

AIRBORNE HYPERSPECTRAL DATA FOR ESTIMATION AND MAPPING OF FOREST LEAF AREA INDEX

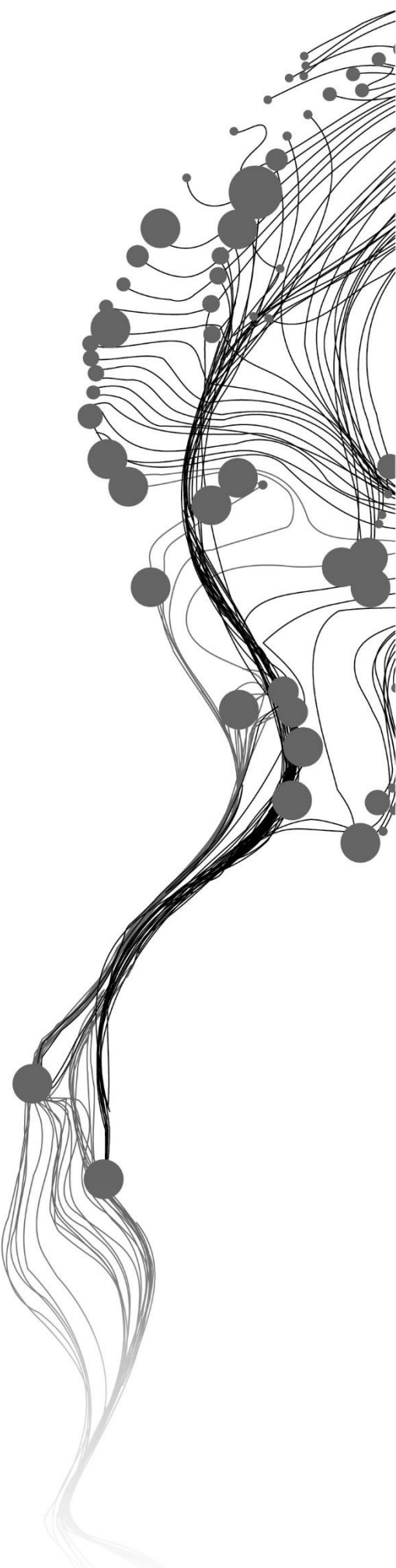
RUI XIE

June, 2020

SUPERVISORS:

Dr. R. (Roshanak) Darvishzadeh

Prof. dr. A.K. (Andrew) Skidmore

An abstract graphic on the left side of the page, resembling a stylized tree or a complex network. It features a central vertical axis from which numerous thin, curved lines branch out. At the ends of these lines are various sized dark gray circles, some of which are significantly larger than others, creating a sense of depth and complexity. The overall shape is roughly triangular, with the base at the bottom left and the top right.

AIRBORNE HYPERSPECTRAL DATA FOR ESTIMATION AND MAPPING OF FOREST LEAF AREA INDEX

RUI XIE

Enschede, The Netherlands, June, 2020

Thesis submitted to the Faculty of Geo-Information Science and Earth
Observation of the University of Twente in partial fulfilment of the
requirements for the degree of Master of Science in Spatial Engineering

SUPERVISORS:

Dr. R. Darvishzadeh

Prof. dr. A.K. Skidmore

THESIS ASSESSMENT BOARD:

Prof. dr. A.D. Nelson (Chair)

Prof. Moses Azong Cho (External Examiner, Council for Scientific and
Industrial Research, Pretoria, South Africa)

DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

ABSTRACT

Forest ecosystems cover about 30% of the earth's land surface and provide a significant contribution to the terrestrial biodiversity, biomass and carbon storage, as well as timber production. Quantitative timely information about the forest canopy cover and characteristics is important for ecologists and decision-makers to assess the influence of climate change and expanding human activities on forest ecosystems. However, traditional field sampling of plant traits is often laborious and limited to small areas. Remote sensing, because of its repetitiveness, cost-effectiveness, and non-destructive characterization of land surfaces, has been recognized as a prevalent technology and a practical mean for monitoring forest canopy characteristics over a large scale. Among many characteristics, leaf area index (LAI) is a widely used biophysical parameter to quantify forest health and growth. Thus, accurate estimating LAI and mapping its spatial distribution is crucial for forest management and many ecological studies. Among existing remote sensing-based methods, machine learning algorithms, in particular, kernel-based machine learning methods, such as Gaussian processes regression (GPR), have shown to be promising alternatives to conventional empirical methods for retrieving vegetation parameters. However, the performance of GPR in predicting forest biophysical parameters has hardly been examined in the literature. The main objective of this study was to evaluate the potential of GPR to estimate forest 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. The 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 m² m⁻²) compared to the best performing narrowband vegetation indices SAVI2 ($R_{cv}^2 = 0.54$, RMSE_{cv} = 0.63 m² m⁻²), PLSR ($R_{cv}^2 = 0.74$, RMSE_{cv} = 0.73 m² m⁻²) and ANN ($R_{cv}^2 = 0.68$, RMSE = 0.54 m² m⁻²). Consequently, when a spectral subset obtained from the analysis of VIs was used as input, the predictive accuracies were generally improved (GPR RMSE_{cv} = 0.52 m² m⁻²; ANN RMSE_{cv} = 0.55 m² m⁻²; PLSR RMSE_{cv} = 0.69 m² 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. The results of this study demonstrated the potential utility 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. The generated trait maps can offer spatially explicit and continuous information of vegetation to effectively support sustainable forest management and resource decision-making.

Keywords: hyperspectral data, leaf area index, GPR, VIs, PLSR, ANN, uncertainty; spectral subset; temperate mixed forests.

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest gratitude to my supervisors, Dr. Roshanak Darvishzadeh and Prof. Andrew Skidmore. Thanks for their tremendous help and instrumental advice in supervising the whole process of my MSc thesis. Without their continuous encouragement and invaluable support, this work could not have been done smoothly.

I also would like to thank Prof. Andy Nelson, the chair of my thesis assessment board, for his constructive suggestion and feedback during the inception and mid-term evaluation phase.

I am grateful to all the teachers, mentors, and tutors for sharing their knowledge in the lectures and providing guidance for our group projects during my MSc study. Thank you for supporting and helping me along the way.

My gratitude extends to my fellow mates in Master Spatial Engineering. I appreciate your presence and collaborations during the case studies and international module. The time spent with you was the most meaningful and cheerful period during my MSc study life. I will cherish our friendship forever.

Moreover, I also appreciate the ITC Foundation Scholarship committee for providing me this excellent opportunity to peruse my MSc degree.

Last but not least, I want to thank my family that formed a stable basis during all these years. Thank you for your unconditional support and encouragement. There are not enough words to express my emotion, I miss you so much. I would also thank my beloved girlfriend, Keke, who accompanied me, encouraged me, supported me all the time.

*Rui Xie
Enschede,
The Netherlands, June 2020*

TABLE OF CONTENTS

1.	INTRODUCTION.....	1
1.1.	Background.....	1
1.2.	Research objective and research questions	3
2.	MATERIALS AND METHODS	5
2.1.	Study area.....	6
2.2.	Data.....	7
2.3.	Image processing	7
2.4.	Narrowband vegetation indices	8
2.5.	Partial least square regression.....	9
2.6.	Machine learning algorithms	9
2.7.	Model validation and mapping.....	11
3.	RESULTS	12
3.1.	LAI measurements and extracted plot reflectance	12
3.2.	Narrowband vegetation indices	12
3.3.	Partial least square regression.....	14
3.4.	Artificial neural network.....	15
3.5.	Gaussian processes regression	16
3.6.	Use of spectral subset in predicting LAI.....	17
3.7.	Mapping forest leaf area index	17
4.	DISCUSSION.....	19
5.	CONCLUSIONS AND RECOMMENDATIONS.....	23
	APPENDIX.....	30

LIST OF FIGURES

Figure 1. General analytical framework developed for this study.....	5
Figure 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).....	6
Figure 3. Examples of excluded plots. (a) fall closely to the border with different vegetation cover; (b) located in a very heterogeneous area; (c) covered by cloud (shadow).....	8
Figure 4. Schematic diagram of ANN architecture. White, blue, and yellow circles represent input data, neurons, output data, respectively, arrows represent links (weights) between them.....	10
Figure 5. Mean canopy reflectance for broadleaf, conifer, and mixed forest stands measured using the Specim AISA Fenix hyperspectral sensor.....	12
Figure 6. 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.	13
Figure 7. Measured LAI and estimated LAI calculated from entire reflectance data from Fenix airborne sensor using the SAVI2.	14
Figure 8. Regression coefficient (B) of each wavelength for the partial least square regression model. Dashed lines represent the standard deviation of B	14
Figure 9. Measured LAI and estimated LAI calculated from entire reflectance data from Fenix airborne sensor using partial least square regression (No. of factors = 6)..	15
Figure 10. Measured LAI and estimated LAI calculated from entire reflectance data from Fenix airborne sensor using artificial neural network (No. of neurons = 4).....	15
Figure 11. Measured LAI and estimated LAI using Gaussian processes regression 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.	16
Figure 12. Uncertainties (σ) of LAI for sample plots ($n = 30$) estimated using Gaussian processes regression from the entire spectral reflectance.....	16
Figure 13. (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.	18
Figure 14. Measured LAI and estimated LAI calculated from entire reflectance data from Fenix airborne sensor using the narrowband VIs (i.e., NDVI, RVI, TSAVI)	30
Figure 15. Boxplot of <i>in situ</i> LAI and predicted LAI values using four methods.....	30

LIST OF TABLES

Table 1. Summary statistics of the measured LAI of sample plots in BFNP ($n = 40$).	7
Table 2. Vegetation indices used in this study and their broadband forms in the literature	9
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 ($R^2 > 0.5$).	13
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..	13
Table 5. Cross-validated results ($RMSE_{cv}$ and R_{cv}^2) for estimating LAI using entire reflectance data versus the spectral subset obtained from the analysis of VIs.	17
Table 6. Paired t -test results between measured LAI and evaluated methods in this study (i.e., SAVI2, PLSR, ANN, and GPR).	30

1. INTRODUCTION

1.1. Background

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, such as habitat for plant and animal species, biomass and carbon storage, timber products, as well as climate adaptations (Canadell et al., 2000; FAO, 2010; Hill et al., 2019). However, forest ecosystems face great challenges as a result of global warming and expanding human activities (Klos et al., 2009; Birdsey and Pan, 2011). The important processes which affect forest ecosystems include, deforestation caused by the conversion of forests to agricultural fields and logging activities, drought stress, biotic stress, and wildfire (Ciais et al., 2005; Anderegg et al., 2013). Moreover, due to the impact of heat waves and dry spells, forest ecosystems are also becoming more susceptible to insects infestation. These combined effects posed a severe risk of reduction of forest ecosystem services (Hill et al., 2019). In order to keep track of these effects and effectively manage forest ecosystems, monitoring the forest dynamics is of exceptional importance.

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 models (Fischer et al., 1997), and recently has been proposed as one of the essential biodiversity variables (EBVs) to remotely track biodiversity change in ecosystems from space (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).

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 affect 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 shown 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 (Wang et al., 2019). Lacking information about variable uncertainties can lead to errors in subsequent analysis when such vegetation variable maps are used in ecological applications. Pixels with high prediction uncertainties can be improved by collecting additional calibration data from these under-represented regions. Uncertainties mapping is thus crucial for improving model performance and mapping quality.

While GPR has been recently used for the estimation of plant canopy traits by a few studies, its applications have been mainly limited in the agricultural fields (Caicedo et al., 2014; 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 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.

1.2. Research objective and research questions

The aim of this study is to accurately estimate LAI and map its spatial distribution in the Bavarian Forest National Park (BFNP) using full-spectrum Fenix airborne hyperspectral data. To achieve this aim, the objectives and associated research questions (RQ) are as follows:

(1) To determine the best narrowband vegetation indices (VIs) (in terms of RMSE and R^2) for estimating LAI using airborne hyperspectral data in BFNP;

RQ 1. Which spectral narrowband combinations can provide the most accurate LAI estimation?

RQ 2. Which spectral regions are more important for LAI estimation using narrowband VIs?

(2) To investigate the utility of Gaussian processes regression (GPR) as a representative of kernel-based machine learning algorithms for estimating LAI using airborne hyperspectral data in BFNP;

RQ 3. What are the main factors causing high LAI retrieval uncertainties?

(3) To compare the performance of GPR with other commonly used empirical methods namely narrowband VIs, artificial neural network (ANN) and partial least square regression (PLSR);

RQ 4. What are the most informative bands when estimating LAI using the PLSR model?

RQ 5. What is the impact of utilizing a spectral subset (obtained from the analysis of VIs) as an alternative to the entire spectra on the predictive performance of studied models?

(4) To map the spatial distribution of LAI in BFNP using the best-performing approach (in terms of prediction accuracy);

RQ 6. How does the distribution of LAI vary within the BFNP?

2. MATERIALS AND METHODS

Figure 1 presents the general workflow adopted in this study. First of all, field measurements were collected in the study area as the validation dataset. Fenix airborne hyperspectral imagery was acquired paralleled with the field data collection and preprocessed. Once the reflectance of the field plots were properly extracted, Gaussian processes regression together with other three widely used empirical approaches (i.e. narrowband vegetation indices, partial least square regression, and artificial neural network) were then calibrated and validated by applying the leave-one-out cross-validation using the LAI field measurements and corresponding extracted reflectance. The suitability for each retrieval method was then analysed. Moreover, to examine the impact of reducing data dimensionality, a spectral subset obtained from the analysis of VIs was used as input for prediction of the other models. Finally, a forest LAI map was generated using the most accurate method, together with its accompanying prediction uncertainties produced by GPR.

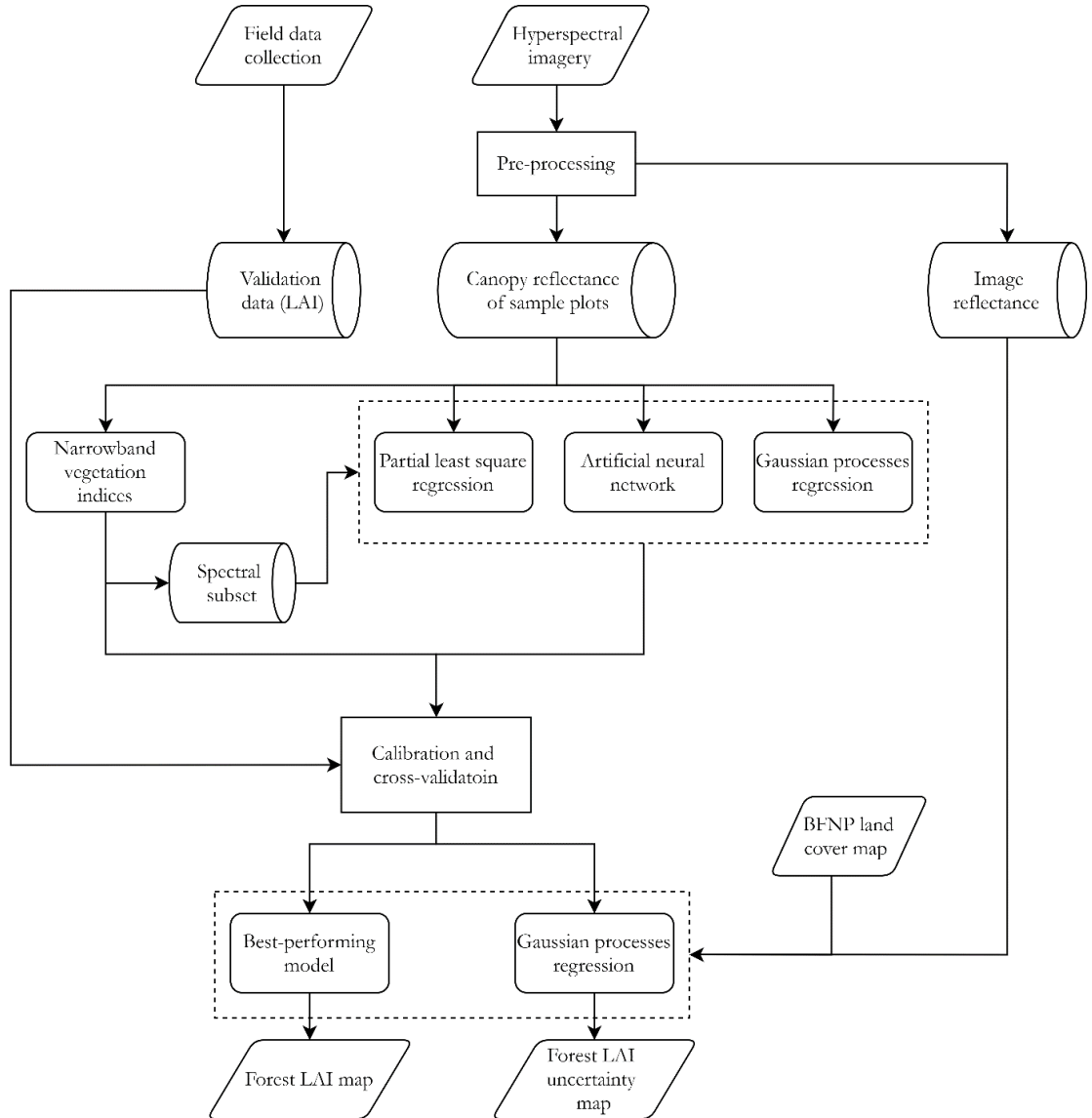


Figure 1. General analytical framework developed for this study.

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 (Figure 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 1453 m above sea level (a.s.l) and is composed of terrains varying from the low valley, hillsides, to highlands. This forest national park has a temperate climate with the mean annual temperature between 6.5 °C in the valleys and 2 °C at highlands. The average annual precipitation ranges from 1200 mm to 1800 mm. The predominant soil type for lower areas (below 900 m a.s.l) is brown soils, while for higher altitude areas, the main soil types are brown soils and brown podzolic soils (Heurich et al., 2010).

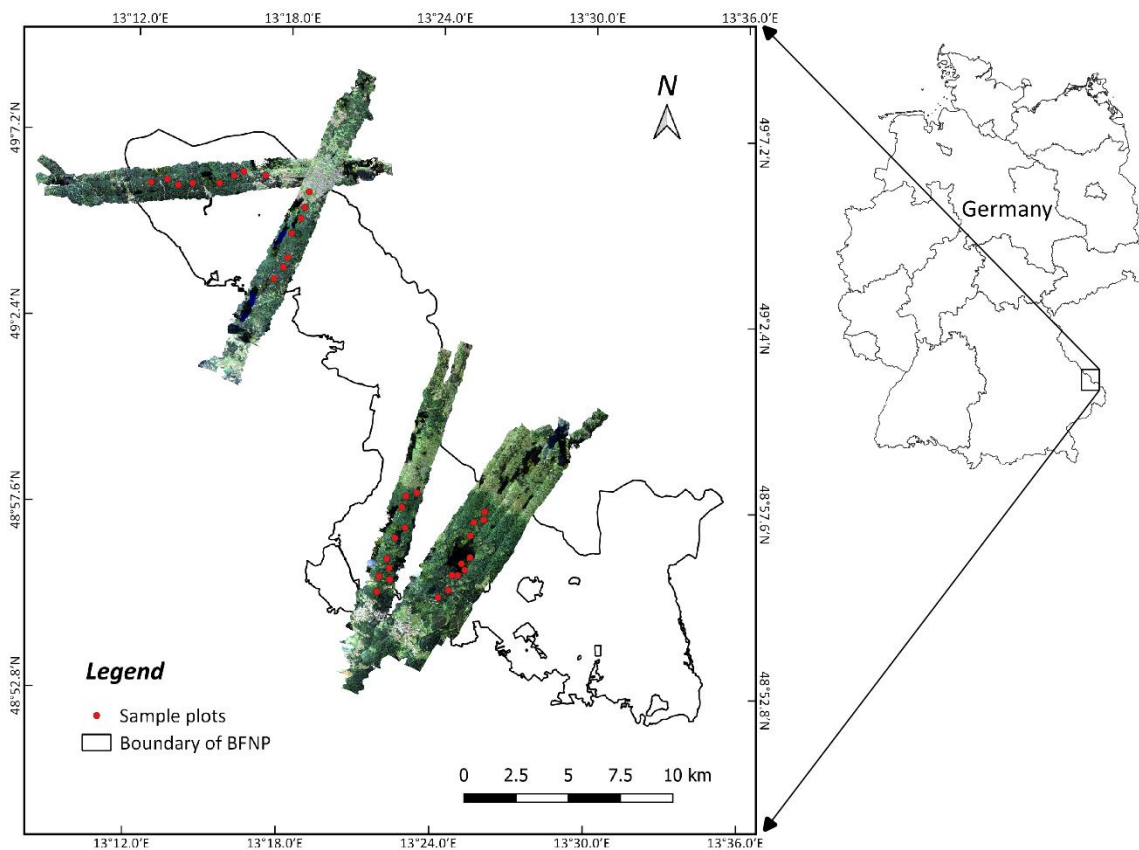


Figure 2. The location of Bavarian Forest National Park (BFNP) and the mosaic of Fenix hyperspectral data of four transects acquired on 6 July 2017 using a true colour composite (bands 469, 549, 640 nm). Red dots represent the locations of sample plots.

Tree species are mainly distributed as a function of altitude. In the peak regions, the majority of trees are Norway spruce (*Picea abies*) and with a few existences of sub-alpine spruce forests and Mountain ash (*Sorbus aucuparia*). Mountain slope areas, characterized by mixed forest, 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). Since the mid-1990s, the forested area in BFNP has been affected by the proliferation of the spruce bark beetles. By 2012, the bark beetle attack had resulted in a massive die-off of about 6000 ha of spruce stands (Lausch et al., 2013).

BFNP is also an important ecosystem that provides habitats for a broad range of animals, plants, and fungi. Moreover, BFNP plays an essential role in the protection of large wildlife species, including lynx (*Lynx lynx*), capercaillie (*Tetrao urogallus*), and also large ungulates such as red deer (*Cervus elaphus*) and roe deer (*Capreolus capreolus*) (Cailleret et al., 2014). Many efforts, such as wildlife population control, reducing disturbance by restricting public access to certain areas, or reduction of winter feeding, have been made to preserve the natural diversity from human interference and activities and to protect privately-owned forests that border the BFNP (Heurich et al., 2011).

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 (Figure 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 measurement

Field measurements of LAI were collected simultaneously with the airborne data acquisition (Figure 2) (Gara et al., 2019). The stratified random sampling strategy was performed within the major cover types in order to select 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 (30m $\times$ 30m). 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 and other forest structural parameters were measured. For each plot, three above-canopy observations were taken using Li-Cor LAI-2200 canopy analyser (Li-Cor, 1992) 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. The summary statistics of LAI field measurements are presented in Table 1.

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

2.3. Image processing

The image strips were preprocessed by the NERC Airborne Research Facility (NERC-ARF). Image data were converted from digital number (DN) to at-sensor radiance. 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 Schl pfer, 2019). Rough terrain model, ASTER Digital Elevation Model (DEM), and rural 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. 27m by 27m) 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, for each sample plot, the average spectra from different adjacent image strips were calculated. Then the spectrum which had a large deviation (greater than one standard deviation) from other spectrums was disregarded. Finally, the mean of rest spectrums was calculated to represent the canopy reflectance of the sample plot.

Except for field plots that are outside the image transects, six plots were excluded from further analysis. Figure 3 presents some examples of the three types of excluded plots. Three plots that either are located on the border between different vegetation cover (Figure 3 (a)) or are within a very heterogeneous region were strongly affected by the geo-referencing error (Figure 3 (b)). Moreover, another three plots covered by cloud (shadow) were also disregarded by this study (Figure 3 (c)). Taking into account another four plots that were located outside the image boundary, the remaining 30 plots were used for further analysis. Data processing and analysis were performed using MATLAB R2019a (The MathWorks, Inc.)

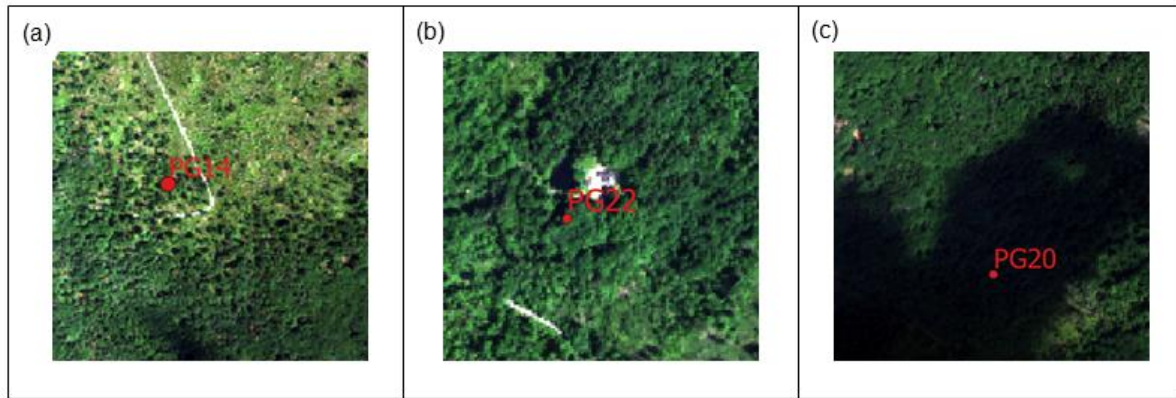


Figure 3. Examples of excluded plots. (a) fall closely to the border with different vegetation cover; (b) located in a very heterogeneous area; (c) covered by cloud (shadow).

2.4. Narrowband vegetation indices

Four widely used vegetation indices (VIs) were utilized as representatives of ratio-based and soil-based vegetation indices 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 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.

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)	$TSAVI = \frac{a(\rho_{NIR} - a\rho_{red} - b)}{a\rho_{NIR} + \rho_{red} - ab}$	$TSAVI_{narrow} = \frac{a(\rho_{\lambda_1} - a\rho_{\lambda_2} - b)}{a\rho_{\lambda_1} + \rho_{\lambda_2} - ab}$

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 most sensitive regions can be identified based on R^2 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 input for all other methods used in this study for LAI estimation.

2.5. Partial least square regression

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 approach 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, 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 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^2_{CV} and reduces the $RMSE_{CV}$ by $> 2\%$ (Kooistra et al., 2004; Cho et al., 2007). Moreover, the importance of each band was evaluated based on the standardized regression coefficient (B coefficient) (Haaland and Thomas, 1988). PLSR analysis was carried out in TOMCAT toolbox 1.01 within MATLAB (Daszykowski et al., 2007).

2.6. Machine learning algorithms

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 artificial neural network (ANN) as a representative of the most common machine learning approach, and Gaussian processes regression (GPR) which is being recently introduced as kernel-based machine learning algorithms were evaluated respectively.

2.6.1. Artificial neural network

ANN is a commonly used approach to develop nonlinear non-parametric models for the estimation of vegetation parameters (Kimes et al., 1998; Rivera et al., 2014; Neinavaz et al., 2016). In general, ANN is an interconnected structure of neurons organized in one or more hidden layers (Figure 4). In this study, a standard multi-layer perceptron with one hidden layer (tan-sigmoid 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.

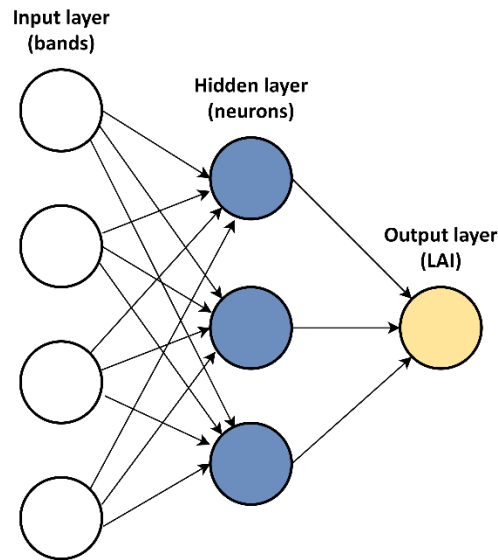


Figure 4. Schematic diagram of ANN architecture. White, blue, and yellow circles represent input data, neurons, output data, respectively, arrows represent links (weights) between them.

It is known that increasing number of neurons 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 early 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 initialized randomly 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 proven to be useful for extracting vegetation parameters by previous studies (Verrelst et al., 2013; Wang et al., 2019):

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

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 (also called 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

Leave-one-out cross-validation (LOOCV) was utilized to validate the studied methods. 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. Cross-validated root mean square error ($RMSE_{CV}$) and cross-validated coefficient of determination (R_{CV}^2) 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 LAI and estimated 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 RGB airborne imagery.

3. RESULTS

3.1. LAI measurements and extracted plot reflectance

The values of LAI field measurements cover a range from 1.33 to 5.42, presenting a relatively wide variety of LAI values involved in this study. Analysis of field data per species type showed that the mean LAI value for broadleaf was 4.10, while the conifer had a mean LAI value of 3.46. The mean reflectance spectra for the three dominant tree species extracted from the airborne hyperspectral data are presented in Figure 5, which also confirms the mean LAI values of the species. As can be seen in this figure, broadleaved stands exhibit the highest mean spectra, while the mean reflectance for conifer stands is relatively low.

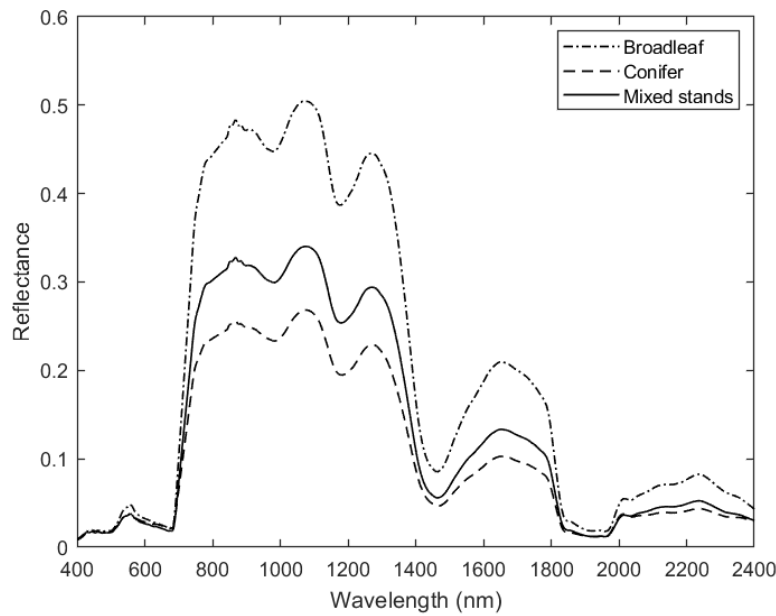


Figure 5. Mean canopy reflectance for broadleaf, conifer, and mixed forest stands measured using the Specim AISA Fenix hyperspectral sensor.

3.2. Narrowband vegetation indices

The selected four narrowband vegetation indices (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. Figure 6 presents the results for the two best-performing indices (i.e., SAVI2 and RVI) of each narrowband VI type 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^2_{CV} and $RMSE_{CV}$ were used to evaluate model accuracy. As presented in Table 4, four studied narrowband VIs overall produced reasonable prediction for LAI, while soil-based narrowband VIs performs slightly better comparing to ratio-based ones. The relationship between the measured LAI and the best-performing indices (SAVI2) is presented in Figure 7. The results for the other three narrowband VIs are shown in Figure 14, Appendix A.

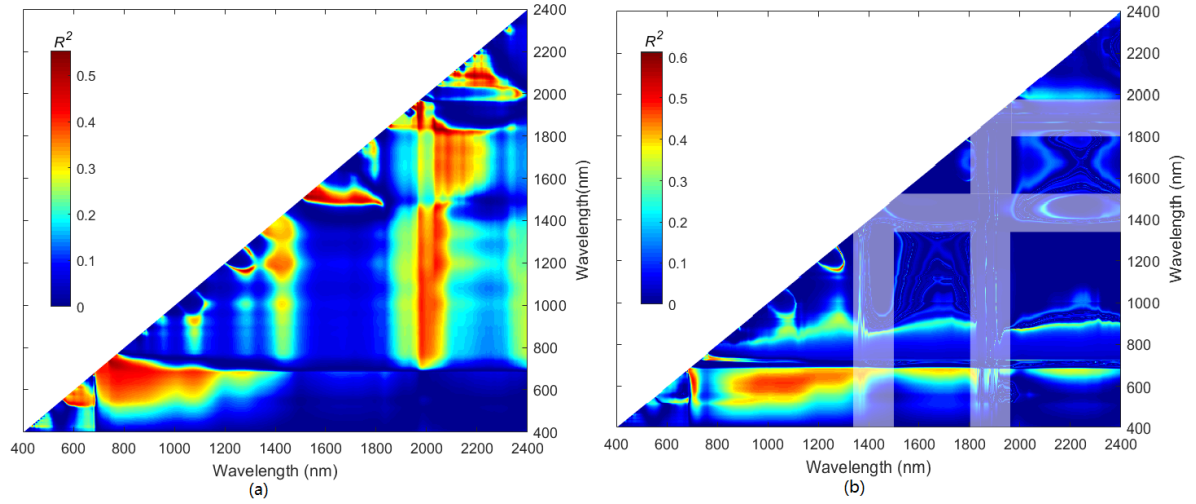


Figure 6. 2-D correlation plot presenting the coefficient of determination (R^2) between measured LAI and (a) narrowband RVI and (b) narrowband SAVI2 calculated using Fenix airborne hyperspectral data. The highlighted regions in (b) refer to the strongly noisy bands existed in the measured understory reflectance.

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)	
			$\lambda_1(nm)$	$\lambda_2(nm)$
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
Soil-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
	SAVI2	0.62	682-688	683-690
			742-771	729-731
			1167-1185	1290-1302
			1238-1274	1275-1290
			1246-1266	1272-1290
	TSAVI	0.61	682-689	684-690
			740-765	729-730
			1162-1180	1282-1296
			1246-1266	1272-1290

Table 4. R^2_{CV} 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^2_{CV}	$RMSE_{CV}$	The best bands combination	
			$\lambda_1(nm)$	$\lambda_2(nm)$
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

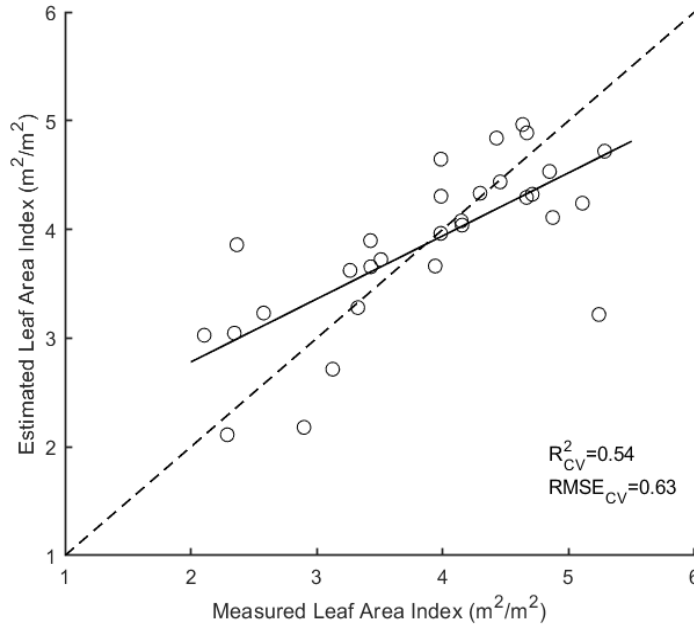


Figure 7. Measured LAI and estimated LAI calculated from entire reflectance data from Fenix airborne sensor using the SAVI2. The dashed line shows the 1:1 relationship, while the solid line indicates the relationship between the field measured and estimated values of LAI.

3.3. Partial least square regression

The regression coefficients (B coefficient) in Figure 8 represents the relative contribution of each waveband to the LAI prediction. Important wavebands can be identified where the corresponding B coefficient is greater than the standard deviation. When we considered the criterion that added component increases R^2_{cv} 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 Figure 9. As it can be observed, PLSR yields higher R^2_{cv} value compared to narrowband vegetation indices (Table 4) but failed to reduce the $RMSE_{cv}$.

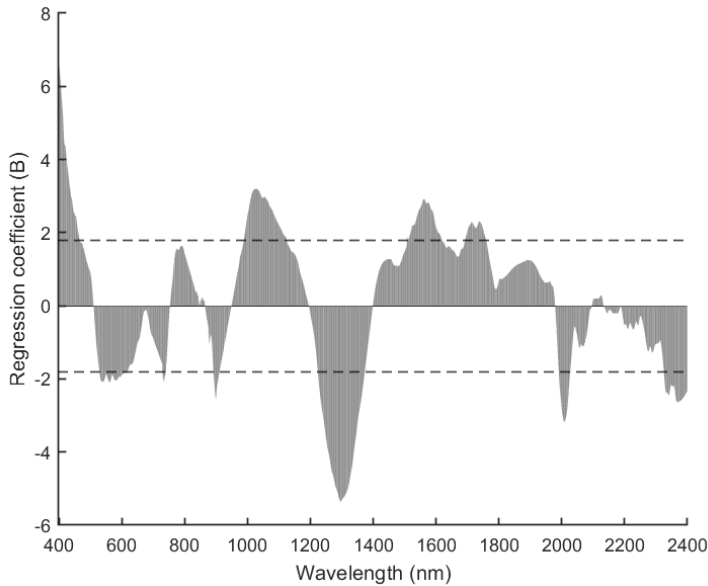


Figure 8. Regression coefficient (B) of each wavelength for the partial least square regression model. Dashed lines represent the standard deviation of B . The regression coefficients larger than its standard deviation (absolute value) indicate a larger influence of the spectral data on the regression model.

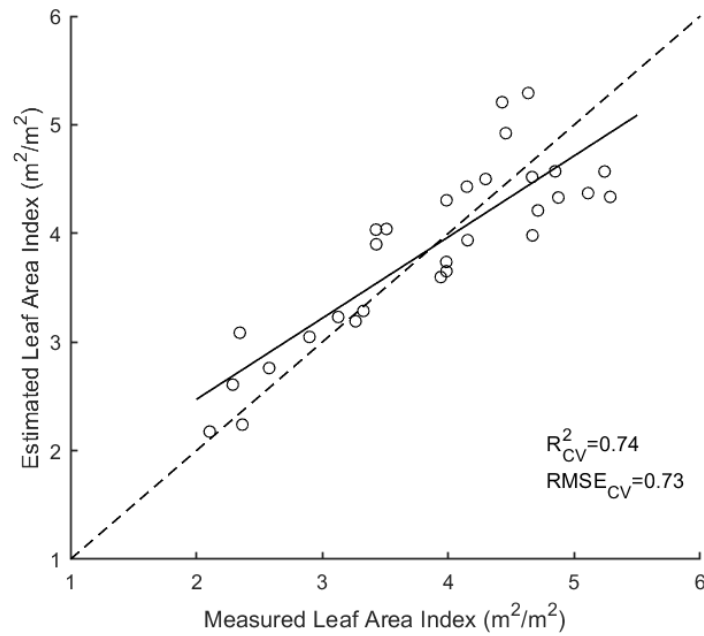


Figure 9. Measured LAI and estimated LAI calculated from entire reflectance data from Fenix airborne sensor using partial least square regression (No. of factors = 6). The dashed line shows the 1:1 relationship, while the solid line indicates the relationship between the field measured and estimated values of LAI.

3.4. 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 determined 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 Figure 10. In comparison with narrowband VIs and PLSR model, ANN significantly improved $RMSE_{CV}$ and obtained a relatively high correlation with LAI (R^2_{CV}).

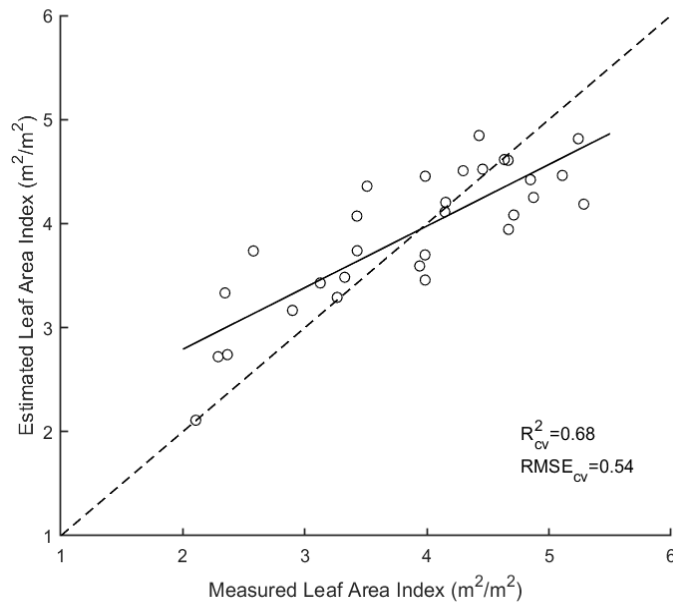


Figure 10. Measured LAI and estimated LAI calculated from entire reflectance data from Fenix airborne sensor using artificial neural network (No. of neurons = 4). R^2_{CV} and $RMSE_{CV}$ are averaged results of 1,000 random initializations. The dashed line shows the 1:1 relationship, while the solid line indicates the relationship between the field measured and estimated values of LAI.

3.5. Gaussian processes regression

The performance of GPR was firstly evaluated from the original hyperspectral reflectance data. The cross-validated results are shown in Figure 11. As can be seen from Figure 11, both the $RMSE_{cv}$ and R^2_{cv} 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 Figure 12. 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.

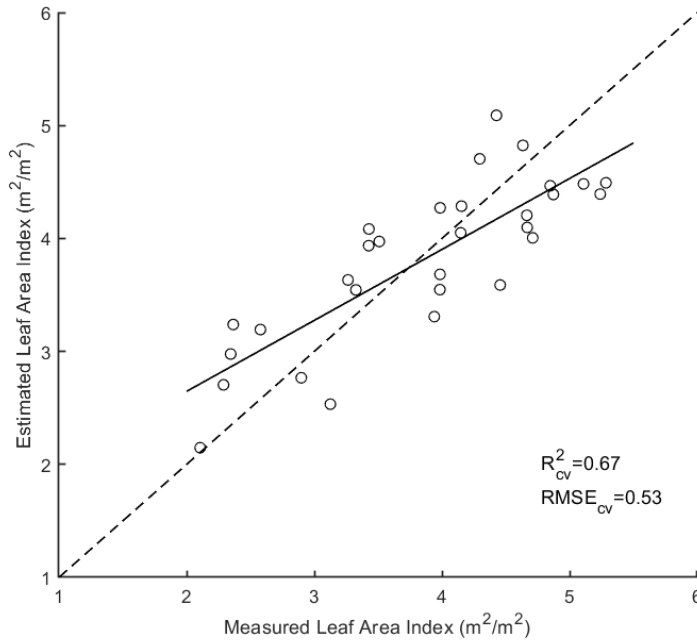


Figure 11. Measured LAI and estimated LAI using Gaussian processes regression calculated from the entire reflectance data from the Fenix airborne sensor. The dashed line shows the 1:1 relationship, while the solid line indicates the relationship between the field measured and estimated values of LAI.

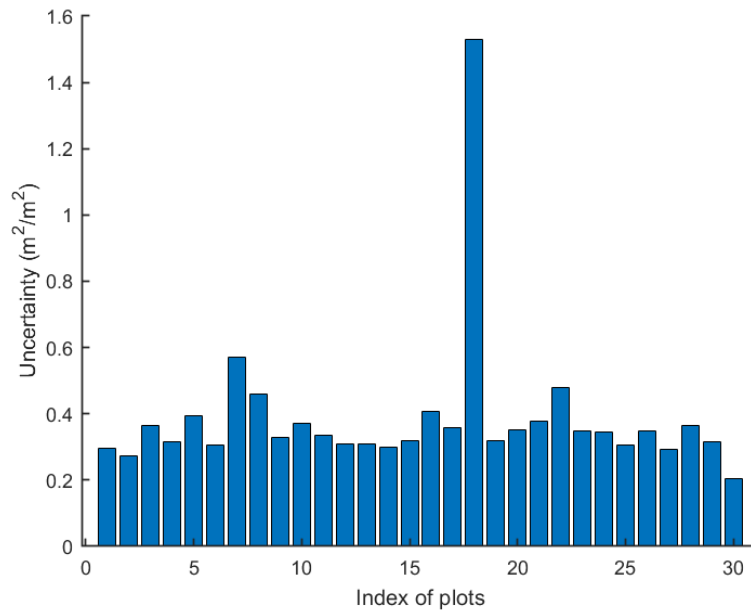


Figure 12. Uncertainties (σ) of LAI for sample plots (n = 30) estimated using Gaussian processes regression from the entire spectral reflectance.

3.6. 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 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.

Table 5. Cross-validated results ($RMSE_{CV}$ and R_{CV}^2) 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_{CV}^2 and lowest $RMSE_{CV}$ is boldfaced.

Input	Regression methods	R_{CV}^2	$RMSE_{CV}$
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

3.7. Mapping forest leaf area index

GPR with the identified spectral subset which produced the highest accuracy (Table 5) was applied to the Fenix image strip with the lowest cloud coverage (i.e., Flight line 29) to generate the LAI map of the BFNP. The generated LAI map and its associated uncertainty map are presented in Figure 13 (c) and (d). The non-forest area was masked out in advance using the BFNP land use map. To check the consistency and analyse the output map, the true colour (RGB) Fenix image and the forest classification map are also shown in Figure 13 (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 reported in Table 1 ($LAI_{mean} = 3.81$). In the obtained LAI uncertainty map, pixels with lower σ values indicate more confident estimations retrieved by the trained GPR model. Generally, the uncertainty levels mapped by GPR were low across the entire image (Figure 13 (d)).

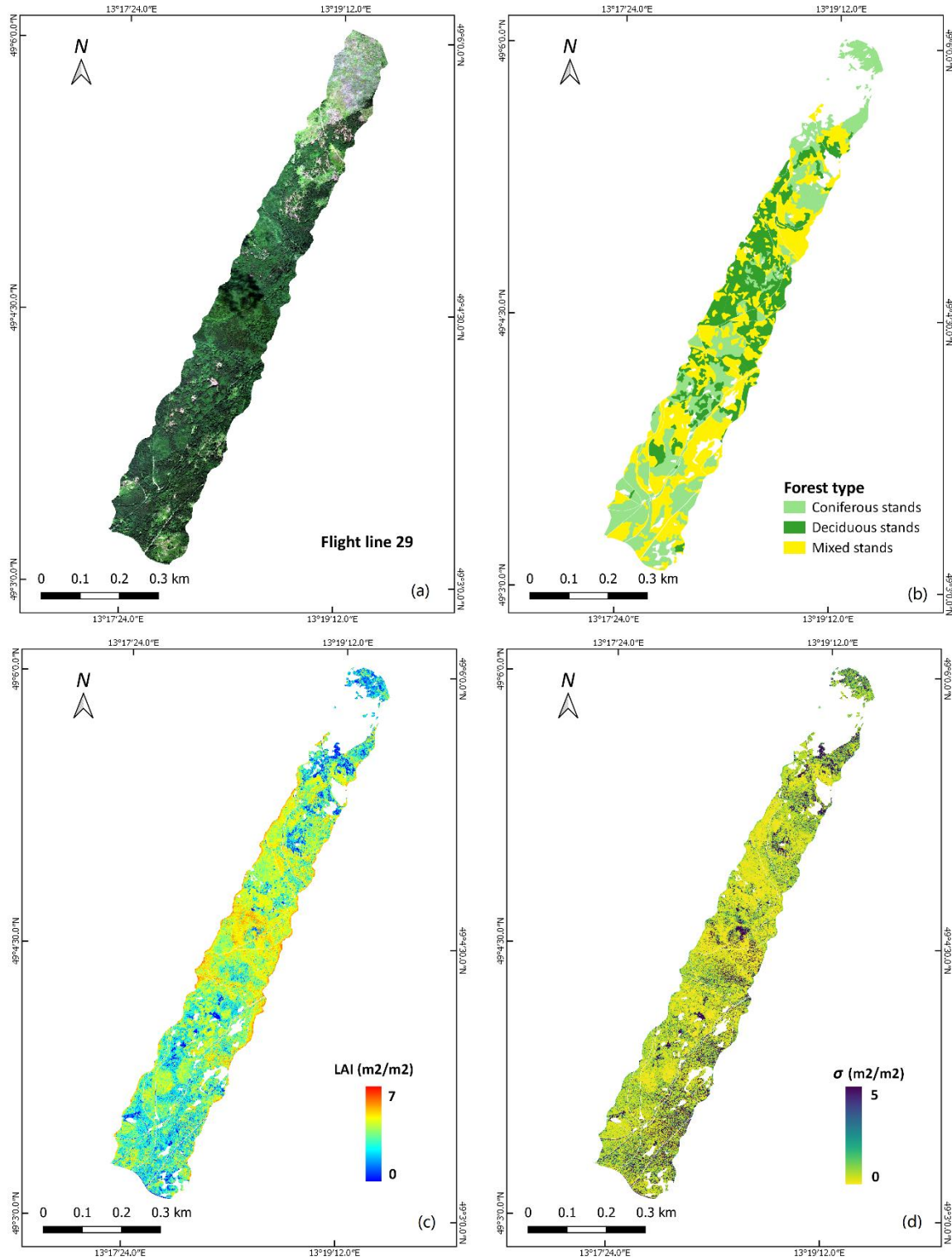


Figure 13. (a) Fenix airborne hyperspectral data (Flight line 29), (b) forest classification map, (c) modelled LAI in BFNP using GPR and the selected spectral subset, and (d) associated uncertainty estimates.

4. DISCUSSION

This study examined the performance of Gaussian processes regression (GPR), a novel kernel-based machine learning algorithm in comparison to the most commonly used empirical methods (i.e., narrowband vegetation indices, partial least square regression, and artificial neural network) for estimating forest LAI using Fenix airborne hyperspectral data. The results showed that GPR outperformed the other methods and yielded the most precise prediction for forest LAI ($\text{RMSE}_{\text{CV}} = 0.53 \text{ m}^2 \text{ m}^{-2}$). 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). The results are consistent with those of Verrelst et al., (2012a) and Rivera et al. (2014) 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 (Figure 15, Appendix B) and paired *t*-test (Table 6, Appendix C) showed that, in general, there was an agreement between measured and estimated LAI using different approaches ($p > 0.05$). Although this suggests a comparable predictive accuracy for different empirical approaches in retrieving LAI, GPR may be a preferred method for its more precise estimation of LAI. 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 narrowband VIs performed slightly better (RMSE 0.63-0.64 $\text{m}^2 \text{ m}^{-2}$) comparing to ratio-based ones (RMSE 0.69-0.71 $\text{m}^2 \text{ m}^{-2}$). 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. This finding is also in line with previous studies, which demonstrated the importance of soil-based vegetation indices, 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 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 vegetation indices 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 entire spectra) with six latent factors resulted in a significantly increased correlation coefficient (R_{CV}^2) 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., 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 Figure 8, the informative spectral bands in the PLSR model are mainly located within the SWIR region and correspond well to those identified as sensitive wavebands 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 Figure 8 are mainly attributed to the poor signal-to-noise ratio at these regions which is probably caused by the silicon photodiode detector of the sensor (Milton et al., 2009). The obtained results broadly support the work of other studies (Brown et al., 2000; Cohen and Goward, 2004; Darvishzadeh et al., 2011; Rivera et al., 2014; 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. 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 make the developed model less robust when inverted and used for prediction (Rivera et al., 2014; 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 to 4) were somehow overestimated (Figure 10). 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 result 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 of agricultural fields. 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. (2014) 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 Figure 12, 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 the

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 coniferous and mixed stands), which is in agreement with the observation during the field campaign. The same distribution pattern for canopy chlorophyll content was found in BFNP (Ali et al., 2020). The artefacts on the edge of the map (Figure 13 (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 (Figure 13 (d)). Regions with high prediction uncertainties infer poorly predicted results and can be either masked or improved by collecting additional training data from these areas. As can be interpreted from Figure 13 ((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 the training samples, this was somehow expected. These results also match the finding of a previous study 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 often 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.

It has long been a consensus for ecologists and forest managers to identify suitable methods for accurately mapping the spatial distribution of forest biophysical variables from remote sensing (RS) data. Maps of plant traits play an important role in monitoring forest ecosystems and are mandatory inputs for spatially ecological modelling such as biodiversity assessment, examining productivity, and ecosystem dynamics. Forest monitoring can further support innovations in forest management, ecology conservation, and resource decision-making in the context of increasing pressure due to climate change and expanding human activities (Asner et al., 2016). The continuous and explicit information of vegetation traits can facilitate assessing the long-term impact of increasing drought events (Hill et al., 2019). By comparing and analysing the changes in vegetation traits, it can also contribute to post-evaluate the effect of forest damages such as wildfire. Moreover, environmental decision-making usually needs to consider resource management for both the conservation of wildlife and privately-owned areas, especially in edges of ecological zones where stakeholders may have conflicting interests. Therefore, regional vegetation mapping products can be an effective tool for monitoring the expanding human activities and wildlife damages, which can further provide decision support to strategic spatial planning.

Previous studies were carried out to develop numerous models and approaches for improving the estimation of vegetation parameters. As demonstrated by this study, GPR seems to be a promising candidate for the retrieval of vegetation parameters compared to conventional empirical methods due to its excellent predictive performance. However, generated trait maps usually can not be directly used for

ecological applications because the absence of uncertainties can lead to errors in subsequent analysis. This is especially important when models are being applied to new sites without ground truth data. Therefore, the generated variable maps are strongly required to check with its uncertainties to assess the utility of the resulting maps. The trait uncertainty maps like we produced in this study provide an opportunity to fill this knowledge gap and give additional insights for managers and researchers to evaluate and improve the model and mapping results.

5. CONCLUSIONS AND RECOMMENDATIONS

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. 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., 2014). Comparing to narrowband vegetation indices (VIs), partial least square regression (PLSR), and artificial neural network (ANN), the most accurate LAI prediction was obtained using GPR in this study ($\text{RMSE}_{\text{CV}} = 0.53 \text{ m}^2 \text{ m}^{-2}$). 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 forested areas which were under-represented in the collected sample plots and low LAI region (i.e. low vegetation cover). Therefore, comprehensive data sampling in regions associated with high uncertainties is recommended in future fieldwork to improve the 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 future retrieval applications on a global scale. Therefore, further studies can investigate the utility of GPR for vegetation parameter retrieval using simulated spaceborne hyperspectral data, such as EnMAP end-to-end simulation tool (Segl et al., 2012). These efforts are critical in preparation for the forthcoming satellite missions and the imminent availability of full-range spectral data from space and contribute to monitor plant function and diversity globally. Moreover, the GPR model investigated in this study could be potentially applied in other vegetation types to generate trait maps and uncertainty estimates. Once the model is validated and uncertainties are evaluated, the generated trait maps can be used for ecological applications.

LIST OF REFERENCES

- Abdi, H. (2003). Partial least square regression (PLS regression). *Encyclopedia for Research Methods for the Social Sciences*, 6(4), 792–795.
- 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. *Agricultural and Forest Meteorology*, 236, 162–174.
- 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 Journal of Photogrammetry and Remote Sensing*, 122, 68–80. <https://doi.org/10.1016/J.ISPRSJPRS.2016.09.015>
- Anderegg, W. R. L., Kane, J. M., & Anderegg, L. D. L. (2013). Consequences of widespread tree mortality triggered by drought and temperature stress. *Nature Climate Change*, 3(1), 30–36. <https://doi.org/10.1038/nclimate1635>
- Asner, G. P., Brodrick, P. G., Anderson, C. B., Vaughn, N., Knapp, D. E., & Martin, R. E. (2016). Progressive forest canopy water loss during the 2012–2015 California drought. *Proceedings of the National Academy of Sciences*, 113(2), E249–E255.
- 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., 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. *International Journal of Applied Earth Observation and Geoinformation*. <https://doi.org/10.1016/j.jag.2015.01.009>
- Atzberger, C., Darvishzadeh, R., Schlerf, M., & Le Maire, G. (2013). Suitability and adaptation of PROSAIL radiative transfer model for hyperspectral grassland studies. *Remote Sensing Letters*, 4(1), 55–64.
- 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. *Computers and Electronics in Agriculture*, 73(2), 165–173. <https://doi.org/https://doi.org/10.1016/j.compag.2010.05.006>
- 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 Sensing of Environment*, 105(4), 313–325. <https://doi.org/10.1016/J.RSE.2006.07.014>
- Baret, F., & Guyot, G. (1991). Potentials and limits of vegetation indices for LAI and APAR assessment. *Remote Sensing of Environment*, 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, & Buis, S. (2008). Estimating Canopy Characteristics from Remote Sensing Observations: Review of Methods and Associated Problems. In S. Liang (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, 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. *Nature Climate Change*, 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. *Bulletin of the American Meteorological Society*, 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 Sensing of Environment*, 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 Sensing of Environment*, 71(1), 16–25. [https://doi.org/https://doi.org/10.1016/S0034-4257\(99\)00035-8](https://doi.org/https://doi.org/10.1016/S0034-4257(99)00035-8)
- Caicedo, J. P. R., Verrelst, J., Munoz-Mari, J., Moreno, J., & Camps-Valls, G. (2014). Toward a semiautomatic machine learning retrieval of biophysical parameters. *IEEE Journal of Selected Topics in*

- Applied Earth Observations and Remote Sensing*, 7(4), 1249–1259.
<https://doi.org/10.1109/JSTARS.2014.2298752>
- Cailleret, 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. *Forest Ecology and Management*, 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., ... others. (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 & Environment*, 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. *Journal of Geophysical Research: Atmospheres*, 102(D24), 29429–29443. <https://doi.org/10.1029/97jd01107>
- Chen, J. M., Rich, P. M., Gower, S. T., Norman, J. M., & Plummer, S. (2004). Leaf area index of boreal forests: Theory, techniques, and measurements. *Journal of Geophysical Research: Atmospheres*, 102(D24), 29429–29443. <https://doi.org/10.1029/97jd01107>
- 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. *International Journal of Applied Earth Observation and Geoinformation*, 9(4), 414–424. <https://doi.org/10.1016/J.JAG.2007.02.001>
- Cho, M. A., & Skidmore, A. K. (2006). A new technique for extracting the red edge position from hyperspectral data: The linear extrapolation method. *Remote Sensing of Environment*, 101(2), 181–193. <https://doi.org/10.1016/J.RSE.2005.12.011>
- Ciais, P., Reichstein, M., Viovy, N., Granier, A., Ogée, J., Allard, V., ... Valentini, R. (2005). Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature*, 437(7058), 529–533. <https://doi.org/10.1038/nature03972>
- Cohen, W. B., & Goward, S. N. (2004). *Landsat's Role in Ecological Applications of Remote Sensing*.
- Cohen, W. B., Maersperger, T. K., Gower, S. T., & Turner, D. P. (2003). An improved strategy for regression of biophysical variables and Landsat ETM+ data. *Remote Sensing of Environment*, 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., ... Wang, L. (2003). Retrieval of canopy biophysical variables from bidirectional reflectance: Using prior information to solve the ill-posed inverse problem. *Remote Sensing of Environment*, 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 Sensing of Environment*, 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., Schlerf, M., & Atzberger, C. (2008a). Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland. *Remote Sensing of Environment*, 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 Journal of Photogrammetry and Remote Sensing*, 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. *International Journal of Remote Sensing*, 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 Journal of Photogrammetry and Remote Sensing*, 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. *Chemometrics and Intelligent Laboratory Systems*, 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 Payne experiment. *International Journal of Remote Sensing*, 24(7), 1505–1520.
- Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., ... Lautenbach, S. (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., ... Verhoef, W. (2017). The FLuorescence EXplorer Mission Concept—ESA's Earth Explorer 8. *IEEE Transactions on Geoscience and Remote Sensing*, 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 Sensing of Environment*, 61(1), 34–45. [https://doi.org/https://doi.org/10.1016/S0034-4257\(96\)00238-6](https://doi.org/https://doi.org/10.1016/S0034-4257(96)00238-6)
- Féret, J.-B., Gitelson, A. A., Noble, S. D., & Jacquemoud, S. (2017). PROSPECT-D: Towards modeling leaf optical properties through a complete lifecycle. *Remote Sensing of Environment*, 193, 204–215. <https://doi.org/https://doi.org/10.1016/j.rse.2017.03.004>
- 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 Sensing Reviews*, 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 Sensing*, 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 Journal of Photogrammetry and Remote Sensing*, 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. *Analytica Chimica 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 Sensing*, 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. *Geophysical Research Letters*, 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 Transactions on Geoscience and Remote Sensing*, 41(6 PART I), 1355–1362. <https://doi.org/10.1109/TGRS.2003.812910>
- Gonzalez, R. S., 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. *International Journal of Remote Sensing*, 39(23), 8859–8884. <https://doi.org/10.1080/01431161.2018.1500071>
- 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 Sensing of Environment*, 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. *Analytical Chemistry*, 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 Sensing of Environment*, 90(3), 337–352. <https://doi.org/https://doi.org/10.1016/j.rse.2003.12.013>
- Halme, E., Pellikka, P., & Möttöus, M. (2019). Utility of hyperspectral compared to multispectral remote sensing data in estimating forest biomass and structure variables in Finnish boreal forest. *International Journal of Applied Earth Observation and Geoinformation*, 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 Sensing of Environment*, 86(4), 542–553. [https://doi.org/https://doi.org/10.1016/S0034-4257\(03\)00131-7](https://doi.org/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., Baierl, F., Günther, S., & Sinner, K. F. (2011). Management and conservation of large

- mammals in the Bavarian Forest National Park. *Silva Gabreta*, 17(1), 1–18.
- Heurich, M., Beudert, B., Rall, H., & Křivánek, 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 F. Müller, C. Baessler, H. Schubert, & S. Klotz (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. *Surveys in Geophysics*, 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 Sensing of Environment*, 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 Sensing of Environment*, 93(1), 18–29. <https://doi.org/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. *International Journal of Remote Sensing*, 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. *Ecological Applications*, 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 Sensing of Environment*, 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. *Environmental Pollution*, 127(2), 281–290. [https://doi.org/https://doi.org/10.1016/S0269-7491\(03\)00266-5](https://doi.org/https://doi.org/10.1016/S0269-7491(03)00266-5)
- Labate, D., Ceccherini, M., Cisbani, A., De Cosmo, V., Galeazzi, C., Giunti, L., ... Stagi, M. (2009). The PRISMA payload optomechanical design, a high performance instrument for a new hyperspectral mission. *Acta Astronautica*, 65(9–10), 1429–1436. <https://doi.org/10.1016/j.ACTAASTRO.2009.03.077>
- Lausch, A., Heurich, M., & Fahse, L. (2013). Spatio-temporal infestation patterns of *Ips typographus* (L.) in the Bavarian Forest National Park, Germany. *Ecological Indicators*, 31, 73–81.
- le Maire, G., François, C., Soudani, K., Berveiller, D., Pontailier, J.-Y., Bréda, N., ... Dufrêne, E. (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 Sensing of Environment*, 112(10), 3846–3864. <https://doi.org/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., Maersperger, T. K., & Gower, S. T. (2004). Hyperspectral versus multispectral data for estimating leaf area index in four different biomes. *Remote Sensing of Environment*, 91(3–4), 508–520. <https://doi.org/10.1016/j.rse.2004.04.010>
- Li-Cor, I. (1992). LAI-2000 plant canopy analyzer instruction manual. *LI-COR Inc., Lincoln, Nebraska, USA*.
- Li, X., Zhang, Y., Bao, Y., Luo, J., Jin, X., Xu, X., ... Yang, G. (2014). Exploring the best hyperspectral features for LAI estimation using partial least squares regression. *Remote Sensing*, 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 Sensing of Environment*, 165, 123–134.
- Ma, Z., Peng, C., Zhu, Q., Chen, H., Yu, G., Li, W., ... Zhang, W. (2012). Regional drought-induced reduction in the biomass carbon sink of Canada's boreal forests. *Proceedings of the National Academy of Sciences*, 109(7), 2423–2427.
- Major, D. J., Baret, F., & Guyot, G. (1990). A ratio vegetation index adjusted for soil brightness. *International Journal of Remote Sensing*, 11(5), 727–740. <https://doi.org/10.1080/01431169008955053>
- MATLAB. (2019). *R2019a*. Natick, Massachusetts: The MathWorks Inc.
- Milton, E. J., Schaepman, M. E., Anderson, K., Kneubühler, M., & Fox, N. (2009). Progress in field

- spectroscopy. *Remote Sensing of Environment*, 113, S92–S109.
<https://doi.org/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) Vol. 5*(67–72), 2.
- Mutanga, O., & Skidmore, A. K. (2004). Narrow band vegetation indices overcome the saturation problem in biomass estimation. *International Journal of Remote Sensing*, 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 Journal of Photogrammetry and Remote Sensing*, 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. In *Advanced neural computers* (pp. 11–19). Elsevier.
- Norman, J. M., & Campbell, G. S. (1989). Canopy structure. In *Plant physiological ecology* (pp. 301–325). Springer.
- Nowlan, S. J., & Hinton, G. E. (1992). Simplifying Neural Networks by Soft Weight-Sharing. *Neural Computation*, 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 Phytologist*, 189(2), 375–394. <https://doi.org/10.1111/j.1469-8137.2010.03536.x>
- Pasolli, L., Melgani, F., & Blanzieri, E. (2010). Gaussian Process Regression for Estimating Chlorophyll Concentration in Subsurface Waters From Remote Sensing Data. *IEEE Geoscience and Remote Sensing Letters*, 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., & Schlöpfer, D. (2019). Atmospheric/Topographic Correction for Airborne Imagery: ATCOR-4 User Guide. In *DLR report DLR-IB 565-02/2019*. Wessling, Germany.
- Rivera, J. P., Verrelst, J., Muñoz-Marí, J., Camps-Valls, G., & Moreno, J. (2017). Hyperspectral dimensionality reduction for biophysical variable statistical retrieval. *ISPRS Journal of Photogrammetry and Remote Sensing*, 132, 88–101. <https://doi.org/https://doi.org/10.1016/j.isprsjprs.2017.08.012>
- Rivera, J. P., Verrelst, J., Delegido, J., Veroustraete, F., & Moreno, J. (2014). On the Semi-Automatic Retrieval of Biophysical Parameters Based on Spectral Index Optimization. *Remote Sensing*, 6(6), 4927–4951. <https://doi.org/10.3390/rs6064927>
- Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W. and 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. *Analytical Chemistry*, 36(8), 1627–1639. <https://doi.org/10.1021/ac60214a047>
- Schlerf, M., Atzberger, C., & Hill, J. (2005). Remote sensing of forest biophysical variables using HyMap imaging spectrometer data. *Remote Sensing of Environment*, 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. *International Journal of Applied Earth Observation and Geoinformation*, 12(1), 17–26.
<https://doi.org/10.1016/J.JAG.2009.08.006>
- 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*.
- Segl, K., Guanter, L., Rogass, C., Kuester, T., Roessner, S., Kaufmann, H., ... Hofer, S. (2012). EeteS—The EnMAP end-to-end simulation tool. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(2), 522–530.
- 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. *Ecological Applications*, 25(8), 2180–2197.
- 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>
- Skidmore, A. K. (2002). Taxonomy of environmental models in the spatial sciences. In A. Skidmore (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.
- Stuffer, T., Kaufmann, C., Hofer, S., Förster, K. P., Schreier, G., Mueller, A., ... Haydn, R. (2007). The EnMAP hyperspectral imager—An advanced optical payload for future applications in Earth observation programmes. *Acta Astronautica*, 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 Sensing of Environment*, 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 Sensing of Environment*, 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 Transactions on Geoscience and Remote Sensing*, 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 Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 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 Journal of Photogrammetry and Remote Sensing*, 108, 273–290.
<https://doi.org/10.1016/j.isprsjprs.2015.05.005>
- Verrelst, J., Malenovsky, Z., Van der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J. P., Lewis, P., ... Moreno, J. (2019). Quantifying Vegetation Biophysical Variables from Imaging Spectroscopy Data: A Review on Retrieval Methods. *Surveys in Geophysics*, 40(3), 589–629.
<https://doi.org/10.1007/s10712-018-9478-y>
- 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. *International Journal of Applied Earth Observation and Geoinformation*, 52, 554–567. <https://doi.org/10.1016/j.jag.2016.07.016>
- Wang, Z., Skidmore, A. K., Wang, T., Darvishzadeh, R., Heiden, U., Heurich, M., ... Hearne, J. (2017). Canopy foliar nitrogen retrieved from airborne hyperspectral imagery by correcting for canopy structure effects. *International Journal of Applied Earth Observation and Geoinformation*, 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 Sensing of Environment*, 221, 405–416.
<https://doi.org/https://doi.org/10.1016/j.rse.2018.11.016>
- Wang, Z., Wang, T., Darvishzadeh, R., Skidmore, A. K., Jones, S., Suarez, L., ... Hearne, J. (2016). Vegetation Indices for Mapping Canopy Foliar Nitrogen in a Mixed Temperate Forest. *Remote Sensing*, 8(6). <https://doi.org/10.3390/rs8060491>
- Weiss, M., & Baret, F. (1999). Evaluation of Canopy Biophysical Variable Retrieval Performances from the Accumulation of Large Swath Satellite Data. *Remote Sensing of Environment*, 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. *Agricultural and Forest Meteorology*, 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., ... Yang, X. (2017). Retrieving soybean leaf area index from unmanned aerial vehicle hyperspectral remote sensing: Analysis of RF, ANN, and SVM regression models. *Remote Sensing*, 9(4). <https://doi.org/10.3390/rs9040309>

APPENDIX

A. The relationship between measured LAI and estimated LAI predicted using narrowband vegetation indices.

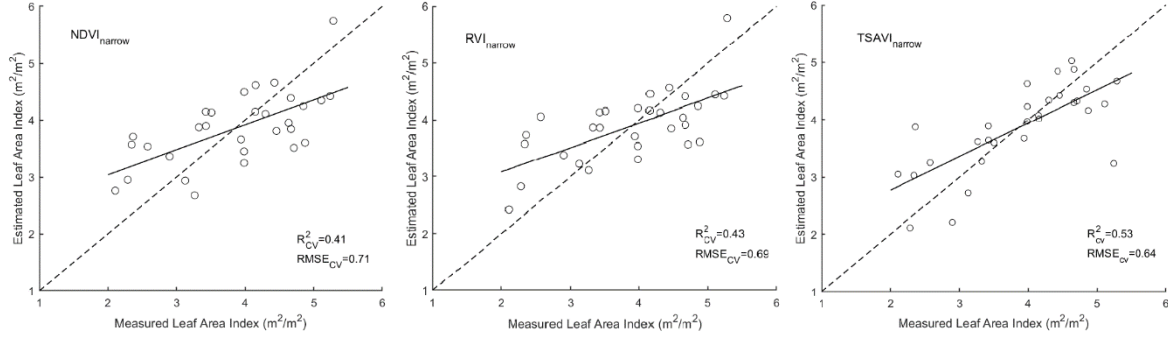


Figure 14. Measured LAI and estimated LAI calculated from entire reflectance data from Fenix airborne sensor using the narrowband VIs (i.e., NDVI, RVI, TSAVI). The dashed line shows the 1:1 relationship, while the solid line indicates the relationship between the field measured and estimated values of LAI.

B. Illustration of the central tendency of the four methods predictions compared to *in situ* LAI data.

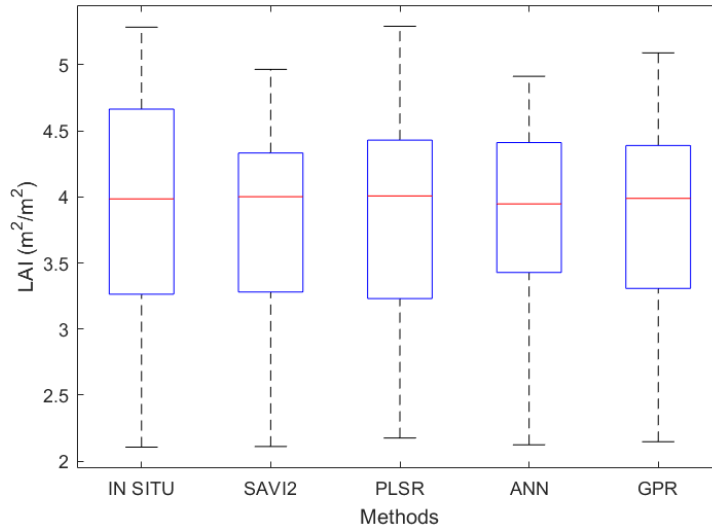


Figure 15. Boxplot of *in situ* LAI and predicted LAI values using four methods.

C. Statistically significant difference for predictions obtained using four different approaches against *in situ* LAI data.

Table 6. Paired *t*-test results between measured LAI and evaluated methods in this study (i.e., SAVI2, PLSR, ANN, and GPR).

Method	Std. dev	95% confidence interval	<i>t</i> -stat	p-value
SAVI2	0.65	(-0.24, 0.25)	0.03	0.97
PLSR	0.47	(-0.18, 0.18)	0.01	0.99
ANN	0.47	(-0.16, 0.20)	0.24	0.71
GPR	0.53	(-0.15, 0.25)	0.39	0.63