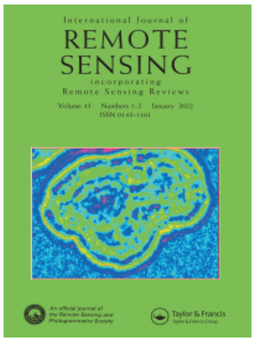
Quantifying mangrove leaf area index from Sentinel-2 imagery using hybrid models and active learning

Binh Nguyen, A.; Hauser, L.T.; Viet, H.P.; Thi Phuong, T.G.; An Nguyen, N.; Nhut Huynh, S.; ... ; Verrelst, J.



International Journal of Remote Sensing

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tres20

Quantifying mangrove leaf area index from Sentinel-2 imagery using hybrid models and active learning

Nguyen An Binh, Leon T. Hauser, Pham Viet Hoa, Giang Thi Phuong Thao, Nguyen Ngoc An, Huynh Song Nhut, Tran Anh Phuong & Jochem Verrelst

INTERNATIONAL JOURNAL OF REMOTE SENSING
https://doi.org/10.1080/01431161.2021.2024912

Taylor & Francis
Taylor & Francis Group

Quantifying mangrove leaf area index from Sentinel-2 imagery using hybrid models and active learning

Nguyen An Binh[a], Leon T. Hauser[b], Pham Viet Hoa[a], Giang Thi Phuong Thao[a], Nguyen Ngoc An[a], Huynh Song Nhut[a], Tran Anh Phuong[a] and Jochem Verrelst[c]

[a]Ho Chi Minh City Institute of Resources Geography, Vietnam Academy of Science and Technology, Ho Chi Minh, Vietnam; [b]Department of Environmental Biology, Institute of Environmental Sciences, Leiden University, Leiden, The Netherlands; [c]Image Processing Laboratory (IPL), Parc Científic, Universitat de València, Paterna, Valéncia, Spain

ABSTRACT

Mangrove forests provide vital ecosystem services. The increasing threats to mangrove forest extent and fragmentation can be monitored from space. Accurate spatially explicit quantification of key vegetation characteristics of mangroves, such as leaf area index (LAI), would further advance our monitoring efforts to assess ecosystem health and functioning. Here, we investigated the potential of radiative transfer models (RTM), combined with active learning (AL), to estimate LAI from Sentinel-2 spectral reflectance in the mangrove-dominated region of Ngoc Hien, Vietnam. We validated the retrieval of LAI estimates against in-situ measurements based on hemispherical photography and compared against red-edge NDVI and the Sentinel Application Platform (SNAP) biophysical processor. Our results highlight the performance of physics-based machine learning using Gaussian processes regression (GPR) in combination with AL for the estimation of mangrove LAI. Our AL-driven hybrid GPR model substantially outperformed SNAP (R^2 = 0.77 and 0.44 respectively) as well as the red-edge NDVI approach. Comparing two canopy RTMs, the highest accuracy was achieved by PROSAIL (RMSE = 0.13 $m^2.m^{-2}$, NRMSE = 9.57%, MAE = 0.1 $m^2.m^{-2}$). The successful retrieval of mangrove LAI from Sentinel-2 can overcome extensive reliance on scarce in-situ measurements for training seen in other approaches and present a more scalable applicability by relying on the universal principles of physics in combination with uncertainty estimates. AL-based GPR models using RTM simulations allow us to adapt the genericity of RTMs to the peculiarities of distinct ecosystems such as mangrove forests with limited ancillary data. These findings bode potential for retrieving a wider range of vegetation variables to quantify large-scale mangrove ecosystem dynamics in space and time.

ARTICLE HISTORY
Received 15 August 2021
Accepted 27 December 2021

CONTACT Nguyen An Binh nabinh@hcmig.vast.vn Ho Chi Minh City Institute of Resources Geography, Vietnam Academy of Science and Technology, 700000 Ho Chi Minh City, Vietnam

1. Introduction

Within the transition zone of land and sea of (sub)tropical coastal regions, mangroves have carved out a distinct niche to thrive and provide vital ecosystem services including the protection of coastal communities (Kuenzer and Tuan 2013; Brander et al. 2012). Mangroves are among the most productive and carbon-rich ecosystems worldwide (Donato et al. 2011; Siikamäki, Sanchirico, and Jardine 2012). Nonetheless, in many regions, mangrove forests are under severe pressure due to forest loss and land degradation caused by overexploitation and land-use change driven by human development (Duke et al. 2007; Giri et al. 2011). Remote sensing through airborne and satellite observations has become a primary instrument to monitor the health and dynamics of these ecosystems, especially given the inaccessible, dynamic and extensive nature of these mangroves that complicate frequent and extensive field visits (Hauser et al. 2020; Heumann 2011; Kuenzer et al. 2011).

Spatially explicit quantification of these ecosystems allows us to determine the extent, cover, and fragmentation of mangrove cover across regions, and serves as a reference for temporal comparisons (Hauser et al. 2020, 2017). While global maps of mangroves are available (Bunting et al. 2018; Giri et al. 2011), these maps are often limited solely to the presence or absence of mangroves (Younes Cárdenas, Joyce, and Maier 2017). To gain a better understanding of how the ecosystem reacts to environmental pressures, it is important to combine information on the extent and fragmentation of mangroves with biophysical variables such as forest density, leaf area index, chlorophyll content, and other functional attributes and vegetation characteristics (Pham et al. 2019). Mapping of biophysical variables that inform us about vegetation characteristics of mangroves would allow us to assess the ecosystems' health, phenology, functional attributes, and diversity thereof (Aguirre-gutiérrez et al. 2021; Lausch et al. 2016). However, such detailed indicators of ecosystem integrity are still rarely measured in mangroves at large spatial scales (Younes Cárdenas, Joyce, and Maier 2017).

One of the key biophysical variables in monitoring vegetation is Leaf Area Index (LAI) – defined as the area of leaf material per unit of ground surface area (Chen and Black 1992). LAI is strongly related to several key plant structural and functional variables such as the fraction of photosynthetically active radiation, leaf mass per area nitrogen content, biomass, aboveground net primary productivity, and stem density (Castillo et al. 2017; Fang et al. 2019; Yuan et al. 2015). Therefore, LAI is considered a fundamental vegetation attribute (Fang et al. 2019), both by itself as it is directly linked to primary productivity and competitive and complementary light use, transpiration, and energy exchange (Asner, Scurlock, and Hicke 2003; Zheng and Moskal 2009), as well as an important characteristic to scale up leaf traits to canopy traits (Asner, 1998). Furthermore, LAI is commonly used as an important modelling input for biosphere processes (Baret and Buis 2008). As such, LAI is recognized as an essential climate variable (GCOS 2011) and also proposed as an essential biodiversity variable (Skidmore et al. 2021).

In the last five decades, a broad variety of retrieval methods have been proposed and developed to estimate vegetation traits from earth observation data, including LAI. These methods range from vegetation indices (Delegido et al. 2011), to data-driven parametric and nonparametric regressions, including sophisticated machine learning methods, to physically based radiative transfer modelling and hybrid approaches, i.e. data-driven methods fed by Radiative Transfer Models (RTMs) (Pham et al. 2019; Verrelst et al. 2015).

Generally, the usage of statistical learning approaches to estimate biophysical vegetation properties depends heavily on comprehensive field measurements for training, which tend to be site- and time-specific (Verrelst et al. 2015). The in-situ measurements needed for data-driven approaches are scarce and difficult to obtain, especially for mangrove ecosystems and at the spatial scaling of satellite observations (Darvishzadeh et al. 2019b; Younes Cárdenas, Joyce, and Maier 2017).

To overcome the need for large but difficult-to-acquire field datasets to set up and train retrieval algorithms, the universal principles of the physics of light and its interaction with vegetation offer potential to simulate training datasets to complement costly in-situ field campaigns (Vane and Goetz 1988). The physical basis of light–vegetation interactions is commonly ensured through the exploitation of RTMs. RTMs relate incident radiation to vegetation canopies through a suite of angular, structural, biochemical, and biophysical characteristics (Jacquemoud et al. 2009; Jacquemoud 2019; Verhoef 1998). In turn, RTM simulations can generate training datasets to develop inversion models for retrieval of key vegetation characteristics from spectral reflectance. However, the universality of these models is bounded by strong assumptions and heavy parameterization, simplifying the heterogeneous canopies and vegetation types encountered in the field. Moreover, the practice of canopy RTM inversion to estimate plant traits from vegetation spectral reflectance is not trivial, but ill-posed and prone to a range of equally possible solutions, especially in multispectral settings (Combal et al. 2003; Koetz et al. 2007; Musavi et al. 2015).

Despite these challenges, LAI estimation based on RTM simulations has been shown to be viable when applied to satellite remote sensing in different (semi-)natural environments (e.g. LAI estimation using Sentinel-2; Brede et al. 2020; Brown, Ogutu, and Dash 2019; Darvishzadeh et al., 2019b; Padalia et al. 2020; Vinué, Camacho, and Fuster 2018). Yet, in the case of mangrove ecosystems, we observe that the retrieval of biophysical variables, including LAI, using optical satellite data has thus far mostly relied on vegetation indices (Kamal, Phinn, and Johansen 2016; Mafi-Gholami et al. 2019; Manna and Raychaudhuri 2020; Parida and Kumari 2020) or machine learning (data-driven) approaches that depend on large training dataset based on in-situ measurements (Castillo et al. 2017; Pham et al. 2019; Zhu et al. 2017).

One challenge and uncertainty to overcome is that commonly applied RTMs, such as PROSAIL (Jacquemoud et al. 2009) or INFORM (Atzberger 2000), have not been developed for mangrove ecosystems. Mangrove ecosystems have very specific conditions; inundated, muddy and tidal forests with limited species with a distinct heterogeneous canopy structure and minimal understory growth (Kuenzer et al. 2011). At present, little is known about the performance of commonly used RTMs to train models for retrieval of LAI in mangrove ecosystems from broadband satellite reflectance data. The development of active learning (AL) approaches in hybrid retrieval methods offers promising and adaptive solutions to overcome the common mismatch between the genericity and assumptions of RTM models against noisy real-world spectra while accounting for the particularities of ecosystems such as mangroves. In particular, AL heuristics involve intelligent sub-sampling pathways for fast inference of biophysical variables (Berger et al. 2021). AL-based sub-sampling intuitively selects the most informative samples out of large training RTM simulated data pools. Such an approach facilitates optimization of the regression accuracy based on an automated selection of simulations and trait ranges that are actually useful for the ecosystem under study (mangroves) and discards redundant information that might inflate ill-posedness (Berger et al. 2021).

In this study, we aim to extend the use of RTM simulations for the training of a hybrid model to estimate LAI from actual Sentinel-2 imagery in a typical mangrove ecosystem. We used AL heuristics to feed our model with an optimized simulation subset for mangrove ecosystems to overcome the mismatch between the genericity of RTMs applied and the specificity of the ecosystem under study. To examine its performance, we studied and mapped the spatial LAI patterns against precisely matched field data in the mangrove-rich Ngoc Hien District, Ca Mau province in the Vietnamese Mekong Delta.

Our implemented approach uses AL-based hybrid models based on simulations originating from two canopy RTMs: INFORM (Atzberger 2000) and PROSAIL (Jacquemoud et al. 2009). We compare the performance of AL-based estimates against a vegetation index (red-edge NDVI), retrieval of LAI through the pre-trained the Sentinel Application Platform (SNAP) global biophysical processor (Weiss and Baret 2016), and most importantly against in-situ hemispherical field measurements of LAI. The assessments open up discussion on the potential and challenges of AL-based hybrid models for the retrieval of canopy traits in mangrove ecosystems across larger scales with limited ancillary (training) data to support better ecological monitoring of mangroves.

2. Methods

Figure 1 represents the proposed methodology for quantifying mangrove leaf area index from Sentinel-2 imagery using hybrid models and active learning. Each step is further detailed in the sections below.

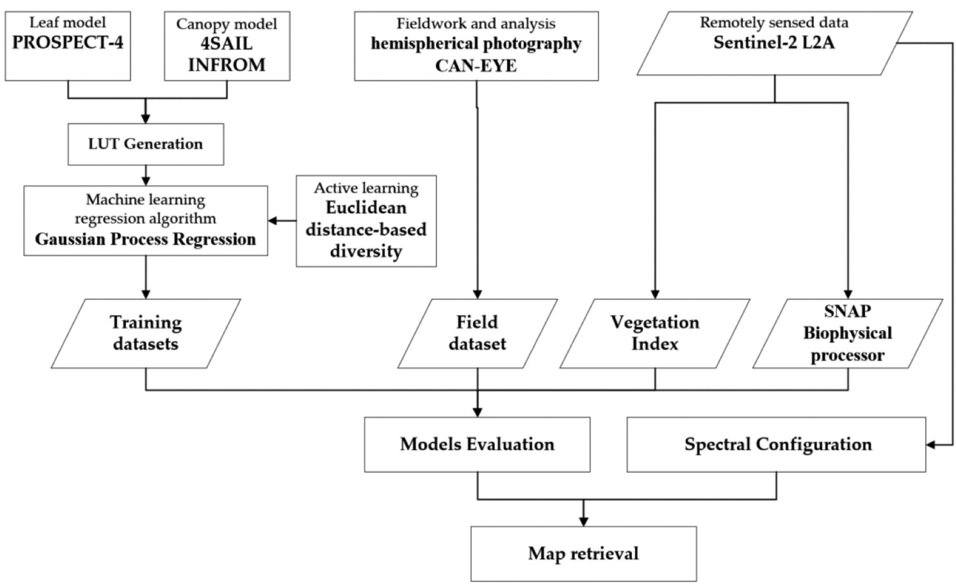


Figure 1. Workflow for mangrove leaf area index retrieval using hybrid models and active learning.

2.1. Study area

The study area is located in Vietnam's southernmost district, Ngoc Hien, Ca Mau province, situated in the Southern Mekong Delta between latitude 8°33′–8°45′N and longitude 104° 42′45″–105°3′54″E, spanning an area of 743 km^2 (see Figure 2). The district has been well studied for its importance as a major aquaculture hub and its significant reserves of Vietnam's largest and last remaining old-growth mangrove forests, including the internationally acknowledged RAMSAR site of Mui Ca Mau (2012) and UNESCO Biosphere Reserve (2009) (Ha, van Dijk, and Visser 2014; Tue et al. 2014). The region therefore plays a key role for conservation and provision of a range of flora, fauna and ecosystem services found in mangroves within Vietnam (Quoc Vo et al., 2015). The landscape supports both ecologically important mangrove ecosystems and the economic livelihoods based on aquaculture. The most common mangrove species listed in the region include *Avicennia alba*, *Avicennia officinalis* and *Rhizophora piculata* (Nguyen et al. 2020). From the observation of optical satellite sensors, *Rhizophor aapiculata* plantations and natural areas are most dominant.

2.2. Trait measurements

Fieldwork was carried out from 6 to 10 January, 2021. In the study area, a total of 65 randomly distributed plots with 10 × 10 m were precisely scaled to match Sentinel-2's resolution and pixel raster using GPS Trimble. For each of the plots, we used hemispherical photography to specify in-situ LAI similar to approaches by (Garrigues et al. 2008; Weiss et al. 2004). For consistent measurements across all sites, we took three hemispherical photos per plot that we combined to plot-wise mean LAI measurements. The hemispherical images were taken using a RICOH THETA M15 device. Images were retaken

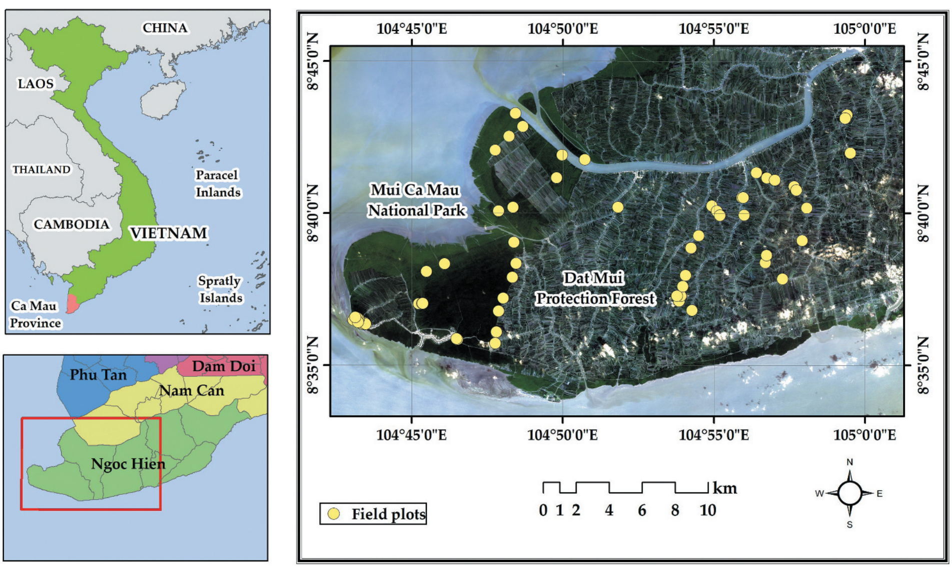


Figure 2. Location of study area.

in case of the presence of sunbeams or sun fleck problems. After the field campaign, we processed the three RGB hemispherical photographs using CAN-EYE v6.495 open-source software to retrieve effective LAI measurements (CE V6.1) (Weiss and Baret 2010).

2.3. Sentinel-2 data

The Sentinel-2 satellite constellation (2A and 2B) consists of two wide-swath, medium-high spatial resolution (10, 20, and 60 m), multi-spectral (13 bands) optical imagers with a combined 5–10 days revisit time (ESA 2015). The location of our study area in the humid subtropical region of Southern Vietnam implies year-long high occurrence of cloudy conditions. Despite frequent cloudiness, one Sentinel 2B Level 2A image was selected closely after the field work campaign (25 January 2021) which was largely free of suboptimal conditions of acquisition (e.g. presence of clouds, cirrus, shadows, aerosols). Among 13 bands with different resolutions ranging from 10–60 m, 10 bands were selected and the spectral information was extracted along with the field information to create the field dataset. We excluded the 60 m broadbands from the analysis. The ten remaining bands were resampled to 10 m (following Brodu 2017) in order to sync with the field-designed plot in the area (Table 1).

2.4. LUT generation using radiative-transfer model

We used radiative-transfer modelling for generation of a Look-up Table (LUT) training database. The training database allowed us to model the relevant relationships between spectra, geometry, and the soil and vegetation biophysical variables including LAI in which we are interested in for retrieval. We used both PROSAIL (Jacquemoud et al. 2009) and INFORM (Schlerf and Atzberger 2006) as RTMs to simulate canopy-scale observations.

PROSAIL combines the leaf model PROSPECT (Jacquemoud and Baret 1990) and the canopy model 4SAIL (Verhoef 1984). PROSAIL assumes the canopy to be a homogeneous turbid medium where absorption is defined by soil, canopy, and leaf properties (Jacquemoud et al. 2006). Such homogeneous canopies are a simplification of the complex heterogeneous architecture of mangrove forests. Therefore, simulations are subject to discrepancies between underlying assumptions of the extended 1-D columnar model and the complex reality of the heterogeneous canopies observed in the field (Jacquemoud et al. 2009).

Table 1. Sensor information for Sentinel 2 including full width at half maximum (FWHM) and signal-to-noise ratio (SNR) (ESA 2015).

Band	Min	Max	Center	FWHM	SNR
2 (Blue)	457.5	522.5	490	65	154
3 (Green)	542.5	577.5	560	35	168
4 (Red)	650	680	665	30	142
5 (Red edge 1)	697.5	712.5	705	15	117
6 (Red edge 2)	732.5	747.5	740	15	89
7 (Red edge 3)	773	793	783	20	105
8 (NIR)	784.5	899.5	842	115	174
8A (Red edge 4)	855	875	865	20	72
11 (SWIR1)	1565	1655	1610	90	100
12 (SWIR2)	2100	2280	2190	180	100

The Invertible Forest Reflectance Model 'INFORM' (Schlerf and Atzberger 2006) is a combination of the forest light interaction model (Rosema et al. 1992) and SAIL (Verhoef 1984) canopy RTMs with the PROSPECT leaf RTM (Jacquemoud and Baret 1990). INFORM considers the one-dimensional turbid medium radiative-transfer within the crowns and the three-dimensional characteristics such as clumping of the leaves and shadows of the crowns (Schlerf and Atzberger 2006). INFORM offers an appealing trade-off between the realism of simulation of the forest canopy and inversion feasibility (Darvishzadeh et al. 2019a). Its effectiveness in modelling LAI has been demonstrated in both broadleaf and conifer stands with varying levels of success (Brown, Ogutu, and Dash 2019; Schlerf and Atzberger 2006; Yuan et al. 2015). To our knowledge, INFORM has not yet been applied to the retrieval of plant traits in mangrove forests.

Here, we ran the RTMs models PROSAIL (PROSPECT4 and 4SAIL) and INFORM (PROSPECT4 and INFORM) to generate a LUT training database. Its execution is largely automated in the ARTMO toolbox (Verrelst, Romijn, and Kooistra 2012). For both models, we established a training set of 1500 different combinations from randomly drawing all from the ranges of the input parameters using a Latin Hypercube Sampling method. The parameterization (e.g. input for LUT simulations) of both PROSAIL and INFORM is described in Table 2. We used the full default ranges available in ARTMO. In addition, fixed parameters were collected from satellite image metadata, which consist of solar zenith angle 36.6°, observer zenith angle 3.4°, and relative azimuth angle 10°. The spectral characteristics of the generated LUTs were adapted according to the band selection and spectral layout of Sentinel-2 (see Table 1).

2.5. Machine learning regression algorithm (MLRA) with active learning (AL)

Sentinel-2 L2A reflectance data served as the foundation for estimating LAI. The generated LUTs served a hybrid inversion approach to deduce estimated biophysical variables corresponding from the Sentinel-2 spectra based on machine learning regression

Table 2. Input variable for PROSPECT-4, 4SAIL and INFORM.

Variables	Abbreviation	Units	Range	Distribution
PROSPECT 4				
Leaf structure	N		1.3–1.6	Uniform
Chlorophyll a + b	Cab	ug/cm^2	0–100	Uniform
Water thickness	Cw	g/cm^2	0.0001–0.08	Uniform
Dry matter	Cm	g/cm^2	0.0001–0.05	Uniform
4SAIL				
Leaf area index	LAI	m^2.m^{-2}	0–6	Uniform
Average leaf angle	ALA	degree	0–90	Uniform
Diffuse/Direct radiation			0–100	Uniform
Soil brightness	psoil		0–1	Uniform
Hot spot effect	hot		0–1	Uniform
INFORM				
Single tree leaf area index	LAI	m^2.m^{-2}	0–4	Uniform
Stem density	SD		2000–3000	Uniform
Tree height	H	m	15–20	Uniform
Scale factor for soil reflectance	Sky	Fraction	0–1	Uniform
Leaf area index of understorey	LAI_{under}		0–3	Uniform
Average leaf angle	ALA	degree	15–75	Uniform
Crown diameter	CD	m	2–4	Uniform
Fraction of diffuse radiation	Sky	Fraction	0–1	Uniform

algorithms. To train our hybrid model, we relied on Gaussian Process Regression (GPR), which is non-parametric regression modelling framed in a Bayesian inference (Rasmussen and Williams 2006). GPR relies on a pre-set covariance kernel functions that is optimized to fit the given data. The model was fitted with our RTM-generated LUT training database, i.e. biophysical variables linked with spectral reflectance data.

Despite the advantages of hybrid methods (see Verrelst et al. 2019), the inversion of RTMs using real-world spectral data, like our Sentinel-2 observations, remains ill-posed; multiple configurations of leaves and/or canopy variables can produce identical or similar spectral responses (Verrelst et al. 2016a). This problem is further amplified when the number of bands is limited or by the presence of noise (de Sá et al. 2021). Commonly, ill-posedness is minimized through a reasonable pre-selection of expected biophysical trait range of values, but this has an obvious implication on the generality and scalability of the trained models (Verrelst et al. 2019, 2015). Pre-selection is difficult in situations where no field data is available and relevant trait ranges are unknown.

Active learning (AL) methods integrate new samples based on uncertainty or diversity criteria to improve the accuracy of the model in the context of Earth observation regression problems (Berger et al. 2021). AL approaches have been shown to enhance retrieval accuracy and mitigate some of the ill-posedness within hybrid inversion methods (Berger et al. 2021). Hence, rather than using the full LUT, we used an AL method based on a GPR to select the most informative spectral samples and thus develop a subset relevant for further training, e.g. a machine learning GPR-based 'pre-selection.'

The employed AL heuristics started with an initially annotated dataset (30 samples or 2%), which was incrementally extended by choosing from the large data pool (1500 simulations). The 2% initialization is randomly selected. Since only a few samples were used as a starting pool (30 samples), its influence is relatively negligible as it maximizes the role of the AL algorithm. Earlier studies demonstrated promising results when keeping the initialization low (see Berger et al. 2021; Verrelst, Berger, and Rivera-Caicedo 2020).

The AL algorithm assumes the simulated training database as an unlabelled data pool, and hence iteratively tests a new sample according to a pre-defined query strategy, in our case Euclidean distance-based diversity (EBD). EBD was selected based on previous demonstration of high accuracy with an apt running time and small number of samples (Verrelst, Berger, and Rivera-Caicedo 2020; Verrelst et al. 2016a).

In the AL iteration, a new sample is only added when it fulfils the requirements to improve the regression model (reduce RMSE against validation data); otherwise, the algorithm proceeds to evaluate the next sample. Optionally, a stopping criterion can be defined, e.g. terminating after 300 samples. Taken together, the AL algorithm using EBD was applied, as implemented in ARTMO, to curate the simulated dataset to provide the most informative samples as well as reduce the size of the dataset. We ran 1501 iterations with the EBD algorithm for both PROSAIL and INFORM to ensure consistency. Detailed information on the MLRA and AL procedures are presented in Table 3.

2.6. Model performance and evaluation

The initial pre-AL training dataset (1500 simulated samples) was generated based on the range of input parameters of the RTMs (Table 2). Then, the initial training dataset was optimized by using AL to select the most informative samples. Finally, the resulting

Table 3. Input parameters of MLRA and AL.

MLRA	
MLRA approaches	Gaussian Processes Regression – Matlab
Parameter Gaussian Noise	0–3
Spectral Gaussian Noise	0–3
Starting data LUT	2%
Active learning	
Method	Euclidean Diversity (EBD)
Number of samples to add in iteration	1
Stop at max number of samples	300

reduced dataset was used for training the final GPR model, and ultimately to map LAI from Sentinel-2 data. The retrieved LAI estimates were then validated against the in-situ measurements using hemispherical photography. For validation, we assessed its performance by comparison of the remotely sensed LAI estimates against the entirety of the in-situ LAI measurements (n = 65).

The LAI estimated and measured values were evaluated based on four commonly applied metrics (Richter and Hank 2021): Pearson's linear correlation coefficient (R), mean absolute error (MAE), root mean square error (RMSE), as well as the normalized root mean square error (NRMSE) were computed as the equation given in (1) – (4)

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_{\mathrm{estimated}} - y_{\mathrm{measured}}\right| \tag{1}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{\mathrm{estimated}} - y_{\mathrm{measured}}\right)^2} \tag{2}$$

$$\mathrm{NRMSE} = \frac{\mathrm{RMSE}}{\max(y_{\mathrm{measured}}) - \min(y_{\mathrm{measured}})} \tag{3}$$

$$R^2 = \left(\frac{\sum_{i=1}^{n}\left(y^i_{\mathrm{estimated}} - \overline{y_{\mathrm{estimated}}}\right)\left(y^i_{\mathrm{measured}} - \overline{y_{\mathrm{measured}}}\right)}{\sqrt{\sum_{i=1}^{n}\left(y^i_{\mathrm{estimated}} - \overline{y_{\mathrm{estimated}}}\right)\sum_{i=1}^{n}\left(y^i_{\mathrm{measured}} - \overline{y_{\mathrm{measured}}}\right)}}\right)^2 \tag{4}$$

where, n is the number of samples, $y_{\mathrm{estimated}}$ is the LAI values estimated from satellite data and, y_{measured} is the LAI values measured in-situ.

Based on these metrics, we evaluated both AL-based hybrid RTM inversion approaches, trained on either PROSAIL or INFORM LUT databases. Besides the AL approaches, we ran two commonly applied and relatively straightforward alternative approaches for cross-comparison; (1) the use of a modified NDVI (Normal Difference Vegetation Index) to derive LAI (Delegido et al. 2011), and (2) the use of the ESA's Sentinel Application Platform (SNAP) biophysical processor (Weiss and Baret 2016).

Vegetation indices are among the oldest and most widely used tools to estimate LAI (Delegido et al. 2011; Fang et al. 2019; Zheng and Moskal 2009; Zhu et al. 2017). Vegetation indices are simple numerical indicators that reduce multispectral (two or more spectral bands) data to a single variable for predicting and assessing vegetation characteristics. In the analysis by (Delegido et al. 2011), the authors proposed a red-edge NDVI index. This red-edge NDVI with one band in the red-edge instead of the NIR is sensitive to green LAI from Sentinel-2 data, as the equation given in (5):

$$LAI = a \times \left(\frac{R_{706} - R_{664}}{R_{706} + R_{664}} \right) \quad (5)$$

where a is the experimental coefficient, R_{664} and R_{706} represent the surface reflectance from band 4 and band 5 in Sentinel-2 (see Table 1).

In addition to vegetation indices, we compared the performance of the AL-based inversion approaches against LAI estimates based on the SNAP biophysical processor. SNAP too relies on a hybrid approach combining physical modelling and machine learning. SNAP uses an artificial neural network (ANN) inversion pre-trained on a PROSAIL simulated database including canopy reflectance and the corresponding set of input parameters. The value, range and distribution followed for each input parameter of the models are fixed and described in (Weiss and Baret 2016). SNAP can be considered as the benchmark for the estimation of vegetation biophysical variables, as it is publicly available and easily applicable without strong expertise. SNAP includes an unreleased version of PROSPECT prior to PROSPECT-4, coupled with the SAIL model (Fourty and Baret 1997).

2.7. Generating full LAI maps

Ultimately, the best-performing GPR algorithm for LAI estimation was applied to map spatial LAI patterns for our study area. An appealing property about the use of the GPR implementation is that, in addition to mapping LAI estimates, it allowed us to map the absolute and relative uncertainty of the retrieval given the trained model for which we produced respective maps. The initial AL-based GPR model was solely trained on ranges that are presumed to be vegetation spectra; however, of course, our study area also consists of non-vegetated areas. To map the full area, we added reflectance data of major land cover types in the study area to the training dataset. These include spectra from built-up areas, water, aquaculture, and bare soil. We proceeded to map LAI for the entire study area using the full field dataset for training the model and applied a range of noise (between 0% and 3% of the observations). Pixels containing clouds and shadow cloud cover were few in our selected image, and those remaining contaminated pixels were masked. All locations sampled in-situ were cloud-free.

3. Results and analysis

3.1. Active learning performance

Overall, the active learning EBD algorithm combined with GPR for both PROSAIL and INFORM training datasets resulted in sharp initial improvements in performance (RMSE, R^2) as the subset for training grew (Figure 3). Notably, both models exhibited significant improvements of RMSE and MAE already within the first 20 samples. The final subset comprising the most informative samples for training in PROSAIL and INFORM consisted

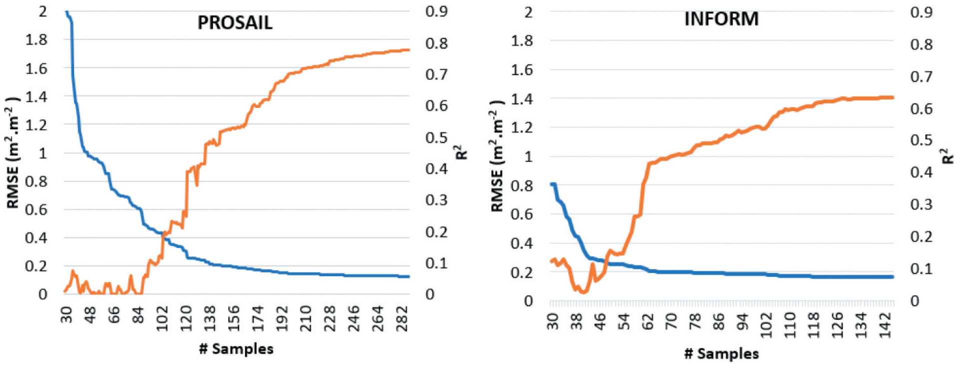


Figure 3. Progress of active learning algorithm estimating LAI using EBD with PROSAIL (left) and INFORM (right) simulations for training.

of 258 and 115 samples, respectively. The higher number of selected samples might be indicative of the suitability of the model to the forest structure of the mangroves in the study area.

3.2. Model evaluation

We evaluated LAI estimates retrievals trained on RTM simulations, PROSAIL and INFORM, respectively, against in-situ measurements of LAI. Figure 4(a,b) overview the two AL-based GPR approaches based on PROSAIL and INFORM simulations. Across all performance metrics, AL-based PROSAIL (GPR) resulted in the most accurate estimates (R^2: 0.77, RMSE: 0.13 $m^2.m^{-2}$ and nRMSE: 9.6%) (Figure 4a). AL-based INFORM (GPR) retrieval performed slightly less accurately (R^2: 0.66, RMSE: 0.16 $m^2.m^{-2}$ and nRMSE: 12.2%) (Figure 4b).

In addition to AL-based approaches, we compared retrievals derived from SNAP (Figure 4c) and a red-edge NDVI index used to estimate green LAI (Delegido et al. 2011) (Figure 4d) to assess how the AL-based approaches compare in its capability to accurately estimate in-situ observations to common alternative methods. For SNAP, we still observed an acceptable estimation of the relative distribution of LAI in our study area ($R^2 = 0.45$). However, in terms of bias, SNAP estimates of LAI were far off from the actual in-situ measurements (RMSE = 2.19 $m^2.m^{-2}$ and nRMSE: 163.2%) (Figure 4c). Both AL-based methods of PROSAIL and INFORM inversion performed dramatically better compared to SNAP. The implementation of vegetation indices, in this case red-edge NDVI green LAI estimator, appeared relatively unsuccessful, returning a relationship of $R^2 = 0.2$ to the in-situ LAI observations.

3.3. Mapping LAI spatial variability in study area

The best performing RTM, i.e. PROSAIL, using a hybrid approach based on EBD-GPR was applied for full mapping of LAI in the study area. This resulted in maps of the spatial patterns of LAI across the entire study area derived from pixel-based mangrove forest canopy reflectance retrieved from Sentinel-2 imagery.

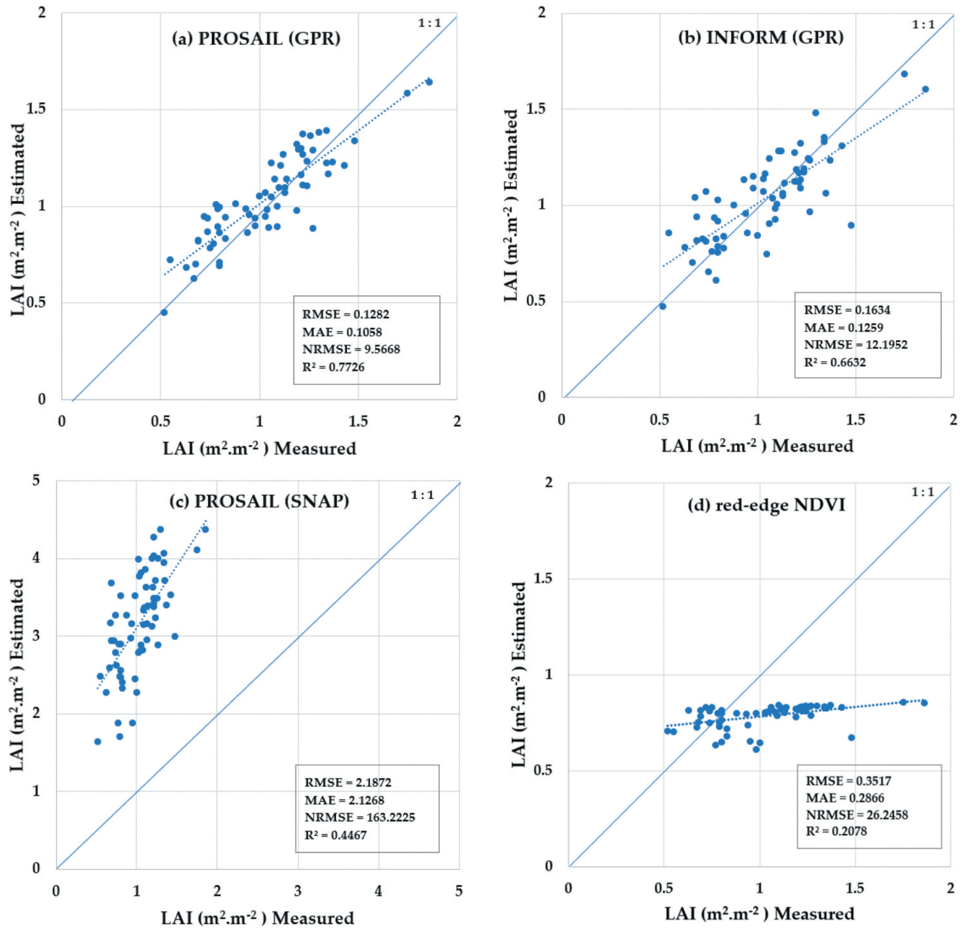


Figure 4. Performance of measured vs estimated data: (a) PROSAIL(GRP), (b) INFORM(GPR), (c) PROSAIL (SNAP) and (d) red-edge NDVI.

LAI map retrieval from the model trained on PROSAIL simulations shows high LAI values (>2) in the protected core forest in the Ca Mau National Park (Figure 5a). Lower values (<2) dominate the majority of the study region located outside the Ca Mau National Park (Figure 5a). This includes the large part of the study area belonging to Dat Mui region, which is characterized by integrated aquaculture (shrimp ponds) and sparse mangrove forest production areas. These patterns are in line with what would be expected based on biomass, forest fragmentation, protection measures and the forest use (Hauser et al. 2017; Pham et al. 2020).

Distribution of LAI absolute uncertainty values, expressed by standard deviation (SD), revealed substantial differences between LAI values in each pixel and mean values calculated in the whole area. Higher SD values indicate higher uncertainty tied to a higher amount of variation or dispersion while applying the retrieval model. The relative uncertainty (Figure 5c) describes the coefficient of variation (CV: SD/estimate × 100) of the LAI map in the study area. Since the SD directly depends on LAI mean value, it is more useful

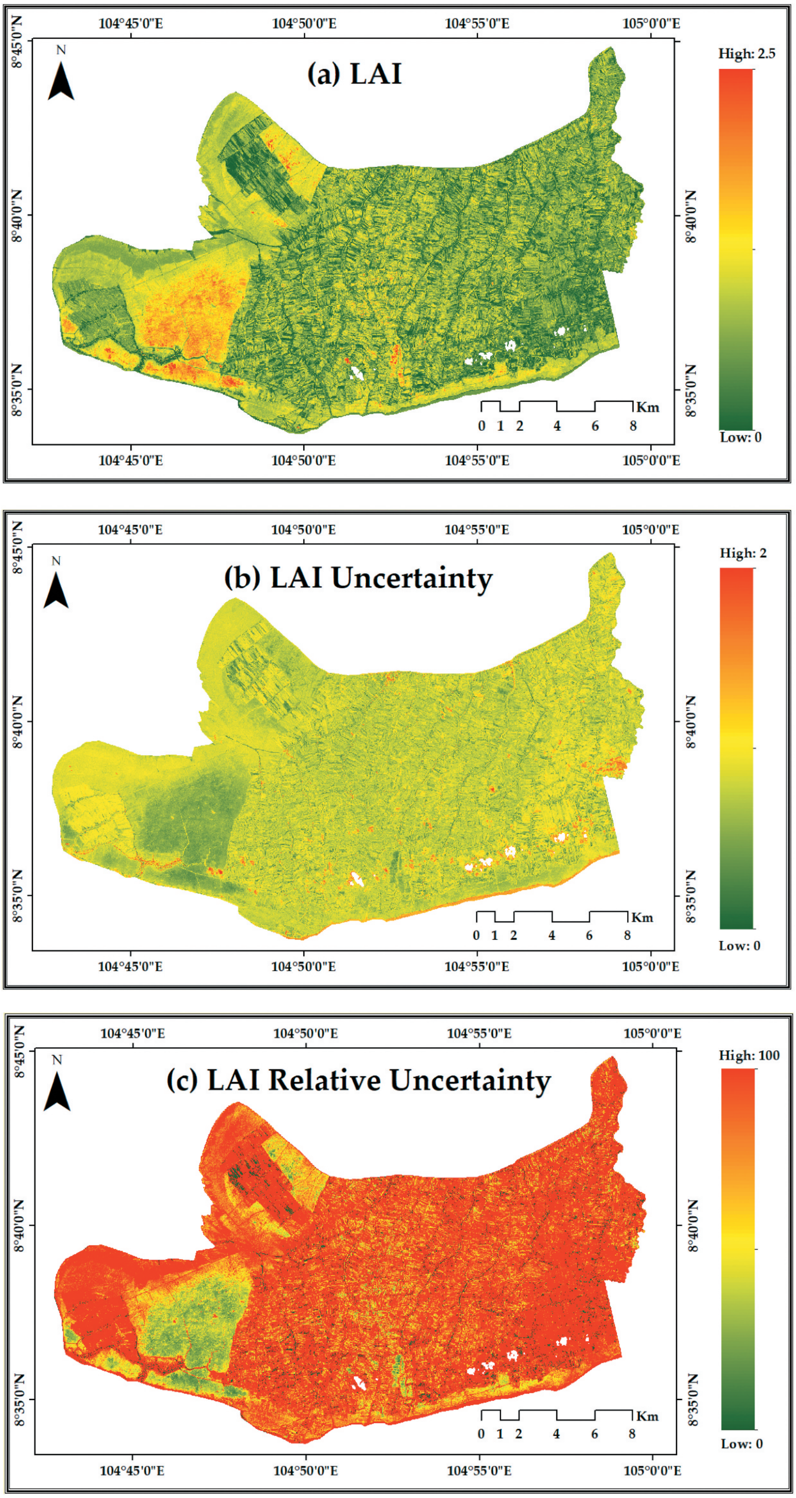


Figure 5. Maps of LAI (a), Uncertainty (b) and Relative uncertainty (c) in Ngoc Hien district, Ca Mau.

to use SD to show the relative uncertainties, enabling accounting for the amplitude of variability caused by spatially biased distributions of higher or lower LAI means. The map of LAI's relative uncertainty suggested that mapping LAI in core forest areas, which belong to Mui Ca Mau National Park in the West, was done with higher confidence, while difficulties are observed to estimate mangrove LAI in the production areas characterized by mixed pixels of shrimp and sparse mangrove cover, which was the main practice livelihood model in the majority of the Eastern regions of the Ngoc Hien region.

4. Discussion

Optical remote sensing approaches can provide rapid insights into mangrove ecosystems over broad regions. Thus far, most studies have commonly focused on mangrove forest extent and fragmentation, yet further characterization of biophysical vegetation variables will be crucial to deepen understanding of the health, phenology, functional attributes, and diversity thereof in mangrove ecosystems. Here, we focussed on LAI as a key vegetation attribute directly linked to primary productivity and competitive and complementary light use, transpiration, and energy exchange (Asner, Scurlock, and Hicke 2003; Fang et al. 2019; Zheng and Moskal 2009).

The few studies engaged in LAI mapping in mangrove ecosystems thus far relied on spectral indices (e.g. Kamal, Phinn, and Johansen 2016; Mafi-Gholami et al. 2019; Manna and Raychaudhuri 2020; Parida and Kumari 2020) and data-driven approaches to estimate spatial patterns of LAI (e.g. Castillo et al. 2017; Pham et al. 2019; Zhu et al. 2017). We investigated the potential of RTM simulations, combined with AL algorithms, to estimate LAI to overcome the shortcomings in scalability of spectral indices and purely data-driven approaches.

Significant differences were found in the performance of LAI estimates between AL approaches versus the generic SNAP biophysical processor, as well as between different implementations of RTMs (PROSAIL and INFORM) to simulate canopy reflectance. Moreover, spatial patterns across the study area indicate distinct patterns aligned with zonal management regimes (areas reserved for conservation versus integrated mangrove-aquaculture production areas) and associated uncertainties in estimates. We discuss the interpretation of these results and its implications for further extension of RTM simulations and AL to improve ecological monitoring of mangroves.

4.1. Performance of GPR and EBD active learning

In the light of various hybrid methods that combine machine learning methods and RTMs (Danner et al. 2021; Sinha et al. 2020), we deployed EBD-GPR to retrieve mangrove LAI using Sentinel-2 reflectance data validated against in-situ LAI data (n = 65) precisely matching Sentinel-2's 10 m spatial resolution. The results showed a superior LAI estimation in both PROSAIL and INFORM models with R^2 values of 0.77 and 0.66, respectively, compared to straightforward commonly applied SNAP retrievals (R^2: 0.45) and a red-edge NDVI vegetation index to estimate green LAI (R^2: 0.20) (see Figure 3).

Mangrove forests present a very distinct type of ecosystem and niche of vegetation type. In addition, the frequent inundation of these tidal forests results in water and mud spectra being influential to aggregate canopy reflectance observations from satellite inference. Common RTMs are not designed with this kind of specific ecosystem characteristics in mind. For instance, PROSAIL has two main soil parameters: psoil (Dry/Wet soil factor) and rsoil (Soil Brightness factor) (Huang et al. 2019). The combination of these soil parameters allows us to model a wide range of different types of background spectra including an approximation of water-like spectra. Importantly, active learning approaches allow to optimize the use of RTM simulations by subsetting those simulations that are relevant for training a model that fits LAI estimates for mangroves.

The optimization of the training dataset through AL heuristics likely drives the superior performance of the presented approaches compared to SNAP when applied to mangrove forests specifically. Furthermore, different versions of PROSPECT and 4SAIL between SNAP's RTM implementation and our integration of later versions in PROSAIL and INFORM may influence the accuracy of estimates due to differences in parameterization. Moreover, although both SNAP and our EBD-GPR approach rely on hybrid methods, the implementation of neural networks (SNAP) versus Gaussian process regression (GPR) leads to differences in mathematical intrinsic properties in model training.

The workflow demonstrated here deploys a hybrid approach using AL to subset RTM simulations in order to optimize training for retrieval of LAI in mangrove ecosystems. Through physical-based simulation, we are able to generate a training data set sufficiently large to train a model for retrieval of LAI. These simulations help overcome the difficulty and scarcity of acquiring high-quality and harmonized in-situ measurements. Therefore, RTM-based approaches may present higher transferability than purely data-driven statistical learning methods, which heavily rely on extensive field datasets to establish empirical relationships (e.g. spectral indices, partial least squares regression). Similarly, the use of spectral indices deals with sensor-, and time-specific calibration and the non-linear saturation of LAI, which limits transferability and scalability (Verrelst et al. 2019).

RTMs, on the other hand, are similarly subject to strong assumptions, heavy parameterization, and ill-posedness (Combal et al. 2003; Koetz et al. 2007; Musavi et al. 2015). Multiple configurations of the RTMs produce identical or similar spectral responses (Verrelst et al. 2016b). This problem is further amplified when the number of bands is limited or by the presence of noise common in satellite observations (Brede et al. 2020; de Sá et al. 2021). This ill-posedness can be minimized through a priori selection (i.e. optimization) of expected biophysical trait range of values (Verrelst et al. 2015). Such pre-selection requires in-depth expertise of the trait ranges found in an ecosystem. AL methods, as demonstrated here, can help guide to optimize a subset of RTM simulations to improve the accuracy of the model, which may be due to mitigation of the ill-posedness and/or a relevant optimization of a generic RTM to the specificity of mangrove vegetation. Although based on a hybrid framework including a broad range of simulated data, a limitation of the AL model for mangrove LAI presented remains that validation and AL tuning were only possible against one in-situ dataset. This raises questions regarding the independence of AL approaches from field data. Of interest would be to apply the model to another mangrove region in Vietnam with independent field measurements to assess its transferability. The labour-intensiveness of such campaigns currently still limits the availability of field data of nearby mangrove ecosystems.

Taken together, the study illustrates that we can use RTM simulations to generate large general training datasets for the retrieval of LAI or potentially other vegetation characteristics. Our AL implementation then allows us to subset these general training datasets to adapt for ecosystem-specific learning and overcome some of the ill-posedness. Here, we demonstrated that general and relatively simplistic RTMs can be used to map LAI, even in a distinct ecosystem such as the mangrove-dominated Ngoc Hien region in Vietnam, when combined with AL algorithms. Further testing is needed to assess its transferability across other mangrove ecosystems globally, across other mangrove canopy traits, and whether physics-based active learning with general RTMs (e.g. PROSAIL) can be successful in other niche and distinct ecosystem types.

4.2. Biophysical variables estimating from PROSAIL and INFORM

Two different canopy RTMs were used for simulation of canopy reflectance. The forest model INFORM performance exhibited slightly lower accuracy than standard vegetation model PROSAIL. INFORM is parameterized with more forestry-specific input parameters such as stem density, crown diameters, and LAI of the understory. ARTMO's INFORM parameterization allows up to a stem density maximum of 3000 ha^{-1}, while some forest areas in Ca Mau are reported to have stem densities up to 20,000 ha^{-1} (Nguyen et al. 2020). In addition, INFORM's understory LAI parameter might overcomplicate mangrove forests since it is largely dominated by single story vertical vegetation composition. The dense canopy cover of mangroves allows little light for understory vegetation and the tidal and saline mud and water below further creates a hostile environment for understory growth (Janzen 1985). As such, despite the more limited parameterization and idealized assumptions, PROSAIL might actually be better equipped for LAI estimates given the characteristics of mangrove ecosystems when combined with AL intelligent sampling.

4.3. Spatial variability across Ngoc Hien

It is necessary to incorporate wider understanding of the characteristics of mangroves traits, not only in leaf but also in canopy layer for interpretation of these results. Specific to our study area, the native species *Rhizophora piculata* were dominant as coverage within Sentinel-2 canopy observation, and its heterogeneous tree structure in different areas can affect the inversion of models. Within the premises of the delineated national park, natural mangrove forests are mainly conserved, while in the rest of Ngoc Hien mangrove forests have a productive function subject to integrated shrimp farming under sparse mangrove roots. Our results indicate that land cover, including the integration of aquaculture, affects the core mangrove spectral reflectance due to the mixture with soil and water, resulting in lower confidence in retrieval estimates (see Figure 4(b,c)). Distinct spatial patterns of absolute LAI estimates and the relative uncertainty of estimates can be observed from the produced maps. These spatial patterns correspond with the zonal management of Ngoc Hien district, which has delineated zones for conservation and aquaculture production integrated within sparse mangrove forests. Indeed, the integrated aquaculture-mangrove forest land-use tends to exhibit lower LAI values. While low estimates easily lead to higher relative uncertainties (CV: SD/estimate × 100), the high relative uncertainties also indicate difficulties and potentially inadequacy of our hybrid GPR model to map plant traits within integrated aquaculture-mangrove forest land use and mixed land uses. These high uncertainties could relate to insufficient adaptation of the RTM models to reflect sparse integrated mangrove areas as input to train the GPR model for application to the production areas.

4.4. Future implementation

In order to achieve a profound understanding of mangrove ecosystems, although LAI plays a key role, also other traits need to be quantitatively estimated (Parida and Kumari 2020; Pham et al. 2019; Younes Cárdenas, Joyce, and Maier 2017). Our demonstration of a hybrid model combining the benefits of physical-based RTM approaches and AL algorithms opens new venues for future research to include other biophysical variables of mangrove forest beyond LAI, for instance: leaf and canopy chlorophyll content, leaf mass per area, equivalent water thickness, as well as towards the functional diversity of mangrove forests (Hauser et al. 2021).

Our model combining GPR and AL estimating highlights significant improvements of model performance as opposed to the SNAP LAI products and using the red-edge NDVI. However, the evaluation of different MLRA coupled with AL techniques deserves further attention to optimize the most-suited hybrid approach for large-scale applications not only at national-level but also regional level in relevant coastal countries to improve biophysical and biochemical monitoring of mangrove ecosystems.

We presented only one LAI map in time, which falls short in elucidating the phenological and cyclic processes that are fundamental to ecosystem functioning. Exploiting time series of satellite remote-sensed data remains to be done, to enable temporal monitoring over the large-scale mangrove forest areas, and so to gain knowledge of natural drives (such as seasonal cycles and trends) as well as the impacts of anthropogenic activities (agriculture and aquaculture). Quantifying the temporal dynamics and the impact of human activities over time can help support decision-making with comprehensive management solutions for mangrove forests (Hauser et al. 2020).

In addition, cloud cover is a major concern for optical remote sensing of coastal countries, especially when located in the tropical monsoon climate. Therefore, gap-filling and data assimilation tools are essential using temporal analytics, as well as seasonal characteristics by mangrove landscape, and the integration of different sensors. Uncertainty maps (standard deviation around the mean or relative uncertainty) of biophysical retrieval can guide further the optimization and transferability of the mapping process. For instance, when applying the GPR model to other sites or at other moments, the robustness of the model can be evaluated by inspecting how the uncertainties behave; i.e. if uncertainties stay in the same order as the region where the validation took place, it then implies that the quality of the retrieval is also in the same order (Verrelst et al. 2013).

5. Conclusion

Our study demonstrated the performance of remotely sensed LAI estimated through a hybrid model validated against in-situ LAI measurements based on hemispherical photography. Robust hybrid approaches based on physical-based RTMs together with active learning (AL) and machine learning appear to improve the capacities of mangrove LAI retrieval, with significant gains in accuracy compared to the SNAP LAI model and red-edge NDVI. AL techniques show promise in selecting the most informative data while learning from big data sources, which seem to be the key in advanced biophysical variable retrieval in heterogeneous landscapes. Using RTM simulations allows us to overcome the difficulties and scarcity of obtaining large datasets of empirical observations for model training needed for machine learning. The AL algorithms facilitate an optimal subset of RTM simulations to guide model training. This could mitigate the ill-posedness in retrieval of biophysical variables and accommodate for the peculiarities of the ecosystem under study. Given the demonstrated performance in this study, future research should expand and further assess the application of AL-driven physics-based hybrid models for retrieval of ecologically relevant plant traits at both leaf and canopy level, its transferability across other mangrove ecosystems globally, and whether physics-based AL with general RTMs (e.g. PROSAIL) can be successfully applied in other niche and distinct ecosystem types. Ultimately, accurate retrieval of mangrove traits can help us better understand the functioning and status of these ecosystems to guide conservation, rehabilitation and the management of mangrove forests and its vital ecosystem services.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was supported by the project ĐLTE00.06/20-21. We thank the Vietnam Academy of Science and Technology (VAST) for funding the research project grant number ĐLTE00.06/20-21. The authors also would like to thank Mui Ca Mau National Park and Dat Mui Protection Forest Management Board located in Ngoc Hien district, Ca Mau Province, Vietnam, for the support of field survey. J. Verrelst was additionally funded by the European Research Council (ERC) under the ERC-2017-STG SENTIFLEX project [grant agreement 755617] and by Ramón y Cajal Contract (Spanish Ministry of Science, Innovation and Universities).

ORCID

Leon T. Hauser http://orcid.org/0000-0003-1408-9942

References

Aguirre-gutiérrez, J., S. Rifai, A. Shenkin, I. Oliveras, L. Bentley, M. Svátek, C. A. J. Girardin, et al. 2021. "Pantropical Modelling of Canopy Functional Traits Using Sentinel-2 Remote Sensing Data." *Remote Sensing of Environment* 252: 112122. doi:10.1016/j.rse.2020.112122.

Asner, G. P. 1998. "Biophysical and Biochemical Sources of Variability in Canopy Reflectance." *Remote Sensing of Environment* 64: 234–253. doi:10.1016/S0034-4257(98)00014-5.

Asner, G. P., J. M. O. Scurlock, and J. A. Hicke. 2003. "Global Synthesis of Leaf Area Index Observations: Implications for Ecological and Remote Sensing Studies."*Global Ecology and Biogeography* 12 (3): 191–205. doi:10.1046/j.1466-822X.2003.00026.x.

Atzberger, C., 2000. "Development of an Invertible Forest Reflectance Model the INFORM-Model." in: A Decade of Trans-European Remote Sensing Cooperation, Proceedings of the 20th EARSeL Symposium, Rotterdam, the Netherlands.

Baret, F., and S. Buis. 2008. "Estimating Canopy Characteristics from Remote Sensing Observations: Review of Methods and Associated Problems." In *Advances in Land Remote Sensing*, 173–201. Dordrecht: Springer. https://doi.org/10.1007/978-1-4020-6450-0_7

Berger, K., J. Pablo, R. Caicedo, L. Martino, M. Wocher, and T. Hank. 2021. "A Survey of Active Learning for Quantifying Vegetation Traits from Terrestrial Earth Observation Data." *Remote Sensing* 13: 1–23. doi:10.3390/rs13020287.

Brander, L. M., A. J. Wagtendonk, S. S. Hussain, A. McVittie, P. H. Verburg, R. S. de Groot, and S. van der Ploeg. 2012. "Ecosystem Service Values for Mangroves in Southeast Asia: A Meta-analysis and Value Transfer Application." *Ecosystem Services* 1: 62–69. doi:10.1016/j.ecoser.2012.06.003.

Brede, B., J. Verrelst, J. P. Gastellu-Etchegorry, J. G. P. W. Clevers, L. Goudzwaard, J. den Ouden, J. Verbesselt, and M. Herold. 2020. "Assessment of Workflow Feature Selection on Forest LAI Prediction with sentinel-2A MSI, Landsat 7 ETM+ and Landsat 8 OLI." *Remote Sensing* 12. doi:10.3390/rs12060915.

Brodu, N. 2017. "Super-Resolving Multiresolution Images with Band-Independent Geometry of Multispectral Pixels." *IEEE Transactions on Geoscience and Remote Sensing* 55: 4610–4617. doi:10.1109/TGRS.2017.2694881.

Brown, L. A., B. O. Ogutu, and J. Dash. 2019. "Estimating Forest Leaf Area Index and Canopy Chlorophyll Content with Sentinel-2: An Evaluation of Two Hybrid Retrieval Algorithms." *Remote Sensing* 11: 1752. doi:10.3390/rs11151752.

Bunting, P., A. Rosenqvist, R. M. Lucas, L. M. Rebelo, L. Hilarides, N. Thomas, A. Hardy, T. Itoh, M. Shimada, and C. M. Finlayson. 2018. "The Global Mangrove Watch - A New 2010 Global Baseline of Mangrove Extent." *Remote Sensing* 10: 1669. doi:10.3390/rs10101669.

Castillo, J., A. A. Apan, T. N. Maraseni, and S. G. Salmo. 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.

Chen, J. M., and T. A. Black. 1992. "Defining Leaf Area Index for Non-flat Leaves." *Plant, Cell & Environment* 15: 421–429. doi:10.1111/j.1365-3040.1992.tb00992.x.

Combal, B., F. Baret, M. Weiss, A. Trubuil, D. Macé, A. Pragnère, R. Myneni, Y. Knyazikhin, and L. Wang. 2003. "Retrieval of Canopy Biophysical Variables from Bidirectional Reflectance." *Remote Sensing of Environment* 84: 1–15. doi:10.1016/S0034-4257(02)00035-4.

Danner, M., K. Berger, M. Wocher, W. Mauser, and T. Hank. 2021. "Efficient RTM-based Training of Machine Learning Regression Algorithms to Quantify Biophysical & Biochemical Traits of Agricultural Crops." *ISPRS Journal of Photogrammetry and Remote Sensing* 173: 278–296. doi:10.1016/j.isprsjprs.2021.01.017.

Darvishzadeh, R., A. Skidmore, H. Abdullah, E. Cherenet, A. Ali, T. Wang, W. Nieuwenhuis, et al. 2019a. "Mapping Leaf Chlorophyll Content from Sentinel-2 and RapidEye Data in Spruce Stands Using the Invertible Forest Reflectance Model." *International Journal of Applied Earth Observation and Geoinformation* 79: 58–70. doi:10.1016/j.jag.2019.03.003.

Darvishzadeh, R., T. Wang, A. Skidmore, A. Vrieling, B. O'Connor, T. W. Gara, B. J. Ens, and M. Paganini. 2019b. "Analysis of Sentinel-2 and rapidEye for Retrieval of Leaf Area Index in a Saltmarsh Using a Radiative Transfer Model." *Remote Sensing* 11: 671. doi:10.3390/rs11060671.

de Sá, N. C., M. Baratchi, L. T. Hauser, and P. M. van Bodegom. 2021. "Exploring the Impact of Noise on Hybrid Inversion of PROSAIL RTM on Sentinel-2 Data." *Remote Sensing* 13 (4): 1–20. doi:10.3390/rs13040648.

Delegido, J., J. Verrelst, L. Alonso, and J. Moreno. 2011. "Evaluation of Sentinel-2 Red-edge Bands for Empirical Estimation of Green LAI and Chlorophyll Content." *Sensors* 11: 7063–7081. doi:10.3390/s110707063.

Donato, D. C., J. B. Kauffman, D. Murdiyarso, S. Kurnianto, M. Stidham, and M. Kanninen. 2011. "Mangroves among the Most Carbon-rich Forests in the Tropics." *Nature Geoscience* 4: 293–297. doi:10.1038/ngeo1123.

Duke, N. C., J.-O. Meynecke, S. Dittmann, A. M. Ellison, K. Anger, U. Berger, S. Cannicci, K. Diele, K. C. Ewel, and C. D. Field. 2007. "A World without Mangroves?" *Science* 80-.). 317: 41–42. doi:10.1126/science.317.5834.41b.

ESA. 2015. *Sentinel-2 User Handbook*. European Space Agency (ESA). https://sentinels.copernicus.eu/documents/247904/685211/Sentinel-2_User_Handbook.pdf/8869acdf-fd84-43ec-ae8c-3e80a436a16c?t=1438278087000

Fang, H., F. Baret, S. Plummer, and G. Schaepman-Strub. 2019. "An Overview of Global Leaf Area Index (LAI): Methods, Products, Validation, and Applications." *Reviews of Geophysics* 57: 739–799. doi:10.1029/2018RG000608.

Fourty, T., and F. Baret. 1997. *Amélioration de la Précision Des Coefficients D'absorption Spécifique de la Matière Sèche et Des Pigments Photosynthétiques*. Avignon: INRA Bioclimatol.

Garrigues, S., N. V. Shabanov, K. Swanson, J. T. Morisette, F. Baret, and R. B. Myneni. 2008. "Intercomparison and Sensitivity Analysis of Leaf Area Index Retrievals from LAI-2000, AccuPAR, and Digital Hemispherical Photography over Croplands." *Agricultural and Forest Meteorology* 148: 1193–1209. doi:10.1016/j.agrformet.2008.02.014.

GCOS. 2011. "Systematic Observation Requirements for Satellite–BASED Data Products for Climate, 154 Document."

Giri, C., E. Ochieng, L. L. Tieszen, Z. Zhu, A. Singh, T. Loveland, J. Masek, and N. Duke. 2011. "Status and Distribution of Mangrove Forests of the World Using Earth Observation Satellite Data." *Global Ecology and Biogeography* 20: 154–159. doi:10.1111/j.1466-8238.2010.00584.x.

Ha, T. T. P., H. van Dijk, and L. Visser. 2014. "Impacts of Changes in Mangrove Forest Management Practices on Forest Accessibility and Livelihood: A Case Study in Mangrove-shrimp Farming System in Ca Mau Province, Mekong Delta." *Vietnam. Land Use Policy* 36: 89–101. doi:10.1016/j.landusepol.2013.07.002.

Hauser, L. T., N. A. Binh, P. V. Hoa, N. H. Quan, and J. Timmermans. 2020. "Gap-free Monitoring of Annual Mangrove Forest Dynamics in Ca Mau Province, Vietnamese Mekong Delta, Using the Landsat-7-8 Archives and Post-classification Temporal Optimization." *Remote Sensing* 12: 1–16. doi:10.3390/rs12223729.

Hauser, L. T., J.-B. Féret, N. An Binh, N. van der Windt, Â. F. Sil, J. Timmermans, N. A. Soudzilovskaia, and P. M. van Bodegom. 2021. "Towards Scalable Estimation of Plant Functional Diversity from Sentinel-2: In-situ Validation in a Heterogeneous (Semi-)natural Landscape." *Remote Sensing of Environment* 262: 112505. doi:10.1016/j.rse.2021.112505.

Hauser, L. T., G. Nguyen Vu, B. A. Nguyen, E. Dade, H. M. Nguyen, T. T. Q. Nguyen, T. Q. Le, L. H. Vu, A. T. H. Tong, and H. V. Pham. 2017. "Uncovering the Spatio-temporal Dynamics of Land Cover Change and Fragmentation of Mangroves in the Ca Mau Peninsula, Vietnam Using Multi-temporal SPOT Satellite Imagery (2004–2013)." *Applied Geography* 86: 197–207. doi:10.1016/j.apgeog.2017.06.019.

Heumann, B. W. 2011. "Satellite Remote Sensing of Mangrove Forests: Recent Advances and Future Opportunities." *Progress in Physical Geography: Earth and Environment* 35: 87–108. doi:10.1177/0309133310385371.

Huang, J., H. Ma, F. Sedano, P. Lewis, S. Liang, Q. Wu, W. Su, X. Zhang, and D. Zhu. 2019. "Evaluation of Regional Estimates of Winter Wheat Yield by Assimilating Three Remotely Sensed Reflectance Datasets into the Coupled WOFOST–PROSAIL Model." *European Journal of Agronomy* 102. doi:10.1016/j.eja.2018.10.008.

Jacquemoud, S., and F. Baret. 1990. "PROSPECT: A Model of Leaf Optical Properties Spectra." *Remote Sensing of Environment* 34: 75–91. doi:10.1016/0034-4257(90)90100-Z.

Jacquemoud, S., W. Verhoef, F. Baret, C. Bacour, P. J. Zarco-Tejada, G. P. Asner, C. François, and S. L. Ustin. 2009. "PROSPECT + SAIL Models: A Review of Use for Vegetation Characterization." *Remote Sensing of Environment* 113: S56–S66. doi:10.1016/j.rse.2008.01.026.

Jacquemoud, S., W. Verhoef, F. Baret, P. J. Zarco-Tejada, G. P. Asner, C. François, and S. L. Ustin. 2006. "PROSPECT + SAIL : 15 Years of Use for Land Surface Characterization." *Geoscience and Remote Sensing Symposium* 00:1992–1995. doi:10.1109/IGARSS.2006.516. 2006. IEEE Int. Conf.

Jacquemoud, U. 2019. *Leaf Optical Properties*. UK: Cambridge University Press. ISBN-10:1108481264

Janzen, D. H. 1985. "Mangroves: Where's the Understory?" *Journal of Tropical Ecology* 1: 89–92. doi:10.1017/S0266467400000122.

Kamal, M., S. Phinn, and K. Johansen. 2016. "Assessment of Multi-resolution Image Data for Mangrove Leaf Area Index Mapping." *Remote Sensing of Environment* 176: 242–254. doi:10.1016/j.rse.2016.02.013.

Koetz, B., G. Sun, F. Morsdorf, K. J. Ranson, M. Kneubühler, K. Itten, and B. Allgöwer. 2007. "Fusion of Imaging Spectrometer and LIDAR Data over Combined Radiative Transfer Models for Forest Canopy Characterization." *Remote Sensing of Environment* 106: 449–459. doi:10.1016/j.rse.2006.09.013.

Kuenzer, C., A. Bluemel, S. Gebhardt, T. V. Quoc, and S. Dech. 2011. "Remote Sensing of Mangrove Ecosystems: A Review." *Remote Sensing* 3: 878–928. doi:10.3390/rs3050878.

Kuenzer, C., and V. Q. Tuan. 2013. “Assessing the Ecosystem Services Value of Can Gio Mangrove Biosphere Reserve: Combining Earth-observation- and Household-survey-based Analyses.” *Applied Geography* 45: 167–184. doi:10.1016/j.apgeog.2013.08.012.

Lausch, A., S. Erasmi, D. J. King, P. Magdon, and M. Heurich. 2016. “Understanding Forest Health with Remote sensing-Part I-A Review of Spectral Traits, Processes and Remote-sensing Characteristics.” *Remote Sensing* 8: 1–44. doi:10.3390/rs8121029.

Mafi-Gholami, D., E. K. Zenner, A. Jaafari, and R. D. Ward. 2019. “Modeling Multi-decadal Mangrove Leaf Area Index in Response to Drought along the Semi-arid Southern Coasts of Iran.” *Science of the Total Environment* 656: 1326–1336. doi:10.1016/j.scitotenv.2018.11.462.

Manna, S., and B. Raychaudhuri. 2020. “Retrieval of Leaf Area Index and Stress Conditions for Sundarban Mangroves Using Sentinel-2 Data.” *International Journal of Remote Sensing* 41: 1019–1039. doi:10.1080/01431161.2019.1655174.

Musavi, T., M. D. Mahecha, M. Migliavacca, M. Reichstein, M. J. van de Weg, P. M. van Bodegom, M. Bahn, et al. 2015. “The Imprint of Plants on Ecosystem Functioning: A Data-driven Approach.” *International Journal of Applied Earth Observation and Geoinformation* 43: 119–131. doi:10.1016/j.jag.2015.05.009.

Nguyen, L. T., H. T. Hoang, H. V. Ta, and P. S. Park. 2020. “Comparison of Mangrove Stand Development on Accretion and Erosion Sites in Ca Mau Vietnam.” *Forests* 11(6): 615. doi:10.3390/f11060615.

Padalia, H., S. K. Sinha, V. Bhave, N. K. Trivedi, and A. Senthil Kumar. 2020. “Estimating Canopy LAI and Chlorophyll of Tropical Forest Plantation (North India) Using Sentinel-2 Data.” *Advances in Space Research* 65: 458–469. doi:10.1016/j.asr.2019.09.023.

Parida, B. R., and A. Kumari. 2020. “Mapping and Modeling Mangrove Biophysical and Biochemical Parameters Using Sentinel-2A Satellite Data in Bhitarkanika National Park, Odisha.” *Modeling Earth Systems and Environment* 7: 2463–2474. doi:10.1007/s40808-020-01005-3.

Pham, T. D., N. N. Le, N. T. Ha, L. V. Nguyen, J. Xia, N. Yokoya, T. T. To, H. X. Trinh, L. Q. Kieu, and W. Takeuchi. 2020. “Estimating Mangrove Above-ground Biomass Using Extreme Gradient Boosting Decision Trees Algorithm with Fused Sentinel-2 and ALOS-2 PALSAR-2 Data in Can Gio Biosphere Reserve Vietnam.” *Remote Sensing* 12: 230. doi:10.3390/rs12050777.

Pham, T. D., N. Yokoya, D. T. Bui, K. Yoshino, and D. A. Friess. 2019. “Remote Sensing Approaches for Monitoring Mangrove Species, Structure, and Biomass: Opportunities and Challenges.” *Remote Sensing* 11: 1–24. doi:10.3390/rs11030230.

Rasmussen, C. E., and C. K. I. Williams. 2006. *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press.

Richter, K., and T. B. Hank. 2021. “Derivation of Biophysical Variables from Earth Observation Data : Validation and Statistical Measures.” *Journal of Applied Remote Sensing* 6(1): 063557. doi:10.1117/1.JRS.6.063557.

Rosema, A., W. Verhoef, H. Noorbergen, and J. J. Borgesius. 1992. “A New Forest Light Interaction Model in Support of Forest Monitoring.” *Remote Sensing of Environment* 42: 23–41. doi:10.1016/0034-4257(92)90065-R.

Schlerf, M., and C. Atzberger. 2006. “Inversion of a Forest Reflectance Model to Estimate Structural Canopy Variables from Hyperspectral Remote Sensing Data.” *Remote Sensing of Environment* 100: 281–294. doi:10.1016/j.rse.2005.10.006.

Siikamäki, J., J. N. Sanchirico, and S. L. Jardine. 2012. “Global Economic Potential for Reducing Carbon Dioxide Emissions from Mangrove Loss.” *Proceedings of the National Academy of Sciences of the United States of America* 109: 14369–14374. doi:10.1073/pnas.1200519109.

Sinha, S. K., H. Padalia, A. Dasgupta, J. Verrelst, and J. P. Rivera. 2020. “Estimation of Leaf Area Index Using PROSAIL Based LUT Inversion, MLRA-GPR and Empirical Models: Case Study of Tropical Deciduous Forest Plantation, North India.” *International Journal of Applied Earth Observation and Geoinformation* 86: 102027. doi:10.1016/j.jag.2019.102027.

Skidmore, A. K., N. C. Coops, E. Neinavaz, A. Ali, M. E. Schaepman, M. Paganini, W. D. Kissling, et al. 2021. “Priority List of Biodiversity Metrics to Observe from Space.” *Nature Ecology and Evolution* 5: 1639. doi:10.1038/s41559-021-01451-x.

Tue, N. T., L. V. Dung, M. T. Nhuan, and K. Omori. 2014. "Carbon Storage of a Tropical Mangrove Forest in Mui Ca Mau National Park Vietnam." *Catena* 121: 119–126. doi:10.1016/j.catena.2014.05.008.

Vane, G., and A. F. H. Goetz. 1988. "Terrestrial Imaging Spectroscopy." *Remote Sensing of Environment* 24: 1–29. doi:10.1016/0034-4257(88)90003-X.

Verhoef, W. 1984. "Light Scattering by Leaf Layers with Application to Canopy Reflectance Modeling: The SAIL Model." *Remote Sensing of Environment* 16: 125–141. doi:10.1016/0034-4257(84)90057-9.

Verhoef, W. 1998. *Theory of Radiative Transfer Models Applied in Optical Remote Sensing of Vegetation Canopies*. Wageningen, The Netherlands: Wageningen University.

Verrelst, J., K. Berger, and J. P. Rivera-Caicedo. 2020. "Intelligent Sampling for Vegetation Nitrogen Mapping Based on Hybrid Machine Learning Algorithms." *IEEE Geoscience and Remote Sensing Letters*. 108: 273–290. doi:10.1109/lgrs.2020.3014676.

Verrelst, J., G. Camps-Valls, J. Muñoz-Marí, J. P. Rivera, F. Veroustraete, J. G. P. W. P. W. Clevers, and J. Moreno. 2015. "Optical Remote Sensing and the Retrieval of Terrestrial Vegetation Bio-geophysical Properties - A Review." *ISPRS Journal of Photogrammetry and Remote Sensing* 108: 273–290. doi:10.1016/j.isprsjprs.2015.05.005.

Verrelst, J., S. Dethier, J. P. Rivera, J. Munoz-Mari, G. Camps-Valls, and J. Moreno. 2016a. "Active Learning Methods for Efficient Hybrid Biophysical Variable Retrieval." *IEEE Geoscience and Remote Sensing Letters* 13: 1012–1016. doi:10.1109/LGRS.2016.2560799.

Verrelst, J., Z. Malenovský, C. Van der Tol, G. Camps-Valls, J. P. Gastellu-Etchegorry, P. Lewis, P. North, and J. Moreno. 2019. "Quantifying Vegetation Biophysical Variables from Imaging Spectroscopy Data: A Review on Retrieval Methods." *Survey in Geophysics* 40: 589–629. doi:10.1007/s10712-018-9478-y.

Verrelst, J., J. P. Rivera, A. Gitelson, J. Delegido, J. Moreno, and G. Camps-Valls. 2016b. "Spectral Band Selection for Vegetation Properties Retrieval Using Gaussian Processes Regression." *International Journal of Applied Earth Observation and Geoinformation* 52: 554-567. doi:10.1016/j.jag.2016.07.016.

Verrelst, J., J. P. Rivera, J. Moreno, and G. Camps-Valls. 2013. "Gaussian Processes Uncertainty Estimates in Experimental Sentinel-2 LAI and Leaf Chlorophyll Content Retrieval." *ISPRS Journal of Photogrammetry and Remote Sensing* 86: 157-167. doi:10.1016/j.isprsjprs.2013.09.012.

Verrelst, J., E. Romijn, and L. Kooistra. 2012. "Mapping Vegetation Density in a Heterogeneous River Floodplain Ecosystem Using Pointable CHRIS/PROBA Data." *Remote Sensing* 4: 2866–2889. doi:10.3390/rs4092866.

Vinué, D., F. Camacho, and B. Fuster. 2018. "Validation of Sentinel-2 LAI and FAPAR Products Derived from SNAP Toolbox over a Cropland Site in Barrax and over an Agroforested Site in Liria (Spain)." In *Fifth Recent Advances in Quantitative Remote Sensing*, Spain: Universitat de València. 248. http://wind.uv.es/RAQRS/RAQRS2017_Proceedings.pdf

Quoc Vo, T., C., Künzer, and N., Oppelt. 2015. "How remote sensing supports mangrove ecosystem service valuation: a case study in Ca Mau province, Vietnam." *Ecosystem Services* 14: 67–75.

Weiss, M., and F. Baret. 2010. *CAN-EYE V6. 1 User Manual*. Avignon, France: INRA.

Weiss, M., and F. Baret, 2016. "Sentinel-2 ToolBox Level 2 Biophysical Product Algorithms. Version 1.1."

Weiss, M., F. Baret, G. J. Smith, I. Jonckheere, and P. Coppin. 2004. "Review of Methods for in Situ Leaf Area Index (LAI) Determination Part II. Estimation of LAI, Errors and Sampling." *Agricultural and Forest Meteorology* 121: 37–53. doi:10.1016/j.agrformet.2003.08.001.

Younes Cárdenas, N., K. E. Joyce, and S. W. Maier. 2017. "Monitoring Mangrove Forests: Are We Taking Full Advantage of Technology?" *International Journal of Applied Earth Observation and Geoinformation* 63: 1–14. doi:10.1016/j.jag.2017.07.004.

Yuan, H., R. Ma, C. Atzberger, F. Li, S. A. Loiselle, and J. Luo. 2015. "Estimating Forest fAPAR from Multispectral Landsat-8 Data Using the Invertible Forest Reflectance Model INFORM." *Remote Sensing* 7: 7425–7446. doi:10.3390/rs70607425.

Zheng, G., and L. M. Moskal. 2009. "Retrieving Leaf Area Index (LAI) Using Remote Sensing: Theories, Methods and Sensors." *Sensors* 9: 2719–2745. doi:10.3390/s90402719.

Zhu, Y., K. Liu, L. Liu, S. W. Myint, S. Wang, H. Liu, and Z. He. 2017. "Exploring the Potential of World View-2 Red-edge Band-based Vegetation Indices for Estimation of Mangrove Leaf Area Index with Machine Learning Algorithms." *Remote Sensing* 9: 1060. doi:10.3390/rs9101060.