



Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Mapping leaf area index in a mixed temperate forest using Fenix airborne hyperspectral data and Gaussian processes regression

Rui Xie^{a,*}, Roshanak Darvishzadeh^a, Andrew K. Skidmore^{a,b}, Marco Heurich^{c,d}, Stefanie Holzwarth^e, Tawanda W. Gara^f, Ils Reusen^g

^a Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, P.O. Box 217, 7500 AE, Enschede, the Netherlands

^b Department of Environmental Science, Macquarie University, NSW, 2106, Australia

^c Bavarian Forest National Park, Freyunger Straße 2, 94481, Grafenau, Germany

^d Chair of Wildlife Ecology and Wildlife Management, University of Freiburg, Tennenbacher Straße 4, Germany

^e German Aerospace Center (DLR), German Remote Sensing Data Center (DFD) Oberpfaffenhofen, 82234, Wessling, Germany

^f Department of Geography and Environmental Science, University of Zimbabwe, P.O. Box MP167, Mt Pleasant, Harare, Zimbabwe

^g Center for Remote Sensing and Earth Observation Processes (VITO-TAP), BE-2400, Mol, Belgium

ARTICLE INFO

Keywords:

Airborne hyperspectral data
Leaf area index
Gaussian processes regression
Uncertainty
Spectral subset
Temperate mixed forests

ABSTRACT

Machine learning algorithms, in particular, kernel-based machine learning methods such as Gaussian processes regression (GPR) have shown to be promising alternatives to traditional empirical methods for retrieving vegetation parameters from remotely sensed data. However, the performance of GPR in predicting forest biophysical parameters has hardly been examined using full-spectrum airborne hyperspectral data. The main objective of this study was to evaluate the potential of GPR to estimate forest leaf area index (LAI) using airborne hyperspectral data. To achieve this, field measurements of LAI were collected in the Bavarian Forest National Park (BFNP), Germany, concurrent with the acquisition of the Fenix airborne hyperspectral images (400–2500 nm) in July 2017. The performance of GPR was further compared with three commonly used empirical methods (i.e., narrowband vegetation indices (VIs), partial least square regression (PLSR), and artificial neural network (ANN)). The cross-validated coefficient of determination (R_{cv}^2) and root mean square error ($RMSE_{cv}$) between the retrieved and field-measured LAI were used to examine the accuracy of the respective methods. Our results showed that using the entire spectral data (400–2500 nm), GPR yielded the most accurate LAI estimation ($R_{cv}^2 = 0.67$, $RMSE_{cv} = 0.53 \text{ m}^2 \text{ m}^{-2}$) compared to the best performing narrowband VIs SAVI2 ($R_{cv}^2 = 0.54$, $RMSE_{cv} = 0.63 \text{ m}^2 \text{ m}^{-2}$), PLSR ($R_{cv}^2 = 0.74$, $RMSE_{cv} = 0.73 \text{ m}^2 \text{ m}^{-2}$) and ANN ($R_{cv}^2 = 0.68$, $RMSE_{cv} = 0.54 \text{ m}^2 \text{ m}^{-2}$). Consequently, when a spectral subset obtained from the analysis of VIs was used as model input, the predictive accuracies were generally improved (GPR $RMSE_{cv} = 0.52 \text{ m}^2 \text{ m}^{-2}$; ANN $RMSE_{cv} = 0.55 \text{ m}^2 \text{ m}^{-2}$; PLSR $RMSE_{cv} = 0.69 \text{ m}^2 \text{ m}^{-2}$), indicating that extracting the most useful information from vast hyperspectral bands is crucial for improving model performance. In general, there was an agreement between measured and estimated LAI using different approaches ($p > 0.05$). The generated LAI map for BFNP using GPR and the spectral subset endorsed the LAI spatial distribution across the dominant forest classes (e.g., deciduous stands were generally associated with higher LAI values). The accompanying LAI uncertainty map generated by GPR shows that higher uncertainties were observed mainly in the regions with low LAI values (low vegetation cover) and forest areas which were not well represented in the collected sample plots. This study demonstrated the potential of GPR for estimating LAI in forest stands using airborne hyperspectral data. Owing to its capability to generate accurate predictions and associated uncertainty estimates, GPR is evaluated as a promising candidate for operational retrieval applications of vegetation traits.

* Corresponding author.

E-mail address: rui.xie4@gmail.com (R. Xie).

<https://doi.org/10.1016/j.jag.2020.102242>

Received 21 June 2020; Received in revised form 17 September 2020; Accepted 21 September 2020

Available online 17 October 2020

0303-2434/© 2020 The Authors.

Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Forests, one of the most dominant terrestrial ecosystem of Earth, hold more than three-quarters of the world's terrestrial biodiversity and provide a wide variety of environmental materials and ecosystem services (Canadell et al., 2000; FAO, 2010). However, climate change in the past decades has caused a severe impact on forest ecosystems and posed major challenges to forest management (Klos et al., 2009; Birdsey and Pan, 2011). Monitoring the forest dynamics requires spatially, temporally, and accurate quantification of forest biophysical variables (Gower et al., 1999; Hansen and Schjoerring, 2003; Hill et al., 2019). Among many biophysical variables, leaf area index (LAI) is a primary measure since it controls many physiological processes within vegetation canopies, such as photosynthesis, transpiration, evapotranspiration, as well as rainfall interception (Chen and Black, 1992; Weiss et al., 2004). In a broader context, LAI is recognized as one of the essential climate variables (ECVs) to be implemented in the Global Climate Observing System (GCOS) (Bojinski et al., 2014). Moreover, LAI is also a critical input in ecosystem modelling (Fischer et al., 1997), and recently has been proposed as one of the remote sensing-enabled essential biodiversity variables (EBVs) for satellite monitoring of progress towards the Aichi Biodiversity Targets (Skidmore et al., 2015).

Traditionally, estimation of forest LAI relied on field surveys such as destructive sampling, leaf traps, and plant canopy analysers (Chen et al., 1997). Although these methods are perhaps the most accurate pathway for determining LAI, they are time-consuming, costly, and impractical to extrapolate over large spatial extents (Norman and Campbell, 1989). Remote sensing (RS) provides an opportunity for quantifying biophysical variables over large areas being fast, repeatable, synoptic, and cost-effective (Cohen et al., 2003; Atzberger, 2000; Yuan et al., 2017). In particular, hyperspectral remote sensing that can capture detailed vegetation information from hundreds of contiguous spectral narrow bands has significantly improved the prediction of vegetation parameters (Hansen and Schjoerring, 2003; Lee et al., 2004; Mutanga and Skidmore, 2004).

In general, there are two common approaches to estimate LAI from remotely sensed data as described in the RS literature: (1) physically-based models; and (2) empirical approaches (Baret and Buis, 2008; Atzberger et al., 2015). According to Skidmore (2002), these approaches can also be characterized as inductive and deductive by their logics, or as deterministic and stochastic by their processing methods. The physically-based approach involves using radiative transfer models (RTMs) to explicitly simulate the interaction between spectral radiation and vegetation biophysical and biochemical parameters (also referred to as plant traits) (Houborg et al., 2007). However, a major drawback of using RTMs is their ill-posed nature which causes different sets of biophysical input variables to yield similar spectral reflectance (Weiss and Baret, 1999; Combal et al., 2003; Atzberger et al., 2013). Moreover, RTMs requires prior knowledge of several input variables to calibrate and run the model in the forward mode (Darvishzadeh et al., 2008a, Darvishzadeh et al., 2019).

Compared to physically-based models, empirical approaches aim to establish relationships between spectral observations and the target biophysical variable (e.g., LAI). Empirical methods can incorporate parametric or non-parametric regression methods (Verrelst et al., 2015). In parametric regression methods, empirical/inductive models are used to fit a function between the spectral reflectance or its transformation and plant traits (Skidmore, 2002; Haboudane et al., 2004). Parametric regression methods have been frequently used for retrieving vegetation biophysical variables from remote sensing data. Plenty of studies have demonstrated the importance of spectral vegetation indices (VIs) (Gong et al., 2003; Lee et al., 2004; Schlerf et al., 2005; Ali et al., 2017). While some other studies have focused on quasi-continuous spectral band configurations, such as red-edge position and continuum removal (Cho and Skidmore, 2006; Darvishzadeh et al., 2009; Schlerf et al., 2010; Cho et al., 2007). Parametric regression methods are outstanding for their

intrinsic simplicity and fast processing speed. However, these methods do not exploit the complete spectral information from 400–2500 nm, and their established parametric models are often sensitive to site, sensor, and sampling conditions, thus lack robustness and generalization (Baret and Guyot, 1991; Broge and Leblanc, 2001).

Unlike parametric regression methods, the non-parametric regression methods make use of full-spectrum information, and therefore, an explicit selection of spectral bands or transformation is not required. The non-parametric models are usually optimized through a learning phase based on training data. The non-parametric regression methods can be further divided into linear and nonlinear models based on different formulations (Verrelst et al., 2015). Linear non-parametric regression methods such as stepwise multiple linear regression (SMLR) and principal component regression (PCR), have effectively enhanced the estimation of vegetation parameters compared to parametric regression methods (Kokaly and Clark, 1999; Atzberger et al., 2010). This type of method, however, is usually hampered by the multicollinearity problem especially when the sample size is smaller than the number of hyperspectral bands (Curran, 1989; De Jong et al., 2003). By contrast, partial least square regression (PLSR) which has been widely used in chemometrics was specifically developed as a better alternative to conventional linear non-parametric regression methods for quantifying vegetation parameters. An important property of PLSR is that it decomposes the spectra by also considering the response variable information (Geladi and Kowalski, 1986). Several studies have confirmed the feasibility of PLSR for estimating vegetation biophysical variables using hyperspectral data in grasslands (Cho et al., 2007; Darvishzadeh et al., 2011) and agricultural areas (Li et al., 2014). Recently, PLSR was used for predicting canopy foliar nitrogen in a mixed temperate forest using airborne hyperspectral data (Wang et al., 2016).

Nonlinear non-parametric methods, also referred to as machine learning algorithms, have been developed rapidly during the last few decades (Verrelst et al., 2015). Conventional machine learning algorithms applied in the vegetation remote sensing domain include, for instance, artificial neural network (ANN) and decision-tree (DT) based learning (e.g., random forest) (Verrelst et al., 2019). Such methods are popular for their capability in establishing robust and adaptive nonlinear relationships between biophysical variables and the reflected spectrum (Hastie et al., 2009). Successful applications using machine learning algorithms such as ANN include the estimation of foliage nitrogen concentrations (Huang et al., 2004), shrubland LAI prediction (Neinavaz et al., 2016), and crop LAI estimation (Liang et al., 2015). Nevertheless, some limitations of these methods remain to be addressed, for instance, the overly complex model tuning process may largely impact the robustness of the ANN model (Verrelst et al., 2012b).

Among nonlinear non-parametric methods, recently, a group of kernel-based machine learning algorithms has emerged as a potential alternative to conventional machine learning methods in the retrieval of vegetation parameters. Such kernel-based methods owe their names to use kernel functions to transfer training data into a higher dimensional feature space, in which the nonlinear relationships can be modelled by quantifying similarities between input samples of a dataset (Verrelst et al., 2013). The main advantage of kernel methods is their flexibility in performing input-output mapping and thus generate robust relationships. In particular, Gaussian processes regression (GPR), which is developed based on statistical learning and Bayesian theory (Williams and Rasmussen, 2006), has been found to outperform other machine learning models in estimating vegetation variables (Pasolli et al., 2010; Verrelst et al., 2012b). Compared to other machine learning approaches, GPR has the benefit of simple implementation and requires a relatively small training dataset (Verrelst et al., 2013). Moreover, GPR automatically provides uncertainty estimates (also called confidence intervals) along with their mean estimates (Williams and Rasmussen, 2006). Uncertainty estimates which are often absent in the empirical models are especially important to evaluate the reliability of the generated model and assess the utility of the mapping results. Lacking information about

variable uncertainties can lead to errors in subsequent analysis when such vegetation variable maps are used in ecological applications (Wang et al., 2019). Thus, uncertainties mapping is crucial for improving model performance and mapping quality.

While GPR has been recently used for the estimation of canopy traits from hyperspectral RS data by a few studies, its applications have been mainly limited in the agricultural fields (Rivera et al., 2014a, 2014b; Verrelst et al., 2013, 2016) and grassland ecosystems (Wang et al., 2019). To the best of our knowledge, only the study by Halme et al. (2019) has examined GPR and support vector regression (SVR) for LAI estimation in a boreal forest. However, the data which was used in their study was limited to visible and NIR regions (400–1000 nm), and therefore they did not explore the full spectral range (400–2500 nm). Thus, the utility of GPR on full-spectrum hyperspectral imagery for estimating LAI in mixed temperate forest stands remains under-explored. Moreover, although several previous studies have investigated the performance of different empirical methods in estimating forest LAI from airborne hyperspectral data, to date, the comparisons among different methods are still missing.

Therefore, the aim of this study is to examine the performance of GPR as a representative of kernel-based machine learning methods, in comparison to the most commonly used empirical methods (i.e., narrowband VIs, PLSR, and ANN) in estimating forest LAI from Fenix airborne hyperspectral data (400–2500 nm). The performance of the studied methods was validated in terms of prediction accuracy against field LAI measurements, and the suitability for each retrieval method was then analysed. To examine the impact of reducing data dimensionality, a spectral subset obtained from the analysis of VIs was also used as input for model prediction. Finally, a forest LAI map was generated using the method with the highest accuracy.

2. Materials and methods

Fig. 1 presents the general workflow adopted in this study to examine the performance of GPR in estimating forest LAI in comparison with other three widely used empirical approaches (i.e. narrowband VIs, PLSR, and ANN). The following subsections present the details of the field data collection and methods used in this study.

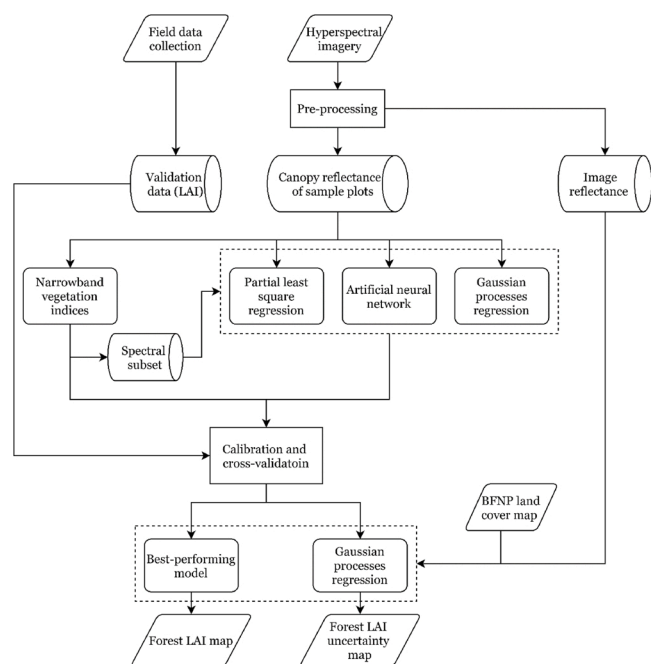


Fig. 1. General analytical framework for examining the performance of GPR in estimating forest LAI in comparison with narrowband VIs, PLSR, and ANN.

2.1. Study area

The study area for this research is Bavarian Forest National Park (BFNP) (49°3'19" N, 13°12'9" E), which is located in the south-eastern part of Germany, close to the border of the Czech Republic (Fig. 2). BFNP is Germany's first designated national park (founded in 1970) and covers an area around 24,250 ha. Together with the neighbouring Czech Sumava National Park, they form the largest, strictly protected contiguous forest area (called the Bohemian Forest Ecosystem) in Central Europe (Heurich et al., 2010). The elevation in BFNP ranges from 600 m to 1,453 m above sea level (a.s.l) and is composed of terrains varying from the low valley, hillsides to highlands. Tree species are mainly distributed as a function of altitude. In the peak regions, the majority of trees are Norway spruce (*Picea abies*) with a few existences of sub-alpine spruce forests and Mountain ash (*Sorbus aucuparia*). Mountain slope areas, characterized by mixed forests, mainly consist of Norway spruce, silver fir (*Abies alba*) and European beech (*Fagus sylvatica*). In valley depressions, Norway spruce is the dominant species with some silver fir and mixture of Birches (*Betula spp.*) (Cailleret et al., 2014).

2.2. Data

2.2.1. Airborne hyperspectral data

Airborne images of the study area (along permanent transects) were acquired during a field campaign on 6 July 2017 (Fig. 2). The data were collected based on 29 flight lines, covering 68.24 km², with an average 35 percent overlap with each adjacent strips. The Specim AISA Fenix sensor comprises two detectors covering the visible and near-infrared (VNIR) and short wave infrared (SWIR) regions. It contains 623 narrow spectral bands ranging from 380 to 2500 nm. The average spectral resolution is 3.5 nm over the VNIR region and 12 nm over the SWIR region. The spatial resolution of the imagery is approximately 3 m based on the average flight height of 2087.3 m above ground level. A pair of black bodies, which are mechanically moved in front of the sensor lens one by one, were used for calibration of the sensor. Most of the flight lines were acquired under a cloud-free condition.

2.2.2. Field measurements

The LAI field measurements were collected between 14 July and 14 August 2017. The stratified random sampling was performed within the major cover types in order to select field samples. This resulted in 13 plots of broadleaf, 14 plots of conifer, and 13 plots of mixed stands (n = 40). Sample plots located outside the image strips were not considered in this study (four plots). The size of each square plot is approximately 900 m² (30 m × 30 m). Their precise positions were recorded based on the centre coordinate of each plot using Leica GPS 1200 (Leica Geosystems AG, Heerburgg, Switzerland), and reached less than 1 m positioning accuracy after post-processing.

Within each sample plot, LAI was measured using a Li-Cor LAI-2200 canopy analyser (Li-Cor, 1992). For each plot, three above-canopy observations were taken as a reference reading in a nearby opening forest to minimize the difference of incoming radiation. Next, five below-canopy LAI observations were measured in each plot, and then the average value was computed to present the LAI value for the sample plot (Gara et al., 2019). Effort was made to keep the illumination conditions as constant for taking the above and below canopy LAI-2200 readings. The summary statistics of LAI field measurements are presented in Table 1.

2.3. Image processing

The image strips were preprocessed by NERC Airborne Research Facility (NERC-ARF). A MODTRAN-4 based radiative transfer model was employed to atmospherically correct each image line using the ATCOR4 software, resulting in reflectance images (Richter and Schläpfer, 2019). Rough terrain model, ASTER Digital Elevation Model (DEM), and rural

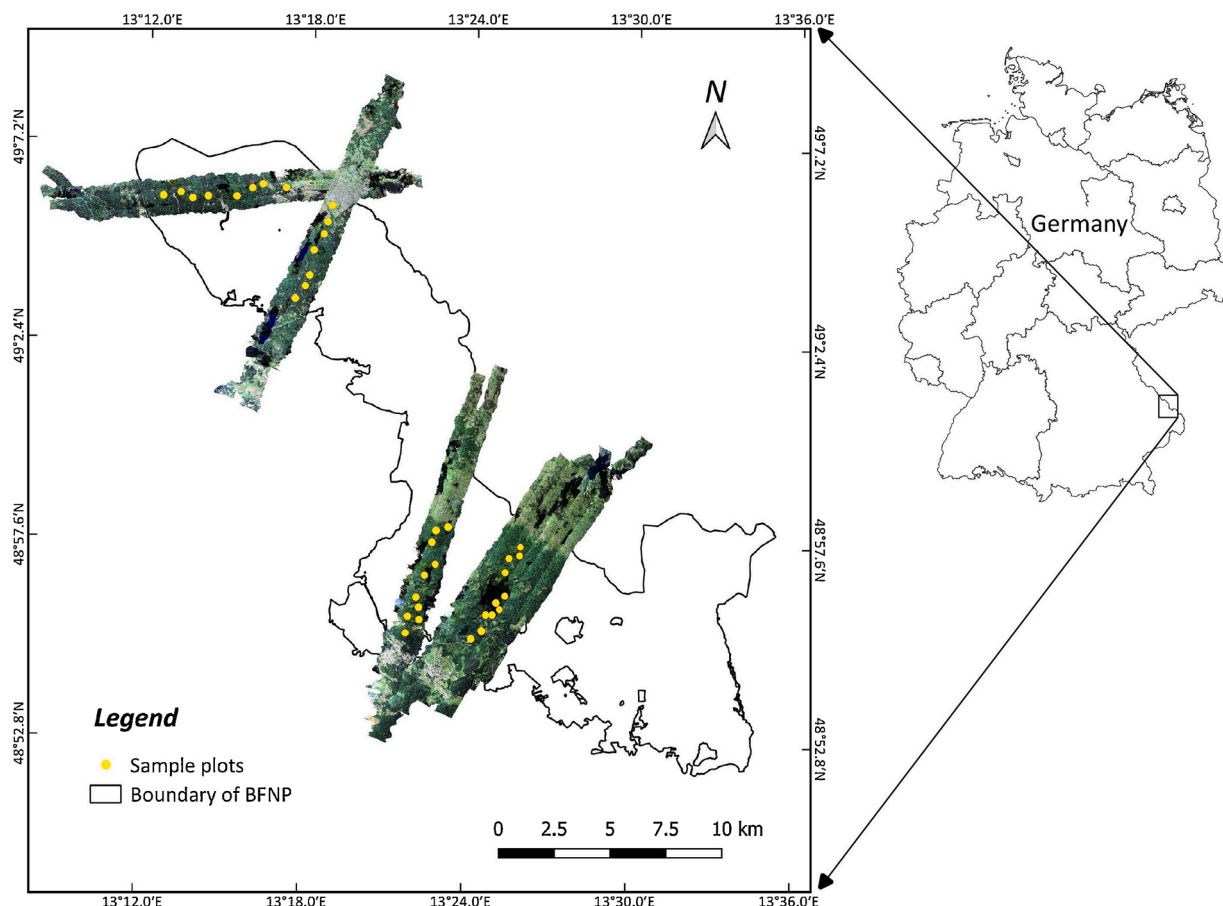


Fig. 2. The location of Bavarian Forest National Park (BFNP) and the mosaic of Fenix hyperspectral data of four transects acquired on 6 July 2017 using a true colour composite (bands 469, 549, 640 nm). Yellow dots represent the locations of sample plots (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

Table 1
Summary statistics of the measured LAI of sample plots in BFNP (n = 40).

Measured variable	Min	Max	Mean	Std.dev
LAI (m ² m ⁻²)	1.33	5.42	3.82	1.01

aerosol model were utilized for geometric and atmospheric correction, respectively. Since, the corrected image data still contained some systematic noise, a moving Savitzky-Golay filter with a frame size of 11 data points (second-degree polynomial) was used to eliminate the noise of the canopy reflectance spectra (Savitzky and Golay, 1964). The spectral data from 380–400 and 2400–2500 nm was not utilized due to their low signal-to-noise ratio.

The mean spectral reflectance for each plot was extracted using a 9 × 9 pixel window (i.e., 27 m by 27 m) to assure that we extracted true representatives of sample plots and avoided edge disturbance (Darvishzadeh et al., 2011). Since each plot may have been covered by a couple of image strips, the average reflectance spectra from adjacent image strips were calculated for each sample plot to represent the canopy reflectance of the plot. From a total of 40 plots, six plots were excluded from further analysis due to the geo-referencing error and cloud (shadow) cover. Taking into account another four plots that were located outside the image boundary, the remaining 30 plots were used for the analysis. The mean reflectance for three dominant tree species obtained from the airborne hyperspectral data are presented in Fig. 3. Data processing and analysis were performed using MATLAB, R2019a (The MathWorks, Inc.)

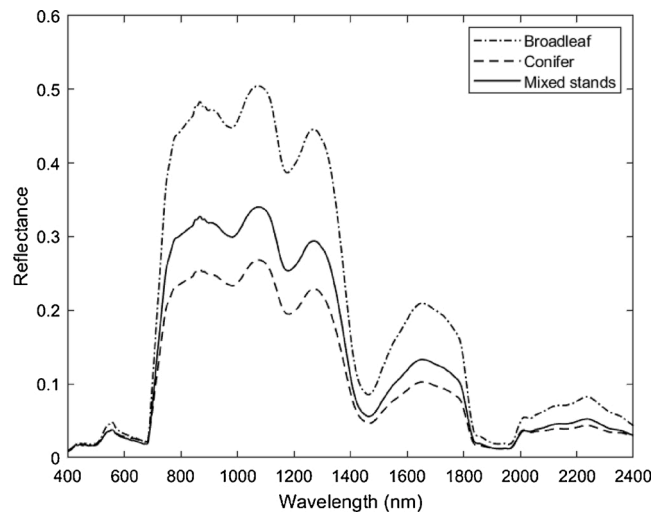


Fig. 3. Mean canopy reflectance for broadleaf, conifer, and mixed forest stands measured using the Specim AISA Fenix hyperspectral sensor.

2.4. Narrowband vegetation indices

Four widely used VIs were utilized as representatives of ratio-based and soil-based VIs to estimate LAI. These indices are normalized difference vegetation index (NDVI), ratio vegetation index (RVI), second soil-adjusted vegetation index (SAVI2), and transformed soil-adjusted

vegetation index (TSAVI). Narrowband indices were calculated using the equations of these broadband indices (Table 2) and the hyperspectral wavebands extracted from the processed image spectra.

To calculate the soil line parameters from spectral measurement, two assumptions were made: (1) the soil line concept which is originally defined for the red-NIR feature space could be transferred to other spectral domain (Thenkabail et al., 2000; Darvishzadeh et al., 2008b), and (2) since there was hardly any bare soil on the forest floor, the soil parameters (a and b) were calculated based on the mean reflectance of different understory layers found in the study area (Ali et al., 2016). A Savitzky-Golay filter with a frame size of 11 data points (second-degree polynomial) was applied to eliminate the noise of the measured background reflectance.

In order to find the optimal bands for narrowband VIs, all possible pairwise wavebands were used for systematically calculating selected narrowband VIs. The coefficients of determination (R^2) between narrowband VIs and measured LAI were used to evaluate the performance of the indices. The results are presented through a 2-D correlation matrix, in which the more sensitive regions can be identified based on a threshold where R^2 is greater than 0.5. The optimal bands which generated the maximum R^2 were selected to compute the narrowband indices. The linear regression model was employed to establish relationships between narrowband VIs and LAI. Further, these sensitive wavebands were selected as a spectral subset to be used as input for all other methods for LAI estimation.

2.5. Partial least square regression

PLSR has been widely used for the retrieval of vegetation parameters (Cho et al., 2007; Darvishzadeh et al., 2011; Singh et al., 2015). PLSR is a multivariate non-parametric regression method designed to alleviate multicollinearity which is an inherent problem in hyperspectral data. Using PLSR, a linear model was built between the response variable (LAI) and predictors (spectral reflectance). The observed collinear predictors were concentrated on a few non-correlated latent variables and the less informative variables were eliminated. The iterative decomposition was then performed on both explanatory and response variables to maximize the fit of Principal Component Analysis (PCA) on the response variables (Abdi, 2003; Schlerf et al., 2003). Further details about the PLSR model can be found in Geladi and Kowalski (1986).

In conditions where the input variables are highly correlated, a feature selection on input data is known to improve the model prediction (Dormann et al., 2013; Rivera et al., 2017). In this study, the PLSR was performed using the entire reflectance spectra (400–2400 nm) and the spectral subset, which was identified to be sensitive for LAI prediction using narrowband VIs in the above section. The spectral data were mean centred before applying PLSR analysis. Leave-one-out cross-validation was used to identify the optimal number of components (also

called factors) to calibrate the model. To avoid the overfitting problem, the number of components was determined according to a standard criterion that the added component increases the R_{cv}^2 and reduces the $RMSE_{cv}$ by $> 2\%$ (Kooistra et al., 2004; Cho et al., 2007). Moreover, the standardized regression coefficient (B coefficient) and the variable importance of projection (VIP) values were calculated to evaluate the importance of each waveband on the regression model (Haaland and Thomas, 1988). Important wavebands can be identified where the corresponding B coefficient is greater than the standard deviation and the VIP value is greater than one. PLSR analysis was carried out in TOMCAT toolbox 1.01 within MATLAB (Daszykowski et al., 2007).

2.6. Machine learning algorithm

Machine learning algorithms were used to learn the relationship between spectral reflectance and vegetation parameters through fitting a nonlinear transformation (Verrelst et al., 2012b). In this study, the performance of the ANN as a representative of the most common machine learning approach and GPR, which is recently introduced as a powerful kernel-based machine learning algorithm were evaluated respectively.

2.6.1. Artificial neural network

ANN is a commonly used approach to develop nonlinear non-parametric models for estimation of vegetation parameters (Kimes et al., 1998; Rivera et al., 2014a, 2014b; Neinavaz et al., 2016). In this study, a standard multi-layer perceptron with one hidden layer (tansigmoid transfer function) was adopted to connect the input (reflectance data) and output layer (corresponding LAI). To test the impact of input data on the performance of ANN, the entire reflectance and the spectral subset obtained from the analysis of narrowband VIs were separately used as inputs for model predictions. The widely used Levenberg-Marquardt learning algorithm in backpropagation with a squared loss function was utilized for training the nonlinear relationship between input and output datasets.

It is known that increasing number of neurons can usually enhance the prediction power of the network, but it would also make the model computationally demanding (Skidmore et al., 1997; Bacour et al., 2006). Thus, in this study, the optimal number of neurons was determined through testing model performance with different neuron numbers. To avoid overfitting, the training stops as soon as the validation sets fail to improve cross-validated results in the iterative procedure (so-called “early-stopping”) (Nowlan and Hinton, 1992). During the training process, the layer weights and biases were randomly initialized using the Nguyen-Widrow method (Nguyen and Widrow, 1990). To alleviate the effect of random model initialization, the cross-validated results were then averaged based on multiple trials following the suggestion of Neinavaz et al. (2016). The ANN analysis was performed with the MATLAB R2019a neural network toolbox (The MathWorks, Inc.)

2.6.2. Gaussian processes regression

In recent years, GPR has been gradually introduced as a powerful regression tool to retrieve vegetation parameters in remote sensing community (Verrelst et al., 2013; Halme et al., 2019; Gewali et al., 2019). GPR is a probabilistic (Bayesian) approach that provides predictions through kernel (covariance) functions, which is calculated by evaluating the similarity between pairs of testing and training input values. A proper kernel function always plays a vital role in successful prediction in GPR. For this study, a scaled squared exponential covariance function was employed, which has been found to be useful for extracting vegetation parameters by previous studies (Verrelst et al., 2013; Wang et al., 2019):

Table 2

Vegetation indices used in this study and their broadband forms in the literature. ρ_{red} and ρ_{NIR} are the reflectance in the red and NIR region. ρ_{λ_1} represents the reflectance at the wavelength λ_1 , and ρ_{λ_2} stands for the reflectance at the wavelength λ_2 ($\lambda_1 \neq \lambda_2$). a and b denote the slope and intercept of the soil line, respectively.

No.	References	Broadband VI	Narrowband VI
1	(Pearson and Miller, 1972)	$RVI = \frac{\rho_{NIR}}{\rho_{red}}$	$RVI_{narrow} = \frac{\rho_{\lambda_1}}{\rho_{\lambda_2}}$
2	(Rouse et al., 1974)	$NDVI = \frac{\rho_{NIR} - \rho_{red}}{\rho_{NIR} + \rho_{red}}$	$NDVI_{narrow} = \frac{\rho_{\lambda_1} - \rho_{\lambda_2}}{\rho_{\lambda_1} + \rho_{\lambda_2}}$
3	(Major et al., 1990)	$SAVI2 = \frac{\rho_{NIR}}{\rho_{red} + (b/a)}$	$SAVI2_{narrow} = \frac{\rho_{\lambda_1}}{\rho_{\lambda_2} + (b/a)}$
4	(Baret et al., 1989)	$TSAVI = \frac{a(\rho_{NIR} - a\rho_{red} - b)}{a\rho_{NIR} + \rho_{red} - ab}$	$TSAVI_{narrow} = \frac{a(\rho_{\lambda_1} - a\rho_{\lambda_2} - b)}{a\rho_{\lambda_1} + \rho_{\lambda_2} - ab}$

$$K(x_i, x_j) = \nu \exp \left(-\frac{1}{2} \sum_{b=1}^B \frac{(x_i^b - x_j^b)^2}{\sigma_b^2} \right) + \delta_{ij} \sigma_n^2 \quad (1)$$

Where ν is a scaling factor, B is the number of bands, σ_b is the length scale, σ_n is the noise standard deviation and δ_{ij} is the Kronecker’s symbol. In GPR, each output of training and testing data is assumed to be a noisy observation composed of true output value and additive Gaussian noise, so-called “prior distribution”. Based on the assumptions and Bayesian inference, the posterior probabilistic estimates (predictive mean and variance) can be obtained in the Gaussian distribution by conditioning the training data. It should be noted that the input spectra need to be normalized before they are used for learning. Further details on GPR theories can be found in Williams and Rasmussen (2006).

GPR does not require prior knowledge about the model hyperparameters (ν, σ_n, σ_b), because these parameters and model weights can be automatically optimized through maximizing the negative log marginal likelihood (Williams and Rasmussen, 2006). Another important advantage is a GPR model provides not only a prediction for each input spectra, but also a corresponding predictive variance (i.e., uncertainty estimates). The delivery of uncertainties allows us to post-evaluate the mapping results when GPR is applied to map variables of interest from images.

In this study, GPR was first applied to the entire spectra (i.e., 594 bands) to test the performance of the original reflectance. The reduction of data dimensionality is known to improve the GPR performance (van der Maaten et al., 2009; Rivera et al., 2017). To study the importance of removing redundant information on the model performance, the spectral subset obtained from the analysis of narrowband VIs was used as input to the GPR model. The GPR analysis was implemented using the GPML package within MATLAB (<http://www.gaussianprocess.org/gpml>).

2.7. Model validation and mapping

All the models were validated using leave-one-out cross-validation (LOOCV) method. In LOOCV, each sample is left out and estimated by a model developed based on the other remaining samples, and this procedure was repeated for all the samples (30 times). Cross-validated root mean square error ($RMSE_{cv}$) and cross-validated coefficient of determination (R^2_{cv}) between estimated LAI and measured LAI were selected as indicators of model accuracy of all studied methods. The LAI map of BFNP was then generated using the best-performing method (with the lowest $RMSE_{cv}$). Boxplot and paired t -test were employed to evaluate the

statistically significant difference between measured and predicted LAI using different retrieval methods. Before mapping LAI, the forest cover was extracted from the BFNP land use map provided by the national park administration (Silveyra Gonzalez et al., 2018). Then the non-forested area was masked out from the hyperspectral imagery using the extracted forest map. The masked image with the lowest cloud (shadow) coverage (i.e., Flight line 29) was used as input to the model for LAI prediction. To analyse the mapping output, the results were further compared with the forest type map and the true colour airborne imagery.

3. Results

3.1. Narrowband vegetation indices

The selected four narrowband VIs were systematically computed from canopy reflectance using all possible paired wavebands. The coefficients of determination (R^2) between narrowband VIs and measured forest LAI were calculated. Fig. 4 presents the results for the two best-performing indices (i.e., SAVI2 and RVI) of each narrowband VI type

Table 3

The highest R^2 between measured LAI and narrowband VIs calculated using Fenix airborne hyperspectral data and obtained sensitive regions for predicting LAI using different narrowband indices ($R^2 > 0.5$).

Type	Narrowband VI	Maximum R^2	Sensitive spectral range (nm)	
			λ_1 (nm)	λ_2 (nm)
Ratio-based narrowband VIs	RVI	0.55	739–767	718–743
			1158–1167	1290–1298
			1517–1561	1507–1525
			1942–1981	1870–1963
			2029–2085	1822–1867
Ratio-based narrowband VIs	NDVI	0.54	739–761	720–743
			1286–1299	1160–1168
			1517–1562	1507–1525
			1942–1981	1868–1964
			2029–2085	1822–1868
Ratio-based narrowband VIs	SAVI2	0.62	682–688	683–690
			742–771	729–731
			1167–1185	1290–1302
			1238–1274	1275–1290
			682–689	684–690
Soil-based narrowband VIs	TSAVI	0.61	740–765	729–730
			1162–1180	1282–1296
			1246–1266	1272–1290

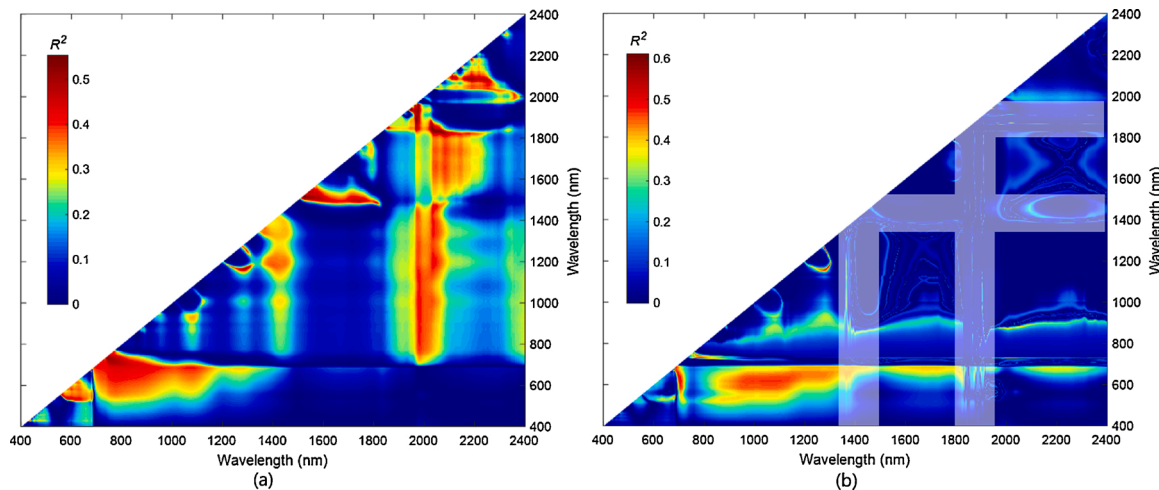


Fig. 4. 2-D correlation plot presenting the coefficient of determination (R^2) between measured LAI and (a) narrowband RVI and (b) narrowband SAVI2 calculated using Fenix airborne hyperspectral data. The highlighted regions in (b) refer to the strongly noisy bands existed in the measured understorey reflectance.

in the 2-D correlation matrix (where the meeting points represent the R^2 values between the measured LAI and narrowband indices). The highest R^2 between LAI and narrowband VIs, as well as the sensitive wavebands range (where $R^2 > 0.5$), are reported in Table 3. Based on the identified sensitive spectral wavebands in Table 3, a spectral subset was formulated.

The narrowband VIs generated by the optimum band combinations were further used to estimate LAI through linear regression models, and R_{cv}^2 and RMSE_{cv} were used to evaluate the model accuracy. As presented in Table 4, four studied narrowband VIs overall produced reasonable predictions for LAI, while soil-based narrowband VIs perform slightly better comparing to ratio-based ones.

3.2. Partial least square regression

The regression coefficients (B coefficients) and VIP values in Fig. 5 represent the relative contribution of each waveband in the PLSR model. The results show that contributing bands are mostly located in the NIR and SWIR region. When we considered the criterion that added component increases R_{cv}^2 and meanwhile reduces RMSE_{cv} by $> 2\%$, six components were determined to establish the final model. The performance of the PLSR model using the full spectrum for LAI estimation is presented in Fig. 6. As it can be observed, PLSR yields higher R_{cv}^2 value compared to narrowband VIs (Table 4) but failed to reduce the RMSE_{cv}.

3.3. Artificial neural network

In ANN, four neurons of the hidden layer produced the most accurate LAI prediction. By applying the early-stopping technique, the optimum model complexity was selected by tuning the training parameter values as soon as the validation sets fail to improve cross-validated results. Based on the optimum neuron size and selected model complexity, the relationship between measured and estimated LAI was calculated and is present in Fig. 7. In comparison with narrowband VIs and PLSR model, ANN significantly improved RMSE_{cv} and obtained a relatively high correlation with LAI (R_{cv}^2).

3.4. Gaussian processes regression

The performance of GPR was firstly evaluated from the original hyperspectral reflectance data. The cross-validated results are shown in Fig. 8. As can be seen from Fig. 8, both the RMSE_{cv} and R_{cv}^2 outperformed the majority of other assessed approaches in this study. GPR not only provides prediction means (μ) but also their corresponding uncertainties (σ) (i.e., variance of the mean prediction) (Verrelst et al., 2012a). The associated variance of LAI predictions are presented in Fig. 9. It can be observed that most of the plots were estimated with a high confidence level, while Plot 24 (index = 18) was predicted with a relatively large uncertainty interval.

3.5. Use of spectral subset in predicting LAI

In addition to predicting LAI using entire reflectance data, a spectral subset obtained from the analysis of narrowband VIs in Table 3 was used

Table 4

R_{cv}^2 and RMSE_{cv} between estimated and measured LAI using narrowband VIs calculated using Fenix airborne hyperspectral data and the wavelength of optimum bands combination.

Narrowband VI	R_{cv}^2	RMSE _{cv}	The best bands combination	
			λ_1 (nm)	λ_2 (nm)
RVI	0.43	0.69	1161	1295
NDVI	0.41	0.71	1167	1296
SAVI2	0.54	0.63	1259	1281
TSAVI	0.53	0.64	1260	1280

as input to all studied models for LAI prediction. The cross-validated results are shown in Table 5. A positive impact of using these informative wavebands for the PLSR model was observed, where R_{cv}^2 slightly increased from 0.74 to 0.75, and RMSE_{cv} decreased from 0.73 to 0.69. Also, applying the spectral subset to the GPR model decreased the RMSE_{cv} from 0.53 to 0.52 and improved R_{cv}^2 from 0.67 to 0.69. Nevertheless, using a spectral subset was not able to further improve the retrieval accuracy of the ANN model compared to the full-spectrum range.

3.6. Mapping leaf area index

The spatial distribution of the predicted LAI by GPR (using the identified spectral subset which produced the highest accuracy (Table 5)) and its associated uncertainty map are presented in Fig. 10 (c) and (d). To check the consistency and analyse the output map, the true colour (RGB) Fenix image and the forest classification map are also shown in Fig. 10 (a) and (b). As it can be compared between the modelled LAI and forest cover map, the variation of LAI corresponds well with the distribution of deciduous, coniferous, and mixed forest stands. Higher LAI values can be observed predominantly in the deciduous stands, followed by mixed stands, while lower LAI values are found in the coniferous area. The average LAI value of all the masked forest pixels is 3.67, which is quite close to the mean LAI of the field sample measurements ($LAI_{mean} = 3.81$). In the obtained LAI uncertainty map (Fig. 10 (d)), pixels with lower σ value indicate more confident estimations retrieved by the trained GPR model. Generally, the uncertainty levels mapped by GPR were low across the entire image.

4. Discussion

This study examined the performance of GPR, a novel kernel-based machine learning algorithm in comparison to the most commonly used empirical methods (i.e., narrowband VIs, PLSR, and ANN) for estimating forest LAI using Fenix airborne hyperspectral data. Our results showed that GPR outperformed the other methods and yielded the most accurate prediction for forest LAI (RMSE_{cv} = 0.53 m² m⁻²). This was despite complex forest structure, signal distortion, and contribution of optical properties from understory layers which usually have undesired impacts on the LAI retrieval processes (Gitelson et al., 2005; Schlerf et al., 2005; Ollinger, 2011). Our results are consistent with those of Verrelst et al. (2012a) and Rivera et al. (2014a, 2014b) who reported better predictive performance of GPR compared with narrowband VIs and classical machine learning algorithms for biophysical parameter estimation in agricultural fields. Predictions obtained using four different approaches were compared for their statistically significant difference from *in situ* data. The results of the boxplot (Fig. A1) and paired *t*-test (Table B1) showed that, in general, there was an agreement between measured and estimated LAI using different approaches ($p > 0.05$). Although this suggests the comparable predictive accuracy for different empirical approaches in retrieving LAI, GPR may be a preferred method for its more accurate estimation of LAI (particularly, when a spectral subset was used). In addition to accurate predictions, using GPR, the variance of the predicted LAI provided insights on the uncertainty level of the model retrievals, which could also contribute to assess the reliability of resulting plant traits map (Wang et al., 2019).

All narrowband VIs produced reasonable LAI predictions, while soil-based VIs performed slightly better (RMSE 0.63–0.64 m² m⁻²) comparing to ratio-based ones (RMSE 0.69–0.71 m² m⁻²). As most of the field plots were within the open canopies where the contribution of the background was pronounced, consideration of soil (understory) parameters improved the LAI estimation. Our finding is also in line with previous studies, which demonstrated the importance of soil-based VIs, particularly in open canopies (Broge and Leblanc, 2001; Darvishzadeh et al., 2009). We also observed that when sample plots were stratified according to individual species, the results of LAI estimation were highly

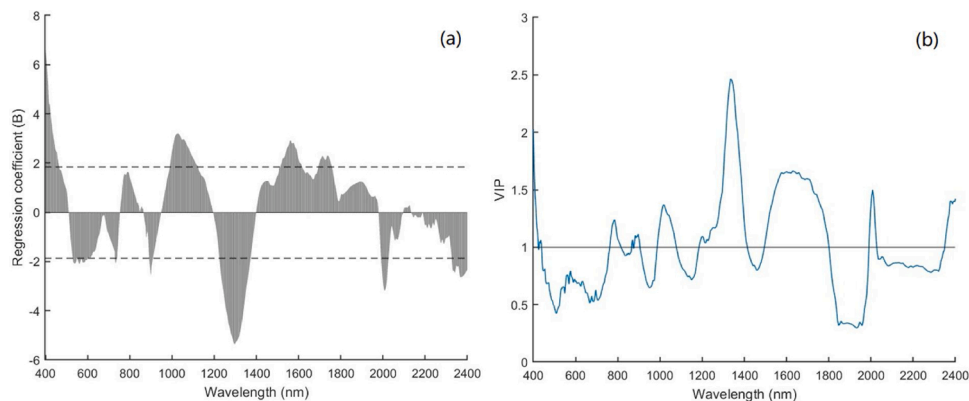


Fig. 5. Importance of wavebands to the PLSR model. (a) Regression coefficient (B) of each wavelength for the PLSR model. Dashed lines represent the standard deviation of B . The regression coefficients larger than its standard deviation (absolute value) indicate a larger influence of the spectral data on the regression model. (b) Variable importance projection (VIP) values for the PLSR model. The VIP values greater than one indicates greater importance in the model.

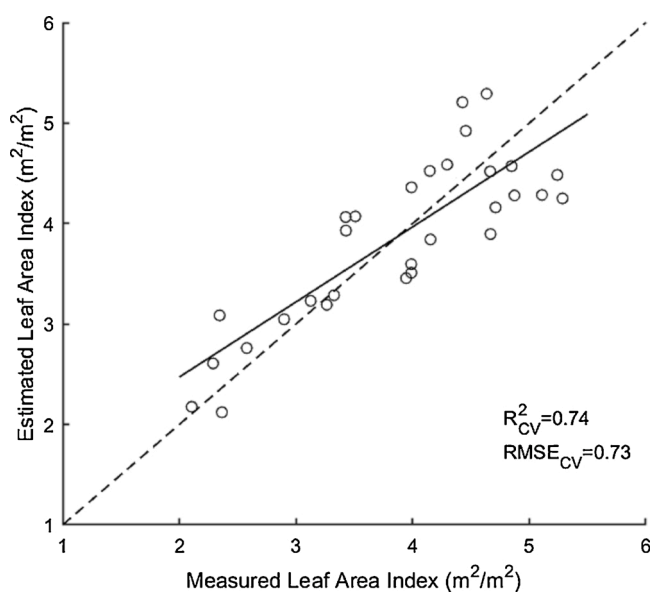


Fig. 6. Measured LAI and estimated LAI calculated from entire reflectance data from Fenix airborne sensor using PLSR (No. of factors = 6). The dashed line shows the 1:1 relationship, while the solid line indicates the relationship between the field measured and estimated values of LAI.

improved than when the pooled data was used (not shown). This finding can be explained as the heterogeneous nature of the mixed forest species which may have different crown density, canopy structure, and leaf angle distribution (Ollinger, 2011; Wang et al., 2016, 2017; Gara et al., 2018).

Narrowband VIs computed from wavebands located in red-edge and SWIR spectral regions yielded higher correlations with LAI (Table 3). The identified important bands are consistent with previous studies that observed canopy reflectance especially in the red-edge and SWIR spectrum are important for LAI estimation (Brown et al., 2000; Gong et al., 2003; Lee et al., 2004; Darvishzadeh et al., 2008b; Verrelst et al., 2016). Spectral data from these wavebands were further used as a spectral subset in PLSR, ANN, and GPR. Although narrowband VIs are simple and computationally cheap retrieval methods of vegetation biophysical properties, their limited use of the full spectral information, as well as sensitivity to sensor configuration and sampling sites, makes them less attractive for quantitative estimation of vegetation parameters (Haboudane et al., 2004; Darvishzadeh et al., 2011; Atzberger et al., 2015).

Compared to narrowband VIs, applying the PLSR model (using the

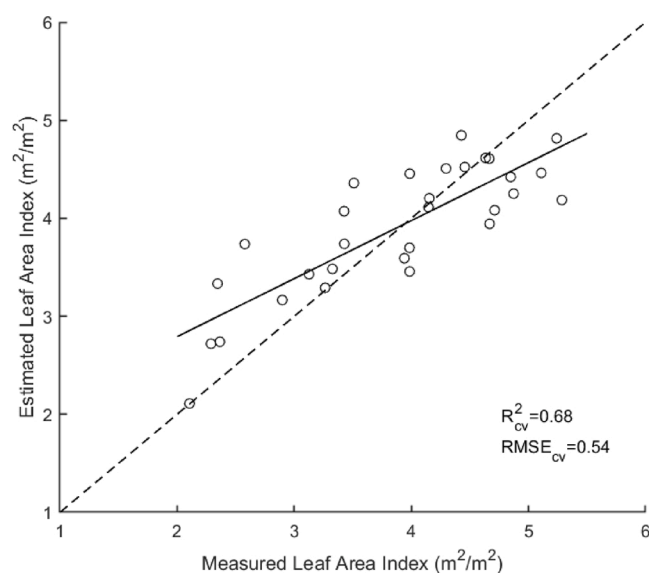


Fig. 7. Measured LAI and estimated LAI calculated from entire reflectance data from Fenix airborne sensor using ANN (No. of neurons = 4). R^2_{cv} and $RMSE_{cv}$ are averaged results of 1,000 random initializations. The dashed line shows the 1:1 relationship, while the solid line indicates the relationship between the field measured and estimated values of LAI.

entire spectra) with six latent factors resulted in a significantly increased correlation coefficient (R^2_{cv}) between measured and estimated LAI by 0.21 to 0.33, but did not reduce the prediction error ($RMSE_{cv}$). Although spectral information from additional wavebands contains more comprehensive information related to the plant traits (Lee et al., 2004), the inclusion of the full-spectrum may also introduce data of noisy bands leading to a deterioration of the model quality (Rivera et al., 2014a, 2014b, 2017; Darvishzadeh et al., 2011). When including the spectral subset obtained from the analysis of narrowband VIs to the PLSR model, the predictive performance of PLSR was improved (Table 5). This result highlights the importance of selecting relevant bands for enhancing LAI estimation using the PLSR model (Cho et al., 2007; Darvishzadeh et al., 2008b). As may be observed in Fig. 5, most of the identified informative bands in the PLSR model are concentrated in the SWIR region and well covered by the former identified important bands in the analysis of narrowband VIs (Table 3). The irregular peaks that are observed at the shortest wavelength (400 nm) and towards longest wavelength (2400 nm) in Fig. 5 can be attributed to the poor signal-to-noise ratio at these regions which is probably caused by the silicon photodiode detector of

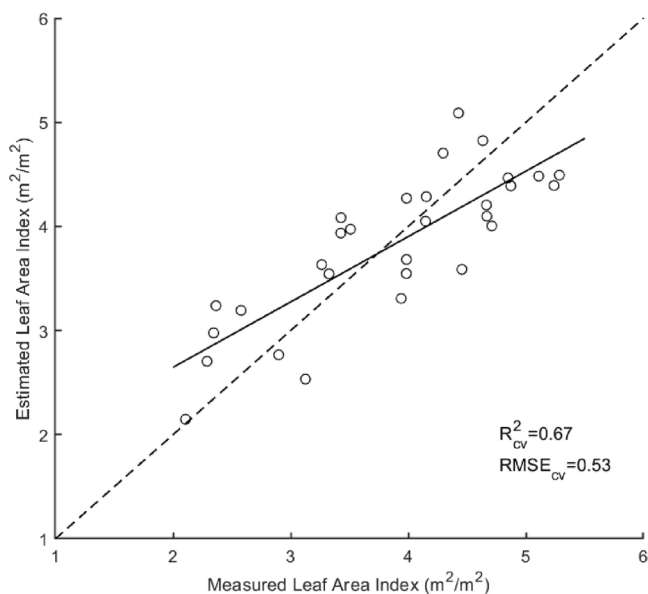


Fig. 8. Measured LAI and estimated LAI using GPR calculated from the entire reflectance data from the Fenix airborne sensor. The dashed line shows the 1:1 relationship, while the solid line indicates the relationship between the field measured and estimated values of LAI.

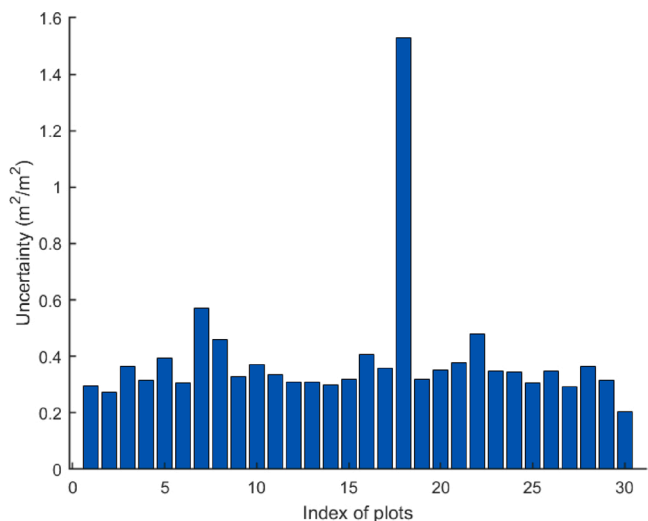


Fig. 9. Uncertainties (σ) of LAI for sample plots ($n = 30$) estimated using GPR from the entire spectral reflectance.

Table 5

Cross-validated results ($RMSE_{cv}$ and R^2_{cv}) for estimating LAI using entire reflectance data versus the spectral subset obtained from the analysis of VIs. Note that the statistical results of ANN are averaged results of 1,000 random initializations. The best model with the highest R^2_{cv} and lowest $RMSE_{cv}$ is boldfaced.

Input	Regression methods	R^2_{cv}	$RMSE_{cv}$
Entire reflectance	PLSR	0.74	0.73
	ANN	0.68	0.54
	GPR	0.67	0.53
Spectral subset	PLSR	0.75	0.69
	ANN	0.68	0.55
	GPR	0.69	0.52

the sensor (Milton et al., 2009). Our results broadly support the work of other studies (Brown et al., 2000; Cohen and Goward, 2004; Darvishzadeh et al., 2011; Rivera et al., 2014a, 2014b; Schlerf et al., 2005) that

demonstrated SWIR region to be an important spectral domain for modelling leaf area index.

The optimum neural net node size for estimating LAI in this study was determined through a comparative analysis of ANN (not shown) and the most accurate results (in terms of both $RMSE_{cv}$ and R^2_{cv}) were achieved using four neurons within the hidden layer. A similar number of neurons was used in the study by Neinavaz et al. (2016) when modelling LAI using a field spectrometer. Compared to the narrowband VIs and PLSR model, LAI estimations were significantly improved when the ANN was applied. This is probably due to the sophisticated training process and highly specialized model developed by ANN, though this property can often make the developed model less robust when inverted and used for prediction (Rivera et al., 2014a, 2014b; Verrelst et al., 2015). Nevertheless, using ANN, the high LAI values appeared to be underestimated (typically for LAI greater than 4.5), while the intermediate LAI values (LAI range of 2–4) were somehow overestimated (Fig. 7). According to Bacour et al. (2006) who observed a similar trend, ANN applies a global training strategy to estimate the variables of interest, therefore the underestimation in the intermediate range can logically be compensated by the overestimation occur for the higher values.

Further, the ANN retrieval accuracy was not improved by reducing the data dimensions using the spectral subset (Table 5). This is in agreement with the finding of Rivera et al. (2017), who showed that adopting the dimensionality reduction methods in the ANN model could not improve the LAI estimation. In the backpropagation training phase, the ANN model is adjusted to minimize the distance between the model output and the training targets. Therefore, the spectral bands which are poorly-related to the target LAI are iteratively identified and adjusted to lower weights during the optimization process (Kimes et al., 1998; Bacour et al., 2006), hence, eliminating them may not have a positive impact on the prediction results.

The higher performance of GPR confirmed the findings of Rivera et al. (2014a, 2014b) and Verrelst et al. (2015) which demonstrated the superiority of GPR over linear parametric methods and classic machine learning algorithms for estimating LAI and LCC (leaf canopy chlorophyll) in agricultural fields. The accuracy of GPR was further improved ($RMSE_{cv} = 0.52 \text{ m}^2 \text{ m}^{-2}$, $R^2_{cv} = 0.69$) using the selected spectral subset obtained from the analysis of narrowband VIs. This increase concurs with Verrelst et al. (2016) and Rivera et al. (2017) who observed that removing less relevant spectral data may enhance the GPR prediction compared to using all hyperspectral bands. The existence of much redundant information could make the model unnecessarily complex and computationally demanding, therefore lead to reduced predictive power (van der Maaten et al., 2009).

An important benefit of GPR is its property of providing the associated uncertainty estimates. Such a property is of interest to the remote sensing community to assess the model reliability and post-evaluate the performance of the calibrated model on mapping products. As can be seen in Fig. 9, most of the plots were estimated with a high confidence level, except Plot 24 (index = 18). This high uncertainty of Plot 24, might be due to the proximity of the plot to a deadwood area which is under-represented in our training dataset, thus yielding a greater uncertainty. We note that the low confidence does not necessarily mean that the prediction is wrong, it only indicates that the input reflectance deviates from the spectrum provided in the training phase, thus leading to an uncertain estimation (Verrelst et al., 2013).

The generated LAI map confirmed the spatial variation of LAI across different forest classes (i.e., higher LAI are observed in deciduous stands than mixed and coniferous stands), which is in agreement with our observation during the field campaign (mean LAI for broadleaf 4.10, mean LAI for conifer 3.46). The same distribution pattern for canopy chlorophyll content was found in BFNP (Ali et al., 2020). The artefacts on the edge of the map (Fig. 10 (c)) are possibly due to the wide FOV sensor and the related BRDF. Uncertainties generated by GPR were generally low across the entire resulting map (Fig. 10 (d)). Regions with high prediction uncertainties infer poorly predicted results and can be

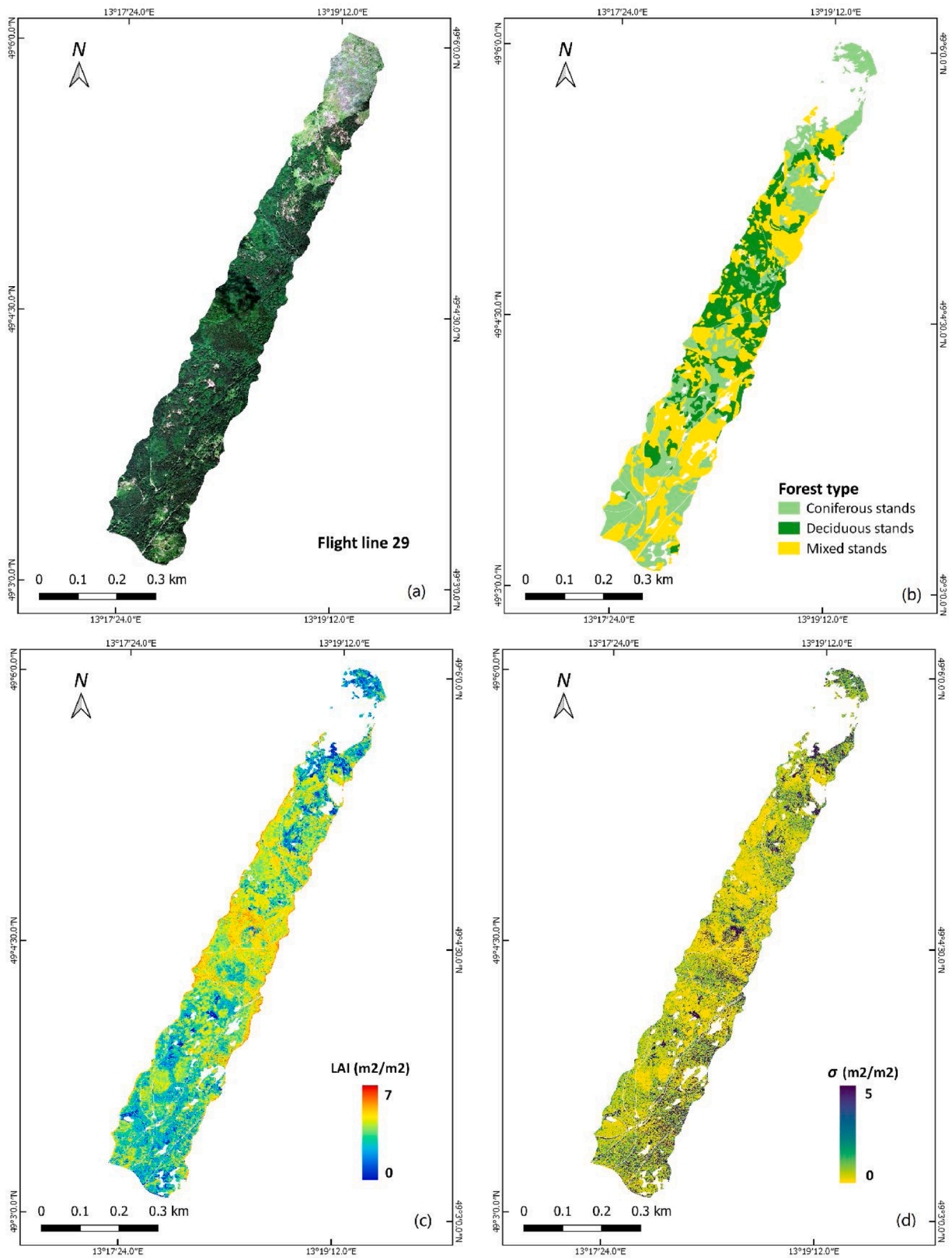


Fig. 10. (a) Fenix airborne hyperspectral data (Flight line 29), (b) forest classification map, (c) modelled LAI in BFNP using GPR and the selected spectral subset, and (d) associated uncertainty estimates.

either masked or improved by collecting additional training data from these areas. As can be interpreted from Fig. 10 ((a), (c), and (d)), higher uncertainties are mainly located in the deadwood area (displayed as pink colour) and the regions with cloud cover or regions with low LAI values (i.e., low vegetation cover). Since these areas were not considered in our training samples, this was somehow expected. Our results also match the finding of previous studies in a grassland experiment that concluded high uncertainties are largely associated with additional treatments (i.e., not represented in the field sampling) or with low vegetation cover (Wang et al., 2019).

Owing to the vast number of contiguous spectral channels, a general challenge of using hyperspectral data is the intrinsic problem of multicollinearity. In the present study, this problem was alleviated by forming a spectral subset obtained from the analysis of narrowband VIs as an alternative to using the whole spectral range. This is perhaps the most typical way for defining a spectral subset to represent the most useful information of the original reflectance (le Maire et al., 2008) compared to extracting a small set of new features obtained from latent factors in principal component analysis or partial least squares (Motoda and Liu, 2002; Lee and Verleysen, 2007), which is limited to be suitable for retrieval of vegetation parameters which have a broad sensitive spectral response (e.g., LAI) (Rivera et al., 2017). Alternatively, selection of important wavelengths from literature to form a spectral subset that can present similar information of the full spectrum (Cho et al., 2007; Faurtyot and Baret, 1997; Darvishzadeh et al., 2008b) would require, a careful literature search, particularly of controlled experimental studies to ensure that the selected wavelengths are truly sensitive to the studied vegetation parameters.

5. Conclusions

This study demonstrated the potential of a recently introduced machine learning algorithm, namely Gaussian processes regression (GPR), for estimating LAI using Fenix airborne hyperspectral data (400–2500 nm) in a heterogeneous mixed mountain forest. Comparing to narrowband VIs, PLSR, and ANN, the most accurate LAI prediction was obtained using GPR ($RMSE_{cv} = 0.53 \text{ m}^2 \text{ m}^{-2}$) in this study. Our findings confirmed the outperformance of GPR over conventional empirical methods in estimating crop LAI as reported in previous studies (Verrelst et al., 2012b; Rivera et al., 2014a, 2014b). The LAI map generated by GPR demonstrated a spatial variation of LAI across forest types. Another important benefit of using GPR is its property of providing the prediction uncertainty estimates along with the predicted values. The uncertainty map shows that LAI uncertainties were generally low across the entire images. Higher uncertainties were observed mainly in the forest areas which were under-represented in the collected sample plots and regions with low LAI values (i.e., low vegetation cover). Therefore, comprehensive data sampling in regions associated with high uncertainties is recommended in future fieldwork to improve our model. Moreover, a spectral subset obtained from the analysis of narrowband VIs generally improved model performance of the studied approaches. This emphasized the importance of utilizing the most useful information and eliminating irrelevant bands for estimating vegetation parameters from hyperspectral data.

With the development of spaceborne hyperspectral missions, global full-range spectral data will be soon available for vegetation monitoring (Labate et al., 2009; Stuffer et al., 2007; Drusch et al., 2017). The upcoming big data stream would thus require methods that can cope well with the hundreds of hyperspectral bands and provide accurate, robust, fast predictions for operational vegetation parameter retrieval. Gaussian processes regression, being able to generate adaptive and robust relationships between image spectral reflectance and target variables and the accompanying uncertainty estimates, shows great potentials to be implemented in the future retrieval applications on a global scale.

CRedit authorship contribution statement

Rui Xie: Conceptualization, Methodology, Formal analysis, Software, Investigation, Writing - original draft. **Roshanak Darvishzadeh:** Conceptualization, Methodology, Supervision, Writing - review & editing. **Andrew K. Skidmore:** Methodology, Supervision, Writing - review & editing. **Marco Heurich:** Data curation, Writing - review & editing. **Stefanie Holzwarth:** Data curation, Writing - review & editing. **Tawanda W. Gara:** Data curation, Writing - review & editing. **Ils Reusen:** Data curation, Writing - review & editing.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

This research work was carried out as a part of the MSc dissertation by the first author (RX) in Faculty ITC, University of Twente. The research leading to these results has received funding from the European Facility for Airborne Research in Environmental and Geo-sciences (EUFAR), a project of the EC's 7th Framework Programme (FP7/2014-2018) under grant agreement n° 312609. The data for this study has been acquired within the framework of the 9th EUFAR Training Course "RS4ForestEBV - Airborne remote sensing for monitoring essential biodiversity variables in forest ecosystems". We are grateful for the support from Dr. Abebe Ali for providing the background reflectance data in the BFN park. The authors would also like to thank the anonymous reviewers for their valuable comments to improve the manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.jag.2020.102242>.

References

- Abdi, H., 2003. Partial least square regression (PLS regression). *Encyclopedia Res. Methods Soc. Sci.* 6 (4), 792–795.
- Ali, A.M., Skidmore, A.K., Darvishzadeh, R., van Duren, I., Holzwarth, S., Mueller, J., 2016. Retrieval of forest leaf functional traits from HySpex imagery using radiative transfer models and continuous wavelet analysis. *ISPRS J. Photogramm. Remote Sens.* 122, 68–80. <https://doi.org/10.1016/j.isprsjprs.2016.09.015>.
- Ali, A.M., Darvishzadeh, R., Skidmore, A.K., Gara, T.W., O'Connor, B., Rocolesi, C., Heurich, M., Paganini, M., et al., 2020. Comparing methods for mapping canopy chlorophyll content in a mixed mountain forest using Sentinel-2 data. *Int. J. Appl. Earth Obs. Geoinf.* 87 <https://doi.org/10.1016/j.jag.2019.102037>.
- Ali, A.M., Darvishzadeh, R., Skidmore, A.K., van Duren, I., 2017. Specific leaf area estimation from leaf and canopy reflectance through optimization and validation of vegetation indices. *Agric. For. Meteorol.* 236, 162–174.
- Atzberger, C., 2000. Development of an invertible forest reflectance model: the INFOR-model. A Decade of Trans-European Remote Sensing Cooperation, Proceedings of the 20th EARSEL Symposium 39–44.
- Atzberger, C., Guérif, M., Baret, F., Werner, W., 2010. Comparative analysis of three chemometric techniques for the spectroradiometric assessment of canopy chlorophyll content in winter wheat. *Comput. Electron. Agric.* 73 (2), 165–173. <https://doi.org/10.1016/j.compag.2010.05.006>.
- Atzberger, C., Darvishzadeh, R., Schlerf, M., Le Maire, G., 2013. Suitability and adaptation of PROSAIL radiative transfer model for hyperspectral grassland studies. *Remote Sens. Lett.* 4 (1), 55–64.
- Atzberger, C., Darvishzadeh, R., Immitzer, M., Schlerf, M., Skidmore, A.K., le Maire, G., 2015. Comparative analysis of different retrieval methods for mapping grassland leaf area index using airborne imaging spectroscopy. *Int. J. Appl. Earth Obs. Geoinf.* <https://doi.org/10.1016/j.jag.2015.01.009>.
- Bacour, C., Baret, F., Béal, D., Weiss, M., Pavageau, K., 2006. Neural network estimation of LAI, fAPAR, fCover and LAI×Cab, from top of canopy MERIS reflectance data: principles and validation. *Remote Sens. Environ.* 105 (4), 313–325. <https://doi.org/10.1016/j.rse.2006.07.014>.
- Baret, Frédéric, Buis, S., 2008. Estimating canopy characteristics from remote sensing observations: review of methods and associated problems. In: Liang, S. (Ed.), *Advances in Land Remote Sensing: System, Modeling, Inversion and Application*, pp. 173–201. https://doi.org/10.1007/978-1-4020-6450-0_7.

- Baret, F., Guyot, G., 1991. Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sens. Environ.* 35 (2–3), 161–173. [https://doi.org/10.1016/0034-4257\(91\)90009-U](https://doi.org/10.1016/0034-4257(91)90009-U).
- Baret, Frédéric, Guyot, G., Major, D., 1989. TSAVI: a vegetation index which minimizes soil brightness effects on LAI and APAR estimation. 12th Canadian Symposium on Remote Sensing Geoscience and Remote Sensing Symposium 3, 1355–1358.
- Birdsey, R., Pan, Y., 2011. Drought and dead trees. *Nat. Clim. Chang.* 1 (9), 444–445.
- Bojinski, S., Verstraete, M., Peterson, T.C., Richter, C., Simmons, A., Zemp, M., 2014. The concept of essential climate variables in support of climate research, applications, and policy. *Bull. Am. Meteorol. Soc.* 95 (9), 1431–1443.
- Broge, N., Leblanc, E., 2001. Comparing prediction power and stability of broadband and hyperspectral vegetation indices for estimation of green leaf area index and canopy chlorophyll density. *Remote Sens. Environ.* 76 (2), 156–172. [https://doi.org/10.1016/S0034-4257\(00\)00197-8](https://doi.org/10.1016/S0034-4257(00)00197-8).
- Brown, L., Chen, J.M., Leblanc, S.G., Cihlar, J., 2000. A shortwave infrared modification to the simple ratio for LAI retrieval in boreal forests: an image and model analysis. *Remote Sens. Environ.* 71 (1), 16–25. [https://doi.org/10.1016/S0034-4257\(99\)00035-8](https://doi.org/10.1016/S0034-4257(99)00035-8).
- Caillieret, M., Heurich, M., Bugmann, H., 2014. Reduction in browsing intensity may not compensate climate change effects on tree species composition in the Bavarian Forest National Park. *For. Ecol. Manage.* 328, 179–192. <https://doi.org/10.1016/J.FORECO.2014.05.030>.
- Canadell, J.G., Mooney, H.A., Baldocchi, D.D., Berry, J.A., Ehleringer, J.R., Field, C.B., et al., 2000. Commentary: carbon metabolism of the terrestrial biosphere: a multitechnique approach for improved understanding. *Ecosystems* 3 (2), 115–130.
- Chen, J.M., Black, T.A., 1992. Defining leaf area index for non-flat leaves. *Plant Cell Environ.* 15 (4), 421–429. <https://doi.org/10.1111/j.1365-3040.1992.tb00992.x>.
- Chen, J.M., Rich, P.M., Gower, S.T., Norman, J.M., Plummer, S., 1997. Leaf area index of boreal forests: theory, techniques, and measurements. *J. Geophys. Res. Atmos.* 102 (D24), 29429–29443. <https://doi.org/10.1029/97jd01107>.
- Cho, M.A., Skidmore, A.K., 2006. A new technique for extracting the red edge position from hyperspectral data: the linear extrapolation method. *Remote Sens. Environ.* 101 (2), 181–193. <https://doi.org/10.1016/J.RSE.2005.12.011>.
- Cho, M.A., Skidmore, A.K., Corsi, F., van Wieren, S.E., Sobhan, I., 2007. Estimation of green grass/herb biomass from airborne hyperspectral imagery using spectral indices and partial least squares regression. *Int. J. Appl. Earth Obs. Geoinf.* 9 (4), 414–424. <https://doi.org/10.1016/J.JAG.2007.02.001>.
- Cohen, W.B., Goward, S.N., 2004. Landsat's Role in Ecological Applications of Remote Sensing.
- Cohen, W.B., Maierperger, T.K., Gower, S.T., Turner, D.P., 2003. An improved strategy for regression of biophysical variables and Landsat ETM+ data. *Remote Sens. Environ.* 84 (4), 561–571. [https://doi.org/10.1016/S0034-4257\(02\)00173-6](https://doi.org/10.1016/S0034-4257(02)00173-6).
- Combal, B., Baret, F., Weiss, M., Trubuil, A., Macé, D., Pragnère, A., et al., 2003. Retrieval of canopy biophysical variables from bidirectional reflectance: using prior information to solve the ill-posed inverse problem. *Remote Sens. Environ.* 84 (1), 1–15. [https://doi.org/10.1016/S0034-4257\(02\)00035-4](https://doi.org/10.1016/S0034-4257(02)00035-4).
- Curran, P.J., 1989. Remote sensing of foliar chemistry. *Remote Sens. Environ.* 30 (3), 271–278. [https://doi.org/10.1016/0034-4257\(89\)90069-2](https://doi.org/10.1016/0034-4257(89)90069-2).
- Darvishzadeh, R., Skidmore, A.K., Abdullah, H., Cherenet, E., Ali, A.M., Wang, T., et al., 2019. Mapping leaf chlorophyll content from Sentinel-2 and RapidEye data in spruce stands using the invertible forest reflectance model. *Int. J. Appl. Earth Obs. Geoinf.* 79, 58–70. <https://doi.org/10.1016/j.jag.2019.03.003>.
- Darvishzadeh, R., Skidmore, A.K., Schlerf, M., Atzberger, C., 2008a. Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland. *Remote Sens. Environ.* 112 (5), 2592–2604. <https://doi.org/10.1016/J.RSE.2007.12.003>.
- Darvishzadeh, R., Skidmore, A.K., Schlerf, M., Atzberger, C., Corsi, F., Cho, M., 2008b. LAI and chlorophyll estimation for a heterogeneous grassland using hyperspectral measurements. *ISPRS J. Photogramm. Remote Sens.* 63 (4), 409–426. <https://doi.org/10.1016/j.isprsjprs.2008.01.001>.
- Darvishzadeh, R., Atzberger, C., Skidmore, A.K., Abkar, A.A., 2009. Leaf Area Index derivation from hyperspectral vegetation indices and the red edge position. *Int. J. Remote Sens.* 30 (23), 6199–6218. <https://doi.org/10.1080/01431160902842342>.
- Darvishzadeh, R., Atzberger, C., Skidmore, A.K., Schlerf, M., 2011. Mapping grassland leaf area index with airborne hyperspectral imagery: a comparison study of statistical approaches and inversion of radiative transfer models. *ISPRS J. Photogramm. Remote Sens.* 66 (6), 894–906. <https://doi.org/10.1016/j.isprsjprs.2011.09.013>.
- Daszykowski, M., Serneels, S., Kaczmarek, K., Van Espen, P., Croux, C., Walczak, B., 2007. TOMCAT: a MATLAB toolbox for multivariate calibration techniques. *Chemom. Intell. Lab. Syst.* 85 (2), 269–277. <https://doi.org/10.1016/J.CHEMOLAB.2006.03.006>.
- De Jong, S.M., Pebesma, E.J., Lacaze, B., 2003. Above-ground biomass assessment of Mediterranean forests using airborne imaging spectrometry: the DAIS Peyne experiment. *Int. J. Remote Sens.* 24 (7), 1505–1520.
- Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., et al., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* 36 (1), 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>.
- Drusch, M., Moreno, J., Del Bello, U., Franco, R., Goulas, Y., Huth, A., et al., 2017. The FLUorescence Explorer mission concept—ESA's earth explorer 8. *IEEE Trans. Geosci. Remote Sens.* 55 (3), 1273–1284. <https://doi.org/10.1109/TGRS.2016.2621820>.
- Faurtyot, T., Baret, F., 1997. Vegetation water and dry matter contents estimated from top-of-the-atmosphere reflectance data: a simulation study. *Remote Sens. Environ.* 61 (1), 34–45. [https://doi.org/10.1016/S0034-4257\(96\)00238-6](https://doi.org/10.1016/S0034-4257(96)00238-6).
- Fischer, A., Kergoat, L., Dedieu, G., 1997. Coupling satellite data with vegetation functional models: review of different approaches and perspectives suggested by the assimilation strategy. *Remote Sens. Rev.* 15 (1–4), 283–303. <https://doi.org/10.1080/02757259709532343>.
- Food and Agriculture Organization of the United Nations, 2010. *Global Forest Resources Assessment 2010*. Desk Reference, 378.
- Gara, T.W., Darvishzadeh, R., Skidmore, A.K., Wang, T., 2018. Impact of vertical canopy position on leaf spectral properties and traits across multiple species. *Remote Sens.* 10 (2), 346.
- Gara, T.W., Darvishzadeh, R., Skidmore, A.K., Wang, T., Heurich, M., 2019. Accurate modelling of canopy traits from seasonal Sentinel-2 imagery based on the vertical distribution of leaf traits. *ISPRS J. Photogramm. Remote Sens.* 157, 108–123. <https://doi.org/10.1016/J.ISPRSJPRS.2019.09.005>.
- Geladi, P., Kowalski, B.R., 1986. Partial least-squares regression: a tutorial. *Anal. Chim. Acta* 185, 1–17. [https://doi.org/10.1016/0003-2670\(86\)80028-9](https://doi.org/10.1016/0003-2670(86)80028-9).
- Gewali, U.B., Monteiro, S.T., Saber, E., 2019. Gaussian processes for vegetation parameter estimation from hyperspectral data with limited ground truth. *Remote Sens.* 11 (13) <https://doi.org/10.3390/rs11131614>.
- Gitelson, A.A., Vina, A., Ciganda, V., Rundquist, D.C., Arkebauer, T.J., 2005. Remote estimation of canopy chlorophyll content in crops. *Geophys. Res. Lett.* 32 (8).
- Gong, P., Pu, R., Biging, G.S., Larrieu, M.R., 2003. Estimation of forest leaf area index using vegetation indices derived from Hyperion hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* 41 (6 PART 1), 1355–1362. <https://doi.org/10.1109/TGRS.2003.812910>.
- Gower, S.T., Kucharik, C.J., Norman, J.M., 1999. Direct and indirect estimation of leaf area index, fAPAR, and net primary production of terrestrial ecosystems. *Remote Sens. Environ.* 70 (1), 29–51. [https://doi.org/10.1016/S0034-4257\(99\)00056-5](https://doi.org/10.1016/S0034-4257(99)00056-5).
- Haaland, D.M., Thomas, E.V., 1988. Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information. *Anal. Chem.* 60 (11), 1193–1202. <https://doi.org/10.1021/ac00162a020>.
- Haboudane, D., Miller, J.R., Pattey, E., Zarco-Tejada, P.J., Strachan, I.B., 2004. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: modeling and validation in the context of precision agriculture. *Remote Sens. Environ.* 90 (3), 337–352. <https://doi.org/10.1016/j.rse.2003.12.013>.
- Halme, E., Pellikka, P., Mõttus, M., 2019. Utility of hyperspectral compared to multispectral remote sensing data in estimating forest biomass and structure variables in Finnish boreal forest. *Int. J. Appl. Earth Obs. Geoinf.* 83 (August), 101942. <https://doi.org/10.1016/j.jag.2019.101942>.
- Hansen, P.M., Schjoerring, J.K., 2003. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sens. Environ.* 86 (4), 542–553. [https://doi.org/10.1016/S0034-4257\(03\)00131-7](https://doi.org/10.1016/S0034-4257(03)00131-7).
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media.
- Heurich, M., Beudert, B., Rall, H., Kvenová, Z., 2010. national parks as model regions for interdisciplinary Long-term ecological research: the Bavarian Forest and Šumavá national parks underway to transboundary ecosystem research. In: Müller, F., Baessler, C., Schubert, H., Klotz, S. (Eds.), *Long-Term Ecological Research: Between Theory and Application*, pp. 327–344. https://doi.org/10.1007/978-90-481-8782-9_23.
- Hill, J., Buddenbaum, H., Townsend, P.A., 2019. Imaging spectroscopy of forest ecosystems: perspectives for the use of space-borne hyperspectral earth observation systems. *Surv. Geophys.* 40 (3), 553–588. <https://doi.org/10.1007/s10712-019-09514-2>.
- Houborg, R., Soegaard, H., Boegh, E., 2007. Combining vegetation index and model inversion methods for the extraction of key vegetation biophysical parameters using Terra and Aqua MODIS reflectance data. *Remote Sens. Environ.* 106 (1), 39–58. <https://doi.org/10.1016/J.RSE.2006.07.016>.
- Huang, Z., Turner, B.J., Dury, S.J., Wallis, I.R., Foley, W.J., 2004. Estimating foliage nitrogen concentration from HYMAP data using continuum removal analysis. *Remote Sens. Environ.* 93 (1), 18–29. <https://doi.org/10.1016/j.rse.2004.06.008>.
- Kimes, D.S., Nelson, R.F., Manry, M.T., Fung, A.K., 1998. Review article: attributes of neural networks for extracting continuous vegetation variables from optical and radar measurements. *Int. J. Remote Sens.* 19 (14), 2639–2663. <https://doi.org/10.1080/014311698214433>.
- Klos, R.J., Wang, G.G., Bauerle, W.L., Rieck, J.R., 2009. Drought impact on forest growth and mortality in the southeast USA: an analysis using Forest Health and monitoring data. *Ecol. Appl.* 19 (3), 699–708. <https://doi.org/10.1890/08-0330.1>.
- Kokaly, R.F., Clark, R.N., 1999. Spectroscopic determination of leaf biochemistry using band-depth analysis of absorption features and stepwise multiple linear regression. *Remote Sens. Environ.* 67 (3), 267–287.
- Kooistra, L., Salas, E.A.L., Clevers, J.G.P., Wehrens, R., Leuven, R.S.E., Nienhuis, P.H., Buydens, L.M.C., 2004. Exploring field vegetation reflectance as an indicator of soil contamination in river floodplains. *Environ. Pollut.* 127 (2), 281–290. [https://doi.org/10.1016/S0269-7491\(03\)00266-5](https://doi.org/10.1016/S0269-7491(03)00266-5).
- Labate, D., Ceccherini, M., Cisbani, A., De Cosmo, V., Galeazzi, C., Giunti, L., et al., 2009. The PRISMA payload optomechanical design, a high performance instrument for a new hyperspectral mission. *Acta Astronaut.* 65 (9–10), 1429–1436. <https://doi.org/10.1016/J.ACTAASTRO.2009.03.077>.
- le Maire, G., François, C., Soudani, K., Berweiler, D., Pontailleur, J.-Y., Bréda, N., et al., 2008. Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass. *Remote Sens. Environ.* 112 (10), 3846–3864. <https://doi.org/10.1016/j.rse.2008.06.005>.

- Lee, J.A., Verleysen, M., 2007. *Nonlinear Dimensionality Reduction*. Springer Science & Business Media.
- Lee, K.S., Cohen, W.B., Kennedy, R.E., Maierberger, T.K., Gower, S.T., 2004. Hyperspectral versus multispectral data for estimating leaf area index in four different biomes. *Remote Sens. Environ.* 91 (3–4), 508–520. <https://doi.org/10.1016/j.rse.2004.04.010>.
- Li, X., Zhang, Y., Bao, Y., Luo, J., Jin, X., Xu, X., et al., 2014. Exploring the best hyperspectral features for LAI estimation using partial least squares regression. *Remote Sens. (Basel)* 6 (7), 6221–6241.
- Liang, L., Di, L., Zhang, L., Deng, M., Qin, Z., Zhao, S., Lin, H., 2015. Estimation of crop LAI using hyperspectral vegetation indices and a hybrid inversion method. *Remote Sens. Environ.* 165, 123–134.
- Li-Cor, L., 1992. LAI-2000 Plant Canopy Analyzer Instruction Manual. LI-COR Inc., Lincoln, Nebraska, USA.
- Major, D.J., Baret, F., Guyot, G., 1990. A ratio vegetation index adjusted for soil brightness. *Int. J. Remote Sens.* 11 (5), 727–740. <https://doi.org/10.1080/01431169008955053>.
- MATLAB, 2019. R2019a. The MathWorks Inc., Natick, Massachusetts.
- Milton, E.J., Schaeppman, M.E., Anderson, K., Kneubühler, M., Fox, N., 2009. Progress in field spectroscopy. *Remote Sens. Environ.* 113, S92–S109. <https://doi.org/10.1016/j.rse.2007.08.001>.
- Motoda, H., Liu, H., 2002. Feature selection, extraction and construction. *Communication of IICM (Institute of Information and Computing Machinery, Taiwan)* 5 (67–72), 2.
- Mutanga, O., Skidmore, A.K., 2004. Narrow band vegetation indices overcome the saturation problem in biomass estimation. *Int. J. Remote Sens.* 25 (19), 3999–4014.
- Neinavaz, E., Skidmore, A.K., Darvishzadeh, R., Groen, T.A., 2016. Retrieval of leaf area index in different plant species using thermal hyperspectral data. *ISPRS J. Photogramm. Remote. Sens.* 119, 390–401. <https://doi.org/10.1016/j.isprsjprs.2016.07.001>.
- Nguyen, D., Widrow, B., 1990. The truck backer-upper: an example of self-learning in neural networks. *Advanced Neural Computers*. Elsevier, pp. 11–19.
- Norman, J.M., Campbell, G.S., 1989. *Canopy structure*. *Plant Physiological Ecology*. Springer, pp. 301–325.
- Nowlan, S.J., Hinton, G.E., 1992. Simplifying neural networks by Soft weight-Sharing. *Neural Comput.* 4 (4), 473–493. <https://doi.org/10.1162/neco.1992.4.4.473>.
- Ollinger, S.V., 2011. Sources of variability in canopy reflectance and the convergent properties of plants. *New Phytol.* 189 (2), 375–394. <https://doi.org/10.1111/j.1469-8137.2010.03536.x>.
- Pasolli, F., Melgani, F., Blanzieri, E., 2010. Gaussian process regression for estimating chlorophyll concentration in subsurface waters from remote sensing data. *Ieee Geosci. Remote. Sens. Lett.* 7 (3), 464–468. <https://doi.org/10.1109/LGRS.2009.2039191>.
- Pearson, R.L., Miller, L.D., 1972. *Remote Mapping of Standing Crop Biomass for Estimation of the Productivity of the Shortgrass Prairie*. Pawnee National Grasslands, Colorado.
- Richter, R., Schläpfer, D., 2019. *Atmospheric/Topographic correction for airborne imagery: ATCOR-4 user Guide*. DLR Report DLR-IB 565-02/2019. Wessling, Germany.
- Rivera, J.P., Verrelst, J., Delegido, J., Veroustraete, F., Moreno, J., 2014a. On the semi-automatic retrieval of biophysical parameters based on spectral index optimization. *Remote Sens.* 6 (6), 4927–4951. <https://doi.org/10.3390/rs6064927>.
- Rivera, J.P., Verrelst, J., Muñoz-Mari, J., Moreno, J., Camps-Valls, G., 2014b. Toward a semiautomatic machine learning retrieval of biophysical parameters. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* 7 (4), 1249–1259. <https://doi.org/10.1109/JSTARS.2014.2298752>.
- Rivera, J.P., Verrelst, J., Muñoz-Marí, J., Camps-Valls, G., Moreno, J., 2017. Hyperspectral dimensionality reduction for biophysical variable statistical retrieval. *ISPRS J. Photogramm. Remote. Sens.* 132, 88–101. <https://doi.org/10.1016/j.isprsjprs.2017.08.012>.
- Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., Harlan, J.C., 1974. *Monitoring the Vernal Advancement of Retrogradation of Natural Vegetation*. NASA/GSFC.
- Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36 (8), 1627–1639. <https://doi.org/10.1021/ac60214a047>.
- Schlerf, M., Atzberger, C., Udelhoven, T., Jarmer, T., Mader, S., Werner, W., Hill, J., 2003. Spectrometric Estimation of Leaf Pigments in Norway Spruce Needles Using Band - Depth Analysis, Partial Least - Square Regression and Inversion of a Conifer Leaf Model.
- Schlerf, M., Atzberger, C., Hill, J., 2005. Remote sensing of forest biophysical variables using HyMap imaging spectrometer data. *Remote Sens. Environ.* 95 (2), 177–194. <https://doi.org/10.1016/j.rse.2004.12.016>.
- Schlerf, M., Atzberger, C., Hill, J., Buddenbaum, H., Werner, W., Schüler, G., 2010. Retrieval of chlorophyll and nitrogen in Norway spruce (*Picea abies* L. Karst.) using imaging spectroscopy. *Int. J. Appl. Earth Obs. Geoinf.* 12 (1), 17–26. <https://doi.org/10.1016/J.JAG.2009.08.006>.
- Silveira Gonzalez, R., Latifi, H., Weinacker, H., Dees, M., Koch, B., Heurich, M., 2018. Integrating LIDAR and high-resolution imagery for object-based mapping of forest habitats in a heterogeneous temperate forest landscape. *Int. J. Remote Sens.* 39 (23), 8859–8884. <https://doi.org/10.1080/01431161.2018.1500071>.
- Singh, A., Serbin, S.P., McNeil, B.E., Kingdon, C.C., Townsend, P.A., 2015. Imaging spectroscopy algorithms for mapping canopy foliar chemical and morphological traits and their uncertainties. *Ecol. Appl.* 25 (8), 2180–2197.
- Skidmore, A.K., 2002. Taxonomy of environmental models in the spatial sciences. In: Skidmore, A. (Ed.), *Environmental Modelling With GIS and Remote Sensing*, pp. 8–24. https://doi.org/10.1201/9780203302217_ch2.
- Skidmore, A.K., Turner, B.J., Brinkhof, W., Knowles, E., 1997. Performance of a neural network: mapping forests using GIS and remotely sensed data. *Photogrammetric Engineering and Remote Sensing: PE&RS* 63 (5), 501–514.
- Skidmore, A.K., Pettorelli, N., Coops, N.C., Geller, G.N., Matthew, H., 2015. Agree on biodiversity metrics to track from space: ecologists and space agencies must forge a global monitoring strategy. *Nature* 523, 403–406. <https://doi.org/10.1038/523403a>.
- Stuffer, T., Kaufmann, C., Hofer, S., Förster, K.P., Schreier, G., Mueller, A., et al., 2007. The EnMAP hyperspectral imager—an advanced optical payload for future applications in Earth observation programmes. *Acta Astronaut.* 61 (1–6), 115–120. <https://doi.org/10.1016/J.ACTAASTRO.2007.01.033>.
- Thenkabail, P.S., Smith, R.B., De Pauw, E., 2000. Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. *Remote Sens. Environ.* 71 (2), 158–182. [https://doi.org/10.1016/S0034-4257\(99\)00067-X](https://doi.org/10.1016/S0034-4257(99)00067-X).
- van der Maaten, L., Postma, E.O., van den Herik, J., 2009. *Dimensionality Reduction: A Comparative Review*.
- Verrelst, J., Muñoz, J., Alonso, L., Delegido, J., Rivera, J.P., Camps-Valls, G., Moreno, J., 2012a. Machine learning regression algorithms for biophysical parameter retrieval: opportunities for Sentinel-2 and -3. *Remote Sens. Environ.* 118, 127–139. <https://doi.org/10.1016/J.RSE.2011.11.002>.
- Verrelst, J., Alonso, L., Camps-Valls, G., Delegido, J., Moreno, J., 2012b. Retrieval of vegetation biophysical parameters using Gaussian process techniques. *IEEE Trans. Geosci. Remote. Sens.* 50 (5 PART 2), 1832–1843. <https://doi.org/10.1109/TGRS.2011.2168962>.
- Verrelst, J., Alonso, L., Rivera Caicedo, J.P., Moreno, J., Camps-Valls, G., 2013. Gaussian process retrieval of chlorophyll content from imaging spectroscopy data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens.* 6 (2), 867–874. <https://doi.org/10.1109/JSTARS.2012.2222356>.
- Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J.P., Veroustraete, F., Clevers, J.G.P.W., Moreno, J., 2015. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties - A review. *ISPRS J. Photogramm. Remote. Sens.* 108, 273–290. <https://doi.org/10.1016/j.isprsjprs.2015.05.005>.
- Verrelst, J., Rivera, J.P., Gitelson, A., Delegido, J., Moreno, J., Camps-Valls, G., 2016. Spectral band selection for vegetation properties retrieval using Gaussian processes regression. *Int. J. Appl. Earth Obs. Geoinf.* 52, 554–567. <https://doi.org/10.1016/j.jag.2016.07.016>.
- Verrelst, J., Malenovsky, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J.P., Lewis, P., et al., 2019. Quantifying vegetation biophysical variables from imaging spectroscopy data: a review on retrieval methods. *Surv. Geophys.* 40 (3), 589–629. <https://doi.org/10.1007/s10712-018-9478-y>.
- Wang, Z., Wang, T., Darvishzadeh, R., Skidmore, A.K., Jones, S., Suarez, L., et al., 2016. Vegetation indices for mapping canopy foliar nitrogen in a mixed temperate forest. *Remote Sens.* 8 (6) <https://doi.org/10.3390/rs8060491>.
- Wang, Z., Skidmore, A.K., Wang, T., Darvishzadeh, R., Heiden, U., Heurich, M., et al., 2017. Canopy foliar nitrogen retrieved from airborne hyperspectral imagery by correcting for canopy structure effects. *Int. J. Appl. Earth Obs. Geoinf.* 54, 84–94.
- Wang, Z., Townsend, P.A., Schweiger, A.K., Couture, J.J., Singh, A., Hobbie, S.E., Cavender-Bares, J., 2019. Mapping foliar functional traits and their uncertainties across three years in a grassland experiment. *Remote Sens. Environ.* 221, 405–416. <https://doi.org/10.1016/j.rse.2018.11.016>.
- Weiss, M., Baret, F., 1999. Evaluation of canopy biophysical variable retrieval performances from the accumulation of large swath satellite data. *Remote Sens. Environ.* 70 (3), 293–306. [https://doi.org/10.1016/S0034-4257\(99\)00045-0](https://doi.org/10.1016/S0034-4257(99)00045-0).
- Weiss, M., Baret, F., Smith, G.J., Jonckheere, I., Coppin, P., 2004. Review of methods for in situ leaf area index (LAI) determination Part II. Estimation of LAI, errors and sampling. *Agric. For. Meteorol.* 121 (1–2), 37–53. <https://doi.org/10.1016/j.agrformet.2003.08.001>.
- Williams, C.K.I., Rasmussen, C.E., 2006. *Gaussian Processes for Machine Learning*, Vol. 2. MIT press Cambridge, MA.
- Yuan, H., Yang, G., Li, C., Wang, Y., Liu, J., Yu, H., et al., 2017. Retrieving soybean leaf area index from unmanned aerial vehicle hyperspectral remote sensing: analysis of RF, ANN, and SVM regression models. *Remote Sens.* 9 (4) <https://doi.org/10.3390/rs9040309>.