From spectra to plant functional traits: Transferable multi-trait models from heterogeneous and sparse data

Eya Cherif[a,b,*], Hannes Feilhauer[a,b,c,d], Katja Berger[e,f], Phuong D. Dao[i], Michael Ewald[g], Tobias B. Hank[h], Yuhong He[j], Kyle R. Kovach[i], Bing Lu[k], Philip A. Townsend[i], Teja Kattenborn[a,c]

[a] Remote Sensing Centre for Earth System Research (RSC4Earth), Leipzig University, Talstr. 35, 04103 Leipzig, Germany
[b] Center for scalable data analytics and artificial intelligence (ScaDS.AI), Leipzig University, Humboldtstr. 25, 04105 Leipzig, Germany
[c] German Centre for Integrative Biodiversity Research (iDiv), Halle-Jena-Leipzig, Germany
[d] Helmholtz-Centre for Environmental Research (UFZ), Leipzig, Germany
[e] Image Processing Laboratory (IPL), Parc Científic, Universitsat de València, València, Spain
[f] Mantle Labs GmbH, Vienna, Austria
[g] Institute of Geography and Geoecology, Karlsruhe Institute of Technology (KIT), Kaiserstr. 12, 76131 Karlsruhe, Germany
[h] Department of Geography, Faculty of Geosciences, Ludwig-Maximilians-Universität München (LMU), Luisenstr. 37, 80333, Munich, Germany
[i] Department of Forest and Wildlife Ecology, University of Wisconsin-Madison, 1630 Linden Drive, Madison, WI 53706, USA
[j] Department of Geography, Geomatics and Environment, University of Toronto, 3359 Mississauga Road, Mississauga, ON L5L 1C6, Canada
[k] Department of Geography, Simon Fraser University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada
*Corresponding author

This is a preprint. The final and accepted article can be found in Remote Sensing of Environment (https://doi.org/10.1016/j.rse.2023.113580).

Abstract

Large-scale information on several vegetation properties ('plant traits') is critical to assess ecosystem functioning, functional diversity and their role in the Earth system. Hyperspectral remote sensing of plant canopies offers a key tool to map multiple plant traits. However, we are still lacking generalized methods to translate hyperspectral reflectance into a suite of relevant plant traits across biomes, land cover and sensor types. The absence of globally representative data sets and the gap between the available reflectance data with corresponding in-situ measurements have hampered such approaches. In recent years, the scientific community acquired multiple data sets encompassing canopy hyperspectral reflectance and plant traits from different plant types and sensors. To combine these heterogeneous data sets, we propose three multi-trait modeling approaches based on Convolutional Neural Networks (CNNs) to simultaneously infer a broad set of 20 structural and chemical traits (e.g. leaf mass per area, leaf area index, pigments, nitrogen). The performance of these multi-trait CNN models predicting these traits was

compared against single-trait CNN as well as single-trait partial least squares regression (PLSR). We found that the multi-trait CNNs performances significantly increased from single-trait CNNs (nRMSE = 0.027-19.61%) and the state-of-the-art PLSR models (nRMSE = 1.94 - 40.07%) across a broad range of vegetation types (crops, forest, tundra, grassland, shrubland) and sensor types. Thus, providing a single model for multiple traits not only proved to be computationally more efficient, but also more accurate, since it enabled the model to incorporate traits' co-variation. Despite the data heterogeneity of the merged data set, our models performances' were comparable or exceeded those of previous studies. Overall, this study highlights the potential of weakly supervised approaches to overcome the scarcity of in-situ measurements and take a step forward in creating efficient predictive models of multiple biochemical and biophysical vegetation properties.

Keywords: Hyperspectral remote sensing, Plant trait retrieval, Deep learning, Biophysical variables, Imaging spectroscopy, Canopy properties, Weakly supervised learning, Multi-task regression

1. Introduction

Plant functional traits are key for assessing and monitoring terrestrial ecosystem properties. They provide insights on functional diversity and can enhance our understanding of ecosystem functioning (Lavorel and Garnier, 2002; Migliavacca et al., 2021). Traits determine plant productivity and stress resistance and thus also how plants compete for growth and survival in different environments (Funk et al., 2017). For example, leaf mass per area (LMA) is positively related to photosynthetic productivity and negatively to structural robustness and depends on resource availability and environmental conditions (Díaz et al., 2016; Grime, 1988; Poorter et al., 2009). Leaf pigments (e.g., chlorophyll, carotenoids) determine photosynthetic capacities and their variations can indicate changes in plant health due to stress (Feret et al., 2008; Zarco-Tejada et al., 2019, 2018, Berger et al., 2022). Other leaf constituents such as nitrogen and carbon are directly linked to biosphere-atmosphere cycles (de Bello et al., 2010) and are important to parameterize vegetation in Earth system models (Yang et al., 2015). A comprehensive set of quantitative trait measurements is thus desirable to understand the functioning of ecosystems.

Still, despite the efforts towards compiling field observations from a myriad of studies into global databases (e.g. TRY, Kattge et al., 2020), the available data are sparse in terms of geographical coverage, species and range of traits (Asner et al., 2015; Kattge et al., 2020). In this context, hyperspectral remote sensing data offer an efficient proxy to map plant traits (Cavender-Bares et al., 2020; Jetz et al., 2016). Such data enable repeatable and non-destructive optical observations using numerous platforms and sensors providing information on spectral reflectance across a wide range of the electromagnetic spectrum via continuous narrow bands. Given the mechanistic interactions of light with leaf and canopy traits (Billings and Morris, 1951; Gates et al., 1965; Kattenborn and Schmidtlein, 2019; Ustin and Gamon, 2010), hyperspectral observations have a high potential to reveal plant traits over remote and large areas (Hank et al., 2019, Asner and Martin, 2016; Homolová et al., 2013; Singh et al., 2015; van Cleemput et al., 2018, Danner et al., 2021; Wocher et al., 2022). Recently launched and forthcoming hyperspectral space missions such as PRecursore IperSpettrale della Missione Applicativa (PRISMA, Cogliati et al., 2021), Environmental Mapping and Analysis Program (EnMAP, Guanter et al., 2015) and Surface Biology and Geology (SBG, Cawse-Nicholson et al., 2021) along with the higher-resolution proximal and airborne instruments, support this potential and will provide an unprecedented source of data. However, in view of

the varieties of these hyperspectral data sources and potential applications, we are missing transferable retrieval methods across sensors, acquisition settings, ecosystems and plant functional types.

From a methodological perspective, available retrieval methods range from data-driven statistical methods to the inversion of radiative transfer models (RTM) to hybrid methods (see Verrelst et al., 2019 for a review). RTMs simulate the interaction of light with vegetation properties and thus their inversion can represent a promising approach for plant trait retrieval (Berger et al., 2018; Dorigo et al., 2007; Feilhauer et al., 2018, 2017; Jacquemoud et al., 2009). Yet, plant trait retrieval by RTM inversion is only possible for traits that are considered in the RTM itself. Moreover, RTM inversions are very sensitive to the RTM's configuration and thus have to be specifically parameterized for different vegetation types, canopy structures, phenological stages or use cases (Dorigo et al., 2007; Atzberger and Richter, 2012; Verrelst et al., 2013). Conversely, data-driven approaches automatically learn the statistical relation between the spectral data and plant traits. Partial Least Squares regression (PLSR) (Geladi and Kowalski, 1986; Wold et al., 2001) can be considered as the benchmark approach given its long history in imaging spectroscopy (Asner and Martin, 2008; Feilhauer et al., 2010; Singh et al., 2015; Wang et al., 2020). In recent years, new machine learning algorithms emerged as powerful approaches to solve retrieval tasks from hyperspectral data (Wang et al., 2020, Prilianti et al., 2021, Shi et al., 2022).

Despite the potentials of data-driven methods, there are multiple constraints:

1) Commonly, data-driven models are trained with data sets representing limited variation in ecosystem properties, plant functional types, sensor systems and acquisition settings, thus limiting their transferability. For instance, previous studies (Asner et al., 2015; Berger et al., 2020; Wang et al., 2019) have concentrated on individual ecosystems such as croplands, forests, or grasslands using specific data sets. However, models developed from these data sets may produce significant uncertainties when employed on a new data set, making them less transferable to other ecosystems or alternative data sets (Wang et al. 2020).

2) Data-driven models are often built independently for different traits. This prevents exploiting interrelationships between certain traits. For example, different traits may be driven by the same processes or may manifest in overlapping absorption features such as pigments or resource-investment related traits. Consequently, taking the trait interrelations into account might improve the overall retrieval quality.

Moreover, the simultaneous prediction of multiple traits may also enlarge computational efficiency. It is thus compelling to aim for a data-driven approach that is capable of predicting a set of traits simultaneously. We further refer to such an approach as 'multi-trait' model.

3) Furthermore, most data-driven approaches for plant trait retrieval cannot easily be extended with new training data, which hinders continuous model improvements and knowledge transferability.

Deep learning and particularly Convolutional Neural Networks (CNNs) may pave new avenues to alleviate such issues (Sosnin et al., 2019; Yosinski et al., 2014; Zhang and Yang, 2021). CNNs are a powerful method for automatic feature engineering and are increasingly being applied to remote sensing data (Kattenborn et al., 2021; Zhu et al., 2017). Due to their depth and large number of neurons such models are capable of learning complex relationships. Accordingly, given sufficient representativeness of the input data, such models may learn transferable relationships across application domains, sensor types and acquisition settings. Moreover, CNNs are commonly trained iteratively, enabling to exploit very large data sets and allowing for continuous updating and fine-tuning of models with new, unseen data (Shin et al., 2016).

The availability of canopy spectra and their corresponding trait observations from different studies encompassing different plant types and sensors constantly increases (e.g. Rogers et al., 2021; van Cleemput et al., 2019). This opens a way to harness the scalability of deep learning and test the robustness of the models when integrating multi-source hyperspectral and plant trait data (e.g. EcoSIS, Wagner et al., 2018). However, due to the different context of these studies, a combination of such data sets is naturally sparse, meaning not all potential traits are covered across different data sets. Therefore, the objective of this study is to explore the potential of weakly supervised approaches to train models on sparse data for simultaneously predicting multiple traits (n = 20) from canopy hyperspectral data. This analysis is based on a combination of 42 data sets from heterogeneous data of different vegetation and sensor types. We implement three weakly-supervised multi-trait CNN approaches to investigate the hypothesis of whether the incorporation of trait-trait correlation in models' calibration would improve the trait estimations. The performance of these strategies is compared to common single-trait PLSR and single-trait CNN models.

2. Material and Methods

2.1 Data merging and cleaning

We employed 42 data sets of full range canopy spectra (400-2500 nm) with corresponding trait measurements (e.g. LMA, pigments) that were available upon request or from archives (e.g. EcoSIS, PANGEA) (Burnett et al., 2021; Cerasoli et al., 2018; Chlus et al., 2020; Ewald et al., 2018b, 2020; Hank et al., 2016, 2015; Herrmann et al., 2011; Kattenborn et al., 2019a; Pottier et al., 2014; Rogers et al., 2021; Singh et al., 2015; van Cleemput et al., 2019; Wang et al., 2020; Wocher et al., 2018, Dao et al., 2021). The sites of the collected samples are distributed across different continents (America, Asia and Europe see Fig. SA.1) and vary in climate and vegetation type (see details Table SA.1). The data comprise observations from different natural and semi-natural ecosystems (forest, grassland, tundra and shrubland), agricultural sites (crops and pastures), as well as plant-pot-experiments. Hence, the resulting data represent an aggregation of large and heterogeneous multi-site and multi-ecosystem sources.

2.1.1 Hyperspectral data

Canopy reflectance spectra were acquired with proximal and airborne hyperspectral sensors (e.g. ASD FieldSpec, Spectra Vista Corporation, SVC; AVIRIS; NEON Airborne Observation Platform AOP) and have different spectral properties. Still, they cover a comparable wavelength range of the solar electromagnetic spectrum (see Table SA.1). A forward and backward linear interpolation was performed to unify the diverse measurements in the full range of 400-2500 nm in 1 nm steps. To deal with the known issues of atmospheric water absorption in open-sky canopy reflectance spectra, we removed water absorption regions (1351-1430, 1801-2050 and 2451-2501 nm) and independently smoothed the three remaining parts of the spectra with a Savitzky-Golay filter (Savitzky and Golay, 1964) using a window size of 65 nm. Finally, 1721 interpolated spectral bands were retained for the analysis. Given the heterogeneity of the different data sets, the 5573 processed spectra cover a wide range of reflectance values (Fig. 1).The data were checked for overall spectral artifacts or inconsistencies (Supplement B). Despite the heterogeneity in land cover and vegetation types, we observed smooth transitions between the spectral features of the different data sets originating from sensor or pre-processing procedures (Fig. SB.1).

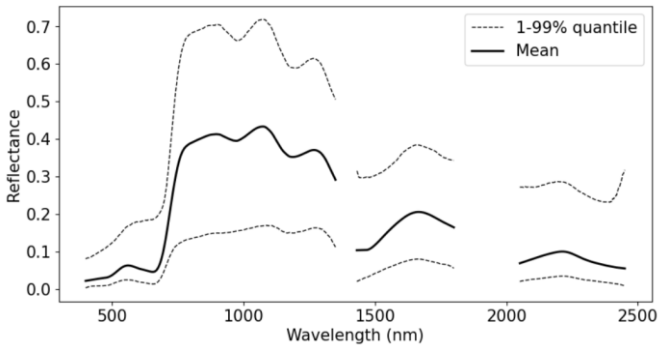

Fig. 1. Distribution of canopy reflectance of the available samples (N = 5573).

2.1.2. Leaf and canopy traits

From the available reference data, we selected a variety of traits (Table 1) related to light harvesting and growth,e.g. leaf pigments, nitrogen, structure and resource investments, leaf area index (LAI), equivalent water thickness (EWT) and leaf mass per area (LMA). Where necessary, leaf traits were converted to area basis, in contrast to mass-based measurements, to avoid the high correlation with LMA (see also Fig. SB.2) and to facilitate the model to learn the actual absorption features of the respective constituent (Hill et al., 2019; Kattenborn et al., 2019b; Ustin and Gamon, 2010, Zhao et al., 2021) (Fig. SB.2). Main conversions were based on the common relationships from literature (Kattenborn and Schmidtlein, 2019; Lichtenthaler, 1987) (Table SA.2). Table 1 summarizes the quantity of observations for each trait, their value ranges and the number of associated data sets. For the further analyses and for the sake of the training stability and computational efficiency, the trait values were rescaled. For this, we used the power transformation, which is a monotonic transformation to transform and normalize the data to a more-Gaussian-like distribution.

Table 1. Statistics of 20 selected functional traits available across 42 data sets. More details on the data sets can be found in Table SA.1. N = Number of samples, N Data sets = Number of data sets including the trait, Std = standard deviation, Min = minimum, Max = maximum.

Trait name	Trait description	Unit	N	N Data sets	Mean	Std	Min	Max
Anth	Anthocyanin content	(µg/cm²)	644	2	1.27	0.41	0.56	2.98

Boron	Boron content	(µg/cm²)	1086	14	0.39	0.26	0.01	2.34
C	Carbon content	(mg/cm²)	1876	23	5.84	4.44	0.10	37.29
Ca	Calcium content	(µg/cm²)	1045	16	107.25	101.97	0.69	988.73
Car	Carotenoid content	(µg/cm²)	1859	21	8.75	2.77	1.18	40.44
Cellulose	Cellulose	(mg/cm²)	1402	15	2.35	1.87	0.35	15.22
Chl	Chlorophyll content	(µg/cm²)	2141	24	38.57	14.53	4.45	229.50
Copper	Copper content	(µg/cm²)	1101	14	0.07	0.03	0.01	0.28
EWT	Equivalent Water Thickness	(mg/cm²)	1918	19	15.65	9.27	0.23	80.62
Fiber	Fiber	(mg/cm²)	1385	15	5.23	4.57	0.14	29.81
LAI	Leaf Area Index	(m²/m²)	1643	15	3.35	1.64	0.06	7.67
Lignin	Lignin	(mg/cm²)	1415	16	2.69	2.41	0.05	14.58
LMA	Leaf Mass per Area	(g/m²)	3328	32	92.05	68.08	5.72	663.81
Magnesium	Magnesium content	(µg/cm²)	1099	15	24.09	16.16	0.25	141.54
Manganese	Manganese content	(µg/cm²)	894	14	3.09	2.31	0.01	15.19
N	Nitrogen content	(mg/cm²)	2193	26	0.19	0.10	0.01	0.95
NSC	Non-Structural Carbohydrates	(mg/cm²)	1093	14	3.21	2.85	0.28	21.83
Phosphorus	Phosphorus content	(µg/cm²)	1289	16	14.42	9.45	0.29	73.43
Potassium	Potassium content	(µg/cm²)	1008	15	102.64	62.73	0.40	470.07
Sulfur	Sulfur content	(µg/cm²)	1039	14	13.31	9.13	0.62	57.23

2.2. Multi-trait model development

2.2.1. CNN implementation and training

Given the one-dimensional nature of the spectral data, we used one-dimensional Convolutional Neural Networks (1D-CNN). The context of neighboring wavelengths of the spectra makes CNN-based models

196 preferable here to the naive multilayer perceptron architecture (MLP). CNNs can have a deep structure and
197 conventionally include blocks (convolutional block) of successive layers including convolution, pooling, and
198 activation layers. The convolution operation is a sliding dot product of a filter (kernel) applied to the spectral
199 signal. Several filters are used in the convolutional layer where they serve as a feature extractor and are
200 iteratively learned during the CNN training process. The kernel's sliding fashion enables feature detection
201 to be applied across the full signal range. Subsequently, the pooling layers enable to condense information
202 from the output of the convolutional layers and facilitate a hierarchical feature extraction at multiple
203 wavelength scales. For more details about CNNs, the reader is referred to Goodfellow et al. (2016).

204 As backbone architecture, we used an adapted version of EfficientNet-B0 (Tan and Le, 2019), which is
205 modified for one dimensional input data. EfficientNet architectures are composed of a sequence of the
206 previously described CNN blocks with skip connections. They are designed to improve accuracy and
207 efficiency by using a scalable structure that allows the network to learn effectively from larger resolutions
208 while reducing computation costs. This is done through a combination of depthwise separable convolutions,
209 1x1 convolutions and network scaling methods (Tan and Le, 2019). The output layer of the implemented
210 architecture comprised 20 units corresponding to the number of traits to be predicted.

211 The learning process of the model was based on the stochastic gradient descent algorithm, where the
212 Adaptive Momentum estimation (Adam) optimizer was used to update the weights (Kingma and Ba, 2014).
213 The number of epochs was set to 300 with a batch size of 32. We employed the Hubert loss function to
214 reduce the effect of outliers. Given sparsity and resulting imbalance of trait observations in the merged data
215 set, we used a weighted loss version. The weights of the samples w (%) were calculated for each sample
216 as the complement of the number of non-null trait samples n_{nn} to the total number of samples n_{tot} in the
217 corresponding original data set (Eq 1). Additionally, a random up-sampling with replacement was
218 performed to have an equal number of samples from each data set on the training set.

219 $$w = 100 - n_{nn} / n_{tot} * 100 \quad (1)$$

220 To avoid over-fitting, two conventional regularization techniques were used: data augmentation and drop
221 out. Data augmentation introduces artificial variation in the data to help regulate the learning process. We

applied two random modifications for every epoch (a training cycle using all observations) with a 15% chance. This included 1) an addition of random noise with +/-30% of the spectral standard deviation per wavelength derived from all training samples and 2) an amplitude multiplication of the entire reflectance spectra with a random value between 0.98 and 1.02. As additional model regularization, we applied dropout (Hinton et al., 2012) after each block, which randomly drops learning units with a defined probability.

Within the 300 epochs, we selected the final model according to the lowest root mean squared error of a 20% hold-out from the training data. All CNN models were implemented in Python (3.9.5) with the TensorFlow (2.7.0) and Keras (2.7.0) frameworks.

2.2.2. CNN multi-trait and weakly supervised learning

In view of the sparsity of the merged data set (Table. 1), we tested three different strategies to train multi-trait models using the above-mentioned CNN architecture: The first strategy, $CNN_{multiIncomplete}$, was trained on the original sparse data set. To overcome data sparsity, we modified the loss function to only update the weights according to traits where a corresponding reference observation was present (i.e. not a missing value). This approach falls within the incomplete supervision category in the context of weakly supervised learning (Zhou, 2018). This strategy is considered as the baseline approach in this study.

The second strategy, $CNN_{multiInexact}$ aims to maximize the identification of trait-trait relations during the learning process from all data samples and, hence, includes a gap-filling of missing trait values. The gap-filling process is based on the predictions of the $CNN_{multiIncomplete}$. To avoid unrealistic values, trait predictions lower than the 1% quantile and exceeding the 99% quantile of the original data set (Table 1) were not considered for gap filling. This automated gap-filling approach does not require data on species or ecosystem characteristics, which might be missing or hard to define (Schrodt et al., 2015; Shan et al., 2012). Instead, it directly learns trait-trait relationships from available hyperspectral data. $CNN_{multiInexact}$ falls within the two weak supervision categories: incomplete and inexact supervision. The incomplete supervision is related to the gap-filling procedure, and the inexact supervision is performed when training on the completed but noisy labels (i.e. reference data with gap-filled values).

The third strategy, $CNN_{multiIncompleteTRY}$ aims to fill data gaps with trait observations obtained from the TRY plant trait database (Kattge et al., 2020). The TRY database (version 5), includes more than 11.8 million trait observations across more than 270.000 taxa. For each dominant species found in the reference data, trait observations were queried from TRY using the species name. We applied fuzzy matching to deal with minor inconsistencies in the spelling of the species names with a Damerau-Levenshtein-Edit distance >89 (Damerau, 1964; Konstantinidis, 2005). The dominant species mapping resulted therefore in 144 correspondences with TRY species. For these species, the mean median trait values were then used to fill the missing values. This gap-filled data set was then used to train the multi-trait CNN models ($CNN_{multIincompleteTRY}$). This strategy falls also within the inexact and incomplete weak supervision categories as the model is trained on sparse and noisy labels (i.e. median trait values within species).

2.3. Comparison to single-trait models

To evaluate the benefit of the multi-trait models and the uncertainty introduced from the weakly supervised approaches (i.e. inexact and incomplete), we additionally trained single-trait CNN (CNN_{single}) models, where a separate model was trained for each individual trait. Apart from the final layer (number of output units), the architecture for these models was the same as for the multi-trait models (Section 2.2.1). Moreover, we compared the CNN-based single and multi-trait models to partial least squares regression (Wold et al., 1984). PLSR is currently one of the most frequently applied algorithms for imaging spectroscopy (Feilhauer et al., 2010; Homolová et al., 2013). For training PLSR models for each trait ($PLSR_{single}$), we used scikit-learn (version 0.24.2) Python libraries. To avoid over-fitting, the optimal PLSR number of latent components was selected by minimizing the predicted residual sum of squares (PRESS) in cross-validation (Chen et al., 2004).

2.4. Model evaluation

Using trait measurements and the canopy reflectance data from 42 data sets described in (section 2.1), we compared the predictive performance of the 1) multi-trait CNN models to 2) single-trait CNN and 3) single-trait PLSR models (Fig. 2). For a fair comparison, the same input data settings were adopted for the training and evaluation of all modeling approaches including data splitting, transformation and up-sampling. The up-sampling procedure is a random sampling with replacement and was applied to all samples in the training

set to make sure that a comparable number of samples is included from each data set and to reduce the effect of bias towards data sets with more samples.

After training, the models were evaluated for their performance 1) within the domain of the training data (internal validation) using randomly sampled hold-outs, and 2) with regard to their transferability to new domains (external validation), where each individual data set was once retained from model training. For the internal evaluation, we adopted a 5-fold cross-validation (CV) for all models. Given the unbalanced sampling frequency of the individual data sets, we performed a stratified cross-validation based on the data set provenance (data sets). This procedure ensures equal distribution of trait samples across the folds. For the hold-out test sets, only the original (and not the gap-filled) samples were used. The external validation consisted of training the models repetitively on 41 out of 42 data sets while keeping one data set as hold-out for testing. To reduce computational load, the data set-CV was only applied for $CNN_{multiIncomplete}$ and $PLSR_{single}$..We evaluated the model performances using the coefficient of determination R^2 and the normalized root mean squared error (nRMSE, %). The nRMSE was derived by normalizing the root mean square error over the range of the observations (1-99% quantile). The final model performance was obtained by averaging the R^2 and nRMSE values over the 5 folds of the CV.

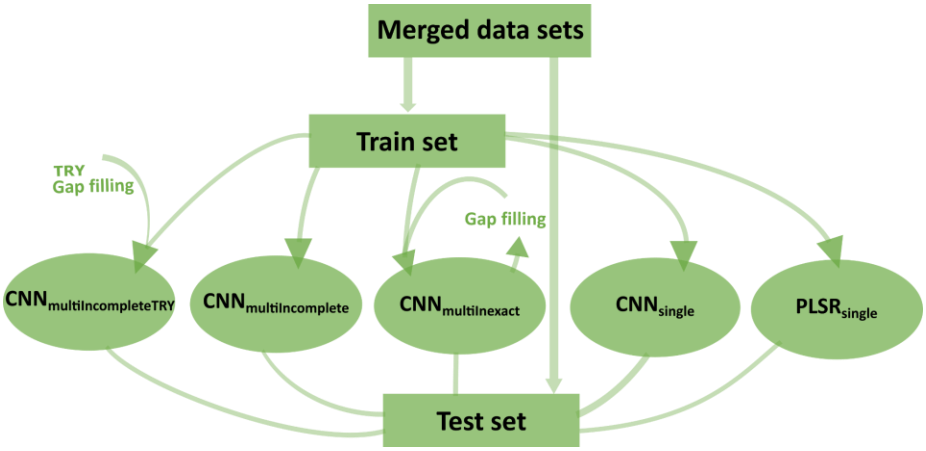

Fig. 2. Model performance assessment (internal validation) of multi-trait and single-trait models. The evaluation is based on a stratified 5-fold cross-validation.

2.5. Feature attribution

To visualize the spectral features learned by the $CNN_{multiIncomplete}$ model, we estimated the feature importance of each input wavelength, which were interpreted and compared with known spectral plant features. As feature importance metrics, we derived medians of SHapley Additive exPlanations (SHAP, Lundberg and Lee, 2017) absolute values. The SHAP values present a unified approach to explain model predictions based on the optimal game theory Shapley values. The Shapley values represent the local marginal contribution (i.e. for individual samples in the data) of each feature in the input for a specific prediction. They attribute the change in the expected model prediction when conditioning on one feature by calculating the difference from the prediction in which the feature in question is not included (Lundberg and Lee, 2017). These values can be approximated with different algorithms. We used the gradient explainer class, which combines the integrated gradients (Sundararajan et al., 2017), SHAP and SmoothGrad (Smilkov et al., 2017) methods. All SHAP coefficients were rescaled between 0-1 and normalized by the mean SHAP value of all traits to eliminate the effect of the learned trait covariance. For comparison, we also displayed the PLSR regression coefficients.

2.6. Uncertainty estimation

As indicated by earlier studies, transferability of machine learning-based models to new, unseen data depends on the distance in feature space (Kattenborn et al., 2022; Ludwig et al., 2023; Mila et al., 2022). Therefore, we implemented an uncertainty estimation procedure to reveal the effect of spectral dissimilarity between new data and data used in model training. Such a procedure is particularly valuable in view of large-scale mapping across ecosystems and sensors.

Inspired by Janet et al. (2019) and Meyer and Pebesma (2021), the implemented uncertainty estimation was based on the relationship between 1) CNN model residuals obtained from the internal evaluation and 2) the distance in feature space (dissimilarity of training vs. test sets). To reveal spectral dissimilarity from the eye of the CNN, the feature space was obtained from the CNN model embedding space of the global pooling of the last convolutional layer. Based on this feature space, the dissimilarity for each test sample was calculated as the average distance to the five nearest neighbors of the training data. The model uncertainty was then estimated using the calculated dissimilarity as predictor in a 95% quantile regression. The predicted values can be seen as the worst-case error prediction of the model. This procedure was tested for the $CNN_{multiIncomplete}$ model.

3. Results

3.1. Summary of the merged data set

3.1.1. Trait variations

The trait values across the merged data sets varied highly due to the heterogeneity in vegetation types and species (Table 1). This yields a large range in the trait values. LMA, Chl and EWT showed the highest variability (Coefficient of Variation CV = 47.97, 42,91, 38.44%, see Fig SB. 3, 4) in the original data while all other traits had similar variations (on average 35%). The correlation analysis based on Spearman coefficient of the merged data set revealed high correlation between several traits (Fig. 3). As expected, leaf constituents related to plant resource investments showed a large correlation (e.g. LMA, Carbon, Lignin, Fiber, Cellulose). These resource-investment related traits were rather independent from leaf pigments, which in turn were highly correlated among each other (Chl, Car, Anth). Both resource investment related traits and pigments showed a considerable correlation with leaf N. Overall, rather weak correlations were found for LAI and leaf constituents, whereas for N and C a positive relationship was observed. Water content overall also showed a positive correlation with other leaf constituents.

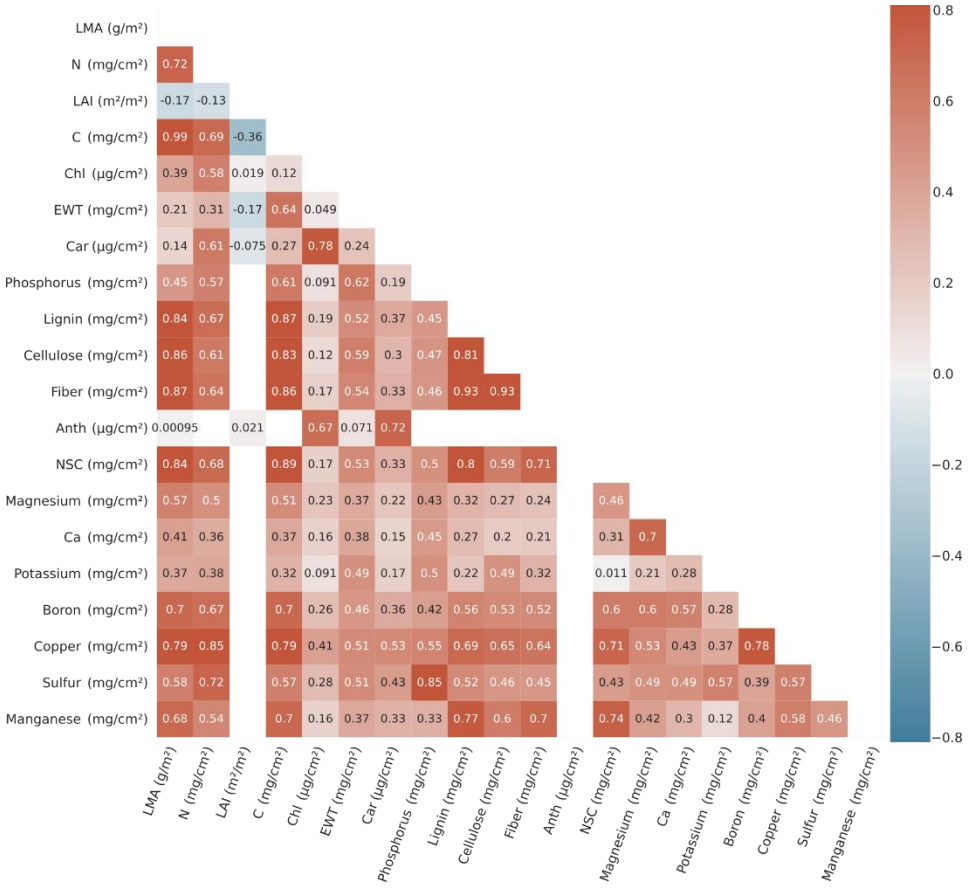

Fig. 3. Correlation plot of traits based on Spearman's rank correlation coefficient. Refer to Table 1 for an explanation of the traits. A correlation of leaf traits on a mass-basis is given in Fig. SB.2.

3.1.2. Canopy spectra

The canopy reflectance spectra were relatively similar when averaged across land cover types (Fig. 4) and we found smooth transitions across data sets and biomes (Fig. SB.1). Higher reflectance values were observed for the Tundra data in the NIR region (Fig. 4a). Largest coefficients of variations were found in the SWIR 2 region (2000-2500 nm) followed by the VIS region (400-750 nm). Most of the spectral variation was found in the crop related samples whilst forest samples had the lowest spectral variation (Fig. 4b).

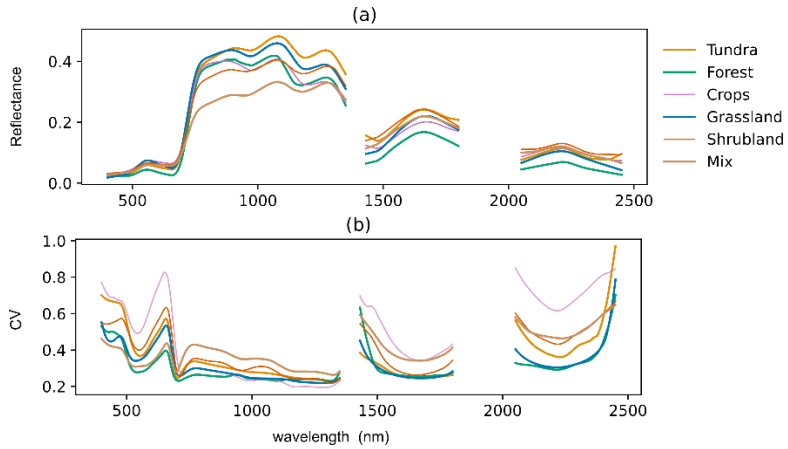

Fig. 4. Canopy average reflectance and the corresponding spectral variation (Coefficient of Variation, CV) across the different land cover types

3.2. Trait predictions

3.2.1. Prediction performances

The model performances derived from the 5-fold cross-validation showed the overall predictive performance varied greatly for the different traits (Fig. 5). With all CNN-based models, the goodness-of-fit of the predictions was higher for LMA, C, NSC (Non-structural carbon) (R^2> 0.69). Lower predictive performances of these models were obtained for EWT, N, Pigments, LAI, Cellulose, Lignin, Fiber, Copper and Phosphorus (R^2: 0.46 – 0.69 and nRMSE: 12 – 17%). Overall, the trait estimation performances of the CNN-based models exceeded those of the PLSR models (R^2: 0.18 to 0.66 and nRMSE: 11 – 22%). The PLSR models showed bias with high values for some traits, including LMA, Pigments and Carbon related traits (See Fig. SD.1). Only for a few traits, i.e. Boron, Ca and Manganese, the PLSR models showed higher performances than CNN models. Ca, Boron, Magnesium, Sulfur, Potassium and Manganese obtained the lowest validation performance for all models, especially with single-trait models (R^2 < 0.44 and nRMSE >15 %).

According to a Wilcoxon signed-rank test, the multi-trait models performed significantly better than single-trait models across all traits (e.g. $CNN_{multiIncomplete}$ $p < 0.001$, $w = 205$, details see Supplement C). In comparison to CNN_{single}, CNN-based, multi-trait models clearly improved the prediction performance for most of the traits. The prediction performance was particularly improved for traits where fewer samples were available or where a comparably lower correlation with spectral bands was observed (Fig SB.4), including

Anth, Sulfur, Ca and Potassium (Fig. 5b). Overall, the R^2 across all traits was higher for CNN multi-trait models than for CNN_{single} except for LMA, C and NSC (Fig. 5c, d).

Similar performance was obtained among the different CNN-based multi-trait models, i.e. $CNN_{multiIncomplete}$, $CNN_{multiInexact}$ and $CNN_{multiIncompleteTRY}$. The predictive performance for the $CNN_{multiInexact}$ ranged from R^2 of 0.21 – 0.70 and nRMSE of 10.41 – 18.79% , for $CNN_{multiIncomplete}$ R^2 of 0.29 – 0.77 and nRMSE of 9.17 – 17.81% and $CNN_{multiIncompleteTRY}$ R^2 of 0.29 – 0.78 and nRMSE of 8.92 – 17.85%. Overall, the $CNN_{multiIncompleteTRY}$ performed slightly better than the other two multi-trait strategies for most of the traits (Fig. 5). The $CNN_{multiIncompleteTRY}$ procedure is further discussed in section 3.2.2.

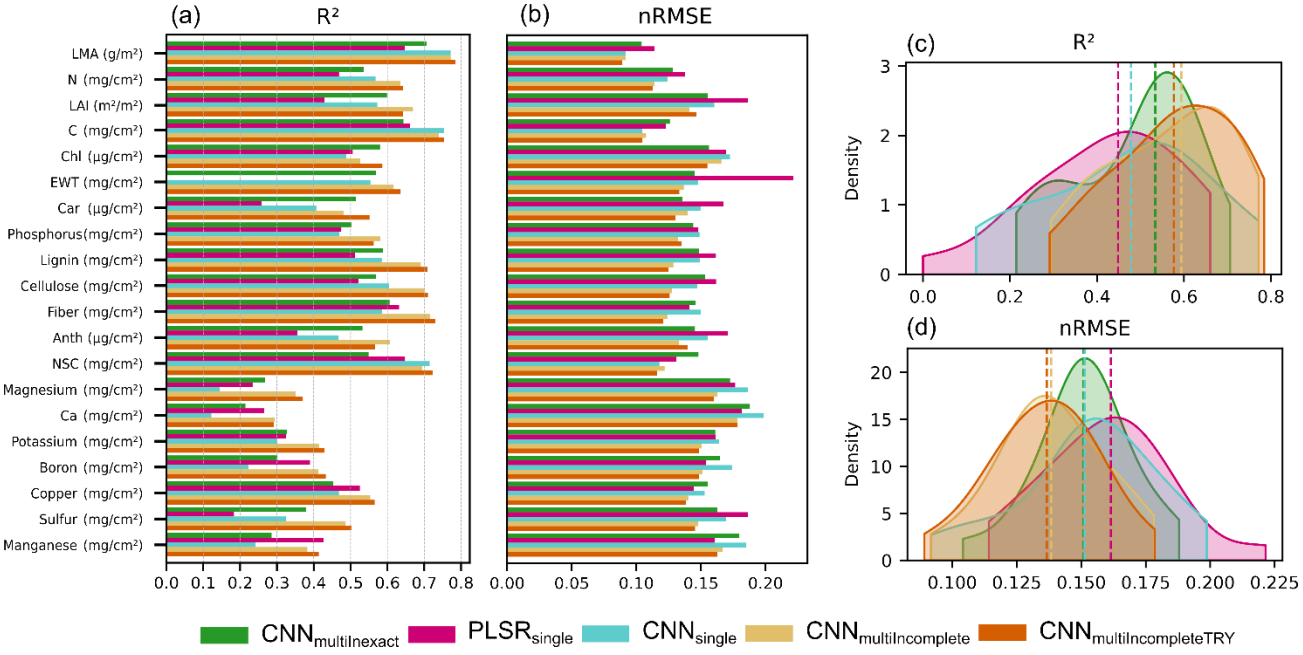

Fig. 5. (a) and (b): Comparative predictive accuracies (R^2 (a) and nRMSE (b)) from the stratified 5-fold cross validation of the $CNN_{multiIncomplete}$, $CNN_{multiInexact}$ and $CNN_{multiIncompleteTRY}$ models as well as $PLSR_{single}$ and CNN_{single} models for 20 traits. (c) and (d): The kernel density estimate (KDE) of the trait-based metric distributions (R^2 (c) and nRMSE (d)) with the associated median values (dashed lines). Refer to Table 1 for an explanation of the traits. Detailed performances can be found in Table SC.1 and 2 Supplement C.

All multi-trait approaches resulted in relatively robust and similar prediction performances across the different vegetation types (Fig. 6, SD.3, 4). For some traits (e.g. LMA, N, EWT), the values are slightly clustered according to vegetation types, but we did not observe a prominent or systematic bias in predictive

385 performance across these classes. For most traits, the model predictions are evenly scattered around the
386 1:1 line, which is also underlined by slopes of the linear fit close to 1 between the predicted and observed
387 trait values (Fig. 6, SD.3, 4).
388
389

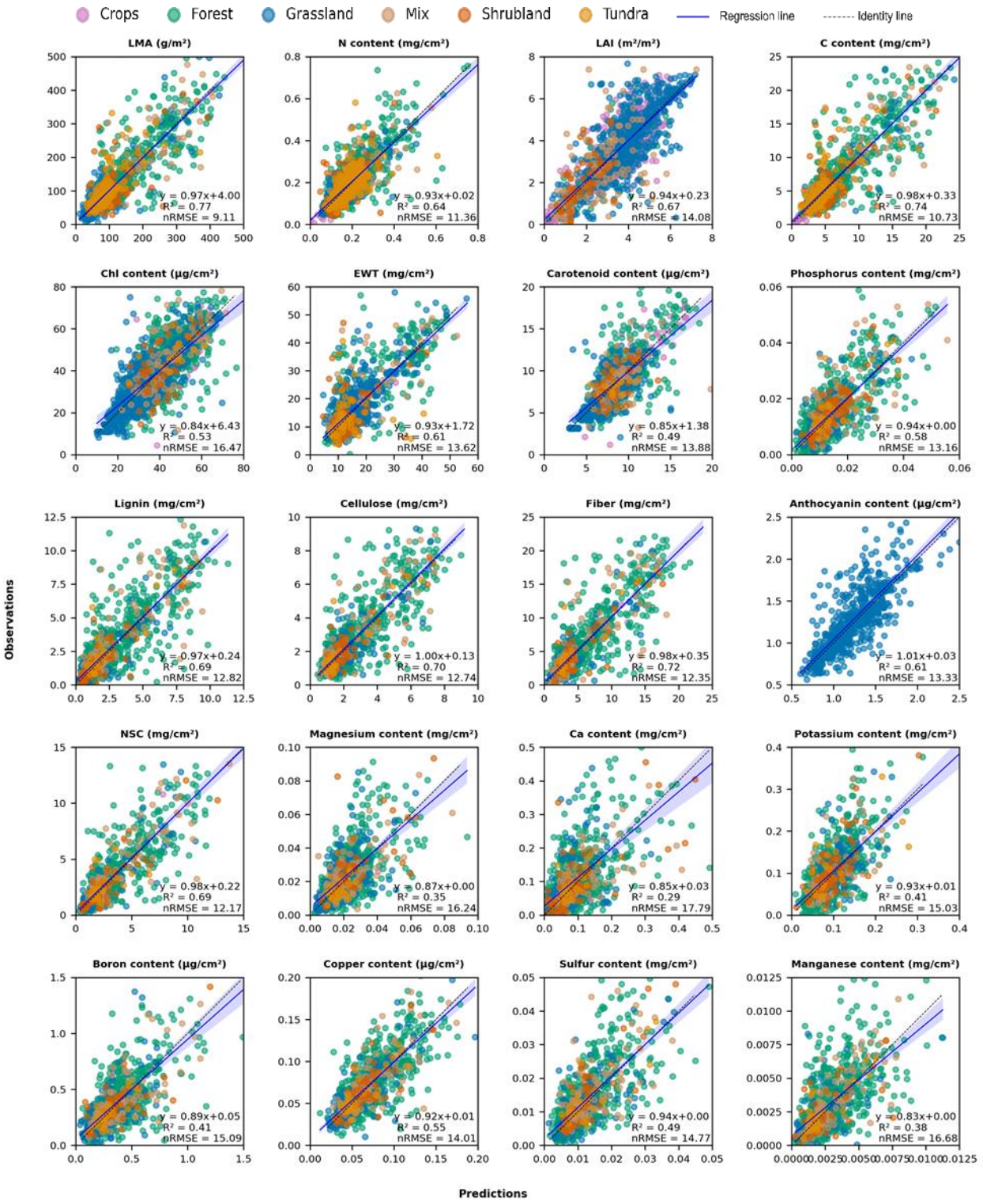

Fig. 6. Internal validation: Correlation between observed and predicted values of 20 traits from the multi-trait model $CNN_{multiIncomplete}$. The shown vegetation types only refer to the available types in the original associated data sets (not all land cover types are covered for each trait). Refer to Table 1 for an explanation of the traits. Scatter plots for the other models are given in supplement D.

Similar performance results were followed with the external evaluation, where CNN multi-trait model surpassed the performance of $PLSR_{single}$ models (Fig. 7, SE.1). With both modeling approaches, the performance across all traits with the external validation was lower than the internal validation,especially with $PLSR_{single}$ (Fig. 7, SE.1, Table SE.1). For $CNN_{multiIncomplete}$ LMA and C were the most transferable traits with R^2 higher than 0.6 which is consistent with the internal validation, while for PLSR Copper and Chl had the highest goodness-of-fit with $R^2 > 0.39$. However, the baseline multi-trait model ($CNN_{multiIncomplete}$) showed a bias in high trait values with N and LAI for example.

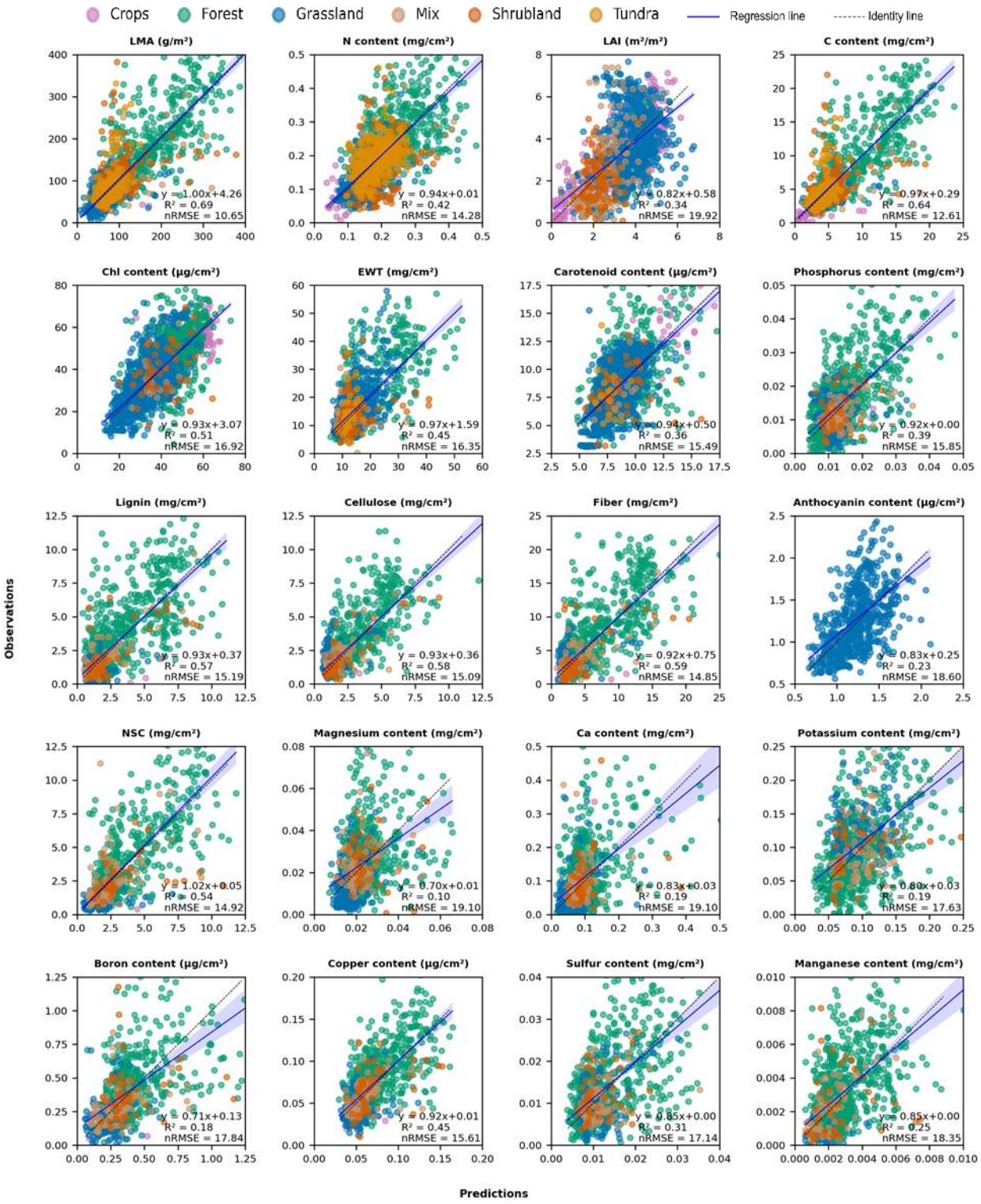

Fig. 7. External validation: Correlation between observed and predicted values of 20 traits from the multi-trait model $CNN_{multiIncomplete}$. The shown vegetation types only refer to the available types in the original associated data sets. Scatter plots for $PLSR_{single}$ are given in supplement E.

406 *3.2.2. Details on trait database integration*

407 Due to data availability, the gap filling of the $CNN_{multiIncompleteTRY}$ procedure was limited to 13 out of 20 traits
408 (Tab. 3). The model performance significantly improved for all the gap-filled traits (p = 0.004, w = 82,
409 Wilcoxon signed-rank test). Surprisingly, the $CNN_{multiIncompleteTRY}$ approach resulted even in significantly
410 improved performance for traits where no gap-filling could be performed, i.e. EWT, Car, Fiber, NSC and S
411 (p = 0.0313, w = 15). While the filling rate was not an important factor for model improvement, the
412 introduced variation from the species-based trait values had the largest effect on traits that already had
413 less sparse trait observations samples in the data set. For instance, Chl had the highest improvement in
414 performance and even surpasses the results of the baseline model $CNN_{multiIncomplete}$ (Table 3).
415
416 **Table 3.** Comparative nRMSE values (%) of the $CNN_{multiIncompleteTRY}$ with $CNN_{multiIncomplete}$ and CNN_{single}
417 models. $CNN_{multiIncompleteTRY.}$ Filling rates = (n obs. after - n obs. before) * 100 / old metric. Refer to Table. 1
418 for an explanation of the traits and to Table. SC.3 for more detailed metrics.

Traits	Filling rate (%)	nRMSE (%) (CNN_{single})	nRMSE (%) ($CNN_{multiIncomplete}$)	nRMSE (%) ($CNN_{multiIncompleteTRY}$)
Potassium	118.14	16.42	15.04	14.84
Phosphorus	99.07	14.89	13.23	13.51
Ca	97.42	19.87	17.82	17.85
Magnesium	93.64	18.65	16.26	16.00
C	92.33	10.45	10.76	10.48
Manganese	64.54	18.49	16.69	16.26
N	50.69	12.40	11.39	11.29
Copper	50.27	15.29	14.02	13.83
Chl	34.13	17.25	16.58	15.50
LMA	23.20	9.18	9.18	8.92
Lignin	12.43	14.91	12.86	12.48
Cellulose	7.77	14.71	12.78	12.58
Boron	0.55	17.39	15.11	14.86

419
420 *3.2.3. Feature importance*

421 The feature importance for $CNN_{multiIncomplete}$ and $PLSR_{Single}$ showed a clear correspondence in the overall
422 patterns (Fig. 8). For LMA, the relevant wavelengths in the CNN multi-trait model were spread across the
423 entire spectrum, with higher values in the longer wavelengths of the NIR SWIR regions (1200 - 2450 nm).
424 As expected, very similar patterns were found for traits that directly contribute to LMA, namely C, Cellulose,
425 Fiber and Lignin. The CNN multi-trait estimation of Chl and Car mostly relied on spectral bands in the VIS
426 and red-edge region (approx. 500 - 800 nm). For LAI, high SHAP values were found in the NIR.

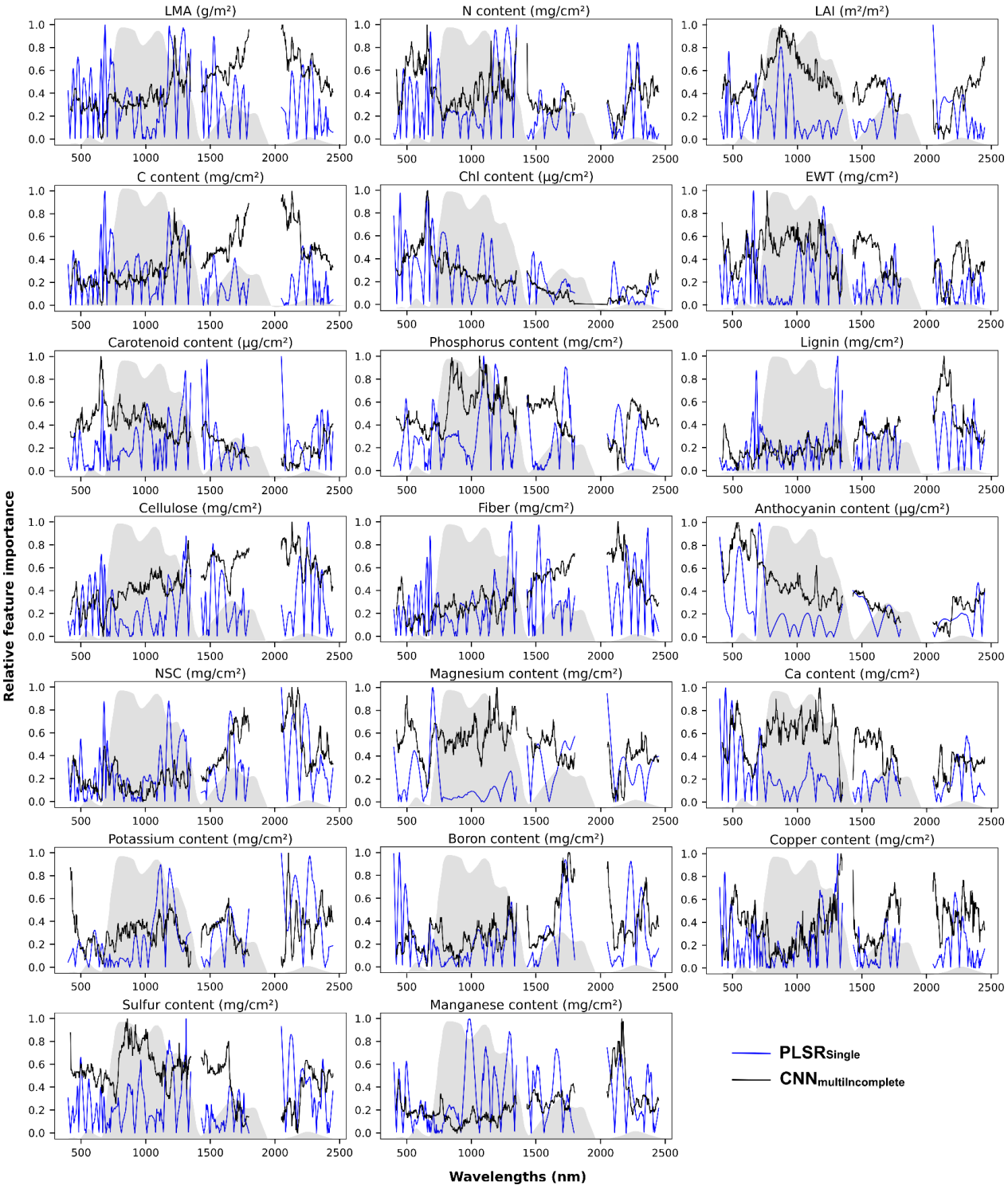

Fig. 8. Relative importance of spectral bands for the prediction of 20 traits using the $CNN_{multiIncomplete}$ and $PLSR_{single}$ models. The importance metric of $CNN_{multiIncomplete}$ (Black) is based on the SHAP scores with the gradient explainer, as for $PLSR_{single}$ the regression coefficients are shown (Blue). The gray shaded polygon represents a sample vegetation spectrum for orientation.

4. Discussion

4.1. Considerations on the merged data set

The transferability of statistical models to predict plant traits from new reflectance spectra is a major challenge (Ainsworth et al., 2014; Heckmann et al., 2017; Silva-Perez et al., 2018). Few previous studies have demonstrated that the transferability of models can be enhanced when the model training includes plots from different species and sites (Asner et al. 2015, Serbin et al., 2019; Wang et al., 2020, Kothari et al., 2022). Here, we merged 42 canopy reflectance data sets (from 28 studies) to assess the robustness of retrieval models when calibrated on heterogeneous data not only from different ecosystem types but also experimental settings (e.g. hyperspectral data acquisition and in-situ measurement protocols). This procedure provides an opportunity to address common shortages of reference data while also increasing the representativeness in terms of geographical coverage and diversity in vegetation type in the training data. Yet, it should be noted that the temporal coverage of the data is biased towards the peak of the vegetation period, while the senescence is underrepresented. This may affect for example the inter-correlations between traits as displayed in Fig. 3.

Merging the data sets required expert knowledge and a considerable effort for checking, cleaning, and converting trait observations. Although a large share of the data used here was acquired from the EcoSIS database, the available data often include errors and inconsistencies, e.g. assignment of wrong dimensions or units. In consideration of future initiatives for data integration, these experiences emphasize the need for a harmonization of plant trait observations, including units and dimensions, e.g. area or mass based, as well as quality assessments, terminology and sampling protocols.

As this merged data set incorporates various ecosystems and land cover types, its trait variability exceeds those of previous studies (Table 1, Fig. 3, Asner et al., 2015; Schiefer et al., 2021; Serbin et al., 2019; Wang et al., 2020, 2019). We assume that merging the different data sets is a compelling requirement for developing models that are transferable and robust across different traits, ecosystems, and vegetation types in the context of global mapping. Here, the baseline multi-trait model ($CNN_{multiIncomplete}$) appeared to generalize well over the individual data sets (Fig. 7). It should be noted, however, that the data only represent a small portion of the Earth's flora and its spatio-temporal variation. Hence, despite the

unprecedented trait variability realized here, the presented study should be regarded as a pioneering study in terms of model transferability and performance.

In the merged data set, not only the trait values but also the reflectance data showed considerable variability, which could be attributed not only to the spectral properties of the vegetation itself but also to differences in pre-processing modes with related uncertainties (e.g. during atmospheric correction procedures), remote sensing data acquisition settings (e.g. sun-observer-relationship) and instruments (e.g. airborne vs. field spectrometer data). We could not investigate in depth to what extent such factors limited the transferability of the models as information on such factors was not available for all individual data sets. Yet, we did not observe a significant difference in performance of our baseline multi-trait model ($CNN_{multiIncomplete}$) across the different remote sensing platforms (p = 0.17, u = 72, Mann–Whitney-U test) (Fig. SE.3, 4).

Merging data from multiple sources may improve model performances and transferability, but the sparsity and imbalance of trait observations challenged the model evaluations. For example, the number of data sets per trait ranges from 2 to 32 (Table 1), as most studies are application-specific and, hence, trait-specific. Likewise, the number of observations per data set ranged from 22 to 549. Thus, the relative performance of the model for the different traits is not necessarily directly comparable. Similarly, for some ecosystems or vegetation types only a few samples were available, which limited a conclusive performance evaluation in this regard. These challenges are expected to be resolved as more data may become openly available in the future.

4.2. Comparison of modeling approaches

Overall, the model performances of CNN-based models outperformed the widely used PLSR based models. This is consistent with previous studies that used hyperspectral data to retrieve vegetation and soil properties (Cui and Fearn, 2018; Ng et al., 2019; Pullanagari et al., 2021). The increased performance of CNN over PLSR may be explained by its ability to represent nonlinear relationships with an overall increased number of parameters, enabling the algorithm to learn more complex relationships. For example, the large trait-ranges of the merged data set presented in this study may inherit several non-linearities

between spectra-trait-relationships. Such nonlinearities may result from saturation effects, where a change in high trait values results in little change of spectral reflectance, as observed in the present study for chlorophyll, LMA or LAI. The linearity of PLSR models appeared to be less suitable to resolve such effects, as indicated by a clear saturation of PLSR-based predictions for high values for these traits (Fig. SD.1, SE.1). In such cases PLSR models tended to include more predictors (latent vectors) but this did not necessarily improve the model performance. Similar issues with PLSR-based models and saturation effects were also reported with leaf-scale reflectance data in Kothari et al. (2022). In contrast to the PLSR-based predictions, the predictions of the CNN models did not show saturation effects and no obvious systematic biases could be observed across the trait range (Fig. SD.2 – 4).

In addition to the model performance, CNNs are known to be less reliant on feature engineering and are effective to identify automatically relevant features from the input data (Goodfellow et al. 2016). Previous studies in the context of variable retrieval from hyperspectral data showed that shallower machine learning methods were more dependent on pre-processing of input data (Cui and Fearn, 2018; Ng et al., 2019). Another advantage of CNNs and other batch-compatible deep learning methods over previous machine learning methods (e.g. PLSR, Random Forest) is that the data are exposed iteratively to the model, which potentially enables training models with an infinite amount of data without exceeding the memory. The latter aspect may become very relevant in the near future that promises an increase in data availability, e.g. via more data acquisitions from spaceborne spectrometers and a growing culture of open data through initiatives such as ecosis.org.

The multi-trait CNN models clearly outperformed the single-trait models. This is consistent with other studies in different areas which employed multi-task CNN models (Ng et al., 2019; Padarian et al., 2019; Ramsundar et al., 2015; Tsakiridis et al., 2020). In comparison of the CNN_{single} model the retrieval of pigments, N, LAI, EWT, Phosphorus, Lignin, Cellulose, Fiber, Magnesium, Ca, Potassium, Boron, Copper, Sulfur was improved with our baseline multi-trait model ($CNN_{multiIncomplete}$). Even for traits that were only represented in a few data sets, the multi-trait models performed better than the single-trait models (e.g. Anth, Sulfur, Copper, Boron, Magnesium). We assume that multi-trait models not only allow for simultaneous and thus efficient trait retrieval, but also allow the model to indirectly learn trait-trait relationships.

519 Such trait-trait relationships may also explain the observed feature importances (Fig. 8). For instance, the
520 spectral features for N were consistent with known protein features in the SWIR region (Féret et al., 2021)
521 and others near the red-edge region related to pigments (Ustin et al., 2009). As expected, we also observed
522 very similar spectral features across all wavelengths among traits related to leaf resource investments (LMA,
523 Lignin, Fiber, Cellulose, and C; compare Kokaly et al., 2009), which may also explain higher model
524 performance for several of these traits when predicted in a multi-trait setting. For Anth, we observed
525 relatively accurate predictions and rather broad absorption features, although previous studies revealed that
526 Anth pigments have rather subtle and narrow spectral absorption properties (Féret et al., 2017). We assume
527 that the broad features obtained here result from the high correlation with Chl and Car (Fig. 3), which in turn
528 have more broad spectral absorption features and may indirectly facilitate Anth estimation (Jacquemoud
529 and Ustin, 2019, Ollinger, 2011; Ustin et al., 2009). Similarly, nutrients such as Copper, Sulfur, Potassium
530 and Boron do not have distinct spectral absorption features in canopy spectra, but their surprisingly high
531 retrieval performance may be explained by their correlation with other leaf traits that are related to leaf
532 resource investments (Fig. 3, 6) and that have a more explicit spectral response, such as LMA or C
533 (Domínguez et al., 2012, Kothari and Schweiger, 2022).
534
535 Largest improvements from single- to multi-trait estimates were found for Lignin, Cellulose and Fiber (Fig.
536 5), which can be attributed to the high correlation with LMA and C (Fig. 3). Conversely, for LMA, C and NSC
537 the multi-trait approaches did not result in notable improvements. This may be explained by the fact that
538 these three traits are already very tightly related (chemically and spectrally) and a covariance among these
539 traits does not add further benefit. Moreover, compared to other traits, LMA, C and NSC can be predicted
540 most accurately (Fig. 6), so the covariance with other traits that cannot be predicted as accurately is less
541 likely to facilitate the predictive performance. Similar findings for LMA estimation were found by Furbank et
542 al. (2021) when including the inter-correlation with photosynthetic traits.
543
544 We tested three weakly supervised strategies for training the multi-trait models in the context of the data
545 sparsity, i.e. $CNN_{multiIncompleteTRY}$, $CNN_{multiIncomplete}$ and $CNN_{multiInexact}$. The three strategies resulted in similar
546 model performance across the traits (e.g. for each strategy, LMA, C and NSC were most accurate and
547 macronutrients least accurate). Yet, $CNN_{multiInexact}$ resulted in the lowest model performance. This is

548 explained by the uncertainty introduced during the spectrally-based gap-filling procedure. However, even
549 with the propagated uncertainty from the gap-filling process, $CNN_{multiInexact}$ outperformed the single-trait
550 models. This demonstrates that such gap-filling strategies are promising to enrich existing sparse data sets,
551 especially as no external knowledge on species or ecosystem type is required. Future attempts may apply
552 a more conservative gap-filling, where data gaps are only filled if the estimated traits are assumed to have
553 a low uncertainty. The uncertainty assessment presented in this study (see Fig SG.1 for details) may be a
554 promising avenue.
555
556 The gap filling strategy based on trait databases ($CNN_{multiIncompleteTRY}$) significantly improved the
557 performance (compared to $CNN_{multiIncomplete}$) for those traits that were gap-filled *(*p = 0.004, w = 82 Wilcoxon
558 signed-rank test, Table 3, SC.3), even when using median trait values by species which do not account for
559 the within-species trait variations. Nonetheless, for the scope of this analysis this does not affect the
560 interpretation of the results as most of the collected samples were taken in the growing season and the
561 results were only evaluated with the original trait obsevations (i.e. no gap filling). Interestingly,
562 $CNN_{multiIncompleteTRY}$ even improved the model performance for those traits where no gap-filling was
563 performed (due to missing observations in the TRY database, p = 0.0313, w = 15, Wilcoxon signed-rank
564 test, Fig. 5, Table SC.3). This not only underlines the potential of incorporating ancillary trait information,
565 but also highlights the overall value of the multi-trait and corresponding trait-trait relationship. For instance,
566 this has surprisingly influenced the retrieval of Chl and Car, with an improvement of 12 – 16 % in R^2 and 7
567 – 7.88 % in nRMSE; as well as EWT, Fiber, NSC and Sulfur by 2.00 – 4.10% in R^2 and 1.68 – 4.95% in
568 nRMSE. We assume that the growth of trait databases as TRY will even increase the potential of this gap-
569 filling approach.
570
571 *4.3. Model performance across plant traits*
572 Across all traits, highest model performance was observed for LMA (Fig. 5). This is in line with a series of
573 previous studies highlighting the transferability of models for estimating LMA across data sets at leaf and
574 canopy scale (Serbin et al., 2019; Silva-Perez et al., 2018; Wang et al., 2019, 2020, Helsen et al., 2021,
575 Schiefer et al. 2021, Kothari et al., 2022). In contrast to these previous studies, the CNN models used here
576 resulted in comparable or even higher model performances although we tested our models using a more

diverse data set and exclusively on canopy spectra. The high performance of the LMA estimation is partly a surprise given its broad and overlapping absorption features with water content and scattering components at the canopy scale (Homolová et al., 2013). The high performance of LMA may be partially supported by the ample samples across most of the used data sets (32 data sets out of 42 had LMA observations). Moreover, the robustness of the LMA estimation may also be explained by the overall high correlation of LMA with individual bands across the entire spectrum (Fig. SB. 4).

Particularly for LMA but also for most of the other traits, our results suggest that the performances of the multi-trait models are often on par to those of previous studies. For instance, for LAI, Chl, Car and EWT, our models obtained higher performances than Schiefer et al. (2021), who used PLSR models on a data set of canopy spectra across grassland species, which was also integrated in our study. Overall, model performances were comparable to Wang et al. (2020), who used airborne canopy spectra across biomes and to Wang et al. (2019), who used canopy spectra in grasslands. EWT performances were lower than in Wang et al. 2020, where water content was one of the most accurately retrieved traits. The fact that the estimation of EWT was comparably low in the present study may result from the different protocols used across the merged data sets.

In this study we focused on area-based leaf traits due to multiple reasons: Firstly, as highlighted across different studies in the context of the radiative transfer theory (Dawson et al., 1998; Ganapol et al., 1998; Jacquemoud and Baret, 1990; Vilfan et al., 2016), the retrieval of leaf constituents from spectral signals depends on how much of a leaf constituent (mass) in a given leaf area interacts with light (area-based). In contrast, relative ratios of leaf constituents to LMA (mass-based traits) are not directly related to spectral absorption features (also discussed in Kattenborn et al. 2019b, Zhao et al. 2021). Secondly, normalizing traits on a mass-basis may overshadow the original variation of leaf traits and introduce unrealistic trait-trait-relationships. For instance, photosynthetic traits (e.g. pigments) are generally assumed to be largely independent of leaf resource investments (LMA) (Lloyd et al., 2013; Osnas et al., 2013). This was confirmed for the present data set (Spearman rho < 0.4) - but only if the data was scaled on an area-basis (Fig. SB.2). As soon as pigments were scaled on a mass-basis, ill-founded correlations were introduced (Spearman rho < -0.73, see Lloyd et al. 2013 for a statistical rationale). Likewise, traits that directly contribute to the

606 total leaf mass were obviously highly correlated to LMA when compared on an area-basis (spearmans´s
607 rho > 0.84 for Carbon, NSC, Lignin, Fiber, Cellulose), while a comparably weak relationship was found on
608 a mass-basis (Spearman´s rho < 0.51). Moreover, we found unrealistically high variation of these LMA-
609 related traits (Carbon, NSC, Lignin, Fiber, Cellulose) when assessed on a mass-basis, which may have
610 mis-lead model calibration (Fig. SB.3). Thus, to comply with the physical principles of radiative transfer
611 theory but also reasonable trait-trait relationships, the modeling in the present study was performed
612 exclusively on an area basis.
613
614 Note, however, that our proposed models can also be used to derive mass-based traits through
615 normalizing the respective trait prediction by LMA predictions ($trait_{mass}$ = $trait_{area}$ / LMA). We applied this
616 procedure to compare our model performances to previous studies that performed trait retrieval on a mass-
617 basis (Supplement F). The performances of our baseline multi-trait model ($CNN_{multiIncomplete}$) with mass-
618 based N and Phosphorus were comparable to studies reviewed in Homolová et al. (2013), while exceeding
619 those of Wang et al. (2020), Asner et al. (2015), Chadwick and Asner (2016), Ewald et al. (2018a) and
620 Wang et al. (2019). The predictive performance for the converted pigments, Fiber, Lignin and Cellulose
621 was lower or comparable to Wang et al. (2020) and Singh et al. (2015) and exceeded those of Asner et al.
622 (2015) and Martin et al. (2018) for tropical forest.
623
624 Nevertheless, it should be highlighted that it is often not possible to directly and quantitatively compare
625 model performances across studies, since they frequently differ in vegetation type, modeling approach,
626 model performance metrics and validation strategy, remote sensing platform and sensor, temporal and
627 spatial resolution and extent, simulated and real data, plant traits or a combination of these. Also, the aim
628 and thus the setting of the individual modeling attempts largely differs: some studies aimed to predict traits
629 in a very specific domain and from a very specific platform, while here we aimed to predict traits across
630 different platforms, sensors and vegetation types.
631
632 *4.4 Model performance across data sets (transferability)*
633 While the 5-fold CV evaluated the model performance with observations that are similar to those
634 observations used in training (internal validation), the model transferability specifically estimated the model

performance towards entirely unseen data sets (external validation). The model performances for the transferability evaluation were lower than the internal 5-fold CV (decline of 32% R^2 and 18% nRMSE (mean across traits), Fig. 7, Table SE.1). This decline in performance is expected given the large heterogeneity among the data sets (Tab. 1, Fig.4) which might largely differ from the training data, e.g. in terms of sensor, platform, illumination conditions, calibration procedure, trait sampling protocol or vegetation type. Overall, in terms of transferability the $CNN_{multiIncomplete}$ model clearly outperformed the $PLSR_{single}$ model (Fig. 7, SE.1). This may be explained by the larger number of parameters in the CNN-based models, which may facilitate learning more abstract spectral features and to resolve spectral features across different sensor or calibration settings. Both CNN- and PLSR-based traits, whose predictions had higher performances with the random internal CV, corresponded to those that had on average the most accurate prediction with the site transferability evaluation. Similar findings have been obtained in Kothari et al. (2022) at the leaf-scale.

Overall, the CNN-based transferability across data sets in this study can be considered as relatively high when compared with previous studies. Even at the leaf-level where spectrally-based trait retrieval is generally less challenging than at the canopy-scale, several studies reported similar or larger drops in performance across traits (Serbin et al. 2019, Helsen et al. 2021, Kothari et al. 2022). For instance, the LMA PLSR multi-biome model of Serbin et al. (2019) resulted in R^2 of 0.89 for the internal calibration and dropped to 0.66 when validated externally with LOPEX (Hosgood et al., 1995) and ANGERS data sets (Feret et al., 2008) and to 0.68 with the CABO data set (https://data. caboscience.org/leaf, Kothari et al., 2022, Kothari et al. 2022a). Wang et al. (2020) showed a very high model transferability with PLSR models across different vegetation types particularly for LMA and EWT. Likewise, the CNN-based model in Pullanagari et al. (2021) resulted in a robust transferability performance for N retrieval from grasslands where the authors claimed that this can be attributed to the richness of samples from multi-year and multi-site in the training set. However, these studies were based on a consistent sensor and data calibration and processing procedure. This underscores the challenge to train models that are transferable across remote sensing data acquisition settings. However, despite these challenges stemming from the diversity of integrated data sets, the transferability in this study is surprisingly high and we anticipate that with ever increasing data availability more generalized models can be trained in the future.

Eventually, the transferability of models will depend on the feature space distance between the new, unseen data to the training data (Ludwig et al. 2023). This is confirmed by the model uncertainty estimation procedure developed in this study (Fig. SG1), which is based on this principle and estimates the model uncertainty from the internal CNN embedding, i.e. the feature space viewed from the perspective of the model itself. Such an approach is assumed to be very promising to reveal the area of applicability of a model to new observations and domains (Meyer and Pebesma, 2021).

4.5. Outlook

As demonstrated in the present study, multi-trait models may not only facilitate high model performances due to the incorporated trait interrelationships, but also provide a tool to simultaneously and, hence, efficiently track multiple traits from remotely sensed hyperspectral data. The multi-trait approach presented here is expandable to more traits and can continuously be improved as new data become available. Instead of retraining the model from scratch, the model weights can be easily updated by retraining the model on new data. In the near future, a large increase in the availability of hyperspectral and trait data can be expected through the availability of operationally scheduled large-scale hyperspectral observations from spaceborne platforms. This goes along with a generally increased incentive for data sharing by the community and institutions and will include future in-situ and airborne campaigns that contribute to the success of global missions such as PRISMA, EnMAP, CHIME and SBG (Guanter et al., 2015; Labate et al., 2009). Upcoming approaches may also test the integration of simulated data from soil-leaf-canopy RTMs, in the context of hybrid retrieval models (e.g. Wocher et al., 2022; Verrelst et al., 2021). Such an approach might be particularly promising for traits, vegetation types or states for which only few data are available. In addition, such a physically based approach also takes information about the soil background as well as viewing and observation geometries into account, which may be neglected by empirical approaches.

5. Conclusion

From terrestrial platforms up to satellites, hyperspectral remote sensing is advancing as an important tool for future global monitoring applications. Currently, a significant bottleneck to unleash this potential is the lack of scalable and transferable models. Here, we compiled a large and sparse data set with a wide

variability in vegetation types and traits. Our results showed that multi-trait CNN models trained on these data can be more performant than CNN models trained for single traits individually. All tested CNN model approaches outperformed widely-used PLSR models. For multiple traits, the model performances obtained using the CNN multi-trait models were on par to those obtained in previous studies – although the model performances here were estimated from a more diverse data set. This highlights that building robust models requires substantial data variability and only a collaborative effort by the remote sensing community can significantly advance our ability to create models that are transferable across sensors, scales, domains, and ecosystems.

Author Contributions

Eya Cherif: Conceptualization, Methodology, Formal analysis, Data Curation, Writing - Original Draft, Visualization

Hannes Feilhauer: Data collection, Conceptualization, Writing - Original Draft, Supervision

Katja Berger: Data collection, Writing - Review & Editing

Phuong D. Dao: Data collection, Writing - Review & Editing

Michael Ewald: Data collection, Writing - Review & Editing

Tobias B. Hank: Data collection, Writing - Review & Editing

Yuhong He: Data collection, Writing - Review & Editing

Kyle R. Kovach: Data collection, Writing - Review & Editing

Bing Lu: Data collection, Writing - Review & Editing

Philip A. Townsend: Data collection, Writing - Review & Editing

Teja Kattenborn: Data collection, Conceptualization, Writing - Original Draft, Supervision

Acknowledgements

We thank all data owners for sharing the data either by request (in particular the consortium of the EU BiodivERsA project DIARS) or through the public Ecological Spectral Information System (EcoSIS), Data Publisher for Earth & Environmental Science (PANGEA) and DRYAD platforms. The authors acknowledge the financial support by the Federal Ministry of Education and Research of Germany and by the Sächsische Staatsministerium für Wissenschaft Kultur und Tourismus in the program Center of Excellence for AI-

research "Center for Scalable Data Analytics and Artificial Intelligence Dresden/Leipzig", project identification number: ScaDS.AI. Further support of this work was granted by the Federal Ministry for Economic Affairs and Climate Action (BMWK) and the German Aerospace Center (DLR) through the project AIResVeg (grant 50EE2203A) and the EnMAP scientific preparation program (grants 50EE0947, 50EE1308 and 50EE1623). Support for Dao, Kovach and Townsend was provided by US National Science Foundation (NSF), Macrosystems Biology and NEON-Enabled Science grant DEB-1638720 and NSF Biology Integration Institute award DBI-2021898.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships.

Code and data availability

The code, the data and trained model objects are available on Gitlab: https://gitlab.com/eya95/multi-traitretrieval

References

Ainsworth, E.A., Serbin, S.P., Skoneczka, J.A., Townsend, P.A., 2014. Using leaf optical properties to detect ozone effects on foliar biochemistry. Photosynth Res 119, 65–76.

Asner, G.P., Martin, R.E., 2016. Spectranomics: Emerging science and conservation opportunities at the interface of biodiversity and remote sensing. Glob Ecol Conserv 8, 212–219.

Asner, G.P., Martin, R.E., 2008. Spectral and chemical analysis of tropical forests: Scaling from leaf to canopy levels. Remote Sens Environ 112, 3958–3970.

Asner, G.P., Martin, R.E., Anderson, C.B., Knapp, D.E., 2015. Quantifying forest canopy traits: Imaging spectroscopy versus field survey. Remote Sens Environ 158, 15–27.

Atzberger, C., Richter, K., 2012. Spatially constrained inversion of radiative transfer models for improved LAI mapping from future Sentinel-2 imagery. Remote Sens Environ 120, 208–218.

Bergstra, J., Yamins, D., Cox, D., 2013. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures, in: International Conference on Machine Learning. pp. 115–123.

Berger, K., Machwitz, M., Kycko, M., Kefauver, S.C., van Wittenberghe, S., & others, 2022. Multi-sensor spectral synergies for crop stress detection and monitoring in the optical domain: A review. Remote Sens Environ 280, 113198.

Berger, K., Verrelst, J., Feret, J.-B., Wang, Z., Wocher, M., Strathmann, M., Danner, M., Mauser, W., Hank, T., 2020. Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions. Remote Sens Environ 242, 111758.

Berger, K., Atzberger, C., Danner, M., D'Urso, G., Mauser, W., & others, 2018. Evaluation of the PROSAIL model capabilities for future hyperspectral model environments: A review study. Remote Sens (Basel) 10, 85.

Billings, W.D., Morris, R.J., 1951. Reflection of visible and infrared radiation from leaves of different ecological groups. Am J Bot 327–331.

Burnett, A.C., Serbin, S.P., Rogers, A., 2021. Source: sink imbalance detected with leaf-and canopy-level spectroscopy in a field-grown crop. Plant Cell Environ 44, 2466–2479.

Cavender-Bares, J., Gamon, J.A., Townsend, P.A., 2020. The use of remote sensing to enhance biodiversity monitoring and detection: A critical challenge for the twenty-first century, in: Remote Sensing of Plant Biodiversity. Springer, Cham, pp. 1–12.

Cawse-Nicholson, K., Townsend, P.A., Schimel, D., Assiri, A.M., Blake, P.L., & others, 2021. NASA's surface biology and geology designated observable: A perspective on surface imaging algorithms. Remote Sens Environ 257, 112349.

Cerasoli, S., Campagnolo, M., Faria, J., Nogueira, C., Caldeira, M. da C., 2018. On estimating the gross primary productivity of Mediterranean grasslands under different fertilization regimes using vegetation indices and hyperspectral reflectance. Biogeosciences 15, 5455–5471.

Chadwick, K.D., Asner, G.P., 2016. Organismic-scale remote sensing of canopy foliar traits in lowland tropical forests. Remote Sens (Basel) 8, 87.

Chen, S., Hong, X., Harris, C.J., Sharkey, P.M., 2004. Sparse modeling using orthogonal forward regression with PRESS statistic and regularization. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 34, 898–911.

Chlus, A., Kruger, E.L., Townsend, P.A., 2020. Mapping three-dimensional variation in leaf mass per area with imaging spectroscopy and lidar in a temperate broadleaf forest. Remote Sens Environ 250, 112043.

Cogliati, S., Sarti, F., Chiarantini, L., Cosi, M., Lorusso, R. & others, 2021. The PRISMA imaging spectroscopy mission: Overview and first performance analysis. Remote Sens Environ 262, 112499.

Cornelissen, J.H.C., Lavorel, S., Garnier, E., D\'\iaz, S., Buchmann, N., & others, 2003. A handbook of protocols for standardised and easy measurement of plant functional traits worldwide. Aust J Bot 51, 335–380.

Cui, C., Fearn, T., 2018. Modern practical convolutional neural networks for multivariate regression: Applications to NIR calibration. Chemometrics and Intelligent Laboratory Systems 182, 9–20.

Damerau, F.J., 1964. A technique for computer detection and correction of spelling errors. Commun ACM 7, 171–176.

Danner, M., Berger, K., Wocher, M., Mauser, W., Hank, T., 2021. Efficient RTM-based training of machine learning regression algorithms to quantify biophysical & biochemical traits of agricultural crops. ISPRS Journal of Photogrammetry and Remote Sensing 173, 278–296.

Dao, P.D., Axiotis, A., He, Y., 2021. Mapping native and invasive grassland species and characterizing topography-driven species dynamics using high spatial resolution hyperspectral imagery. International Journal of Applied Earth Observation and Geoinformation 104, 102542. https://doi.org/https://doi.org/10.1016/j.jag.2021.102542

Dawson, T.P., Curran, P.J., Plummer, S.E., 1998. LIBERTY—Modeling the effects of leaf biochemical concentration on reflectance spectra. Remote Sens Environ 65, 50–60.

de Bello, F., Lavorel, S., D\'\iaz, S., Harrington, R., Cornelissen, J.H.C., & others, 2010. Towards an assessment of multiple ecosystem processes and services via functional traits. Biodivers Conserv 19, 2873–2893.

Díaz, S., Kattge, J., Cornelissen, J.H.C., Wright, I.J., Lavorel, S., & others, 2016. The global spectrum of plant form and function. Nature 529, 167–171.

Domínguez, M.T., Aponte, C., Pérez-Ramos, I.M., Garc\'\ia, L. v, Villar, R., Marañón, T., 2012. Relationships between leaf morphological traits, nutrient concentrations and isotopic signatures for Mediterranean woody plant species and communities. Plant Soil 357, 407–424.

Dorigo, W.A., Zurita-Milla, R., de Wit, A.J.W., Brazile, J., Singh, R., Schaepman, M.E., 2007. A review on reflective remote sensing and data assimilation techniques for enhanced agroecosystem modeling. International journal of applied earth observation and geoinformation 9, 165–193.

Ewald, M., Aerts, R., Lenoir, J., Fassnacht, F.E., Nicolas, M., & others, 2018a. LiDAR derived forest structure data improves predictions of canopy N and P concentrations from imaging spectroscopy. Remote Sens Environ 211, 13–25.

Ewald, M., Skowronek, S., Aerts, R., Dolos, K., Lenoir, J., & others, 2018b. Analyzing remotely sensed structural and chemical canopy traits of a forest invaded by Prunus serotina over multiple spatial scales. Biol Invasions 20, 2257–2271.

Ewald, M., Skowronek, S., Aerts, R., Lenoir, J., Feilhauer, H. & others, 2020. Assessing the impact of an invasive bryophyte on plant species richness using high resolution imaging spectroscopy. Ecol Indic 110, 105882.

Feilhauer, H., Schmid, T., Faude, U., Sanchez-Carrillo, S., Cirujano, S., 2018. Are remotely sensed traits suitable for ecological analysis? A case study of long-term drought effects on leaf mass per area of wetland vegetation. Ecol Indic 88, 232–240.

Feilhauer, H., Somers, B., van der Linden, S., 2017. Optical trait indicators for remote sensing of plant species composition: Predictive power and seasonal variability. Ecol Indic 73, 825–833.

Feilhauer, H., Asner, G.P., Martin, R.E., Schmidtlein, S., 2010. Brightness-normalized partial least squares regression for hyperspectral data. J Quant Spectrosc Radiat Transf 111, 1947–1957.

Féret, J.-B., Berger, K., de Boissieu, F., Malenovsk\`y, Z., 2021. PROSPECT-PRO for estimating content of nitrogen-containing leaf proteins and other carbon-based constituents. Remote Sens Environ 252, 112173.

Féret, J.-B., Gitelson, A.A., Noble, S.D., Jacquemoud, S., 2017. PROSPECT-D: Towards modeling leaf optical properties through a complete lifecycle. Remote Sens Environ 193, 204–215.

Feret, J.-B., François, C., Asner, G.P., Gitelson, A.A., Martin, R.E., & others, 2008. PROSPECT-4 and 5: Advances in the leaf optical properties model separating photosynthetic pigments. Remote Sens Environ 112, 3030–3043.

Funk, J.L., Larson, J.E., Ames, G.M., Butterfield, B.J., Cavender-Bares, J., & others, 2017. Revisiting the Holy Grail: using plant functional traits to understand ecological processes. Biological Reviews 92, 1156–1173.

Furbank, R.T., Silva-Perez, V., Evans, J.R., Condon, A.G., Estavillo, G.M., & others, 2021. Wheat physiology predictor: predicting physiological traits in wheat from hyperspectral reflectance measurements using deep learning. Plant Methods 17, 1–15.

Gates, D.M., Keegan, H.J., Schleter, J.C., Weidner, V.R., 1965. Spectral properties of plants. Appl Opt 4, 11–20.

Ganapol, B.D., Johnson, L.F., Hammer, P.D., Hlavka, C.A., Peterson, D.L., 1998. LEAFMOD: a new within-leaf radiative transfer model. Remote Sens Environ 63, 182–193.

Geladi, P., Kowalski, B.R., 1986. Partial least-squares regression: a tutorial. Anal Chim Acta 185, 1–17.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep learning. MIT press.

Grime, J.P., 1988. The CSR model of primary plant strategies—origins, implications and tests, in: Plant Evolutionary Biology. Springer, pp. 371–393.

Guanter, L., Kaufmann, H., Segl, K., Foerster, S., Rogass, C., & others, 2015. The EnMAP spaceborne imaging spectroscopy mission for earth observation. Remote Sens (Basel) 7, 8830–8857.

Hank, T.B., Berger, K., Bach, H., Clevers, J.G.P.W., Gitelson, A., & others, 2019. Spaceborne imaging spectroscopy for sustainable agriculture: Contributions and challenges. Surv Geophys 40, 515–551.

Hank, T., Locherer, M., Richter, K., Mauser, W., Consortium, E., others, 2016. Neusling (Landau ad Isar) 2012-a multitemporal and multisensoral agricultural EnMAP Preparatory Flight Campaign.

Hank, T., Richter, K., Mauser, W., Consortium, E., others, 2015. Neusling (Landau ad Isar) 2009-an agricultural EnMAP Preparatory Flight Campaign using the HyMap instrument.Heckmann, D., Schlüter, U., Weber, A.P.M., 2017. Machine learning techniques for predicting crop photosynthetic capacity from leaf reflectance spectra. Mol Plant 10, 878–890.

Helsen, K., Bassi, L., Feilhauer, H., Kattenborn, T., Matsushima, H., & others, 2021. Evaluating different methods for retrieving intraspecific leaf trait variation from hyperspectral leaf reflectance. Ecol Indic 130, 108111.

Herrmann, I., Pimstein, A., Karnieli, A., Cohen, Y., Alchanatis, V., Bonfil, D.J., 2011. LAI assessment of wheat and potato crops by VENμS and Sentinel-2 bands. Remote Sens Environ 115, 2141–2151.

Hill, J., Buddenbaum, H., Townsend, P.A., 2019. Imaging spectroscopy of forest ecosystems: perspectives for the use of space-borne hyperspectral earth observation systems. Surv Geophys 40, 553–588.

Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R., 2012. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580.

Homolová, L., Malenovsk\`y, Z., Clevers, J.G.P.W., Garc\'\ia-Santos, G., Schaepman, M.E., 2013. Review of optical-based remote sensing for plant trait mapping. Ecological Complexity 15, 1–16.

Hosgood, B., Jacquemoud, S., Andreoli, G., Verdebout, J., Pedrini, G., Schmuck, G., 1995. Leaf optical properties experiment 93 (LOPEX93). Report EUR 16095.

Jacquemoud, S., Baret, F., 1990. PROSPECT: A model of leaf optical properties spectra. Remote Sens Environ 34, 75–91.

Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P., & others, 2009. Use of prospect+ sail to estimate canopy biochemistry at different scales. Remote Sens Environ 113, S56–S66.

Jacquemoud, S., Ustin, S., 2019. Leaf optical properties. Cambridge University Press.

Janet, J.P., Duan, C., Yang, T., Nandy, A., Kulik, H.J., 2019. A quantitative uncertainty metric controls error in neural network-driven chemical discovery. Chem Sci 10, 7913–7922.

Jetz, W., Cavender-Bares, J., Pavlick, R., Schimel, D., Davis, F.W., Asner, G.P., & others, 2016. Monitoring plant functional diversity from space. Nat Plants 2, 1–5.

880 Kattenborn, T., Schiefer, F., Frey, J., Feilhauer, H., Mahecha, M.D., Dormann, C.F., 2022. Spatially
881 autocorrelated training and validation samples inflate performance assessment of convolutional neural
882 networks. ISPRS Open Journal of Photogrammetry and Remote Sensing 5, 100018.

883 Kattenborn, T., Leitloff, J., Schiefer, F., Hinz, S., 2021. Review on Convolutional Neural Networks (CNN)
884 in vegetation remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing 173, 24–49.

885 Kattenborn, T., Schiefer, F., Zarco-Tejada, P., Schmidtlein, S., 2019b. Advantages of retrieving pigment
886 content [µg/cm2] versus concentration [%] from canopy reflectance. Remote Sens Environ 230, 111195.

887 Kattenborn, T., Fassnacht, F.E., Schmidtlein, S., 2019a. Differentiating plant functional types using
888 reflectance: which traits make the difference? Remote Sens Ecol Conserv 5, 5–19.

889 Kattenborn, T., Schmidtlein, S., 2019. Radiative transfer modelling reveals why canopy reflectance follows
890 function. Sci Rep 9, 1–10.

891 Kattge, J., Bönisch, G., Díaz, S., Lavorel, S., Prentice, I.C., & others, 2020. TRY plant trait database -
892 enhanced coverage and open access. Glob Chang Biol 26, 119–188. https://doi.org/10.1111/gcb.14904

893 Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

894 Kokaly, R.F., Asner, G.P., Ollinger, S. v, Martin, M.E., Wessman, C.A., 2009. Characterizing canopy
895 biochemistry from imaging spectroscopy and its application to ecosystem studies. Remote Sens Environ
896 113, S78–S91.

897 Konstantinidis, S., 2005. Computing the Levenshtein distance of a regular language, in: IEEE Information
898 Theory Workshop, 2005. pp. 4–pp.

899 Kothari, S., Schweiger, A.K., 2022b. Plant spectra as integrative measures of plant phenotypes. Journal of
900 Ecology 110, 2536–2554.

901 Kothari, S., Beauchamp-Rioux, R., Blanchard, F., L. Crofts, A., Girard, A., Guilbeault-Mayers, X., W.
902 Hacker, P., Pardo, J., K. Schweiger, A., Demers-Thibeault, S., Bruneau, A., C. Coops, N., Kalacska, M.,
903 Vellend, M., Lalibert, E., 2022a. CABO 2018-2019 Leaf-Level Spectra.

904 Kothari, S., Beauchamp-Rioux, R., Blanchard, F., Crofts, A.L., Girard, A., Guilbeault-Mayers, X., Hacker,
905 P.W., Pardo, J., Schweiger, A.K., Demers-Thibeault, S., others, 2022. Predicting leaf traits across
906 functional groups using reflectance spectroscopy. bioRxiv 2022–2027.

907 Krogh, A., Hertz, J., 1991. A simple weight decay can improve generalization. Adv Neural Inf Process Syst
908 4.

909 Labate, D., Ceccherini, M., Cisbani, A., de Cosmo, V., Galeazzi, & others, 2009. The PRISMA payload
910 optomechanical design, a high performance instrument for a new hyperspectral mission. Acta Astronaut
911 65, 1429–1436.

912 Lavorel, S., Garnier, É., 2002. Predicting changes in community composition and ecosystem functioning
913 from plant traits: revisiting the Holy Grail. Funct Ecol 16, 545–556.

914 Lichtenthaler, H.K., 1987. [34] Chlorophylls and carotenoids: pigments of photosynthetic biomembranes,
915 in: Methods in Enzymology. Elsevier, pp. 350–382.

916 Lloyd, J., Bloomfield, K., Domingues, T.F., Farquhar, G.D., 2013. Photosynthetically relevant foliar traits
917 correlating better on a mass vs an area basis: of ecophysiological relevance or just a case of
918 mathematical imperatives and statistical quicksand? New Phytologist 199, 311–321.

919 Ludwig, M., Moreno-Martinez, A., Hölzel, N., Pebesma, E., Meyer, H., 2023. Assessing and improving the
920 transferability of current global spatial prediction models. Global Ecology and Biogeography.

921 Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. Adv Neural Inf
922 Process Syst 30.

923 Meyer, H., Pebesma, E., 2021. Predicting into unknown space? Estimating the area of applicability of
924 spatial prediction models. Methods Ecol Evol 12, 1620–1633.

925 Migliavacca, M., Musavi, T., Mahecha, M.D., Nelson, J.A., Knauer, J., Baldocchi, D.D., Perez-Priego, O.,
926 Christiansen, R., Peters, J., Anderson, K., others, 2021. The three major axes of terrestrial ecosystem
927 function. Nature 598, 468–472.

928 Mila, C., Mateu, J., Pebesma, E., Meyer, H., 2022. Nearest neighbour distance matching Leave-One-Out
929 Cross-Validation for map validation. Methods Ecol Evol 13, 1304–1316.

930 Miles, J., 2005. R-squared, adjusted R-squared. Encyclopedia of statistics in behavioral science.

931 Ng, W., Minasny, B., Montazerolghaem, M., Padarian, J., Ferguson, R., Bailey, S., McBratney, A.B., 2019.
932 Convolutional neural network for simultaneous prediction of several soil properties using visible/near-
933 infrared, mid-infrared, and their combined spectra. Geoderma 352, 251–267.

934 Ollinger, S. v, 2011. Sources of variability in canopy reflectance and the convergent properties of plants.
935 New Phytologist 189, 375–394.

936 Osnas, J.L.D., Lichstein, J.W., Reich, P.B., Pacala, S.W., 2013. Global leaf trait relationships: mass, area,
937 and the leaf economics spectrum. Science (1979) 340, 741–744.

938 Padarian, J., Minasny, B., McBratney, A.B., 2019. Using deep learning to predict soil properties from
939 regional spectral data. Geoderma Regional 16, e00198.

940 Poorter, H., Niinemets, Ü., Poorter, L., Wright, I.J., Villar, R., 2009. Causes and consequences of variation
941 in leaf mass per area (LMA): a meta-analysis. New phytologist 182, 565–588.

942 Pottier, J., Malenovsk\`y, Z., Psomas, A., Homolová, L., Schaepman, M.E., & others, 2014. Modelling
943 plant species distribution in alpine grasslands using airborne imaging spectroscopy. Biol Lett 10,
944 20140347.

945 Prilianti, K.R., Setiyono, E., Kelana, O.H., Brotosudarmo, T.H.P., 2021. Deep chemometrics for
946 nondestructive photosynthetic pigments prediction using leaf reflectance spectra. Information Processing
947 in Agriculture 8, 194–204.

948 Pullanagari, R.R., Dehghan-Shoar, M., Yule, I.J., Bhatia, N., 2021. Field spectroscopy of canopy nitrogen
949 concentration in temperate grasslands using a convolutional neural network. Remote Sens Environ 257,
950 112353.

951 Ramsundar, B., Kearnes, S., Riley, P., Webster, D., Konerding, D., Pande, V., 2015. Massively multitask
952 networks for drug discovery. arXiv preprint arXiv:1502.02072.

953 Rogers, A., Serbin, S., Ely, K., 2021. Leaf Mass Area, Leaf Carbon and Nitrogen Content, Kougarok Road
954 and Teller Road, Seward Peninsula, Alaska, 2016.

955 Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares
956 procedures. Anal Chem 36, 1627–1639.

957 Schiefer, F., Schmidtlein, S., Kattenborn, T., 2021. The retrieval of plant functional traits from canopy
958 spectra through RTM-inversions and statistical models are both critically affected by plant phenology. Ecol
959 Indic 121, 107062.

960 Schrodt, F., Kattge, J., Shan, H., Fazayeli, F., Joswig, J., & others, 2015. BHPMF–a hierarchical B
961 ayesian approach to gap-filling and trait prediction for macroecology and functional biogeography. Global
962 Ecology and Biogeography 24, 1510–1521.

963 Serbin, S.P., Wu, J., Ely, K.S., Kruger, E.L., Townsend, P.A., & others, 2019. From the Arctic to the
964 tropics: multibiome prediction of leaf mass per area using leaf reflectance. New Phytologist 224, 1557–
965 1568.

966 Shan, H., Kattge, J., Reich, P., Banerjee, A., Schrodt, F., Reichstein, M., 2012. Gap Filling in the Plant
967 Kingdom—Trait Prediction Using Hierarchical Probabilistic Matrix Factorization. arXiv preprint
968 arXiv:1206.6439.

969 Shi, S., Xu, L., Gong, W., Chen, Bowen, & others, 2022. A convolution neural network for forest leaf
970 chlorophyll and carotenoid estimation using hyperspectral reflectance. International Journal of Applied
971 Earth Observation and Geoinformation 108, 102719.

972 Shin, H.-C., Roth, H.R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D., Summers, R.M., 2016.
973 Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset
974 characteristics and transfer learning. IEEE Trans Med Imaging 35, 1285–1298.

975 Silva-Perez, V., Molero, G., Serbin, S.P., Condon, A.G., Reynolds, M.P., & others, 2018. Hyperspectral
976 reflectance as a tool to measure biochemical and physiological traits in wheat. J Exp Bot 69, 483–496.

977 Singh, A., Serbin, S.P., McNeil, B.E., Kingdon, C.C., Townsend, P.A., 2015. Imaging spectroscopy
978 algorithms for mapping canopy foliar chemical and morphological traits and their uncertainties. Ecological
979 Applications 25, 2180–2197.

980 Smilkov, D., Thorat, N., Kim, B., Viégas, F., Wattenberg, M., 2017. Smoothgrad: removing noise by adding
981 noise. arXiv preprint arXiv:1706.03825.

982 Sosnin, S., Vashurina, M., Withnall, M., Karpov, P., Fedorov, M., Tetko, I. v, 2019. A survey of multi-task
983 learning methods in chemoinformatics. Mol Inform 38, 1800108.

984 Sundararajan, M., Taly, A., Yan, Q., 2017. Axiomatic attribution for deep networks, in: International
985 Conference on Machine Learning. pp. 3319–3328.

986 Tan, M., Le, Q., 2019. Efficientnet: Rethinking model scaling for convolutional neural networks, in:
987 International Conference on Machine Learning. pp. 6105–6114.

988 Tsakiridis, N.L., Keramaris, K.D., Theocharis, J.B., Zalidis, G.C., 2020. Simultaneous prediction of soil
989 properties from VNIR-SWIR spectra using a localized multi-channel 1-D convolutional neural network.
990 Geoderma 367, 114208.

991 Ustin, S.L., Gamon, J.A., 2010. Remote sensing of plant functional types. New Phytologist 186, 795–816.

992 Ustin, S.L., Gitelson, A.A., Jacquemoud, S., Schaepman, M., Asner, G.P., 2009. Retrieval of foliar
993 information about plant pigment systems from high resolution spectroscopy. Remote Sens Environ 113,
994 S67–S77.

995 van Cleemput, E., Roberts, D.A., Honnay, O., Somers, B., 2019. A novel procedure for measuring
996 functional traits of herbaceous species through field spectroscopy. Methods Ecol Evol 10, 1332–1338.

997 van Cleemput, E., Vanierschot, L., Fernández-Castilla, B., Honnay, O., Somers, B., 2018. The functional
998 characterization of grass-and shrubland ecosystems using hyperspectral remote sensing: trends,
999 accuracy and moderating variables. Remote Sens Environ 209, 747–763.

1000 Verrelst, J., Rivera-Caicedo, J.P., Reyes-Muñoz, P., Morata, M., Amin, E., & others, 2021. Mapping
1001 landscape canopy nitrogen content from space using PRISMA data. ISPRS Journal of Photogrammetry
1002 and Remote Sensing 178, 382–395.

1003 Verrelst, J., Malenovsk\`y, Z., der Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J.-P., & others, 2019.
1004 Quantifying vegetation biophysical variables from imaging spectroscopy data: a review on retrieval
1005 methods. Surv Geophys 40, 589–629.

1006 Verrelst, J., Rivera, J.P., Leonenko, G., Alonso, L., Moreno, J., 2013. Optimizing LUT-based RTM
1007 inversion for semiautomatic mapping of crop biophysical parameters from Sentinel-2 and-3 data: Role of
1008 cost functions. IEEE Transactions on Geoscience and Remote Sensing 52, 257–269.

1009 Vilfan, N., der Tol, C., Muller, O., Rascher, U., Verhoef, W., 2016. Fluspect-B: A model for leaf
1010 fluorescence, reflectance and transmittance spectra. Remote Sens Environ 186, 596–615.

1011 Wagner, E.P., Merz, J., Townsend, P.A., 2018. Ecological spectral information system: an open spectral
1012 library, in: AGU Fall Meeting Abstracts. pp. B41L–2878.

1013 Wang, Z., Chlus, A., Geygan, R., Ye, Z., Zheng, T., & others, 2020. Foliar functional traits from imaging
1014 spectroscopy across biomes in eastern North America. New Phytologist 228, 494–511.

1015 Wang, Z., Townsend, P.A., Schweiger, A.K., Couture, J.J., Singh, A., & others, 2019. Mapping foliar
1016 functional traits and their uncertainties across three years in a grassland experiment. Remote Sens
1017 Environ 221, 405–416.

1018 Wocher, M., Berger, K., Verrelst, J., Hank, T., 2022. Retrieval of carbon content and biomass from
1019 hyperspectral imagery over cultivated areas. ISPRS Journal of Photogrammetry and Remote Sensing
1020 193, 104–114.

1021 Wocher, M., Berger, K., Danner, M., Mauser, W., Hank, T., 2018. Physically-based retrieval of canopy
1022 equivalent water thickness using hyperspectral data. Remote Sens (Basel) 10, 1924.

1023 Wold, S., Ruhe, A., Wold, H., Dunn Iii, W.J., 1984. The collinearity problem in linear regression. The
1024 partial least squares (PLS) approach to generalized inverses. SIAM Journal on Scientific and Statistical
1025 Computing 5, 735–743.

1026 Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. Chemometrics
1027 and intelligent laboratory systems 58, 109–130.

1028 Yang, Y., Zhu, Q., Peng, C., Wang, H., Chen, H., 2015. From plant functional types to plant functional
1029 traits: A new paradigm in modelling global vegetation dynamics. Prog Phys Geogr 39, 514–535.

1030 Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural
1031 networks? Adv Neural Inf Process Syst 27.

1032 Zarco-Tejada, P.J., Camino, C., Beck, P.S.A., Calderon, R., Hornero, A., & others, others, 2018. Previsual
1033 symptoms of Xylella fastidiosa infection revealed in spectral plant-trait alterations. Nat Plants 4, 432–439.

1034 Zarco-Tejada, P.J., Hornero, A., Beck, P.S.A., Kattenborn, T., Kempeneers, P., Hernández-Clemente, R.,
1035 2019. Chlorophyll content estimation in an open-canopy conifer forest with Sentinel-2A and hyperspectral
1036 imagery in the context of forest decline. Remote Sens Environ 223, 320–335.

1037 Zhang, Y., Yang, Q., 2021. A survey on multi-task learning. IEEE Trans Knowl Data Eng.

1038 Zhao, Y., Sun, Y., Lu, X., Zhao, X., Yang, L., Sun, Z., Bai, Y., 2021. Hyperspectral retrieval of leaf
1039 physiological traits and their links to ecosystem productivity in grassland monocultures. Ecol Indic 122,
1040 107267.

1041 Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., Fraundorfer, F., 2017. Deep learning in remote
1042 sensing: A comprehensive review and list of resources. IEEE Geosci Remote Sens Mag 5, 8–36.