







Article

Monitoring forest health using hyperspectral imagery: Does feature selection improve the performance of machine-learning techniques?

Patrick Schratz ¹, Jannes Muenchow ¹, Eugenia Iturritxa ², José Cortés ¹, Bernd Bischl ³ and Alexander Brenning ¹

¹ Friedrich Schiller University Jena, Department of Geography, GIScience group, Germany

² NEIKER Tecnalia, Spain

³ Ludwig-Maximilians-Universität München, Department of Statistics, Computational Statistics group, Germany

* Correspondence: patrick.schratz@uni-jena.de

Abstract: This study analyzed highly correlated, feature-rich datasets from hyperspectral remote sensing data using multiple statistical and machine-learning methods. The effect of filter-based feature selection methods on predictive performance was compared. Also, the effect of multiple expert-based and data-driven feature sets, derived from the reflectance data, was investigated. Defoliation of trees (%), derived from in-situ measurements from fall 2016, was modeled as a function of reflectance. Variable importance was assessed using permutation-based feature importance. Overall, the support vector machine (SVM) outperformed other algorithms such as random forest (RF), extreme gradient boosting (XGBoost), and lasso (L1) and ridge (L2) regressions by at least three percentage points. The combination of certain feature sets showed small increases in predictive performance while no substantial differences between individual feature sets were observed. For some combinations of learners and feature sets, filter methods achieved better predictive performances than using no feature selection. Ensemble filters did not have a substantial impact on performance. The most important features were located around the red edge. Additional features in the near-infrared region (800 nm - 1000 nm) were also essential to achieve the overall best performances. Filter methods have the potential to be helpful in high-dimensional situations and are able to improve the interpretation of feature effects in fitted models, which is an essential constraint in environmental modeling studies. Nevertheless, more training data and replication in similar benchmarking studies are needed to be able to generalize the results.

Keywords: hyperspectral imagery; forest health monitoring; machine learning; feature selection; model comparison

Citation: Schratz, P.; Muenchow, J.; Cortés, J.; Iturritxa, E.; Bischl, B.; Brenning, A. Title. *Remote Sens.* **2021**, *1*, 0. <https://doi.org/>

Received:

Accepted:

Published:

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Submitted to *Remote Sens.* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The use of machine learning (ML) algorithms for analyzing remote sensing data has seen a huge increase in the last decade [1]. This coincided with the increased availability of remote sensing imagery, especially since the launch of the first Sentinel satellite in the year 2014. At the same time, the implementation and usability of learning algorithms has been greatly simplified with many contributions from the open-source community. Scientists can nowadays process large amounts of (environmental) information with relative ease using various learning algorithms. This makes it possible to easily extend benchmark comparison matrices of studies in a semi-automated way, possibly stumbling upon unexpected findings such as process settings that would not have been explored otherwise [2].

ML methods in combination with remote sensing data are used in many environmental fields such as vegetation cover analysis or forest carbon storage mapping [3,4].

The ability to predict these environmental variables in unknown regions qualifies ML algorithms as a helpful tool for such environmental analyses. One aspect of this research field is to enhance the understanding of biotic and abiotic stress triggers, for example by analyzing tree defoliation [5]. Defoliation is known to be a proxy for pathogen and insect damage [6]. While common symptoms are observable across species, some effects and their degree of severity are species-specific [7]. Defoliation has also been shown to increase predisposition of tree death from secondary biotic factors up to ten years after the actual defoliation event [8]. Other approaches for analyzing forest health include change detection [9], or describing the current health status of forests on a stand level [10].

Vegetation indices have shown the potential to provide valuable information when analyzing forest health [11,12]. Most vegetation indices were developed with the aim of being sensitive to changes in specific wavelength regions, serving proxies for underlying plant physiological processes. In some cases, indices developed for different purposes than the one to be analyzed (e.g., defoliation of pine trees) may help to explain complex underlying relationships that are not obvious at first. This emphasizes the need to extract as much information as possible from the available input data to generate promising features which can help to understand the modeled relationship [13]. A less known index type that can be derived from spectral information is the normalized ratio index (NRI). In contrast to most vegetation indices, NRIs do not use an expert-based formula following environmental heuristics; instead, they make use of a data-driven feature engineering approach by combining (all possible) combinations of spectral bands [14]. When working with hyperspectral data, thousands of NRI features can be derived this way.

Even though ML algorithms are capable of handling highly correlated input variables, model fitting becomes computationally more demanding, and model interpretation more challenging. Feature selection approaches can help to address this issue, reducing possible noise in the feature space, simplifying model interpretability, and possibly enhancing predictive performance [15]. Instead of using wrapper feature selection methods, which add a substantial overhead to a nested model optimization approach, especially for datasets with many features, this study focuses on (ensemble) filter methods, which can be directly integrated into the hyperparameter optimization step during model construction.

Overall, this study aims to show how high-dimensional datasets can be handled effectively with ML methods while still allowing the interpretation of the fitted models. The predictive power of non-linear methods and their ability to handle highly correlated predictors is combined with common and new approaches for assessing feature importance and feature effects. Yet, this study focuses mainly on investigating the effects of filter methods and feature set types on predictive performance rather than interpreting individual feature effects.

Considering these opportunities and challenges, the research questions of this study are the following:

- How do different feature selection methods influence the predictive performance of ML models of the defoliation of trees?
- Do different (environmental) feature sets show differences in performance?
- Can predictive performance be substantially improved by combining feature sets?
- Which features are most important and how can these be interpreted in this context?

In recent years, various studies that have used hyperspectral data to analyze pest/fungi infections on trees have been published: Pinto et al. [16] successfully used hyperspectral imagery to characterize pest infections on peanut leaves using random forest while Yu et. al [17] aimed to detect pine wilt disease in pine plots in China using vegetation indices derived from hyperspectral data. Other studies which applied hyperspectral data for forest health monitoring are [18–20]. Building upon these successful applications of hyperspectral remote sensing usage in the field of leaf and tree health monitoring, this work analyses tree defoliation in northern Spain using airborne hyperspectral data. The

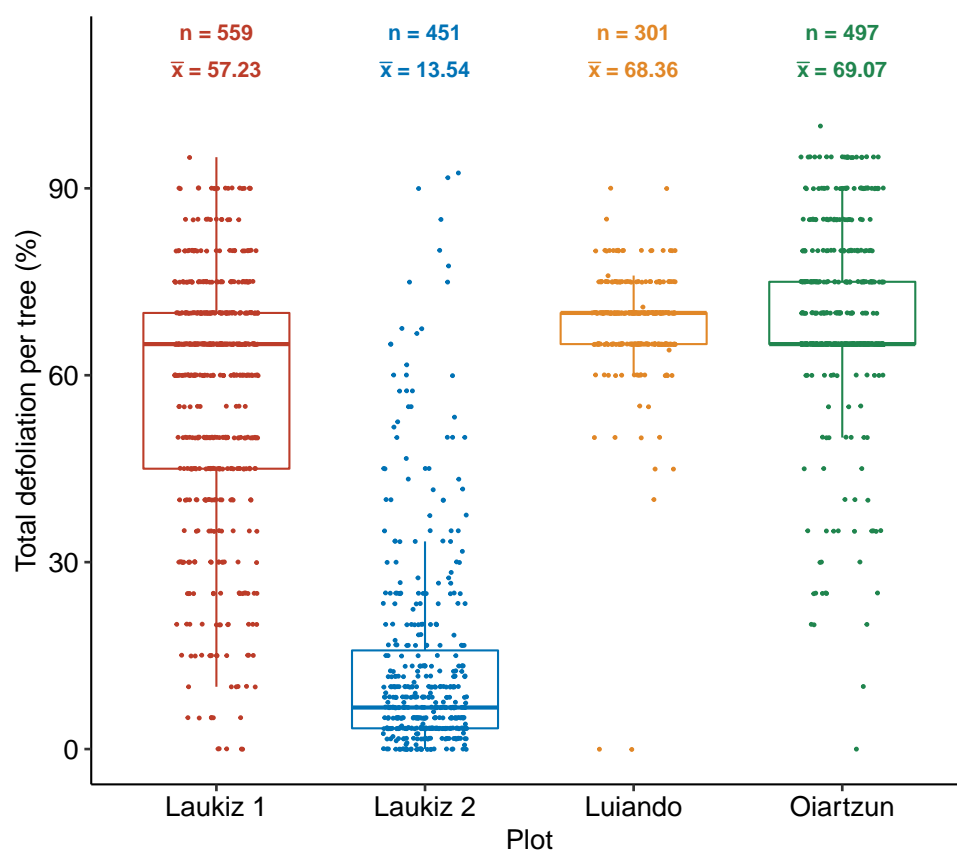


Figure 1. Response variable “defoliation of trees” for plots Laukiz1, Laukiz2, Luiando and Oiartzun. n is the total number of trees in each plot, and \bar{x} the mean defoliation. Values for Laukiz1, Luiando and Oiartzun were observed in 5% intervals; for Laukiz2 defoliation was observed at multiple heights and then averaged, leading to smaller defoliation differences than 5%.

methodology of this study uses ML methods in combination with feature selection and hyperparameter tuning. In addition, feature importance was analyzed. Incorporating the idea of creating data-driven NRIs, this study also discusses the practical problems of high dimensionality in environmental modeling [21,22].

2. Materials and Methods

2.1. Data and study area

Airborne hyperspectral data with a spatial resolution of one meter and 126 spectral bands was available for four Monterey Pine (*Pinus radiata* D. Don) plantations in northern Spain. The trees in the plots suffer from infections with pathogens such as *Diplodia sapinea* (Fr.) Fuckel, *Fusarium circinatum* Nirenberg & O'Donnell, *Armillaria mellea* (Vahl) P. Kumm, *Heterobasidion annosum* (Fr.) Bref, *Lecanosticta acicola* (Thüm) Syd. and *Dothistroma septosporum* (Dorogin) M. Morelet causing (among others) needle blight, pitch canker, and root diseases [23,24]. The first two fungi are mainly responsible for the foliation loss of the trees analyzed in this study [25]. In-situ measurements of defoliation of trees (serving as a proxy for tree health) were collected by visual inspection by experts. Defoliation in percent was used as the response variable (Figure 1).

It is assumed that the fungi infect the trees through open wounds, possibly caused by previous hail damage [25]. The dieback of these trees, which are mainly used as timber, causes substantial economic losses [26].

Table 1: Specifications of hyperspectral data.

Characteristic	Value
Geometric resolution	1 m
Radiometric resolution	12 bit
Spectral resolution	126 bands (404.08 nm — 996.31 nm)
Correction:	Radiometric, geometric, atmospheric

2.1.1. In-situ data

The *Pinus radiata* plots of this study, namely Laukiz1, Laukiz2, Luiando and Oiartzun, are located in the northern part of the Basque Country (Figure 2). Oiartzun has the largest number of observations ($n = 559$ trees) while Laukiz2 is the largest in area (1.44 ha). All plots besides Luiando are located within 100 km from the coast (Figure 2). A total of 1808 observations are available (Laukiz1: 559, Laukiz2: 451, Luiando: 301, Oiartzun: 497). Field surveys were conducted in September 2016 by experienced forest pathologists. Defoliation was measured in 5% steps through visual inspection with the help of a score card. For Laukiz2, values at three height levels (bottom, mid, top) were available and averaged into an overall defoliation value, resulting in values that are not multiples of 5% (e.g., 8.33%). The magnitude of human observer errors in such surveys, including the present one, is not precisely known and has been discussed for many years [27]. [28] estimated human observer errors in defoliation surveys to range between 7% and 18%.

2.1.2. Hyperspectral data

The airborne hyperspectral data was acquired by an AISA EAGLE-II sensor during two flight campaigns on September 28th and October 5, 2016 at noon. All preprocessing steps (geometric, radiometric, atmospheric) were conducted by the Institut Cartogràfic i Geològic de Catalunya (ICGC). The first four bands were corrupted, leaving 122 bands with valid information. Additional metadata information is available in Table 1.

2.2. Derivation of indices

To use the full potential of the hyperspectral data, all possible vegetation indices supported by the R package *hsdar* (89 in total) as well as all possible NRI combinations were calculated. NRIs follow the optimized multiple narrow-band reflectance (OMNBR) concept of data-driven information extraction from narrow-band indices of hyperspectral data [13,14]. While various index formulations such as band ratios or normalized ratios are available, indices involving the same bands are strongly correlated. Since the widely used NDVI index belongs to the group of normalized ratio indices (NRIs), which are implemented in the *hsdar* R package, we used the following normalized difference index (NDI) formula to combine all pairs of reflectances:

$$NRI_{i,j} = \frac{band_i - band_j}{band_i + band_j} \quad (1)$$

where i and j are the respective band numbers.

To account for geometric offsets within the hyperspectral data, which were reported by ICGC to be potentially up to one meter, a buffer of one meter radius around the centroid of each tree was used when extracting the reflectance values. A pixel was considered to fall into a tree's buffer zone if the centroid of the respective pixel was touched by the buffer. The pixel values within a buffer zone were averaged and formed the final reflectance value of a single tree and were used as the base information to derive all additional feature sets. In total, $\frac{121 \times 122}{2} = 7381$ NRIs were calculated.

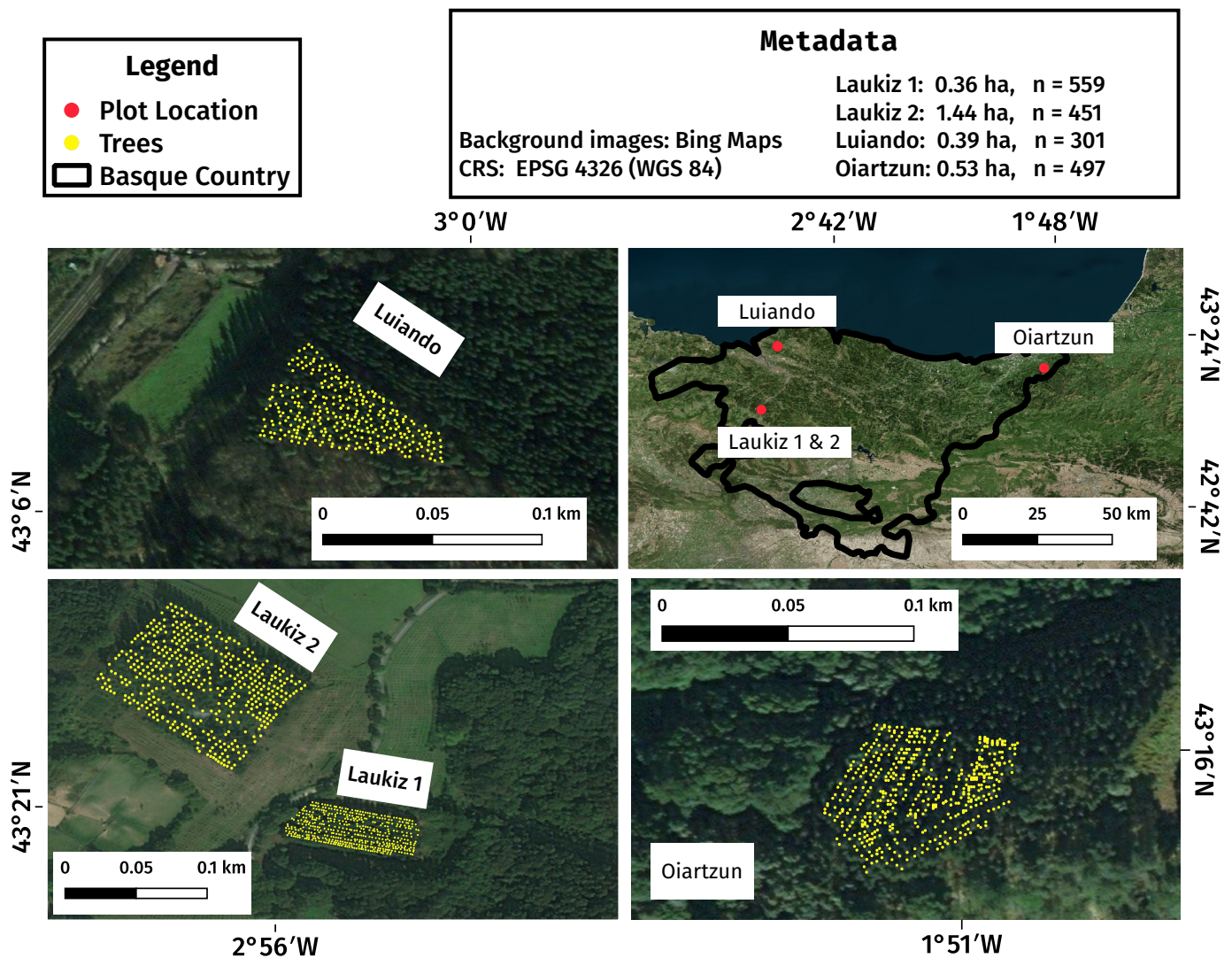


Figure 2. Study area maps showing information about location, size and spatial distribution of trees for all plots (Laukiz1, Laukiz2, Luiando, Oiartzun). The background maps give a visual impression of the individual plot area but do not necessarily represent the plot's state during data acquisition.

2.3. Feature selection

High-dimensional, feature-rich datasets come with several challenges for both model fitting and evaluation.

- Model fitting times increase.
- Noise is possibly introduced into models by highly correlated variables [29].
- Model interpretation and prediction become more challenging [29].

To reduce the feature space of a dataset, several conceptually differing approaches exist: wrapper methods, filters, penalization methods (lasso and ridge), and principal component analysis (PCA) [30–33]. In contrast to wrapper methods, filters have a much lower computational cost, and their tuning can be added to the hyperparameter optimization step. In addition, filters are less known than wrapper methods and in recent years, ensemble filters were introduced which have shown promising results compared to single filter algorithms [34]. These two points mainly led to the decision to focus on filter methods for this work and evaluate their effectiveness on highly correlated,

high-dimensional datasets. Due to this focus, only this subgroup of feature selection methods will be introduced in greater detail in the following sections.

2.3.1. Filter methods

The concept of filters originates from the idea of ranking features following a score calculated by an algorithm [32]. Some filter methods can only deal with specific types of variables (e.g., numeric or nominal). Filters only rank features, they do not decide which covariates to drop or keep [35]. The decision which features to keep for model fitting can be integrated into the optimization phase during model fitting, along with hyperparameter tuning. Thus, the number of top-ranked features to be included in the model is treated as an additional hyperparameter of the model. This hyperparameter is tuned to optimize the model's performance.

Beyond the use of individual filter methods to rank and select features, recent studies have shown that combining several filters by using statistical operations such as 'minimum' or 'mean' may enhance the predictive performance of the resulting models, especially when applied to multiple datasets [34,36]. This approach is referred to as 'ensemble filtering' [37]. Ensemble filters align with the recent rise of the ensemble approach in ML, which uses the idea of stacking to combine the predictions of multiple models, aiming to enhance predictive performance [38–40]. In this work the Borda ensemble filter was used [34]. Its overall feature order is determined by the sum of filter ranks of all individual filters in the ensemble.

Filter methods can be categorized based on three binary criteria: multivariate or univariate feature use, correlation or entropy-based importance weighting, and linear and non-linear filter methodology. Care needs to be taken to not weigh certain classes more than others in the ensemble as otherwise the ranking will be biased. In this study this was taken care of by checking the rank correlations (Spearman's correlation) of the generated feature rankings of all methods against each other. If filter pairs showed a correlation of 0.9 or higher, only one of the two was included in the ensemble filter, selecting it at random.

2.3.2. Description of used filter methods

Filter methods can be classified as follows (Table 2):

- Univariate/multivariate (scoring based on a single variable/multiple variables).
- Linear/non-linear (usage of linear/non-linear calculations).
- Entropy/correlation (scoring based on derivations of entropy or correlation-based approaches).

The filter 'Information Gain' in its original form is only defined for nominal response variables:

$$H(Class) + H(Attribute) - H(Class, Attribute) \quad (2)$$

where H is the conditional entropy of the response variable (class or Y) and the feature (attribute or X), respectively. $H(X)$ is Shannon's Entropy [47] for a variable X

Table 2: List of filter methods used in this work, their categorization and scientific reference.

Name	Group	Ref.
Linear correlation (Pearson)	univariate, linear, correlation	[41]
Information gain	univariate, non-linear, entropy	[42]
Minimum redundancy, maximum relevance	multivariate, non-linear, entropy	[43]
Carscore	multivariate, linear, correlation	[44]
Relief	multivariate, linear, entropy	[45]
Conditional minimal information maximization	multivariate, linear, entropy	[46]

and $H(X, Y)$ is a joint Shannon's Entropy for a variable X with a condition to Y . $H(X)$ itself is defined as

$$H(X) = - \sum_{i=1}^n P(x_i) \log_b P(x_i) \quad (3)$$

where b is the base of the logarithm used, most commonly 2.

In order to use this method with a numeric response (percentage defoliation of trees), the variable was discretized into equal bins $n_{bin} = 10$ and treated as a categorical variable.

2.4. Benchmarking design

2.4.1. Algorithms

The following learners were used in this work:

- Extreme Gradient Boosting (XGBoost)
- Random Forest (RF)
- Penalized Regression (with L1/lasso and L2/ridge penalties)
- Support Vector Machine (SVM, radial basis function Kernel)
- Featureless learner

RF and SVM are well-established algorithms and widely used in environmental remote sensing. Extreme Gradient Boosting (commonly abbreviated as XGBoost) has shown promising results in benchmarking studies in recent years. Penalized regression is a statistical modeling technique capable of dealing with highly correlated covariates by penalizing the model coefficients [48]. Common penalties are 'lasso' (L1) and 'ridge' (L2). Ridge regression does not remove variables from the model (penalization to zero), but it shrinks them towards zero, keeping them in the model. A featureless learner was included for a baseline comparison.

In total, the benchmarking grid consisted of 156 experiments (6 feature sets \times 3 ML algorithms \times 8 feature-selection methods and for the L1/L2 models, 6 feature sets \times 2 models). The selected hyperparameter settings are shown in appendix Table 8. All code and data are included in the research compendium of this study (<https://doi.org/10.5281/zenodo.2635403>).

2.4.2. Feature sets

Three feature sets were used in this study, each representing a different approach to feature engineering:

- The raw hyperspectral band information (HR): no feature engineering
- Vegetation Indices (vegetation index (VI)s): expert-based feature engineering
- Normalized Ratio Indices (NRIs): data-driven feature engineering

The idea of splitting the features into different sets originated from the question whether feature-engineered indices derived from reflectance values have a positive effect on model performance. Peña et al. 2017 [49] is an exemplary study which used this approach in a spectro-temporal setting. Benchmarking learners on these feature sets while keeping all other variables such as model type, tuning strategy and partitioning method fixed makes it possible to draw conclusions on their individual impact. Each feature set has distinct capabilities that differentiate it from the others. This can have both a positive and negative effect on the resulting performance, which is one question this study aims to explore. For example, feature set VI misses certain parts of the spectral range as the chosen indices only use specific spectral bands. Feature set NRI will introduce highly correlated features, for which some algorithms may be more suitable than others.

In addition to these individual feature sets, the following combinations of feature sets were also compared:

- HR + VI

- 245 • HR + NRI
- 246 • HR + VI + NRI

247 Some individual features were removed before using the datasets for modeling when
 248 being numerically equivalent to another feature based on the pairwise correlation being
 249 greater than $1 - 10^{-10}$. This preprocessing step reduced the number of covariates for
 250 feature set VI to 86 (from 89). This decision was made to prevent numerical issues
 251 that may occur in the subsequent tuning, filtering and model fitting steps when offering
 252 features with a pairwise correlation of (almost) one. The remaining features were then
 253 used as input for the filter-based feature selection within the CV.

254 2.4.3. Hyperparameter optimization

255 Hyperparameters were tuned using model-based optimization (MBO) within a
 256 nested spatial cross-validation (CV) [50–52]. In MBO, first n hyperparameter settings
 257 are randomly chosen from a user-defined search space. After these n settings have
 258 been evaluated, one new setting, which is going to be evaluated next, is proposed by a
 259 fitted surrogate model (by default, a kriging method). This strategy continues until a
 260 user-defined stopping criterion is satisfied [53,54].

261 In this work, an initial design of 30 randomly drawn hyperparameter settings in
 262 combination with a stopping criterion of 70 iterations was used, resulting in a total budget
 263 of 100 evaluated hyperparameter settings per fold. The advantage of this tuning approach
 264 is a substantial reduction of the tuning budget that is required to find a setting close
 265 to the global optimization minimum. MBO may outperform methods that do not use
 266 information from previous iterations, such as random search or grid search [55]. Tuning
 267 ranges used in this work are shown in Table 8.

268 To optimize the number of features used for model fitting, the percentage of features
 269 was added as a hyperparameter during the optimization stage [51]. For PCA, the number
 270 of principal components was tuned. The RF hyperparameter m_{try} was re-expressed as
 271 $m_{try} = p_{sel}^t$, a function of the number of selected features, p_{sel} . It was thus tuned on
 272 a logarithmic scale by varying t between 0 (i.e., $m_{try} = 1$) and 0.5 (i.e., $m_{try} = \sqrt{p_{sel}}$).
 273 This was necessary to ensure that m_{try} did not exceed the number of features available
 274 after optimizing the feature percentage during tuning.

275 2.4.4. Spatial resampling

276 A spatial nested cross-validation on the plot level was chosen to account for spatial
 277 autocorrelation within the plots and assess model transferability to different plots [52,56].
 278 The root mean square error (RMSE) was chosen as the error measure. Each plot served
 279 as one cross-validation fold, resulting in four iterations in total. The inner level of
 280 cross-validation for hyperparameter tuning also used plot-level cross-validation.

281 2.5. Feature importance and feature effects

282 Estimating feature importance for datasets with highly correlated features is a
 283 challenging task for which numerous model-specific and model-agnostic approaches exist
 284 [48,57,58]. The strong correlations among features make it challenging to calculate an
 285 unbiased estimate for single features [59]. Methods like partial dependence plots (PDP)
 286 or permutation-based approaches may produce unreliable estimates in such scenarios
 287 because unrealistic combinations of feature values are created [59]. The development
 288 of robust methods that enable an unbiased estimation of feature importance for highly
 289 correlated variables are subject to current research [60].

290 In this work, permutation-based feature importance was calculated to estimate
 291 feature importance or effects [61]. With the limitations in mind when applied to cor-
 292 related features, the aim was to get a general overview of the feature importance of
 293 the hyperspectral bands while trying to avoid an over-interpretation of results. The
 294 best-performing algorithm on the HR task (i.e., SVM) was used for the feature importance
 295 calculation.

To facilitate interpretation, the ten most important indices of the best performing models using feature sets HR and VI were linked to the spectral regions of the hyperspectral data that went into their calculation. The aim was to visualize the most important features along the spectral curve of the plots to better understand which spectral regions were most important for the model.

2.6. Research compendium

All tasks of this study were conducted using the open-source statistical programming language R [62]. A complete list of all R packages used in this study can be found in the linked repositories mentioned in the next paragraph. Due to space limitations, only selected packages with high impact on this work will be explicitly cited.

The algorithm implementations of the following packages have been used: xgboost [63] (*Extreme Gradient Boosting*), kernlab [64] (Support Vector Machine) and glmnet [65] (penalized regression). The filter implementations of the following packages have been used: praznik [66], FSelectorRcpp [67]. Package mlr [68] was used for all modeling related steps. drake [69] was used for structuring the work and for reproducibility. This study is available as a research compendium on Zenodo ([10.5281/zenodo.2635403](https://doi.org/10.5281/zenodo.2635403)). Apart from the availability of code and manuscript sources, a static webpage is available at (<https://pat-s.github.io/2019-feature-selection>), which includes additional side analyses that were carried out during the creation of this study.

3. Results

3.1. Principal component analysis of feature sets

PCA was used to assess the complexity of the three feature sets. Depending on the feature set, 95% of the variance is explained by two (HR), twelve (VI) and 42 (NRI) principal component (PC)s. HR features are therefore highly redundant, while the applied feature transformations enrich the data set, at least from an exploratory linear perspective.

3.2. Predictive performance

Overall, the response variable “tree defoliation” could be modeled with an RMSE of 28 percentage points (p.p.) (Figure 3). SVM showed almost no differences in RMSE across feature sets whereas other learners (RF, SVM, XGBoost, lasso and ridge) differed up to five p.p. (Figure 3). SVM showed the best overall performance with a mean difference of around three p.p. to the next best model (XGBoost) (Table 5). Performance differences between test folds were large: Predicting on Luiando resulted in an RMSE of 9.0 p.p. for learner SVM (without filter) but up to 54.3 p.p. when testing on Laukiz2 (Table 6).

The combination of feature sets showed small increases in performance for some learners. XGBoost scored slightly better on the combined datasets HR-NRI, NRI-VI and HR-NRI-VI, respectively, compared to their standalone variants (NRI, VI) (Figure 3). However, the best performances for RF and XGBoost were scored on NRI and HR, respectively. RF showed a substantial performance increase when using only NRI compared to all other feature sets whereas for XGBoost the worst performances were associated with the VI and NRI only feature sets (Figure 3).

SVM combined with the “Information Gain” filter achieved the best overall performance (RMSE of 27.915 p.p.) (Table 3). Regression with ridge (L2) and lasso (L1) penalties showed their best performances on the NRI feature set (Table 5). Combining feature sets for lasso and ridge did not help to increase performance, and while there was no substantial difference for lasso, the performance of ridge improved by around two percentage points. XGBoost showed very poor performances for some feature sets and fills the last ten places of the ranking (Table 4). Especially when combined with PCA, the algorithm was not able to achieve adequate performances.

Effects of filter methods on performance differed greatly between algorithms: SVM showed no variation in performance across filters (Figure 4). The use of filters for RF resulted in a substantial increase in performance in all tasks, especially on the HR feature set where all filters showed an improved score compared to using no filter (Figure 4). XGBoost's performance depended strongly on feature selection: In two out of six tasks (HR, VI), using no filter resulted in the worst performance. XGBoost showed the highest overall differences between filters for a single task: for feature set HR, the range is up to 13 p.p. ("CMIM" vs. "no filter") (Figure 4).

When comparing the usage of filters against using no filter at all, there were no instances in which a non-filtered model scored a better performance than the best filtered one (Figure 4). For SVM, all filters and "no filter" achieved roughly the same performance on all tasks.

The Borda ensemble filter was not able to score the best performance in any learner/filter setting (Figure 5). For RF and XGBoost, it most often ranked within the better half among all filters of a specific task.

The number of features selected during model optimization strongly varied across learners and plots. RF selected the least features of all three learners, and with the exception of Oiartzun, only one or two variables were selected. SVM used 210 features or more in all instances and selected between 16% (Laukiz1) and 81% (Oiartzun) of the features (Table 7). XGBoost also favored using several hundred features with the exception of Laukiz2 for which only 14 (0.96%) were selected.

Table 3: Best ten results among all learner-task-filter combinations, sorted in decreasing order of RMSE (p.p.) and their respective standard error (SE).

	Task	Model	Filter	RMSE	SE
1	NRI-VI	SVM	Info Gain	27.915	18.970
2	NRI	SVM	CMIM	28.044	19.101
3	VI	SVM	Relief	28.082	19.140
4	NRI-VI	SVM	Borda	28.102	19.128
5	HR	SVM	CMIM	28.119	19.123
6	HR	SVM	MRMR	28.119	19.123
7	VI	SVM	Info Gain	28.121	19.123
8	NRI	SVM	PCA	28.121	19.123
9	HR-NRI	SVM	PCA	28.121	19.123
10	HR-NRI-VI	SVM	PCA	28.121	19.123

Table 4: Worst ten results among all learner-task-filter combinations, sorted in decreasing order of RMSE (p.p.) and their respective standard error (SE).

	Task	Model	Filter	RMSE	SE
1	VI	XGBoost	No Filter	45.366	6.672
2	HR	XGBoost	No Filter	44.982	5.378
3	VI	XGBoost	PCA	44.539	8.187
4	HR	XGBoost	PCA	44.032	6.183
5	NRI	XGBoost	PCA	43.433	9.543
6	HR-NRI	XGBoost	PCA	43.220	2.557
7	HR-NRI-VI	XGBoost	PCA	41.076	9.862
8	VI	RF	CMIM	39.980	10.144
9	VI	RF	Info Gain	39.623	10.616
10	NRI	XGBoost	Pearson	39.492	11.548

Table 5: The overall best individual learner performance across any task and filter method for RF, SVM, XGBoost, Lasso and Ridge, sorted ascending by RMSE (p.p.) including the respective standard error (SE) of the cross-validation run. For extttregr.featureless the Task is no applicable and was therefore removed.

	Task	Model	Filter	RMSE	SE
1	NRI-VI	SVM	Info Gain	27.915	18.970
2	NRI	RF	Relief	30.842	12.028
3	HR	XGBoost	Info Gain	31.165	15.025
4	NRI	Lasso-MBO	No Filter	31.165	15.025
5	NRI	Ridge-MBO	No Filter	31.165	15.025
6	-	regr.featureless	No Filter	31.165	15.025

Table 6: Test fold performances in RMSE (p.p.) for learner SVM on the HR dataset without using a filter, showcasing performance variance on the fold level. For each row, the model was trained on observations from all others plots but the given one and tested on the observations of the given plot.

	RMSE	Test Plot
1	28.12	Laukiz1
2	54.26	Laukiz2
3	9.00	Luiando
4	21.17	Oiartzun

Table 7: Selected feature portions during tuning for the best performing learner-filter settings (SVM Relief, RF Relief, XGBoost CMIM) across folds for task HR-NRI-VI, sorted by plot name. 'Features (#)' denotes the absolute number of features selected and 'Features (%)' refers to the percentage relative to the overall features available in the training sets for each plot (Laukiz1 = 1249, Laukiz2 = 1357, Luiando = 1507, Oiartzun = 1311). Results were estimated in a separate model tuning step, not within the main cross-validation comparison.

Learner	Test Plot	Features (%)	Features (#)
RF	Laukiz1	0.00245	1/1249
	Laukiz2	0.00359	1/1357
	Luiando	0.12448	2/1507
	Oiartzun	2.80356	37/1311
SVM	Laukiz1	16.76686	210/1249
	Laukiz2	40.77700	554/1357
	Luiando	43.80604	661/1507
	Oiartzun	81.23205	1065/1311
XGB	Laukiz1	79.54091	994/1249
	Laukiz2	0.96545	14/1357
	Luiando	66.27871	999/1507
	Oiartzun	41.89759	550/1311

3.3. Variable importance

3.3.1. Permutation-based Variable Importance

The most important features for datasets HR and VI showed an average decrease in RMSE of 1.06 p.p. (HR, B65) and 1.93 p.p. (VI, Vogelmann2) when permuted (Figure 6). For the HR dataset, most (i.e., six out of ten) relevant features clustered around the infrared region (920 nm - 1000 nm), while for VI, eight out of ten concentrate within the wavelength range of 700 nm - 750 nm (the so/called "red edge"). For HR, four features in the infrared region (920 nm - 1000 nm) were identified by the model to be most important, being associated with a mean decrease in RMSE of around 1 p.p.

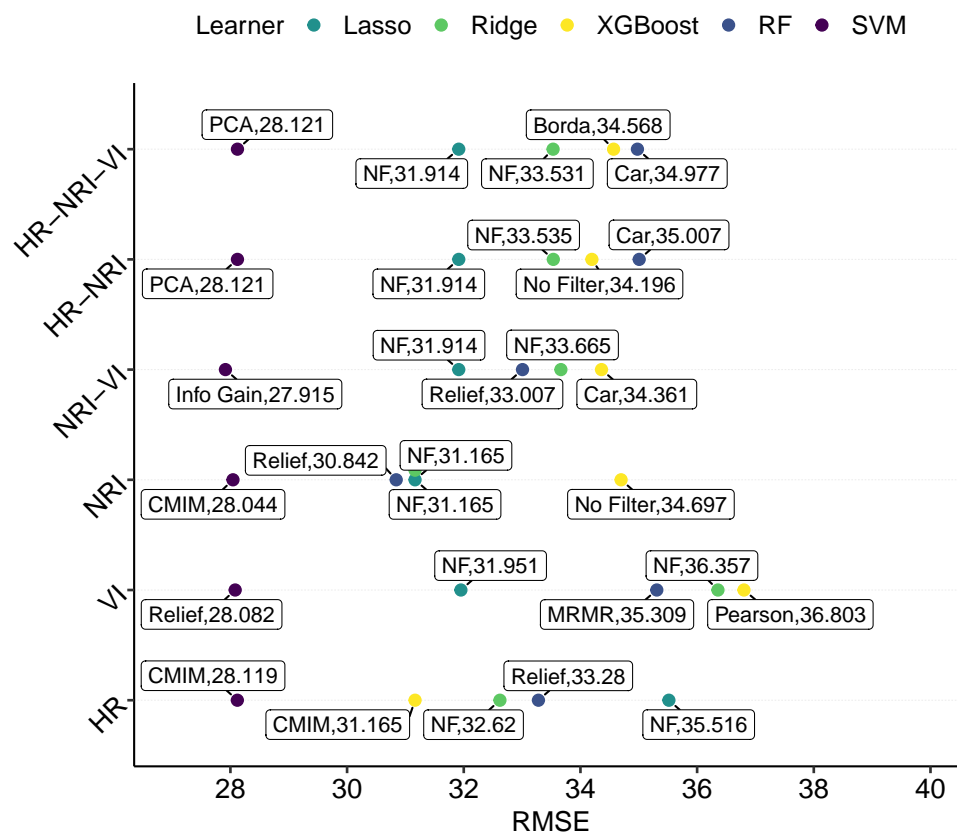


Figure 3. Predictive performance in RMSE (p.p.) of models across tasks. Different feature sets are shown on the y-axis. Labels show the feature selection method (e.g., NF = no filter, Car = 'Carscore', Info Gain = 'Information Gain', Borda = 'Borda').

Overall, apart from the top five features, the vast majority of features showed only a small importance with average decreases in RMSE below 0.5 p.p.

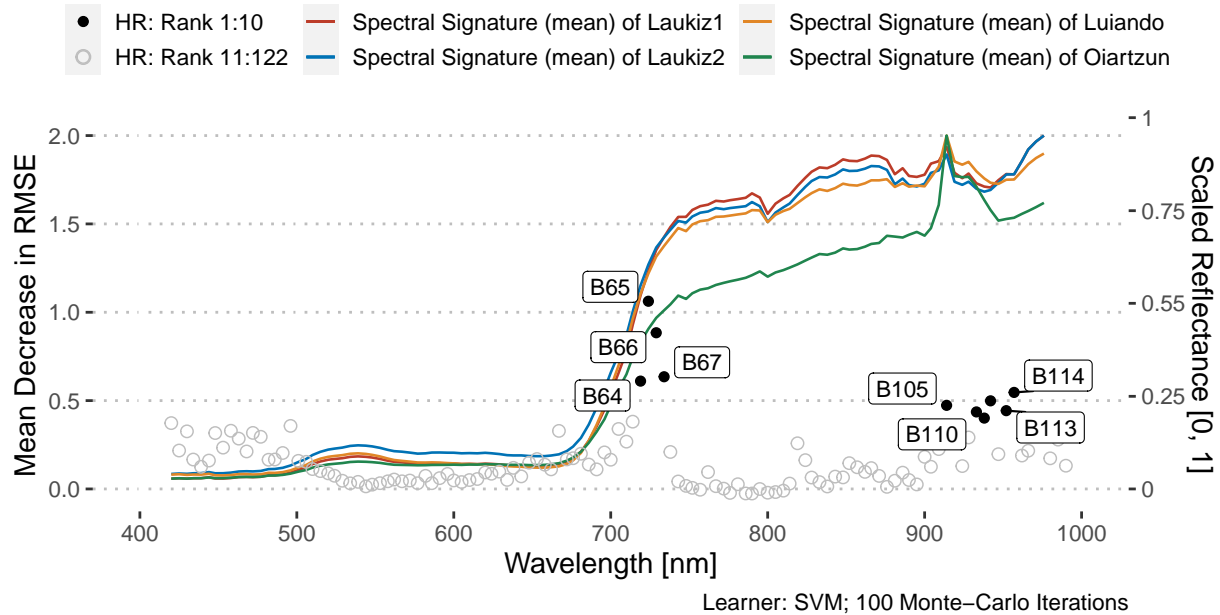
4. Discussion

4.1. Predictive Performance

The best overall performance achieved in this study (SVM with the “Info Gain” filter, RMSE 27.915 p.p.) has to be seen in the light of model overfitting (see subsection 4.2). Leaving out the performance on Laukiz2 when aggregating results, the mean RMSE would be around 19 p.p. However, leaving out a single plot would also change the prediction results for the other plots because the observations from Laukiz2 would not be available for model training. Due to the apparent presence of model overfitting in this study, it is suggested that more training data representing a greater variety of situations is needed. A model can only make robust predictions if it has learned relationships across the whole range of the response. Hence, care should be taken when predicting on the landscape scale using models fitted on this dataset due to their lack of generalizability caused by the limitations of the available training data. However, when inspecting the fold/level performances, it can be concluded that the models performed reasonably well predicting defoliation greater than 50%, but they failed for lower levels. This applied to all learners of this study. In this study, the overall performance across all learners can be classified as “poor” given that only the SVM learner was able to substantially outperform the featureless learner (Table 5). It is worth noting that data quality issues may have affecting model performances, as discussed below in detail (subsection 4.5).

Permutation-based Variable Importance for dataset 'HR'

The ten most important features are labeled by their band number.



Permutation-based Variable Importance for dataset 'VI'

The ten most important features are labeled by their index name.

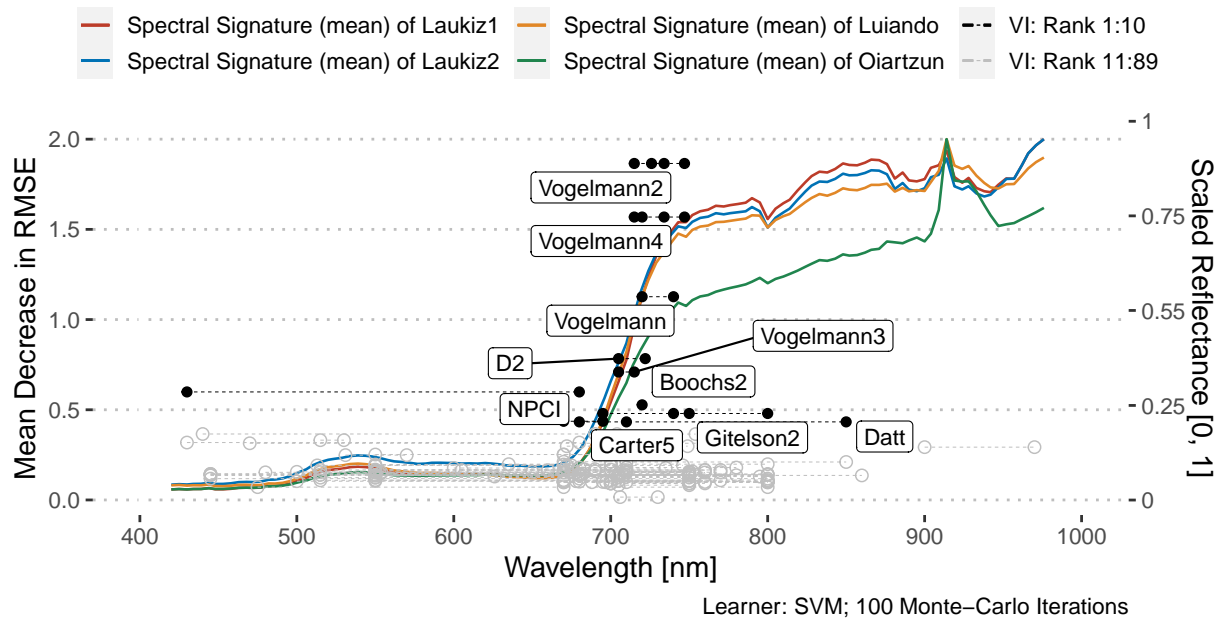


Figure 6. Variable importance for feature sets HR and VI: Mean decrease in RMSE for one hundred feature permutations using the SVM learner. The wavelength range on the x-axis matches the range of the hyperspectral sensor (400 nm - 1000 nm). For each dataset, the ten most important features were highlighted as black dots and labeled by name. Gray dots represent features from importance rank 11 to last. The spectral signature (mean) of each plot was added as a reference on a normalized reflectance scale [0, 1] (secondary y-axis). VI features were decomposed into their individual formula parts; all instances being connected via dashed lines. Each VI feature is composed out of at least two instances.

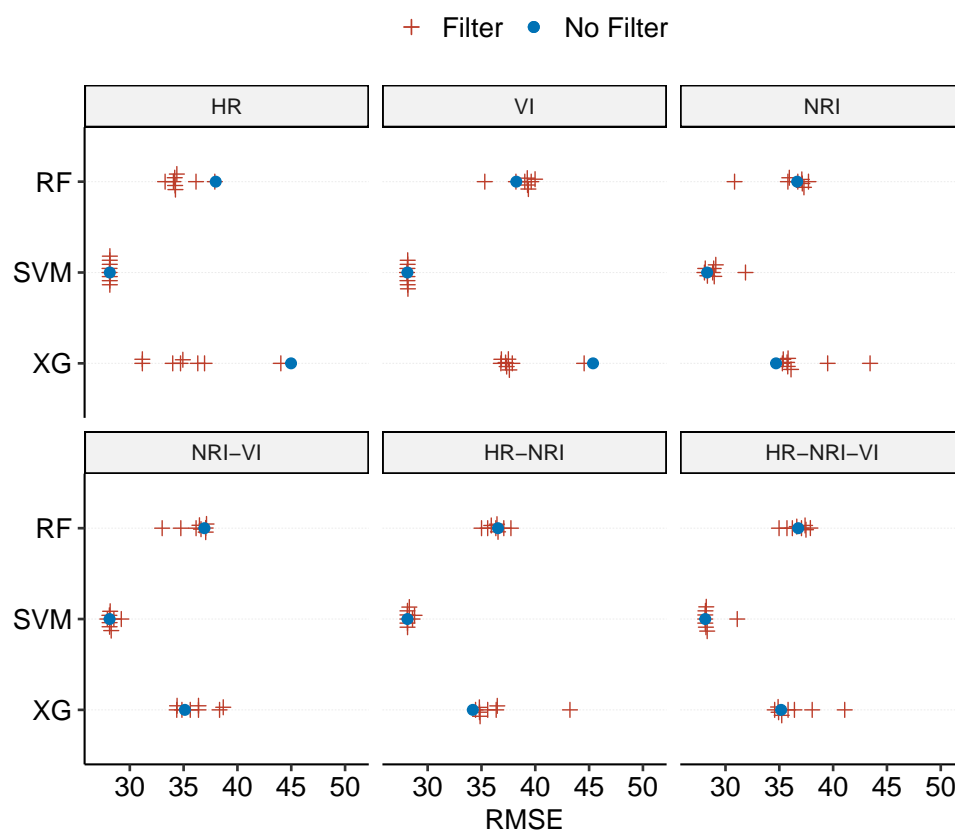


Figure 4. Model performances in RMSE across all tasks, split up in facets, when using no filter method (blue dot) compared to any other filter method (red cross) for learners RF, SVM and XGBoost (XG)

4.1.1. Model differences

An interesting finding is the strength of the SVM algorithm when comparing its predictive performance to its competitors (Table 5). These cluster around a performance of 31 p.p. while SVM scored about three p.p. better than all other methods. However, we refrain from comparing these results (both relatively and absolute) to other studies since many study design points have an influence on the final result (optimization strategy, data characteristics, feature selection methods, etc.).

A potential limiting factor in this study could be the upper limit of 70 iterations used for the XGBoost algorithm (hyperparameter rounds), especially for feature sets including NRIs (Table 8). This setting was a compromise between runtime and tuning space extension with the goal to work well for most feature sets. It may be recommendable to increase this upper limit to a value closer to the number of features in the dataset in order to be able to exploit the full potential of this hyperparameter.

4.1.2. Feature set differences

One objective of this study was to determine whether expert-based and data-driven feature engineering have a positive influence on model performance. With respect to Figure 3, no overall positive or negative pattern related to specific feature sets was found that would be valid across all models. The performance of RF and XGBoost on the VI feature set was around four to six p.p. lower than on others. One reason could be the lack of coverage in the wavelength range between 810 nm and 1000 nm (Figure 6). In addition, for all learners but SVM a better performance was observed when NRI indices were included in the feature set (i.e., NRI-VI, HR-NRI, HR-NRI-VI).

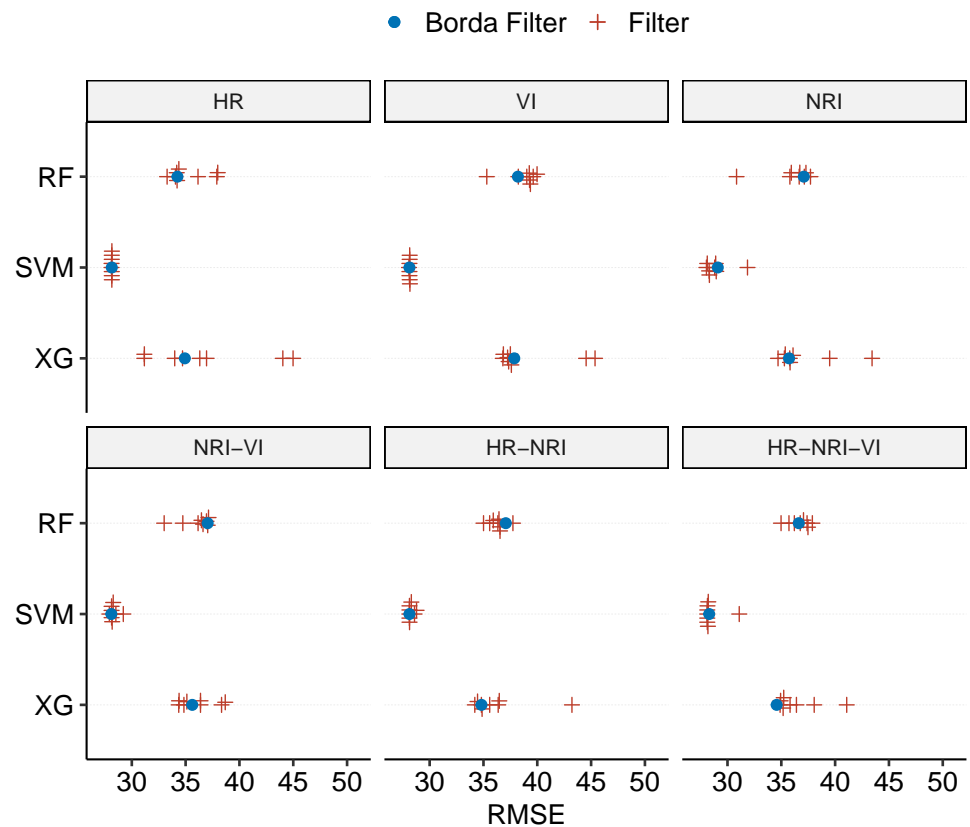


Figure 5. Predictive performances in RMSE (p.p.) when using the Borda filter method (blue dot) compared to any other filter (red cross) for each learner across all tasks.

4.2. Performance vs. plot characteristics

The large differences in RMSE obtained on different test folds can be attributed to model overfitting (Table 6). An RMSE of 54.26 p.p. reveals the model's inability to predict tree defoliation on this plot (Laukiz2). Laukiz2 differs highly in the distribution of the response variable defoliation compared to all other plots (Figure 1). In the prediction scenario for Laukiz2, the model was trained on data containing mostly medium-to-high defoliation values and only few low ones. This caused overfitting on the medium-to-high values, degrading the model's predictive performance in other scenarios. When Laukiz2 was in the training set, the overall mean RMSE was reduced by up to 50% with single fold performances as good as 9 p.p. RMSE (with Luiando as test set).

There was also no clear pattern in the percentage of features selected based on hyperparameter tuning (Table 7). The optimal value for the number of features (interpreted as a percentage of available features), which are selected from the ranked filter results, is determined by the internal surrogate learner of the MBO tuning method using the results from the previous tuning iterations. Due to this iterative approach MBO is in some ways able to evaluate how well a learning algorithm plays together with a certain amount of selected features and is subsequently able to adjust the number of variables to an optimal value. All in all, considering SVM's relative success, the use of at least a few hundred features from the combined feature set appears to be beneficial, or at least not harmful when the model's built-in regularization is capable of dealing with the resulting high-dimensional situation.

Realizing early during hyperparameter optimization that only few features are needed to reach adequate performances can reduce the overall computational runtime substantially. Hence, regardless of the potential advantage of using filters for increased

predictive performance, these can have a strong positive effect on runtime, especially of models making use of hyperparameters that depend on the available number of features, such as RF with m_{try} .

Ultimately, the results of Table 7 should be taken with care as they rely on single model-filter combinations and are subject to random variation. More in-depth research is needed to investigate the effect of filters on other criteria than performance (such as runtime), leading to a multi-criteria optimization problem.

4.3. Feature selection methods

The usefulness of filters with respect to predictive performance in this study varied. While the performance of some models (up to five p.p. for RF and XGBoost) was improved by specific filters, some models achieved a poorer performance with filters than without them (Figure 4). There was no pattern of specific filters consistently resulting in better scores. Hence, it is recommended to test multiple filters in a study if it is intended to use filters. While filters can improve the performance of models, they may be more appealing based on other aspects than performance: Reducing variables can reduce computational efforts in high-dimensional scenarios and may enhance the interpretability of models. Filters are a lot cheaper to compute than wrapper methods, and the final feature subset selection can be integrated as an additional hyperparameter into the model optimization stage.

Models that used the Borda ensemble method in this study did not perform better on average than models which used a single filter or no filter at all. Ensemble methods have higher stability and robustness than single ones and have shown promising results in [34]. Hence, their expected main advantage is stable performances across datasets with varying characteristics. Single filter methods might yield better model performances on certain datasets but fail on others. The fact that this study used multiple feature sets but only one dataset and tested many single filters could be a potential explanation why in all cases a single filter outperformed the ensemble filter. However, studies which used ensemble filters are still rare, and these are usually not compared against single filters [70]. In summary, in this study Borda did not perform better than a randomly selected filter method. More case studies applying ensemble filter methods are needed to verify this finding. Nevertheless, ensemble filters can be a promising addition to a ML feature-selection portfolio.

PCA, acting as a filter method, more often showed less than optimal results, especially for algorithms RF and XGBoost. Especially XGBoost had substantial problems when using PCA as a filter method and scored four of the six worst results (Table 4). However, PCA was able to reduce model fitting times substantially across all algorithms. Depending on the use case, PCA can be an interesting option to reduce dimensionality while keeping runtime low. However, information about the total number of features used by the model is lost when applying this technique. Since filter scores only need to be calculated once for a given dataset in a benchmark setting, the runtime advantage of a PCA vs. filter methods might in fact be negligible in practice.

4.4. Linking feature importance to spectral characteristics

Not surprisingly the most important features for both HR and VI datasets were identified around the red edge of the spectra, specifically in the range of 680 nm to 750 nm.

This area has the highest ability to distinguish between reflectances related to a high density/high foliage density and thus the health status of vegetation and its respective counterpart [71]. However, four out of ten of the most important features of dataset HR are located between 920 nm and 1000 nm. Looking at the spectral curves of the plots, apparent reflectance differences can be observed in this spectral range—especially for plot Oiartzun—, which might explain why these features were considered important by the model.

A possible explanation for the poor performances of most models scored on the VI dataset compared to all other feature sets could be the lack of features covering the area between 850 nm and 1000 nm (Figure 6). The majority of VI features covers the range between 550 nm - 800 nm. Only one index (PWI) covers information in the range beyond 900 nm.

4.5. Data quality

Environmental datasets always come with some constraints that can have potential influence on the modeling process and its outcome. Finding a suitable approach to extract the remote sensing information from each tree was a complex process. Due to the reported geometric offset of up to one meter within the hyperspectral data, the risk of assigning a value to an observation that would actually refer to a different, possibly non-tree, pixel was reasonably high. It was concluded that using a buffer radius of one meter can be a good compromise between the inclusion of information from too many surrounding trees and an under-coverage of the tree crown. With the chosen radius, we are confident that we were able to map individual tree crowns while accounting for a possible geometric offset. This results in all cases in four contributing pixels (= four square meters) for the extraction of hyperspectral information for a given tree. Even though no results showing the influence of different buffer values on the extraction were provided, it is hypothesized that the relationships between features would not change substantially, leading to almost identical model results. Instead of using a buffer to extract the hyperspectral information, segmentation could have been considered. However, this method would have required more effort for no clear added value in our view and would have moved the focus of this manuscript more to data preprocessing and away from feature selection methods.

Trees located within grid cells on the border of a plot are a notable exception where the exact number of pixels contributing to the observation's feature value may be reduced since the image was cropped to the plot's extent. Cropping was applied to avoid the accidental inclusion of background data such as forest roads. However, this effect was deemed to be of negligible importance.

The available hyperspectral data covered a wavelength between 400 nm and 1000 nm. Hence, the spectral range of the shortwave infrared (SWIR) region is not covered in this study. Given that this range is often used in forest health studies [72], e.g., when calculating the normalized difference moisture index (NDMI) index [73], this marks a clear limitation of the dataset at hand.

The dataset consists of in-situ data collected during September 2016, which was matched against remote sensing data acquired at the end of September 2016. A multi-temporal dataset consisting of in-situ data from different phenology stages would possibly improve the achieved model performances. However, this would also require the costly acquisition of hyperspectral data of these additional timestamps.

The R package *hsdar* was used for the calculation of vegetation indices [74]. All indices that could be calculated with the given spectral range of the data (400 nm - 1000 nm) were used. This means even though Table 9 lists all indices available in the package, not all listed indices were used in this study. Although this selection included a large number of indices, some possibly helpful indices might have been missed due to the restriction of the hyperspectral data.

Overall, the magnitude of uncertainty introduced by the mentioned effects during index derivation cannot be quantified. Such limitations and uncertainties apply to most environmental studies and cannot be completely avoided.

4.6. Practical implications on defoliation and tree health mapping

Even though this work has a strong methodological focus by comparing different benchmark settings on highly correlated feature sets, implications on tree health should be briefly discussed in the following. Due to the outlined dataset issues in subsection 4.5, which are mainly responsible for the resulting poor model performances, the trained

models are not suited for practical use, e.g., to predict defoliation in unknown areas, due to the high mapping uncertainty. Yet the general approach of utilizing hyperspectral data to classify the health status of trees partly led to promising results: For example, due to the narrow bandwidth of the hyperspectral sensor, small parts of the spectrum (mainly in the infrared region) were of higher importance to the models (e.g., see Figure 6), meaning that they helped the models to distinguish between low and high tree defoliation. If spatial offset errors of the image data and possible background noise can be reduced (possibly by making use of image segmentation), we believe that model performances could be substantially enhanced. Such improved models, starting around an RMSE of 20% and less, should be able to provide added value to support the long-term monitoring of forest health and early detection of fungi-affected tree plots. Nevertheless, overall the use of defoliation as a proxy for forest health should be critically reconsidered as it comes with various practical issues, starting from potential offsets during data collection, varying leaf density due to tree age and differing effects between tree species, to just name a few.

4.7. Comparison to other studies

While most defoliation studies operate on the plot level using coarser resolution multispectral satellite data [10,75,76], there are also several recent studies using airborne or ground-based sensors at the tree level. Among these, [77,78] used ground-level methods such as airborne laser scanning (ALS) or light detection and ranging (LiDAR).

Studies focusing on tree-level defoliation mainly used ground-level methods such as ALS or LiDAR [77,78]. [77] used ordinary least squares (OLS) regression methods while [78] retrieved information from ground-level RGB photos using convolutional neural networks (CNN). However, neither of them used spatial CV for model assessment, and [78] did not perform feature selection (FS). [79] used a partial least squares (PLS) model with high-resolution digital aerial photogrammetry (DAP) to predict cumulative defoliation caused by the spruce budworm. Study results indicated that spectral features were found to be most helpful for the model. Incorporating such features (both spectral and structural) could be a possible enhancement for future works. No studies were found modeling defoliation caused by *Diplodia sapinea* (Fr.) Fuckel with remote sensing data, and most studies focused on describing the tree conditions based on local sampling [80,81].

The field of (hyperspectral) remote sensing has a strong focus on using RF for modeling in recent years [82]. However, in high-dimensional scenarios, tuning the parameter m_{try} becomes computationally expensive. To account for this and the high dimensionality in general, studies used feature selection approaches like semi-supervised feature extraction [83], wrapper methods [84–86], PCA and adjusted feature selection [87]. In general, applying feature selection methods on hyperspectral datasets has shown to be effective, regardless of the method used [88,89]. However, no studies were found that made explicit use of filter methods in combination with hyperparameter tuning in the field of (hyperspectral) remote sensing. Potential reasons for this absence could be an easier programmatic access to wrapper methods, and a higher general awareness of these compared to filter methods. Applying the filter-based feature selection methodology shown in this study and its related code provided in the research compendium might be a helpful reference for future studies using hyperspectral remote sensing data.

When looking for remote sensing studies that compare multiple models, it turned out that these often operate in a low-dimensional predictor space [90] or use wrapper methods explicitly [86].

[91,92] are more similar in their methodology but focus on a different response variable (woody cover). [91] used machine learning with ALS data to study dieback of trees in eucalyptus forests. A grid search was used for hyperparameter tuning and forward feature selection (FFS) for variable selection. [92] analyzed woody cover in South Africa using spatial CV and FS approach [93] with an RF classifier. [94] shows

a similar setup: they used hyperspectral vegetation indices, a nested CV approach for performance estimation and estimated variable importance targeting woody biomass as the response. In the results, lasso showed the best performance among the chosen methods. However, the authors did not optimize the hyperparameters of RF which makes a fair comparison problematic since the other models used perform internal optimization. The discussion section of [94] lists additional studies that made use of shrinkage models for high dimensional remote sensing modeling.

In summary, no studies could be found that used filter methods for FS or made use of NRI indices in their work and had a relation to tree health. This might relate to the fact that most environmental datasets are not high-dimensional. In fact, many studies use fewer than ten features and issues related to correlations are often solved manually instead of relying on an automated approach. These manual approaches might suffer from subjectivity and may limit the reproducibility of results.

Other fields (e.g., bioinformatics) face high-dimensional datasets more often. Hence more studies using (filter-based) feature-selection approaches can be found for this field [95,96]. Yet bioinformatics differs conceptually in many ways from environmental modeling and therefore no greater focus was put into comparing studies of this field. The availability of high/dimensional feature sets will increase in the future due to higher temporal and spectral resolutions of sensors. In addition, a high spatial resolution comes with the possibility to calculate many textural features. Hence, the ability to deal with high/dimensional datasets becomes more important and unbiased robust approaches are needed. We hope that this work and its methodology raises awareness about the application of filter methods to tackle high-dimensional problems in the environmental modeling field.

5. Conclusions

This study analyzed the effectiveness of filter-based feature selection in improving various machine-learning models of defoliation of trees in northern Spain based on hyperspectral remote-sensing data. Substantial differences in performance occurred depending on which feature-selection and machine-learning methods were combined. SVM showed the most robust behavior across all highly correlated datasets and was able to predict the response variable of this study substantially better than other methods.

Filter methods were able to improve the predictive performance on datasets in some instances, although there was no clear and systematic pattern. Their effectiveness depends on the algorithm and the dataset characteristics. Ensemble filter methods did not show a substantial improvement over individual filter methods in this study.

The addition of derived feature sets was in most cases able to improve predictive performance. In contrast, feature sets which focused on only a small fraction of the available spectral range (i.e., dataset VI) showed a worse performance than the ones which covered a wider range (400 nm - 1000 nm; HR, NRI). NRIs can be seen as a valuable addition for optimizing predictive performance in remote sensing of vegetation.

Features along the red edge wavelength region were most important for models during prediction. With respect to dedicated vegetation indices, all versions of the Vogelmann index were seen as the most important index for the best performing SVM model. This matches well with the actual purpose of these indices: These were invented to detect defoliation on sugar maple trees (*Acer saccharum* Marsh.) caused by pear thrips (*Taeniothrips inconsequens* Uzel) [97]. However, assessing the feature importance for highly correlated features remains a challenging task. Results might be biased and should be taken with care to avoid overgeneralizing from individual studies.

Finally, the potential of predicting defoliation with the given study design was rather limited with respect to the average RMSE of 28 p.p. scored by the best performing model. More training data covering a wider range of defoliation values in a larger number of forest plantations is needed to train better models which can create more robust predictions.

Author Contributions: Conceptualization, Patrick Schratz and Alexander Brenning; Data curation, Patrick Schratz and Eugenia Iturritxa; Formal analysis, Patrick Schratz, José Cortés, Bernd Bischl and Alexander Brenning; Funding acquisition, Alexander Brenning; Investigation, Patrick Schratz; Methodology, Patrick Schratz, Jannes Münchow, José Cortés and Bernd Bischl; Project administration, Alexander Brenning; Resources, Eugenia Iturritxa; Software, Patrick Schratz; Supervision, Bernd Bischl and Alexander Brenning; Validation, Patrick Schratz and Alexander Brenning; Visualization, Patrick Schratz; Writing – original draft, Patrick Schratz; Writing – review & editing, Jannes Münchow, José Cortés and Alexander Brenning.

Funding: This work was funded by the EU LIFE Healthy Forest project (LIFE14 ENV/ES/000179) and the German Scholars Organization/Carl Zeiss Foundation.

Data Availability Statement: The data presented in this study are openly available in Zenodo at <https://doi.org/10.5281/zenodo.2635403>.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

AGB	above-ground biomass
ALE	accumulated local effects
ALS	airborne laser scanning
ANN	artificial neural network
AUROC	area under the receiver operating characteristics curve
BRT	boosted regression trees
CART	classification and regression trees
CNN	convolutional neural networks
CV	cross-validation
DAP	digital aerial photogrammetry
ENM	environmental niche modeling
FFS	forward feature selection
FPR	false positive rate
FS	feature selection
GAM	generalized additive model
GBM	gradient boosting machine
GLM	generalized linear model
ICGC	Institut Cartografic i Geologic de Catalunya
IQR	interquartile range
LiDAR	light detection and ranging
LOWESS	locally weighted scatter plot smoothing
MARS	multivariate adaptive regression splines
MBO	model-based optimization
MEM	maximum entropy model
670 ML	machine learning
NDII	normalized difference infrared index
NDMI	normalized difference moisture index
NIR	near-infrared
NRI	normalized ratio index
OLS	ordinary least squares
OMNBR	optimized multiple narrow-band reflectance
PCA	principal component analysis
PDP	partial dependence plots
PISR	potential incoming solar radiation
PLS	partial least-squares
POV	proportion of variance explained
RBF	radial basis function
RF	random forest
RMSE	root mean square error
RR	ridge regression
RSS	residual sum of squares
SAR	synthetic aperture radar
SDM	species distribution modeling
SMBO	sequential-based model optimization
SVM	support vector machine
TPR	true positive rate
VI	vegetation index
XGBoost	extreme gradient boosting

671 Appendix A Appendices

672 Appendix A.1

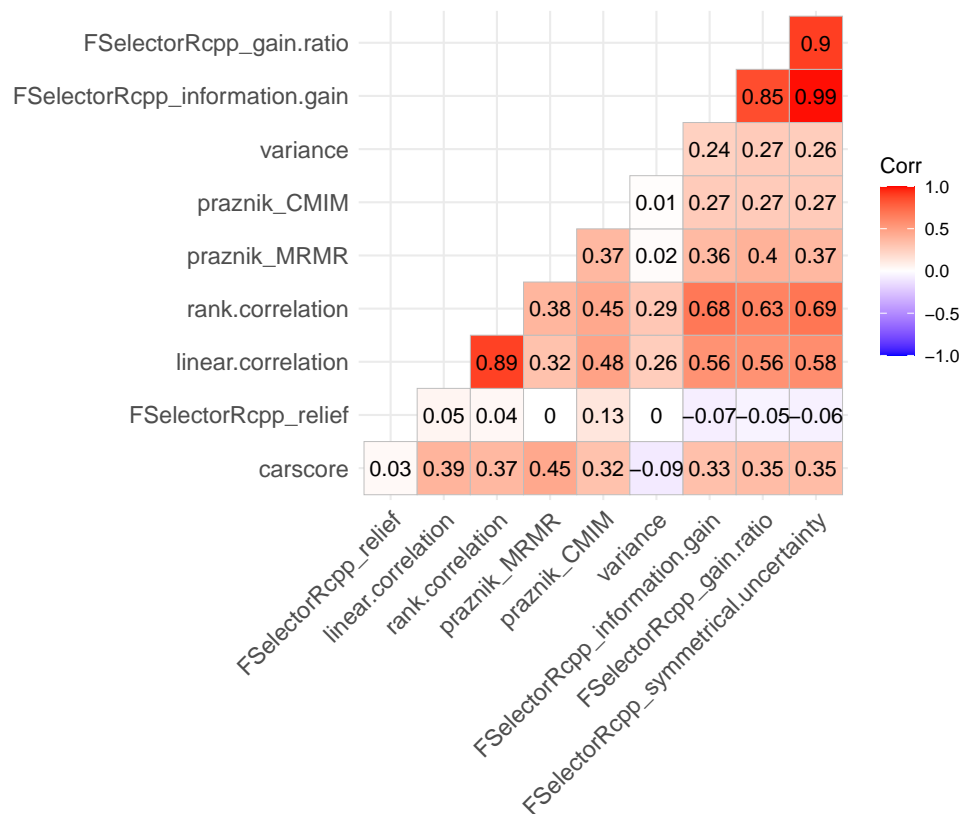


Figure A1. Spearman correlations of NRI feature rankings obtained with different filters. Filter names refer to the nomenclature used by the mlr R package. Underscores in names divide the terminology into their upstream R package and the actual filter name.

673 *Appendix A.2*

Table 8: Hyperparameter ranges and types for each model. Hyperparameter notations from the respective R packages are shown.

Model (package)	Hyperparameter	Type	Start	End	Default
RF (ranger)	x_{try}	dbl	0	0.5	-
	min.node.size	int	1	10	1
	sample.fraction	dbl	0.2	0.9	1
SVM (kernlab)	C	dbl	2^{-10}	2^{10}	1
	σ	dbl	2^{-5}	2^5	1
XGBoost (xgboost)	nrounds	int	10	70	-
	colsample_bytree	dbl	0.6	1	1
	subsample	dbl	0.6	1	1
	max_depth	int	3	15	6
	gamma	int	0.05	10	0
	eta	dbl	0.1	1	0.3
	min_child_weight	int	1	7	1

674 *Appendix A.3*Table 9: List of available vegetation indices in the *hsdar* package.

Name	Formula	Reference*
Boochs	D_{703}	[98]
Boochs2	D_{720}	[98]
CAI	$0.5 \times (R_{2000} + R_{2200}) - R_{2100}$	[99]
CARI	$a = (R_{700} - R_{550}) / 150$ $b = R_{550} - (a \times 550)$ $\frac{R_{700} \times (a \times 670 + R_{670} + b) }{R_{670} \times (a^2 + 1)^{0.5}}$	[100]
Carter	R_{695} / R_{420}	[101]
Carter2	R_{695} / R_{760}	[101]
Carter3	R_{605} / R_{760}	[101]
Carter4	R_{710} / R_{760}	[101]
Carter5	R_{695} / R_{670}	[101]
Carter6	R_{550}	[101]
CI	$R_{675} \times R_{690} / R_{683}^2$	[102]
675 CI 2	$R_{760} / R_{700} - 1$	[103]
CIInt	$\int_{600nm}^{735nm} R$	[104]
CRI1	$1 / R_{515} - 1 / R_{550}$	[103]
CRI2	$1 / R_{515} - 1 / R_{770}$	[103]
CRI3	$1 / R_{515} - 1 / R_{550} \times R_{770}$	[103]
CRI4	$1 / R_{515} - 1 / R_{700} \times R_{770}$	[103]
D1	D_{730} / D_{706}	[102]
D2	D_{705} / D_{722}	[102]
Datt	$(R_{850} - R_{710}) / (R_{850} - R_{680})$	[105]
Datt2	R_{850} / R_{710}	[105]
Datt3	D_{754} / D_{704}	[105]
Datt4	$R_{672} / (R_{550} \times R_{708})$	[106]
Datt5	R_{672} / R_{550}	[106]
Datt6	$(R_{860}) / (R_{550} \times R_{708})$	[106]
Datt7	$(R_{860} - R_{2218}) / (R_{860} - R_{1928})$	[107]

	Datt8	$(R_{860} - R_{1788}) / (R_{860} - R_{1928})$	[107]
	DD	$(R_{749} - R_{720}) - (R_{701} - R_{672})$	[108]
	DDn	$2 \times (R_{710} - R_{660} - R_{760})$	[109]
	DPI	$(D_{688} * D_{710}) / D_{697}^2$	[102]
	DWSI1	R_{80} / R_{1660}	[110]
	DWSI2	R_{1660} / R_{550}	[110]
	DWSI3	R_{1660} / R_{680}	[110]
	DWSI4	R_{550} / R_{680}	[110]
	DWSI5	$(R_{800} + R_{550}) / (R_{1660} + R_{680})$	[110]
	EGFN	$\frac{(\max(D_{650:750}) - \max(D_{500:550}))}{(\max(D_{650:750}) + \max(D_{500:550}))}$	[111]
	EGFR	$\max(D_{650:750}) / \max(D_{500:550})$	[111]
	EVI	$\frac{2.5 \times ((R_{800} - R_{670}))}{(R_{800} - (6 \times R_{670}) - (7.5 \times R_{475}) + 1)}$	[112]
	GDVI	$(R_{800}^n - R_{680}^n) / (R_{800}^n + R_{680}^n)^{**}$	[113]
	GI	R_{554} / R_{677}	[114]
	Gitelson	$1 / R_{700}$	[115]
676	Gitelson2	$(R_{750} - R_{800} / R_{695} - R_{740}) - 1$	[103]
	GMI1	R_{750} / R_{550}	[103]
	GMI2	R_{750} / R_{700}	[103]
	Green NDVI	$\frac{R_{800} - R_{550}}{R_{800} + R_{550}}$	[116]
	LWVI_1	$\frac{(R_{1094} - R_{983})}{(R_{1094} + R_{983})}$	[117]
	LWVI_2	$\frac{R_{1094} - R_{1205}}{R_{1094} + R_{1205}}$	[117]
	Maccioni	$\frac{R_{780} - R_{710}}{R_{780} - R_{680}}$	[118]
	MCARI	$((R_{700} - R_{670}) - 0.2 \times (R_{700} - R_{550})) \times (R_{700} / R_{670})$	[119]
	MCARI2	$((R_{750} - R_{705}) - 0.2 \times (R_{750} - R_{550})) \times (R_{750} / R_{705})$	[120]
	mND705	$\frac{(R_{750} - R_{705})}{R_{750} + R_{705} - 2 \times R_{445}}$	[121]
	mNDVI	$\frac{(R_{800} - R_{680})}{R_{800} + R_{680} - 2 \times R_{445}}$	[121]
	MPRI	$\frac{R_{515} - R_{530}}{R_{515} + R_{530}}$	[122]
	MSAVI	$0.5 \times ((2 \times R_{800} + 1)^2 - 8 \times (R_{800} - R_{670}))^{0.5}$	[123]
	MSI	$\frac{R_{1600}}{R_{817}}$	[124]
	mSR	$\frac{R_{800} - R_{445}}{R_{680} - R_{445}}$	[121]

mSR2	$\frac{(R_{750}/R_{705})-1}{R_{750}/R_{705}+1)^{0.5}}$	[125]
mSR705	$\frac{R_{750}-R_{445}}{R_{705}-R_{445}}$	[121]
MTCI	$\frac{R_{754}-R_{709}}{R_{709}-R_{681}}$	[126]
MTVI	$1.2 \times (1.2 \times (R_{800} - R_{550}) - 2.5 \times (R_{670} - R_{550}))$	[127]
NDLI	$\frac{\log(1/R_{1754})-\log(1/R_{1680})}{\log(1/R_{1754})+\log(1/R_{1680})}$	[128]
NDNI	$\frac{\log(1/R_{1510})-\log(1/R_{1680})}{\log(1/R_{1510})+\log(1/R_{1680})}$	[128]
NDVI	$\frac{R_{800}-R_{680}}{R_{800}+R_{680}}$	[129]
NDVI2	$\frac{R_{750}-R_{705}}{R_{750}+R_{705}}$	[130]
NDVI3	$\frac{R_{682}-R_{553}}{R_{682}+R_{553}}$	[131]
NDWI	$\frac{R_{860}-R_{1240}}{R_{860}+R_{1240}}$	[73]
NPCI	$\frac{R_{680}-R_{430}}{R_{680}+R_{430}}$	[111]
OSAVI	$\frac{(1+0.16) \times (R_{800}-R_{670})}{R_{800}+R_{670}+0.16}$	[132]
OSAVI2	$\frac{(1+0.16) \times (R_{750}-R_{705})}{R_{750}+R_{705}+0.16}$	[120]
PARS	$\frac{R_{746}}{R_{513}}$	[133]
PRI	$\frac{R_{531}-R_{570}}{R_{531}+R_{570}}$	[134]
677 PRI_norm	$\frac{PRI \times (-1)}{RDVI \times R_{700}/R_{670}}$	[135]
PRI*CI2	$PRI * CI2$	[136]
PSRI	$\frac{R_{678}-R_{500}}{R_{750}}$	[137]
PSSR	$\frac{R_{800}}{R_{635}}$	[138]
PSND	$\frac{R_{800}-R_{470}}{R_{800}-R_{470}}$	[138]
PWI	$\frac{R_{900}}{R_{970}}$	[139]
RDVI	$\frac{R_{800}-R_{670}}{\sqrt{R_{800}+R_{670}}}$	[140]
REP_LE	Red-edge position through linear extrapolation	[141]
REP_Li	$R_{re} = \frac{R_{670}+R_{780}}{2}$ $\frac{700+40 \times ((R_{re}-R_{700})}{(R_{740}-R_{700}))}$	[142]
SAVI	$\frac{(1+L) \times (R_{800}-R_{670})}{(R_{800}+R_{670}+L)}$	[143]
SIPI	$\frac{R_{800}-R_{445}}{R_{800}-R_{680}}$	[144]
SPVI	$0.4 \times 3.7 \times (R_{800} - R_{670}) - 1.2 \times ((R_{530} - R_{670})^2)^{0.5}$	[145]
SR	$\frac{R_{800}}{R_{680}}$	[146]
SR1	$\frac{R_{750}}{R_{700}}$	[147]

SR2	$\frac{R_{752}}{R_{690}}$	[147]
SR3	$\frac{R_{750}}{R_{550}}$	[147]
SR4	$\frac{R_{700}}{R_{670}}$	[148]
SR5	$\frac{R_{675}}{R_{700}}$	[133]
SR6	$\frac{R_{750}}{R_{710}}$	[149]
SR7	$\frac{R_{440}}{R_{690}}$	[150]
SR8	$\frac{R_{515}}{R_{550}}$	[151]
SRPI	$\frac{R_{430}}{R_{680}}$	[144]
SRWI	$\frac{R_{850}}{R_{1240}}$	[102]
Sum_Dr1	$\sum_{i=626}^{795} D1_i$	[152]
Sum_Dr2	$\sum_{i=680}^{780} D1_i$	[153]
SWIR FI	$\frac{R_{2133}^2}{R_{2225} \times R_{2209}^3}$	[154]
SWIR LI	$3.87 \times (R_{2210} - R_{2090}) - 27.51 \times (R_{2280} - R_{2090}) - 0.2$	[155]
SWIR SI	$-41.59 \times (R_{2210} - R_{2090}) + 1.24 \times (R_{2280} - R_{2090}) + 0.64$	[155]
SWIR VI	$37.72 \times (R_{2210} - R_{2090}) + 6.27 \times (R_{2280} - R_{2090}) + 0.57$	[155]
TCARI	$3 * ((R_{700} - R_{670}) - 0.2 \times R_{700} - R_{550}) \times (R_{700} / R_{670})$	[127]
TCARI/OSAVI	TCARI/OSAVI	[127]
TCARI2	$3 \times ((R_{750} - R_{705}) - 0.2 \times (R_{750} - R_{550}) \times (R_{750} / R_{705}))$	[120]
TCARI2/OSAVI2	TCARI2/OSAVI2	[120]
TGI	$-0.5(190(R_{670} - R_{550}) - 120(R_{670} - R_{480}))$	[156]
TVI	$0.5 \times (120 \times (R_{750} - R_{550}) - 200 \times (R_{670} - R_{550}))$	[157]
Vogelmann	$\frac{R_{740}}{R_{720}}$	[97]
Vogelmann2	$\frac{R_{734} - R_{747}}{R_{715} + R_{726}}$	[97]
Vogelmann3	$\frac{D_{715}}{D_{705}}$	[97]
Vogelmann4	$\frac{R_{734} - R_{747}}{R_{715} + R_{720}}$	[97]

679 *Appendix A.4*

680 The following information was provided by the Institut Carogràfic i Geològic de
681 Catalunya, which was in charge of image acquisition and data preprocessing.

682 The AISA EAGLE-II sensor was used for airborne image acquisition with a field of
683 view of 37.7°. Its spectral resolution is 2.4 nm and ranged from 400 nm to 1000 nm.

684 The conversion of Digital Numbers (DN) to spectral radiance was made by using
685 software designed for the instrument. Images were originally scaled in 12 bits but were
686 radiometrically calibrated to 16 bits, reserving the highest value (65535) for null values.
687 The procedure was applied to the 23 previously selected images. Last, the geometric and
688 atmospheric corrections were applied to the images.

689 The aim of this procedure was to reduce the positional errors of the images. The
690 cartographic reference system in use was EPSG 25830. Positioning was done by coupling
691 an Applanix POS AV 410 system to the sensor, integrating GPS and IMU systems. The
692 system provides geographic coordinates of the terrain and relative coordinates of the
693 aircraft (attitude) at each scanned line. Additionally a DSM from GeoEuskadi with a
694 spatial resolution of 1 m was used. The orthorectified hyperspectral images were compared
695 to orthoimages (1:5000) from GeoEuskadi. This comparison was used as the base to
696 calculate RMSE, which was below the ground sampling distance in the across and along
697 track directions.

698 The radiance measured by an instrument depends on the illumination geometry and
699 the reflective properties of the observed surface. Radiation may be absorbed or scattered
700 (Rayleigh and Mie scattering). Scattering is responsible for the adjacency effect, i.e.,
701 radiation coming from neighbors' areas to the target pixel. The MODTRAN algorithm
702 was used to model the effect of the atmosphere on the radiation. To represent the aerosols
703 of the study area, the rural model was used. In addition the optical thickness was
704 estimated on pixels with a high vegetation cover. Columnar water vapor was estimated
705 by a linear regression ratio where the spectral radiance of each pixel at the band of
706 the maximum water absorption (906 nm) is compared to its theoretical value in the
707 absence of absorption. Nonetheless, this technique is unreliable in the presence of a
708 spectral resolution as in this case. To resolve this, the water vapor parameter was selected
709 manually according to the smoothness observed on the reflectance peak at 960 nm.
710 This was combined with a mid-latitude summer atmosphere model. The output of this
711 procedure was reflectance from the target pixel scaled between 0 and 10,000.

712 The image acquisitions were originally attempted during one day (29.10.2016). Due
713 to the variable meteorological conditions, some stands had to be imaged one day later.

References

1. Lary, D.J.; Alavi, A.H.; Gandomi, A.H.; Walker, A.L. Machine Learning in Geosciences and Remote Sensing. *Geoscience Frontiers* **2016**, *7*, 3–10. doi:10.1016/j.gsf.2015.07.003.
2. Ma, Y.; Wu, H.; Wang, L.; Huang, B.; Ranjan, R.; Zomaya, A.; Jie, W. Remote Sensing Big Data Computing: Challenges and Opportunities. *Future Generation Computer Systems* **2015**, *51*, 47–60. doi:10/gdq98h.
3. Mascaro, J.; Asner, G.P.; Knapp, D.E.; Kennedy-Bowdoin, T.; Martin, R.E.; Anderson, C.; Higgins, M.; Chadwick, K.D. A Tale of Two "Forests": Random Forest Machine Learning Aids Tropical Forest Carbon Mapping. *PLOS ONE* **2014**, *9*, e85993. doi:10/f5vmrb.
4. Urban, M.; Berger, C.; Mudau, T.E.; Heckel, K.; Trukenbrodt, J.; Onyango Odipo, V.; Smit, I.P.J.; Schmulius, C. Surface Moisture and Vegetation Cover Analysis for Drought Monitoring in the Southern Kruger National Park Using Sentinel-1, Sentinel-2, and Landsat-8. *Remote Sensing* **2018**, *10*, 1482. doi:10/gf3xhw.
5. Hawryło, P.; Bednarz, B.; Wezyk, P.; Szostak, M. Estimating Defoliation of Scots Pine Stands Using Machine Learning Methods and Vegetation Indices of Sentinel-2. *European Journal of Remote Sensing* **2018**, *51*, 194–204. doi:10/gf3rbq.
6. Pollastrini, M.; Feducci, M.; Bonal, D.; Fotelli, M.; Gessler, A.; Grossiord, C.; Guyot, V.; Jactel, H.; Nguyen, D.; Radoglou, K.; Bussotti, F. Physiological Significance of Forest Tree Defoliation: Results from a Survey in a Mixed Forest in Tuscany (Central Italy). *Forest Ecology and Management* **2016**, *361*, 170–178. doi:10.1016/j.foreco.2015.11.018.
7. Gottardini, E.; Cristofolini, F.; Cristofori, A.; Pollastrini, M.; Camin, F.; Ferretti, M. A Multi-Proxy Approach Reveals Common and Species-Specific Features Associated with Tree Defoliation in Broadleaved Species. *Forest Ecology and Management* **2020**, *467*, 118151. doi:10.1016/j.foreco.2020.118151.

8. Oliva, J.; Stenlid, J.; Grönkvist-Wichmann, L.; Wahlström, K.; Jonsson, M.; Drobyshev, I.; Stenström, E. Pathogen-Induced Defoliation of *Pinus Sylvestris* Leads to Tree Decline and Death from Secondary Biotic Factors. *Forest Ecology and Management* **2016**, *379*, 273–280. doi:10.1016/j.foreco.2016.08.011.
9. Zhang, K.; Thapa, B.; Ross, M.; Gann, D. Remote Sensing of Seasonal Changes and Disturbances in Mangrove Forest: A Case Study from South Florida. *Ecosphere* **2016**, p. e01366. doi:10.1002/ecs2.1366@10.1002/(ISSN)2150-8925.ExtremeColdSpells.
10. Townsend, P.A.; Singh, A.; Foster, J.R.; Rehberg, N.J.; Kingdon, C.C.; Eshleman, K.N.; Seagle, S.W. A General Landsat Model to Predict Canopy Defoliation in Broadleaf Deciduous Forests. *Remote Sensing of Environment* **2012**, *119*, 255–265. doi:10.1016/j.rse.2011.12.023.
11. Jiang, Y.; Wang, T.; de Bie, C.A.J.M.; Skidmore, A.K.; Liu, X.; Song, S.; Zhang, L.; Wang, J.; Shao, X. Satellite-Derived Vegetation Indices Contribute Significantly to the Prediction of Epiphyllous Liverworts. *Ecological Indicators* **2014**, *38*, 72–80. doi:10.1016/j.ecolind.2013.10.024.
12. Adamczyk, J.; Osberger, A. Red-Edge Vegetation Indices for Detecting and Assessing Disturbances in Norway Spruce Dominated Mountain Forests. *International Journal of Applied Earth Observation and Geoinformation* **2015**, *37*, 90–99. doi:10.1016/j.jag.2014.10.013.
13. Thenkabail, P.S.; Lyon, J.G.; Huete, A., Eds. *Hyperspectral Indices and Image Classifications for Agriculture and Vegetation*; CRC Press, 2018. doi:10.1201/9781315159331.
14. Thenkabail, P.S.; Smith, R.B.; De Pauw, E. Hyperspectral Vegetation Indices and Their Relationships with Agricultural Crop Characteristics. *Remote sensing of Environment* **2000**, *71*, 158–182.
15. Cai, J.; Luo, J.; Wang, S.; Yang, S. Feature Selection in Machine Learning: A New Perspective. *Neurocomputing* **2018**, *300*, 70–79. doi:10.1016/j.neucom.2017.11.077.
16. Pinto, J.; Powell, S.; Peterson, R.; Rosalen, D.; Fernandes, O. Detection of Defoliation Injury in Peanut with Hyperspectral Proximal Remote Sensing. *Remote Sensing* **2020**, *12*, 3828. doi:10.3390/rs12223828.
17. Yu, R.; Ren, L.; Luo, Y. Early Detection of Pine Wilt Disease in *Pinus Tabuliformis* in North China Using a Field Portable Spectrometer and UAV-Based Hyperspectral Imagery. *Forest Ecosystems* **2021**, *8*, 44. doi:10.1186/s40663-021-00328-6.
18. Lin, H.; Yan, E.; Wang, G.; Song, R. Analysis of Hyperspectral Bands for the Health Diagnosis of Tree Species. 2014 Third International Workshop on Earth Observation and Remote Sensing Applications (EORSA), 2014, pp. 448–451. doi:10.1109/EORSA.2014.6927931.
19. Kayet, N.; Pathak, K.; Chakrabarty, A.; Singh, C.P.; Chowdary, V.M.; Kumar, S.; Sahoo, S. Forest Health Assessment for Geo-Environmental Planning and Management in Hilltop Mining Areas Using Hyperion and Landsat Data. *Ecological Indicators* **2019**, *106*, 105471. doi:10.1016/j.ecolind.2019.105471.
20. Dash, J.P.; Watt, M.S.; Pearce, G.D.; Heaphy, M.; Dungey, H.S. Assessing Very High Resolution UAV Imagery for Monitoring Forest Health during a Simulated Disease Outbreak. *ISPRS Journal of Photogrammetry and Remote Sensing* **2017**, *131*, 1–14. doi:10.1016/j.isprsjprs.2017.07.007.
21. Trunk, G.V. A Problem of Dimensionality: A Simple Example. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1979**, *PAMI-1*, 306–307. doi:10/bb2j4t.
22. Xu, H.; Caramanis, C.; Mannor, S. Statistical Optimization in High Dimensions. *Operations Research* **2016**, *64*, 958–979. doi:10/f8zbmj.
23. Mesanza, N.; Iturritxa, E.; Patten, C.L. Native Rhizobacteria as Biocontrol Agents of *Heterobasidion Annosum* s.s. and *Armillaria Mellea* Infection of *Pinus Radiata*. *Biological Control* **2016**, *101*, 8–16. doi:10.1016/j.biocontrol.2016.06.003.
24. Iturritxa, E.; Trask, T.; Mesanza, N.; Raposo, R.; Elvira-Recuenco, M.; Patten, C.L. Biocontrol of *Fusarium Circinatum* Infection of Young *Pinus Radiata* Trees. *Forests* **2017**, *8*, 32. doi:10.3390/f8020032.
25. Iturritxa, E.; Mesanza, N.; Brenning, A. Spatial Analysis of the Risk of Major Forest Diseases in Monterey Pine Plantations. *Plant Pathology* **2014**, *64*, 880–889. doi:10.1111/ppa.12328.
26. Ganley, R.J.; Watt, M.S.; Manning, L.; Iturritxa, E. A Global Climatic Risk Assessment of Pitch Canker Disease. *Canadian Journal of Forest Research* **2009**, *39*, 2246–2256. doi:10.1139/x09-131.
27. Innes, J. Methods to Estimate Forest Health. *Silva Fennica* **1993**, *27*. doi:10.14214/sf.a15668.
28. MacLean, D.A.; Lidstone, R.G. Defoliation by Spruce Budworm: Estimation by Ocular and Shoot-Count Methods and Variability among Branches, Trees, and Stands. *Canadian Journal of Forest Research* **1982**, *12*, 582–594. doi:10.1139/x82-090.
29. Johnstone Iain M.; Titterton D. Michael. Statistical Challenges of High-Dimensional Data. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2009**, *367*, 4237–4253. doi:10/cbk4qk.
30. Bommert, A.; Sun, X.; Bischl, B.; Rahnenführer, J.; Lang, M. Benchmark for Filter Methods for Feature Selection in High-Dimensional Classification Data. *Computational Statistics & Data Analysis* **2020**, *143*, 106839. doi:10.1016/j.csda.2019.106839.
31. Das, S. Filters, Wrappers and a Boosting-Based Hybrid for Feature Selection. ICML. Citeseer, 2001.
32. Guyon, I.; Elisseeff, A. An Introduction to Variable and Feature Selection. *Journal of Machine Learning Research* **2003**, *3*, 1157–1182.
33. Jolliffe, I.; Cadima, J. Principal Component Analysis: A Review and Recent Developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **2016**, *374*, 20150202. doi:10/gcsfk7.
34. Drotár, P.; Šimoňák, S.; Pietriková, E.; Chovanec, M.; Chovancová, E.; Ádám, N.; Szabó, C.; Baláz, A.; Biñas, M. Comparison of Filter Techniques for Two-Step Feature Selection. *Computing and Informatics* **2017**, *36*, 597–617. doi:10.4149/cai_2017_3_597.

35. Drotár, P.; Gazda, J.; Smékal, Z. An Experimental Comparison of Feature Selection Methods on Two-Class Biomedical Datasets. *Computers in Biology and Medicine* **2015**, *66*, 1–10. doi:10.1016/j.compbimed.2015.08.010.
36. Abeel, T.; Helleputte, T.; Van de Peer, Y.; Dupont, P.; Saeys, Y. Robust Biomarker Identification for Cancer Diagnosis with Ensemble Feature Selection Methods. *Bioinformatics* **2010**, *26*, 392–398. doi:10/dfznxv.
37. Dietterich, T.G. Ensemble Methods in Machine Learning. Proceedings of the First International Workshop on Multiple Classifier Systems. Springer-Verlag, 2000, pp. 1–15.
38. Polikar, R. Ensemble Learning. In *Ensemble Machine Learning: Methods and Applications*; Zhang, C.; Ma, Y., Eds.; Springer US: Boston, MA, 2012; pp. 1–34. doi:10.1007/978-1-4419-9326-7_1.
39. Feurer, M.; Klein, A.; Eggensperger, K.; Springenberg, J.; Blum, M.; Hutter, F. Efficient and Robust Automated Machine Learning. In *Advances in Neural Information Processing Systems 28*; Cortes, C.; Lawrence, N.D.; Lee, D.D.; Sugiyama, M.; Garnett, R., Eds.; Curran Associates, Inc., 2015; pp. 2962–2970.
40. Bolón-Canedo, V.; Alonso-Betanzos, A. Ensembles for Feature Selection: A Review and Future Trends. *Information Fusion* **2019**, *52*, 1–12. doi:10.1016/j.inffus.2018.11.008.
41. Pearson, K. LIII. On Lines and Planes of Closest Fit to Systems of Points in Space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **1901**, *2*, 559–572. doi:10/dd63n4.
42. Quinlan, J.R. Induction of Decision Trees. *Machine Learning* **1986**, *1*, 81–106. doi:10/ctd6mv.
43. Zhao, X.M. Maximum Relevance/Minimum Redundancy (MRMR). In *Encyclopedia of Systems Biology*; Dubitzky, W.; Wolkenhauer, O.; Cho, K.H.; Yokota, H., Eds.; Springer New York: New York, NY, 2013; pp. 1191–1192. doi:10.1007/978-1-4419-9863-7_432.
44. Zuber, V.; Strimmer, K. High-Dimensional Regression and Variable Selection Using CAR Scores. *Statistical Applications in Genetics and Molecular Biology* **2011**, *10*, 1–27. doi:10/c62xhp.
45. Kira, K.; Rendell, L.A. The Feature Selection Problem: Traditional Methods and a New Algorithm. Proceedings of the Tenth National Conference on Artificial Intelligence. AAAI Press, 1992, pp. 129–134.
46. Fleuret, F. Fast Binary Feature Selection with Conditional Mutual Information. *The Journal of Machine Learning Research* **2004**, *5*, 1531–1555.
47. Shannon, C.E. A Mathematical Theory of Communication. *The Bell System Technical Journal* **1948**, *27*, 379–423. doi:10.1002/j.1538-7305.1948.tb01338.x.
48. Hastie, T.; Friedman, J.; Tibshirani, R. *The Elements of Statistical Learning*; Springer New York, 2001. doi:10.1007/978-0-387-21606-5.
49. Peña, M.; Liao, R.; Brenning, A. Using Spectrotemporal Indices to Improve the Fruit-Tree Crop Classification Accuracy. *ISPRS Journal of Photogrammetry and Remote Sensing* **2017**, *128*, 158–169.
50. Bischl, B.; Richter, J.; Bossek, J.; Horn, D.; Thomas, J.; Lang, M. mlrMBO: A Modular Framework for Model-Based Optimization of Expensive Black-Box Functions. *ArXiv e-prints* **2017**, [1703.03373].
51. Binder, M.; Moosbauer, J.; Thomas, J.; Bischl, B. Multi-Objective Hyperparameter Tuning and Feature Selection Using Filter Ensembles. *arXiv:1912.12912 [cs, stat]* **2020**, [arXiv:cs, stat/1912.12912].
52. Schratz, P.; Muenchow, J.; Iturritxa, E.; Richter, J.; Brenning, A. Hyperparameter Tuning and Performance Assessment of Statistical and Machine-Learning Algorithms Using Spatial Data. *Ecological Modelling* **2019**, *406*, 109–120. doi:10.1016/j.ecolmodel.2019.06.002.
53. Hutter, F.; Hoos, H.H.; Leyton-Brown, K. Sequential Model-Based Optimization for General Algorithm Configuration. In *Lecture Notes in Computer Science*; Springer Berlin Heidelberg, 2011; pp. 507–523. doi:10.1007/978-3-642-25566-3_40.
54. Jones, D.R.; Schonlau, M.; Welch, W.J. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization* **1998**, *13*, 455–492. doi:10.1023/a:1008306431147.
55. Bergstra, J.; Bengio, Y. Random Search for Hyper-Parameter Optimization. *J. Mach. Learn. Res.* **2012**, *13*, 281–305.
56. Brenning, A. Spatial Cross-Validation and Bootstrap for the Assessment of Prediction Rules in Remote Sensing: The R Package Sperrorest. 2012 IEEE International Geoscience and Remote Sensing Symposium; IEEE: Munich, Germany, 2012. R package version 2.1.0, doi:10.1109/igarss.2012.6352393.
57. Friedman, J.H. Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics* **2001**, *29*, 1189–1232. doi:10.1214/aos/1013203451.
58. Greenwell, B.M.; Boehmke, B.C.; McCarthy, A.J. A Simple and Effective Model-Based Variable Importance Measure. *arXiv:1805.04755 [cs, stat]* **2018**, [arXiv:cs, stat/1805.04755].
59. Molnar, C. *Interpretable Machine Learning - A Guide for Making Black Box Models Explainable*; self-published: Munich, Germany, 2019.
60. Brenning, A. Transforming Feature Space to Interpret Machine Learning Models. *arXiv:2104.04295 [cs, stat]* **2021**, [arXiv:cs, stat/2104.04295].
61. Apley, D.W.; Zhu, J. Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models. *arXiv:1612.08468 [stat]* **2019**, [arXiv:stat/1612.08468].
62. R Core Team. *R: A Language and Environment for Statistical Computing*; 2019.

63. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; ACM: New York, NY, USA, 2016; KDD '16, pp. 785–794. doi:10.1145/2939672.2939785.
64. Karatzoglou, A.; Smola, A.; Hornik, K.; Zeileis, A. Kernlab – An S4 Package for Kernel Methods in R. *Journal of Statistical Software* **2004**, *11*, 1–20. doi:10.18637/jss.v011.i09.
65. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* **2010**, *33*, 1–22. doi:10.18637/jss.v033.i01.
66. Kursa, M.B. *Praznik: Collection of Information-Based Feature Selection Filters*; 2018.
67. Zawadzki, Z.; Kosinski, M. *FSelectorRcpp: 'Rcpp' Implementation of 'FSelector' Entropy-Based Feature Selection Algorithms with a Sparse Matrix Support*; 2019.
68. Bischl, B.; Lang, M.; Kotthoff, L.; Schiffner, J.; Richter, J.; Studerus, E.; Casalicchio, G.; Jones, Z.M. mlr: Machine learning in R. *Journal of Machine Learning Research* **2016**, *17*, 1–5.
69. Landau, W.M. The Drake R Package: A Pipeline Toolkit for Reproducibility and High-Performance Computing. *Journal of Open Source Software* **2018**, *3*.
70. Ghosh, M.; Adhikary, S.; Ghosh, K.K.; Sardar, A.; Begum, S.; Sarkar, R. Genetic Algorithm Based Cancerous Gene Identification from Microarray Data Using Ensemble of Filter Methods. *Medical & Biological Engineering & Computing* **2019**, *57*, 159–176. doi:10.1007/s11517-018-1874-4.
71. Horler, D.N.H.; Dockray, M.; Barber, J. The Red Edge of Plant Leaf Reflectance. *International Journal of Remote Sensing* **1983**, *4*, 273–288. doi:10.1080/01431168308948546.
72. Hais, M.; Hellebrandová, K.N.; Šrámek, V. Potential of Landsat Spectral Indices in Regard to the Detection of Forest Health Changes Due to Drought Effects. *Journal of Forest Science* **2019**, *65* (2019), 70–78. doi:10.17221/137/2018-JFS.
73. Gao, B.C. NDWI—A Normalized Difference Water Index for Remote Sensing of Vegetation Liquid Water from Space. *Remote sensing of environment* **1996**, *58*, 257–266.
74. Lehnert, L.W.; Meyer, H.; Bendix, J. *Hsdar: Manage, Analyse and Simulate Hyperspectral Data in R*, 2016.
75. de Beurs, K.M.; Townsend, P.A. Estimating the Effect of Gypsy Moth Defoliation Using MODIS. *Remote Sensing of Environment* **2008**, *112*, 3983–3990. doi:10/fpqhrc.
76. Rengarajan, R.; Schott, J.R. Modeling Forest Defoliation Using Simulated BRDF and Assessing Its Effect on Reflectance and Sensor Reaching Radiance. Remote Sensing and Modeling of Ecosystems for Sustainability XIII; International Society for Optics and Photonics: San Diego, California, United States, 2016; Vol. 9975, p. 997503. doi:10.1117/12.2235391.
77. Meng, R.; Dennison, P.E.; Zhao, F.; Shendryk, I.; Rickert, A.; Hanavan, R.P.; Cook, B.D.; Serbin, S.P. Mapping Canopy Defoliation by Herbivorous Insects at the Individual Tree Level Using Bi-Temporal Airborne Imaging Spectroscopy and LiDAR Measurements. *Remote Sensing of Environment* **2018**, *215*, 170–183. doi:10/gd67bc.
78. Kälín, U.; Lang, N.; Hug, C.; Gessler, A.; Wegner, J.D. Defoliation Estimation of Forest Trees from Ground-Level Images. *Remote Sensing of Environment* **2019**, *223*, 143–153. doi:10/gf3wzz.
79. Goodbody, T.R.H.; Coops, N.C.; Hermosilla, T.; Tompalski, P.; McCartney, G.; MacLean, D.A. Digital Aerial Photogrammetry for Assessing Cumulative Spruce Budworm Defoliation and Enhancing Forest Inventories at a Landscape-Level. *ISPRS Journal of Photogrammetry and Remote Sensing* **2018**, *142*, 1–11. doi:10.1016/j.isprsjprs.2018.05.012.
80. Hlebarska, S.; Georgieva, M. Distribution of the invasive pathogen *Diplodia sapinea* on *pinus* spp. in Bulgaria. *Conference proceedings, 90 years forest research institute - for the society and nature, Sofia, Bulgaria, 24-26 October 2018* **2018**, pp. 61–70.
81. Kaya, A.G.A.; Yeltekin, Ş.; Lehtijarvi, T.D.; Lehtijarvi, A.; Woodward, S. Severity of *Diplodia* Shoot Blight (Caused by *Diplodia* *Sapinea*) Was Greatest on *Pinus Sylvestris* and *Pinus Nigra* in a Plantation Containing Five Pine Species. *Phytopathologia Mediterranea* **2019**, *58*, 249–259.
82. Belgiu, M.; Drăguţ, L. Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS Journal of Photogrammetry and Remote Sensing* **2016**, *114*, 24–31. doi:10.1016/j.isprsjprs.2016.01.011.
83. Xia, J.; Liao, W.; Chanussot, J.; Du, P.; Song, G.; Philips, W. Improving Random Forest With Ensemble of Features and Semisupervised Feature Extraction. *IEEE Geoscience and Remote Sensing Letters* **2015**, *12*, 1471–1475. doi:10.1109/LGRS.2015.2409112.
84. Fassnacht, F.E.; Neumann, C.; Förster, M.; Buddenbaum, H.; Ghosh, A.; Clasen, A.; Joshi, P.K.; Koch, B. Comparison of Feature Reduction Algorithms for Classifying Tree Species With Hyperspectral Data on Three Central European Test Sites. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* **2014**, *7*, 2547–2561. doi:10.1109/JSTARS.2014.2329390.
85. Feng, J.; Jiao, L.; Liu, F.; Sun, T.; Zhang, X. Unsupervised Feature Selection Based on Maximum Information and Minimum Redundancy for Hyperspectral Images. *Pattern Recognition* **2016**, *51*, 295–309. doi:10.1016/j.patcog.2015.08.018.
86. Georganos, S.; Grippa, T.; Vanhuyse, S.; Lennert, M.; Shimoni, M.; Kalogirou, S.; Wolff, E. Less Is More: Optimizing Classification Performance through Feature Selection in a Very-High-Resolution Remote Sensing Object-Based Urban Application. *GIScience & Remote Sensing* **2018**, *55*, 221–242. doi:10.1080/15481603.2017.1408892.

87. Rochac, J.F.R.; Zhang, N. Feature Extraction in Hyperspectral Imaging Using Adaptive Feature Selection Approach. 2016 Eighth International Conference on Advanced Computational Intelligence (ICACI), 2016, pp. 36–40. doi:10.1109/ICACI.2016.7449799.
88. Pal, M.; Foody, G.M. Feature Selection for Classification of Hyperspectral Data by SVM. *IEEE Transactions on Geoscience and Remote Sensing* **2010**, *48*, 2297–2307. doi:10.1109/TGRS.2009.2039484.
89. Keller, S.; Braun, A.C.; Hinz, S.; Weinmann, M. Investigation of the Impact of Dimensionality Reduction and Feature Selection on the Classification of Hyperspectral EnMAP Data. 2016 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2016, pp. 1–5. doi:10.1109/WHISPERS.2016.8071759.
90. Xu, S.; Zhao, Q.; Yin, K.; Zhang, F.; Liu, D.; Yang, G. Combining Random Forest and Support Vector Machines for Object-Based Rural-Land-Cover Classification Using High Spatial Resolution Imagery. *Journal of Applied Remote Sensing* **2019**, *13*, 014521. doi:10.1117/1.JRS.13.014521.
91. Shendryk, I.; Broich, M.; Tulbure, M.G.; McGrath, A.; Keith, D.; Alexandrov, S.V. Mapping Individual Tree Health Using Full-Waveform Airborne Laser Scans and Imaging Spectroscopy: A Case Study for a Floodplain Eucalypt Forest. *Remote Sensing of Environment* **2016**, *187*, 202–217. doi:10/gf3wz4.
92. Ludwig, M.; Morgenthal, T.; Detsch, F.; Higginbottom, T.P.; Lezama Valdes, M.; Nauß, T.; Meyer, H. Machine Learning and Multi-Sensor Based Modelling of Woody Vegetation in the Molopo Area, South Africa. *Remote Sensing of Environment* **2019**, *222*, 195–203. doi:10/gf3wz8.
93. Meyer, H.; Reudenbach, C.; Hengl, T.; Katurji, M.; Nauss, T. Improving Performance of Spatio-Temporal Machine Learning Models Using Forward Feature Selection and Target-Oriented Validation. *Environmental Modelling & Software* **2018**, *101*, 1–9. doi:10.1016/j.envsoft.2017.12.001.
94. Zandler, H.; Brenning, A.; Samimi, C. Quantifying Dwarf Shrub Biomass in an Arid Environment: Comparing Empirical Methods in a High Dimensional Setting. *Remote Sensing of Environment* **2015**, *158*, 140–155. doi:10.1016/j.rse.2014.11.007.
95. Guo, Y.; Chung, F.L.; Li, G.; Zhang, L. Multi-Label Bioinformatics Data Classification With Ensemble Embedded Feature Selection. *IEEE Access* **2019**, *7*, 103863–103875. doi:10.1109/ACCESS.2019.2931035.
96. Radovic, M.; Ghalwash, M.; Filipovic, N.; Obradovic, Z. Minimum Redundancy Maximum Relevance Feature Selection Approach for Temporal Gene Expression Data. *BMC Bioinformatics* **2017**, *18*, 9. doi:10.1186/s12859-016-1423-9.
97. Vogelmann, J.E.; Rock, B.N.; Moss, D.M. Red Edge Spectral Measurements from Sugar Maple Leaves. *International Journal of Remote Sensing* **1993**, *14*, 1563–1575. doi:10.1080/01431169308953986.
98. Boochs, F.; Kupfer, G.; Dockter, K.; Kühbauch, W. Shape of the Red Edge as Vitality Indicator for Plants. *International Journal of Remote Sensing* **1990**, *11*, 1741–1753. doi:10.1080/01431169008955127.
99. Nagler, P.L.; Inoue, Y.; Glenn, E.P.; Russ, A.L.; Daughtry, C.S.T. Cellulose Absorption Index (CAI) to Quantify Mixed Soil–Plant Litter Scenes. *Remote Sensing of Environment* **2003**, *87*, 310–325. doi:10.1016/j.rse.2003.06.001.
100. Walthall, C.L.; Daughtry, C.S.T.; Chappelle, E.W.; McMurtrey, J.E.; Kim, M.S. The Use of High Spectral Resolution Bands for Estimating Absorbed Photosynthetically Active Radiation (A Par). *NASA Gov* **1994**.
101. Carter, G.A. Ratios of Leaf Reflectances in Narrow Wavebands as Indicators of Plant Stress. *International Journal of Remote Sensing* **1994**, *15*, 697–703. doi:10.1080/01431169408954109.
102. Zarco-Tejada, P.J.; Pushnik, J.C.; Dobrowski, S.; Ustin, S.L. Steady-State Chlorophyll a Fluorescence Detection from Canopy Derivative Reflectance and Double-Peak Red-Edge Effects. *Remote Sensing of Environment* **2003**, *84*, 283–294. doi:10.1016/s0034-4257(02)00113-x.
103. Gitelson, A.A.; Gritz, Y.; Merzlyak, M.N. Relationships between Leaf Chlorophyll Content and Spectral Reflectance and Algorithms for Non-Destructive Chlorophyll Assessment in Higher Plant Leaves. *Journal of Plant Physiology* **2003**, *160*, 271–282. doi:10.1078/0176-1617-00887.
104. Oppelt, N.; Mauser, W. Hyperspectral Monitoring of Physiological Parameters of Wheat during a Vegetation Period Using AVIS Data. *International Journal of Remote Sensing* **2004**, *25*, 145–159. doi:10.1080/0143116031000115300.
105. Datt, B. Visible/near Infrared Reflectance and Chlorophyll Content in Eucalyptus Leaves. *International Journal of Remote Sensing* **1999**, *20*, 2741–2759. doi:10.1080/014311699211778.
106. Datt, B. Remote Sensing of Chlorophyll a, Chlorophyll b, Chlorophyll A+B, and Total Carotenoid Content in Eucalyptus Leaves. *Remote Sensing of Environment* **1998**, *66*, 111–121. doi:10.1016/s0034-4257(98)00046-7.
107. Datt, B. Remote Sensing of Water Content in Eucalyptus Leaves. *Australian Journal of Botany* **1999**, *47*, 909. doi:10.1071/bt98042.
108. le Maire, G.; François, C.; Dufrêne, E. Towards Universal Broad Leaf Chlorophyll Indices Using PROSPECT Simulated Database and Hyperspectral Reflectance Measurements. *Remote Sensing of Environment* **2004**, *89*, 1–28. doi:10.1016/j.rse.2003.09.004.
109. Lemaire, G.; Francois, C.; Soudani, K.; Berveiller, D.; Pontailier, J.; Breda, N.; Genet, H.; Davi, H.; Dufrene, E. 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* **2008**, *112*, 3846–3864. doi:10.1016/j.rse.2008.06.005.
110. Apan, A.; Held, A.; Phinn, S.; Markley, J. Detecting Sugarcane ‘orange Rust’ Disease Using EO-1 Hyperion Hyperspectral Imagery. *International Journal of Remote Sensing* **2004**, *25*, 489–498. doi:10.1080/01431160310001618031.

111. Peñuelas, J.; Gamon, J.A.; Fredeen, A.L.; Merino, J.; Field, C.B. Reflectance Indices Associated with Physiological Changes in Nitrogen- and Water-Limited Sunflower Leaves. *Remote Sensing of Environment* **1994**, *48*, 135–146. doi:10.1016/0034-4257(94)90136-8.
112. Huete, A.R.; Liu, H.Q.; Batchily, K.; van Leeuwen, W. A Comparison of Vegetation Indices over a Global Set of TM Images for EOS-MODIS. *Remote Sensing of Environment* **1997**, *59*, 440–451. doi:10.1016/s0034-4257(96)00112-5.
113. Wu, W. The Generalized Difference Vegetation Index (GDVI) for Dryland Characterization. *Remote Sensing* **2014**, *6*, 1211–1233. doi:10.3390/rs6021211.
114. Smith, R.C.G.; Adams, J.; Stephens, D.J.; Hick, P.T. Forecasting Wheat Yield in a Mediterranean-Type Environment from the NOAA Satellite. *Australian Journal of Agricultural Research* **1995**, *46*, 113. doi:10.1071/ar9950113.
115. Gitelson, A.A.; Buschmann, C.; Lichtenthaler, H.K. The Chlorophyll Fluorescence Ratio F735/F700 as an Accurate Measure of the Chlorophyll Content in Plants. *Remote Sensing of Environment* **1999**, *69*, 296–302. doi:10.1016/s0034-4257(99)00023-1.
116. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a Green Channel in Remote Sensing of Global Vegetation from EOS-MODIS. *Remote Sensing of Environment* **1996**, *58*, 289–298. doi:10.1016/s0034-4257(96)00072-7.
117. Galvão, L.S.; Formaggio, A.R.; Tisot, D.A. Discrimination of Sugarcane Varieties in Southeastern Brazil with EO-1 Hyperion Data. *Remote Sensing of Environment* **2005**, *94*, 523–534. doi:10.1016/j.rse.2004.11.012.
118. Maccioni, A.; Agati, G.; Mazzinghi, P. New Vegetation Indices for Remote Measurement of Chlorophylls Based on Leaf Directional Reflectance Spectra. *Journal of Photochemistry and Photobiology B: Biology* **2001**, *61*, 52–61. doi:10.1016/s1011-1344(01)00145-2.
119. Daughtry, C. Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance. *Remote Sensing of Environment* **2000**, *74*, 229–239. doi:10.1016/s0034-4257(00)00113-9.
120. Wu, C.; Niu, Z.; Tang, Q.; Huang, W. Estimating Chlorophyll Content from Hyperspectral Vegetation Indices: Modeling and Validation. *Agricultural and Forest Meteorology* **2008**, *148*, 1230–1241. doi:10.1016/j.agrformet.2008.03.005.
121. Sims, D.A.; Gamon, J.A. Relationships between Leaf Pigment Content and Spectral Reflectance across a Wide Range of Species, Leaf Structures and Developmental Stages. *Remote Sensing of Environment* **2002**, *81*, 337–354. doi:10.1016/s0034-4257(02)00010-x.
122. Hernández-Clemente, R.; Navarro-Cerrillo, R.M.; Suárez, L.; Morales, F.; Zarco-Tejada, P.J. Assessing Structural Effects on PRI for Stress Detection in Conifer Forests. *Remote Sensing of Environment* **2011**, *115*, 2360–2375. doi:10.1016/j.rse.2011.04.036.
123. Qi, J.; Chehbouni, A.; Huete, A.R.; Kerr, Y.H.; Sorooshian, S. A Modified Soil Adjusted Vegetation Index. *Remote Sensing of Environment* **1994**, *48*, 119–126. doi:10.1016/0034-4257(94)90134-1.
124. Hunt, E.; Rock, B. Detection of Changes in Leaf Water Content Using Near- and Middle-Infrared Reflectances. *Remote Sensing of Environment* **1989**, *30*, 43–54. doi:10.1016/0034-4257(89)90046-1.
125. Chen, J.M. Evaluation of Vegetation Indices and a Modified Simple Ratio for Boreal Applications. *Canadian Journal of Remote Sensing* **1996**, *22*, 229–242. doi:10.1080/07038992.1996.10855178.
126. Dash, J.; Curran, P. Evaluation of the MERIS Terrestrial Chlorophyll Index (MTCI). *Advances in Space Research* **2007**, *39*, 100–104. doi:10.1016/j.asr.2006.02.034.
127. Haboudane, D.; Miller, J.R.; Tremblay, N.; Zarco-Tejada, P.J.; Dextraze, L. Integrated Narrow-Band Vegetation Indices for Prediction of Crop Chlorophyll Content for Application to Precision Agriculture. *Remote Sensing of Environment* **2002**, *81*, 416–426. doi:10.1016/s0034-4257(02)00018-4.
128. Serrano, L.; Peñuelas, J.; Ustin, S.L. Remote Sensing of Nitrogen and Lignin in Mediterranean Vegetation from AVIRIS Data. *Remote Sensing of Environment* **2002**, *81*, 355–364. doi:10.1016/s0034-4257(02)00011-1.
129. Tucker, C.J. Red and Photographic Infrared Linear Combinations for Monitoring Vegetation. *Remote Sensing of Environment* **1979**, *8*, 127–150. doi:10.1016/0034-4257(79)90013-0.
130. Gitelson, A.; Merzlyak, M.N. Quantitative Estimation of Chlorophyll-a Using Reflectance Spectra: Experiments with Autumn Chestnut and Maple Leaves. *Journal of Photochemistry and Photobiology B: Biology* **1994**, *22*, 247–252. doi:10.1016/1011-1344(93)06963-4.
131. Guanter, L.; Alonso, L.; Moreno, J. A Method for the Surface Reflectance Retrieval from PROBA/CHRIS Data over Land: Application to ESA SPARC Campaigns. *IEEE Transactions on Geoscience and Remote Sensing* **2005**, *43*, 2908–2917. doi:10.1109/tgrs.2005.857915.
132. Rondeaux, G.; Steven, M.; Baret, F. Optimization of Soil-Adjusted Vegetation Indices. *Remote Sensing of Environment* **1996**, *55*, 95–107. doi:10.1016/0034-4257(95)00186-7.
133. Chappelle, E.W.; Kim, M.S.; McMurtrey, J.E. Ratio Analysis of Reflectance Spectra (RARS): An Algorithm for the Remote Estimation of the Concentrations of Chlorophyll A, Chlorophyll B, and Carotenoids in Soybean Leaves. *Remote Sensing of Environment* **1992**, *39*, 239–247. doi:10.1016/0034-4257(92)90089-3.
134. Gamon, J.; Peñuelas, J.; Field, C. A Narrow-Waveband Spectral Index That Tracks Diurnal Changes in Photosynthetic Efficiency. *Remote Sensing of Environment* **1992**, *41*, 35–44. doi:10.1016/0034-4257(92)90059-s.
135. Zarco-Tejada, P.J.; González-Dugo, V.; Williams, L.E.; Suárez, L.; Berni, J.A.J.; Goldhamer, D.; Fereres, E. A PRI-Based Water Stress Index Combining Structural and Chlorophyll Effects: Assessment Using Diurnal Narrow-Band Airborne Imagery and the CWSI Thermal Index. *Remote Sensing of Environment* **2013**, *138*, 38–50. doi:10.1016/j.rse.2013.07.024.

136. Garrity, S.R.; Eitel, J.U.; Vierling, L.A. Disentangling the Relationships between Plant Pigments and the Photochemical Reflectance Index Reveals a New Approach for Remote Estimation of Carotenoid Content. *Remote Sensing of Environment* **2011**, *115*, 628–635. doi:10.1016/j.rse.2010.10.007.
137. Merzlyak, M.N.; Gitelson, A.A.; Chivkunova, O.B.; Rakitin, V.Y. Non-Destructive Optical Detection of Pigment Changes during Leaf Senescence and Fruit Ripening. *Physiologia Plantarum* **1999**, *106*, 135–141. doi:10.1034/j.1399-3054.1999.106119.x.
138. Blackburn, G.A. Quantifying Chlorophylls and Carotenoids at Leaf and Canopy Scales. *Remote Sensing of Environment* **1998**, *66*, 273–285. doi:10.1016/s0034-4257(98)00059-5.
139. Penuelas, J.; Pinol, J.; Ogaya, R.; Filella, I. Estimation of Plant Water Concentration by the Reflectance Water Index WI (R900/R970). *International Journal of Remote Sensing* **1997**, *18*, 2869–2875. doi:10.1080/014311697217396.
140. Roujean, J.L.; Breon, F.M. Estimating PAR Absorbed by Vegetation from Bidirectional Reflectance Measurements. *Remote Sensing of Environment* **1995**, *51*, 375–384. doi:10.1016/0034-4257(94)00114-3.
141. Cho, M.A.; Skidmore, A.K. A New Technique for Extracting the Red Edge Position from Hyperspectral Data: The Linear Extrapolation Method. *Remote Sensing of Environment* **2006**, *101*, 181–193. doi:10.1016/j.rse.2005.12.011.
142. Guyot, G.; Baret, F. Utilisation de La Haute Resolution Spectrale Pour Suivre l'etat Des Couverts Vegetaux. Spectral Signatures of Objects in Remote Sensing; Guyenne, T.D.; Hunt, J.J., Eds., 1988, Vol. 287, *ESA Special Publication*, p. 279.
143. Huete, A. A Soil-Adjusted Vegetation Index (SAVI). *Remote Sensing of Environment* **1988**, *25*, 295–309. doi:10.1016/0034-4257(88)90106-x.
144. Penuelas, J.; Filella, I.; Lloret, P.; Munoz, F.; Vilaleliu, M. Reflectance Assessment of Mite Effects on Apple Trees. *International Journal of Remote Sensing* **1995**, *16*, 2727–2733. doi:10.1080/01431169508954588.
145. Vincini, M.; Frazzi, E.; D'Alessio, P. Angular Dependence of Maize and Sugar Beet VIs from Directional CHRIS/Proba Data. Proc. 4th ESA CHRIS PROBA Workshop, 2006, Vol. 2006, pp. 19–21.
146. Jordan, C.F. Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology* **1969**, *50*, 663–666. doi:10.2307/1936256.
147. Gitelson, A.A.; Merzlyak, M.N. Remote Estimation of Chlorophyll Content in Higher Plant Leaves. *International Journal of Remote Sensing* **1997**, *18*, 2691–2697. doi:10.1080/014311697217558.
148. McMurtrey, J.; Chappelle, E.; Kim, M.; Meisinger, J.; Corp, L. Distinguishing Nitrogen Fertilization Levels in Field Corn (*Zea Mays* L.) with Actively Induced Fluorescence and Passive Reflectance Measurements. *Remote Sensing of Environment* **1994**, *47*, 36–44. doi:10.1016/0034-4257(94)90125-2.
149. Zarco-Tejada, P.J.; Miller, J.R. Land Cover Mapping at BOREAS Using Red Edge Spectral Parameters from CASI Imagery. *Journal of Geophysical Research: Atmospheres* **1999**, *104*, 27921–27933. doi:10.1029/1999jd900161.
150. Lichtenthaler, H.; Lang, M.; Sowinska, M.; Heisel, F.; Miehe, J. Detection of Vegetation Stress via a New High Resolution Fluorescence Imaging System. *Journal of Plant Physiology* **1996**, *148*, 599–612. doi:10.1016/s0176-1617(96)80081-2.
151. Hernández-Clemente, R.; Navarro-Cerrillo, R.M.; Zarco-Tejada, P.J. Carotenoid Content Estimation in a Heterogeneous Conifer Forest Using Narrow-Band Indices and PROSPECTDART Simulations. *Remote Sensing of Environment* **2012**, *127*, 298–315. doi:10.1016/j.rse.2012.09.014.
152. Elvidge, C.D.; Chen, Z. Comparison of Broad-Band and Narrow-Band Red and near-Infrared Vegetation Indices. *Remote Sensing of Environment* **1995**, *54*, 38–48. doi:10.1016/0034-4257(95)00132-k.
153. Filella, I.; Penuelas, J. The Red Edge Position and Shape as Indicators of Plant Chlorophyll Content, Biomass and Hydric Status. *International Journal of Remote Sensing* **1994**, *15*, 1459–1470. doi:10.1080/01431169408954177.
154. Levin, N.; Kidron, G.J.; Ben-Dor, E. Surface Properties of Stabilizing Coastal Dunes: Combining Spectral and Field Analyses. *Sedimentology* **2007**, *54*, 771–788. doi:10.1111/j.1365-3091.2007.00859.x.
155. Lobell, D.B.; Asner, G.P.; Law, B.E.; Treuhaft, R.N. Subpixel Canopy Cover Estimation of Coniferous Forests in Oregon Using SWIR Imaging Spectrometry. *Journal of Geophysical Research: Atmospheres* **2001**, *106*, 5151–5160. doi:10.1029/2000jd900739.
156. Hunt, E.R.; Doraiswamy, P.C.; McMurtrey, J.E.; Daughtry, C.S.; Perry, E.M.; Akhmedov, B. A Visible Band Index for Remote Sensing Leaf Chlorophyll Content at the Canopy Scale. *International Journal of Applied Earth Observation and Geoinformation* **2013**, *21*, 103–112. doi:10.1016/j.jag.2012.07.020.
157. Broge, N.; Leblanc, E. 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* **2001**, *76*, 156–172. doi:10.1016/s0034-4257(00)00197-8.

Short Biography of Authors



Patrick Schratz is a Data Scientist working in Zurich, Switzerland. His area of expertise is applied machine learning, more specifically the field of environmental modeling. He is a Ph.D. candidate in geographic information science in the Department of Geography at Friedrich Schiller University Jena where he conducts research in environmental modeling. At cynkra GmbH in Zurich, Patrick is an R consultant with many years of experience in the following fields: CI/CD, package development, DevOps tasks, machine learning and spatiotemporal data handling.



Jannes Muenchow is a GIScientist working in tropical ecology since 2007 with a special interest in ENSO, biodiversity, species distribution modeling and predictive mapping. He received his Ph.D. from Friedrich-Alexander-Universität Erlangen-Nürnberg (Germany) in 2013. He joined the business location department of a large consulting company as a geo-data scientist for more than two years until the prediction of a strong El Niño event brought him back to academia from 2016 to 2019 (Friedrich Schiller University Jena). His research focuses on developing open-source tools for ecology, geomorphology and qualitative GIS. He is a co-author of the book “Geocomputation with R”.



Eugenia Iturritxa received the Ecology and Ph.D. degree in Plant Protection from the University of the Basque Country in 2001. Since 1999 her main research has focused on forest health, on the study of diverse species of native and introduced pathogenic fungi in forests and forest plantations. Her research the NEIKER research center includes the distribution of diseases, analysis of predisposing factors for them, genetic and phenotypic studies of populations and their epidemiology.



José Cortés received his B.S. in Mathematics and M.S. in Statistics from Arizona State University, Arizona, USA, in 2014 and 2016, respectively. Currently he is a Ph.D. student at Friedrich Schiller University Jena, Germany. He is a member of the International Max Planck Research School on Global Biogeochemical Cycles (IMPRS-gBGC), a joint program with the Max Planck Institute for Biogeochemistry. His research focuses on spatiotemporal trend detection in environmental data.



Bernd Bischl obtained his Ph.D. from TU Dortmund University in 2013. He is a professor of Statistical Learning and Data Science in the Department of Statistics at Ludwig-Maximilians-Universität in Munich, Germany, and a director of the Munich Center for Machine Learning. His research interests include AutoML, model selection, interpretable ML and XAI.



Alexander Brenning (Ph.D. 2005) graduated in mathematics at Technical University of Freiberg, Germany, and received his Ph.D. in geography from Humboldt-Universität zu Berlin. He served as an assistant professor and tenured associate professor in geomatics at the University of Waterloo, Ontario, Canada from 2007 until 2015, when he was appointed as a full professor in geographic information science at Friedrich Schiller University Jena, Germany. Dr. Brenning's research interests include statistical and machine-learning techniques for environmental modeling and remote sensing. He has also contributed to open-source geocomputation by developing R packages for spatial cross-validation and GIS coupling.