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

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

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

***Abstract*—This study analyzed highly-correlated, feature-rich datasets from hyperspectral remote sensing data using multiple machine and statistical-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 (%) was modeled as a function of reflectance, and variable importance was assessed using permutation-based feature importance. Overall support vector machine (SVM) outperformed others such as random forest (RF), extreme gradient boosting (XGBoost), lasso (L1) and ridge (L2) regression 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 the unfiltered feature sets, while ensemble filters did not have a substantial impact on performance. Permutation-based feature importance estimated features around the red edge to be most important for the models. However, the presence of features in the near-infrared region (800 nm - 1000 nm) was essential to achieve the best performances. More training data and replication in similar benchmarking studies is needed for more generalizable conclusions. 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.**

***Index Terms*—hyperspectral imagery, forest health monitoring, machine learning, feature selection, feature effects, model comparison**

I. Introduction

THE use of machine learning (ML) algorithms for analyzing remote sensing data has seen a huge increase in the last decade [1]. Naturally, 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 extend the benchmark comparison matrix of studies in a semi-automated way, possibly stumbling across unexpected findings of process settings that would never have been explored otherwise [2].

Machine-learning 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 of predicting into unknown space qualifies these tools as a promising toolset for such tasks. One aspect of this research field is to enhance the understanding of biotic and abiotic stress triggers, for example by analyzing tree defoliation [5].

Other approaches for analyzing forest health include change detection [6] or describing the current health status of forests on a stand level [7]. In such studies, the defoliation of trees serves as a proxy for forest health by describing the impact of biotic and abiotic pest triggers [7], [8].

Vegetation indices have shown the potential to provide valuable information when analyzing forest health [9], [10]. Most vegetation indices were developed with the aim of being sensitive to changes of specific wavelength regions, serving as a proxy for underlying plant processes. However, often enough indices developed for different purposes than the one to be analyzed can help to explain complex relationships. 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. A less known index type which 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 but instead makes use of a data-driven feature engineering approach by combining (all possible) combinations of spectral bands. Especially when working with hyperspectral data, thousands of NRI features can be derived this way.

Despite their popularity in environmental modeling, there are no studies so far which used machine-learning algorithms in combination with remote sensing data to analyze defoliation at the tree level. This study aims to close this gap by analyzing tree defoliation in northern

P.Schratz, J.Muenchow, J.Cortés and A.Brenning are with the Department of Geography, GIScience group, Friedrich-Schiller-University of Jena, Germany.

B.Bischl is head of the computational statistics group at the Department of Statistics, Ludwig-Maximilian-University Munich.

E.Iturritxa is with NEIKER Tecnalia, Vitoria-Gasteiz, Arab, Spain.

Spain using airborne hyperspectral data. The methodology of this study uses machine-learning methods in combination with feature selection and hyperparameter tuning. In addition, feature importance and feature effects are evaluated. Incorporating the idea of creating data-driven NRIs, this study also discusses the practical problems of high dimensionality in environmental modeling [11], [12].

Even though ML algorithms are capable of handling highly correlated input variables, model fitting becomes computationally more demanding, and model interpretation more complex. Feature selection approaches can help to address this issue, reducing possible noise in the feature space, simplify model interpretability and possibly enhance predictive performance [13].

This study shows how high-dimensional datasets can be handled effectively with machine-learning methods while still being able to interpret 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. However, this study clearly focuses on investigating the effects of filter methods and feature set types on predictive performance rather than on interpreting feature effects.

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

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

II. 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 of invasive 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 *Dothisthroma septosporum (Dorogin) M. Morelet* causing a spread of cankers or defoliation [14], [15]. The last two fungi are mainly responsible for the foliation loss of the trees analyzed in this study [16]. In-situ measurements of defoliation of trees (serving as a proxy for tree health) were collected to serve as the response variable *defoliation* which ranges from 0 - 100 (in %) (Figure 1).

It is assumed that the fungi infect the trees through open wounds, possibly caused by previous hail damage [16]. The dieback of these trees, which are mainly used as timber, causes high economic damages [17].

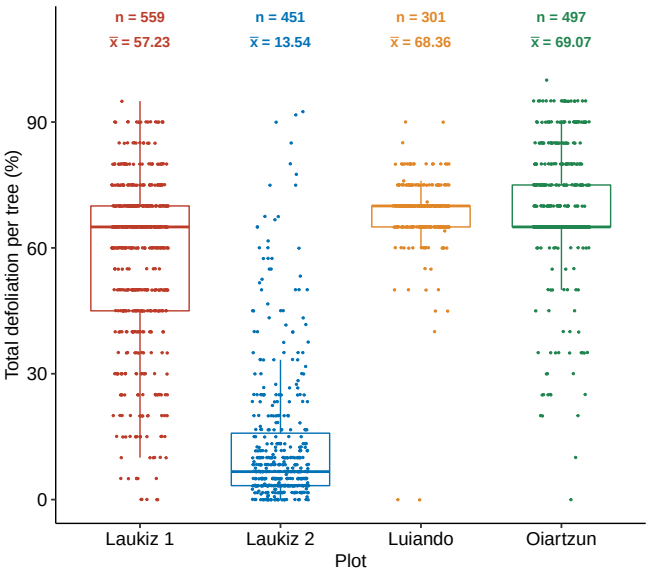

Fig. 1. Response variable *defoliation* at trees for plots Laukiz1, Laukiz2, Luiando and Oiartzun. n corresponds to the total number of trees in the plot, $\bar{x}$ refers to the mean defoliation.

A. 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 most observations (n = 559) while Laukiz2 shows the largest area size (1.44 ha). All plots besides Luiando are located within 100 km from the coast (Figure 2). In total 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 via visual inspection using 5% intervals at three height levels (bottom, mid, top) with the help of a dedicated score card. Estimating the human observer error when assessing defoliation is an issue which is being discussed since many years [18]. Even though no estimation error was recorded in this study, [19] estimated human observer errors when assessing defoliation to range between 7% - 18%.

B. Hyperspectral data

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

III. Methods

A. Derivation of indices

To use the full potential of the hyperspectral data, all possible vegetation indices supported by the R package

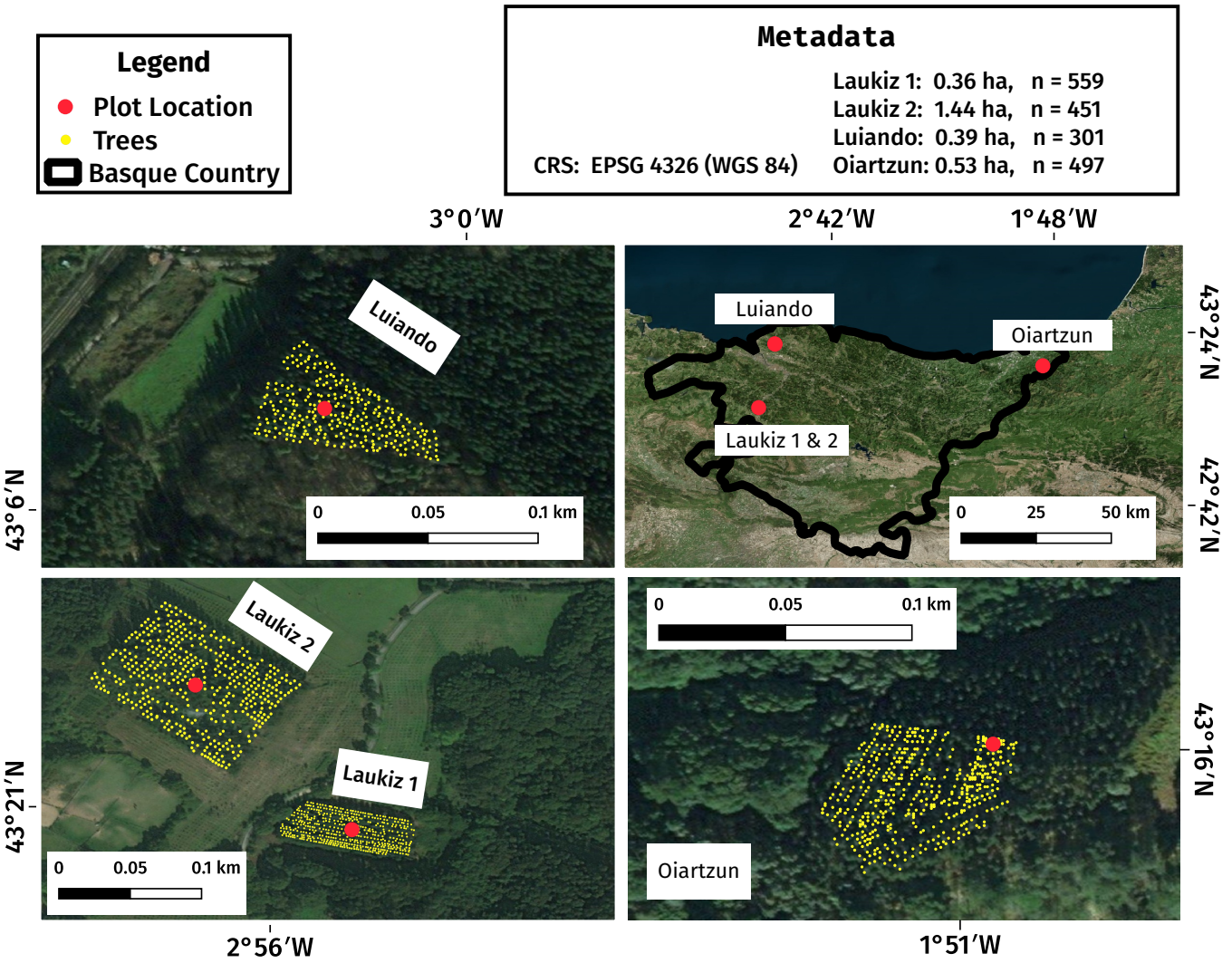

Fig. 2. Information about location, size and spatial distribution of trees for all plots used in this study.

TABLE I
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

hsdar (89 in total) as well as all possible NRIs were calculated.

NRIs follow the optimized multiple narrow-band reflectance (OMNBR) concept of data-driven information extraction from narrow-band indices of hyperspectral data [20], [21]. 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 family of normalized ratio indices (NRIs), which are implemented in the hsdar R package, we used the following 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 with up to one meter from ICGC, a buffer of one meter 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. All of those pixels formed the final reflectance value of a single tree and were used as the base information to derive all additional feature sets. In total, $\frac{121*122}{2} = 7471$ NRIs were calculated.

B. 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 [22].
- Model interpretation and prediction become more challenging [22].

To reduce the feature space of a dataset, conceptually differing approaches exist: wrapper methods, filters, penalization methods (lasso and ridge) or principal component analysis (PCA) [23]–[26]. In contrast to wrapper methods, filters can be added to the hyperparameter optimization step and have a lower computational footprint. Due to the focus on filter methods in this manuscript, only this subgroup of feature selection methods will be introduced in greater detail in the following subsections.

1) Filter methods: The concept of filters originates from the idea of ranking features following a score calculated by an algorithm [25]. Some filter methods can only deal with specific types of variables (numeric or nominal). Filters only rank features, they do not decide which covariates to drop or keep [27]. The selection which features to keep for model fitting is usually done within the optimization phase of the model fitting, along with the hyperparameter tuning. Essentially, the number of covariates in the model is treated as a additional hyperparameter of the model. The goal is to optimize the number of ranked features to the point at which the model achieves the best performance.

Besides the concept of choosing a specific filter method to rank variables, studies showed that combining several filters using statistical operations such as 'minimum' or 'mean' can enhance the predictive performance of the resulting models, especially when applied to multiple datasets [28], [29]. This approach is referred to as 'ensemble filtering' [30]. Ensemble filters align with the recent rise of the 'ensemble' approach in machine learning which uses the idea of stacking to combine the predictions of multiple models, aiming to enhance predictive performance [31]–[33]. In this work the 'Borda' ensemble filter was applied [29]. Its final feature order is determined by the sum of all single filters ranks.

Filter methods can be grouped into groups which are formed out of three binary classes: 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 final ranking result 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 into the ensemble filter, selected at random. This ensured that the ensemble filter composition was not biased towards a certain group of filter methods.

2) Description of used filter methods: Filter methods can be classified as follows (Table II):

- 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' 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) or the feature (attribute or X), respectively. $H(X)$ is Shannon's Entropy [40] for a variable X 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 (with b commonly being set to 2).

In order to use this method with a numeric response (percentage defoliation of trees), the variable was discretized into equal bins and treated as a class variable. After feature rank correlations of > 0.9 between different bin sizes were observed in a side analysis, $n_{bin} = 10$ was found to be a reasonable setting to go with.

C. Benchmarking design

1) Algorithms: The following learners were used in this work:

- Extreme Gradient Boosting (XGBoost)
- Random Forest (RF)
- Penalized Regression (both L1 (lasso) and L2 (ridge))
- Support Vector Machine (SVM, RBF Kernel)

Random forest and SVM are well established algorithms widely used in (environmental) remote sensing. Extreme Gradient Boosting (commonly abbreviated as XGBoost) has shown promising results in benchmark studies in recent years. Penalized regression is a statistical modeling technique capable of dealing with highly-correlated covariates by penalizing the coefficients of the model [41]. Common penalties are 'lasso' (L1) and 'ridge' (L2). Ridge does not remove variables from the model (penalization to zero) but just shrinks them to effectively zero, keeping them in the model.

TABLE II
LIST OF FILTER METHODS USED IN THIS WORK

Name	Group	Ref.
Linear correlation (Pearson)	univariate, linear, correlation	[34]
Information gain	univariate, non-linear, entropy	[35]
Minimum redundancy, maximum relevance	multivariate, non-linear, entropy	[36]
Carscore	multivariate, linear, correlation	[37]
Relief	multivariate, linear, entropy	[38]
Conditional minimal information maximization	multivariate, linear, entropy	[39]

In total the benchmarking grid consisted of 156 experiments (6 feature sets × 3 ML algorithms × 8 feature-selection methods and for the L1/L2 models, 6 feature sets × 2 models. The selected hyperparameter settings are shown in Table VIII. In addition the code which generated the matrix is available in the research compendium of this study (10.5281/zenodo.2635403).

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. [42] 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 constant makes it possible to draw conclusions on their individual impact. However, rather than only looking at these three groups also combinations of such were taken into account:

- HR + VI
- HR + NRI
- HR + VI + NRI

Even though the feature-selection step should be solely left to the filter methods in this study, it was ensured a priori to account for features with a pairwise correlation of 1. Having such features within the data can cause undesired effects during model fitting and feature importance calculation. Hence, after having calculated all pair-wise correlations between features, for pairs which exceeded the threshold of $1 - 10^{-10}$, the feature with the largest mean absolute correlation across all variables was removed from the dataset.

This preprocessing step reduced the number of covariates to 122 (HR), 86 (VI) and 7467 (NRI).

3) Hyperparameter Optimization: Hyperparameters were tuned using model-based optimization (MBO) within a nested spatial cross-validation (CV) [43]–[45]. In MBO first n randomly chosen hyperparameter settings out of a user defined search space are composed. After these n settings have been evaluated, one new setting, which is going to be evaluated next, is proposed by a fitted surrogate model (by default a kriging method). This strategy continues until a termination criterion, defined by the user, is reached [46], [47].

In this work, an initial design of 30 randomly composed hyperparameter settings in combination with a termination criterion of 70 iterations was used, resulting in a total budget of 100 evaluated hyperparameter settings per fold. The advantage of this tuning approach is a substantial reduction of the tuning budget that is required to find a setting close to the global optimization minimum. MBO may outperform methods that do not use information from previous iterations, such as random search or grid search [48]. Tuning ranges are shown in Table VIII.

To optimize the number of features used for model fitting, the percentage of features was added as a hyperparameter during the optimization stage [44]. For PCA, the number of principal components was tuned instead. The RF hyperparameter m_{try} was re-expressed as $m_{\text{try}} = p_{\text{sel}}^t$ as a function of the number of selected features, p_{sel}. It was thus tuned on a logarithmic scale by varying t between 0 (i.e. $m_{\text{try}} = 1$) and 0.5 (i.e. $m_{\text{try}} = \sqrt{p_{\text{sel}}}$). This was necessary to ensure that m_{try} was not chosen higher than the available number of features left after optimizing the feature percentage during tuning.

4) Spatial resampling: A spatial nested cross-validation on the plot level was chosen to reduce the influence of spatial autocorrelation as much as possible [45], [49]. The root mean square error (RMSE) was chosen as the error measure. Each plot served as one fold within the cross-validation setting, resulting in four iterations in total. For the inner level (hyperparameter tuning), $k - 1$ folds were used with k being the number of plots.

D. Feature importance and feature effects

Estimating feature importance for datasets with highly correlated features is a complicated task for which many different approaches, model-specific and agnostic, exist [41], [50], [51]. The correlation between covariates makes it challenging to calculate an unbiased estimate for single features [52]. Methods like partial dependence plots (PDP) or permutation-based approaches may produce unreliable estimates in such scenarios because unrealistic situations between covariates are created [52]. The development of robust methods which enable an unbiased estimation of feature importance for highly correlated variables are subject to current research.

In this work permutation-based feature importance was calculated to estimate feature importance / effects [53]. With the limitations in mind when applied to correlated features, the aim was to get a general overview of the feature importance of the hyperspectral bands while trying to avoid an over-interpretation of results. The best-performing algorithm on the HR task (i.e. SVM) was used for the feature importance calculation.

E. Linking feature importance to wavelength regions

For environmental interpretation purposes the ten most important indices of the best performing models of feature sets HR and VI were linked to the spectral regions of the hyperspectral data. 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.

F. Research compendium

All tasks of this study were conducted using the open-source statistical programming language R [54]. A complete list of all R packages used in this study can be found in linked repositories. Due to space limitations only the selected packages with high impact on this work will be explicitly cited.

The algorithm implementations of the following packages have been used: xgboost [55] (*Extreme Gradient Boosting*), kernlab [56] (Support Vector Machine) and glmnet [57] (penalized regression). The filter implementations of the following packages have been used: praznik [58], FSelectorRcpp [59]. Package mlr [60] was used for all modeling related steps. drake [61] was used for structuring the work and reproducibility. This study is available as a research compendium on Zenodo (10.5281/zenodo.2635403). Besides the availability of code and manuscript sources, a static webpage is available at (https://github.com/pat-s/2019-feature-selection), listing more side-analyses that were carried out during the creation of this study.

IV. Results

A. Correlation analysis

The PCA-based correlation analysis showed that, depending on the feature set, between two (HR) and 42 (NRI) principal component (PC) are needed to explain 95% of variance. The fewer PCs are needed to reach a high proportion of variance explained the more similar the individual features are to each other. Hence, a high correlation can be observed for feature set HR which reaches 96% with only two PCs. Feature sets NRI and VI are more diverse and require 42 (NRI) and 12 (VI) PCs to explain 95% of the variance in the data, respectively.

B. Predictive performance

Overall, the response variable "tree defoliation" could be modeled with an RMSE of 28 percentage points (p.p.). SVM showed no differences in RMSE across feature sets whereas other learners (RF, SVM, XGBoost, lasso and ridge) differed up to five percentage points (Figure 3). SVM showed the best overall performance with a mean difference of around three percentage points to the next best model (XGBoost) (Table V). 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 VI).

The combination of feature sets showed small increases in performance for some learners. RF and 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 both RF and XGBoost were scored on HR only. Datasets containing derived features only (VI, NRI) showed no improvement in performance compared to the raw hyperspectral band information (HR) or combined feature sets.

SVM combined with the "Relief" filter achieved the best overall performance (RMSE of 28.09 p.p.) (Table III). Regression with ridge (L2) and lasso (L1) penalty showed their best performances on the NRI feature set (Table V). Difference to other feature sets were very small for lasso (below one p.p.) and a bit higher for ridge (between two and five p.p. (VI)). XGBoost shows bad performances for some feature sets and fills the ten last places of the ranking (Table IV). 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). Using filters for RF showed a substantial increase in performance for all tasks with the exception of NRI, for which the difference among all filters was also the smallest (Figure 4). XGBoost showed a high dependency on filtering the data: In two out of six tasks using no filter resulted in the worst or second worst performance. XGBoost shows the highest overall differences between filters for a single task: for feature set HR, the range is up to 15 percentage points ("CMIM" vs. "no filter") (Figure 4).

When comparing the usage of filters against using no filter at all, there no instance where a model without filtering 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 filter achieved only in one instance (RF on HR-NRI-VI) the best performance among the used filters (Figure 5). For RF and XGBoost it most often ranked within the first 50% with respect to all filters of a specific task. For XGBoost on the NRI task, the Borda filter scored the second worst performance.

Large differences were observed between the numbers of features selected during tuning for the subsequent fitting process across each learner and plot. While for RF least features were selected when Luiando or Oiartzun were the test set ($n = 1$), more than 90% of all features were used for test sets Laukiz and Laukiz2 (Table VII). RF and XGBoost used very few features only when Oiartzun was the test set (one and three, respectively) whereas this was the plot for which SVM used most features ($n = 119$). Overall, SVM used the least features across all plots among all learners, with 24% for test set Oiartzun being the highest single plot value. In general every learner behaved quite differently for each plot and no overall pattern could be observed.

C. Variable importance

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, B69) and 1.93 p.p (VI, Vogelmann2) (Figure 6). For dataset HR most features cluster around the infrared region (920 nm - 1000 nm) (six out of ten) while for VI eight out of ten

TABLE III
BEST TEN RESULTS AMONG ALL COMBINATIONS, SORTED IN ASCENDING ORDER OF RMSE

	Task	Model	Filter	RMSE	SE
1	HR-NRI-VI	SVM	Relief	28.09	18.95
2	HR	SVM	Borda	28.12	19.12
3	HR	SVM	Car	28.12	19.12
4	HR	SVM	CMIM	28.12	19.12
5	HR	SVM	Info Gain	28.12	19.12
6	HR	SVM	MRMR	28.12	19.12
7	HR	SVM	Relief	28.12	19.12
8	HR-NRI-VI	SVM	CMIM	28.12	19.18
9	HR-NRI-VI	SVM	Car	28.12	19.12
10	NRI-VI	SVM	CMIM	28.12	19.12

TABLE IV
WORST TEN RESULTS AMONG ALL COMBINATIONS, SORTED IN DECREASING ORDER OF RMSE

	Task	Model	Filter	RMSE	SE
1	NRI-VI	XGBoost	PCA	47.19	9.17
2	HR-NRI-VI	XGBoost	PCA	46.30	5.79
3	NRI	XGBoost	PCA	45.96	8.28
4	VI	XGBoost	PCA	45.73	7.54
5	VI	XGBoost	No Filter	45.08	6.95
6	HR	XGBoost	No Filter	45.08	5.26
7	HR-NRI-VI	XGBoost	Relief	44.76	5.54
8	HR-NRI	XGBoost	PCA	43.14	7.21
9	HR-NRI	XGBoost	MRMR	42.84	6.73
10	HR-NRI	XGBoost	Pearson	42.44	6.45

TABLE V
BEST PERFORMANCE OF EACH LEARNER ACROSS ANY TASK AND FILTER METHOD, SORTED ASCENDING BY RMSE

	Task	Model	Filter	RMSE	SE
1	HR-NRI-VI	SVM	Relief	28.09	18.95
2	HR	XGBoost	CMIM	30.88	14.47
3	NRI	Lasso-MBO	No Filter	31.16	15.03
4	NRI	Ridge-MBO	No Filter	31.16	15.03
5	HR	RF	Relief	32.93	13.17

TABLE VI
TEST FOLD PERFORMANCES FOR LEARNER SVM ON THE HR DATASET WITHOUT USING A FILTER. 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.

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

TABLE VII
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 ASCENDING BY RMSE.

Learner	Test Plot	Features (%)	#	RMSE
RF Relief	Laukiz1	0.99973	559	41.73
	Laukiz2	0.93371	422	17.12
	Luiando	0.00216	1	33.21
	Oiartzun	0.00004	1	37.20
XGB CMIM	Laukiz1	0.30849	173	33.36
	Laukiz2	0.98122	443	45.59
	Luiando	0.83613	252	41.01
	Oiartzun	0.00515	3	15.24
SVM Relief	Laukiz1	0.00062	1	35.49
	Laukiz2	0.04106	19	31.03
	Luiando	0.20735	63	37.19
	Oiartzun	0.23928	119	14.89

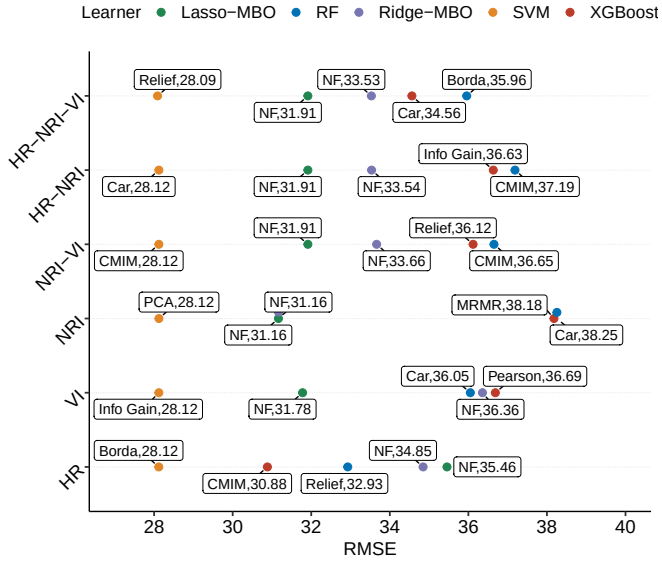

Fig. 3. Predictive performance (RMSE) of models across tasks. Suffix 'MBO' means that model-based optimization was used for hyperparameter optimization. Abbreviations on the vertical axis refer to the combinations of feature sets on which each model was scored on. Labels represent the feature selection method (e.g. NF = no filter, Car = 'Carscore', Info Gain = 'Information Gain', Borda = 'Borda'). The second value of each label shows the RMSE value of the respective setting.

concentrate on the wavelength range of 700 nm - 750 nm (the so called "red edge"). For feature set HR, four features in the infrared region (920 nm - 1000 nm) were identified by the model to be most important (causing a mean decrease in RMSE of around 1 percentage point). Overall, most features (excluding the top five respectively) showed only a small importance with average decreases in RMSE below 0.5 p.p..

V. DISCUSSION

A. Predictive Performance

The best aggregated performance of this study (SVM + "Relief" filter, RMSE 28.09 p.p.) has to be seen in the light of model overfitting (see subsection V-B). Leaving out the performance on Laukiz2 when aggregating results, the mean RMSE would be around 19 percentage points. 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 can be postulated 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 to the landscape scale using models fitted on this dataset due

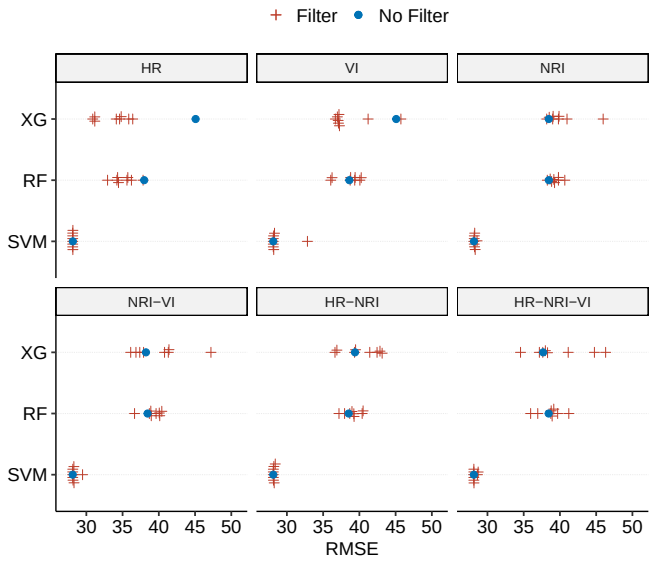

Fig. 4. Model performances in RMSE when using no filter method compared to all other filters across all tasks.

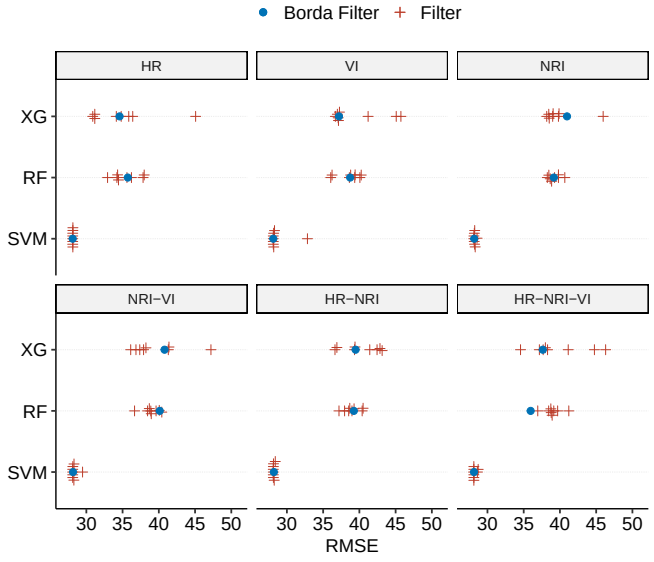

Fig. 5. Predictive performances in RMSE when using the Borda filter method compared to all other filters for each learner across all tasks.

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 model performed reasonably well predicting defoliation greater than 50% but failed for lower levels. This applied to all learners of this study. The overall performance achieved in this study can be classified as "fair".

In addition, data quality issues (subsection V-E) might have an influence on model performances. These include the timing of the acquisition of the hyperspectral data (late phenological phase), field measurement errors when surveying defoliation, the influence of background reflectance (e.g. soil reflectance) and the possible positional offset of measured GPS coordinates of trees.

1) Model differences: An interesting finding is the strength of the SVM algorithm when comparing its predictive performance to its competitors (Table V). These cluster around a performance of 31 p.p while SVM is able to score 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.).

Penalized methods showed promising performances, especially when taking runtime into account. When removing features with a correlation of nearly 1, lasso and ridge are able to score performances around 31 p.p..

A potential limiting factor in this study could be the upper limit of 70 iterations used for the XGBoost algorithm (hyperparameter nrounds), especially for feature sets including NRIs (Table VIII). 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.

2) Feature set differences: One objective of this study was whether expert-based or data-driven feature engineering has a positive influence on model performance. With respect to Figure 3, no overall positive or negative trend was found for all models that related to specific feature sets. The performance of RF and XGBoost on the VI feature set was around four to six percentage points lower than on others. One reason could be the lack of coverage in the wavelength area 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).

B. Performance vs. plot characteristics

The large differences in RMSE obtained on different test folds can be attributed to model overfitting (Table VI). 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).

The large differences of selected features per fold during tuning give interesting insights into internals of the used models (Table VII). While in most cases, SVM and XGBoost require a substantial portion of all available

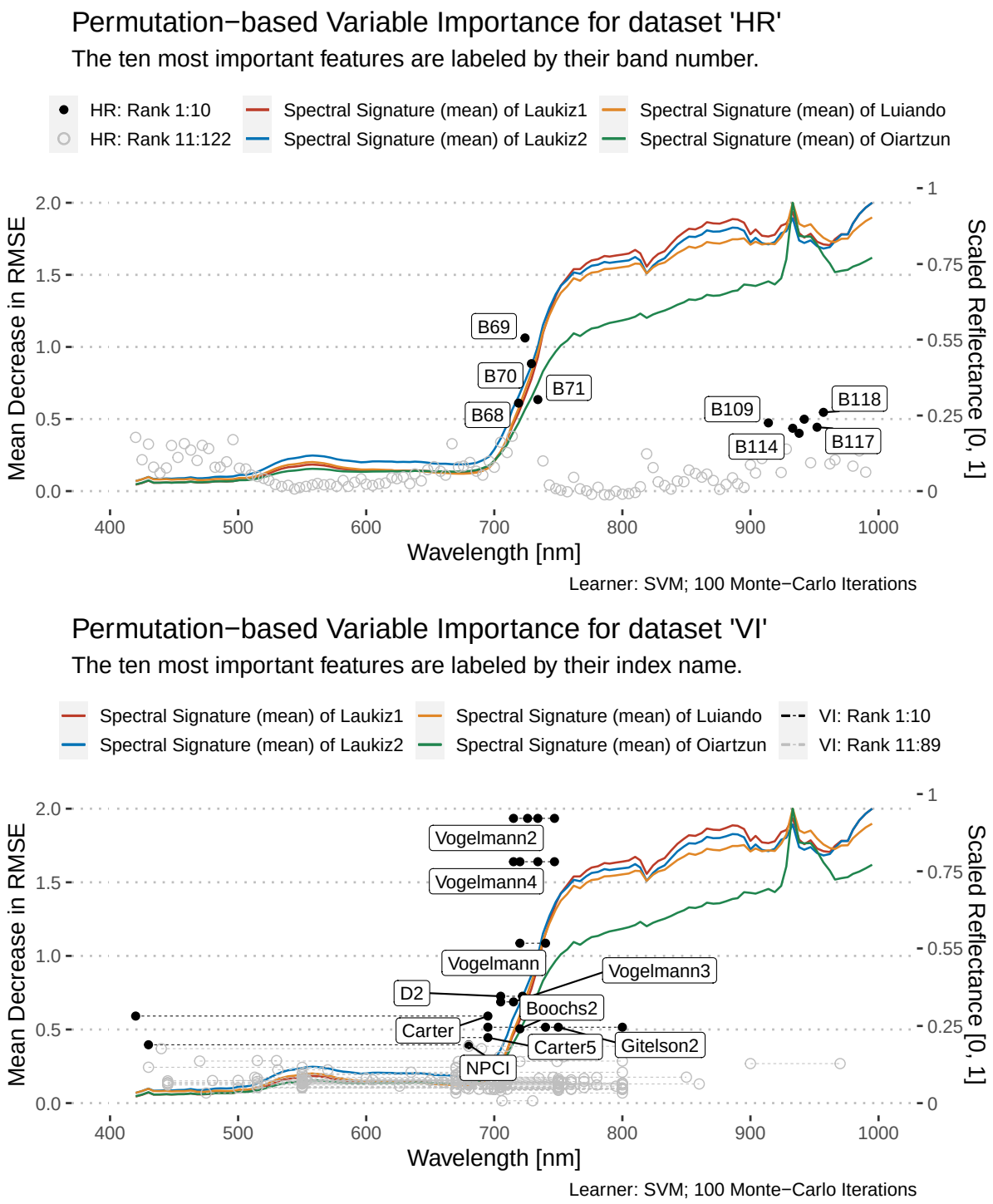

Fig. 6. Variable importance for feature sets HR and VI: Mean decrease in RMSE for one-hundred feature permutations using the SVM learner. The wavelength range on the x-axis matches the range of the hyperspectral sensor (400 nm - 1000 nm). For each dataset, the ten most important features were highlighted as black dots and labeled by name. Grey 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.

features to achieve robust predictions, RF was able to achieve the best results with a relatively small amount of features. Realizing early that few features are needed during tuning to reach adequate performances can reduce the overall computational runtime substantially, especially when iterating over parameters such as m_{try} whose op-

timum (and range) depends on the number of features. Hence, regardless of the potential advantage of using filters for increased predictive performance, it should be noted that these can have a strong positive effect on runtime, at least for RF in this study.

Ultimately, the results of Table VII 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.

C. 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). No pattern with respect to lower scores related to a specific filter method could be found. Hence, it is recommended to test multiple filters in a study if filters are going to be used. While filters can improve the performance of models, they might be more interesting in other aspects than performance: reducing variables can reduce computational efforts in high-dimensional scenarios and might 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.

The models which used the Borda ensemble method in this study did not score 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 [29]. Hence, their main advantage are 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 (besides RF on task HR-NRI-VI) a single filter outperformed the ensemble filter. However, studies which used ensemble filters are still rare and usually these are not compared against single filters [62]. In summary, Borda performs no better than a randomly selected filter method in this study. More case studies applying ensemble filter methods are needed to verify this finding. Nevertheless, ensemble filters can be a promising addition to a machine-learning feature-selection portfolio.

PCA, acting as a filter method, more often showed less optimal results, especially for algorithms RF and XGBoost. Especially XGBoost had substantial problems when using PCA as a filter method and scored the worst four results (Table IV). 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.

D. 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 und thus the health status of vegetation and its respective counterpart [63]. 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 area - especially for plot Oiartzun - which might explain why these features were considered important by the model.

A possible explanation for the worse 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.

E. Data Quality

Environmental datasets always come with some constraints that can have potential influence on the modeling process and its outcome. The following paragraph discusses these.

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 which would actually refer to a different, possibly non-tree, pixel was reasonably high. It was concluded that using a buffer of one meter can be a good compromise between the inclusion of information from too many surrounding trees, mapping a single tree crown accordingly accounting for a possible geometric offset. The applied buffer only included a pixel value if the distance to the centroid of a pixel was smaller than the buffer radius (i.e. ¡= 1m). This results in almost 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. Besides using a buffer to extract the hyperspectral information, a 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 the focus on filter-based feature selection methods.

Another point worth discussing is that the exact number of contributing pixels to the final index value of an observation cannot be determined precisely: it depends on the location of the tree within the pixel grid. According to the extract function of the raster package, a pixel is included if its centroid (and not just any part of the grid cell) falls inside the buffer. As the buffer is circular, the total number of contributing pixels of each tree depends on the exact location of a tree within the pixel grid. If a tree observation is located on the border of the plot, some directions of the buffer will contain no values (because the image coverage was cropped to the borders of the plot) and the subsequent index value will be calculated with fewer pixels than if the tree observation is located in the middle of the plot. However, in most cases each tree information should be composed out of four hyperspectral pixels.

The available hyperspectral data covered a wavelength between 400 nm and 1000 nm. Hence, the wavelength range of the shortwave infrared (SWIR) region is not covered in this study. Given that this wavelength range is often used in forest health studies [64], e.g. when calculating the normalized difference moisture index (NDMI) index [65], this marks a clear limitation of the dataset at hand.

The dataset consists of in-situ data collected within September 2016 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 matching hyperspectral data of these additional timestamps.

The R package hsdar was used for the calculation of vegetation indices [66]. All indices that could be calculated with the given spectral range of the data (400 nm - 1000 nm) were used. This means even though section D lists all available indices of the package, not all listed indices were used in this study. Even though 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.

F. Comparison to other studies

While most defoliation studies operate on the plot level using coarser-resolution multispectral satellite data [7], [67], [68], there are also several recent studies using airborne or ground-based sensors at the tree level. Among these, [69], [70] used ground-level methods such as airborne laser scanning (ALS) or light detection and ranging (LiDAR).

Studies focusing on tree-level defoliation used ground-level methods such as ALS or LiDAR [69], [70]. [69] used ordinary least squares (OLS) regression methods while [70] retrieved information from ground-level RGB photos using convolutional neural networks (CNN). However, both did not use spatial CV and [70] no feature selection (FS). [8] 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 (both spectral and structural) could be a possible enhancement for future works.

The field of (hyperspectral) remote sensing has a strong focus on using random forests for modeling in recent years [71]. However, in high-dimensional scenarios, tuning 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 [72], wrapper methods [73]–[75], PCA and adjusted feature selection [76]. However, no study that made use of filter methods in combination with hyperparameter tuning in the field of (hyperspectral) remote sensing could be found. Potential reasons for this gap could be an easier access to wrapper methods and a higher general awareness of such 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 [77] or use wrapper methods explicitly [75].

[78], [79] are more similar in their methodology but focus on a different response variable (woody cover). [78] used machine learning with ALS data to study dieback of trees for eucalyptus forests. A grid search was used for hyperparameter tuning and forward feature selection (FFS) for variable selection. [79] analyzed woody cover in South Africa using spatial CV and FS approach [80] with a random forest classifier. [81] 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 [81] lists additional studies that made use of shrinkage models for high dimensional remote sensing modeling.

In summary, no studies which used filter methods for FS or made use of NRI indices in their work and had a relation to tree health were found. 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. This can be subjective 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 [82], [83]. 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.

VI. Conclusion

This study analyzed defoliation of trees in northern Spain by using hyperspectral data as input for machine-learning models which used hyperparameter tuning and filter-based feature selection. 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 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*) [84]. 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 percentage points 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.

VII. Acknowledgments

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

Appendix A
Correlation among filter methods

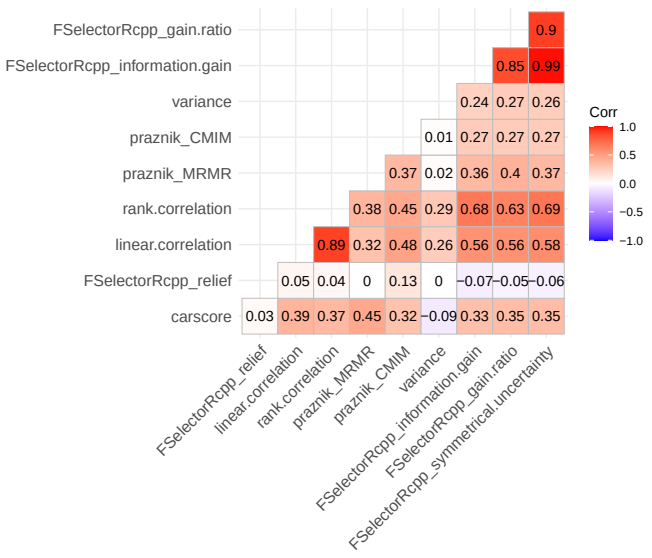

Fig. 7. Spearman correlations of NRI feature rankings obtained with different filters.

Appendix B
Effect of different n_{bins} values on filter 'information gain'

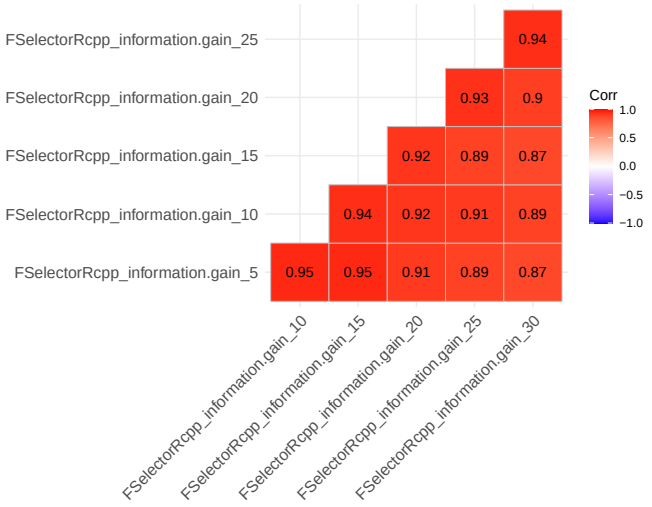

Fig. 8. Spearman correlations of rankings obtained with the information gain filter using different n_{bins} values for discretization of the numeric response.

Appendix C
Hyperparameter tuning ranges

TABLE VIII
Hyperparameter ranges and types for each model. Hyperparameter notations from the respective R packages were used.

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

Appendix D
Vegetation index formulas

Name	**Formula**	**Reference***
Boochs	D_{703}	[85]
Boochs2	D_{720}	[85]
CAI	$0.5 \times (R_{2000} + R_{2200}) - R_{2100}$	[86]
CARI	$a = (R_{700} - R_{550})/150$ $b = R_{550} - (a \times 550)$ $\frac{R_{700} \times \lvert(a \times 670 + R_{670} + b)}{R_{670} \times (a^2+1)\rvert^{0.5}}$	[87]
Carter	R_{695}/R_{420}	[88]
Carter2	R_{695}/R_{760}	[88]
Carter3	R_{605}/R_{760}	[88]
Carter4	R_{710}/R_{760}	[88]
Carter5	R_{695}/R_{670}	[88]
Carter6	R_{550}	[88]
CI	$R_{675} \times R_{690}/R_{683}^2$	[89]
CI2	$R_{760}/R_{700} - 1$	[90]
ClAInt	$\int_{600nm}^{735nm} R$	[91]
CRI1	$1/R_{515} - 1/R_{550}$	[90]
CRI2	$1/R_{515} - 1/R_{770}$	[90]
CRI3	$1/R_{515} - 1/R_{550} \times R_{770}$	[90]
CRI4	$1/R_{515} - 1/R_{700} \times R_{770}$	[90]
D1	D_{730}/D_{706}	[89]
D2	D_{705}/D_{722}	[89]
Datt	$(R_{850} - R_{710})/(R_{850} - R_{680})$	[92]
Datt2	R_{850}/R_{710}	[92]
Datt3	D_{754}/D_{704}	[92]
Datt4	$R_{672}/(R_{550} \times R_{708})$	[93]
Datt5	R_{672}/R_{550}	[93]
Datt6	$(R_{860})/(R_{550} \times R_{708})$	[93]
Datt7	$(R_{860} - R_{2218})/(R_{860} - R_{1928})$	[94]
Datt8	$(R_{860} - R_{1788})/(R_{860} - R_{1928})$	[94]
DD	$(R_{749} - R_{720}) - (R_{701} - R_{672})$	[95]
DDn	$2 \times (R_{710} - R_{660} - R_{760})$	[96]
DPI	$(D_{688} * D_{710})/D_{697}^2$	[89]
DWSI1	R_{80}/R_{1660}	[97]
DWSI2	R_{1660}/R_{550}	[97]
DWSI3	R_{1660}/R_{680}	[97]
DWSI4	R_{550}/R_{680}	[97]
DWSI5	$(R_{800} + R_{550})/(R_{1660} + R_{680})$	[97]
EGFN	$\frac{(\max(D_{650:750}) - \max(D_{500:550}))}{(\max(D_{650:750}) + \max(D_{500:550}))}$	[98]

EGFR	$\max(D_{650:750})/\max(D_{500:550})$	[98]
EVI	$\frac{2.5\times((R_{800}-R_{670})}{(R_{800}-(6\times R_{670})-(7.5\times R_{475})+1)}$	[99]
GDVI	$(R^n_{800}-R^n_{680})/(R^n_{800}+R^n_{680})$**	[100]
GI	R_{554}/R_{677}	[101]
Gitelson	$1/R_{700}$	[102]
Gitelson2	$(R_{750}-R_{800}/R_{695}-R_{740})-1$	[90]
GMI1	R_{750}/R_{550}	[90]
GMI2	R_{750}/R_{700}	[90]
Green NDVI	$\frac{R_{800}-R_{550}}{R_{800}+R_{550}}$	[103]
LWVI_1	$\frac{(R_{1094}-R_{983})}{(R_{1094}+R_{983})}$	[104]
LWVI_2	$\frac{R_{1094}-R_{1205}}{R_{1094}+R_{1205}}$	[104]
Maccioni	$\frac{R_{780}-R_{710})}{R_{780}-R_{680}}$	[105]
MCARI	$((R_{700}-R_{670})-0.2\times(R_{700}-R_{550}))\times(R_{700}/R_{670})$	[106]
MCARI2	$((R_{750}-R_{705})-0.2\times(R_{750}-R_{550}))\times(R_{750}/R_{705})$	[107]
mND705	$\frac{(R_{750}-R_{705})}{R_{750}+R_{705}-2\times R_{445}}$	[108]
mNDVI	$\frac{(R_{800}-R_{680})}{R_{800}+R_{680}-2\times R_{445}}$	[108]
MPRI	$\frac{R_{515}-R_{530}}{R_{515}+R_{530}}$	[109]
MSAVI	$0.5\times((2\times R_{800}+1)^2-8\times(R_{800}-R_{670}))^{0.5}$	[110]
MSI	$\frac{R_{1600}}{R_{817}}$	[111]
mSR	$\frac{R_{800}-R_{445}}{R_{680}-R_{445}}$	[108]
mSR2	$\frac{(R_{750}/R_{705})-1}{R_{750}/R_{705}+1)^{0.5}}$	[112]
mSR705	$\frac{R_{750}-R_{445}}{R_{705}-R_{445}}$	[108]
MTCI	$\frac{R_{754}-R_{709}}{R_{709}-R_{681}}$	[113]
MTVI	$1.2\times(1.2\times(R_{800}-R_{550})-2.5\times(R_{670}-R_{550}))$	[114]
NDLI	$\frac{log(1/R_{1754})-log(1/R_{1680})}{log(1/R_{1754})+log(1/R_{1680})}$	[115]
NDNI	$\frac{log(1/R_{1510})-log(1/R_{1680})}{log(1/R_{1510})+log(1/R_{1680})}$	[115]
NDVI	$\frac{R_{800}-R_{680}}{R_{800}+R_{680}}$	[116]
NDVI2	$\frac{R_{750}-R_{705}}{R_{750}+R_{705}}$	[117]
NDVI3	$\frac{R_{682}-R_{553}}{R_{682}+R_{553}}$	[118]
NDWI	$\frac{R_{860}-R_{1240}}{R_{860}+R_{1240}}$	[65]
NPCI	$\frac{R_{680}-R_{430}}{R_{680}+R_{430}}$	[98]
OSAVI	$\frac{(1+0.16)\times(R_{800}-R_{670})}{R_{800}+R_{670}+0.16}$	[119]
OSAVI2	$\frac{(1+0.16)\times(R_{750}-R_{705})}{R_{750}+R_{705}+0.16)}$	[107]
PARS	$\frac{R_{746}}{R_{513}}$	[120]
PRI	$\frac{R_{531}-R_{570}}{R_{531}+R_{570}}$	[121]
PRI_norm	$\frac{PRI\times(-1)}{RDVI\times R_{700}/R_{670}}$	[122]
PRI*CI2	$PRI*CI2$	[123]
PSRI	$\frac{R_{678}-R_{500}}{R_{750}}$	[124]
PSSR	$\frac{R_{800}}{R_{635}}$	[125]
PSND	$\frac{R_{800}-R_{470}}{R_{800}-R_{470})}$	[125]
PWI	$\frac{R_{900}}{R_{970}}$	[126]
RDVI	$\frac{R_{800}-R_{670}}{\sqrt{R_{800}+R_{670}}}$	[127]
REP_LE	Red-edge position through linear extrapolation	[128]
REP_Li	$R_{re}=\frac{R_{670}+R_{780})}{\frac{700+40\times((R_{re}-R_{700})}{(R_{740}-R_{700}))}}$	[129]
SAVI	$\frac{(1+L)\times(R_{800}-R_{670})}{(R_{800}+R_{670}+L)}$	[130]
SIPI	$\frac{R_{800}-R_{445}}{R_{800}-R_{680}}$	[131]
SPVI	$0.4\times3.7\times(R_{800}-R_{670})-1.2\times((R_{530}-R_{670})^2)^{0.5}$	[132]
SR	$\frac{R_{800}}{R_{680}}$	[133]
SR1	$\frac{R_{750}}{R_{700}}$	[134]
SR2	$\frac{R_{752}}{R_{690}}$	[134]
SR3	$\frac{R_{750}}{R_{550}}$	[134]
SR4	$\frac{R_{700}}{R_{670}}$	[135]
SR5	$\frac{R_{675}}{R_{700}}$	[120]
SR6	$\frac{R_{750}}{R_{710}}$	[136]
SR7	$\frac{R_{440}}{R_{690}}$	[137]
SR8	$\frac{R_{515}}{R_{550}}$	[138]
SRPI	$\frac{R_{430}}{R_{680}}$	[131]
SRWI	$\frac{R_{850}}{R_{1240}}$	[89]
Sum_Dr1	$\sum_{i=626}^{795} D1_i$	[139]
Sum_Dr2	$\sum_{i=680}^{780} D1_i$	[140]
SWIR FI	$\frac{R^2_{2133}}{R_{2225}\times R^3_{2209}}$	[141]
TCARI2/OSAVI2	TCARI2/OSAVI2	[107]
TGI	$-0.5(190(R_{670}-R_{550})-120(R_{670}-R_{480}))$	[143]
TVI	$0.5\times(120\times(R_{750}-R_{550})-200\times(R_{670}-R_{550}))$	[144]
Vogelmann	$\frac{R_{740}}{R_{720}}$	[84]
Vogelmann2	$\frac{R_{734}-R_{747}}{R_{715}+R_{726})}$	[84]
Vogelmann3	$\frac{D_{715}}{D_{705}}$	[84]
Vogelmann4	$\frac{R_{734}-R_{747}}{R_{715}+R_{720}}$	[84]

Appendix E
Radiometric, Geometric and Atmospheric correction of hyperspectral data

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

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

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

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

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

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

References

[1] D. J. Lary, A. H. Alavi, A. H. Gandomi, and A. L. Walker, "Machine learning in geosciences and remote sensing," *Geoscience Frontiers*, vol. 7, no. 1, pp. 3–10, Jan. 2016.

[2] Y. Ma, H. Wu, L. Wang, B. Huang, R. Ranjan, A. Zomaya, and W. Jie, "Remote sensing big data computing: Challenges and opportunities," *Future Generation Computer Systems*, vol. 51, pp. 47–60, Oct. 2015.

[3] J. Mascaro, G. P. Asner, D. E. Knapp, T. Kennedy-Bowdoin, R. E. Martin, C. Anderson, M. Higgins, and K. D. Chadwick, "A Tale of Two "Forests": Random Forest Machine Learning Aids Tropical Forest Carbon Mapping," *PLOS ONE*, vol. 9, no. 1, p. e85993, Jan. 2014.

[4] M. Urban, C. Berger, T. E. Mudau, K. Heckel, J. Truckenbrodt, V. Onyango Odipo, I. P. J. Smit, and C. Schmullius, "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*, vol. 10, no. 9, p. 1482, Sep. 2018.

[5] P. Hawryło, B. Bednarz, P. Wezyk, and M. Szostak, "Estimating defoliation of Scots pine stands using machine learning methods and vegetation indices of Sentinel-2," *European Journal of Remote Sensing*, vol. 51, no. 1, pp. 194–204, Jan. 2018.

[6] K. Zhang, B. Thapa, M. Ross, and D. Gann, "Remote sensing of seasonal changes and disturbances in mangrove forest: A case study from South Florida," *Ecosphere*, p. e01366, 2016.

[7] P. A. Townsend, A. Singh, J. R. Foster, N. J. Rehberg, C. C. Kingdon, K. N. Eshleman, and S. W. Seagle, "A general Landsat model to predict canopy defoliation in broadleaf deciduous forests," *Remote Sensing of Environment*, vol. 119, pp. 255–265, Apr. 2012.

[8] T. R. H. Goodbody, N. C. Coops, T. Hermosilla, P. Tompalski, G. McCartney, and D. A. MacLean, "Digital aerial photogrammetry for assessing cumulative spruce budworm defoliation and enhancing forest inventories at a landscape-level," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 142, pp. 1–11, Aug. 2018.

[9] Y. Jiang, T. Wang, C. A. J. M. de Bie, A. K. Skidmore, X. Liu, S. Song, L. Zhang, J. Wang, and X. Shao, "Satellite-derived vegetation indices contribute significantly to the prediction of epiphyllous liverworts," *Ecological Indicators*, vol. 38, pp. 72–80, Mar. 2014.

[10] J. Adamczyk and A. Osberger, "Red-edge vegetation indices for detecting and assessing disturbances in Norway spruce dominated mountain forests," *International Journal of Applied Earth Observation and Geoinformation*, vol. 37, pp. 90–99, May 2015.

[11] G. V. Trunk, "A Problem of Dimensionality: A Simple Example," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-1, no. 3, pp. 306–307, Jul. 1979.

[12] H. Xu, C. Caramanis, and S. Mannor, "Statistical Optimization in High Dimensions," *Operations Research*, vol. 64, no. 4, pp. 958–979, Jul. 2016.

[13] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70–79, Jul. 2018.

[14] N. Mesanza, E. Iturritxa, and C. L. Patten, "Native rhizobacteria as biocontrol agents of Heterobasidion annosum s.s. and Armillaria mellea infection of Pinus radiata," *Biological Control*, vol. 101, pp. 8–16, Oct. 2016.

[15] E. Iturritxa, T. Trask, N. Mesanza, R. Raposo, M. Elvira-Recuenco, and C. L. Patten, "Biocontrol of Fusarium circinatum infection of young Pinus radiata trees," *Forests*, vol. 8, no. 2, p. 32, Jan. 2017.

[16] E. Iturritxa, N. Mesanza, and A. Brenning, "Spatial analysis of the risk of major forest diseases in Monterey pine plantations," *Plant Pathology*, vol. 64, no. 4, pp. 880–889, 2014.

[17] R. J. Ganley, M. S. Watt, L. Manning, and E. Iturritxa, "A global climatic risk assessment of pitch canker disease," *Canadian Journal of Forest Research*, vol. 39, no. 11, pp. 2246–2256, Nov. 2009.

[18] J. Innes, "Methods to estimate forest health." *Silva Fennica*, vol. 27, no. 2, 1993.

[19] D. A. MacLean and R. G. Lidstone, "Defoliation by spruce budworm: Estimation by ocular and shoot-count methods and variability among branches, trees, and stands," *Canadian Journal of Forest Research*, vol. 12, no. 3, pp. 582–594, Sep. 1982.

[20] P. S. Thenkabail, R. B. Smith, and E. De Pauw, "Hyperspectral vegetation indices and their relationships with agricultural crop characteristics," *Remote sensing of Environment*, vol. 71, no. 2, pp. 158–182, 2000.

[21] P. S. Thenkabail, J. G. Lyon, and A. Huete, Eds., *Hyperspectral Indices and Image Classifications for Agriculture and Vegetation*. CRC Press, Dec. 2018.

[22] Johnstone Iain M. and Titterington D. Michael, "Statistical challenges of high-dimensional data," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 367, no. 1906, pp. 4237–4253, Nov. 2009.

[23] A. Bommert, X. Sun, B. Bischl, J. Rahnenführer, and M. Lang, "Benchmark for filter methods for feature selection in high-dimensional classification data," *Computational Statistics & Data Analysis*, vol. 143, p. 106839, Mar. 2020.

[24] S. Das, "Filters, Wrappers and a Boosting-Based Hybrid for Feature Selection," in *ICML*, 2001.

[25] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of Machine Learning Research*, vol. 3, no. Mar, pp. 1157–1182, 2003.

[26] I. Jolliffe and J. Cadima, "Principal component analysis: A review and recent developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, no. 2065, p. 20150202, Apr. 2016.

[27] P. Drotár, J. Gazda, and Z. Smékal, "An experimental comparison of feature selection methods on two-class biomedical datasets," *Computers in Biology and Medicine*, vol. 66, pp. 1–10, Nov. 2015.

[28] T. Abeel, T. Helleputte, Y. Van de Peer, P. Dupont, and Y. Saeys, "Robust biomarker identification for cancer diagnosis with ensemble feature selection methods," *Bioinformatics*, vol. 26, no. 3, pp. 392–398, Feb. 2010.

[29] P. Drotár, S. Šimoňák, E. Pietriková, M. Chovanec, E. Chovancová, N. Ádám, C. Szabó, A. Baláž, and M. Biňas, "Comparison of Filter Techniques for Two-Step Feature Selection," *Computing and Informatics*, vol. 36, no. 3, pp. 597–617, Jul. 2017.

[30] T. G. Dietterich, "Ensemble Methods in Machine Learning," in *Proceedings of the First International Workshop on Multiple Classifier Systems*. Springer-Verlag, Jun. 2000, pp. 1–15.

[31] R. Polikar, "Ensemble Learning," in *Ensemble Machine Learning: Methods and Applications*, C. Zhang and Y. Ma, Eds. Boston, MA: Springer US, 2012, pp. 1–34.

[32] M. Feurer, A. Klein, K. Eggensperger, J. Springenberg, M. Blum, and F. Hutter, "Efficient and Robust Automated Machine Learning," in *Advances in Neural Information Processing Systems 28*, C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, Eds. Curran Associates, Inc., 2015, pp. 2962–2970.

[33] V. Bolón-Canedo and A. Alonso-Betanzos, "Ensembles for feature selection: A review and future trends," *Information Fusion*, vol. 52, pp. 1–12, Dec. 2019.

[34] K. Pearson, "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*, vol. 2, no. 11, pp. 559–572, Nov. 1901.

[35] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, no. 1, pp. 81–106, Mar. 1986.

[36] X.-M. Zhao, "Maximum Relevance/Minimum Redundancy (MRMR)," in *Encyclopedia of Systems Biology*, W. Dubitzky, O. Wolkenhauer, K.-H. Cho, and H. Yokota, Eds. New York, NY: Springer New York, 2013, pp. 1191–1192.

[37] V. Zuber and K. Strimmer, "High-Dimensional Regression and Variable Selection Using CAR Scores," *Statistical Applications in Genetics and Molecular Biology*, vol. 10, no. 1, 2011.

[38] K. Kira and L. A. Rendell, "The feature selection problem: Traditional methods and a new algorithm," in *Proceedings of the Tenth National Conference on Artificial Intelligence*. AAAI Press, Dec. 1992, pp. 129–134.

[39] F. Fleuret, "Fast Binary Feature Selection with Conditional Mutual Information," *The Journal of Machine Learning Research*, vol. 5, pp. 1531–1555, Jan. 2004.

[40] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[41] T. Hastie, J. Friedman, and R. Tibshirani, *The Elements of Statistical Learning*. Springer New York, 2001.

[42] M. Peña, R. Liao, and A. Brenning, "Using spectrotemporal indices to improve the fruit-tree crop classification accuracy," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 128, pp. 158–169, 2017.

[43] B. Bischl, J. Richter, J. Bossek, D. Horn, J. Thomas, and M. Lang, "mlrMBO: A Modular Framework for Model-Based Optimization of Expensive Black-Box Functions," *ArXiv e-prints*, Mar. 2017.

[44] M. Binder, J. Moosbauer, J. Thomas, and B. Bischl, "Multi-Objective Hyperparameter Tuning and Feature Selection using Filter Ensembles," *arXiv:1912.12912 [cs, stat]*, Feb. 2020.

[45] P. Schratz, J. Muenchow, E. Iturritxa, J. Richter, and A. Brenning, "Hyperparameter tuning and performance assessment of statistical and machine-learning algorithms using spatial data," *Ecological Modelling*, vol. 406, pp. 109–120, Aug. 2019.

[46] F. Hutter, H. H. Hoos, and K. Leyton-Brown, "Sequential model-based optimization for general algorithm configuration," in *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2011, pp. 507–523.

[47] D. R. Jones, M. Schonlau, and W. J. Welch, "Efficient global optimization of expensive black-box functions," *Journal of Global Optimization*, vol. 13, no. 4, pp. 455–492, Dec. 1998.

[48] J. Bergstra and Y. Bengio, "Random Search for Hyper-parameter Optimization," *J. Mach. Learn. Res.*, vol. 13, pp. 281–305, Feb. 2012.

[49] A. Brenning, "Spatial cross-validation and bootstrap for the assessment of prediction rules in remote sensing: The R package sperrorest," in *2012 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, Jul. 2012, R package version 2.1.0.

[50] J. H. Friedman, "Greedy function approximation: A gradient boosting machine." *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001.

[51] B. M. Greenwell, B. C. Boehmke, and A. J. McCarthy, "A Simple and Effective Model-Based Variable Importance Measure," *arXiv:1805.04755 [cs, stat]*, May 2018.

[52] C. Molnar, *Interpretable Machine Learning - A Guide for Making Black Box Models Explainable*, 2019.

[53] D. W. Apley and J. Zhu, "Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models," *arXiv:1612.08468 [stat]*, Aug. 2019.

[54] R Core Team, *R: A Language and Environment for Statistical Computing*, Vienna, Austria, 2019.

[55] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. New York, NY, USA: ACM, 2016, pp. 785–794.

[56] A. Karatzoglou, A. Smola, K. Hornik, and A. Zeileis, "Kernlab – An S4 Package for Kernel Methods in R," *Journal of Statistical Software*, vol. 11, no. 9, pp. 1–20, 2004.

[57] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *Journal of Statistical Software*, vol. 33, no. 1, pp. 1–22, 2010.

[58] M. B. Kursa, *Praznik: Collection of Information-Based Feature Selection Filters*, 2018.

[59] Z. Zawadzki and M. Kosinski, *FSelectorRcpp: 'Rcpp' Implementation of 'FSelector' Entropy-Based Feature Selection Algorithms with a Sparse Matrix Support*, 2019.

[60] B. Bischl, M. Lang, L. Kotthoff, J. Schiffner, J. Richter, E. Studerus, G. Casalicchio, and Z. M. Jones, "mlr: Machine learning in R," *Journal of Machine Learning Research*, vol. 17, no. 170, pp. 1–5, 2016.

[61] W. M. Landau, "The drake R package: A pipeline toolkit for reproducibility and high-performance computing," *Journal of Open Source Software*, vol. 3, no. 21, 2018.

[62] M. Ghosh, S. Adhikary, K. K. Ghosh, A. Sardar, S. Begum, and R. Sarkar, "Genetic algorithm based cancerous gene identification from microarray data using ensemble of filter methods,"

Medical & Biological Engineering & Computing, vol. 57, no. 1, pp. 159–176, Jan. 2019.

[63] D. N. H. Horler, M. Dockray, and J. Barber, "The red edge of plant leaf reflectance," *International Journal of Remote Sensing*, vol. 4, no. 2, pp. 273–288, Jan. 1983.

[64] M. Hais, K. N. Hellebrandová, and V. Šrámek, "Potential of Landsat spectral indices in regard to the detection of forest health changes due to drought effects," *Journal of Forest Science*, vol. 65 (2019), no. No. 2, pp. 70–78, Mar. 2019.

[65] B.-c. Gao, "NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space," *Remote Sensing of Environment*, vol. 58, no. 3, pp. 257–266, Dec. 1996.

[66] L. W. Lehnert, H. Meyer, and J. Bendix, *Hsdar: Manage, Analyse and Simulate Hyperspectral Data in R*, 2016.

[67] K. M. de Beurs and P. A. Townsend, "Estimating the effect of gypsy moth defoliation using MODIS," *Remote Sensing of Environment*, vol. 112, no. 10, pp. 3983–3990, Oct. 2008.

[68] R. Rengarajan and J. R. Schott, "Modeling forest defoliation using simulated BRDF and assessing its effect on reflectance and sensor reaching radiance," in *Remote Sensing and Modeling of Ecosystems for Sustainability XIII*, vol. 9975. International Society for Optics and Photonics, Sep. 2016, p. 997503.

[69] R. Meng, P. E. Dennison, F. Zhao, I. Shendryk, A. Rickert, R. P. Hanavan, B. D. Cook, and S. P. Serbin, "Mapping canopy defoliation by herbivorous insects at the individual tree level using bi-temporal airborne imaging spectroscopy and LiDAR measurements," *Remote Sensing of Environment*, vol. 215, pp. 170–183, Sep. 2018.

[70] U. Kälin, N. Lang, C. Hug, A. Gessler, and J. D. Wegner, "Defoliation estimation of forest trees from ground-level images," *Remote Sensing of Environment*, vol. 223, pp. 143–153, Mar. 2019.

[71] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 114, pp. 24–31, Apr. 2016.

[72] J. Xia, W. Liao, J. Chanussot, P. Du, G. Song, and W. Philips, "Improving Random Forest With Ensemble of Features and Semisupervised Feature Extraction," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 7, pp. 1471–1475, Jul. 2015.

[73] F. E. Fassnacht, C. Neumann, M. Förster, H. Buddenbaum, A. Ghosh, A. Clasen, P. K. Joshi, and B. Koch, "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*, vol. 7, no. 6, pp. 2547–2561, Jun. 2014.

[74] J. Feng, L. Jiao, F. Liu, T. Sun, and X. Zhang, "Unsupervised feature selection based on maximum information and minimum redundancy for hyperspectral images," *Pattern Recognition*, vol. 51, pp. 295–309, Mar. 2016.

[75] S. Georganos, T. Grippa, S. Vanhuysse, M. Lennert, M. Shimoni, S. Kalogirou, and E. Wolff, "Less is more: Optimizing classification performance through feature selection in a very-high-resolution remote sensing object-based urban application," *GIScience & Remote Sensing*, vol. 55, no. 2, pp. 221–242, Mar. 2018.

[76] J. F. R. Rochac and N. Zhang, "Feature extraction in hyperspectral imaging using adaptive feature selection approach," in *2016 Eighth International Conference on Advanced Computational Intelligence (ICACI)*, Feb. 2016, pp. 36–40.

[77] S. Xu, Q. Zhao, K. Yin, F. Zhang, D. Liu, and G. Yang, "Combining random forest and support vector machines for object-based rural-land-cover classification using high spatial resolution imagery," *Journal of Applied Remote Sensing*, vol. 13, no. 1, p. 014521, Feb. 2019.

[78] I. Shendryk, M. Broich, M. G. Tulbure, A. McGrath, D. Keith, and S. V. Alexandrov, "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*, vol. 187, pp. 202–217, Dec. 2016.

[79] M. Ludwig, T. Morgenthal, F. Detsch, T. P. Higginbottom, M. Lezama Valdes, T. Nauß, and H. Meyer, "Machine learning and multi-sensor based modelling of woody vegetation in the Molopo Area, South Africa," *Remote Sensing of Environment*, vol. 222, pp. 195–203, Mar. 2019.

[80] H. Meyer, C. Reudenbach, T. Hengl, M. Katurji, and T. Nauss, "Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation," *Environmental Modelling & Software*, vol. 101, pp. 1–9, Mar. 2018.

[81] H. Zandler, A. Brenning, and C. Samimi, "Quantifying dwarf shrub biomass in an arid environment: Comparing empirical methods in a high dimensional setting," *Remote Sensing of Environment*, vol. 158, pp. 140–155, Mar. 2015.

[82] Y. Guo, F.-L. Chung, G. Li, and L. Zhang, "Multi-Label Bioinformatics Data Classification With Ensemble Embedded Feature Selection," *IEEE Access*, vol. 7, pp. 103 863–103 875, 2019.

[83] M. Radovic, M. Ghalwash, N. Filipovic, and Z. Obradovic, "Minimum redundancy maximum relevance feature selection approach for temporal gene expression data," *BMC Bioinformatics*, vol. 18, no. 1, p. 9, Jan. 2017.

[84] J. E. Vogelmann, B. N. Rock, and D. M. Moss, "Red edge spectral measurements from sugar maple leaves," *International Journal of Remote Sensing*, vol. 14, no. 8, pp. 1563–1575, May 1993.

[85] F. Boochs, G. Kupfer, K. Dockter, and W. Kühbauch, "Shape of the red edge as vitality indicator for plants," *International Journal of Remote Sensing*, vol. 11, no. 10, pp. 1741–1753, Oct. 1990.

[86] P. L. Nagler, Y. Inoue, E. P. Glenn, A. L. Russ, and C. S. T. Daughtry, "Cellulose absorption index (CAI) to quantify mixed soil–plant litter scenes," *Remote Sensing of Environment*, vol. 87, no. 2-3, pp. 310–325, Oct. 2003.

[87] C. L. Walthall, C. S. T. Daughtry, E. W. Chappelle, J. E. Mcmurtrey, and M. S. Kim, "The Use of High Spectral Resolution Bands for Estimating Absorbed Photosynthetically Active Radiation (A Par)," 1994.

[88] G. A. Carter, "Ratios of leaf reflectances in narrow wavebands as indicators of plant stress," *International Journal of Remote Sensing*, vol. 15, no. 3, pp. 697–703, Feb. 1994.

[89] P. J. Zarco-Tejada, J. C. Pushnik, S. Dobrowski, and S. L. Ustin, "Steady-state chlorophyll a fluorescence detection from canopy derivative reflectance and double-peak red-edge effects," *Remote Sensing of Environment*, vol. 84, no. 2, pp. 283–294, Feb. 2003.

[90] A. A. Gitelson, Y. Gritz †, and M. N. Merzlyak, "Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves," *Journal of Plant Physiology*, vol. 160, no. 3, pp. 271–282, Jan. 2003.

[91] N. Oppelt and W. Mauser, "Hyperspectral monitoring of physiological parameters of wheat during a vegetation period using AVIS data," *International Journal of Remote Sensing*, vol. 25, no. 1, pp. 145–159, Jan. 2004.

[92] B. Datt, "Visible/near infrared reflectance and chlorophyll content in Eucalyptus leaves," *International Journal of Remote Sensing*, vol. 20, no. 14, pp. 2741–2759, Jan. 1999.

[93] ——, "Remote Sensing of Chlorophyll a, Chlorophyll b, Chlorophyll a$\mathplus$b, and Total Carotenoid Content in Eucalyptus Leaves," *Remote Sensing of Environment*, vol. 66, no. 2, pp. 111–121, Nov. 1998.

[94] ——, "Remote Sensing of Water Content in Eucalyptus Leaves," *Australian Journal of Botany*, vol. 47, no. 6, p. 909, 1999.

[95] G. le Maire, C. François, and E. Dufrêne, "Towards universal broad leaf chlorophyll indices using PROSPECT simulated database and hyperspectral reflectance measurements," *Remote Sensing of Environment*, vol. 89, no. 1, pp. 1–28, Jan. 2004.

[96] G. Lemaire, C. Francois, K. Soudani, D. Berveiller, J. Pontailler, N. Breda, H. Genet, H. Davi, and E. Dufrene, "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*, vol. 112, no. 10, pp. 3846–3864, Oct. 2008.

[97] A. Apan, A. Held, S. Phinn, and J. Markley, "Detecting sugarcane 'orange rust' disease using EO-1 Hyperion hyperspectral imagery," *International Journal of Remote Sensing*, vol. 25, no. 2, pp. 489–498, Jan. 2004.

[98] J. Peñuelas, J. A. Gamon, A. L. Fredeen, J. Merino, and C. B. Field, "Reflectance indices associated with physiologi-

cal changes in nitrogen- and water-limited sunflower leaves," *Remote Sensing of Environment*, vol. 48, no. 2, pp. 135–146, May 1994.

[99] A. Huete, "A comparison of vegetation indices over a global set of TM images for EOS-MODIS," *Remote Sensing of Environment*, vol. 59, no. 3, pp. 440–451, Mar. 1997.

[100] W. Wu, "The Generalized Difference Vegetation Index (GDVI) for Dryland Characterization," *Remote Sensing*, vol. 6, no. 2, pp. 1211–1233, Jan. 2014.

[101] R. C. G. Smith, J. Adams, D. J. Stephens, and P. T. Hick, "Forecasting wheat yield in a Mediterranean-type environment from the NOAA satellite," *Australian Journal of Agricultural Research*, vol. 46, no. 1, p. 113, 1995.

[102] A. A. Gitelson, C. Buschmann, and H. K. Lichtenthaler, "The Chlorophyll Fluorescence Ratio F735/F700 as an Accurate Measure of the Chlorophyll Content in Plants," *Remote Sensing of Environment*, vol. 69, no. 3, pp. 296–302, Sep. 1999.

[103] A. A. Gitelson, Y. J. Kaufman, and M. N. Merzlyak, "Use of a green channel in remote sensing of global vegetation from EOS-MODIS," *Remote Sensing of Environment*, vol. 58, no. 3, pp. 289–298, Dec. 1996.

[104] L. S. Galvão, A. R. Formaggio, and D. A. Tisot, "Discrimination of sugarcane varieties in Southeastern Brazil with EO-1 Hyperion data," *Remote Sensing of Environment*, vol. 94, no. 4, pp. 523–534, Feb. 2005.

[105] A. Maccioni, G. Agati, and P. Mazzinghi, "New vegetation indices for remote measurement of chlorophylls based on leaf directional reflectance spectra," *Journal of Photochemistry and Photobiology B: Biology*, vol. 61, no. 1-2, pp. 52–61, Aug. 2001.

[106] C. Daughtry, "Estimating Corn Leaf Chlorophyll Concentration from Leaf and Canopy Reflectance," *Remote Sensing of Environment*, vol. 74, no. 2, pp. 229–239, Nov. 2000.

[107] C. Wu, Z. Niu, Q. Tang, and W. Huang, "Estimating chlorophyll content from hyperspectral vegetation indices: Modeling and validation," *Agricultural and Forest Meteorology*, vol. 148, no. 8-9, pp. 1230–1241, Jul. 2008.

[108] D. A. Sims and J. A. Gamon, "Relationships between leaf pigment content and spectral reflectance across a wide range of species, leaf structures and developmental stages," *Remote Sensing of Environment*, vol. 81, no. 2-3, pp. 337–354, Aug. 2002.

[109] R. Hernández-Clemente, R. M. Navarro-Cerrillo, L. Suárez, F. Morales, and P. J. Zarco-Tejada, "Assessing structural effects on PRI for stress detection in conifer forests," *Remote Sensing of Environment*, vol. 115, no. 9, pp. 2360–2375, Sep. 2011.

[110] J. Qi, A. Chehbouni, A. R. Huete, Y. H. Kerr, and S. Sorooshian, "A modified soil adjusted vegetation index," *Remote Sensing of Environment*, vol. 48, no. 2, pp. 119–126, May 1994.

[111] E. Hunt and B. Rock, "Detection of changes in leaf water content using Near- and Middle-Infrared reflectances," *Remote Sensing of Environment*, vol. 30, no. 1, pp. 43–54, Oct. 1989.

[112] J. M. Chen, "Evaluation of vegetation indices and a Modified Simple Ratio for boreal applications," Natural Resources Canada/ESS/Scientific and Technical Publishing Services, Tech. Rep., 1996.

[113] J. Dash and P. Curran, "Evaluation of the MERIS terrestrial chlorophyll index (MTCI)," *Advances in Space Research*, vol. 39, no. 1, pp. 100–104, Jan. 2007.

[114] D. Haboudane, J. R. Miller, N. Tremblay, P. J. Zarco-Tejada, and L. Dextraze, "Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture," *Remote Sensing of Environment*, vol. 81, no. 2-3, pp. 416–426, Aug. 2002.

[115] L. Serrano, J. Peñuelas, and S. L. Ustin, "Remote sensing of nitrogen and lignin in Mediterranean vegetation from AVIRIS data," *Remote Sensing of Environment*, vol. 81, no. 2-3, pp. 355–364, Aug. 2002.

[116] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sensing of Environment*, vol. 8, no. 2, pp. 127–150, May 1979.

[117] A. Gitelson and M. N. Merzlyak, "Quantitative estimation of chlorophyll-a using reflectance spectra: Experiments with autumn chestnut and maple leaves," *Journal of Photochemistry and Photobiology B: Biology*, vol. 22, no. 3, pp. 247–252, Mar. 1994.

[118] L. Guanter, L. Alonso, and J. Moreno, "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*, vol. 43, no. 12, pp. 2908–2917, Dec. 2005.

[119] G. Rondeaux, M. Steven, and F. Baret, "Optimization of soil-adjusted vegetation indices," *Remote Sensing of Environment*, vol. 55, no. 2, pp. 95–107, Feb. 1996.

[120] E. W. Chappelle, M. S. Kim, and J. E. McMurtrey, "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*, vol. 39, no. 3, pp. 239–247, Mar. 1992.

[121] J. Gamon, J. Peñuelas, and C. Field, "A narrow-waveband spectral index that tracks diurnal changes in photosynthetic efficiency," *Remote Sensing of Environment*, vol. 41, no. 1, pp. 35–44, Jul. 1992.

[122] P. Zarco-Tejada, V. González-Dugo, L. Williams, L. Suárez, J. Berni, D. Goldhamer, and E. Fereres, "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*, vol. 138, pp. 38–50, Nov. 2013.

[123] S. R. Garrity, J. U. Eitel, and L. A. Vierling, "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*, vol. 115, no. 2, pp. 628–635, Feb. 2011.

[124] M. N. Merzlyak, A. A. Gitelson, O. B. Chivkunova, and V. Y. Rakitin, "Non-destructive optical detection of pigment changes during leaf senescence and fruit ripening," *Physiologia Plantarum*, vol. 106, no. 1, pp. 135–141, May 1999.

[125] G. A. Blackburn, "Quantifying chlorophylls and caroteniods at leaf and canopy scales," *Remote Sensing of Environment*, vol. 66, no. 3, pp. 273–285, Dec. 1998.

[126] J. Penuelas, J. Pinol, R. Ogaya, and I. Filella, "Estimation of plant water concentration by the reflectance Water Index WI (R900/R970)," *International Journal of Remote Sensing*, vol. 18, no. 13, pp. 2869–2875, Sep. 1997.

[127] J.-L. Roujean and F.-M. Breon, "Estimating PAR absorbed by vegetation from bidirectional reflectance measurements," *Remote Sensing of Environment*, vol. 51, no. 3, pp. 375–384, Mar. 1995.

[128] M. A. Cho and A. K. Skidmore, "A new technique for extracting the red edge position from hyperspectral data: The linear extrapolation method," *Remote Sensing of Environment*, vol. 101, no. 2, pp. 181–193, Mar. 2006.

[129] G. Guyot and F. Baret, "Utilisation de la haute resolution spectrale pour suivre l'etat des couverts vegetaux," in *Spectral Signatures of Objects in Remote Sensing*, ser. ESA Special Publication, T. D. Guyenne and J. J. Hunt, Eds., vol. 287, Apr. 1988, p. 279.

[130] A. Huete, "A soil-adjusted vegetation index (SAVI)," *Remote Sensing of Environment*, vol. 25, no. 3, pp. 295–309, Aug. 1988.

[131] J. Penuelas, I. Filella, P. Lloret, F. Munoz, and M. Vilaleliu, "Reflectance assessment of mite effects on apple trees," *International Journal of Remote Sensing*, vol. 16, no. 14, pp. 2727–2733, Sep. 1995.

[132] M. Vincini, E. Frazzi, and P. D'Alessio, "Angular dependence of maize and sugar beet VIs from directional CHRIS/Proba data," in *Proc. 4th ESA CHRIS PROBA Workshop*, vol. 2006, 2006, pp. 19–21.

[133] C. F. Jordan, "Derivation of leaf-area index from quality of light on the forest floor," *Ecology*, vol. 50, no. 4, pp. 663–666, Jul. 1969.

[134] A. A. Gitelson and M. N. Merzlyak, "Remote estimation of chlorophyll content in higher plant leaves," *International Journal of Remote Sensing*, vol. 18, no. 12, pp. 2691–2697, Aug. 1997.

[135] J. McMurtrey, E. Chappelle, M. Kim, J. Meisinger, and L. Corp, "Distinguishing nitrogen fertilization levels in field corn (Zea mays L.) with actively induced fluorescence and passive reflectance measurements," *Remote Sensing of Environment*, vol. 47, no. 1, pp. 36–44, Jan. 1994.

[136] P. J. Zarco-Tejada and J. R. Miller, "Land cover mapping at BOREAS using red edge spectral parameters from CASI

imagery," *Journal of Geophysical Research: Atmospheres*, vol. 104, no. D22, pp. 27 921–27 933, Nov. 1999.

[137] H. Lichtenthaler, M. Lang, M. Sowinska, F. Heisel, and J. Miehé, "Detection of vegetation stress via a new high resolution fluorescence imaging system," *Journal of Plant Physiology*, vol. 148, no. 5, pp. 599–612, Jan. 1996.

[138] R. Hernández-Clemente, R. M. Navarro-Cerrillo, and P. J. Zarco-Tejada, "Carotenoid content estimation in a heterogeneous conifer forest using narrow-band indices and PROSPECTDART simulations," *Remote Sensing of Environment*, vol. 127, pp. 298–315, Dec. 2012.

[139] C. D. Elvidge and Z. Chen, "Comparison of broad-band and narrow-band red and near-infrared vegetation indices," *Remote Sensing of Environment*, vol. 54, no. 1, pp. 38–48, Oct. 1995.

[140] I. Filella and J. Penuelas, "The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status." *International Journal of Remote Sensing*, vol. 15, no. 7, pp. 1459–1470, May 1994.

[141] N. Levin, G. J. Kidron, and E. Ben-Dor, "Surface properties of stabilizing coastal dunes: Combining spectral and field analyses," *Sedimentology*, vol. 54, no. 4, pp. 771–788, Aug. 2007.

[142] D. B. Lobell, G. P. Asner, B. E. Law, and R. N. Treuhaft, "Subpixel canopy cover estimation of coniferous forests in Oregon using SWIR imaging spectrometry," *Journal of Geophysical Research: Atmospheres*, vol. 106, no. D6, pp. 5151–5160, Mar. 2001.

[143] E. R. Hunt, P. C. Doraiswamy, J. E. McMurtrey, C. S. Daughtry, E. M. Perry, and B. Akhmedov, "A visible band index for remote sensing leaf chlorophyll content at the canopy scale," *International Journal of Applied Earth Observation and Geoinformation*, vol. 21, pp. 103–112, Apr. 2013.

[144] N. Broge and E. Leblanc, "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*, vol. 76, no. 2, pp. 156–172, May 2001.

Patrick Schratz is a Data Scientist working in Zurich, CH. His area of expertise is applied machine learning, more specifically the field of environmental modeling. He is a PhD candidate at the GIScience group, Department of Geography at University of Jena where he acts as a researcher in the environmental modeling field. 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 spatio-temporal 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 has earned his PhD at the Friedrich-Alexander University Erlangen-Nuremberg (Germany) in 2013. He joined the business location department of a large consulting company as a geodata scientist for more than two years until the prediction of a strong El Niño event brought him back to academia in 2016 (Friedrich Schiller University Jena). Currently, 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 Basque country University 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 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 PhD 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 spatio-temporal trend detection in environmental data.

Bernd Bischl obtained his Ph.D from Dortmund Technical University in 2013. He is a professor for "Statistical Learning and Data Science" at the Department of Statistics at the Ludwig-Maximilians-University Munich 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 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.