*Article*

# MISPEL: A Multi-Crop Spectral Library for Statistical Crop Trait Retrieval and Agricultural Monitoring

Peter Borrmann [†], Patric Brandt and Heike Gerighausen *

Julius Kühn Institute (JKI)–Federal Research Centre for Cultivated Plants, Institute for Crop and Soil Science, Bundesallee 58, 38116 Braunschweig, Germany

* Correspondence: heike.gerighausen@julius-kuehn.de
[†] Current address: Department of Remote Sensing, Helmholtz Centre for Environmental Research (UFZ), Permoserstraße 15, 04318 Leipzig, Germany.

**Abstract:** Spatiotemporally accurate estimates of crop traits are essential for both scientific modeling and practical decision making in sustainable agricultural management. Besides efficient and concise methods to derive these traits, site- and crop-specific reference data are needed to develop and validate retrieval methods. To address this shortcoming, this study first includes the establishment of 'MISPEL', a comprehensive spectral library (SpecLib) containing hyperspectral measurements and reference data for six key traits of ten widely grown crops. Secondly, crop-specific statistical leaf area index (LAI) models for winter wheat are developed based on a hyperspectral (MISPEL$_{FR}$) and a simulated Sentinel-2 (MISPEL$_{S2}$) SpecLib applying four nonparametric methods. Finally, an independent Sentinel-2 model evaluation at the DEMMIN test site in Germany is conducted, including a comparison with the commonly used SNAP-LAI product. To date, MISPEL comprises a set of 1411 spectra of ten crops and more than 6800 associated reference measurements. Cross-validations of winter wheat LAI models revealed that Elastic-net generalized linear model (GLMNET) and Gaussian process (GP) regressions outperformed partial least squares (PLS) and random forest (RF) regressions, showing RSQ values up to 0.86 and a minimal NRMSE of 0.21 using MISPEL$_{FR}$. GLMNET and GP models based on MISPEL$_{S2}$ further outperformed SNAP-based LAI estimates derived for the external validation site. Thus, it is concluded that the presented SpecLib 'MISPEL' and applied methodology have a very high potential for deriving diverse crop traits of multiple crops in view of most recent and future multi-, super-, and hyperspectral satellite missions.

**Keywords:** MISPEL; spectral library; crop monitoring; crop traits; hyperspectral remote sensing; Sentinel-2; machine learning; leaf area index (LAI); SNAP; winter wheat

## 1. Introduction

Space-borne observation of crop traits delivers important parameters to monitor crop health and model crop growth, supporting decision making processes towards sustainable agriculture [1]. Among key traits describing the current biophysical crop status are leaf area index (LAI) [2–12], biomass [2,6,13,14], chlorophyll content [3,5,9,10,12,15], nitrogen content [13,15–17], and carbon content [5,14].

Many studies have focused on the remotely sensed retrieval of winter wheat traits [5–7,9,12,18,19]. Wheat, especially the high yielding winter wheat, is an important cereal crop to ensure food security [20]. Germany is among the large grain producing countries in the world [21]. In 2020, winter wheat has been cultivated on 23.7% of the total arable land, accounting for the largest share among all crops being produced [22].

Hyperspectral sensors hold a high potential for monitoring crop traits as they exploit features with great spectral resolution using a high number of wavebands [19]. Most recent and upcoming hyperspectral missions such as the research missions EnMAP [23], HysPIRI [24], and PRISMA [25] provide data at an unprecedented spectral resolution and

are hence predestined for monitoring crop traits. The planned hyperspectral Copernicus mission CHIME will deliver routine hyperspectral data at a high revisit frequency [26]. Multispectral satellite missions such as Sentinel-2 (S2) [27] are equipped with fewer but specific spectral bands to detect crop characteristics. They are better suited for practical and large-scale applications because of their high spatiotemporal resolution.

Robust ground truth data are required to derive crop traits from optical earth observation (EO) sensors [28]. The systematic collection of spectroscopic measurements accompanied by reference measurements of in situ traits to construct spectral libraries (SpecLibs) has therefore played an important role for many sub-disciplines of remote sensing [29–37]. SpecLibs provide an excellent foundation for both multi- and hyperspectral model development because they allow the simulation of arbitrary optical sensor configurations through spectral resampling [6,17,19]. In previous studies, SpecLibs were primarily used for classification purposes of, e.g., shrubland species [35], wetland species [32], rocks [36], urban materials [34], and crops [29,33,37]. Only [17,19] explored SpecLibs for crop nitrogen, grain nitrogen, and yield retrieval. However, methods are not applied to satellite image data in these studies. For the accurate retrieval of crop traits, a particularly extensive in situ database consisting of multi-annual crop spectra and references for key biophysical parameters over the entire phenologiocal cycle is required. Such a basis of data has not been provided by existing SpecLibs so far.

The variety of parameter derivation approaches is frequently distinguished into parametric, nonparametric, physical, and hybrid methods. They are comprehensively reviewed for LAI estimation based on S2 imagery [8,38] and trait retrieval from spectroscopic data [39]. Parametric methods assume explicit relationships between spectral data and biophysical properties. Thus, vegetation indices (VIs) are directly related to a biophysical property [11,40]. In contrast, nonparametric methods have emerged as a more powerful tool—they fit a data-driven, non-explicit regression function that is able to exploit existing relationships between spectral data and the associated parameter [41]. Nonparametric approaches can be further divided into linear and nonlinear methods. The most common linear methods are principal component analysis [42] and partial least squares (PLS) regression [43]. Both are known for their fast performance and ease of application in the presence of collinearity of spectroscopic data by dimensionality reduction, whereas PLS regression can be considered an advancement in terms of regression [39]. PLS regression was used in a variety of studies to derive traits from, in particular, hyperspectral data [6,13,18,44–47]. Fewer studies, which received little attention, applied Ridge [48] and Lasso [49] regression, although they outperformed other methods in some studies [50,51]. With the emergence of nonlinear nonparametric methods, also referred to as machine learning (ML) methods, linear methods have been outperformed in many cases as they do not account for nonlinear relationships. They include decision-tree-based methods such as random forest (RF), kernel-based methods, and artificial neural networks (ANNs). In terms of kernel-based methods, which solve regression problems by transferring data to a higher-dimensional space, Gaussian process (GP) regression in particular proved to be the best performing method surpassing its relatives, support vector regression (SVR) and kernel ridge regression [52]. Increasingly, the special form of variational heteroscedastic GP regression (VHGP) [53] is used, which is additionally able to output the uncertainty of estimates. ANNs are known for their enormous versatility [54,55] but require sophisticated design and a lot of training time to incorporate many bands. Additionally, collinearity inherent to hyperspectral data renders applying ANNs rather difficult and requires preliminary feature selection. Unlike nonparametric methods, physical models are built on physical laws and mechanistic variable relationships. This type of model is predominantly based on the inversion of radiative transfer models (RTMs), enabling generalizability of the retrieval. However, the inversion is generally difficult because the models are mostly inadequately defined and ill-posed, limiting the applicability of RTMs [56,57]. Furthermore, RTMs typically need more powerful computational resources than statistical approaches to perform quickly and require expert knowledge of parameterization and inversion to

achieve good results [58]. Optimization techniques, lookup tables, and the hybrid use of RTMs and ML approaches are employed to enhance the inversion process [9,59,60]. The common LAI product of the Sentinel Application Platform (SNAP) is likewise based on a hybrid approach by training an ANN using RTM-simulated crop spectra [61]. Similar to other LAI products, underlying models have not been calibrated crop-specifically.

Against the backdrop of an insufficient database for crop-specific trait retrieval, this study encompasses three objectives: First, MISPEL, a multi-crop SpecLib is introduced. MISPEL contains reference measurements and spectra of ten crops for six traits that are highly relevant for monitoring the biophysical crop state. Second, four nonparametric models are deployed and cross-validated, demonstrating the potential of the SpecLib to estimate LAI for 10 individual crop types. LAI has been chosen for its broad application as a trait that characterizes crop states over time and because of its extensive use in the context of crop growth modeling. In a two-step approach, deployed models are trained and cross-validated using (i) MISPEL's original full range SpecLib (MISPEL$_{FR}$) and (ii) resampled spectra of MISPEL simulating the S2 band configuration (MISPEL$_{S2}$). Third, an independent validation is conducted to assess the transferability of the methodology applied. During this validation procedure, the four nonparametric LAI models trained using MISPEL$_{S2}$ (objective two) are compared against independent field measurements and the widely used SNAP-LAI product, focusing on winter wheat alone. Winter wheat has been chosen due to the crop's importance in the German agricultural sector and in terms of the crop's role in ensuring the global food supply.

## 2. Materials and Methods

### 2.1. Methodology

Figure 1 shows a flowchart of the methodology applied in this study. Sampling and validation sites, where data were collected to set up MISPEL$_{FR}$ and MISPEL$_{S2}$ and to validate the respective S2 LAI models developed, are shown and described in Figure 2 in Section 2.2. Sections 2.2.1 and 2.2.2 contain a detailed description of the resulting datasets. Section 2.3 elaborates on the preprocessing of hyperspectral field measurements to construct the SpecLib using obtained reference data. The procedure of selection, generation, and accuracy assessment of S2 crop trait models is summarized in Section 2.4. Section 2.5 describes the subsequent application of these models to S2 imagery from the Durable Environmental Multidisciplinary Monitoring Information Network (DEMMIN) test site and the retrieval of LAI using the SNAP toolbox, which are finally validated and compared as described in Section 2.7.
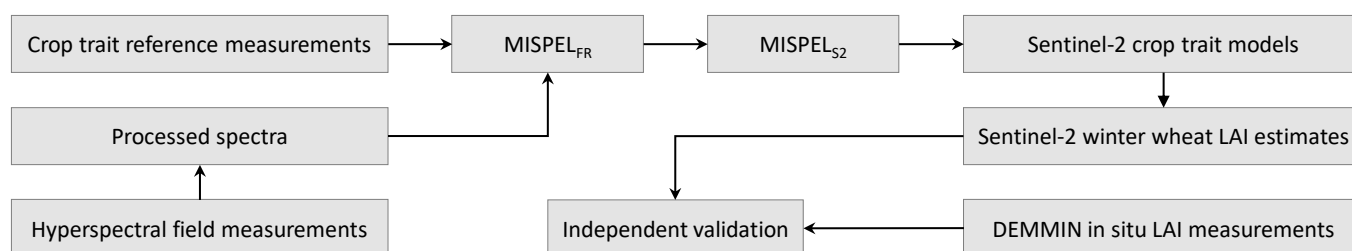


**Figure 1.** Flowchart of the methodology applied to establish the multi-crop SpecLib MISPEL and crop-specific Sentinel-2 (S2) trait models and to validate leaf area index (LAI) estimates of winter wheat derived from S2 imagery.

### 2.2. Data Collection

Figure 2 shows an overview of the sites in Germany where hyperspectral field measurements and corresponding crop trait measurements were collected to establish the multi-crop SpecLib. An additional independent set of in situ LAI measurements to validate remote-sensing-based estimations was gathered on four winter wheat test fields at the DEMMIN site in 2021 and 2022 (site 6). The DEMMIN test site is located near the town of Demmin in northeastern Germany and is part of the joint experiment for crop assessment

and monitoring (JECAM) as well as the terrestrial environmental observatories network (TERENO-NE) [62,63]. The sampling design for the independent validation was composed of elementary sampling units (ESUs), each of which consisted of 13 secondary sampling units (SSUs). At each of the SSUs, a 0.25 m$^2$ sampling frame, in which the samples were taken, was positioned.
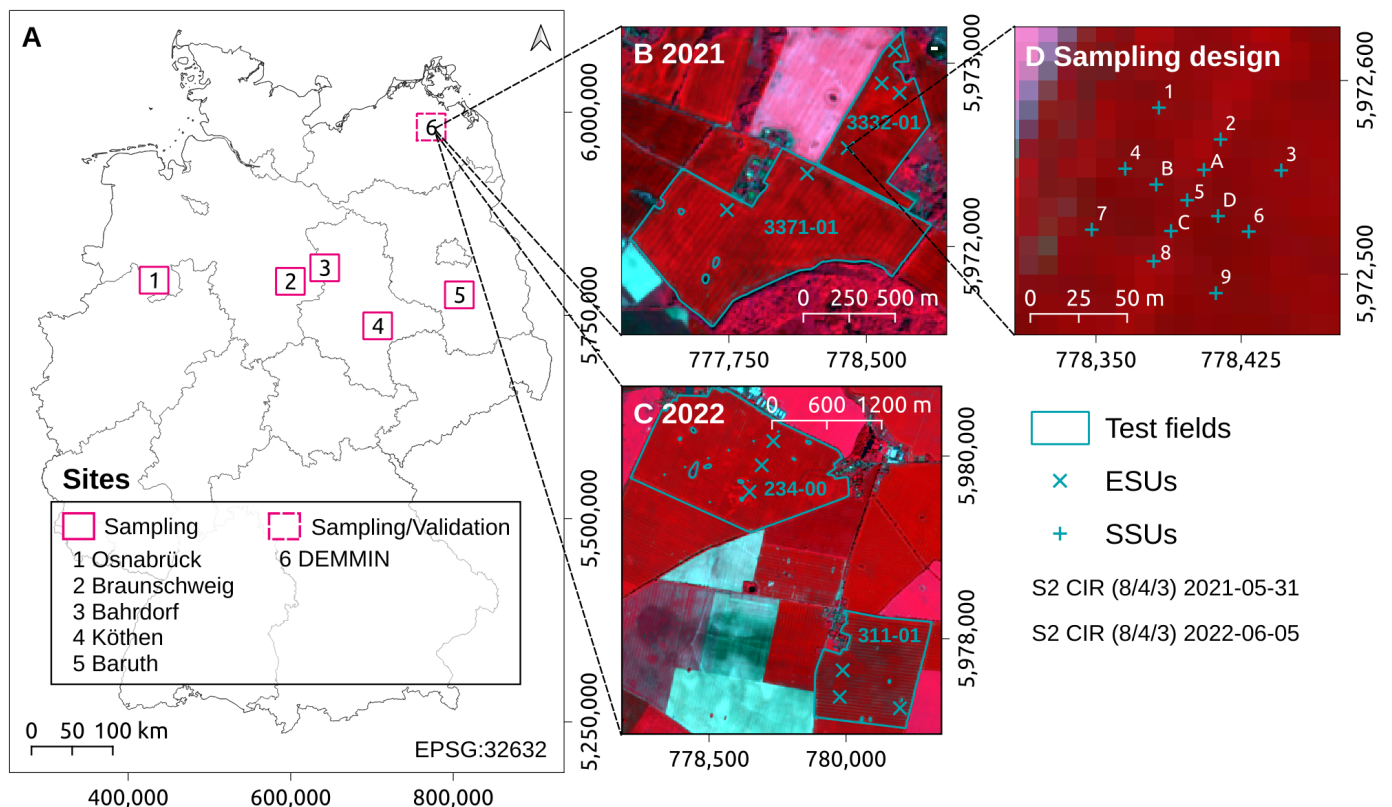


**Figure 2.** (**A**) Sites in Germany where hyperspectral field measurements and in situ references were obtained to establish the multi-crop spectral library MISPEL. (**B**,**C**) Depiction of winter wheat test fields and elementary sampling units (ESUs) at the DEMMIN test site where an additional independent set of in situ leaf area index (LAI) measurements was obtained in 2021 and 2022 and used for independent model validation. (**D**) Sampling design consisting of 13 secondary sampling units (SSUs) per ESU. Sentinel-2 (S2) color-infrared (CIR) false-color composites are shown in the background.

2.2.1. Hyperspectral Measurements and Reference Data Acquisition

Between 2011 and 2021, a total of 1411 field spectra of ten crops (Table 1) were recorded using an SVC HR-1024 (Spectra Vista Corporation, Poughkeepsie, NY, USA) and an ASD FieldSpec® 4 (ASD, Malvern Panalytical, Malvern, UK) spectrometer with a spectral range of 350 to 2500 nm.

Measurements were conducted using a white reference and in fivefold repetition in a 0.25 m$^2$ sampling unit. Within the same plots, LAI was nondestructively gathered using an LI-COR LAI-2200C Plant Canopy Analyzer (LI-COR Biosciences, Lincoln, NE, USA) and a SunScan Canopy Analysis System SS1-Com-R4 (Delta-T Devices, Burwell, Cambridge, UK). Phenology was determined according to the BBCH system of unified coding of phenological growth stages [64]. Height was gathered by threefold repeated and averaged measurements using a ruler from the ground to the plant top. Fresh above-ground biomass was determined by destructive harvest of the entire above-ground biomass present in the 0.25 m$^2$ area enclosed within the sampling frame and recorded in units of t/ha. The fresh above-ground biomass was dried at 60 °C for at least 24 h to obtain dry biomass and water content. The

biomass samples were further used to determine nitrogen content by combustion [65,66] using a vario MAX cube (Elementar Analysesysteme, Langenselbold, Germany).

**Table 1.** Sample size (N) of gathered hyperspectral field measurements per crop.

| Crop | Botanical Name | Abbreviation | N |
|---|---|---|---|
| Broad bean | Vicia faba | BB | 49 |
| Oat | Avena sativa | OA | 39 |
| Potato | Solanum tuberosum | PT | 54 |
| Spring barley | Hordeum vulgare | SBA | 75 |
| Sugar beet | Beta vulgaris | SBE | 106 |
| Triticale | xTriticosecale | TR | 66 |
| Winter barley | Hordeum vulgare | WB | 33 |
| Winter rapeseed | Brassica napus | WRA | 426 |
| Winter rye | Secale cereale | WRY | 157 |
| Winter wheat | Triticum aestivum | WW | 406 |

2.2.2. Independent Validation Data

An independent validation dataset of winter wheat LAI measurements was collected within the Federal Ministry of Food and Agriculture (BMEL)-funded AgriSens DEMMIN 4.0 project in 2021 and 2022 and retrieved from the digital in situ data collection provided by [28]. Between April and July, sampling was conducted on an approximately biweekly basis in a 0.25 m$^2$ sampling frame close to each secondary sampling unit (SSU) according to the sampling design depicted in Figure 2. LAI was collected by means of LI-COR LAI-2200C Plant Canopy Analyzers in a quadruple-repeated ABBBB sequence, where A corresponds to a reference reading above the canopy and B corresponds to a reading below the canopy. The LAI measurements were scatter-corrected using the FV2200 Software [67].

*2.3. Spectral Library Establishment*

The automated preprocessing of crop spectra was performed using the R language and environment of statistical computing [68], including the hsdar package for hyperspectral data analysis [69] and tidyverse [70] functions. The fivefold measurement was averaged per plot, resampled to 1 nm spectral resolution, and corrected by white standard reflectance. Water absorption bands in the wavelength regions of 1355–1425 nm and 1785–1999 nm were omitted and linearly interpolated. A Savitzky–Golay smoothing with 65 nm window size and 4th order polynomial was applied. For S2 models, spectra were resampled according to S2 sensor characteristics, abandoning channels 1, 9, and 10 designated for atmospheric correction. Prior to model development, MISPEL$_{FR}$ and MISPEL$_{S2}$ were normalized by unit vector normalization (UVN, Equation (1)) according to [6].

$$x_i^{norm} = x_i / \sqrt{\sum_{i=1}^{n} x_i^2},\qquad(1)$$

where $x_i$ corresponds to a single value of the spectrum.

*2.4. Model Selection, Implementation, and Accuracy Assessment*

The model selection process was informed by criteria such as suitability in terms of handling multicollinearity inherent to spectral data, proven performance in comparable applications as depicted in Section 1, and the ability for describing potential nonlinear relationships. The selection process resulted in choosing four nonparametric regression approaches: Elastic-net regularized generalized linear model (GLMNET), Gaussian process (GP) regression, partial least squares (PLS) regression, and random forest (RF) regression.

The GLMNET algorithm linearly combines L1 (Lasso) and L2 (Ridge) regularization penalties in a multiple linear regression model. GLMNET was applied using the glmnet package [71] as the engine. The critical hyperparameters of GLMNET are the total amount of regularization (penalty) and the mixture of Lasso and Ridge penalty (mixture). Gaussian

process (GP) regression is a kernel-based method that can solve nonlinear regression tasks by projecting data to a higher-dimensional space using a kernel function [72]. GP was implemented using the kernlab package [73], choosing a linear kernel without tuning parameters after evaluating different kernels. PLS is an ideal model if responses can be predicted using a linear regression. PLS models the data by maximizing the variance of predictors that are explained by unobserved latent variables. Simultaneously, PLS tries to maximize the correlation between those latent variables and the outcomes and therefore approaches the greatest possible variance of predictors and outcomes. PLS regression was implemented using the pls package [74]. The optimal number of components was determined during the hyperparameter tuning process. RF is a well-established and robust ensemble learning algorithm for classification and regression. It constructs a variety of decision or regression trees based on a randomly chosen subset of the data and a randomly chosen subset of features for each of the trees built. In regression, the prediction corresponds to the average prediction of all regression trees [75]. Due to its structure of uncorrelated decision trees, RF is an algorithm that can map nonlinear relationships. The ranger package [76] was used as the RF engine. The tuned hyperparameters were the minimal node size, the number of variables randomly sampled as candidates at each tree split, and the splitting rule.

All models were implemented using the caret framework [77]. Model implementation and accuracy assessment were conducted across all 10 crops and crop-wise for MISPEL$_{FR}$, which served as a hyperspectral best-case scenario in terms of spectral resolution and the simulated S2 dataset MISPEL$_{S2}$ described in Section 2.2.1. The latter is required to derive crop traits such as LAI from S2 imagery. Prior to model building, predictors were centered and scaled. Optimal sets of hyperparameters were determined by fivefold cross-validated grid search tuning. Final model accuracy metrics were averaged over a fivefold cross-validation. The accuracy metrics used for model evaluation were root mean square error (RMSE, Equation (2)), normalized root mean square error (NRMSE, Equation (3)), and the coefficient of determination (RSQ, Equation (4)). The *NRMSE* relates the *RMSE* to the observed range of the variable. It therefore facilitates the comparison of model errors with different scales or variable ranges as they occur between different crop types, but also different studies. In addition, the *BIAS* was used for external model validation to assess systematic prediction errors (Equation (5)).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - x_i)^2}{n}}, \tag{2}$$

$$NRMSE = \frac{RMSE}{\bar{x}}, \tag{3}$$

$$RSQ = 1 - \frac{\sum_{i=1}^{n}(x_i - y_i)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \tag{4}$$

$$BIAS = \frac{\sum_{i=1}^{n}(y_i - x_i)}{n}, \tag{5}$$

$x_i$ corresponds to in situ gained LAI; $\bar{x}$ is the average value of in situ LAI; $y_i$ is the estimated value of LAI; and $n$ is the number of observations.

### 2.5. Model Application to Sentinel-2 Data

S2 L2A imagery for test fields at the DEMMIN test site was retrieved from the CODE-DE cloud computing platform [78]. The imagery was processed using the terra package [79]. The 20 m bands (B5, B6, B7, B8A, B11, B12) were bilinearly resampled to a 10 m spatial resolution. All 60 m bands (B1, B9, B10) were disregarded, keeping consistency with the spectral configuration of MISPEL$_{S2}$ (see Section 2.3). The resulting raster stacks were cloud-masked using the scene classification layer provided by S2 L2A scenes and normalized by UVN according to Equation (1). Models trained as described in Section 2.4 were applied to preprocessed S2 scenes using terra's predict function.

*2.6. SNAP-Based Sentinel-2 LAI Retrieval*

The Sentinel Application Platform (SNAP, v.9.0.0) developed by the European Space Agency (ESA) was used to compute SNAP-based LAI estimates utilizing its integrated biophysical processor. SNAP estimates LAI values based on a pretrained ANN. The ANN was trained for S2A and B individually using a comprehensive database with reflectance and crop traits obtained from ground measurements and RTM [61]. Eight S2 top of canopy reflectance bands (B3, B4, B5, B6, B7, B8A, B11, and B12), viewing zenith, solar zenith, and relative azimuth angles were taken as inputs in order to spatially predict LAI using S2 scenes. Prior to scene-wise LAI calculation, the spatial resolution of all inputs was harmonized by resampling them to 10 m. Areas covered by clouds and cloud shadows were masked using scene classification information being part of each S2 L2A scene.

*2.7. Independent Model Validation*

The S2-based LAI estimates predicted by four developed models and SNAP were extracted for each SSU, shown in Figure 2, using terra's extract function. Subsequently, extracted LAI values were assigned to independent LAI measurements obtained from the DEMMIN test site by inexact acquisition date matching using the fuzzyjoin package [80] with a maximum time difference of 3 days. In situ measurements for which no satellite-based estimates were available in these intervals were discarded. Accuracy metrics described in Section 2.4 were used for external validation.

## 3. Results

### 3.1. Multi-Crop Spectral Library MISPEL and Model Establishment

Figures 3 and 4 give a descriptive overview of collected hyperspectral field measurements and reference measurements for six crop traits, respectively. These measurements were collected to establish the introduced MISPEL. MISPEL comprises 1411 spectra of 10 crops with the number of measurements varying by crop between 33 and 426 (Figure 3). To date, a particularly comprehensive database is available for winter rapeseed, winter wheat, winter rye, and sugar beet. The means, ranges, and standard deviations of reflectance spectra differed considerably depending on the crop. Reference measurements covered a set of 6837 individual in situ observations of six traits for 10 crops (Figure 4). Like the spectra, the traits collected showed a high crop-specific variability.
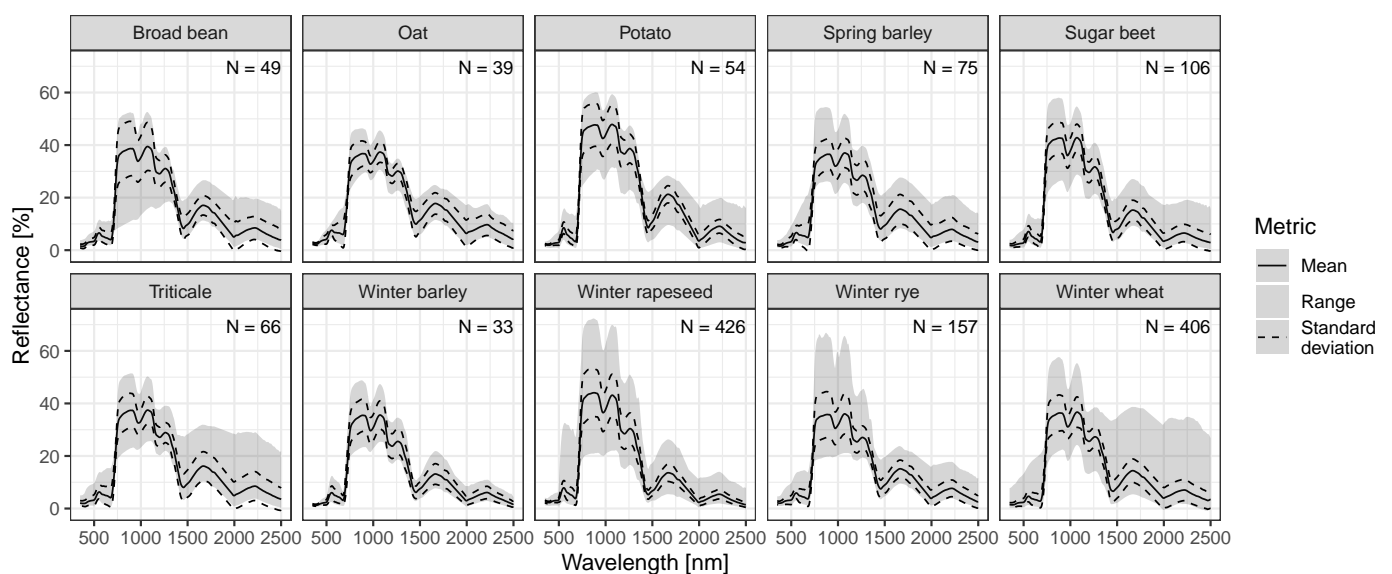


**Figure 3.** Sample size (N), mean, range, and standard deviation of hyperspectral field measurements per crop after preprocessing that, in conjunction with biophysical crop parameter reference measurements shown in Figure 4, form the introduced multi-crop spectral library MISPEL.
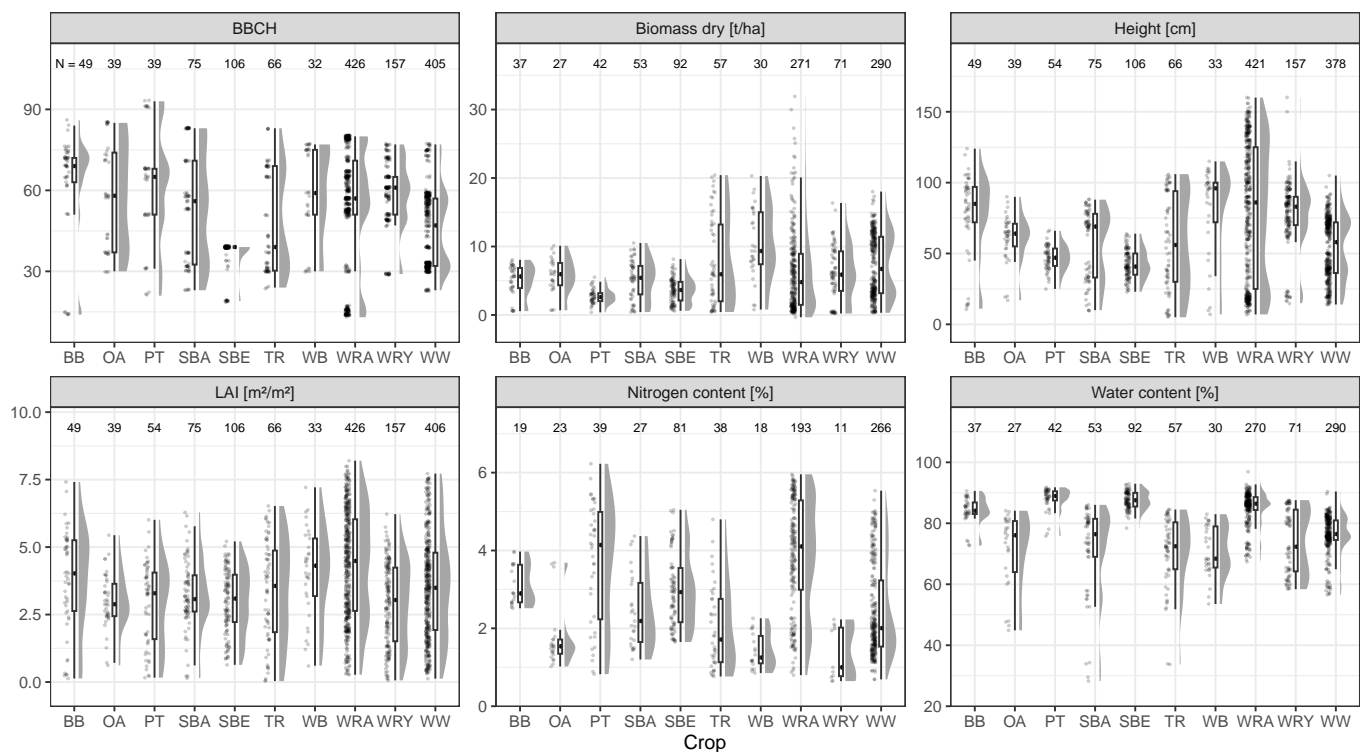
**Figure 4.** Boxplots and density distributions of crop trait reference measurements accompanying hyperspectral field measurements in Figure 3 for spectral library (SpecLib) establishment per trait and crop. BB = broad bean, OA = oat, PT = potato, SBA = spring barley, SBE = sugar beet, TR = triticale, WB = winter barley, WRA = winter rapeseed, WRY = winter rye, WW = winter wheat. Note that numbers vary as not each hyperspectral measurement comes with an entire set of reference measurements.

As illustrated by the plot of phenological development stages (BBCH), a wide range of phenological stages is covered by the reference measurements for all crops. However, the density distributions currently reveal non-uniform sampling distribution across phenological development stages for most of the crops included, as strong weather dependence of hyperspectral field measurements limits the ability to measure crop traits in the field. Winter wheat, for instance, shows a bimodal distribution of reference measurements with peaks at shooting (BBCH ≈ 30) and emergence (BBCH ≈ 50). The distributions of dry biomass references show typical ranges of values corresponding to the considered crops. Here, the winter crops triticale, winter barley, winter rapeseed, winter rye, and winter wheat showed much wider ranges compared to broad bean, oat, potato, spring barley and sugar beet. Note that sugar beet measurements do not exceed a BBCH stage of 39, as this corresponds to the end of rosette growth of above-ground vegetative parts in the first year of cultivation of sugar beet. The highest values of dry biomass were reached by winter rapeseed with outliers of more than 30 t/ha. Plant height is typically also strongly crop-specific. Compared to the cereals, potato and sugar beet showed an expected reduced height with mean values of 47 and 42 cm. Broad bean height varied between 11 and 124 cm. The cereals showed a range from 5 to 115 cm. Merely winter rye reached up to 160 cm of plant height. Again, winter rapeseed was partly grown in small plot trials in different trial sites and years with a wide range of genotypes, resulting in differences in growth habit and biomass development; it showed the widest range of values with 7 to 170 cm of height. The expression of LAI exhibited a fairly even distribution for all crops and ranged from 0.062 to 8.2 m$^2$/m$^2$. The mean values of oat, potato, spring barley, sugar beet, and winter rye were lower than those of broad bean, triticale, winter barley, winter rapeseed, and winter wheat. Nitrogen contents are highly crop-specific and characterized by a generally reduced number of samples because only a subset of biomass samples are further analyzed

by combustion, as described in Section 2.2.1. Broad bean, oat, winter barley, and winter rye showed small value ranges with means of 3.15, 1.75, 1.41, and 1.37%. Average values ranges are shown by spring barley, and sugar beet with mean values of 2.44 and 2.97%. Potato, triticale, winter rapeseed, and winter wheat showed the widest ranges of nitrogen content with mean values of 3.70, 2.01, 3.82, and 2.42%. What is striking is the differentiation of crops into two groups regarding water content. While broad bean, potato, sugar beet, and winter rapeseed showed consistently high values with means of 84.1, 88.0, 87.6, and 85.2%, the cereal crops of oat, spring barley, triticale, winter barley, winter rye, and winter wheat generally showed lower means (71.2, 71.5, 70.7, 70.9, 74.0, and 76.4%) and wider ranges of water content.

Validation results for LAI models established based on MISPEL$_{FR}$ and MISPEL$_{S2}$ are summarized in Table 2, focusing on winter wheat. Individual accuracy metrics for all remaining crops are included in Appendix A. Generally, the MISPEL$_{FR}$ models yielded slightly higher goodness of fits (RSQ = 0.81 to 0.86) and marginally lower errors (NRMSE = 0.20 to 0.23) than the MISPEL$_{S2}$ models (RSQ = 0.82 to 0.83, NRMSE = 0.22 to 0.23). Overall, GP obtained the best result (RSQ = 0.86, NRMSE = 0.20) using MISPEL$_{FR}$, followed by RF, GLMNET, and PLS. Based on MISPEL$_{S2}$, GLMNET performed best (RSQ = 0.83, NRMSE = 0.22), closely followed by GP, RF, and PLS.

**Table 2.** Accuracy metrics averaged over fivefold cross-validated predictions for winter wheat leaf area index (LAI) based on four presented statistical approaches using the two datasets MISPEL$_{FR}$ and MISPEL$_{S2}$. A comprehensive overview of LAI model accuracy metrics for all crops is provided in Appendix A.

| MISPEL | Model | NRMSE | RMSE | RSQ | N |
|--------|-------|-------|------|-----|---|
| FR | GLMNET | 0.22 | 0.78 | 0.83 | 406 |
| FR | GP | 0.20 | 0.70 | 0.86 | 406 |
| FR | PLS | 0.23 | 0.82 | 0.81 | 406 |
| FR | RF | 0.22 | 0.77 | 0.83 | 406 |
| S2 | GLMNET | 0.22 | 0.79 | 0.83 | 406 |
| S2 | GP | 0.23 | 0.79 | 0.83 | 406 |
| S2 | PLS | 0.23 | 0.81 | 0.82 | 406 |
| S2 | RF | 0.23 | 0.79 | 0.83 | 406 |

### 3.2. Crop-Specific S2 LAI Model Accuracies

As the GP LAI model turned out to be the top performing model in independent validation, Figure 5 shows the fivefold cross-validated accuracy metrics for GP LAI models trained on MISPEL$_{S2}$. The entirety of accuracy metrics for all model and SpecLib combinations are provided in Appendix A. The general GP model, trained on all LAI measurements, yielded an RSQ of 0.65, an NRMSE of 0.30, and an RMSE of 1.10. The best crop-specific model fits in terms of RSQ were obtained for triticale, winter rye, and winter wheat, reaching RSQ values ranging between 0.83 and 0.84. The corresponding NRMSEs varied from 0.21 to 0.24 (RMSE = 0.62–0.79). Slightly lower accuracy values were achieved for the winter barley, broad bean, winter rapeseed, and oat models. Models showing moderate accuracy were established for potato, sugar beet, and spring barley. The scatterplot including all crops shows the largest overall deviation of estimated values from the ideal line. Also for winter rapeseed, estimates are widely scattered as it was planted in a small plot trial with a wide range of genotypes, resulting in differences in growth habit and biomass development. In general, high LAI reference values tend to be underestimated in model predictions. This is particularly evident from some of the outliers present in scatterplots for broad bean, potato, spring barley, and the general model.
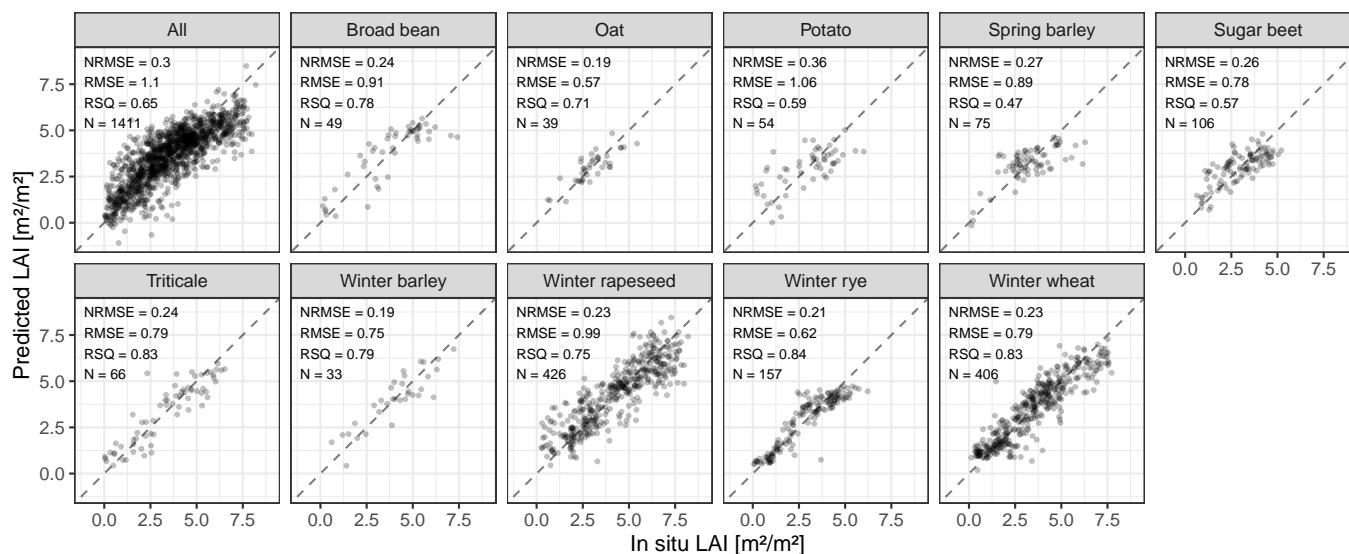
**Figure 5.** Crop-specific fivefold cross-validated Gaussian process (GP) predictions of leaf area index (LAI) versus in situ LAI measurements and accuracy metrics based on MISPEL$_{S2}$. Dashed lines indicate and ideal (1:1) relationship. The entirety of accuracy metrics for all model and SpecLib combinations are provided in Appendix A.

### 3.3. Independent Model Validation at the DEMMIN Test Site

The results of independent model evaluation of winter wheat LAI estimates for the four models developed and SNAP at the DEMMIN site are shown in Figure 6. For the evaluation, 395 pairs of in situ measurements and S2-based estimates were available. With the exception of PLS regression, all MISPEL$_{S2}$-based models outperformed the prediction of SNAP-based LAI estimation. The best estimate was obtained by GP regression, which achieved an RSQ value of 0.63, an NRMSE of 0.32, an RMSE of 1.16, and a BIAS of 0.49. An approximately good estimate was achieved by GLMNET. RF achieved inferior results and, in particular, a very low BIAS. SNAP revealed an RSQ of 0.57, an NRMSE of 0.38, an RMSE of 1.41, and the overall lowest BIAS value of 0.47. The PLS method showed the lowest accuracy values. In general, estimates of orange and reddish dots representing late recording times tended to be too low compared to actual measured values. A detailed examination of the most negatively deviated estimates revealed that artifacts of inadequate cloud masking were present in all corresponding S2 scenes. What is particularly striking is an overall underestimation of LAI present in the case of the RF model, where estimates do not exceed a value of 5.5. Some particularly large overestimates of LAI occurred in the SNAP estimate at DOY 158 on field 234-00.

Figure 7 shows a map series of S2-derived winter wheat LAI estimates for one of the test fields at the DEMMIN site predicted by GP regression. Maps showing LAI estimates at a high spatiotemporal resolution provide timely insights into the differences in crop development. Here, a strong site-specific plant development is evident from high spatial variability in LAI estimates; this is particularly seen in the observed phenological phases of stem elongation in late April and May and heading in early June.
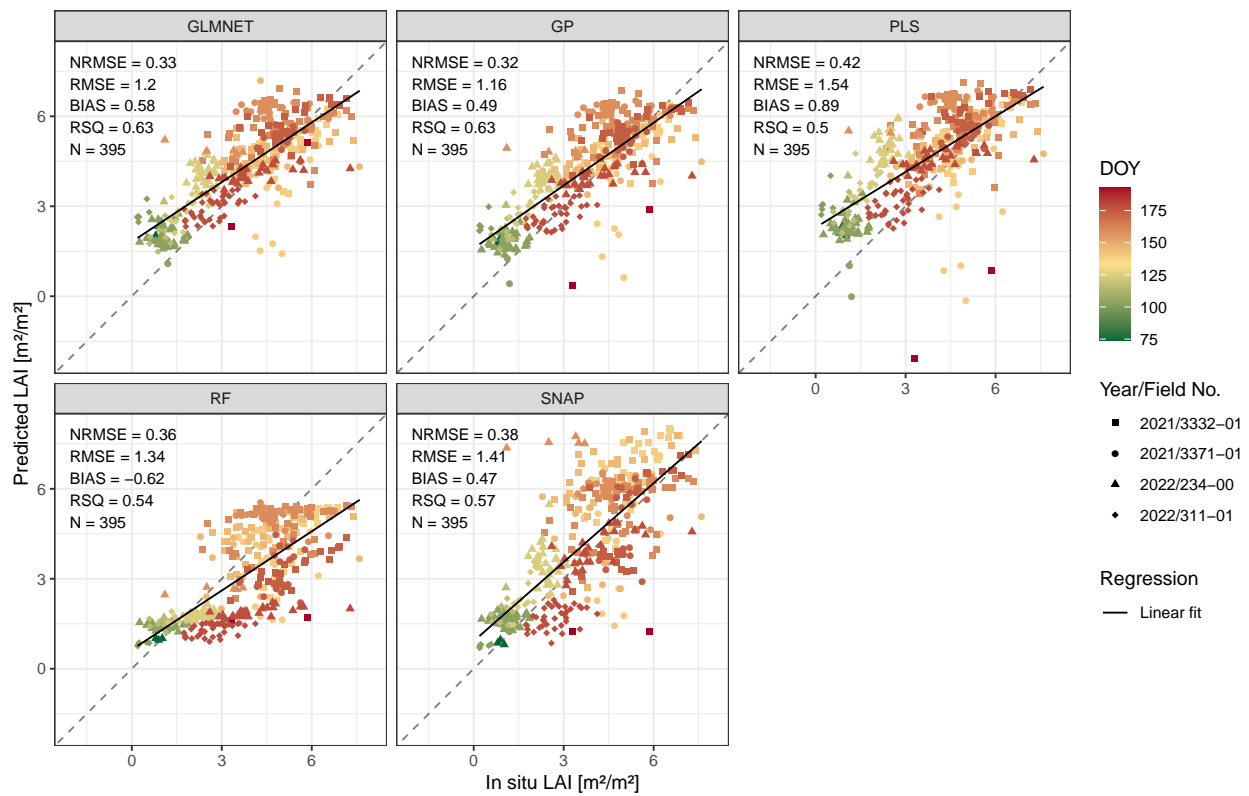
**Figure 6.** Independent validation results for winter wheat leaf area index (LAI) derived from Sentinel-2 (S2) imagery. Model accuracy metrics of the four established statistical models and SNAP versus in situ measurements from field campaigns at the DEMMIN test site in 2021 and 2022. Point colour represents the day of the year (DOY) when in situ sampling was conducted. Dashed lines indicate an ideal (1:1) relationship.
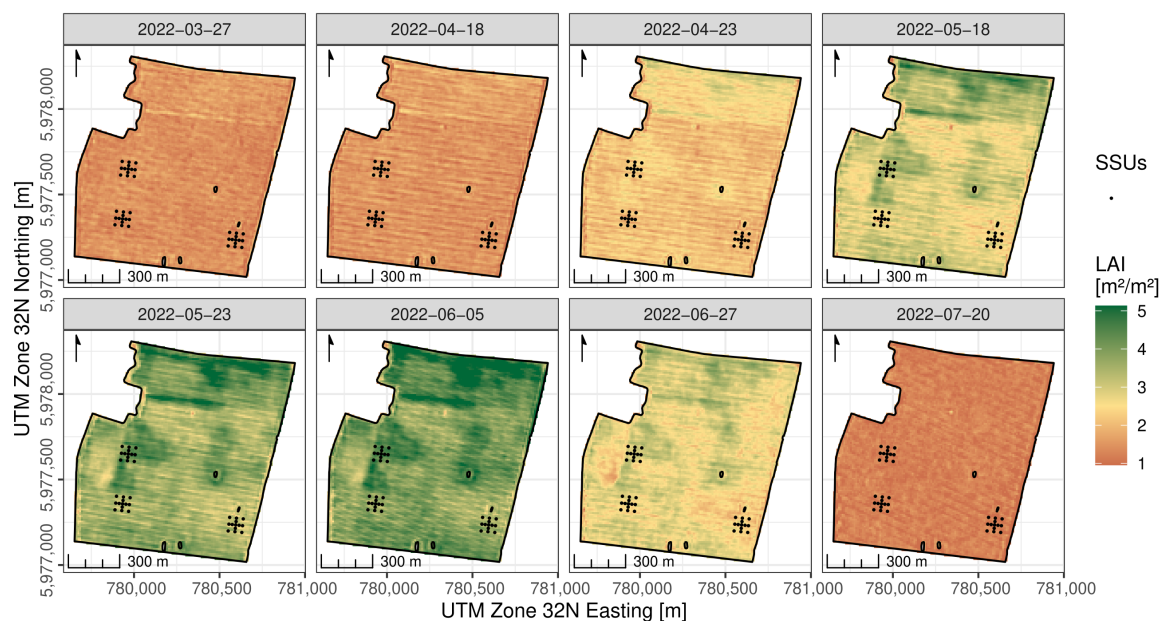


**Figure 7.** Sentinel-2 (S2) winter wheat leaf area index (LAI) estimate series of test field 311-01 at the DEMMIN test site at eight dates in 2022 modeled using MISPEL$_{S2}$ and Gaussian process (GP) regression. All available cloud-free S2 scenes from March to end of July were considered. Earlier estimates were discarded due to strong soil influence and associated unrealistic estimates. Points depict the location of secondary sampling units (SSUs).

## 4. Discussion

### 4.1. Multi-Crop Spectral Library MISPEL and Model Establishment

The introduced multi-crop spectral library MISPEL represents a comprehensive and consistently sampled in situ data collection of 10 widely grown crops. Crop-specific reference measurements of BBCH, biomass, height, LAI, nitrogen content, water content, and associated hyperspectral measurements from multiple growing seasons and regions of Germany provide a solid foundation for remote-sensing-based retrieval of key crop traits. Due to the SpecLib's hyperspectral resolution, arbitrary sensor configuration from the multispectral to the hyperspectral domain can be simulated through resampling of spectra to exploit the library for crop trait derivation. For this, any sensor with bands detecting light within the range of MISPEL's spectra, i.e., between 350 and 2500 nm, can be taken into consideration. Compared to the SpecLibs of [17,19,29,33,37], MISPEL stands out with a particularly extensive in situ database regarding crop-related traits over the entire phenological cycle. The large scope of MISPEL regarding covered crop types and traits, multi-annual spectral and reference measurements along phenological cycles, and spectral resolution enables the creation of robust crop-specific trait models.

The results in Section 3.1 demonstrate that crop-specific models lead to a significant increase in performance compared to the overall crop model. The comprehensive list of crops covered by MISPEL represents the greatest strength of the presented methodology compared to generic approaches such as SNAP, which do not incorporate crop differentiation [61].

### 4.2. Model Validation

Subsequently, we discuss model validation results obtained in light of the recent literature. In addition to comparing model validation metrics, attention is being paid to contextualizing performed validation procedures, i.e., how validations have been carried out (e.g., cross-validation, external validation), putting our results into perspective.

A similar study aiming at developing statistical LAI models for winter wheat based on hyperspectral field measurements was conducted by [7]. Here, the best model reached an RSQ of 0.94 and an RMSE of 0.32. This quality of fit surpasses the best model shown in Table 2 (MISPEL$_{FR}$ with GP). However, considering that Siegmann's result is based on a leave-one-out cross-validation using 124 LAI measurements from two years and two phenological stages only, caution must be taken regarding model transferability. In fact, Siegmann's model performance deteriorated to an RSQ of 0.77 to 0.91 and an RMSE of 0.40 to 0.65 when the model was further validated based on a threefold cross-validation across both years. The study of [81] compared the retrieval of herbaceous plant traits from canopy spectra through RTM inversion and statistical modeling based on hyperspectral field measurements. Fivefold cross-validated LAI PLS regression repeated ten times revealed an RSQ of 0.57, an NRMSE of 0.16 and an RMSE of 1.00. RTM inversion approaches yielded RSQs of 0.33 to 0.52, NRMSEs of 0.26 to 0.29, and RMSEs of 1.58 to 1.79. Phenology had a significant impact on LAI predictability. When flowering canopies were included into cross-validation, the LAI prediction accuracy decreased to RSQ = 0.41, NRMSE = 0.17, and RMSE = 1.20 for PLS regression and RSQ = 0.01 to 0.06, NRMSE = 0.28 to 0.34 and RMSE = 1.91 to 2.37 for inverted RTM approaches. Generally, given the cross-validated results of the presented hyperspectral MISPEL LAI models over multiple phenological phases, years, and regions, the model is assumed to perform reasonably well and is equally transferable.

Many studies have investigated the use of multispectral S2 imagery for LAI retrieval. A series of VIs for parametric winter wheat trait LAI retrieval was evaluated by [11]. The capability of VIs for deriving LAI was obtained by fitting linear and exponential regression models to LAI field measurements. A simple ratio index from the S2 bands of B7 and B6 turned out to be the best-performing VI for LAI estimation, resulting in an NRMSE of 0.45 and an RMSE of 1.00. Thus, the relative error is higher than for estimations using MISPEL and GP regression (NRMSE = 0.32), with a slightly lower RMSE than the MISPEL approach (RMSE = 1.16). VIs, a physical approach, and a hybrid approach for

LAI retrieval were compared against SNAP by [10]. The best VI was validated against in situ LAI measurements in a leave-one-out cross-validation procedure revealing an RSQ of 0.68 and an RMSE of 0.94. The physical and hybrid approaches showed inferior results (RSQ = 0.50, RMSE = 1.60 and RSQ = 0.50, RMSE = 1.35). SNAP estimates yielded an accuracy of RSQ = 0.62 and RMSE = 1.53. Compared to these results, the MISPEL approach does not perform quite as well as the empirical approach, but performs better than the physical and hybrid approaches. Highly accurate validation results (RSQ = 0.92, RMSE = 0.43), exceeding those of the empirical approach with MISPEL, were achieved by the hybrid approach focusing on winter wheat trait retrieval of [12]. The same applies to the hybrid approach of eight different crop types using GP and VHGP regressions and RTM simulations; S2 top-of-atmosphere data were assessed by [82], where estimates were validated against independent LAI measurements, with GP (RSQ = 0.78, NRMSE = 0.11, RMSE = 0.61) and VHGP (RSQ = 0.80, NRMSE = 0.11, RMSE = 0.57) showing the best performances across all eight crops. As in this study, an additional validation of retrieved LAI maps against SNAP LAI was performed and revealed overestimation of LAI in dense vegetation. The hybrid approach of [9] was validated with an RSQ of 0.78, an NRMSE of 0.19, and an RMSE of 0.68. The extent to which these results are comparable with MISPEL's independent validation is, however, limited, since the studies often include validation data into the actual model calibration process. This holds equally true for the studies of, e.g., [10–12]. Given the fully independent validation of trained LAI models by applying them in a spatiotemporally separated fashion presented in this study suggests that the modeling approach using MISPEL data shows a considerable degree of transferability.

The MISPEL$_{FR}$ LAI models showed only slightly higher accuracy values compared to the MISPEL$_{S2}$ models. Therefore, it is assumed that the considered LAI trait can in fact be well described by the high specificity of S2 bands [27] compared to MISPEL$_{FR}$ models. Through resampling, MISPEL is adjustable and flexible for crop monitoring applications relying on versatile multispectral optical EO satellites besides the S2 flotilla. Assuming that the traits covered by MISPEL can be adequately described by multispectral sensing, there is further high exploitation potential from spatiotemporal fusion of different EO missions, e.g., S2 and Landsat-8 [83] or S2 and Sentinel-3 [84].

The particular good performance of the GP method in terms of model accuracy in independent validation (Section 3.3) supports the findings of [39,52], highlighting the superior accuracy of kernel-based GP regression towards other ML algorithms for crop trait retrieval. Still, GP regression did not exclusively produce the best results during model calibration of the remaining crops covered by MISPEL (Appendix A). Interestingly, the linear GLMNET method that couples Ridge and Lasso regression as promoted by [50,51,85] yielded only slightly lower accuracy results compared to GP when applying the GLMNET model to S2 imagery. The independent validation further reveals that three of the developed models, namely GP, GLMNET, and PLS, using MISPEL outperformed SNAP in estimating LAIs of winter wheat (Figure 6). The main difference of the presented methodology is that the three models were trained crop-specifically using MISPEL's distinct hyperspectral field and reference measurements, whereas SNAP's underlying ANN has been pre-trained using PROSAIL-simulated spectra, and LAI measurements did not take into account crop identity [61]. In some cases, SNAP estimates deviated strongly from LAI values measured in situ. The fact that SNAP's ANN lacked goodness of fit could result from large variability inherent to the underlying training data. In contrast, the low BIAS of SNAP estimates can be interpreted as an indication of the ANN's capability to generalize reasonably well. The approach presented in this study is therefore judged to be particularly suited for trait retrieval applications ranging from local to regional or even supra-regional scales focusing on the 10 crops covered by MISPEL.

A close examination of outliers present in Figure 6 revealed that the remaining cloud artifacts resulting from insufficient S2 L2A scene classification affected S2-based LAI estimates, leading to an increase in model prediction errors. Thus, implausible value ranges of the satellite input data contaminated by cloud remnants may impact overall the model

performance. PLS regression turned out to be most sensitive in this regard. Consequently, a more conservative cloud masking through the use of alternative atmospheric correction algorithms such as the Framework for Operational Radiometric Correction for Environmental monitoring (FORCE) [86] should positively affect the quality of trait estimates regardless of the model used, yet may reduce observational density of S2 time series. The striking limitation of RF regression to predict high in situ LAI values could be attributed to the tree-based algorithm's susceptibility with regard to covariate shifts described by [87].

### 4.3. Practical Use

Information on current crop trait development potentially provides an added value for farmers, supporting them in making targeted management decisions towards more sustainable precision farming, improved resource efficiency, and cost–benefit ratios. For instance, crop trait maps may aid in determining site- and time-specific requirements for fertilizer and plant protection applications, rather than using blanket approaches [88]. LAI maps as provided in Section 3.3 reveal site heterogeneities during critical phenological phases, which can be taken into account when planning the application of agricultural resources such as fertilizers. Crop traits such as LAI and biomass derived from EO satellites further play a role as input variables for crop growth models that estimate crop yields at high resolutions, taking into account site-specific variability [89]. As described in Section 4.1, the approach presented in this study provides the flexibility to fit models based on spectral data adjusted to arbitrary optical sensor configurations and even integrated EO products such as Landsat and S2 imagery fused by applying FORCE [86]. Fused EO products ensure condensation of the time series, which provides regular data availability despite eventual cloud cover.

In terms of model applicability, an advantage of linear nonparametric approaches such as PLSR and GLMNET is their simple output structure in the form of linear regression coefficients for each of the spectral bands involved. They can be easily applied to external satellite imagery without the need for pre-trained models or re-training these models with only slightly reduced performance. The ability to transfer the models by a linear combination of model coefficients and satellite data renders an application of these models to be very comfortable.

### 4.4. Caveats

Despite the collection of SpecLib measurements in multiple seasons and regions, the transferability of the presented approach remains limited to the crops covered by MISPEL. MISPEL still shows gaps in data collection along the phenological cycle for some crops such as oat, potato, and winter barley. Increasingly, these gaps need to be closed during future sampling in order to allow precise trait retrieval at the corresponding developmental stages.

Generally, physical approaches are considered to be more transferable than empirical approaches because of their mechanistic structure. The nature of the SpecLib presented here is that the data acquisition is performed in different environments over several years and the entire phenological cycle. This, however, inheres great potential for the transferability in a larger geographical context. Yet, this remains to be explored in greater depth.

Many machine learning methods suffer from spectral collinearity of hyperspectral inputs. Given the simplicity of the approach presented here, no dimensionality reduction was performed prior to model building based on MISPEL$_{FR}$. A combination of ML algorithms and dimension reduction techniques could further improve estimation in hyperspectral applications [39,90]. Other than RTM-based approaches, statistical retrieval methods do not account for anisotropies such as the bidirectional reflectance distribution function [60].

### 4.5. Outlook

With steady expansion of the introduced multi-crop SpecLib MISPEL, phenology-specific modeling becomes possible. Taking into account phenological crop development, the specificity of this approach will further increase by modeling stage-dependent crop trait estimates along phenological cycles. Apart from estimating LAI, MISPEL and the applied

methodology enable the mapping of further crop traits. Given MISPEL's consistent data collection, the use of a multi-trait model that accounts for trait covariance, as provided by [91], is promising. Furthermore, the use of MISPEL to derive crop traits from hyperspectral missions such as EnMAP [23] and CHIME [26], but also from fused sensor products from S2, Landsat, and Sentinel-3, will be pursued. According to the statistical nature of the presented approach, spatial transferability remains to be evaluated in greater spatial and temporal contexts to prove its validity in other regions. Large-scale crop maps [92–94] can serve as a foundation for this. Following the proof of MISPEL's usability for large-scale use cases, a data publication is considered, making the spectral library openly available. Using MISPEL in a hybrid approach, including inversion methods of an RTM such as PROSAIL [57], seems particularly promising to increase transferability. Here, crop-specific parameterization and the creation of LUTs based on SpecLib references can be applied. A hybrid use would further pave the way for the integration of soil spectra into modeling and thus improve early season trait retrieval. Further, the assimilation of remotely sensed crop traits into crop growth models facilitates seamless crop monitoring and yield forecasting [1,95,96]. As the model validation revealed, S2-based crop trait estimates are strongly influenced by cloud artifacts of the L2A scene classification. The impact of more conservative cloud masks as provided by FORCE data available on the CODE-DE platform on the quality of trait estimates remains to be investigated.

## 5. Conclusions

In this study, MISPEL, a SpecLib that addresses the deficiency of in situ data for establishing crop-specific models to derive key crop traits from satellite imagery, was introduced. To date, MISPEL covers a set of 1411 spectra of 10 crops and more than 6800 related reference measurements of six crop traits. Based on MISPEL, four nonparametric S2 models were trained to explore their capability for LAI retrieval of winter wheat, one of the most important crops cultivated in Germany. GP regression turned out to be the best-performing model based on MISPEL$_{FR}$ (RSQ = 0.86, NRMSE = 0.20, RMSE = 0.70). Using MISPEL$_{S2}$, GLMNET regression performed best (RSQ = 0.83, NRMSE = 0.22, RMSE = 0.79). The models were independently validated and compared with the hybrid SNAP-LAI product using in situ measurements gathered during field campaigns at the DEMMIN test site in Germany. The evaluation revealed that three models developed superseded SNAP in terms of LAI estimation accuracy. GP and GLMNET performed particularly well at this task. We therefore conclude that MISPEL and the straightforward modeling methodology applied enable the mapping of crop-specific traits, leveraging the use of EO data provided, e.g., by S2 satellites in support of resource-efficient and sustainable agriculture.

**Author Contributions:** Conceptualization, P.B. (Peter Borrmann), P.B. (Patric Brandt) and H.G.; methodology, P.B. (Peter Borrmann), P.B. (Patric Brandt) and H.G.; software, P.B. (Peter Borrmann) and P.B. (Patric Brandt); validation, P.B. (Peter Borrmann); formal analysis, P.B. (Peter Borrmann); investigation, P.B. (Peter Borrmann); data curation, P.B. (Peter Borrmann); writing—original draft preparation, P.B. (Peter Borrmann); visualization, P.B. (Peter Borrmann); supervision, H.G.; project administration, H.G.; funding acquisition, H.G. All authors have read and agreed to the published version of the manuscript.

**Data Availability Statement:** Publicly available Sentinel-2 data was analysed in this study and can be accessed via the Copernicus Open Access Hub (https://scihub.copernicus.eu/, accessed on 14 July 2023). Following the comprehensive examination and proof of MISPEL's usability for large-scale use cases, a data publication is considered, making the spectral library openly available.

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

## Appendix A. Fivefold Cross-Validated Crop-Specific Leaf Area Index (LAI) Model Metrics Based on MISPEL

| Crop | Abbreviation | MISPEL | Model | NRMSE | RMSE | RSQ | N |
|------|-------------|--------|-------|-------|------|-----|---|
| All | ALL | FR | GLMNET | 0.28 | 1.02 | 0.7 | 1411 |
| All | ALL | FR | GP | 0.25 | 0.9 | 0.76 | 1411 |
| All | ALL | FR | PLS | 0.29 | 1.06 | 0.67 | 1411 |
| All | ALL | FR | RF | 0.25 | 0.89 | 0.77 | 1411 |
| All | ALL | S2 | GLMNET | 0.3 | 1.1 | 0.65 | 1411 |
| All | ALL | S2 | GP | 0.3 | 1.1 | 0.65 | 1411 |
| All | ALL | S2 | PLS | 0.31 | 1.12 | 0.63 | 1411 |
| All | ALL | S2 | RF | 0.26 | 0.92 | 0.75 | 1411 |
| Broad bean | BB | FR | GLMNET | 0.24 | 0.9 | 0.78 | 49 |
| Broad bean | BB | FR | GP | 0.26 | 0.99 | 0.76 | 49 |
| Broad bean | BB | FR | PLS | 0.25 | 0.95 | 0.76 | 49 |
| Broad bean | BB | FR | RF | 0.22 | 0.85 | 0.8 | 49 |
| Broad bean | BB | S2 | GLMNET | 0.22 | 0.82 | 0.82 | 49 |
| Broad bean | BB | S2 | GP | 0.24 | 0.91 | 0.78 | 49 |
| Broad bean | BB | S2 | PLS | 0.24 | 0.92 | 0.79 | 49 |
| Broad bean | BB | S2 | RF | 0.22 | 0.84 | 0.8 | 49 |
| Oat | OA | FR | GLMNET | 0.17 | 0.5 | 0.76 | 39 |
| Oat | OA | FR | GP | 0.19 | 0.56 | 0.73 | 39 |
| Oat | OA | FR | PLS | 0.19 | 0.57 | 0.73 | 39 |
| Oat | OA | FR | RF | 0.22 | 0.66 | 0.59 | 39 |
| Oat | OA | S2 | GLMNET | 0.19 | 0.57 | 0.71 | 39 |
| Oat | OA | S2 | GP | 0.19 | 0.57 | 0.71 | 39 |
| Oat | OA | S2 | PLS | 0.19 | 0.57 | 0.72 | 39 |
| Oat | OA | S2 | RF | 0.2 | 0.6 | 0.67 | 39 |
| Potato | PT | FR | GLMNET | 0.32 | 0.93 | 0.68 | 54 |
| Potato | PT | FR | GP | 0.41 | 1.2 | 0.52 | 54 |
| Potato | PT | FR | PLS | 0.35 | 1.02 | 0.63 | 54 |
| Potato | PT | FR | RF | 0.32 | 0.94 | 0.7 | 54 |
| Potato | PT | S2 | GLMNET | 0.36 | 1.04 | 0.58 | 54 |
| Potato | PT | S2 | GP | 0.36 | 1.06 | 0.59 | 54 |
| Potato | PT | S2 | PLS | 0.36 | 1.06 | 0.59 | 54 |
| Potato | PT | S2 | RF | 0.32 | 0.92 | 0.72 | 54 |
| Spring barley | SBA | FR | GLMNET | 0.24 | 0.77 | 0.66 | 75 |
| Spring barley | SBA | FR | GP | 0.23 | 0.75 | 0.66 | 75 |
| Spring barley | SBA | FR | PLS | 0.27 | 0.89 | 0.53 | 75 |
| Spring barley | SBA | FR | RF | 0.19 | 0.6 | 0.81 | 75 |
| Spring barley | SBA | S2 | GLMNET | 0.26 | 0.85 | 0.52 | 75 |
| Spring barley | SBA | S2 | GP | 0.27 | 0.89 | 0.47 | 75 |
| Spring barley | SBA | S2 | PLS | 0.27 | 0.89 | 0.5 | 75 |
| Spring barley | SBA | S2 | RF | 0.22 | 0.71 | 0.72 | 75 |
| Sugar beet | SBE | FR | GLMNET | 0.25 | 0.76 | 0.6 | 106 |
| Sugar beet | SBE | FR | GP | 0.3 | 0.91 | 0.5 | 106 |
| Sugar beet | SBE | FR | PLS | 0.26 | 0.77 | 0.59 | 106 |
| Sugar beet | SBE | FR | RF | 0.27 | 0.81 | 0.55 | 106 |
| Sugar beet | SBE | S2 | GLMNET | 0.26 | 0.77 | 0.59 | 106 |
| Sugar beet | SBE | S2 | GP | 0.26 | 0.78 | 0.57 | 106 |
| Sugar beet | SBE | S2 | PLS | 0.27 | 0.81 | 0.56 | 106 |
| Sugar beet | SBE | S2 | RF | 0.26 | 0.78 | 0.57 | 106 |
| Triticale | TR | FR | GLMNET | 0.24 | 0.8 | 0.82 | 66 |
| Triticale | TR | FR | GP | 0.28 | 0.92 | 0.78 | 66 |
| Triticale | TR | FR | PLS | 0.25 | 0.84 | 0.8 | 66 |

| Crop | Code | Sensor | Model | | | | N |
|---|---|---|---|---|---|---|---|
| Triticale | TR | FR | RF | 0.27 | 0.89 | 0.78 | 66 |
| Triticale | TR | S2 | GLMNET | 0.23 | 0.78 | 0.83 | 66 |
| Triticale | TR | S2 | GP | 0.24 | 0.79 | 0.83 | 66 |
| Triticale | TR | S2 | PLS | 0.24 | 0.79 | 0.82 | 66 |
| Triticale | TR | S2 | RF | 0.25 | 0.84 | 0.81 | 66 |
| Winter barley | WB | FR | GLMNET | 0.2 | 0.8 | 0.77 | 33 |
| Winter barley | WB | FR | GP | 0.22 | 0.86 | 0.75 | 33 |
| Winter barley | WB | FR | PLS | 0.2 | 0.81 | 0.75 | 33 |
| Winter barley | WB | FR | RF | 0.21 | 0.83 | 0.76 | 33 |
| Winter barley | WB | S2 | GLMNET | 0.18 | 0.74 | 0.83 | 33 |
| Winter barley | WB | S2 | GP | 0.19 | 0.75 | 0.79 | 33 |
| Winter barley | WB | S2 | PLS | 0.19 | 0.75 | 0.79 | 33 |
| Winter barley | WB | S2 | RF | 0.22 | 0.87 | 0.7 | 33 |
| Winter rapeseed | WRA | FR | GLMNET | 0.22 | 0.97 | 0.76 | 426 |
| Winter rapeseed | WRA | FR | GP | 0.2 | 0.88 | 0.81 | 426 |
| Winter rapeseed | WRA | FR | PLS | 0.25 | 1.07 | 0.71 | 426 |
| Winter rapeseed | WRA | FR | RF | 0.21 | 0.92 | 0.79 | 426 |
| Winter rapeseed | WRA | S2 | GLMNET | 0.23 | 0.99 | 0.75 | 426 |
| Winter rapeseed | WRA | S2 | GP | 0.23 | 0.99 | 0.75 | 426 |
| Winter rapeseed | WRA | S2 | PLS | 0.24 | 1.03 | 0.73 | 426 |
| Winter rapeseed | WRA | S2 | RF | 0.22 | 0.96 | 0.77 | 426 |
| Winter rye | WRY | FR | GLMNET | 0.22 | 0.63 | 0.83 | 157 |
| Winter rye | WRY | FR | GP | 0.17 | 0.5 | 0.9 | 157 |
| Winter rye | WRY | FR | PLS | 0.22 | 0.64 | 0.82 | 157 |
| Winter rye | WRY | FR | RF | 0.17 | 0.49 | 0.9 | 157 |
| Winter rye | WRY | S2 | GLMNET | 0.19 | 0.54 | 0.87 | 157 |
| Winter rye | WRY | S2 | GP | 0.21 | 0.62 | 0.84 | 157 |
| Winter rye | WRY | S2 | PLS | 0.22 | 0.63 | 0.83 | 157 |
| Winter rye | WRY | S2 | RF | 0.18 | 0.51 | 0.89 | 157 |
| Winter wheat | WW | FR | GLMNET | 0.22 | 0.78 | 0.83 | 406 |
| Winter wheat | WW | FR | GP | 0.2 | 0.7 | 0.86 | 406 |
| Winter wheat | WW | FR | PLS | 0.23 | 0.82 | 0.81 | 406 |
| Winter wheat | WW | FR | RF | 0.22 | 0.77 | 0.83 | 406 |
| Winter wheat | WW | S2 | GLMNET | 0.22 | 0.79 | 0.83 | 406 |
| Winter wheat | WW | S2 | GP | 0.23 | 0.79 | 0.83 | 406 |
| Winter wheat | WW | S2 | PLS | 0.23 | 0.81 | 0.82 | 406 |
| Winter wheat | WW | S2 | RF | 0.23 | 0.79 | 0.83 | 406 |

## References

1. Huang, J.; Gómez-Dans, J.L.; Huang, H.; Ma, H.; Wu, Q.; Lewis, P.E.; Liang, S.; Chen, Z.; Xue, J.H.; Wu, Y.; et al. Assimilation of remote sensing into crop growth models: Current status and perspectives. *Agric. For. Meteorol.* **2019**, *276–277*, 107609. [CrossRef]
2. Gitelson, A.A.; Gritz, Y.; Merzlyak, M.N. Relationships between leaf chlorophyll content and spectral reflectance and algorithms for non-destructive chlorophyll assessment in higher plant leaves. *J. Plant Physiol.* **2003**, *160*, 271–282. [CrossRef]
3. Darvishzadeh, R.; Skidmore, A.; Schlerf, M.; Atzberger, C. Inversion of a radiative transfer model for estimating vegetation LAI and chlorophyll in a heterogeneous grassland. *Remote Sens. Environ.* **2008**, *112*, 2592–2604. [CrossRef]
4. Zheng, G.; Moskal, L.M. Retrieving Leaf Area Index (LAI) Using Remote Sensing: Theories, Methods and Sensors. *Sensors* **2009**, *9*, 2719–2745. [CrossRef] [PubMed]
5. Delegido, J.; Verrelst, J.; Alonso, L.; Moreno, J. Evaluation of Sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. *Sensors* **2011**, *11*, 7063–7081. [CrossRef] [PubMed]
6. Gerighausen, H.; Lilienthal, H.; Jarmer, T.; Siegmann, B. Evaluation of leaf area index and dry matter predictions for crop growth modelling and yield esti-mation based on field reflectance measurements. In Proceedings of the 9th EARSeL Imaging Spectroscopy Workshop, Luxembourg, 14–16 April 2015; Volume 14, pp. 71–90.
7. Siegmann, B.; Jarmer, T. Comparison of different regression models and validation techniques for the assessment of wheat leaf area index from hyperspectral data. *Int. J. Remote Sens.* **2015**, *36*, 4519–4534. [CrossRef]
8. Verrelst, J.; Rivera, J.P.; Veroustraete, F.; Muñoz-Marí, J.; Clevers, J.G.; Camps-Valls, G.; Moreno, J. Experimental Sentinel-2 LAI estimation using parametric, non-parametric and physical retrieval methods—A comparison. *ISPRS J. Photogramm. Remote Sens.* **2015**, *108*, 260–272. [CrossRef]
9. Upreti, D.; Huang, W.; Kong, W.; Pascucci, S.; Pignatti, S.; Zhou, X.; Ye, H.; Casa, R. A Comparison of Hybrid Machine Learning Algorithms for the Retrieval of Wheat Biophysical Variables from Sentinel-2. *Remote Sens.* **2019**, *11*, 481. [CrossRef]
10. Xie, Q.; Jadu, D.; Huete, A.; Jiang, A.; Yin, G.; Ding, Y.; Peng, D.; C. Hall, C.; Brown, L.; Shi, Y.; et al. Retrieval of crop biophysical parameters from Sentinel-2 remote sensing imagery. *Int. J. Appl. Earth Obs. Geoinf.* **2019**, *80*, 187–195. [CrossRef]
11. Kamenova, I.; Dimitrov, P. Evaluation of Sentinel-2 vegetation indices for prediction of LAI, fAPAR and fCover of winter wheat in Bulgaria. *Eur. J. Remote Sens.* **2021**, *54*, 89–108. [CrossRef]

12. Caballero, G.; Pezzola, A.; Winschel, C.; Casella, A.; Angonova, P.S.; Rivera-Caicedo, J.P.; Berger, K.; Verrelst, J.; Delegido, J. Seasonal Mapping of Irrigated Winter Wheat Traits in Argentina with a Hybrid Retrieval Workflow Using Sentinel-2 Imagery. *Remote Sens.* **2022**, *14*, 4531. [CrossRef]

13. Hansen, P.M.; Schjoerring, J.K. Reflectance measurement of canopy biomass and nitrogen status in wheat crops using normalized difference vegetation indices and partial least squares regression. *Remote Sens. Environ.* **2003**, *86*, 542–553. [CrossRef]

14. Wocher, M.; Berger, K.; Verrelst, J.; Hank, T. Retrieval of carbon content and biomass from hyperspectral imagery over cultivated areas. *ISPRS J. Photogramm. Remote Sens.* **2022**, *193*, 104–114. [CrossRef]

15. Clevers, J.G.P.W.; Kooistra, L. Using Hyperspectral Remote Sensing Data for Retrieving Canopy Chlorophyll and Nitrogen Content. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2012**, *5*, 574–583. [CrossRef]

16. Liang, L.; Di, L.; Huang, T.; Wang, J.; Lin, L.; Wang, L.; Yang, M. Estimation of Leaf Nitrogen Content in Wheat Using New Hyperspectral Indices and a Random Forest Regression Algorithm. *Remote Sens.* **2018**, *10*, 1940. [CrossRef]

17. Perich, G.; Aasen, H.; Verrelst, J.; Argento, F.; Walter, A.; Liebisch, F. Crop Nitrogen Retrieval Methods for Simulated Sentinel-2 Data Using In-Field Spectrometer Data. *Remote Sens.* **2021**, *13*, 2404. [CrossRef] [PubMed]

18. Yue, J.; Yang, G.; Li, C.; Li, Z.; Wang, Y.; Feng, H.; Xu, B. Estimation of Winter Wheat Above-Ground Biomass Using Unmanned Aerial Vehicle-Based Snapshot Hyperspectral Sensor and Crop Height Improved Models. *Remote Sens.* **2017**, *9*, 708. [CrossRef]

19. Prey, L.; Schmidhalter, U. Simulation of satellite reflectance data using high-frequency ground based hyperspectral canopy measurements for in-season estimation of grain yield and grain nitrogen status in winter wheat. *ISPRS J. Photogramm. Remote Sens.* **2019**, *149*, 176–187. [CrossRef]

20. Grote, U.; Fasse, A.; Nguyen, T.T.; Erenstein, O. Food Security and the Dynamics of Wheat and Maize Value Chains in Africa and Asia. *Front. Sustain. Food Syst.* **2021**, *4*, 617009. [CrossRef]

21. FAO. *Crops and Livestock Products*; FAO: Rome, Italy, 2023.

22. Bundesministerium für Ernährung und Landwirtschaft. *Daten und Fakten: Land-, Forst- und Ernährungswirtschaft mit Fischerei und Wein- und Gartenbau*; Bundesministerium für Ernährung und Landwirtschaft: Berlin, Germany, 2022.

23. Guanter, L.; Kaufmann, H.; Segl, K.; Foerster, S.; Rogass, C.; Chabrillat, S.; Kuester, T.; Hollstein, A.; Rossner, G.; Chlebek, C.; et al. The EnMAP Spaceborne Imaging Spectroscopy Mission for Earth Observation. *Remote Sens.* **2015**, *7*, 8830–8857. [CrossRef]

24. Lee, C.M.; Cable, M.L.; Hook, S.J.; Green, R.O.; Ustin, S.L.; Mandl, D.J.; Middleton, E.M. An introduction to the NASA Hyperspectral InfraRed Imager (HyspIRI) mission and preparatory activities. *Remote Sens. Environ.* **2015**, *167*, 6–19. [CrossRef]

25. Labate, D.; Ceccherini, M.; Cisbani, A.; de Cosmo, V.; Galeazzi, C.; Giunti, L.; Melozzi, M.; Pieraccini, S.; Stagi, M. The PRISMA payload optomechanical design, a high performance instrument for a new hyperspectral mission. *Acta Astronaut.* **2009**, *65*, 1429–1436. [CrossRef]

26. Nieke, J.; Rast, M. Towards the Copernicus Hyperspectral Imaging Mission For The Environment (CHIME). In Proceedings of the IGARSS, Valencia, Spain, 22–27 July 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 157–159. [CrossRef]

27. Drusch, M.; Del Bello, U.; Carlier, S.; Colin, O.; Fernandez, V.; Gascon, F.; Hoersch, B.; Isola, C.; Laberinti, P.; Martimort, P.; et al. Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sens. Environ.* **2012**, *120*, 25–36. [CrossRef]

28. Teucher, M.; Thürkow, D.; Alb, P.; Conrad, C. Digital In Situ Data Collection in Earth Observation, Monitoring and Agriculture—Progress towards Digital Agriculture. *Remote Sens.* **2022**, *14*, 393. [CrossRef]

29. Rao, N.R. Development of a crop–specific spectral library and discrimination of various agricultural crop varieties using hyperspectral imagery. *Int. J. Remote Sens.* **2008**, *29*, 131–144. [CrossRef]

30. Baldridge, A.M.; Hook, S.J.; Grove, C.I.; Rivera, G. The ASTER spectral library version 2.0. *Remote Sens. Environ.* **2009**, *113*, 711–715. [CrossRef]

31. Hueni, A.; Nieke, J.; Schopfer, J.; Kneubühler, M.; Itten, K.I. The spectral database SPECCHIO for improved long-term usability and data sharing. *Comput. Geosci.* **2009**, *35*, 557–565. [CrossRef]

32. Zomer, R.J.; Trabucco, A.; Ustin, S.L. Building spectral libraries for wetlands land cover classification and hyperspectral remote sensing. *J. Environ. Manag.* **2009**, *90*, 2170–2177. [CrossRef]

33. Nidamanuri, R.R.; Zbell, B. Existence of characteristic spectral signatures for agricultural crops—Potential for automated crop mapping by hyperspectral imaging. *Geocarto Int.* **2012**, *27*, 103–118. [CrossRef]

34. Kotthaus, S.; Smith, T.E.; Wooster, M.J.; Grimmond, C. Derivation of an urban materials spectral library through emittance and reflectance spectroscopy. *Isprs J. Photogramm. Remote Sens.* **2014**, *94*, 194–212. [CrossRef]

35. Jiménez, M.; Díaz-Delgado, R. Towards a Standard Plant Species Spectral Library Protocol for Vegetation Mapping: A Case Study in the Shrubland of Doñana National Park. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 2472–2495. [CrossRef]

36. Fuchs, M.; Awan, A.A.; Akhtar, S.S.; Ahmad, I.; Sadiq, S.; Razzak, A.; Haider, N. Lithological mapping with multispectral data—Setup and application of a spectral database for rocks in the Balakot area, Northern Pakistan. *J. Mt. Sci.* **2017**, *14*, 948–963. [CrossRef]

37. Zhang, J.; Wang, C.; Yuan, L.; Liu, P.; Zhang, Y.; Wu, K. Construction of a plant spectral library based on an optimised feature selection method. *Biosyst. Eng.* **2020**, *195*, 1–16. [CrossRef]

38. de Peppo, M.; Taramelli, A.; Boschetti, M.; Mantino, A.; Volpi, I.; Filipponi, F.; Tornato, A.; Valentini, E.; Ragaglini, G. Non-Parametric Statistical Approaches for Leaf Area Index Estimation from Sentinel-2 Data: A Multi-Crop Assessment. *Remote Sens.* **2021**, *13*, 2841. [CrossRef]

39. Verrelst, J.; Malenovský, Z.; van der Tol, C.; Camps-Valls, G.; Gastellu-Etchegorry, J.P.; Lewis, p.; North, P.; Moreno, J. Quantifying Vegetation Biophysical Variables from Imaging Spectroscopy Data: A Review on Retrieval Methods. *Surv. Geophys.* **2019**, *40*, 589–629. [CrossRef] [PubMed]

40. Pasqualotto, N.; Delegido, J.; van Wittenberghe, S.; Rinaldi, M.; Moreno, J. Multi-Crop Green LAI Estimation with a New Simple Sentinel-2 LAI Index (SeLI). *Sensors* **2019**, *19*, 904. [CrossRef]

41. Verrelst, J.; Muñoz, J.; Alonso, L.; Delegido, J.; Rivera, J.P.; Camps-Valls, G.; Moreno, J. Machine learning regression algorithms for biophysical parameter retrieval: Opportunities for Sentinel-2 and -3. *Remote Sens. Environ.* **2012**, *118*, 127–139. [CrossRef]

42. Wold, S.; Esbensen, K.; Geladi, P. Principal Component Analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52. [CrossRef]

43. Geladi, P.; Kowalski, B.R. Partial Least-Squares Regression: A Tutorial. *Anal. Chim. Acta* **1986**, *185*, 1–17. [CrossRef]

44. Fu, Y.; Yang, G.; Wang, J.; Song, X.; Feng, H. Winter wheat biomass estimation based on spectral indices, band depth analysis and partial least squares regression using hyperspectral measurements. *Comput. Electron. Agric.* **2014**, *100*, 51–59. [CrossRef]

45. Feilhauer, H.; Asner, G.P.; Martin, R.E. Multi-method ensemble selection of spectral bands related to leaf biochemistry. *Remote Sens. Environ.* **2015**, *164*, 57–65. [CrossRef]

46. Foster, A.J.; Kakani, V.G.; Mosali, J. Estimation of bioenergy crop yield and N status by hyperspectral canopy reflectance and partial least square regression. *Precis. Agric.* **2017**, *18*, 192–209. [CrossRef]

47. Wang, B.; Chen, J.; Ju, W.; Qiu, F.; Zhang, Q.; Fang, M.; Chen, F. Limited Effects of Water Absorption on Reducing the Accuracy of Leaf Nitrogen Estimation. *Remote Sens.* **2017**, *9*, 291. [CrossRef]

48. Hoerl, A.E.; Kennard, R.W. Ridge Regression: Applications to Nonorthogonal Problems. *Technometrics* **1970**, *12*, 69–82. . [CrossRef]

49. Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. Methodol.* **1996**, *58*, 267–288. [CrossRef]

50. Lazaridis, D.C.; Verbesselt, J.; Robinson, A.P. Penalized regression techniques for prediction: A case study for predicting tree mortality using remotely sensed vegetation indicesThis article is one of a selection of papers from Extending Forest Inventory and Monitoring over Space and Time. *Can. J. For. Res.* **2011**, *41*, 24–34. [CrossRef]

51. Zandler, H.; Brenning, A.; Samimi, C. Quantifying dwarf shrub biomass in an arid environment: Comparing empirical methods in a high dimensional setting. *Remote Sens. Environ.* **2015**, *158*, 140–155. [CrossRef]

52. Caicedo, J.P.R.; Verrelst, J.; Munoz-Mari, J.; Moreno, J.; Camps-Valls, G. Toward a Semiautomatic Machine Learning Retrieval of Biophysical Parameters. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2014**, *7*, 1249–1259. [CrossRef]

53. Lázaro-Gredilla, M.; Titsias, M.K.; Verrelst, J.; Camps-Valls, G. Retrieval of Biophysical Parameters With Heteroscedastic Gaussian Processes. *IEEE Geosci. Remote Sens. Lett.* **2013**, *11*, 2013. [CrossRef]

54. Chen, B.; Wu, Z.; Wang, J.; Dong, J.; Guan, L.; Chen, J.; Yang, K.; Xie, G. Spatio-temporal prediction of leaf area index of rubber plantation using HJ-1A/1B CCD images and recurrent neural network. *ISPRS J. Photogramm. Remote Sens.* **2015**, *102*, 148–160. [CrossRef]

55. Neinavaz, E.; Skidmore, A.K.; Darvishzadeh, R.; Groen, T.A. Retrieval of leaf area index in different plant species using thermal hyperspectral data. *ISPRS J. Photogramm. Remote Sens.* **2016**, *119*, 390–401. [CrossRef]

56. Combal, B.; Baret, F.; Weiss, M.; Trubuil, A.; Myneni, R.; Knyazikhin, Y.; Wang, L. Retrieval of canopy biophysical variables from bidirectional reflectance: Using prior information to solve the ill-posed inverse problem. *Remote Sens. Environ.* **2002**, *84*, 1–15. [CrossRef]

57. Jacquemoud, S.; Verhoef, W.; Baret, F.; Bacour, C.; Zarco-Tejada, P.J.; Asner, G.P.; François, C.; Ustin, S.L. PROSPECT+SAIL models: A review of use for vegetation characterization. *Remote Sens. Environ.* **2009**, *113*, S56–S66. [CrossRef]

58. Berger, K.; Atzberger, C.; Danner, M.; D'Urso, G.; Mauser, W.; Vuolo, F.; Hank, T. Evaluation of the PROSAIL Model Capabilities for Future Hyperspectral Model Environments: A Review Study. *Remote Sens.* **2018**, *10*, 85. [CrossRef]

59. Weiss, M.; Baret, F.; Myneni, R.B.; Pragnère, A.; Knyazikhin, Y. Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data. *Agronomie* **2000**, *20*, 3–22. .:2000105. [CrossRef]

60. Locherer, M.; Hank, T.; Danner, M.; Mauser, W. Retrieval of Seasonal Leaf Area Index from Simulated EnMAP Data through Optimized LUT-Based Inversion of the PROSAIL Model. *Remote Sens.* **2015**, *7*, 10321–10346. [CrossRef]

61. Weiss, M.; Baret, F.; Jay, S. *S2ToolBox Level 2 Products LAI, FAPAR, FCOVER*; HAL Open Science: Lyon, France, 2020.

62. Borg, E.; Lippert, K.; Zabel, E.; Löpmeier, F.J.; Fichtelmann, B.; Jahncke, D.; Maass, H. *DEMMIN—Teststandort zur Kalibrierung und Validierung von Fernerkundungsmissionen*; Rebenstorf, R.W., Ed.; Self-published: Neubrandenburg, Germany, 2009; Volume 15, pp. 401–419.

63. Zacharias, S.; Bogena, H.; Samaniego, L.; Mauder, M.; Fuß, R.; Pütz, T.; Frenzel, M.; Schwank, M.; Baessler, C.; Butterbach-Bahl, K.; et al. A Network of Terrestrial Environmental Observatories in Germany. *Vadose Zone J.* **2011**, *10*, 955–973. [CrossRef]

64. Meier, U.; Bleiholder, H.; Buhr, L.; Feller, C.; Hack, H.; Heß, M.; Lancashire, P.D.; Schock, U.; Strauß, R.; van den Boom, T.; et al. The BBCH system to coding the phenological growth stages of plants–history and publications. *J. Kult.* **2009**, *61*, 41–52.

65. VDLUFA. Bestimmung von organischem Kohlenstoff durch Verbrennung und Gasanalyse (Differenzmethode). In *Das VDLUFA Methodenbuch*; VDLUFA, Ed.; VDLUFA: Darmstadt, Germany, 1991; Volume 1.

66. VDLUFA. Bestimmung von Gesamtstickstoff nach trockener Verbrennung (Elementaranalyse). In *Das VDLUFA Methodenbuch*; VDLUFA, Ed.; VDLUFA: Darmstadt, Germany, 1991; Volume 1.

67. LI-COR Biosciences. FV2200. 2010. Available online: https://www.licor.com/env/products/leaf_area/LAI-2200C/software (accessed on 14 July 2023).

68. R Core Team. *R: A Language and Environment for Statistical Computing*; R Core Team: Vienna, Austria, 2020.

69. Lehnert, L.W.; Meyer, H.; Obermeier, W.A.; Silva, B.; Regeling, B.; Bendix, J. Hyperspectral Data Analysis in R: The hsdar Package. *J. Stat. Softw.* **2019**, *89*, 1–23. [CrossRef]

70. Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.; François, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J.; et al. Welcome to the Tidyverse. *J. Open Source Softw.* **2019**, *4*, 1686. [CrossRef]

71. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1–13. [CrossRef]

72. Rasmussen, C.E. Gaussian Processes in Machine Learning. In *Advanced Lectures on Machine Learning*; Bousquet, O., von Luxburg, U., Rätsch, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2003; pp. 63–71.

73. Karatzoglou, A.; Smola, A.; Hornik, K.; Zeileis, A. kernlab-an S4 package for kernel methods in R. *J. Stat. Softw.* **2004**, *11*, 1–20. [CrossRef]

74. Liland, K.H.; Mevik, B.H.; Wehrens, R. pls: Partial Least Squares and Principal Component Regression. 2022. Available online: https://cran.r-project.org/web/packages/pls/index.html (accessed on 14 July 2023).

75. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]

76. Wright, M.N.; Ziegler, A. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. *J. Stat. Softw.* **2017**, *77*, 1–17. [CrossRef]

77. Kuhn, M. caret: Classification and Regression Training, 2021.

78. CODE-DE. CODE-DE: The German Access to Copernicus Data. 2022. Available online: https://code-de.org/en/ (accessed on 14 July 2023).

79. Hijmans, R.J. terra: Spatial Data Analysis. 2021. Available online: https://CRAN.R-project.org/package=terra (accessed on 14 July 2023).

80. Robinson, D. fuzzyjoin: Join Tables Together on Inexact Matching. 2020. Available online: https://cran.r-project.org/package=fuzzyjoin (accessed on 14 July 2023).

81. Schiefer, F.; Schmidtlein, S.; Kattenborn, T. The retrieval of plant functional traits from canopy spectra through RTM-inversions and statistical models are both critically affected by plant phenology. *Ecol. Indic.* **2021**, *121*, 107062. [CrossRef]

82. Estevez, J.; Vicent, J.; Rivera-Caicedo, J.P.; Morcillo-Pallarés, P.; Vuolo, F.; Sabater, N.; Camps-Valls, G.; Moreno, J.; Verrelst, J. Gaussian Processes Retrieval of LAI from Sentinel-2 Top-of-Atmosphere Radiance Data. *ISPRS J. Photogramm. Remote Sens.* **2020**, *167*, 289–304. [CrossRef] [PubMed]

83. Shao, Z.; Cai, J.; Fu, P.; Hu, L.; Liu, T. Deep learning-based fusion of Landsat-8 and Sentinel-2 images for a harmonized surface reflectance product. *Remote Sens. Environ.* **2019**, *235*, 111425. [CrossRef]

84. Wang, Q.; Atkinson, P.M. Spatio-temporal fusion for daily Sentinel-2 images. *Remote Sens. Environ.* **2018**, *204*, 31–42. [CrossRef]

85. Bratsch, S.; Epstein, H.; Buchhorn, M.; Walker, D.; Landes, H. Relationships between hyperspectral data and components of vegetation biomass in Low Arctic tundra communities at Ivotuk, Alaska. *Environ. Res. Lett.* **2017**, *12*, 025003. [CrossRef]

86. Frantz, D. FORCE—Landsat + Sentinel-2 Analysis Ready Data and Beyond. *Remote Sens.* **2019**, *11*, 1124. [CrossRef]

87. Y, G.D.; Nair, N.G.; Satpathy, P.; Christopher, J. Covariate Shift: A Review and Analysis on Classifiers. In Proceedings of the Global Conference for Advancement in Technology (GCAT), Bangalore, India, 18–20 October 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6. [CrossRef]

88. Segarra, J.; Buchaillot, M.L.; Araus, J.L.; Kefauver, S.C. Remote Sensing for Precision Agriculture: Sentinel-2 Improved Features and Applications. *Agronomy* **2020**, *10*, 641. [CrossRef]

89. Kang, Y.; Özdoğan, M. Field-level crop yield mapping with Landsat using a hierarchical data assimilation approach. *Remote Sens. Environ.* **2019**, *228*, 144–163. [CrossRef]

90. Rivera-Caicedo, J.P.; Verrelst, J.; Muñoz-Marí, J.; Camps-Valls, G.; Moreno, J. Hyperspectral dimensionality reduction for biophysical variable statistical retrieval. *ISPRS J. Photogramm. Remote Sens.* **2017**, *132*, 88–101. [CrossRef]

91. Cherif, E.; Feilhauer, H.; Berger, K.; Dao, P.D.; Ewald, M.; Hank, T.B.; He, Y.; Kovach, K.R.; Lu, B.; Townsend, P.A.; et al. From spectra to plant functional traits: Transferable multi-trait models from heterogeneous and sparse data. *Remote Sens. Environ.* **2023**, *292*, 113580. [CrossRef]

92. Preidl, S.; Lange, M.; Doktor, D. Introducing APiC for regionalised land cover mapping on the national scale using Sentinel-2A imagery. *Remote Sens. Environ.* **2020**, *240*, 111673. [CrossRef]

93. Blickensdörfer, L.; Schwieder, M.; Pflugmacher, D.; Nendel, C.; Erasmi, S.; Hostert, P. Mapping of crop types and crop sequences with combined time series of Sentinel-1, Sentinel-2 and Landsat 8 data for Germany. *Remote Sens. Environ.* **2022**, *269*, 112831. [CrossRef]

94. d'Andrimont, R.; Verhegghen, A.; Lemoine, G.; Kempeneers, P.; Meroni, M.; van der Velde, M. From parcel to continental scale—A first European crop type map based on Sentinel-1 and LUCAS Copernicus in-situ observations. *Remote Sens. Environ.* **2021**, *266*, 112708. [CrossRef]

95. Dhillon, M.S.; Dahms, T.; Kübert-Flock, C.; Liepa, A.; Rummler, T.; Arnault, J.; Steffan-Dewenter, I.; Ullmann, T. Impact of STARFM on Crop Yield Predictions: Fusing MODIS with Landsat 5, 7, and 8 NDVIs in Bavaria Germany. *Remote Sens.* **2023**, *15*, 1651. [CrossRef]

96. Tewes, A.; Hoffmann, H.; Krauss, G.; Schäfer, F.; Kerkhoff, C.; Gaiser, T. New Approaches for the Assimilation of LAI Measurements into a Crop Model Ensemble to Improve Wheat Biomass Estimations. *Agronomy* **2020**, *10*, 446. [CrossRef]