Satellite remote sensing of plant functional diversity

Hauser, L.T.

Chapter 3.: Towards scalable estimation of plant functional diversity from Sentinel-2 imagery: in-situ validation in a heterogeneous (semi-)natural landscape

Based on:

Hauser, L. T., Féret, J.-B., Nguyen An Binh, van der Windt, N., Sil, Â. F., Timmermans, J., Soudzilovskaia, N. A., & P. M. van Bodegom (2021). Towards scalable estimation of plant functional diversity from Sentinel-2: In-situ validation in a heterogeneous (semi-)natural landscape. Remote Sensing of Environment, 262 (December 2020), 112505. https://doi.org/https://doi.org/10.1016/j.rse.2021.112505. (Impact Factor: 10.2)

Abstract

Large-scale high-resolution satellite observations of plant functional diversity patterns will greatly benefit our ability to study ecosystem functioning. Here, we demonstrate a potentially scalable approach that uses aggregate plant traits estimated from radiative transfer model (RTM) inversion of Sentinel-2 satellite images to calculate community patterns of plant functional diversity. Trait retrieval relied on simulations and Look-up Tables (LUTs) generated by an RTM and hybrid approaches rather than heavily depending on a priori field data and data-driven statistical learning. This independence from in-situ training data benefits its scalability as relevant field data remains scarce and difficult to acquire. We ran a total of three different inversion algorithms that are representative of commonly applied approaches and we used two different metrics to calculate functional diversity.

In tandem with Sentinel-2 image-based estimation of plant traits, we measured Leaf Area Index (LAI), leaf Chlorophyll content (CAB), and Leaf Mass per Area (LMA) in-situ in a (semi-)natural heterogeneous landscape (Montesinho region) located in northern Portugal. Sampling plots were scaled and georeferenced to match the satellite observed pixels and thereby allowed for a direct one-to-one posterior ground truth validation of individual traits and functional diversity.

Across approaches, we observe a reasonable correspondence between the satellite-based retrievals and the in-situ observations in terms of the relative distribution of individual trait means and plant functional diversity across locations despite the heterogeneity of the landscape and canopies. The functional diversity estimates, based on a combination of canopy and leaf traits, were robust against estimation biases in trait means. Particularly, the convex hull volume estimate of functional diversity showed strong concordance with in-situ observations across all three inversion methods (Spearman's ρ: 0.67–0.80). The remotely sensed estimates of functional diversity also related to in-situ taxonomic diversity (Spearman's ρ: 0.55–0.63).

Our work highlights the potential and challenges of RTM-based functional diversity metrics to study spatial community-level ecological patterns using currently operational and publicly

available Sentinel-2 imagery. While further validation and assessment across different ecosystems and larger datasets are needed, the study contributes towards a further maturation of scalable, spatially, and temporally explicit methods for functional diversity assessments from space.

3.1. Introduction

Worldwide biodiversity declines are affecting ecosystem functioning and pose risks to humankind as our existence heavily relies on healthy ecosystems (Cardinale et al., 2012; IPBES, 2019; Rands et al., 2010). In light of this ongoing global biodiversity crisis, the urgency to monitor and map terrestrial biodiversity at large scales has spurred research on adequate quantitative methods for biodiversity assessments (Anderson, 2018; Pereira et al., 2013). Improved monitoring of biodiversity dynamics can equip us to better understand and act upon changes, and halt further exacerbation of the current alarming rates of biodiversity loss (O'Connor et al., 2015; Skidmore, 2015).

A growing body of research highlights the role of functional diversity - rather than species diversity – in linking biodiversity to the functioning of ecosystems (Diaz and Cabido, 2001; Hooper et al., 2005; Violle et al., 2014). Functional diversity describes the range, value, and abundance of organismal traits. Traits are the measurable features of an organism that potentially affect performance, fitness, or resource acquisition strategies (Cadotte et al., 2011). Plant functional diversity integrates both inter- and intraspecific trait variation and has been found enhance ecosystem productivity, stability and resilience (Cardinale et al., 2011; Díaz et al., 2007; Duncan et al., 2015; Funk et al., 2016; Grime, 1998; Hooper, 2002; Isbell et al., 2011; Mori et al., 2013; Ruiz-jaen and Potvin, 2010). As such, the assessment of plant functional diversity patterns is highly relevant to monitoring the health (productivity, stability) and biodiversity of our ecosystems.

Traditionally, trait measurements are acquired by elaborate field campaigns (Baraloto et al., 2010). Such field campaigns are highly valuable but laborious, costly, and inefficient in dealing with the ecological complexity that comes with monitoring spatial and temporal variation of functional diversity (Májeková et al., 2016; Scholes et al., 2012). Field campaigns are particularly laborious if we aim to gather detailed spatially continuous information across large spatial extents for mapping and understanding the spatio-temporal dynamics of functional diversity.

To overcome this challenge, an increasing number of studies has explored the applicability of remote sensing techniques in assessing regional plant functional diversity for different ecosystems to scale up our biodiversity monitoring capabilities (Aguirre-gutiérrez et al., 2021; Jetz et al., 2016; Wang and Gamon, 2019). State-of-the-art studies used airborne data to map multivariate forest functional types (Asner et al., 2017) and plant functional diversity using both optical and LiDAR observations in combination with statistical approaches (Durán et al., 2019) and spectral indices (Schneider et al. 2017). Despite the value of these airborne remote sensing observations, its potential for application at larger extents is limited as airborne campaigns remain costly to organize and are bound in spatial extent and repeatability.

With ongoing technological advances and the launch of higher spatial and spectral resolution sensors in orbit, satellite-based observations present global and timely information that holds

large potential as the next frontier to monitor functional diversity patterns across space and time (Aguirre-gutiérrez et al., 2021; Jetz et al., 2016; Ma et al., 2019). Spaceborne remote sensing, however, generally relies on sensors that operate at spatial and spectral resolutions that are inferior to airborne hyperspectral instruments. These constraints challenge the fine-grained local-scale interpretation and in-situ validation of biodiversity estimates, and increase ill-posedness in retrieving biophysical traits from the spectral broadbands of satellites (Baret and Buis, 2008; Wang and Gamon, 2019).

The search for adequate quantitative methods to monitor functional diversity exploiting satellite earth observations remains in need of further research and development (Ma et al., 2019; Rossi et al., 2020; Torresani et al., 2019). Ideally, these research efforts will provide methods that allow us to accurately retrieve plants traits for functional diversity 1) with current satellite sensors, 2) without heavy reliance on scarce comprehensive ancillary field measurements, and that are 3) scalable across time and location including across vegetation types, and 4) measurable in-situ for validation against ecological field data. To meet these requirements, the use of optical traits that are physically related to spectra is particularly appealing given its universal applicability as opposed to statistical learning approaches and/or spectral indices that heavily depend on comprehensive field measurements for training and that have been found to be site- and time-specific (Verrelst et al. 2015; Clevers 2014; Ali, Darvishzadeh, Skidmore, Gara, et al. 2020).

The physical basis of trait retrieval is commonly ensured through the use of Radiative Transfer Models (RTMs) which relate incident radiation to vegetation canopies through angular, structural, biochemical, and biophysical characteristics (Verhoef 1998; Jacquemoud et al. 2009; Jacquemoud and Ustin 2019). These model parameters include leaf or canopy characteristics of ecological relevance (Anderson, 2018; Feilhauer et al., 2018; Homolová et al., 2013; Ollinger, 2011; Roelofsen et al., 2013). However, the universality of these models is bound by strong assumptions and heavy parameterization, simplifying the heterogeneous canopies and vegetation types encountered in the field. In the end, the practice of canopy RTM inversion to estimate plant traits from vegetation spectral reflectance is not trivial, but ill-posed and prone to a range of equally possible solutions, especially in multispectral settings (Combal et al., 2003; Koetz et al., 2007; Musavi et al., 2015).

Recently, a number of studies have shown success in applying RTM inversion on satellite earth observations (Sentinel-2) to estimate key plant traits in (semi-)natural ecosytems (e.g. Ali, Darvishzadeh, Skidmore, Gara, et al. 2020; Ali, Darvishzadeh, Skidmore, Heurich, et al. 2020; Rossi et al. 2020; Vinué, Camacho, and Fuster 2018). Unfortunately, the step from using trait estimates that are consistent among each other to deriving functional diversity from satellite remote sensing is undertaken less often (Ma et al., 2019; Rossi et al., 2020). To our knowledge, none of the existing satellite-based approaches so far has used RTM inversion to derive multiple traits simultaneously to obtain functional diversity estimates in heterogeneous (semi-)natural landscapes.

In this study, we present satellite-based functional diversity estimates using Sentinel-2 imagery. Our main objective is to examine our current ability to derive multiple plant traits, the local variation in trait aggregates, and ultimately estimate community patterns of functional diversity metrics across a heterogeneous and biodiverse (semi-)natural landscape. Our approach focuses on the use of RTM inversion that does not heavily rely on ‘data-intensive’ or

'a priori' in-situ training data. In support of our main objective, we 1) validated functional diversity estimates as derived by satellite observations using appropriately scaled in-situ measurements, 2) evaluated the robustness of different functional diversity indicators in light of uncertainties in the retrieval of trait values, and 3) assessed how remotely sensed community-based patterns of functional diversity aligns with in-situ community taxonomic diversity, as species are still the most commonly used units of assessment in conservation planning (Gaston, 2010; Meatyard, 2005; Petchey and Gaston, 2002). We ran three different implementations of commonly used RTM inversion approaches to estimate plant traits to more generically evaluate the performance, applicability and robustness of RTM-based functional diversity estimates instead of focusing on tweaking a single inversion method. Taken together, the demonstration and validation of Sentinel-2-derived functional diversity gives insight into the potential of scalable and operational RTM approaches to serve plant functional diversity monitoring from satellite earth observations across large-scale heterogeneous landscapes.

3.2. Methods

3.2.1. Study area

A comprehensive field data campaign was conducted in the Montesinho Natural Park and the Natura 2000 sites in Montesinho-Nogueira located in the Northeast of Portugal along the Spanish border (See Fig. 3.1). With a size of over 1000 km^2, the area plays an important role in the conservation of regionally endemic biodiversity (Aguiar, 2001; Bastos et al., 2018). The study area is characterized as a natural mountainous area with elevation ranging between 371m and 1488m above sea level. The highlands are dominated by a patchy landscape of shrublands, mixed with occasional Pyrenean oak forests and Holm-oak woodlands of which the latter mostly occur on rock outcrops, shallow soils, and steep slopes (Azevedo et al., 2013; Fonseca et al., 2012; Rego et al., 2011). The lowlands consist of agriculture intermingled with chestnut plantations, while pine forests plantations occur at mid-elevation in the eastern part of the area (Sil et al., 2017).

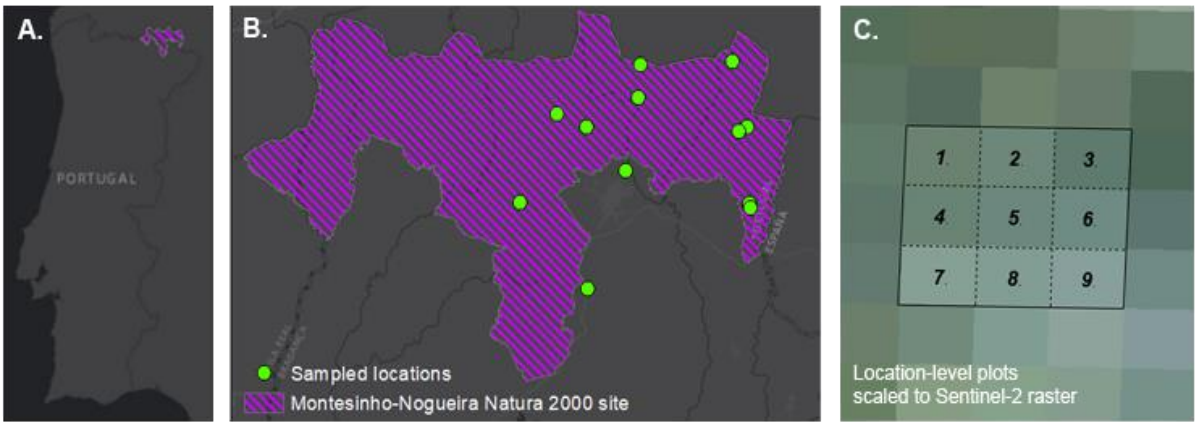

Fig. 3.1: Map of the study area depicting the location of the Montesinho-Nogueira Natura 2000 site in Portugal (panel A.), the distribution of the twelve individual sampling locations characterized by (semi-)natural woody vegetation across the wider national park and Natura 2000 site (panel B.), and an exemplar individual location (panel C.) consisting of nine individual adjacent plots scaled and georeferenced to Sentinel-2's pixel raster which is illustrated by the panchromatic reflectance at 20m spatial resolution as

background. For each plot (total N=97), an average of 17 individual sunlit branches were sampled to collect leaves representative of trait means of the overstory areal composition.

A total of twelve vegetated locations were selected in the Montesinho area (Fig. 3.1). The selection of the locations is representative of the territorially abundant extensively managed (semi-)natural areas dominated by woody vegetation that are of importance to the endemic biodiversity. These twelve locations include six forested and six shrubland locations and include sites that are dominated by single-species canopies as well as sites with a heterogeneous mix of species. Each location consisted of eight to nine plots (8-9×(20m×20m)) that correspond to georeferenced and scaled Sentinel-2 pixels. Fig. S. 8 depicts photos taken during the field campaign illustrative of the typical sampling locations.

3.2.2. Field measured plant traits

In-situ species inventories and leaf and canopy traits of each plot across the twelve locations were assessed during a field study that lasted from 12/06 until 05/07 of 2019. The trait collection resulted in ground measurements of Leaf Area Index (LAI), Chlorophyll A and B content (µg/cm²; CAB) and Leaf Mass per Area (g/cm²; LMA) operationalized in units directly comparable to the implementation of these traits in the PROSAIL RTM (Jacquemoud et al. 2009). The selection of traits was based on their ecological importance in terms of plant functioning (Croft et al. 2017; Wright et al. 2004; Díaz et al. 2016; Asner, Scurlock, and Hicke 2003; Zheng and Moskal 2009), and their importance in the spectral response of vegetation and our understanding thereof (Feret et al. 2008; Jacquemoud and Baret 1990; Serbin et al. 2019). Selected traits include;

- LAI, defined as the area of leaf material per unit of ground surface area, is considered an important canopy trait, both by itself as well as an important characteristic to scale up leaf traits to canopy traits (Asner 1998; Roelofsen et al. 2013). LAI relates directly to primary productivity and to competitive and complementary light use, transpiration, and energy exchange (Asner et al., 2003; Castillo et al., 2017; Fang et al., 2019; Zheng and Moskal, 2009).
- CAB is the surface-based leaf content of green photosynthetic pigments in chloroplasts and plays an important role in the photosynthetic capacity and resource strategy of plants (Croft et al. 2017).
- LMA is the amount of dry mass of a leaf per leaf area and a key feature in capturing leaf economics, reflecting trade-offs between carbon gain and longevity of a plant (Díaz et al., 2016; Wright et al., 2004).

We used hemispherical photography to specify LAI similar to approaches by Garrigues et al. (2008), Hadi et al. (2017), and Weiss et al. (2004). For consistent measurements across all sites, we took five hemispherical photos per plot that we combined to plot-wise mean LAI measurements: one from the centroid and one from the center of each quadrant. Images were retaken in case of the presence of sunbeams or sun fleck problems. After the field campaign, we processed the RGB hemispherical photographs using CAN-EYE v6.4 open-source software to retrieve effective LAI measurements comparable to PROSAIL's interpretation (Weiss and Baret 2010). We cross-validated the LAI measurements from hemispherical photography with above and below canopy measurements taken with a Photosynthetically Active Radiation (PAR) sensor (Apogee MQ-301; handheld device), quantifying the relative quantity of incident

solar radiation absorbed by vegetation. The LAI observations strongly correlated with the PAR measurements (See Fig. S. 9).

For the leaf trait analysis, we collected 17 leaf samples from the healthy unshaded top of the canopy of each individual plot. Prior visual and geometric inspection guided choices to select the most appropriate samples in terms of areal representativeness. As such, the collected samples approximated a representative composite trait mean of the plots' upper canopy layer. Collected leaves were transported on ice after which they were either dried for LMA analysis or stored in a ≤ -18 °C freezer until CAB analysis in the lab. LMA was calculated based on the dry leaf weight and fresh leaf area (as determined in Image J 1.52a software (Schneider, Rasband, and Eliceiri 2012)). CAB was derived using a protocol based on Lichtenthaler (1987).

Simultaneously with the collection of leaf samples and hemispherical photography, an inventory of the overstorey plant species composition and species number was made of each plot of each location. Parallel to trait and species assessments, local soil spectra were collected using an RS-3500 spectroradiometer (350-2500nm, ~8nm average spectral resolution) developed by Spectral Evolution. These soil spectra aid a representative approximation of PROSAIL's soil parameters for the study site.

3.2.3. Remotely sensed plant traits

3.2.3.1. Sentinel-2 data

The Sentinel-2 constellation consists of two wide-swath, medium-high spatial resolution (10, 20, and 60 m), multi-spectral (13 bands) imaging instruments with a combined 5-day revisiting time (ESA, 2015). Corresponding to the study area and timeframe of our field campaign, one scene (29/07/2019) was completely free of problematic quality flags (cloud cover, cloud shadow, cirrus, and other atmospheric contamination). The availability of a single could-free image prevented complications of having to combine scenes from different timestamps and introducing temporal variability. The Sentinel-2 Level-2a (L2a) imagery for this data was acquired atmospherically corrected (Sen2Cor software) and obtained through the European Space Agency (ESA)'s Copernicus open-access Scientific Hub (Gascon et al., 2014; Louis et al., 2016). We excluded the 60m broad bands from the analysis and resampled the 10m spectral bands to 20m spatial resolution to match the scaling of our georeferenced field plots. Ultimately, ten out of thirteen available spectral bands of Sentinel-2 were used.

3.2.3.2. Radiative transfer model inversion

Sentinel-2 L2a reflectance data served as the foundation for estimating key plant traits. The relationship between spectra, geometry, and soil and vegetation biophysical parameters was modelled with the use of the PROSAIL radiative transfer model inversion, which combines the leaf model PROSPECT (Feret et al. 2008; Féret et al. 2017; Jacquemoud and Baret 1990) and the canopy model 4SAIL (Verhoef 1984; Verhoef et al. 2007). PROSAIL assumes the canopy to be a homogenous turbid medium where absorption is defined by soil, canopy, and leaf properties (Jacquemoud et al. 2006). Such homogeneous canopies are an idealized approximation for many ecosystems. Inversion results are therefore subject to discrepancies between underlying assumptions of the extended 1-D columnar model and the complex reality of the heterogeneous canopies observed in the field (Jacquemoud et al., 2009). PROSAIL's relative simplicity with few but ecologically relevant input parameters avoids further complication of ill-posedness in the non-trivial inversion of multispectral reflectance data

acquired by Sentinel-2 (Verrelst et al., 2019b; Yin et al., 2015). Spectral sensitivity of Sentinel-2 for retrieval of PROSAIL's traits under study (LAI, CAB, and LMA) has been demonstrated in previous sensitivity analyses (Gu et al. 2016; Rossi et al. 2020; Verrelst, Vicent, et al. 2019; de Sá, Baratchi, and Hauser 2021). Considering the model's widespread application, research on its generality across different vegetation types is relevant to a growing body of applications (Jacquemoud et al. 2006; Yin et al. 2015).

We implemented three approaches for inversion of the PROSAIL RTM on Sentinel-2 reflectance data representative of different common approaches found in remote sensing (Ali et al., 2020b; Verrelst et al., 2015); a Look-up Table (LUT)-based inversion based on a non-normalized "least-squares estimator" (LSE) cost function (Rivera et al., 2013; Rossi et al., 2020), the biophysical processor module from the ESA's Sentinel application platform (SNAP toolbox) (Weiss and Baret 2016), and a hybrid PROSAIL-D Support Vector Regression approach (SVR) (Féret et al., 2018, 2017).

Table 3.1: Ranges of variable input parameters of the PROSAIL model used to generate the LUTs.

Domain	Parameter	Symbol	Unit	Distribution	Range
Leaf	Leaf structural parameter	N	–	Uniform	1.4-1.7
	Chlorophyll a+b content	CAB	$\mu g/cm^2$	Gaussian	10-60
	Equivalent water thickness	EWT	g/cm^2	Uniform	0.001-0.045
	Leaf dry mass per area	LMA	g/cm^2	Uniform	0.001-0.040
	Brown pigments content	Cbrown	–	Fixed	0.01
Canopy	Leaf area index	LAI	m^2/m^2	Gaussian	0.01-3.5
	Mean leaf inclination angle	ALA	deg	Uniform	30-70
	Hot spot size parameter	hot	m/m	Fixed	0.01
Abiotic	Ratio of diffuse to total incident radiation	SKYL	–	Fixed	18%
	Soil brightness	psoil	–	Fixed	Spectroradiometer
Positional	Solar zenith	tts	o	Fixed	Sentinel-2 geometry
	Observer zenith	tto	o	Fixed	Sentinel-2 geometry
	Relative azimuth	phi	o	Fixed	Sentinel-2 geometry

3.2.3.2.1. Look-up Table (LUT)-based inversion

The Look-up Table (LUT)-based inversion is a two-step approach entailing:

i. generating a large number of simulations using a RTM, using a given sampling strategy for the input parameter ranges, and
ii. identifying the sample or the set of samples minimizing a cost function.

The estimated biophysical parameter corresponding to a reflectance spectrum is deduced from the value or mean value of the LUT samples minimizing the cost function. Here, we first created an extensive LUT of a subset of 10 000 simulations using Latin hypercube sampling to capture an optimal representation of all possibilities within the relevant trait space based on constrained search ranges defined by minimum and maximum trait values found in the field. Fixed values and ranges of parameters not measured in the field were selected based on literature (Bacour et al., 2002; Jay et al., 2017; Spitters et al., 1986). We used PROSPECT-4 coupled with 4SAIL in order to simulate canopy reflectance. The value, range, and distribution followed for each input parameter of the models are provided in Table 3.1.

For the inversion, we used a non-normalized "least-squares estimator" (LSE) cost function (Rivera et al., 2013; Rossi et al., 2020) and implemented the LUT inversion with the ARTMO toolbox V1.14 in Matlab (Verrelst et al. 2011). We applied the LUT inversion approach to estimate leaf CAB, LMA, and LAI. For optimization of its performance, we applied Gaussian noise (0-18%) to account for model and measurement uncertainties (Rivera et al., 2013). We used ARTMO's default parameterization for the noise and the multiple solution binning.

3.2.3.2.2. Sentinel Application Platform Biophysical processor

The Sentinel Application Platform (SNAP) biophysical processor (Weiss and Baret 2016) is based on a hybrid approach combining physical modeling and machine learning. This type of approach consists of training the machine learning algorithm with a LUT to produce regression models for the estimation of a set of parameters. SNAP uses an artificial neural network (ANN) inversion pre-trained on a PROSAIL simulated database including canopy reflectance and the corresponding set of input parameters. The value, range and distribution followed for each input parameter of the models are described in Weiss and Baret (2016). SNAP can be considered as the standard approach and first port of entry for the estimation of vegetation biophysical parameters, as it is publicly available and easily applicable without strong expertise. SNAP includes an unreleased version of PROSPECT prior to PROSPECT-4, coupled with the SAIL model (Fourty and Baret, 1997). Traits retrieved in SNAP are scaled at canopy-level and include LAI, canopy chlorophyll (CAB*LAI) and canopy water (EWT*LAI). We reversed the multiplication by LAI to arrive at leaf trait estimates (CAB, EWT). In order to derive LMA, we coupled LMA to EWT (fixed factor; EWT*0.79=LMA), a strategy adopted more commonly in RTM inversion exercises given the large spectral overlap of LMA and EWT (e.g. Combal et al. 2003; Weiss et al. 2000; Kattenborn et al. 2017). The strong association between EWT and LMA has been found repeatedly in relevant datasets; in our own in-situ observations, but also the LOPEX/ANGERS, and NEON datasets (Hosgood et al. 1994; Jacquemoud et al. 2003; NEON, 2020).

3.2.3.2.3. PROSAIL-D/Support Vector Regression hybrid approach

We tested an alternative hybrid approach including a physical modeling layer and a machine learning algorithm differing from those used in SNAP. The physical modeling layer included the newer leaf model PROSPECT-D (Féret et al., 2017) coupled with the 4SAIL canopy model (Verhoef et al. 2007). The Support Vector Regression (SVR, Vapnik 1998) algorithm was used as a regression model.

This hybrid inversion consists of bagging prediction of biophysical properties: a LUT is simulated with PROSAIL and resampled in order to produce multiple datasets including a limited number of samples. Then a set of individual support vector regression (SVR) models is trained from each reduced dataset. In our study, we produced 5000 samples and trained 50 SVR models of 100 samples with repetition for each biophysical property. The sampling used to produce the LUT followed the same distribution as showed in Table 3.1, with a few exceptions; as opposed to the fixed soil parameter in the ARTMO LUT implementation, the variability of the soil reflectance was introduced by defining minimum and maximum soil reflectance from experimental data. Soil spectra corresponding to weighted sum of these minimum and maximum soil spectra were generated, with the weight defined by the $psoil$ parameter (Table 3.1), which was randomly sampled following a uniform distribution between 0 and 1.

The geometry of acquisition, including solar zenith, observer zenith and relative azimuth was defined based on a random sampling following a uniform distribution between the minimum and maximum angles corresponding to the different acquisitions. The ratio of diffuse to total incident radiation (SKYL) was defined based on the sun zenith angle under clear sky conditions, following the equations proposed by Spitters, Toussaint, and Goudriaan (1986).

Two additional pigments are included in PROSPECT-D compared to the PROSPECT-4 parameterization: carotenoids and anthocyanins. Both pigment contents were defined based on a random sampling following a uniform distribution. The range for carotenoid content was defined between 0 and 15 $\mu g/cm^2$, which corresponds to the extended range observed in-situ, while the range for anthocyanin content was defined between 0 and 10 $\mu g/cm^2$, which corresponds to the range measured for mature leaves, yellow and reddish senescent leaves (Feret et al., 2017).

Finally, a random Gaussian noise was applied on the reflectance data, in order to account for the uncertainty originating from multiple origins, such as atmospheric conditions, sensor calibration, and the radiative transfer model itself (Berger et al., 2018b), or its improper parameterization (Danner et al., 2019).

3.2.4. Functional and taxonomic diversity estimations

The inversion of PROSAIL with Sentinel-2 spectra allowed us to estimate pixel-based (aggregated) trait values of plant canopies. For both in-situ and remote sensing datasets, functional diversity metrics were calculated for each location. While numerous functional diversity metrics exist, we opted for two commonly used metrics of functional diversity: the convex hull volume (CHV) and Rao's quadratic entropy (Rao's Q) (Dahlin, 2016; Gholizadeh et al., 2018; Rocchini et al., 2017; Torresani et al., 2019). Both metrics are straightforward to compute with relatively few observations, relatively easy to interpret and particularly equipped in characterizing multivariate trait space.

The CHV, a construct from computational geometry, provides an n-dimensional measure of the volume of canopy plant trait space within a community. CHV is commonly proposed as an adequate method to capture continuous trait space (Cornwell et al., 2006) and provides the smallest convex hull that encloses all observed traits. The measure is relatively sensitive to outliers and anomalies (Blonder et al., 2014; Schleuter et al., 2010). Functional richness calculated through the CHV has generally been found to hold a strong relationship to species richness (Mouchet et al., 2010; Schleuter et al., 2010).

Rao's Q is one of the most commonly used multivariate measures of functional diversity and its calculation offers relative mathematical simplicity (Botta-Dukat, 2005; Mouchet et al., 2010; Ricotta and Moretti, 2011; Rocchini et al., 2017; Schleuter et al., 2010). A trait-based implementation of Rao's Q depends both on the range of functional space occupied and on the similarity between trait combinations weighted by abundance (Botta-Dukat, 2005). Hence, elements of both functional richness and functional divergence are part of Rao's Q (Mouchet et al., 2010). Rao's Q has been widely applied to analyse patterns of trait convergence or divergence, i.e. quantifying trait dissimilarity compared to a random expectation (Ricotta and Moretti, 2011). In the remote sensing setting here, Rao's Q describes the sum of pairwise distances between pixel-based multivariate values representing trait estimates while accounting for pixel abundance (in this case; p = 1) (Botta-Dukat, 2005; Rocchini et al., 2017):

$$Q = \sum_{i=1}^{L-1} \sum_{j=i+1}^{L} d_{ij} {}^{*} p_i {}^{*} p_j \quad (1)$$

where d_{ij} corresponds to the multivariate distance matrix comprising *i*-th to *j*-th pixel, p is the pixel or plot value abundance (=1 in our case), and L corresponds to the number of pixels or plots sampled per location.

For the calculation of functional diversity, we took the combination of canopy (LAI) and leaf traits (CAB and LMA) given the ecological importance of these traits for plant functioning (see section 2.2). The combination of these traits allows us to partition canopy reflectance through RTM inversion and focus on canopy structure through LAI, foliar morphology through LMA and leaf chemistry (pigments) through CAB (Rossi et al. 2020; Serbin et al. 2019; Croft et al. 2017; Díaz et al. 2016). Functional diversity was calculated per location in the study area based on the eight to nine plot/pixel-wise trait mean estimates. All three traits were standardized prior to the calculations to assure equal weight to each trait.

Lastly, we calculated taxonomic diversity using in-situ species count data. We relied on the commonly applied Shannon's H diversity index as an indicator of local taxonomic diversity of each location;

$$H = -\sum_{i=1}^{s} (p_i \log_2 p_i) \quad (2)$$

where s is the number of species IDs and p_i is the proportion of the community represented by species i. These calculations were conducted using the scikit-bio 0.5.6 (http://scikit-bio.org/) package in Python. Min-max normalization of all metrics provided all values scaled between 0 and 1.

3.2.5. Statistical analysis

The structure of our data complicated a straightforward application of single goodness-of-fit measures such as R^2 or Pearson's r correlation (Khamis, 2008; Schober and Schwarte, 2018). Therefore, we evaluated the relationship between satellite estimates and the corresponding in-situ measurements by different metrics of association, error and correlation (Lee's L statistic, Spatial Error Models, Spearman's rho (ρ) and Root Mean Square Error (RMSE)).

The plots/pixels (N=97) were non-randomly distributed, which might inflate the correlation and lead to biases. An ordinary least squares linear regression model and Moran's I statistics on the residuals indeed confirmed the presence of spatial autocorrelation in our data (Suppl. Mat. Table S. 10). To account for autocorrelation, we ran Lee's L test (instead of Pearson's r). Lee's L statistic captures the spatial co-patterning by integrating a univariate spatial autocorrelation of each variable (Moran's I) and their bivariate point-to-point association (Pearson's r) (Kim et al., 2018; Lee, 2001). In addition, we ran a spatial error regression model implemented with the "spdep" package of R (Bivand et al., 2011), which controls for the bias of spatially autocorrelated errors. We evaluated the model's performance for spatial

autocorrelation by reassessing the residuals with Moran's I, which showed that spatial autocorrelation was no longer significant (Suppl. Mat. Table S. 11).

In addition, we evaluated the correspondence of satellite and in-situ observations at the location-level. The means and standard deviations at the location-level gave insight in local variability of estimates and potential robustness of aggregation against noise and misregistration. Moreover, location-wise trait means eliminate the effects of spatial auto-correlation resulting from multiple plots/pixels per locations. The locations are randomly distributed over our study area. However, the small sample size of locations (N=12) makes it difficult to warrant for normality. In precaution of non-normality, we implemented Spearman's rho, as the non-parametric and rank-based alternative to Pearson's r (Fowler, 1987; Khamis, 2008; Schober and Schwarte, 2018). In addition, we calculated the Root-Mean-Square Error (RMSE) to quantify absolute biases for both plot-/pixel-level and location-level analyses.

The functional diversity metrics (CHV, Rao's Q) derived from satellite observations were compared against in-situ functional diversity observations. In addition, the satellite-based functional diversity metrics were compared against in-situ taxonomic diversity to evaluate the ecological relevance of the selected RTM-based traits for assessing species-based biodiversity and conservation planning. For the analyses of functional diversity patterns, we used Spearman's rho to warrant for possible non-normality in the small sample size (N=12).

3.3. Results

3.3.1. Trait estimates

Table 3.2 shows the performance of the three retrievals against in-situ trait measurements, both at pixel-level and location-level aggregation. All three methods estimated the distribution of trait values from Sentinel-2's spectral reflectance reasonably well. LAI was estimated with relatively high precision across all three methods (Table 3.2). Across approaches, forest locations generally exhibited higher LAI values compared to shrublands in line with field observations.

LMA estimates corresponded well with the variation in in-situ measurements, with SVR and ARTMO performing better than SNAP. Despite relative correspondence with the field data, ARTMO exhibited a strong bias, overestimating absolute in-situ LMA observations. SVR produced accurate estimators for LMA both in a relative and absolute sense. All three approaches consistently identified generally higher LMA values in shrublands as compared to forested locations.

CAB retrievals were still significantly correlated to the in-situ measurements at plot-level, but with weaker relationships and stronger deviations from the 1:1 line compared to the other traits for all three retrieval methods (Table 3.2).

Spatial error (regression) models (Table S. 11) accounting for spatial autocorrelations at the plot level indicated a strong significance of the satellite-based estimates and high overall predictive power of the models (Pseudo-R^2: 0.76-0.93, depending on trait and inversion method) with the exception of ARTMO's CAB estimation (Not significant, Pseudo-R^2: 0.35).

Despite the smaller sample size, the location-level aggregation of trait means retained a strong association between in-situ measurements and the satellite-based estimates for LMA and LAI,

while CAB estimation by SVR's hybrid inversion and SNAP's biophysical processor were no longer significant (>0.05) (Table 3.2).

The retrieval of within-location standard deviation was less convincing overall compared to mean estimates, although LAI variability correlated significantly for all three inversion methods (Table 3.2). Only ARTMO's LUT-based inversion performed was significantly correlated to the in-situ variability of LMA (Fig. 3.3). Retrieval for SVR and ARTMO is illustrated in Fig. 3.2 and Fig. 3.3, respectively, while figures for retrieval based on SNAP's biophysical processor can be found in Fig. S. 12. The patterns of location-wise standard deviations did not differ evidently between forested and shrubland locations.

Table 3.2: An overview of satellite-based single trait estimation of the three RTM inversion algorithms (SVR, ARTMO, and SNAP) compared against in-situ field measurements. Validation was done both at the individual pixel-level (in-situ: plot) and the location-level aggregation. Measures of association indicate the strength of correlation between in-situ measured and satellite estimated trait values. At the pixel/plot-level, Lee's L statistic was implemented as an alternative to Pearson's r to account for spatial autocorrelation within spatially neighbouring plot/pixel observations. At the location-level, Spearman's rho (ρ) was implemented to warrant for the possible non-normality in small sample sizes (N=12). RMSE and nRMSE are indicative of the absolute and relative error found in trait means.

Algorithm	Trait	Pixel-level trait estimates			Location-level trait means		Location-level trait standard deviations	
		Lee's L	RMSE	nRMSE (%)	Spearman's ρ	RMSE	Spearman's ρ	RMSE
SVR (Hybrid)	Leaf Area Index (m2/m2)	0.70**	0.49	15.71	0.73**	0.43	0.67**	0.18
	Leaf Mass per Area (mg/cm²)	0.96**	6.36	17.78	0.77**	5.77	0.46	1.20
	Leaf Chlorophyll (μg/cm²)	0.52**	7.07	17.35	0.53	5.80	0.04	1.87
ARTMO (LUT)	Leaf Area Index (m2/m2)	0.72**	0.51	16.24	0.81**	0.44	0.66*	0.13
	Leaf Mass per Area (mg/cm²)	0.79**	31.06	34.43	0.83**	30.25	0.73**	4.70
	Leaf Chlorophyll (μg/cm²)	0.29**	12.03	28.45	0.71**	9.62	0.49	2.63
SNAP	Leaf Area Index (m2/m2)	0.68**	0.62	19.84	0.73**	0.58	0.76**	0.16
	Leaf Mass per Area (mg/cm²)	0.71**	5.94	23.46	0.71**	5.50	0.24	1.62
	Leaf Chlorophyll (μg/cm²)	0.41**	13.79	28.54	0.39	13.23	-0.24	2.61

**: Significant correlation (p <0.01), *: Significant correlation (p <0.05), ns: Not significant (p>0.05)

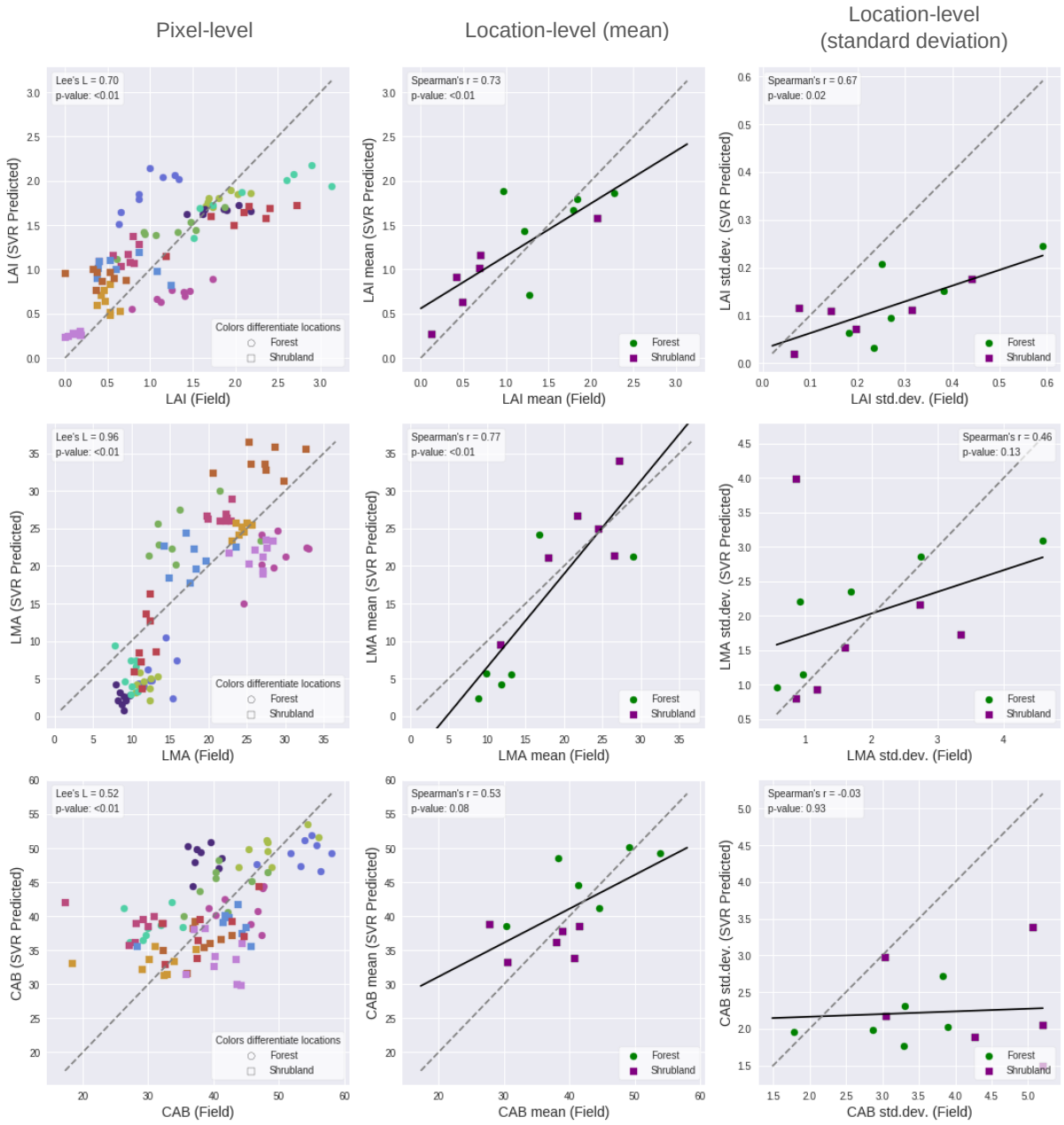

Fig. 3.2: Comparison of Sentinel-2-based trait estimates (y-axis) retrieved using a Support Vector Regression (SVR) hybrid inversion algorithm against in-situ field measurements (x-axis). The left column shows pixel-level (in-situ: plot) comparisons of traits, where different colors indicate plots of respective locations. The middle column depicts trait means per location and the right column presents trait standard deviations per location. The grey dotted line shows the 1:1 relationship, whereas the black line indicates the fitted linear relationship between the remotely sensed estimates and field data. Purple and green markers represent shrubland and forested locations, respectively.

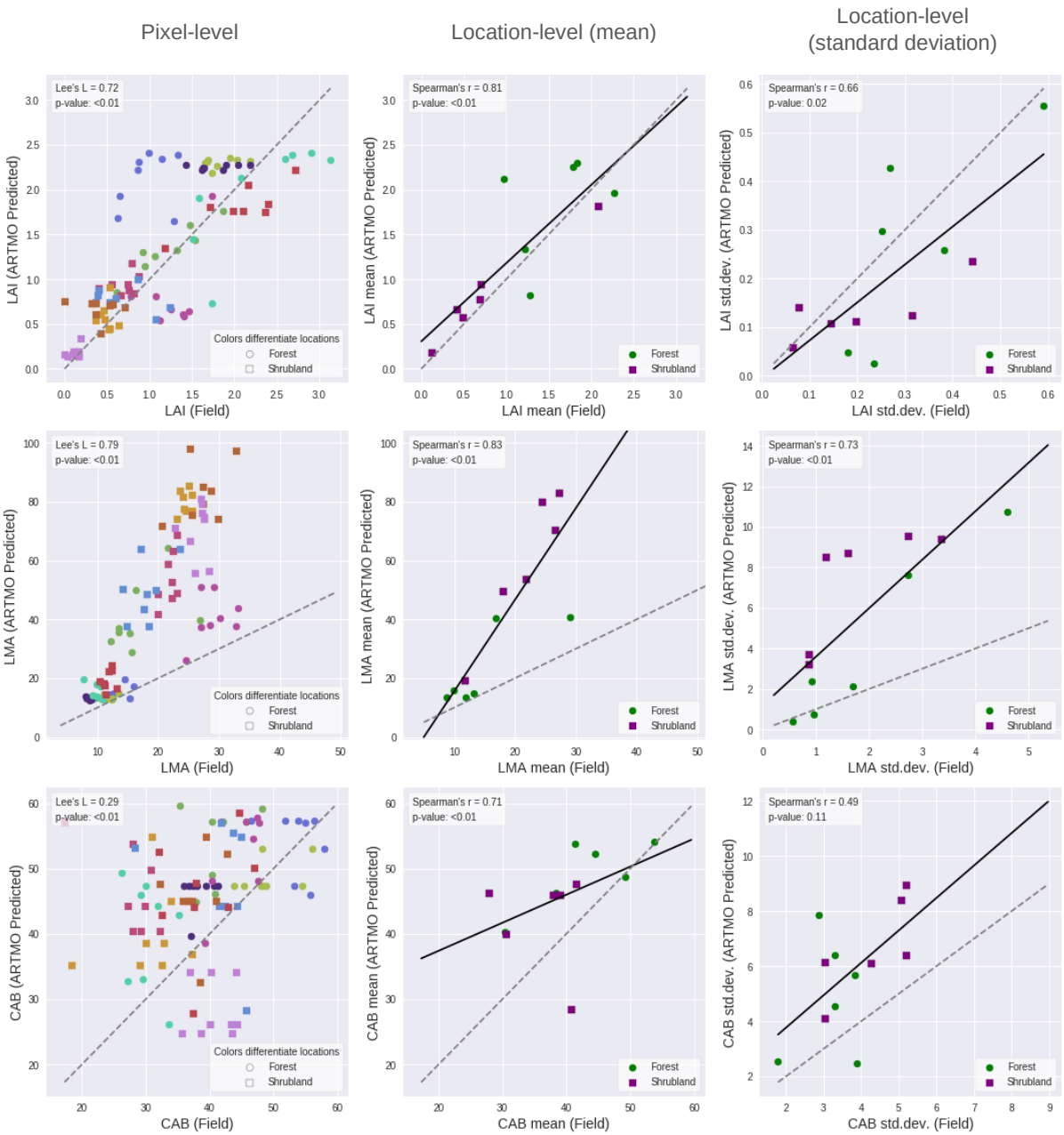

Fig. 3.3: Comparison of Sentinel-2-based trait estimates (y-axis) retrieved using ARTMO LUT-based inversion against in-situ field measurements (x-axis). The left column shows pixel-level (in-situ: plot) comparisons of traits, where different colors indicate plots of respective locations. The middle column depicts trait means per location and the right column presents trait standard deviations per location. The grey dotted line shows the 1:1 relationship, whereas the black line indicates the fitted linear relationship between the remotely sensed estimates and field data. Purple and green markers represent shrubland and forested locations, respectively.

3.3.2. Functional diversity estimates

Despite the small sample size (N= 12) and availability of only 8-9 trait observations per location to calculate functional diversity, functional diversity exhibited a significant relationship between in-situ and satellite-based estimates in most cases. The CHV metric was significantly associated with in-situ functional diversity across all three inversion algorithms. For Rao's Q, this significant relationship only holds for SVR and LUT-based ARTMO trait retrieval. Yet, in general, the three approaches indicate feasibility in predicting in-situ plant functional diversity through satellite-based estimates (Table 3). Fig. 3.4 shows the results of the SVR and ARTMO inversion approaches. The results of the SNAP inversion can be found in Fig. S. 13.

Table 3.3: Rank-based correlation between in-situ observed plant functional diversity (CHV and Rao's Q) and satellite remote sensing observed functional diversity. Calculations of functional diversity combine canopy trait (LAI) and leaf-level traits (LMA and CAB). Significant correlations ($\alpha < 0.05$) are highlighted in bold. RMSE and nRMSE are indicative of the absolute and relative error found in functional diversity estimates.

	Algorithm	Spearman's ρ	Sig.	RMSE	nRMSE (%)
Convex Hull Volume (CHV)	SVR (Hybrid)	0.80	<0.01	0.11	20.48
	ARTMO (LUT)	0.76	<0.01	0.09	20.94
	SNAP	0.67	0.02	0.11	25.74
Rao's quadratic entropy	SVR (Hybrid)	0.75	0.01	0.13	22.08
	ARTMO (LUT)	0.80	<0.01	0.08	19.28
	SNAP	0.19	0.56	0.19	43.06

Remotely sensed functional diversity metrics were also significantly tied to in-situ community taxonomic diversity (Fig. 3.4, rightmost column). This significant relationship indicates that the selected traits are relevant for both trait and species diversity and serves as cross-validation of the capability of RTM inversion of Sentinel-2 spectral information to predict ecologically relevant in-situ plant biodiversity, either directly or through surrogacy. Functional diversity and taxonomic diversity were also related in-situ (Fig. S. 14).

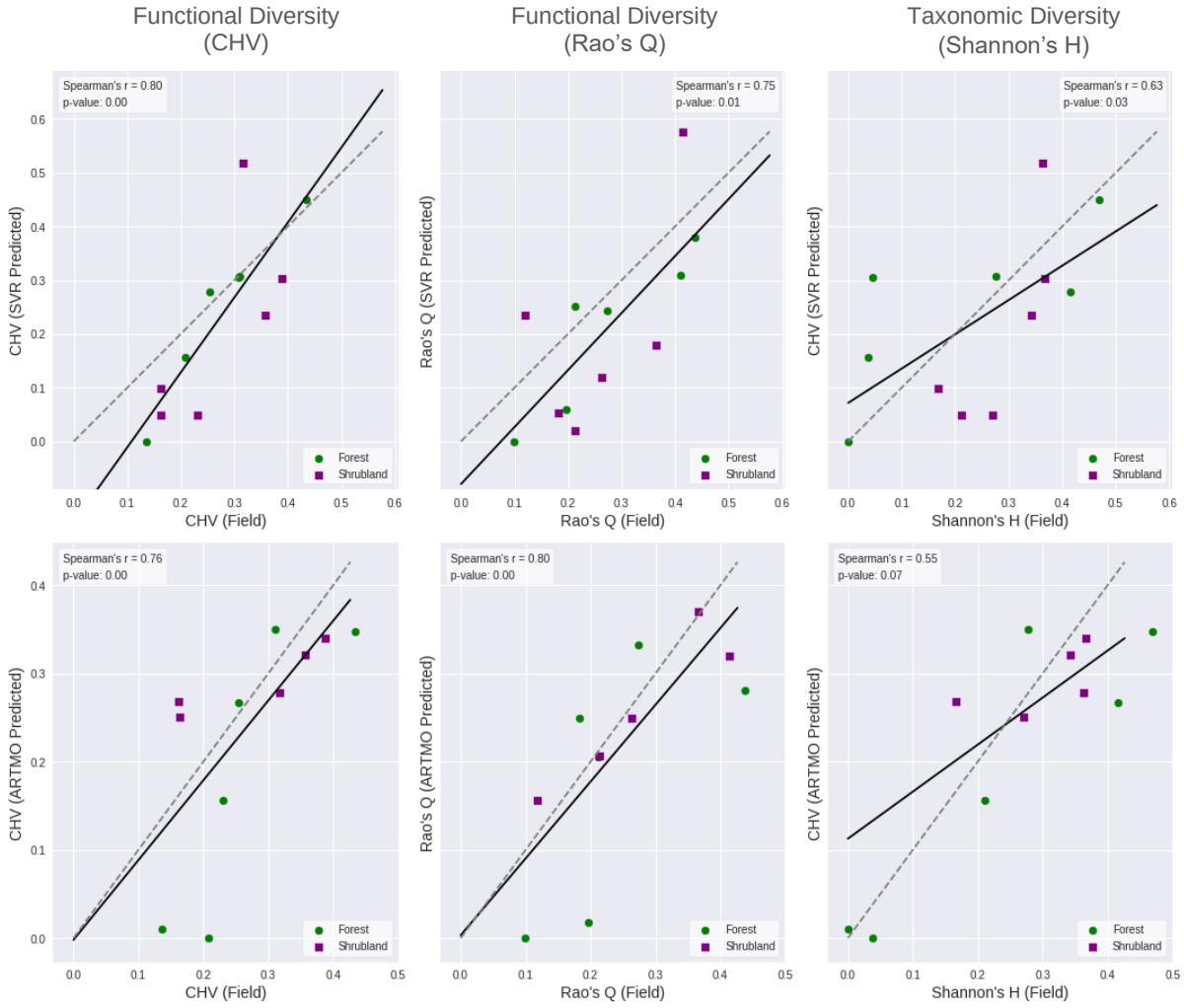

Fig. 3.4: Remotely sensed functional diversity estimates (CHV and Rao's Q) calculated from Sentinel-2 derived traits through the SVR hybrid inversion (top row) and ARTMO LUT-based inversion (bottom row) compared against in-situ functional diversity measurements (left and center columns). The rightmost column compares the remotely sensed functional diversity (CHV) against in-situ taxonomic diversity (Shannon's H). The grey dotted line shows the 1:1 relationship, whereas the black line indicates the linear relationship between the remotely sensed estimates and field data. Purple and green markers represent shrubland and forested locations, respectively.

3.4. Discussion

3.4.1. Estimating canopy and leaf traits

The capability to estimate spatial plant canopy trait patterns from currently operational optical satellite remote sensing, both in terms of mean and variability, serves as an important starting point towards the assessment of satellite-based functional diversity estimates. Estimation of individual plant traits from Sentinel-2 inference using RTM inversion has been shown to be viable in numerous previous studies conducted in relatively homogeneous (semi-)natural environments (Ali et al., 2020a, 2020b; Brede et al., 2020; Brown et al., 2019; Darvishzadeh et al., 2019a, 2019b; Padalia et al., 2020; Rossi et al., 2020; Vinué et al., 2018). Similarly, in our study, we demonstrated the potential of remote sensing to estimate the relative spatial distribution of multiple individual vegetation traits simultaneously, yet in a relatively

heterogeneous landscape using a comparatively simple RTM. Unique to this study, we assessed a multivariate retrieval of three traits (LAI, LMA and CAB) in twelve separate locations consisting of multiple adjacent pixels (N=97) in a set-up designed for further assessment of plant community patterns of functional diversity (Fig. 3.1). The georeferenced and carefully scaled in-situ field measurements allowed us to validate satellite remote sensing directly without relying upon interpolation of point data and/or temporally and/or spatially transposed secondary trait data (e.g. Ma et al. 2019; Butler et al. 2017; Moreno-Martínez et al. 2018).

In light of the multitude of existing inversion methods, we ran three different algorithms to invert PROSAIL on Sentinel-2 reflectance data (Verrelst, Malenovský, et al. 2019; Rivera et al. 2013; Verrelst et al. 2015). Despite the heterogeneity of the canopies and the landscape, all three retrieval approaches (SVR, ARTMO (LUT-based) and SNAP) showed significant correlations between the estimates of the various traits and the actual in-situ measurements (Table, 2, Fig. 3.2 and Fig. 3.3). Significant relations were found for pixel-wise trait estimates as well as aggregated trait means at the location level. The good performance of the latter is noteworthy as aggregation potentially attenuates the influence of noise. The Sentinel-2 retrieval performed specifically well for estimating the relative distribution of LAI and LMA despite biases (RMSE), whereas CAB exhibits a considerably less association with field measurements (Table 3.2). Differences in the performance of CAB may be attributed to its parameterization in different versions of PROSPECT, as well as strong absorption of radiation in the visible range and associated low signal-to-noise ratios.

Differences in performance between the three algorithms is in part due to parameterization and the mathematical intrinsic properties of the individual inversion methods. However, also the role of ‘a priori’ information deserves further attention (Verrelst et al. 2015; Verrelst, Malenovský, et al. 2019). In this study, the ARTMO LUT-based and SVR hybrid approaches both rely on PROSAIL simulations that are generated from the minimum and maximum trait ranges found in the field, while the SNAP biophysical processor runs completely independently of ancillary data (Weiss and Baret 2016). Databases like TRY (Kattge et al. 2020) can aid future analyses to generate locally optimized trait ranges based on species occurrences without the need for dedicated field campaigns. The optimized LUTs, taking into account the trait ranges of the ecosystem under study, enhanced performance in terms of the relative prediction of in-situ trait and the absolute bias.

For an accurate assessment of functional diversity, proper estimation of local trait variability is important, which has received little attention in the body of research on the retrieval of plant traits from satellite remote sensing thus far. In comparison to location-level trait means, our ability to retrieve trait variability is relatively inconsistent (Table 3.2). The observed inconsistencies might require additional accounting of adjacency effects, noise/inconsistencies in field measurements, noise from atmospheric correction (5-10%) and reported multispectral misregistration of Sentinel-2 that complicate the process of deciphering subtle differences at a local scale (Brede et al., 2020; Skakun et al., 2017).

Atmospheric scattering causes light reflected from adjacent landscapes to be observed by the sensor. Without adequate correction of possible adjacency effects (being only an option within the Sen2Cor processor, with a low resolution of 1km), scattering leads to minor biases or spectral convergence (Louis et al., 2016). While the local mean traits would be affected little,

spectral convergence could result in a lower sensitivity to capture the variation between neighboring pixels and thereby affect functional diversity estimates.

As an exception, the results for LAI still indicate high sensitivity across the three retrieval algorithms to accurately depict local variability in LAI between adjacent pixels and plots (Table 3.2) despite a general underestimation. LAI is known to exhibit a relatively strong spectral response across large part of Sentinel-2's spectrum, which could favor the capability in detecting small changes and an improved signal-to-noise ratios as opposed to other traits (Asner 1998).

This validation of multivariate single plant traits demonstrates how an RTM approach can present us with a semi-mechanistic way to retrieve plant traits based on the physical principles of radiative transfer, even if limited field data is available. For now, field data remains critical for validation. More specifically, further independent testing will be needed to examine the scalability of the approach across multiple diverse ecosystems, canopy types and even biomes. Accuracies of simultaneous retrieval of multiple traits can be further improved with RTMs optimized for the specific vegetation types under study. However, RTM selection and configuration needs to be done under consideration of the trade-offs between local optimization and generality across heterogeneous landscapes, as well as the ill-posedness induced by heavy parameterization of complex optimized models in inversion applied to limited spectral information (Huang et al., 2011). In the future, data assimilation developments (Lewis et al. 2012) and the prospective launch of hyperspectral satellite imagers (e.g. EnMAP, SBG, CHIME; Cavender-Bares, Gamon, and Townsend 2020; Lahoz and Schneider 2014) will likely enable more detailed spectral information for inversion to accurately derive more traits for functional diversity assessments.

3.4.2. Estimating multivariate functional diversity

The retrieval and validation of single trait estimates and the relative spatial variability thereof serves our study's core objective of estimating plant functional diversity from satellite remote sensing. The study demonstrates a methodology for one-to-one scaled ground-truthing of functional diversity based on satellite observed mean trait estimates for aggregate canopies. We tested two different metrics (CHV and Rao's Q) to calculate functional diversity. Functional diversity was calculated over three traits combining canopy structure through LAI, and foliar morphology through LMA, and leaf chemistry (pigments) through CAB as key functional traits (Croft et al., 2017; Díaz et al., 2016; Rossi et al., 2020; Serbin et al., 2019).

Across different setups, the majority of satellite-based functional diversity metrics corresponded significantly with in-situ measurements. The convex hull volume (CHV) operationalization of functional diversity showed strong *concordance* with in-situ observations across all three inversion methods (Spearman's ρ: 0.67-0.80). For the interpretation of these results, we highlight the implications of the distance-based nature of the functional diversity metrics and the statistical power of the study in the following paragraphs.

The functional diversity metrics used here, and many of the multivariate alternative metrics (e.g. Schleuter et al. (2010), Mouchet et al. (2010), Aiba et al. (2013)), rely on the quantification of "diversity" or "entropy", based on relative distances or volume between trait observations set in a n-dimensional space in which each axis represents a specific trait. Accordingly, the

accurate prediction of functional diversity is determined predominantly by the normalized relative position of trait combinations rather than the absolute trait value deviation itself. As such, the functional diversity metrics presented here have shown a degree of robustness against biases in RMSE, specifically for CHV calculations. For instance, the SVR hybrid inversion reveals better RMSE for LAI, LMA and CAB (Fig. 3.2), especially compared to the overestimation of LAI by ARTMO (LUT-based) and the inferior accuracy of SNAP, yet the functional diversity estimates show a level of robustness to these errors given the reasonable performance across retrieval algorithms (See Table 3.3, Fig. 3.4, and Fig. S. 13).

The interpretation of the results needs to be done in consideration of the sample size which is dictated by the laboriousness of validating functional diversity at the scale of satellite pixels. Our field campaign involved comprehensive efforts to representatively sample the composition of the dominant canopies, based on an average of 17 sampled sunlit leaves from individual branches, within each of the eight to nine plots across each of the twelve semi-natural shrubland/forested locations. Despite these efforts, the statistical power of the twelve points for functional diversity estimation imposes limitations on the confidence and our ability to attest the patterns found between remotely sensed estimates and in-situ observations. In addition, each functional diversity calculation only relies on eight to nine individual observations. Statistically, we know that a larger number of individual observations allows us to better characterize functional diversity and result in higher robustness against noise, misregistration, and random artifacts (Frank, 2009; Hubbell, 2001; Steinbauer et al., 2012). This robustness is relevant considering that convex hull volumes are particularly sensitive to outliers and that satellite remote sensing, including atmospheric correction, remains susceptible to unfavorable signal-to-noise ratios (Blonder et al. 2014; Lewis et al. 2012; Brede et al. 2020; Verrelst, Vicent, et al. 2019; Skakun et al. 2017).

3.4.3. Ecological implications

The validation results show the potential of using satellite remote sensing to estimate spatial patterns of plant functional diversity. However, for ecological implementation, we need to go beyond validation and delve into the ecological relevance and usefulness of satellite remote sensing products quantifying functional diversity (Feilhauer et al., 2018). Here, we highlight three aspects - the selection, scale, and number of traits, the spatial resolution of the functional diversity metrics, and the need for further validation – as considerations for interpreting and improving the ecological relevance of remotely sensed metrics.

The selection of which and the number of traits to include in functional diversity measurements is a critical step that will impact the patterns ultimately observed (Legras et al., 2020; Petchey and Gaston, 2006; Tsianou and Kallimanis, 2016). The majority of studies render functional diversity using traits largely determined by data availability despite that ideally such decisions are driven by the specific function of interest (Petchey and Gaston, 2006; Tsianou and Kallimanis, 2016). Likewise, in direct retrieval through (optical) remote sensing, we are limited to traits that are spectrally obtainable (Homolová et al., 2013). Exemplar prioritized lists of key functional plant traits that are remotely observable from space have been compiled by e.g. Jetz et al. (2016) and the National Academies of Sciences Engineering and Medicine (2019) among others.

LMA and CAB implemented in this study feature among those prioritized lists (Jetz et al., 2016; National Academies of Sciences Engineering and Medicine, 2018) and are part of the PROSAIL input parameters. These traits allow us to characterize part of the foliar morphology through LMA and leaf chemistry (pigments) through CAB. These capture the tradeoffs of a plant's investment in leaf structure, robustness, versus leaf surface area and per area capacity for photosynthesis (Croft et al., 2017; Díaz et al., 2016; Rossi et al., 2020; Serbin et al., 2019). Functional diversity can both be driven by intra- and interspecific variation. Our field observations indicate that higher taxonomic diversity (Shannon's H) in the study area does translate to higher functional diversity (Fig. S. 14), implying that the observed species exhibit distinct functional profiles shaping trait space. While the implemented traits might not fully capture all aspects of functional importance, the selection does significantly resonate with the taxonomic composition. This relationship is relevant, as species diversity is still the most commonly used indicator by decision-makers in conservation planning and biodiversity monitoring (Gaston, 2010; Petchey and Gaston, 2002).

In agreement with this, functional diversity (based on LAI, LMA, and CAB) captured from satellite remote sensing indeed also showed a relationship to in-situ community taxonomic diversity (Fig. 3.4). Fig. S. 15-Fig. S. 18 indicate that the inclusion of LAI as a key trait of functional diversity actually weakens the relationship between trait diversity and taxonomic diversity both in in-situ and satellite-derived observations. Diversity in LAI seems to be relatively orthogonal to the taxonomic diversity in the dataset of our study area. As such, functional diversity measures only involving the leaf traits (LMA and CAB) resulted in stronger significant relationships to taxonomic diversity, both for in-situ and satellite-based observations (See Fig. S. 15-Fig. S. 18). Given the dominance of LAI on the canopy reflectance signal, proper estimation of LAI and diversity thereof is still highly important to accurately obtain canopy-level traits and structural diversity (Asner et al. 2015; Roelofsen et al. 2013).

These findings illustrate the importance of trait selection considerations in functional diversity assessments and decisions to be made regarding the scale at which we look at traits; at a leaf-level or at the canopy-level. The incorporation or multiplication with LAI facilitates a relatively straightforward approach to upscale leaf traits to the canopy scale (Bacour et al., 2006; Kattenborn et al., 2019; Musavi et al., 2015). The retrieval of canopy traits for functional diversity, either through incorporation (see Fig. 3.4) or multiplication with LAI, resulted in higher accuracies as compared to solely through leaf-level traits as illustrated here by SVR and SNAP inversions (see Fig. Fig. S. 15 and Fig. S. 17). The improved retrieval accuracy of canopy traits compared to leaf traits has also been noted by e.g. Bacour et al., 2006; Homolová et al., 2013; Roelofsen et al., 2013. However, for our study area, this better retrieval seems to come at the cost of lesser relationship to taxonomic diversity.

Sentinel-2's pixel size, with most bands scaled at a 20m spatial resolution, thus resulting in plots of $400m^2$, is significantly larger than most plant canopies. This mismatch between ecology's sampling units and the homogeneous coarse raster offered by satellite remote sensing complicates ecological interpretation (Wang and Gamon, 2019). The calculation of functional diversity requires multiple observations embodied in pixel-scaled plots. The calculations provided a characterization of functional diversity in ecological communities captured through multiple pixels covering around 3200 m^2. As such, for ecological interpretation, the remotely

sensed functional diversity metrics likely capture both elements of alpha and beta-diversity indicative of the local turnover of functional diversity in arbitrary ecological communities (Barton et al., 2013; Rocchini et al., 2015).

Besides ecological interpretation, this scale mismatch also challenges validation efforts. Our field campaign carefully followed Sentinel-2's spatial scaling resulting in comprehensive but laborious sampling efforts of large plots. Our spatial scaling enabled a unique one-to-one validation of remotely sensed estimates against in-situ community functional diversity patterns. The results are indicative of current capabilities in mapping functional diversity from space through RTM inversion approaches (Table 3.3 and Fig. 3.4), yet remain ambiguous in terms of sample size and the number of observations on which functional diversity is calculated. An injection of finer scaled remote sensing approaches, such as a two-tier validation between field, UAV and satellite remote sensing may provide a better characterization of the spatial scaling of functional diversity patterns observed from space, and potentially facilitate feasible validation campaigns that include a larger number of observations. Complementary, wall-to-wall (i.e. spatially explicit) landscape retrieval approaches of functional diversity from remote sensing would allow further assessment of the validity of remotely sensed functional diversity metrics along well-studied ecological gradients (Durán et al., 2019). Finally, interpolation of species-mean trait estimates based on carefully curated trait and species occurrence databases may complement the extensive sampling of field observations presented here to build a more complete validation dataset, possibly covering multiple ecosystem types, regions and or biomes (e.g. Aguirre-gutiérrez et al., 2021; Ma et al., 2019; Serbin et al., 2019).

3.5. Conclusions

We demonstrated the potential of RTM inversion of Sentinel-2 to simultaneously derive multiple relevant traits to calculate satellite remotely sensed functional diversity estimates in a (semi-)natural heterogeneous landscape. The implementation of our study design allowed for a unique direct one-to-one validation of individual traits and community patterns of functional diversity based on in-situ measurements that are precisely scaled and georeferenced to the satellite observed pixels. The implemented general RTM approach is relevant for wider application in ecosystem and biodiversity research as it allows for the retrieval of in-situ measurable multivariate trait estimations from reflectance without heavy reliance on field data for input. The approach is semi-mechanistic and the universal principles behind its physics are in theory scalable, although further research across ecosystems and canopy types is necessary. Across a representative selection of different inversion approaches, the functional diversity metrics (CHV in particular) appear relatively robust against errors in trait retrievals. Taken together, the study provides an important step towards maturation of operational, scalable, spatially and temporally explicit methods and, hopefully, inspires further validation and assessment of satellite-based functional diversity metrics across ecological gradients and larger datasets. A validated assessment will allow monitoring large-scale patterns of plant functional diversity to better understand the dynamics of functioning of our ecosystems.