data from the 24 plots were considered insufficient for machine learning methods. To create training data, we applied image segmentation for all the combinations. A multi-resolution algorithm was applied with an equal weighting of 1 for all image bands and the scale parameter, shape and compactness were set at 10, 0.3 and 0.8 respectively. The ML model training dataset of 216 polygons (segments) and the data for the model accuracy evaluation of 81 polygons were taken from vector layers as results of the image segmentation and labeled based on the segment values at the 24 field measured plots.

- *Model evaluation*

The coefficient of determination (R^2) and the root-mean-square error (RMSE), calculated using equations 9 and 10 respectively, were used to evaluate the performance of the proposed models. These error-indicative indices are frequently utilized to evaluate the differences between measured and predicted MAGB data.

$$R^2 = \frac{\sum_{i=1}^n (mbe_i - \overline{mbe})(mbm_i - \overline{mbm})}{\sqrt{(\sum_{i=1}^n (mbe_i - \overline{mbe})^2)(\sum_{i=1}^n (mbm_i - \overline{mbm})^2)}} \quad (9)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (mbe_i - mbm_i)^2}{n}} \quad (10)$$

Where mbe_i is the model estimated MAGB value, mbm_i is the measured MAGB value obtained from field measurements, n is the total number of sampling plots, and \overline{mbe} and \overline{mbm} are mean values of mbe_i and mbm_i , respectively.

Applying machine learning methods, twelve confusion matrices or contingency matrices, user, producer and overall accuracies, and Kappa statistics were computed for each ML classification (Congalton, 1991).

2.3.5. Applying the best MLR and MARS model for Landsat-X time series

Ideally, every estimate should be trained if a sufficiently comprehensive set of training data is obtainable. However, for a long time series of remote sensing imagery this condition is not always met. Therefore, calibrated models used for other data sets could be a suitable alternative option (Green et al., 1998) to deliver acceptable outputs (Quang et al., 2020). Here, we applied the best-fit MLR and the final MARS models trained for the three different sets of vegetation indices (which produced better results compared to the biophysical variables (Baloloy et al., 2018)) of Sentinel-2, Landsat 8 and SPOT-7 for the time-series Landsat-2, 5 and 8 images to enable a rough estimation of changes in MAGB and mangrove extent over time.

2.3.6. Mapping of mangrove above-ground biomass

The results of the MLR and MARS model are point-based models that have the closest fit to MAGB values from field measured data. To generate maps of the whole forest, we applied the MLR for the group of variables (vegetation indices, biophysical variables and SAR polarizations) and best MARS model for each remote sensing image. Machine learning models are pixel-based models; hence all pixels were ready labeled with predicted values. MAGB grid layers were converted to vector shape-file format and refined in QGIS. The vector mangrove MAGB maps have seven classes

based on the range of MAGB values from lowest to highest. An identical mapping method was applied to the Landsat time-series images.

3. Results

3.1. Estimated MAGB from field measurements

315 **Table 4** presents the plot MAGB values in total calculated for individual mangrove species using the MAGB allometric equations (Table 2), and percentages of dominant mangrove species in each plot. The mangrove plots can be divided into groups of 1-3 which were planted mature mangroves with *S. caseolaris* occupying around 70% and a group of plot 19-25 were newer mangroves planted in 2013 and 2016 with nearly pure *S. caseolaris*. A mixture of the three mangroves in plots 11-18 were natural-original species and in degrading stage, hence the MAGB values (lower than 320 Mg ha⁻¹) lower than the well-growing mangrove trees in the plot 4-9 with dominant of *K. obovata*.

Table 4

Above ground biomass calculations for individual mangrove species and total plots, * new planted mangroves in 2016.

Plot No.	MAGB (Mg ha ⁻¹)	Dominant species	Percentage (%)	Plot No.	MAGB (Mg ha ⁻¹)	Dominant species	Percentage (%)
1	26.05	<i>S. caseolaris</i>	36.20	13	68.19	<i>A. corniculatum</i>	26.61
		<i>K. obovata</i>	63.80			<i>K. obovata</i>	15.58
2	17.49	<i>K. obovata</i>	26.68	14	69.41	<i>S. caseolaris</i>	57.81
		<i>S. caseolaris</i>	73.32			<i>A. corniculatum</i>	20.30
						<i>K. obovata</i>	8.37
3	34.74	<i>K. obovata</i>	20.44	15	43.93	<i>S. caseolaris</i>	71.34
		<i>S. caseolaris</i>	79.56			<i>A. corniculatum</i>	33.66
						<i>K. obovata</i>	14.28
4	43.11	<i>K. obovata</i>	94.76	16	35.34	<i>S. caseolaris</i>	52.06
		<i>A. corniculatum</i>	5.14			<i>K. obovata</i>	33.45
						<i>S. caseolaris</i>	66.55
5	93.25	<i>A. corniculatum</i>	2.25	17	34.42	<i>K. obovata</i>	40.49
		<i>K. obovata</i>	97.75			<i>S. caseolaris</i>	59.51
6	90.96	<i>K. obovata</i>	99.68	18	40.93	<i>K. obovata</i>	40.46
						<i>S. caseolaris</i>	59.54
7	124.77	<i>A. corniculatum</i>	8.26	19	7.88	<i>S. caseolaris</i>	99.89
		<i>K. obovata</i>	86.46				
8	89.84	<i>A. corniculatum</i>	7.58	20	13.56	<i>S. caseolaris</i>	99.24
		<i>K. obovata</i>	86.88				
9	64.82	<i>A. corniculatum</i>	4.56	21	11.04	<i>K. obovata</i>	4.08
		<i>K. obovata</i>	95.44			<i>S. caseolaris</i>	95.92
10	22.26	<i>A. corniculatum</i>	13.37	22*	0.303	<i>S. caseolaris</i>	100
		<i>K. obovata</i>	3.86				
		<i>S. caseolaris</i>	82.76				
11	33.38	<i>A. corniculatum</i>	15.47	23*	0.319	<i>S. caseolaris</i>	100
		<i>S. caseolaris</i>	84.19				
12	31.96	<i>A. corniculatum</i>	17.00	24*	0.25	<i>S. caseolaris</i>	100
		<i>S. caseolaris</i>	82.40				

High correlations were found between MAGB and biophysical parameters (D^2H) of the 9 plots (Fig. 4) with an R^2 of 0.98 for *Sonneratia caseolaris*, 0.89 for *Aegiceras corniculatum* and 0.93 for *Kandelia obovata*. These data were used for training and accuracy assessment of the multiple linear regression (MLR) and MARS model, and in generating the training data set for the machine learning models.

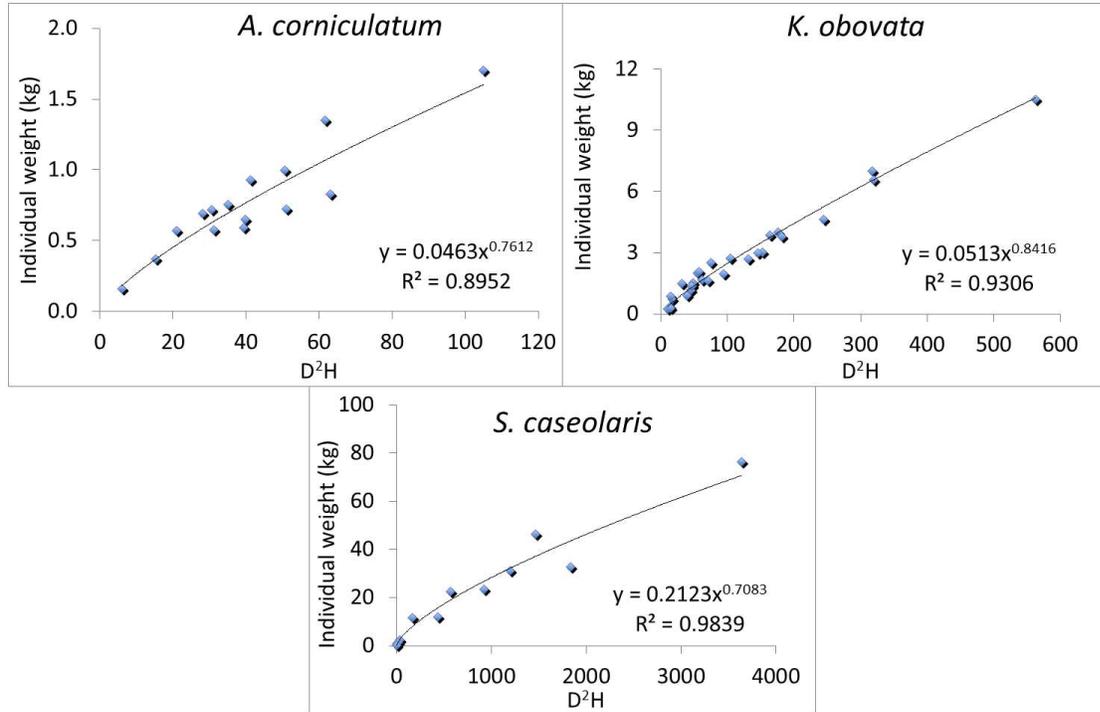


Fig. 4. Correlations between ratio D^2H and mangrove above-ground biomass (MAGB); D is the stem diameter of the stem at 1.3 m above the ground, while for *A. corniculatum* and *K. obovata*, it is the diameter at 10% of the height of the stem; and H is the mangrove tree height.

3.2. Mangrove above ground biomass multiple linear regression

3.2.1. Final multiple linear regression models

The MLR models (Table 5) had various slope values even with the same group of vegetation indices (VI), biophysical variables and SAR backscattering layers. Mostly, the slopes of NDVI and GNDVI were larger than SAVI and SR, and SRre's slopes were lowest in the VI group. Slopes of FVC and GNDVI were higher than those of LAI and NDVI in the biophysical variable group. The Cla's slopes were lowest in all cases. The ratio of $(HH \times VH)^{0.5}$ variable had the most influence on the ALOS-2 MAGB prediction (with slope of -2.24). The errors for the explanatory variables (e_i) varied from a lowest value of ± 5.27 (Sentinel-2) to a highest value of ± 7.75 (ALOS-2). The best fit MAGB model was the model of Landsat 8 using the vegetation indices.

Table 5

Results of final multi linear regression models

Sensor	Multiple linear regression model	Error
Sentinel-2	$MAGB_{Vi} = 67.89 * NDVI + 73.34 * GNDVI + 46.60 * SAVI + 10.617 * SR - 0.86 * SRre - 484.36 + e_i$	$(e_i = \pm 6.07)$
	$MAGB_{BV} = 21.64 * LAI + 27.17 * fAPAR + 55.48 * FVC + 0.49 * Cla - 394.26 + e_i$	$(e_i = \pm 6.42)$
Landsat 8	$MAGB_{Vi} = 29.46 * NDVI + 48.11 * GNDVI + 13.16 * SAVI + 17.55 * SR - 494.61 + e_i \#$	$(e_i = \pm 5.56)$
	$MAGB_{BV} = 59.05 * LAI + 31.96 * fAPAR + 44.74 * FVC + 0.47 * Cla + 54.49 + e_i$	$(e_i = \pm 5.27)$
SPOT-7	$MAGB_{Vi} = 52.34 * NDVI + 90.05 * GNDVI + 15.87 * SAVI + 24.15 * SR - 210.69 + e_i$	$(e_i = \pm 7.67)$
	$MAGB_{BV} = 96.74 * LAI + 60.29 * fAPAR + 189.02 * FVC + 0.03 * Cla - 4236.94 + e_i$	$(e_i = \pm 7.33)$
ALOS-2	$MAGB_{SAR} = 1.77 * HH + 0.33 * VH + 0.53 * (HH + VH) - 2.24 * (HH * VH)^{0.5} + 499.43 + e_i$	$(e_i = \pm 7.75)$

the best multi linear regression model and so was applied to the Landsat-X time-series

3.2.2. Accuracy assessments of the linear/multi linear regression and MAGB estimation

Table 6 shows the coefficients of determination (R^2) and root mean square error (RMSE) values calculated (Eq. 6, 7)

345 for individual vegetation indices, biophysical variables from optical images and the SAR ALOS-2 data, and for the MLR. Overall, Landsat 8 MAGB had the highest coefficient of determination (R^2) with the field measured data, followed by Sentinel-1 and SPOT-7. The ALOS-2 MAGB had the poorest fit to the field measured data ($R^2 = 0.31$ for VH and 0.54 for the ratio $(HH \times VH)^{0.5}$). Among the vegetation indices, SR was the most accurate predictor of MAGB with R^2 greater than 0.7, except for the SPOT-7's SR ($R^2 = 0.53$). In contrast, the Rrep index was not a useful predictor

350 of MAGB, particularly in the case of Sentinel-2's Rrep with an $R^2 = 0.29$. In the biophysical variable group, LAI and GNDVI were the most useful predictor variables, followed by SAVI and fAPAR. The accuracy of all the MLR models was approximately 8% higher than the most accurate VI and BV from the same remote sensing source.

Table 6

Linear/multi linear coefficient of determination (R^2) between in-situ measured and satellite-

355 estimated MAGB and the root mean square error computed for the predictors of vegetation indices, biophysical variables and SAR remote sensing data (NA= not applicable)

Sensor	Predictor	R^2	RMSE (Mg ha ⁻¹)	Resampled/original resolution (m)
Sentinel-2 (Vegetation indices)	SR	0.71	6.84	10/20-60
	GNDVI	0.68	7.46	10/20-60
	NDVI	0.68	7.74	10/20-60
	SAVI	0.62	8.10	10/20-60
	SRre	0.29	10.17	10/20-60
	MLR	0.81	6.07	10/20-60
(Biophysical variables)	Cla	0.76	6.67	10/20-60
	LAI	0.73	7.00	10/20-60

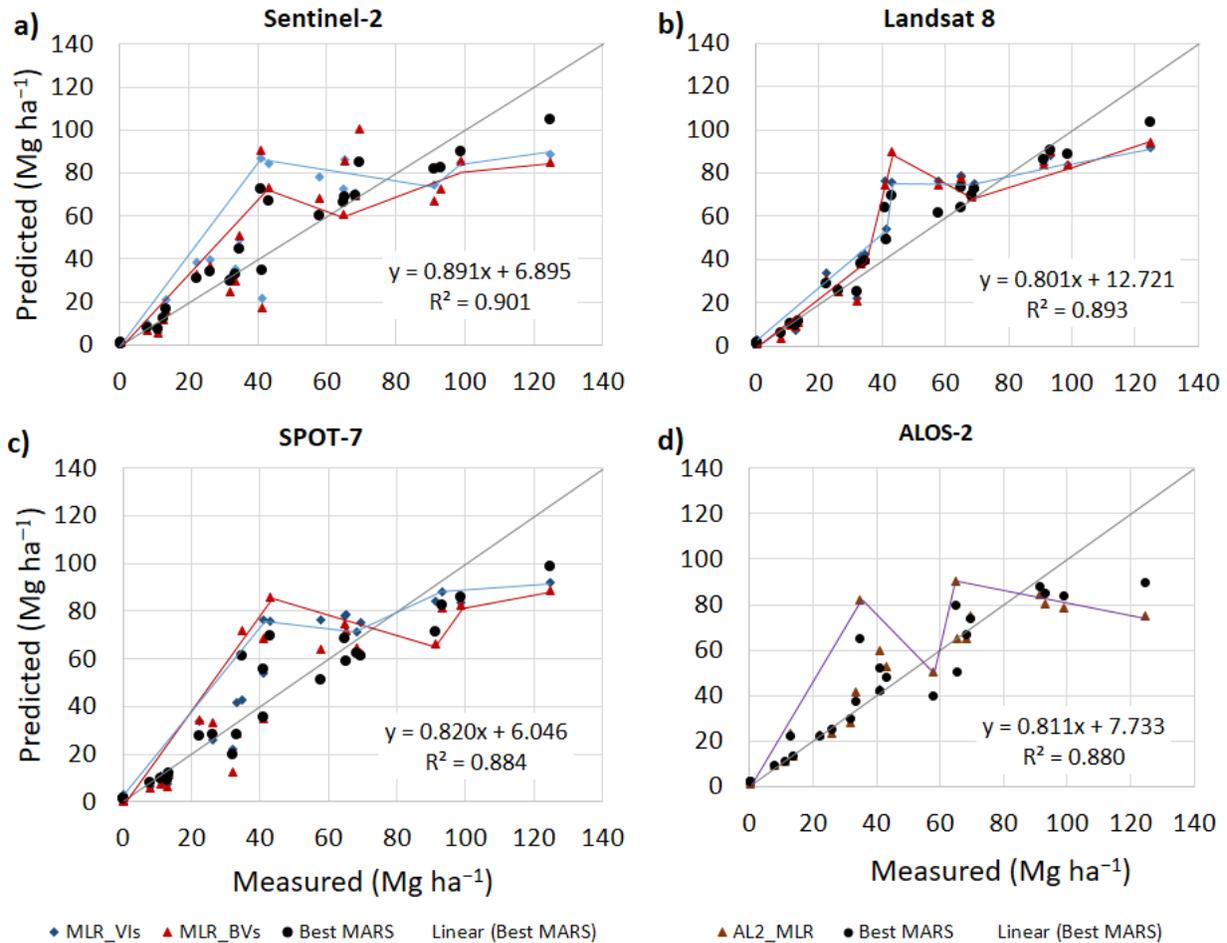
	FVC	0.64	7.81	10/20-60
	fAPAR	0.61	8.25	10/20-60
	MLR	0.80	6.42	10/20-60
Landsat 8 OLI (Vegetation indices)	SR*	0.70	6.66	10/30
	GNDVI	0.68	6.85	10/30
	NDVI	0.67	6.94	10/30
	SAVI	0.67	6.94	10/30
	SRre	NA	NA	NA
	MLR	0.80	5.56	10/30
(Biophysical variables)	LAI	0.70	6.66	10/30
	Cl _a	0.69	6.71	10/30
	FVC	0.69	6.76	10/30
	fAPAR	0.68	6.90	10/30
	MLR	0.82	5.27	10/30
SPOT-7 (Vegetation indices)	GNDVI	0.59	8.00	10/6
	SR	0.53	8.47	10/6
	NDVI	0.53	8.49	10/6
	SAVI	0.53	12.80	10/6
	SRre	NA	NA	NA
	MLR	0.66	7.67	10/6
(Biophysical variables)	LAI _{NDVI}	0.52	8.40	10/6
	FVC _{NDVI}	0.52	8.53	10/6
	CL _{aGreen}	0.51	8.43	10/6
	fAPAR	0.50	8.47	10/6
	MLR	0.65	7.33	10/6
ALOS-2 (polarization and ratios)	(HH×VH) ^{0.5}	0.54	8.21	10/3.125
	HH	0.50	8.57	10/3.125
	HH+VH	0.48	8.78	10/3.125
	VH	0.31	10.05	10/3.125
	MLR	0.60	7.75	10/3.125

* the best fit vegetation index, subsequently applied to the Landsat-X time-series

3.3. MARS model

3.3.1. Mangrove above-ground biomass estimated by MARS model

360 Fig. 5 shows that the MARS's MAGB (black circles) estimated for the Sentinel-2 image and Landsat 8 data had the highest agreement with the field measured data ($R^2 = 0.901$ for the former and $R^2 = 0.893$ for the latter). Using Landsat 8 and SPOT-7 (R^2 of 0.884) and ALOS-2 (R^2 of 0.880) scenes, the MARS model generated MAGB values slightly less accurate than the uses of Sentinel-2 and Landsat 8 images. Compared to the MLR, the MARS model performed better in terms of generating computationally complex models and predicting MAGB, shown by the black circles closer to the 1:1 diagonal line than the red triangles, blue diamonds and brown triangles of the MLRs. The MARS and 365 MLR results deviate most from the measured MAGB of 40 Mg ha⁻¹ using the optical Sentinel-2, Landsat 8 and SPOT-7 images, while the largest deviation using the SAR AL2 was at the measured MAGB of 35 Mg ha⁻¹. The MARS model over-estimated MAGB in the range of 40-60 Mg ha⁻¹ and under-estimated MAGB above 80 Mg ha⁻¹. The MARS model predicted MAGB well at < 40 Mg ha⁻¹ in all cases.



370 **Fig. 5.** The best MARS model results of mangrove above-ground biomass (back circles) on the vertical axis versus
the field measured data on the horizontal axis in comparison with multi linear regression (MLR) using vegetation
indices (light blue diamonds) and biophysical variables (red triangles) calculated from (a) Sentinel-2, (b) Landsat 8,
(c) SPOT-7 images and (d) using backscatter data (brown triangles) of ALOS-2. The red and blue polylines illustrate
the “hinge functions” with knots where the basic functions were changed and applied for each VI (red polylines) and
375 BV (blue polylines) and for the SAR AL2 data (violet polylines).

3.3.3. Accuracy evaluation for the MARS model

Correlations between the MARS MAGB and field measured data, and the most important variables in the MARS
model, are presented in Table 7. Variables are divided into three groups: vegetation indices and biophysical variables
for the optical remote sensing data, and polarizations and ratios for the ALOS-2 data. In general, use of Sentinel-2 and
380 ALOS-2 data produced the most accurate predicted MAGB, with R^2 of 0.9 and $RMSE < 5.33$ and 6.53 Mg ha^{-1} ,
respectively. The Landsat 8 image was the second most accurate and use of SPOT-7 achieved the lowest, although
still good, coefficient of determination (lowest $R^2 = 0.81$, highest $RMSE = 7.57$). Variable importance was identified
based on sensitivity to model outputs. In the vegetation index group, SR, NDVI and GNDVI were the most critical
variables, while SRre and SAVI were less influential on model outputs. In the biophysical variable group, LAI, Cla
385 and FVC were the top three important variables, while fAPAR was of limited use. The combination of SAR

polarization backscatter data (HH+VH and (HH×VH)^{0.5}) was more importance to model output than the single ALOS-2 polarization backscatter layer. Of the three groups of variables, the vegetation indices and SAR data were more correlated to the MARS MAGB than the biophysical variables.

Table 7

390 MARS's versus measured mangrove above-ground biomass, and MARS's important variables

Variables	Satellite data	R ²	RMSE (Mg ha ⁻¹)	Important variables
Vegetation Indices of (NDVI, SAVI, GNDVI, SR, SRre*)	Sentinel-2	0.90	5.33	SR, NDVI, GNDVI, SAVI, SRre
	Landsat 8	0.89	5.42	SR, GNDVI, NDVI, SAVI
	SPOT-7	0.81	7.57	GNDVI, SR, NDVI, SAVI
Biophysical variables of (LAI, fAPAR, FVC, Cla*, CLaGreen**))	Sentinel-2	0.89	5.27	Cla, LAI, FVC, fAPAR
	Landsat 8	0.87	6.42	LAI, Cla, FVC, fAPAR
	SPOT-7	0.85	7.24	LAI, FVC, CLaGreen, fAPAR
Polarizations and ratios (HH, VH, HH+VH and (HH×VH) ^{0.5})	ALOS-2	0.89	6.53	(HH×VH) ^{0.5} HH+VH, HH, VH

*not applicable for SPOT-7, ** for SPOT-7 only

3.4. Machine learning models

Twelve confusion matrices were calculated, examining the goodness of fit of ML classifications with field measured data. These are summarized by the mean producer, user accuracy, and Kappa coefficients in Table 8. Comparing between data sources, Sentinel-2 and Landsat-8 were generally better as most of their producer, user and Kappa coefficients (around 75%) were higher than those of SPOT-7 and ALOS-2 (approximately 63%). In addition, using ALOS-2 data generated MAGB with slightly higher accuracy than SPOT-7. Using higher spatial resolution data did not guarantee that the models would produce better MAGB predictions. Among the ML methods, the SVM produced the best MAGB estimates since all the accuracy indicators were higher than for ANN and RF. The ANN and RF generated MAGB estimates with similar accuracy (differences in R² just 3% and Kappa only 0.1).

400 **Table 8**

Accuracy assessment of machine learning approaches of Artificial Neural Network (ANN), random forest (RF) and Support Vector Machines (SVM) for mangrove above-ground biomass classifications

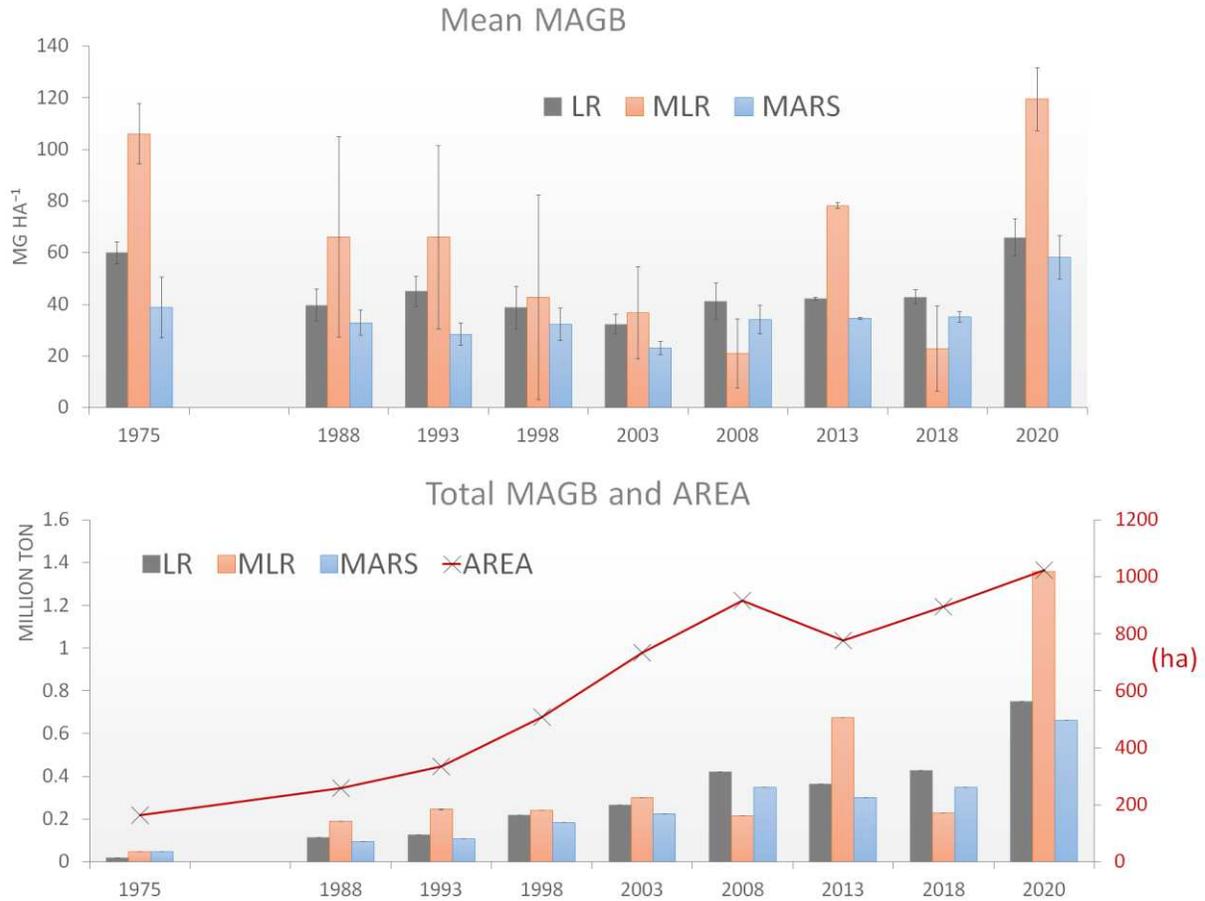
Data	ML methods	Mean producer accuracy (%)	Mean user accuracy (%)	Overall Accuracy (%)	Kappa Coefficient
Sentinel-2 (VI and BV combination)	ANN	59.970	63.771	75.352	0.519
	RF	51.767	43.320	53.535	0.440
	SVM	76.020	71.593	90.335	0.806

Landsat 8 (VI and BV combination)	ANN	64.308	65.652	67.600	0.589
	RF	70.765	76.358	74.568	0.683
	SVM	75.265	81.537	79.969	0.748
SPOT-7 (VI and BV Combination)	ANN	60.233	60.647	62.336	0.542
	RF	58.922	61.050	60.377	0.516
	SVM	63.072	59.392	67.243	0.595
ALOS-2 (HH, VH, HH+VH, (HH×VH) ^{0.5} combination)	ANN	61.504	63.357	68.429	0.550
	RF	60.422	60.344	62.704	0.545
	SVM	71.452	70.841	79.182	0.716

3.5. MAGB estimates for time-series Landsat images

405 The upper graph of Fig. 6 shows the fluctuation in mean unsupervised MAGB (Mg ha^{-1}) estimated by the calibrated linear regression (LR), multi linear regression (MLR) and MARS model over a 45-year time period. All the models estimated higher average MAGB for the 1980s, 1990s and 2020s, with lower average MAGB in 2010s with values of around 40 Mg ha^{-1} . Furthermore, the MLR estimated the highest MAGB, followed by LR and MARS. The standard deviation (STD) bars reveal the magnitude of variation in MAGB. The MLR produced MAGB with highest variation while there were no significant differences in STD between LR and MARS results.

410 The lower graph in Fig. 6 shows a consistent growth of the mangrove forest in terms of total MAGB (from under 0.1 to 1.3 million tons) and a tenfold (50 to 500 ha, based on MARS results) expansion in area, except for a decline in forest area in 2013 (even though total biomass still increased). The MLR repeatedly over estimated the total MAGB compared to the LR and MARS estimations.



415 **Fig. 6.** Time-series mean mangrove above-ground biomass (MAGB) value (Mg ha^{-1}) (upper graph) and total MAGB
 and forest area (lower graph) estimated by the linear regression (LR) using the simple ratios (SR) as the best vegetation
 index from Landsat 8, the best multi linear regression model (MRL) and the best Multivariate adaptive regression
 splines (MARS) model for the mangrove forest of Thuy Truong commune from 1975 to 2020.

3.6. MAGB maps

420 MAGB maps include results of LR for each VI, BV and SAR data layer, MLR and MARS model and the machine
 learning (ML) methods applied for each group of parameters and the maps of Landsat-X time-series images. Hence,
 in this section we focus on the most accurate results of the MARS, ML and Landsat-X time-series images.

3.6.1. MARS model

425 **Fig. 7** depicts the most accurate MAGB values using the vegetation indices calculated from (a) Sentinel-2, (b) Landsat
 8, (c) SPOT-7 images and (d) ALOS-2 backscatter data. It is clearly shown in all maps that MAGB varies spatially,
 and is highest in the center of the forest. Higher values of MAGB (greater than 100 Mg ha^{-1}) were concentrated in the
 south-west of the forest in maps generated using optical images. However, MARS generated more regular values
 across the entire ALOS-2 image. Landsat 8 and SPOT-7 produced lower MAGB values compared to Sentinel-2. The
 highest value for MAGB measured *in situ* was $124.77 \text{ Mg ha}^{-1}$ for plot 7 (not in the forest center). Higher values (up

430 to 160 Mg ha⁻¹) were estimated using the MARS model. These values were located in the forest center, an area that is extremely difficult to physically access because of high tree density.

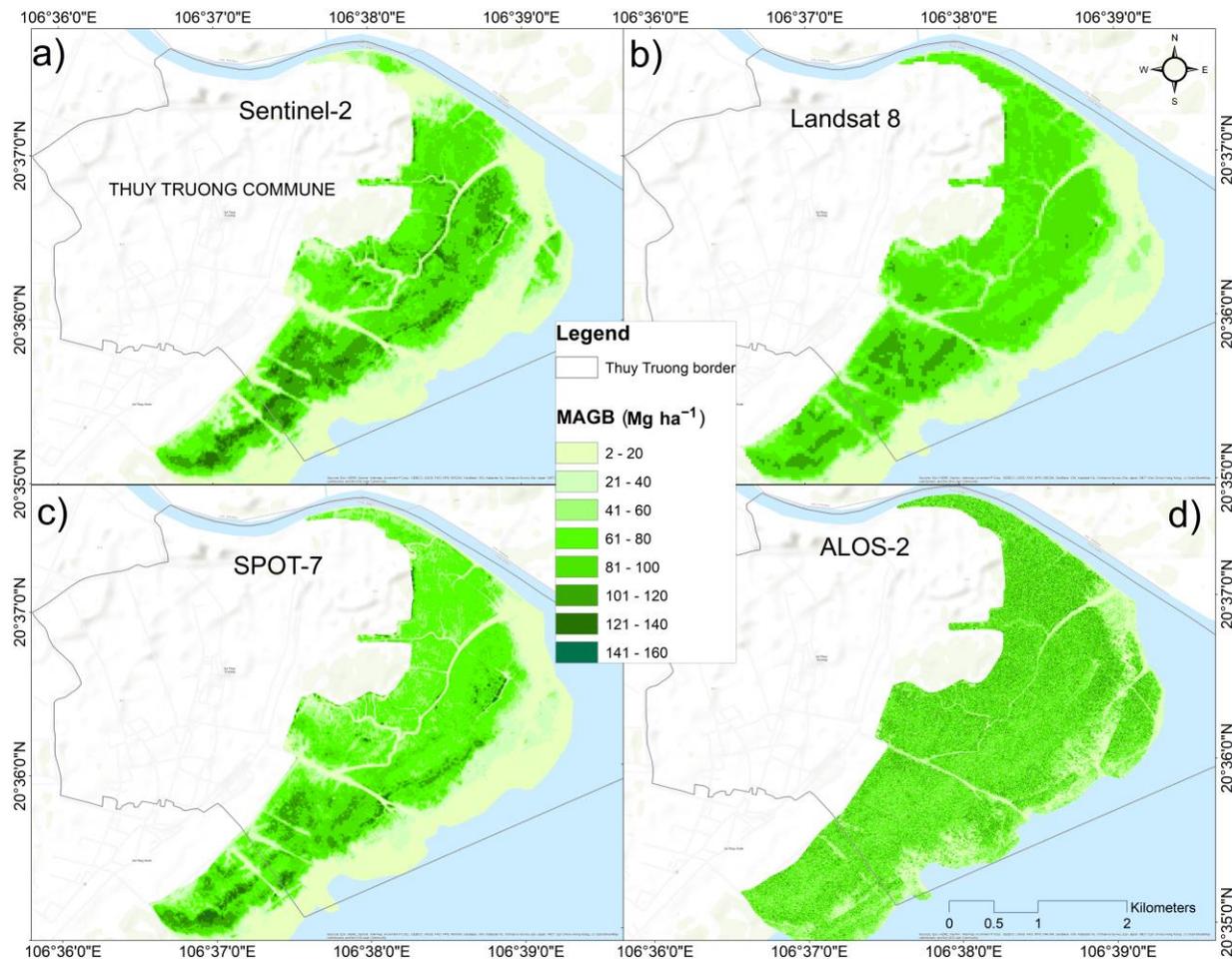
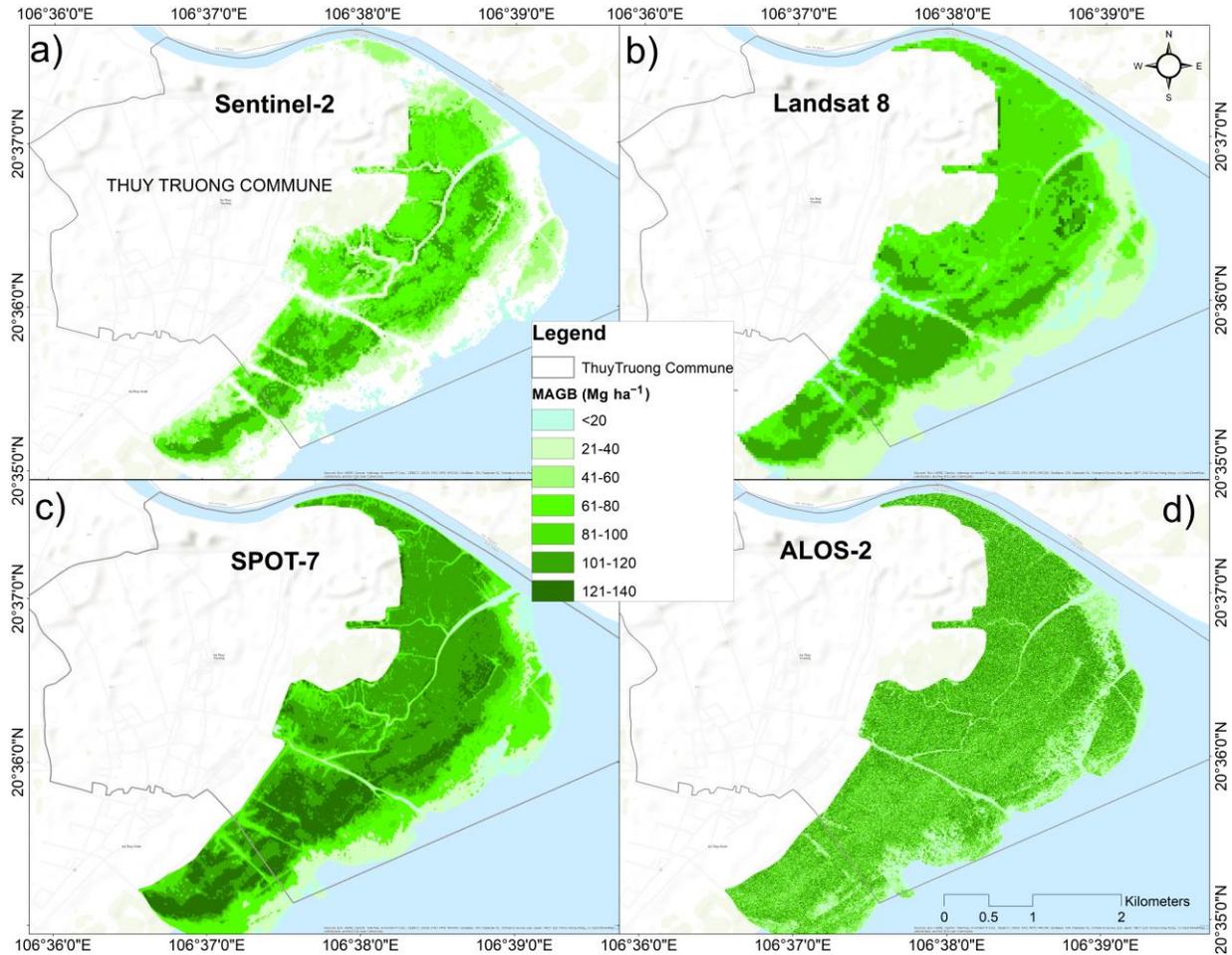


Fig. 7. Mangrove above ground biomass (Mg ha⁻¹) generated by the MARS model using (a) Sentinel-2, (b) Landsat 8, (c) SPOT-7 and (d) ALOS-2

435 3.6.2. Machine learning methods

We selected three machine learning algorithms for MAGB classification and the SVM presented the best performance (based on an accuracy assessment). Therefore, we only mapped the MAGB distributions of the SVM applied to the four remote sensing-based data, as shown in Fig. 8. The highest MAGB values were assigned for the SPOT-7 classification where large areas had values of above 120 Mg ha⁻¹. This may be an overestimate, while the result of Sentinel-2 was assessed as the most accurate prediction (overall accuracy of 90.335% and with lower MAGB values in general). Landsat 8 and ALOS-2 had medium values for MAGB. However, these two classifications had opposite MAGB distributions: higher values were predicted for the mangrove in the south of the Landsat 8 map and lower values were classified for the mangrove in the south of the ALOS-2 map.

440



445 **Fig. 8.** Mangrove above ground biomass (Mg ha^{-1}) generated by the machine learning algorithm of support vector machine (SVM) using combinations of vegetation indices and biophysical variables from (a) Sentinel-2, (b) Landsat 8, (c) SPOT-7 and (d) a combination of polarization of SAR ALOS-2 backscatters

3.6.3. Spatial changes in MAGB over a 45-year time period

450 **Fig. 9** shows an increase in both MAGB (Mg ha^{-1}) and mangrove extent from 1975 to 2020. Growth of the forest between 1975 and 1988 was gradual. Since 1993 the forest has been disturbed by the creation of aquaculture ponds, in the area marked by the red rectangle. Between 1993 and 2020, the forest grew quickly both in extent and MAGB, with a sharp increase between 2018 and 2020. By 2020, the mangrove had nearly disappeared in areas disturbed by aquaculture activities (red rectangle). At the same time the extent of area covered by mangrove in the east and south grew by approximately 80 m y^{-1} and MAGB increased in the center of the mangrove forest (darker green). Although

455 these are unsupervised estimations by the best MARS model, the model was trained using field measured MAGB data. As a result, while the unsupervised MAGB estimations may be uncertain, they were still considered as able to provide valuable information about changing mangrove biomass and extent.

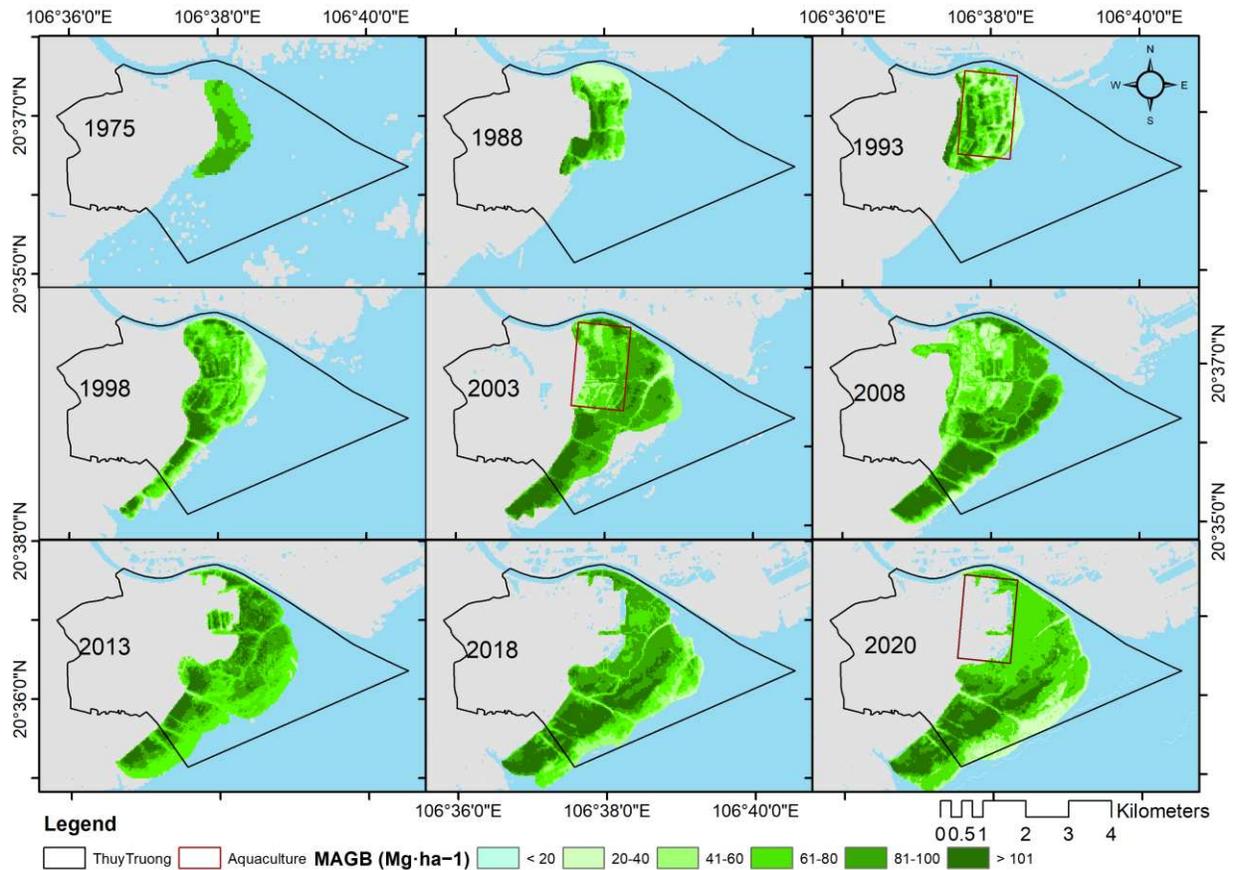


Fig. 9. Time-series of mangrove above ground biomass (Mg ha^{-1}) generated by best MARS model using Landsat 2, 5 and 8 images

4. Discussion

Despite advances in remote sensing, challenges remain. Complex stand structure and variety in species composition in mangroves leads above-ground biomass (AGB) estimation to be problematic (Steininger, 2000). The complexity of vegetation structures determines highly variable standing stocks of MAGB (Lu, 2006). Optical remote sensing sensors measure the reflectance or radiation emission from the vegetation canopy, hence uncertainty underneath the canopy should be taken into account. In struggling to overcome these limitations, research has to mine SAR remote sensing backscatters as some long wavelength SAR systems such as L and P bands penetrate through the canopy and have been found to be highly correlated with AGB (Kurvonen et al., 1999, Sun et al., 2002). However, each remote sensing system has pros and cons and the main issues with SAR data are the speckles known as noise and saturation, that can lead to underestimates of high AGB (Pham et al., 2019). This paper used different remote sensing data (optical and SAR) to compare the results from each method. Our findings offer insights to enable researchers to take the most appropriate approach according to the specific task they are addressing.

Besides finding appropriate data sources, researchers have developed and deployed a large number of methods to estimate AGB in general and MAGB in particular. The most common approaches involve parametric linear regression and these remain useful due to their simplicity and acceptable accuracy (Chave et al., 2005). However, parametric

linear regressions often suffer from rigidity and only consider individual parameters. Multi linear regression models gain advantages from group parameters and often claim better performance than linear regressions (Marill, 2004, Proisy et al., 2007, Soares and Schaeffer-Novelli, 2005). MAGB is complex and may not be in a linear relationship with vegetation parameters, so MAGB models have been developed to “break the lines” and improve the accuracy of
480 MAGB estimation. These efforts include the MARS model (Baloloy et al., 2018, Friedman, 1991, Friedman and Roosen, 1995). Currently, machine learning approaches have proven their effectiveness for MAGB estimation using multispectral imagery (Jachowski et al., 2013) and SAR data (Pham et al., 2020b, Pham et al., 2018). Our results demonstrated that the complex model of MARS and non-parametric machine learning algorithms outperform parametric regression models.

485 Accuracy assessment is often discussed because it determines model reliability. Uncertainty can result from the data used, methods and also the training data. It is sometimes difficult to determine whether the model performance is satisfactory or not (Hutchison et al., 2014) which will be dependent on both the aim and the scale (larger scales in general can accept lower accuracy). To evaluate model performance, it is common to use error-indicative indices, such as R^2 , RMSE, bias, or absolute error values based on reliable information, and commonly these are field measured
490 data. At continental and global scales, frequently R^2 and RMSE show that an accuracy of 50% can be acceptable (Rovai et al., 2016, Tang et al., 2018). In this study we obtained MARS R^2 values of around 0.9 for all four different remote sensing data sources and RMSE around 5 Mg ha⁻¹ at the commune scale with 900 ha mangrove forest. This was considered a satisfactory result and is supported by previous research in the same study site (Pham et al., 2020b). As uncertainty can come from early stage of image pre-processing, choosing an appropriate atmospheric correction
495 model is crucial. FLAASH presented a robust tool with options of inputting local parameters such as local atmospheric and aerosol model and applying for a wide range of hyperspectral and multispectral sensors. Although the surface reflectance data of Landsat and Sentinel-2 are available globally, it is recommended to apply suitable model with local parameter manually for higher accuracy. Besides, remote sensing image resampling techniques enable harmonization of different resolutions when using multiple data sources, but they can have effects on the final results. Hence these
500 techniques should be chosen with care in terms of methods and output resolutions.

Each model may be differently sensitive to a parameter and variable or combination of variables depending on the variation sensitivity of the remote sensing sensor to vegetation index and biophysical metrics (Dube and Mutanga, 2015). Hence, testing model sensitivities is a crucial step in the modelling process and offers useful guidance to enable
505 future studies to select the most important variables to improve model calibration times and the accuracy of outputs (Argamosa et al., 2018). We found that the vegetation indices of Sentinel-2 and Landsat 8 were the most sensitive variables to MAGB, which guided us to select the Landsat 8 VIs for the Landsat-x time-series MAGB extractions. However, we cannot be sure the L8 VIs are the best option for all study sites.

Finally, the increase in MAGB and mangrove forest extent in Thuy Truong commune identified through this work conflicts the general downward trend in mangrove cover in Vietnam, which has been estimated to have decreased by
510 some 62% since 1996 (Macintosh and Ashton, 2002). Extending this work to a national and regional scale would enable a more accurate picture of mangrove biomass and extent to be built, and support endeavors to protect and

conserve them for the benefits that they provide. It is also necessary to emphasize the benefits of the open Landsat archive, without which this work would be extremely difficult.

5. Conclusion

515 This work applied different approaches from simple to complex (linear regression to non-parametric machine learning) on different remotely sensed data (optical and SAR) to estimate mangrove above-ground biomass for a study site in northern Vietnam. We found that the MARS model produced the most accurate estimation of MAGB (with R^2 greater than 0.85, $RMSE \approx 5.5 \text{ Mg ha}^{-1}$ for Sentinel-2 and Landsat 8 images) than the linear and multi linear regression models (mean $R^2 \approx 0.56$ (LR), 0.7 (MLR), $RMSE \approx 6.5 \text{ Mg ha}^{-1}$ (LR), 7.0 (MLR). Of the ML approaches of ANN, RF and 520 SVM, the SVM generated more accurate MAGB estimates (highest overall accuracy of 90% and Kappa coefficient of 0.8) than ANN and RF. For both the MLR and MARS models, Sentinel-2 was most sensitive data to the ML SVM and produced the most accurate MAGB estimates. An unexpected outcome was that the employment of the higher spatial resolution of SPOT-7 and ALOS-2 did not generate the most accurate results. The best estimated MAGB from MARS and ML classifications were mapped for spatial comparison. In general, higher MAGB ($> 120 \text{ Mg ha}^{-1}$) was 525 concentrated in the core of the forest and the lower values ($< 60 \text{ Mg ha}^{-1}$) were found near to the sea. The results of MAGB and forest extent changes was modelled and mapped over 45 years from 1975 to 2020 showing the expansion of the mangrove forest both in extent and mangrove above-ground biomass. Application of the most accurate models identified through this work to national and regional scales could provide important evidence in support of mangrove forest management and conservation efforts.

530 **CRedit authorship contribution statement**

Nguyen Hong Quang: Conceptualization, Methodology, Investigation, Data analysis, Writing - original draft, Writing—review and editing, Resources, Software, Validation, Visualization. **Claire H. Quinn:** Conceptualization, Methodology, Project administration, Supervision, Writing—review and editing. **Rachael Carrie:** Methodology, Visualization, Writing—review and editing. **Lindsay C. Stringer:** Conceptualization, Methodology, Supervision, 535 Writing - review & editing. **Le Thi Van Hue:** Conceptualization, Project administration, Supervision, Writing - review & editing. **Christopher R. Hackney:** Investigation, Data analysis, Validation, Visualization, Writing - review & editing. **Dao Van Tan:** Methodology, Data analysis, Resources, Validation, Visualization, Writing - review & editing.

Acknowledgements: This research was financially supported by the Newton RCUK-SEAMED project “Harnessing 540 multiple benefits from resilient mangrove systems” funded by NAFOSTED RCUK, ESRC reference: ES/R003300/1.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- 545 Ali, M., Montzka, C., Stadler, A., Menz, G., Thonfeld, F., & Vereecken, H. (2015). Estimation and validation of RapidEye-based time-series of leaf area index for winter wheat in the Rur catchment (Germany). *Remote sensing*, 7(3),2808-2831.
- Argamosa, R. J. L., Blanco, A. C., Baloloy, A. B., Candido, C. G., Dumalag, J. B. L. C., DImapilis, L. L. C., & Paringit, E. C. (2018). MODELLING ABOVE GROUND BIOMASS OF MANGROVE
550 FOREST USING SENTINEL-1 IMAGERY. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 4(3).
- Avtar, R., Kumar, P., Oono, A., Saraswat, C., Dorji, S., & Hlaing, Z. (2017). Potential application of remote sensing in monitoring ecosystem services of forests, mangroves and urban areas. *Geocarto International*, 32(8), 874-885.
- 555 Baloloy, A. B., Blanco, A. C., Candido, C. G., Argamosa, R. J. L., Dumalag, J. B. L. C., Dimapilis, L. L. C., & Paringit, E. C. (2018). Estimation of Mangrove Forest Aboveground Biomass Using Multispectral Bands, Vegetation Indices and Biophysical Variables Derived from Optical Satellite Imageries: Rapideye, Planetscope and Sentinel-2. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, IV-3, 29-36. doi: 10.5194/isprs-annals-IV-3-29-2018
- 560 Bilgili, A. V., Van Es, H., Akbas, F., Durak, A., & Hively, W. (2010). Visible-near infrared reflectance spectroscopy for assessment of soil properties in a semi-arid area of Turkey. *Journal of Arid Environments*, 74(2), 229-238.
- Castillo, J. A. A., Apan, A. A., Maraseni, T. N., & Salmo, S. G. (2017). Estimation and mapping of above-ground biomass of mangrove forests and their replacement land uses in the Philippines using Sentinel imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 134, 70-85. doi: 10.1016/j.isprsjprs.2017.10.016
- 565 Chave, J., Andalo, C., Brown, S., Cairns, M. A., Chambers, J. Q., Eamus, D., . . . Kira, T. (2005). Tree allometry and improved estimation of carbon stocks and balance in tropical forests. *Oecologia*, 145(1), 87-99.
- 570 Clark, D. B., & Clark, D. A. (2000). Landscape-scale variation in forest structure and biomass in a tropical rain forest. *Forest ecology and management*, 137(1-3), 185-198.
- Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. *Remote sensing of environment*, 37(1), 35-46.
- Dube, T., & Mutanga, O. (2015). Evaluating the utility of the medium-spatial resolution Landsat 8 multispectral sensor in quantifying aboveground biomass in uMgeni catchment, South Africa. *ISPRS Journal of Photogrammetry and Remote Sensing*, 101, 36-46.
- 575 Friedman, J. H. (1991). Multivariate adaptive regression splines. *The annals of statistics*, 1-67 %@ 0090-5364.
- Friedman, J. H., & Roosen, C. B. (1995). An introduction to multivariate adaptive regression splines: Sage Publications Sage CA: Thousand Oaks, CA.
- 580 Gitelson, A. A., Gritz, Y., & Merzlyak, M. N. (2003). Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *Journal of plant physiology*, 160(3), 271-282.
- Gitelson, A., & Merzlyak, M. N. (1994). Spectral reflectance changes associated with autumn senescence of *Aesculus hippocastanum* L. and *Acer platanoides* L. leaves. Spectral features and relation to chlorophyll estimation. *Journal of plant physiology*, 143(3), 286-292.
- 585 Gitelson, A. A., Vina, A., Ciganda, V., Rundquist, D. C., & Arkebauer, T. J. (2005). Remote estimation of canopy chlorophyll content in crops. *Geophysical Research Letters*, 32(8).
- 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.
- 590 Green, E., Clark, C., Mumby, P., Edwards, A., & Ellis, A. (1998). Remote sensing techniques for mangrove mapping. *International Journal of Remote Sensing*, 19(5), 935-956.
- Herold, M., & Johns, T. (2007). Linking requirements with capabilities for deforestation monitoring in the context of the UNFCCC-REDD process. *Environmental Research Letters*, 2(4), 045025.

- 595 Hoa, N. H. (2016). Using Landsat imagery and vegetation indices differencing to detect mangrove change: A case in Thai Thuy district, Thai Binh province. *J. For. Sci. Technol*(5), 59-66.
- Hoque, A., Sharma, S., & Hagihara, A. (2011). Above and belowground carbon acquisition of mangrove *Kandelia obovata* trees in Manko wetland, Okinawa, Japan. *International Journal of Environment*, 1(1), 7-13.
- 600 Hu, T., Zhang, Y., Su, Y., Zheng, Y., Lin, G., & Guo, Q. (2020). Mapping the global mangrove forest aboveground biomass using multisource remote sensing data. *Remote Sensing*, 12(10), 1690.
- Huete, A. R. (1988). A Soil-Adjusted Vegetation Index (SAVI). *Remote Sensing of Environment*, 25, 295-309.
- Hutchison, J., Manica, A., Swetnam, R., Balmford, A., & Spalding, M. (2014). Predicting Global Patterns in Mangrove Forest Biomass. *Conservation Letters*, 7(3), 233-240. doi: 10.1111/conl.12060
- 605 Iqbal, M. H. (2020). Valuing ecosystem services of Sundarbans Mangrove forest: Approach of choice experiment. *Global Ecology and Conservation*, 24, e01273.
- Jachowski, N. R., Quak, M. S., Friess, D. A., Duangnamon, D., Webb, E. L., & Ziegler, A. D. (2013). Mangrove biomass estimation in Southwest Thailand using machine learning. *Applied Geography*, 45, 311-321.
- 610 Jordan, C. F. (1969). Derivation of leaf-area index from quality of light on the forest floor. *Ecology*, 50(4), 663-666.
- Kganyago, M., Mhangara, P., Alexandridis, T., Laneve, G., Ovakoglou, G., & Mashiyi, N. (2020). Validation of sentinel-2 leaf area index (LAI) product derived from SNAP toolbox and its comparison with global LAI products in an African semi-arid agricultural landscape. *Remote Sensing Letters*, 11(10), 883-892. doi: 10.1080/2150704X.2020.1767823
- 615 Kurvonen, L., Pulliainen, J., & Hallikainen, M. (1999). Retrieval of biomass in boreal forests from multitemporal ERS-1 and JERS-1 SAR images. *IEEE Transactions on Geoscience and Remote Sensing*, 37(1), 198-205.
- 620 Laffoley, D., & Grimsditch, G. D. (2009). *The management of natural coastal carbon sinks: Iucn*.
- Loan, N. T. N., Van Tan, D., Huyen, T. T. T., Quang, N. H., Hue, L. T. V., Nga, P. T. T., . . . Hackney, C. (2020). Comparison of several secondary metabolite and elemental ion contents of leaves from *Kandelia obovata* and *Sonneratia caseolaris* forests located in the Red River Delta. *Academia Journal of Biology*, 87-99.
- 625 Lu, D. (2006). The potential and challenge of remote sensing-based biomass estimation. *International journal of remote sensing*, 27(7), 1297-1328.
- Macintosh, D. J., & Ashton, E. C. (2002). A review of mangrove biodiversity conservation and management. *Centre for tropical ecosystems research, University of Aarhus, Denmark*.
- Marill, K. A. (2004). Advanced statistics: linear regression, part II: multiple linear regression. *Academic emergency medicine*, 11(1), 94-102.
- 630 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.
- Mazda, Y., Magi, M., Kogo, M., & Hong, P. N. (1997). Mangroves as a coastal protection from waves in the Tong King delta, Vietnam. *Mangroves and Salt marshes*, 1(2), 127-135.
- 635 Menéndez, P., Losada, I. J., Torres-Ortega, S., Narayan, S., & Beck, M. W. (2020). The global flood protection benefits of mangroves. *Scientific reports*, 10(1), 1-11.
- Ou, G., Li, C., Lv, Y., Wei, A., Xiong, H., Xu, H., & Wang, G. (2019). Improving aboveground biomass estimation of *Pinus densata* forests in Yunnan using Landsat 8 imagery by incorporating age dummy variable and method comparison. *Remote Sensing*, 11(7), 738.
- 640 Pham, M. H., Do, T. H., Pham, V.-M., & Bui, Q.-T. (2020a). Mangrove forest classification and aboveground biomass estimation using an atom search algorithm and adaptive neuro-fuzzy inference system. *Plos one*, 15(5), e0233110.
- Pham, Q. T. (2007). An analysis of soil characteristics for agricultural land use orientation in Thai Thuy District, Thai Binh Province.

- 645 Pham, T. D., Yokoya, N., Bui, D. T., Yoshino, K., & Friess, D. A. (2019). Remote sensing approaches for monitoring mangrove species, structure, and biomass: Opportunities and challenges. *Remote Sensing*, 11(3), 230.
- Pham, T. D., Yokoya, N., Xia, J., Ha, N. T., Le, N. N., Nguyen, T. T. T., . . . Takeuchi, W. (2020b). 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.
- 650 *Remote Sensing*, 12(8), 1334.
- Pham, T. D., Yoshino, K., Le, N. N., & Bui, D. T. (2018). Estimating aboveground biomass of a mangrove plantation on the Northern coast of Vietnam using machine learning techniques with an integration of ALOS-2 PALSAR-2 and Sentinel-2A data. *International journal of remote sensing*, 39(22),
- 655 7761-7788.
- Powell, N., Osbeck, M., Tan, S. B., & Toan, V. C. (2011). Mangrove restoration and rehabilitation for climate change adaptation in Vietnam. *World Resources Report Case Study. World Resources Report, Washington DC. URL: <http://www.worldresourcesreport.org>.*
- Proisy, C., Coueron, P., & Fromard, F. (2007). Predicting and mapping mangrove biomass from canopy grain analysis using Fourier-based textural ordination of IKONOS images. *Remote sensing of Environment*, 109(3), 379-392.
- 660 Quang, N. H., Quinn, C. H., Stringer, L. C., Carrie, R., Hackney, C. R., Hue, L. T. V., . . . Nga, P. T. T. (2020). Multi-Decadal Changes in Mangrove Extent, Age and Species in the Red River Estuaries of Viet Nam. *Remote Sensing*, 12(14), 2289.
- 665 Rosillo-Calle, F., De Groot, P., Hemstock, S. L., & Woods, J. (2015). *The biomass assessment handbook: energy for a sustainable environment*: Routledge.
- Rouse, J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring vegetation systems in the Great Plains with ERTS. *NASA special publication*, 351(1974), 309.
- Rovai, A., Riul, P., Twilley, R., Castañeda-Moya, E., Rivera-Monroy, V., Williams, A., . . . Crooks, S. (2016). Scaling mangrove aboveground biomass from site-level to continental-scale. *Global Ecology and Biogeography*, 25(3), 286-298.
- 670 Soares, M. L. G., & Schaeffer-Novelli, Y. (2005). Above-ground biomass of mangrove species. I. Analysis of models. *Estuarine, Coastal and Shelf Science*, 65(1-2), 1-18.
- Spawn, S. A., Sullivan, C. C., Lark, T. J., & Gibbs, H. K. (2020). Harmonized global maps of above and belowground biomass carbon density in the year 2010. *Scientific Data*, 7(1), 112. doi: 10.1038/s41597-020-0444-4
- 675 Steinger, M. (2000). Satellite estimation of tropical secondary forest above-ground biomass: data from Brazil and Bolivia. *International journal of remote sensing*, 21(6-7), 1139-1157.
- Sun, G., Ranson, K., & Kharuk, V. (2002). Radiometric slope correction for forest biomass estimation from SAR data in the Western Sayani Mountains, Siberia. *Remote sensing of Environment*, 79(2-3), 279-287.
- 680 Tang, W., Zheng, M., Zhao, X., Shi, J., Yang, J., & Trettin, C. C. (2018). Big Geospatial Data Analytics for Global Mangrove Biomass and Carbon Estimation. *Sustainability*, 10(2), 472.
- Tranmer, M., & Elliot, M. (2008). Multiple linear regression. *The Cathie Marsh Centre for Census and Survey Research (CCSR)*, 5, 30-35.
- 685 VEA. (2016). The Vietnam Environment Administration, Ministry of Natural Resources and Environment.
- Widlowski, J., Pinty, B., Gobron, N., Verstraete, M., Diner, D., & Davis, A. (2004). Canopy structure parameters derived from multi-angular remote sensing data for terrestrial carbon studies. *Climatic Change*, 67(2-3), 403-415.
- 690 Worthington, T. A., Andradi-Brown, D. A., Bhargava, R., Buelow, C., Bunting, P., Duncan, C., . . . Hilarides, L. (2020). Harnessing big data to support the conservation and rehabilitation of mangrove forests globally. *One Earth*, 2(5), 429-443.
- Yu, L., Liang, L., Wang, J., Zhao, Y., Cheng, Q., Hu, L., . . . Zhu, P. (2014). Meta-discoveries from a synthesis of satellite-based land-cover mapping research. *International Journal of Remote Sensing*, 35(13), 4573-4588.
- 695

- Yuan, H., Ma, R., Atzberger, C., Li, F., Loiselle, S. A., & Luo, J. (2015). Estimating Forest fAPAR from Multispectral Landsat-8 Data Using the Invertible Forest Reflectance Model INFORM. *Remote Sensing*, 7(6), 7425-7446.
- 700 Zhang, C., Chen, K., Liu, Y., Kovacs, J. M., Flores-Verdugo, F., & de Santiago, F. J. F. (2012). Spectral response to varying levels of leaf pigments collected from a degraded mangrove forest. *Journal of Applied Remote Sensing*, 6(1), 063501.