Remote Sensing of Environment 335 (2026) 115294

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Testing the accuracy and transferability of remotely sensed biomass models across heterogeneous grasslands

Jan Schweizer [a,b,*], Leon T. Hauser [a,b], Hamed Gholizadeh [c], Anna K. Schweiger [d], Christian Rossi [a,b,c]

[a] Department of Geoinformation, Swiss National Park, Runatsch 124, 7530 Zernez, Switzerland
[b] Department of Geography, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland
[c] Department of Geography, Oklahoma State University, 318 Social Sciences & Humanities, 74078 Stillwater, OK, USA
[d] Department of Land Resources & Environmental Sciences, Montana State University, 334 Leon Johnson Hall, 59717 Bozeman, MT, USA

ARTICLE INFO

Keywords:
Model comparison
Empirical models
Physically-based models
Hybrid retrieval
Active learning
Sentinel-2
LUT inversion
PROSAIL
Gaussian process regression

ABSTRACT

Grassland aboveground biomass provides key insights into ecological processes such as carbon sequestration, animal movement patterns, and agricultural management practices. Different model types have been developed to estimate grassland biomass from satellite imagery. However, differences in model performance across sites with varying management and ecology remain largely understudied. In this study, we compared accuracy and transferability of empirical, physically-based, and hybrid models to estimate grassland biomass from multi-spectral Sentinel-2 data in an agnostic scenario, i.e., the models were not provided with any site-specific information beyond the spectral data. Based on field data from five study sites in Europe and the United States, we assessed (1)site-level accuracy of biomass estimation models, (2) model transferability between sites (domain shift), (3) the performance of models trained or optimized with data from multiple study sites (domain generalization), and (4) the relationship between epistemic uncertainty and model transferability. Our results showed that (1) all models exhibited comparable performance at the site level, (2) physically-based models showed the highest degree of transferability between sites, (3) no model consistently outperformed all other models when trained or optimized with field data from multiple sites, and (4) epistemic uncertainty was not necessarily a reliable measure of model applicability to unseen data. Our findings demonstrate the challenges associated with grassland biomass models under domain shift. This elucidates limits to agnostic inference in targeting diverse grasslands and highlights that model transferability is an integral part of performance assessment towards scalable satellite-based grassland monitoring systems, especially as the community increasingly deploys models at continental to global scales.

Edited by Jing M. Chen

1. Introduction

Grasslands cover up to 40% of Earth's terrestrial surface (White et al., 2000) and two-thirds of the Earth's agricultural land area (O'Mara, 2012). With their extensive land coverage, grasslands store around a third of the global terrestrial carbon (Bai and Cotrufo, 2022) and host many endemic species (Hobohm and Bruchmann, 2009). At the same time, grasslands provide essential ecosystem services (Lemaire et al., 2011; Zhao et al., 2020) including global food production (Bengtsson et al., 2019; O'Mara, 2012). Given their ecological, cultural, and economic importance, accurate monitoring of grasslands is imperative to counteract declines in biodiversity and ecosystem services (Bardgett et al., 2021).

Within the overall objective of grassland monitoring, the accurate and reliable estimation of aboveground biomass (hereafter referred to as biomass) is critical for quantifying numerous ecological processes and effects of human disturbances. Biomass is an important parameter for Earth System Models (e.g., Lawrence et al., 2019) and for estimating the contribution of grasslands to the global carbon cycle (Erb et al., 2018). Grassland biomass is also a key driver of animal movement and grazing patterns (Bailey et al., 1996; Rempfler et al., 2024; Schweiger et al., 2015b), and provides information about management practices such as

* Corresponding author at: Department of Geoinformation, Swiss National Park, Runatsch 124, 7530 Zernez, Switzerland.
E-mail address: jan.schweizer@nationalpark.ch (J. Schweizer).

https://doi.org/10.1016/j.rse.2026.115294
Received 17 April 2024; Received in revised form 19 January 2026; Accepted 2 February 2026
Available online 9 February 2026

mowing (De Vroey et al., 2022), whose timing and frequency are linked to biodiversity (Socher et al., 2012; Van Vooren et al., 2018) and productivity (Zhang et al., 2023).

For large-scale grassland biomass estimation, spaceborne remote sensing enables repeated observations across large spatial domains. Optical sensors measuring surface reflectance, such as those onboard the European Space Agency's (ESA) Sentinel-2 and the National Aeronautics and Space Administration's (NASA)/United States Geological Survey's (USGS) Landsat satellites, are widely used to examine vegetation dynamics of grasslands (Reinermann et al., 2020). Open data policies have supported the availability of many years of archived data (Gascon et al., 2017; Masek et al., 2020; Roy et al., 2014). The spectral layout of Sentinel-2 has shown to hold viable information to estimate foliar properties and canopy structure used to infer biomass (de Sá et al., 2021; Guerini Filho et al., 2020; Hauser et al., 2021a; Rossi et al., 2020). However, many grasslands are subject to high levels of spectral complexity arising from several factors, including effects of non-photosynthetically active vegetation (NPV; Xu et al., 2014), co-occurring plant functional types (Dixon et al., 2014), and management regimes affecting grassland phenology and species composition (Ali et al., 2016; Rossi et al., 2024).

This complexity led to the emergence of different model types to estimate biomass, including the use of vegetation indices (VIs), empirical, physical, and hybrid models. Each of these model types has its specific trade-offs regarding required field data, model complexity, specificity, and transferability (i.e., domain shift). Vegetation indices (VIs), such as the Normalized Difference Vegetation Index (NDVI; Guerini Filho et al., 2020; Li et al., 2016; Wang et al., 2019), are found to correlate to various vegetation properties at both the leaf and canopy level. They are straightforward to use, but can saturate with high amounts of biomass (Huete et al., 2002; Zeng et al., 2023). Empirical models are statistical models trained and validated using field data. Deriving relationships from the training data, they do not rely on prior knowledge about the relationship between input and output variables, and are generally well-equipped to handle non-linearity and noise often present in remote sensing data (Verrelst et al., 2015). Numerous empirical machine and deep learning models such as Random Forest regression (RFR), Support Vector regression (SVR), Extreme Gradient boosting (XGB), Gaussian process regression (GPR) and Deep Neural Networks (DNN) have been successfully used to estimate grassland biomass and vegetation traits from optical Sentinel-2 and Sentinel-1 synthetic aperture radar (SAR) data (Li et al., 2021; Muro et al., 2022; Raab et al., 2020; Schwieder et al., 2020; Verrelst et al., 2012). Physically-based Radiative Transfer Models (RTMs) simulate the interactions between light and matter at leaf- and canopy-scales reducing reliance on field data and improving domain generalization by leveraging universal physical principles (He et al., 2019; Wang et al., 2023). RTM inversion can be achieved by using a look-up table (LUT) approach (Verrelst et al., 2014), which connects simulated or measured spectra with trait combinations linked to those spectra. For example, the PROSAIL RTM (Jacquemoud et al., 2009) can be inverted to estimate grassland biomass derived by multiplying leaf dry matter content with leaf area index (LAI; see e.g., He et al. (2019)). However, the PROSAIL RTM is based on heavily idealized assumptions, such as uniformly distributed leaf constituents and geometrically homogeneous canopies (Jacquemoud and Baret, 1990; Verhoef et al., 2007), that are never met in reality. Hybrid models combine RTM-simulated canopy spectra with machine learning models for model inversion (Verrelst et al., 2015). Active Learning (AL) is often used in combination with hybrid models to select the most informative training samples (Verrelst et al., 2016) to perform the RTM inversion which is inherently ill-posed (Combal et al., 2003), coming however at the potential cost of model transferability (Berger et al., 2021b; Tagliabue et al., 2022). Hybrid models have been successfully used for estimating vegetation properties in croplands (Berger et al., 2021a; Berger et al., 2020; Ranghetti et al., 2022; Tagliabue et al., 2022; Verrelst et al., 2021; Wocher et al., 2022) and forests (Binh et al., 2022; Brown et al., 2019; Hauser et al., 2021a; Yuan et al., 2015). However, the findings from these previous studies cannot be simply extended to grasslands, since grasslands are more chemically and structurally diverse compared to croplands and differ in plant size and canopy characteristics from forests (Habel et al., 2013; Wellstein et al., 2013). Grasslands also tend to violate the assumption of geometrically homogeneous canopies of one-dimensional RTMs (Berger et al., 2018; Rossi et al., 2020), complicating the selection of the most appropriate model for a given application.

In this paper, we aim to compare well-established empirical, physically-based, and hybrid models to estimate biomass across grassland sites with different management regimes, altitude and climate, and determine model performance by assessing their local accuracy, transferability, and epistemic uncertainty (= model uncertainty referring to the confidence of a model about its prediction) sensu Martínez-Ferrer et al. (2022). In doing so, we are focusing on an agnostic scenario, meaning that the models were not provided with any site-specific information beyond the spectral data. The rationale behind this scenario is the endeavor to develop accurate grassland biomass models using only widely available remote sensing data, as site-specific ancillary data often represent an operational bottleneck. Our goals are to develop adequate grassland biomass estimation models for each site and to identify key considerations for domain shift and generalization.

2. Methods

2.1. Study sites and data acquisition

We used field and remote sensing data from five study sites: one in Switzerland, three in Germany, and one in the United States differing in environmental characteristics, including altitude, climate, and management practices (Table 1, Fig. 1). These differences make our compiled dataset particularly valuable for assessing model transferability. The number of samples per study site ranged from 100 to 429, but we based all our models on 100 samples to keep them comparable. For sample selection we used Latin hypercube sampling (LHS) as implemented in the clhs package *v0.9.0* (Roudier, 2021) in R *v4.2.1* (R Core Team, 2021).

2.1.1. Switzerland

The Swiss study site (site code CH) encompasses the Lower Engadine and the Val Müstair in the Canton of Grisons in southeast Switzerland. Plots located in the Swiss National Park (SNP), an IUCN (International Union for the Conservation of Nature) category Ia nature reserve (highest protection level – strict nature reserve), are unmanaged. Plots in the Lower Engadine and the Val Müstair adjacent to the SNP are fertilized, mown, and grazed to varying degrees (Rossi et al., 2020).

At the CH site, biomass sampling for took place on the day of remote sensing data acquisition in late June to early July of 2010–2013 and 2016–2017, respectively. Biomass was clipped approximately 1 cm above the ground in 1 m^2 plots representative for a homogeneous area of 6 × 6 m, and dried at 65° for 48 h for 429 samples (Rossi et al., 2020; Schweiger et al., 2017; Schweiger et al., 2015a, 2015b). From 2010 to 2013 (before the launch of Sentinel-2), remote sensing data were acquired with the Airborne Prism Experiment (APEX) imaging spectrometer (Jehle et al., 2010; Schaepman et al., 2015). APEX data were resampled to 2 m pixel size using nearest neighbor interpolation and the parametric geocoding procedure PARGE (Schläpfer and Richter, 2002) and the airborne atmospheric and topographic correction model ATCOR-4 (Richter and Schläpfer, 2002) were used for geometric and atmospheric correction, respectively (Schweiger et al., 2015b). APEX data were resampled to Sentinel-2 bands (Appendix A Section A.1) using the prospectr R package *v0.2.6* (Stevens and Ramirez-Lopez, 2022) in R *v4.2.1*, and can be considered comparable with Sentinel-2 (Helfenstein et al., 2022). In 2016 and 2017, Sentinel-2 Level-1C (top of atmosphere) images were downloaded from the Copernicus Open Access Hub and

Table 1
Overview of the characteristics of the five study sites covering a wide range of topographic and climate, and management practices. Remote sensing data were acquired with the Airborne Prism Experiment (APEX) and Sentinel-2 between 2010 and 2020. The Köppen-Geiger climate classification is following Beck et al. (2018) for period 1991-2020. Dominant species are provided for Switzerland by the Swiss National Park's long-term permanent grassland monitoring project, for Germany by Bolliger et al. (2020), and for the United States by Gholizadeh et al. (2022). m.a.s.l.: meters above sea level, MAT: mean annual air temperature, MAP: mean annual precipitation.

Country	Switzerland		Germany			United States
Site code	CH		ALB	HAI	SCH	US
Site	Swiss National Park	Lower Engadine, Val Müstair	Schwäbische Alb	Hainich-Dün	Schorfheide-Chorin	The Nature Conservancy's Tallgrass Prairie Preserve (Pawhuska, OK)
Elevation [m.a.s.l.]	1400–2500		460–860	285–550	3–140	252 - 365
MAT [°C]	1 (at 2000 m.a.s.l.)		6–7	6.5–8	8–8.5	17.4
MAP [mm/a]	800		700–1000	500–800	500–600	960
Köppen-Geiger climate class	Cold, no dry season, cold summer (Dfc)		Cold, no dry season, warm summer (Dfb)	Mostly Cfb, small areas Dfb	Temperate, no dry season, warm summer (Cfb)	Temperate, no dry season, hot summer (Cfa)
Dominant species	*Erica carnea, Nardus stricta, Carex sempervirens, Festuca rubra*		*Alopecurus pratensis, Taraxacum* sp., *Festuca rubra* aggr., *Bromus erectus*	*Poa pratensis* aggr., *Taraxacum* sp., *Lolium perenne, Alopecurus pratensis*	*Poa pratensis* aggr., *Lolium perenne, Poa trivialis, Elymus repens*	*Schizachyrium scoparium, Andropogon gerardii, Sorghastrum nutans, Panicum virgatum*
Management	Strict protection	Grazing, mowing, fertilizing	Grazing, mowing, fertilizing	Grazing, mowing, fertilizing	Grazing, mowing, fertilizing	Grazing, burning
Sensor	Resampled APEX	Sentinel-2	Sentinel-2	Sentinel-2	Sentinel-2	Sentinel-2
Years	2010 - 2013	2016 - 2017	2017 - 2020	2017 - 2020	2017 - 2020	2022
Number of available samples	407	22	194	185	146	100
Number of selected samples	78	22	100	100	100	100

processed to Level-2 A (surface reflectance) using Sen2Cor *v2.3* (Müller-Wilm et al., 2013) and the SRTM 90 m digital elevation model (Reuter et al., 2007). The 10 m bands (B2, B3, B4, B8) were aggregated to a spatial resolution of 20 m using the arithmetic mean and bidirectional reflectance distribution function (BRDF) correction was applied following Poortinga et al. (2019). No cloud masking was required. For each plot, the spectral reflectance was sampled in the respective Sentinel-2 image by calculating the weighted mean on 20 m resolution around the plot center coordinate.

2.1.2. Germany

The three German study sites (site codes ALB, HAI, SCH) are part of the *Biodiversity Exploratories* (https://www.biodiversity-exploratories.de/en/). At each site, 50 grassland plots with different management regimes have been closely monitored since 2009 (Fischer et al., 2010; Hinderling et al., 2023; Ostrowski et al., 2020). Management intensity varies from extensive to moderately intensive, e.g., from no mowing or fertilization to three mowing events per year plus fertilization (Blüthgen et al., 2012; Fischer et al., 2010).

Biomass was harvested between late April and mid-July from 2017 to 2020 by clipping biomass approximately 4 cm above the ground on an area of 2 m^2 in plots representative of a homogeneous area of 50 × 50 m and subsequent drying at 80° for 48 h, resulting in 600 samples (Hinderling et al., 2023). We linearly scaled the dry biomass content to 1 m^2 for consistency with the other study sites. Sentinel-2 Level-2 A data closest to the day of biomass harvest were acquired through the Google Earth Engine (GEE, Gorelick et al., 2017) using the "Harmonized Sentinel-2 MSI: MultiSpectral Instrument, Level-2A" collection. The s2cloudless algorithm was used to mask out clouds and cloud shadows with the cloud probability threshold set to 10% (Zupanc, 2017). Again, the 10 m bands were aggregated to 20 m using the function *reduceResolution* in GEE and BRDF correction was applied. Sampling of the spectral reflectance for each plot follows the protocol for the CH site. An NDVI threshold was applied to prevent the inclusion of plots influenced by artifacts such as remaining cloud shadows or inhomogeneous vegetation cover (Appendix A Section A.2) with 525 samples remaining (Table 1).

2.1.3. United States

The study site in the United States (site code US) is located near Pawhuska, Oklahoma, and falls within The Nature Conservancy's Tallgrass Prairie Reserve (TGPP), encompassing an area of approximately 160 km^2 mostly covered by tallgrass prairie with some oak woodland (Hamilton, 2007; The Nature Conservancy, 2023). The TGPP is managed by cattle or bison grazing, and patch burning (Sherrill, 2019). This creates a *"shifting mosaic"* (Fuhlendorf and Engle, 2004) of patches with varying grazing pressure, as bison and cattle tend to primarily graze in recently burned areas with high nutrient availability (Anderson et al., 2006; Fuhlendorf and Engle, 2001).

Biomass sampling was conducted between July and August 2022 across 100 plots of 30 × 30 m, each containing nine 1 m^2 quadrats (Gholizadeh et al., 2024). Biomass was clipped in 80 of the 900 1 m^2 quadrats at approximately 2.54 cm (1 in.) above ground and dried at 65° for 144 h. In the remaining quadrats, biomass was determined using the digital obstruction method using the 80 samples for calibration (Limb et al., 2007). For each of the 100 plots, biomass was calculated as the mean biomass across the nine quadrats. Sentinel-2 Level-2A data for each plot were acquired following the same protocol as for the German study sites.

2.2. Model types

We assessed the accuracy and transferability of three different model types (i.e., empirical, physically-based, and hybrid; Fig. 2). Field data were partitioned into training (for the empirical models) or optimization (for the physically-based and hybrid models) and external testing sets

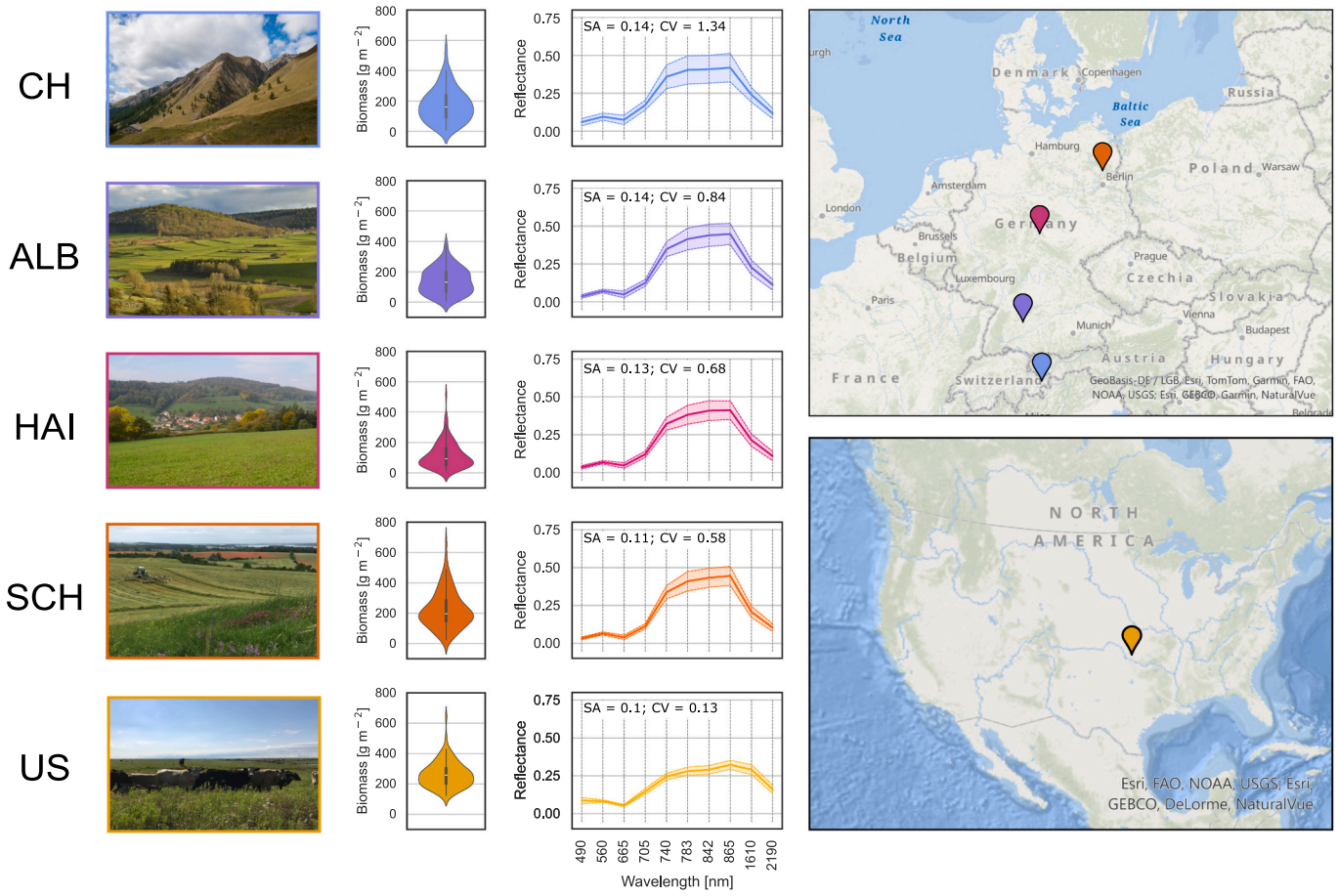

Fig. 1. Overview of the five study sites. Violin plots show dried aboveground biomass for the 100 selected samples per site. Line plots show the mean spectra (± 1 standard deviation) per site for Sentinel-2 or resampled APEX data. Mean spectral angle in radians (SA; Yuhas et al., 1992) of all unique sample combinations and coefficient of variation (CV) for all spectral bands serve as indicators of site-specific spectral variability. CH: Switzerland, image credits: Swiss National Park/Hans Lozza. ALB: Schwäbische Alb, image credits: Biodiversity Exploratories Information System (BExIS)/Martin Fellendorf. HAI: Hainich-Dün, image credits: BExIS/Steffen Both. SCH: Schorfheide-Chorin, image credits: BExIS/Ulrike Garbe. US: United States, image credits: Nicholas McMillan.

using an 80:20% split using LHS (Fig. 2A). We used scikit-learn *v1.5.2* (Pedregosa et al., 2012) and xgboost *v2.1.2* (Chen and Guestrin, 2016) in Python 3.13 for model training, optimization, and validation.

2.2.1. Empirical models

A wide variety of algorithms can be used to train empirical models (Fig. 2B). We used Random Forest regression (RFR), Support Vector regression (SVR), Extreme Gradient boosting regression (XGB), and Gaussian process regression (GPR) models to represent both tree-based and kernel-based methods. These algorithms all have been successfully used to estimate grassland biomass (Li et al., 2021; Muro et al., 2022; Raab et al., 2020; Schwieder et al., 2020; Verrelst et al., 2015; Verrelst et al., 2012). RFR uses an ensemble of decision trees, where each tree is trained on a random subset of samples and features and the final prediction is obtained by averaging the predictions of all trees (Breiman, 2001). SVR was originally introduced by Vapnik et al. (1997) and uses support vectors to fit hyperplanes in the data within a specified margin of tolerance. XGB is based on the concept of gradient boosting of regression trees introduced by Friedman (2001) and incorporates regularization to mitigate overfitting (Chen and Guestrin, 2016). Lastly, GPR uses a prior belief about the latent function describing the relationship between input and output variables, and training data to form the posterior distribution (Rasmussen and Williams, 2006). The advantages of using GPR include automatic hyperparameter optimization during model training and the provision of epistemic uncertainty (Verrelst et al., 2013a). To predict a data point, the model returns the mean of the posterior distribution as the estimated value and the predictive standard deviation (SD) as a measure of epistemic uncertainty.

A 5-fold cross-validation (CV) scheme with negative RMSE for scoring was used to identify the optimal model parametrization from all possible combinations of parameters listed in Table 2.

2.2.2. Physically-based model

Here we used the PROSAIL RTM (Jacquemoud et al., 2009), which combines the PROSPECT-D (Féret et al., 2017) and 4SAIL (Verhoef et al., 2007) RTMs, to simulate grassland canopy reflectance (Fig. 2C). PROSPECT-D simulates leaf level reflectance by considering leaf properties such as chlorophyll content (CHL), leaf mass per area (LMA) and the angle of incoming solar radiation (Féret et al., 2008). Subsequently, 4SAIL computes the bidirectional reflectance at the canopy level, employing canopy properties such as LAI and sun-target-sensor geometry. We used the prosail R package *v1.1.1* (Féret and de Boissieu, 2022) in R *v4.2.1* to create a single LUT for all study sites containing 10,000 simulated canopy reflectance spectra with input parameters selected by means of LHS within their respective value ranges (Table 3) derived from satellite image metadata, prior knowledge, and literature (He et al., 2019; Rossi et al., 2020; Verrelst et al., 2021). The value ranges and size of the LUT were consistent with other studies (Darvishzadeh et al., 2011; Hauser et al., 2021a; Locherer et al., 2015; Punalekar et al., 2018; Rivera et al., 2013; Rossi et al., 2020). We used the psoil parameter as weighting factor for the dry and wet soil spectra with psoil = 0 corresponding to completely wet soil conditions and psoil = 1 corresponding

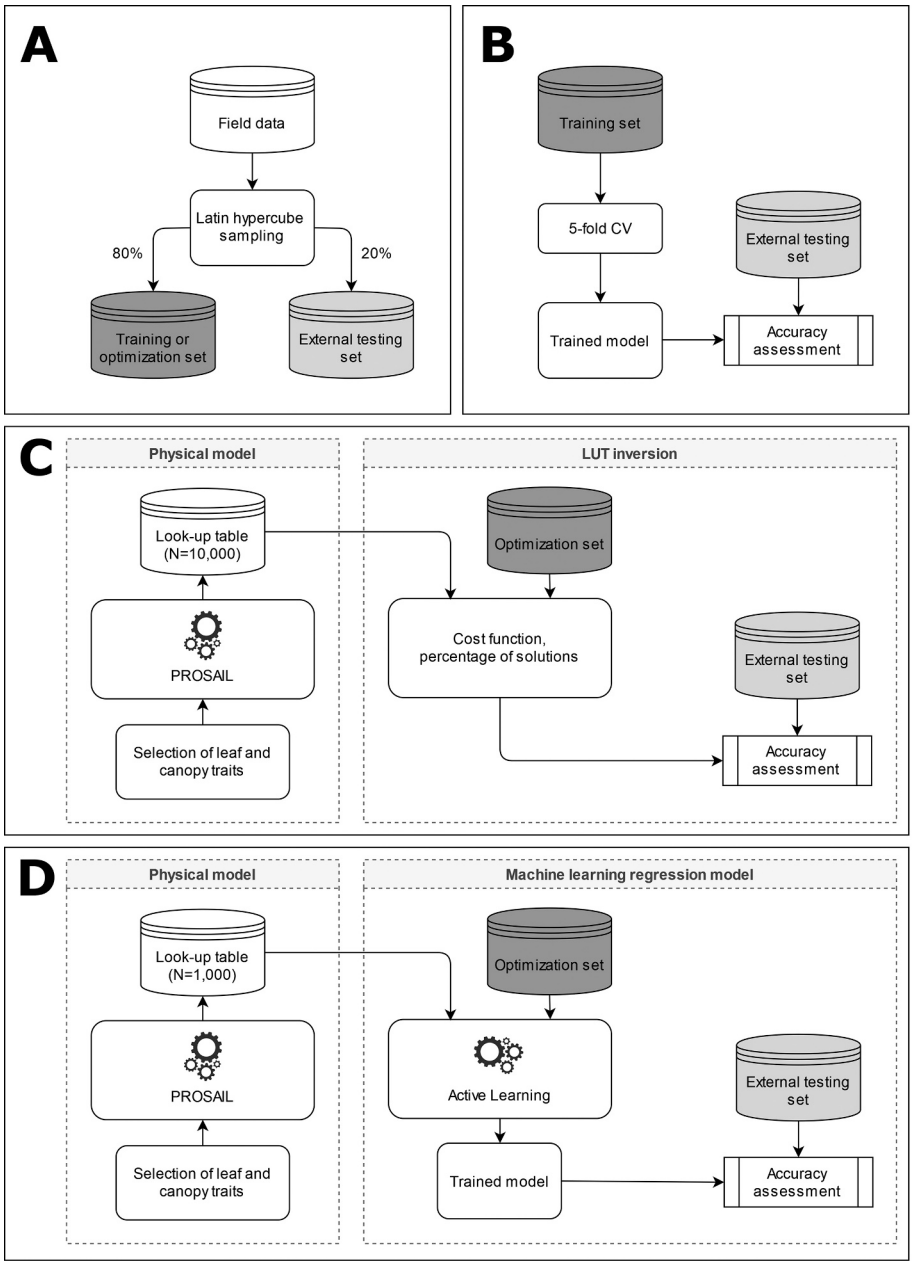

Fig. 2. Functioning of the models used in this study. A: As a prerequisite for all models, the field data need to be split into training (in the case of empirical models) or optimization set (in case of physically-based and hybrid models; colored in dark grey) and external testing set (colored in light grey) using Latin hypercube sampling. B: Empirical models are data-driven and learn data-specific relationships between predictor variables. The 5-fold cross-validation (CV) was performed with each possible combination of parameters listed in Table 2. C: Physically-based models use a radiative transfer model (RTM) such as PROSAIL to simulate canopy reflectance spectra and a cost function is used to find a predefined number of best matches between each field data point and the simulated spectra, a process commonly referred to as look-up table (LUT) inversion. D: Hybrid models are trained with a set of RTM-simulated spectra, which are optionally optimized using Active Learning. Subsequently, a machine learning regression model is used to perform the LUT inversion.

to completely dry soil conditions, respectively. We used the default dry and wet soil spectra of the prosail R package and did not use site-specific soil spectra, as no corresponding reference data were available and their use would have contradicted the agnostic modeling scenario. The simulated reflectance spectra were resampled to Sentinel-2 bands using the prospectr R package and their corresponding biomass content was calculated following (Quan et al., 2017, Eq. 1).

$$biomass\ \left[g\,m^{-2}\right] = LMA\ \left[g\,cm^{-2}\right] * LAI * 10,000 \quad (1)$$

Two parameters must be determined for the inversion of the generated LUT: the cost function and the number of spectra with the lowest cost to consider (hereafter referred to as percentage of solutions). We tested 17 commonly used cost functions listed by Rivera et al. (2013) and different percentages of solutions, namely 0.01% (= 1 solution), 1%, 2%, 5%, and 10% (Punalekar et al., 2018; Rivera et al., 2013; Rossi et al., 2020) to optimize parameter choice. The predicted biomass value was calculated as the mean value of the selected solutions, with the corresponding SD being indicative of epistemic uncertainty (Locherer et al., 2015; Rivera et al., 2013).

2.2.3. Hybrid model

For the hybrid model, we created a LUT containing 1000 simulations (Tagliabue et al., 2022; Verrelst et al., 2021) as described above. But in this case, the simulated canopy reflectance spectra were used to train a

Table 2
Parameter values used for cross-validation of empirical Random Forest regression (RF), Support Vector regression (SVR), Extreme Gradient Boosting regression (XGB), and Gaussian process regression (GPR) models. Nomenclature of parameter names for RFR, SVR, and GPR according to Pedregosa et al. (2012), for XGB according to Chen and Guestrin (2016).

Model	Parameter function	Parameter name	Values
RF	Number of trees	n_estimators	100, 200, 500
	Maximum tree depth	max_depth	None, 5, 10, 15
	Minimum number of samples required to be at a leaf node	min_samples_leaf	1, 2, 5
	Number of features	max_features	'sqrt', 'log2', 10
	Maximum number of leaf nodes	max_leaf_nodes	10, 20, None
SVR	Kernel type	kernel	'rbf','linear'
	Kernel coefficient	gamma	'scale', 'auto', 0.01, 0.1, 1
	Regularization parameter	C	0.1, 1, 10, 100
	Epsilon-tube	epsilon	0.01, 0.1, 0.5, 1
XGB	Number of gradient boosted trees	n_estimators	100, 200, 300
	Maximum tree depth	max_depth	3, 5, 7, 9
	Boosting learning rate	learning_rate	0.01, 0.1, 0.2
	Subsample ratio of training instance	subsample	0.8, 1
	Subsample ratio of columns when constructing each tree	subsample_bytree	0.8, 1
	Minimum loss reduction	gamma	0, 0.1, 0.2
GPR	Kernel	kernel	ConstantKernel() * RBF()
	Length scale bounds of RBF kernel	length_scale_bounds	(−100,100)

regression algorithm for RTM inversion (Fig. 2D). Here we used a GPR model (see Section 2.2.1 for further details) with the parameters indicated in Table 2 for the regression task. We used AL to avoid biophysically unrealistic variable combinations and redundant information, and select the most informative simulations from the LUT (Verrelst et al., 2016). Using the AL-selected subset of simulations for GPR training can help to mitigate the ill-posedness inherent to RTM inversion as different variable combinations can lead to similar spectra (Combal et al., 2003), and increase computational efficiency (Berger et al., 2021b). AL selects a predefined percentage of simulations from the LUT as an initial training dataset. Subsequently, the GPR model is trained using this initial training set, and its predictive training accuracy is assessed via the root-mean-square error (RMSE) computed with the optimization set. By employing a selection heuristic, such as Euclidean distance-based diversity (EBD), a simulation from the remaining LUT is selected, temporarily added to the training set, and only permanently kept if the updated training set leads to an improved RMSE. This optimization process continues until all simulations are evaluated. Finally, the validation accuracy of the GPR model trained with the optimal training set is determined.

In line with previous studies, we used 2% of the data as initial training data (Tagliabue et al., 2022; Wocher et al., 2022). Hybrid models without the use of AL did not lead to meaningful results in the context of this study (Appendix A Section A.3) and were therefore not further analyzed.

2.3. Model comparisons

We conducted three model comparisons to investigate model accuracy and assess model transferability (Fig. 3). First, models were trained (empirical) or optimized (physically-based and hybrid) and validated individually for each study site (hereafter referred to as local models, Fig. 3A). This setting corresponds to that of most local to regional scale studies in which field data of a specific area of interest are available. Second, the local models were applied to the field data of the other study sites to assess their transferability (hereafter referred to as transferred models, Fig. 3B), simulating the case where a model trained or optimized for one area is applied to another area where no field data are available (domain shift). For the physically-based models, this means that the combination of cost function and percentage of solutions identified during local model optimization is transferred to the other study sites (i.e., the LUT remains unchanged). Third, models were trained or optimized with field data from four study sites and validated using the field data of the remaining site to examine any improvement in transferability (hereafter referred to as global models, Fig. 3C). This mimics the case in which a diverse set of field data are available, for example from a compiled database, and used to make predictions for an area not covered by the database. Model transferability in such a setting is also referred to as the model's ability for domain generalization (Zhou et al., 2022). In addition, we compared epistemic uncertainties with model accuracies for all model comparisons where available – namely empirical GPR, physically-based, and hybrid models.

Model performance was assessed by coefficient of determination (R^2), relative root-mean-square errors (RRMSE) calculated as the RMSE divided by the mean value of the external testing set (Richter et al., 2012), and mean bias error (MBE) using the external testing set. In all cases, we used the ten Sentinel-2 bands in the visible and near-infrared (B2 – B8A) and the short-wave infrared regions (B11 – B12) as predictor variables with reflectance values being standardized for empirical and hybrid models and normalized, i.e., treating the spectra as probability distributions summing up to 1, for physically-based models (Rivera et al., 2013), respectively. Accordingly, the models should be considered agnostic, meaning that apart from Sentinel-2 spectral data, no site-

Table 3
Value ranges and distributions of PROSAIL input parameters used in this study.

Parameter	Variable	Unit	Minimum value	Maximum value	Distribution
Leaf structure parameter	N	[-]	1.5	1.9	uniform
Chlorophyll content	CHL	[µg cm^{-2}]	5	75	uniform
Carotenoid content	CAR	[µg cm^{-2}]	2	60	uniform
Anthocyanin content	ANT	[µg cm^{-2}]	0	2	uniform
Brown pigment content	BROWN	[-]	0	1	uniform
Equivalent water thickness	EWT	[cm]	0.001	0.04	uniform
Leaf mass per area	LMA	[g cm^{-2}]	0.002	0.015	uniform
Angle for incident light at leaf surface	alpha	[°]	40	40	fixed
Leaf inclination distribution function	TypeLidf	[-]	2	2	fixed
Average leaf angle	LIDFa	[°]	40	70	uniform
Leaf area index	LAI	[-]	0.1	4	uniform
Hot spot parameter	q	[-]	0.01	0.1	uniform
Sun zenith angle	tts	[°]	25	75	uniform
Observer zenith angle	tto	[°]	0	0	fixed
Relative azimuth angle	psi	[°]	50	180	uniform
Dry/wet soil factor	psoil	[-]	0	1	uniform

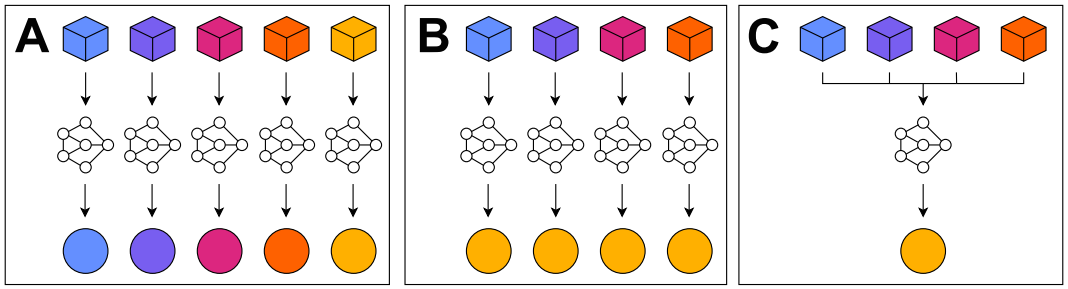

Fig. 3. The three model comparisons conducted in this study. Cubes represent data used for training (of empirical models) or optimization (of physically-based and hybrid models). Circles represent data used for model validation. Colors represent data from different study sites. A: Local models with training/optimization and validation data from the same study site. B: Domain shift: transferred models with training/optimization and validation data from different study site (only one of five cases shown). C: Domain generalization: global models with training/optimization data from four sites and validation data from the remaining site (only one of five cases shown).

specific data were used for model training or prediction.

We note that due to comparatively small sample size ($N = 100$), partitioning of the data into training or calibration and external testing set was repeated 10 times for the local and transferred models to account for stochastic effects (Muro et al., 2022). Correspondingly, local and transferred model performance of all model types were assessed by calculating both mean and SD for R^2, RRMSE, and MBE.

For the empirical models, the best-performing method was selected to be presented in the results as the focus lied on the comparison of model types and not different empirical models. Comprehensive cross-validation and testing performances for all models including additional performance metrics are enclosed in Appendix A Sections A.4 to A.16.

Lastly, the epistemic uncertainty of the local, transferred, and global models was compared to the absolute difference between measured and predicted biomass of the external testing set to test if epistemic uncertainty could potentially be used as an indicator of model transferability for heterogeneous grasslands. Further information about said relation for all models, including the individual repetitions of local and transferred models, is included in Appendix A Section A.17. We tested the relationship between absolute difference and epistemic uncertainty since it has been demonstrated that epistemic uncertainty of GPR models can serve as a *quality indicator* to identify reliable and unreliable predictions of transferred models for croplands, even when applied across spatial scales (Verrelst et al., 2013b; Verrelst et al., 2013a; Verrelst et al., 2012).

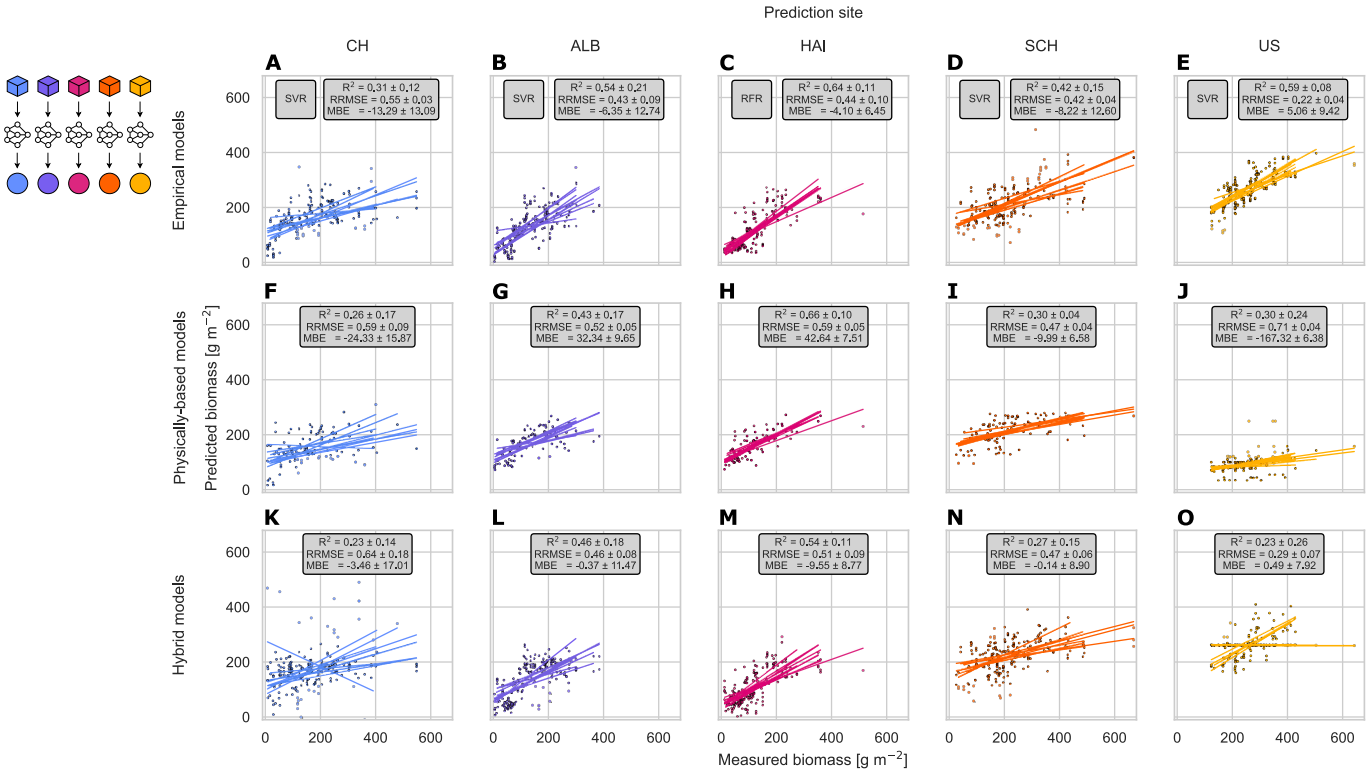

Fig. 4. Scatterplots of measured versus predicted biomass for the local empirical (A-E), physically-based (F-J), and hybrid models (L-O). Textboxes show mean coefficient of determination (R^2), relative root-mean-square error (RRMSE) and mean bias error (MBE) ± 1 standard deviation across 10 repetitions for each model type. For empirical models, only the best-performing model in terms of lowest RRMSE is shown with the corresponding model name added in a separate textbox. CH: Switzerland (A, F, K), ALB: Schwäbische Alb (B, G, L), HAI: Hainich-Dün (C, H, M), SCH: Schorfheide-Chorin (D, I, N), US: United States (E, J, O), RFR: Random Forest regression, SVR: Support Vector regression.

3. Results

3.1. Local models

The accuracy of all models varied with the study sites (Fig. 4). For empirical models (Fig. 4A-E), mean R^2 ranged from 0.31 to 0.64 (CH and HAI, respectively) and mean RRMSE from 0.22 to 0.55 (US and CH, respectively). Mean MBE indicated a systematic underestimation for all sites between -4.1 and -13.29 g/m^2 (HAI and CH, respectively) except for the US, for which a mean overestimation of 5.06 g/m^2 was observed. SVR outperformed the other empirical models for most sites, only the HAI site was best predicted by RFR.

For physically-based models (Fig. 4F-J), mean R^2 ranged from 0.26 to 0.66 (CH and HAI, respectively) and mean RRMSE from 0.47 to 0.71 (SCH and US, respectively). For the ALB and HAI sites, a mean MBE of up to 42.64 g/m^2 was observed while on average, biomass was underestimated for CH and SCH. For the US model, a severe underestimation of -167.32 g/m^2 was reported. Compared to empirical models, model accuracy slightly decreased for most sites; for the US site it decreased substantially.

For hybrid models (Fig. 4K-O), mean R^2 ranged from 0.23 to 0.54 (US and HAI, respectively) while mean RRMSE ranged from 0.29 to 0.64 (US and CH, respectively). Except for the US site, a slight underestimation in terms of MBE could be observed. Model accuracy across sites resembled those of empirical models, although they were again slightly lower.

3.2. Transferred models

In general, model accuracy decreased when local models were applied to other sites, although physical models sustained their predictive power the best (Table 4).

Out of all transferred empirical models, the accuracy of the CH models transferred to the SCH site (mean R^2 = 0.33, mean RRMSE = 0.49) and models transferred among the three German sites came closest to that of the local models, e.g., the ALB models predicting the HAI site (mean R^2 = 0.49, mean RRMSE = 0.54) or the HAI and SCH models predicting the ALB site (mean R^2 = 0.30, mean RRMSE = 0.53 and mean R^2 = 0.49, mean RRMSE = 0.64, respectively). For the CH site, only a mean R^2 of 0.15 with an associated mean RRMSE of 0.65 could be achieved by the transferred empirical model trained at the SCH site. For the US site, all transferred empirical models exhibited a systematic underestimation of the present biomass while the US models themselves overestimated biomass at other sites, e.g., with a mean MBE of 146.09 g/m^2 when predicting the HAI site.

Regarding physically-based models, transferability diverged less strongly between combinations of optimization and prediction sites, with the performance being best for the ALB, HAI, and SCH sites. Overall, variability between the 10 repetitions was comparatively low, as identical combinations of cost function and percentage of solutions were selected for the German sites (Appendix A Section A.5). Best results were achieved for the ALB site with the HAI models having performed similarly to the local models (mean R^2 = 0.43, mean RRMSE = 0.51). A slight decline in performance could be observed for the HAI site, although the ALB model still achieved a mean R^2 of 0.53 and a mean RRMSE of 0.66, with the decline for the SCH site being more pronounced. For the CH site, the models showed a lower mean R^2, but only a slightly higher mean RRMSE. For the US site, in contrast, a direct comparison with the local models was difficult; generally, a higher mean R^2 was achieved, but mean RRMSE and systematic underestimation also increased. Moreover, the US models consistently performed worse for all other prediction sites. In comparison with transferred empirical models, the transferability of physically-based models was higher for the CH, ALB, and HAI sites.

The transferability of hybrid models also varied among different combinations of optimization and validation sites with best results for combinations of the German study sites such as the ALB models for the HAI site (mean R^2 = 0.32, mean RRMSE = 0.63) and vice versa (mean R^2 = 0.34, mean RRMSE = 0.52). The ALB models for the SCH site showed comparatively good values for mean R^2 and RRMSE, but with increased systematic underestimation. For the CH and US sites, no satisfactory performance could be achieved. Overall, the transferred hybrid models exhibited similar patterns to the empirical models, although their performance was somewhat lower. Particularly notable was the comparatively good performance among the German study sites and the systematic underestimation of biomass for the US site.

3.3. Global models

For empirical models, accuracy was generally low (Fig. 5A-E). It was highest for the ALB site, followed by the HAI and SCH sites, but accompanied by a high MBE. For the CH site, the R^2 of the prediction was practically 0.00, while the systematic underestimation for the US site amounted to -195.82 g/m^2. Comparatively poor results were also produced by the physically-based models in most cases (Fig. 5F-J). An exception was the US site, for which an R^2 of 0.45 and an RRMSE of 0.75 were achieved. However, the systematic underestimation also tended to be high with an MBE of -183.99 g/m^2. Regarding hybrid models (Fig. 5K-O), best results were obtained for the models predicting the ALB, HAI, and SCH sites, although the MBE indicated substantial over- and underestimations ranging from 37.30 for the ALB model to -79.07 g/m^2 for the SCH model, respectively. No satisfactory performance was achieved for the CH and US sites. Overall, the accuracy of global models was lower than for the local models (with the exception of the physically-based model for the US site, Fig. 5K).

3.4. Epistemic uncertainty

For none of the comparisons (local, transferred, and global), a systematic relationship between epistemic uncertainty and absolute differences between measured and predicted biomass values was observed when considering all predicted data points (10 repetitions of 20 data points for local and transferred models, 100 data points for global models; Tables 5 and 6). Only isolated weak correlations were found, e.g., for local and transferred physically-based models predicting the US site (R^2 up to 0.31) or the global hybrid model predicting the US site (R^2 = 0.38).

Only for the mean values of epistemic uncertainty and absolute difference, a few patterns could be identified. Regarding the local empirical and hybrid models, a positive correlation between mean epistemic uncertainty and mean absolute difference could be observed. The lowest correlation values were found for the HAI site (for both model types) and the highest values for the SCH and CH sites (for empirical and hybrid models, respectively). The physically-based models displayed a negative correlation; the US site exhibited the lowest mean epistemic uncertainty but the highest mean absolute difference (52.71 and 162.16 g/m^2, respectively).

For the transferred models, a clear pattern between combinations of training or optimization and prediction site was found. The respective empirical and hybrid models of the ALB and HAI sites exhibited low values for both mean epistemic uncertainty and absolute difference for mutual prediction. The highest values for mean epistemic uncertainty and absolute difference were produced by the CH and US models of both model types. For the transferred physically-based models, the lowest values for mean epistemic uncertainty were observed for the US site (38.86 g/m^2 for ALB and HAI models), which did not coincide with the high values for mean absolute difference (182.09 g/m^2 for ALB and HAI models). For the global models, no agreement between mean epistemic uncertainty and mean absolute difference was found for any model type or prediction site.

Table 4
Mean coefficient of determination (R^2), relative root-mean-square error (RRMSE), and mean bias error (MBE) for ±1 standard deviation across 10 repetitions for the transferred empirical, physically-based, and hybrid models. To facilitate an estimation of model transferability, the metrics of the local models were included (in italic). For empirical models, only the best-performing model in terms of lowest RRMSE is shown with the corresponding model name added in brackets. CH: Switzerland, ALB: Schwäbische Alb, HAI: Hainich-Dün, SCH: Schorfheide-Chorin, US: United States, RFR: Random Forest regression, SVR: Support Vector regression, XGB: Extreme Gradient Boosting regression, GPR: Gaussian Process regression.

		Prediction site														
		CH			ALB			HAI			SCH			US		
Model type	Training/ optimization site	R^2	RRMSE	MBE	**R^2**	RRMSE	MBE	R^2	RRMSE	MBE	R^2	RRMSE	MBE	R^2	RRMSE	MBE
Empirical	CH	*0.31 ± 0.12 (SVR)*	*0.55 ± 0.03 (SVR)*	*-13.29 ± 13.09 (SVR)*	0.46 ± 0.12 (SVR)	0.70 ± 0.12 (SVR)	67.46 ± 23.94 (SVR)	0.39 ± 0.04 (SVR)	0.92 ± 0.15 (SVR)	77.08 ± 20.47 (SVR)	0.33 ± 0.08 (SVR)	0.49 ± 0.05 (SVR)	-14.42 ± 23.94 (SVR)	0.08 ± 0.05 (GPR)	0.48 ± 0.00 (GPR)	-88.80 ± 0.88 (GPR)
	ALB	0.04 ± 0.02 (GPR)	0.69 ± 0.00 (GPR)	-37.63 ± 0.98 (GPR)	*0.54 ± 0.21 (SVR)*	*0.43 ± 0.09 (SVR)*	*-6.35 ± 12.74 (SVR)*	0.49 ± 0.05 (SVR)	0.54 ± 0.02 (SVR)	-10.15 ± 5.00 (SVR)	0.30 ± 0.03 (XGB)	0.55 ± 0.01 (XGB)	-68.78 ± 3.22 (XGB)	0.04 ± 0.02 (GPR)	0.57 ± 0.00 (GPR)	-123.02 ± 0.86 (GPR)
	HAI	0.00 ± 0.01 (GPR)	0.73 ± 0.00 (GPR)	-62.02 ± 0.58 (GPR)	0.30 ± 0.03 (GPR)	0.53 ± 0.02 (GPR)	-13.74 ± 3.73 (GPR)	*0.64 ± 0.11 (RFR)*	*0.44 ± 0.10 (RFR)*	*-4.10 ± 6.45 (RFR)*	0.16 ± 0.03 (XGB)	0.63 ± 0.02 (XGB)	-83.11 ± 5.78 (XGB)	0.16 ± 0.04 (GPR)	0.65 ± 0.00 (GPR)	-146.93 ± 0.65 (GPR)
	SCH	0.15 ± 0.02 (RFR)	0.65 ± 0.03 (RFR)	22.32 ± 9.96 (RFR)	0.49 ± 0.04 (SVR)	0.64 ± 0.04 (SVR)	64.65 ± 7.68 (SVR)	0.48 ± 0.05 (SVR)	0.76 ± 0.05 (SVR)	62.53 ± 6.07 (SVR)	*0.42 ± 0.15 (SVR)*	*0.42 ± 0.04 (SVR)*	*-8.22 ± 12.60 (SVR)*	0.09 ± 0.03 (GPR)	0.36 ± 0.00 (GPR)	-39.81 ± 1.23 (GPR)
	US	0.01 ± 0.00 (GPR)	0.81 ± 0.00 (GPR)	87.98 ± 1.09 (GPR)	0.03 ± 0.01 (GPR)	1.07 ± 0.01 (GPR)	122.53 ± 1.11 (GPR)	0.02 ± 0.00 (GPR)	1.46 ± 0.01 (GPR)	146.09 ± 1.31 (GPR)	0.01 ± 0.00 (GPR)	0.57 ± 0.00 (GPR)	39.83 ± 1.14 (GPR)	*0.59 ± 0.08 (SVR)*	*0.22 ± 0.04 (SVR)*	*5.06 ± 9.42 (SVR)*
Physically-based	CH	*0.26 ± 0.17*	*0.59 ± 0.09*	*-24.33 ± 15.87*	0.41 ± 0.00	0.53 ± 0.00	34.97 ± 0.37	0.53 ± 0.00	0.69 ± 0.00	51.81 ± 0.23	0.21 ± 0.00	0.51 ± 0.00	-40.07 ± 0.25	0.45 ± 0.00	0.75 ± 0.01	-182.43 ± 1.66
	ALB	0.14 ± 0.00	0.61 ± 0.00	-24.61 ± 0.00	*0.43 ± 0.17*	*0.52 ± 0.05*	*32.34 ± 9.65*	0.53 ± 0.00	0.66 ± 0.00	45.83 ± 0.00	0.20 ± 0.00	0.53 ± 0.00	-45.64 ± 0.00	0.44 ± 0.00	0.75 ± 0.00	-182.09 ± 0.00
	HAI	0.14 ± 0.00	0.61 ± 0.00	-24.61 ± 0.00	0.43 ± 0.00	0.51 ± 0.00	29.33 ± 0.00	*0.66 ± 0.10*	*0.59 ± 0.05*	*42.64 ± 7.51*	0.20 ± 0.00	0.53 ± 0.00	-45.64 ± 0.00	0.44 ± 0.00	0.75 ± 0.00	-182.09 ± 0.00
	SCH	0.12 ± 0.00	0.60 ± 0.00	-3.44 ± 0.00	0.45 ± 0.00	0.63 ± 0.00	60.49 ± 0.00	0.50 ± 0.00	0.85 ± 0.00	77.44 ± 0.00	*0.30 ± 0.04*	*0.47 ± 0.04*	*-9.99 ± 6.58*	0.43 ± 0.00	0.71 ± 0.00	-170.48 ± 0.00
	US	0.13 ± 0.00	0.64 ± 0.04	-18.58 ± 7.28	0.22 ± 0.21	0.68 ± 0.07	19.05 ± 37.63	0.26 ± 0.23	0.84 ± 0.02	41.44 ± 31.11	0.12 ± 0.12	0.60 ± 0.13	-45.40 ± 31.62	*0.30 ± 0.24*	*0.71 ± 0.04*	*-167.32 ± 6.38*
Hybrid	CH	*0.23 ± 0.14*	*0.64 ± 0.18*	*-3.46 ± 17.01*	0.07 ± 0.07	0.88 ± 0.06	55.91 ± 19.27	0.03 ± 0.03	1.13 ± 0.11	70.71 ± 24.06	0.02 ± 0.02	0.64 ± 0.06	-44.35 ± 27.74	0.13 ± 0.10	0.70 ± 0.10	-148.09 ± 28.52
	ALB	0.02 ± 0.02	0.73 ± 0.05	-31.66 ± 21.59	*0.46 ± 0.18*	*0.46 ± 0.08*	*-0.37 ± 11.47*	0.32 ± 0.04	0.63 ± 0.01	0.42 ± 12.18	0.23 ± 0.04	0.58 ± 0.04	-75.77 ± 12.38	0.12 ± 0.10	0.62 ± 0.15	-132.65 ± 46.57
	HAI	0.02 ± 0.03	0.72 ± 0.06	-36.42 ± 18.12	0.34 ± 0.07	0.52 ± 0.03	-19.74 ± 5.34	*0.54 ± 0.11*	*0.51 ± 0.09*	*-9.55 ± 8.77*	0.17 ± 0.07	0.65 ± 0.03	-94.74 ± 5.45	0.14 ± 0.11	0.59 ± 0.04	-120.22 ± 12.54
	SCH	0.03 ± 0.03	0.79 ± 0.07	36.96 ± 48.54	0.41 ± 0.09	0.77 ± 0.06	83.55 ± 6.82	0.25 ± 0.06	0.95 ± 0.04	78.01 ± 6.21	*0.27 ± 0.15*	*0.47 ± 0.06*	*-0.14 ± 8.90*	0.19 ± 0.05	0.45 ± 0.09	-48.41 ± 44.96
	US	0.01 ± 0.01	0.94 ± 0.20	93.61 ± 20.14	0.03 ± 0.02	1.14 ± 0.20	117.33 ± 42.37	0.02 ± 0.03	1.55 ± 0.27	146.89 ± 38.34	0.02 ± 0.03	0.65 ± 0.12	39.24 ± 43.46	*0.23 ± 0.26*	*0.29 ± 0.07*	*0.49 ± 7.92*

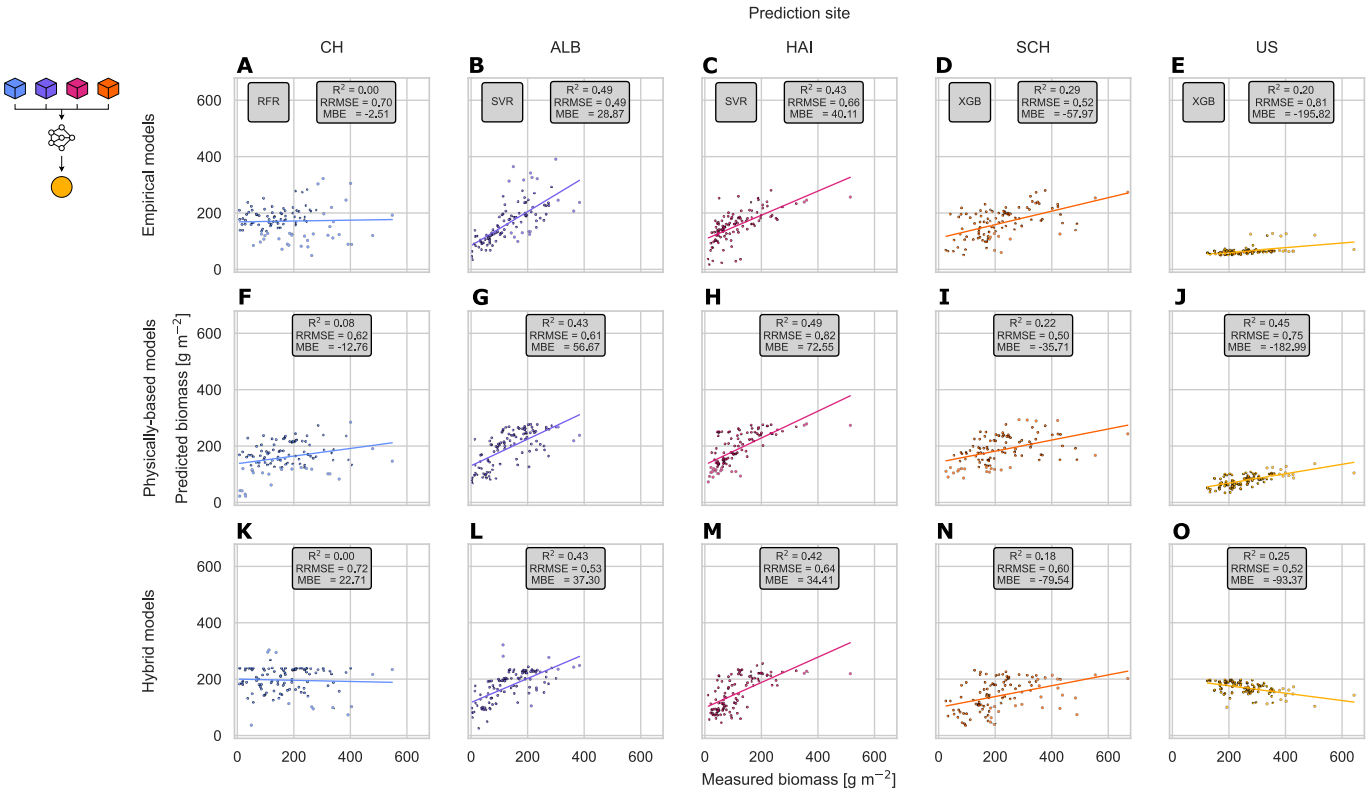

Fig. 5. Scatterplots of measured versus predicted biomass for the global empirical (A-E), physically-based (F-J), and hybrid models (L-O). Textboxes show coefficient of determination (R^2), relative root-mean-square error (RRMSE) and mean bias error (MBE) for each model. For empirical models, only the best-performing model in terms of lowest RRMSE is shown with the corresponding model name added in a separate textbox. CH: Switzerland (A, F, K), ALB: Schwäbische Alb (B, G, L), HAI: Hainich-Dün (C, H, M), SCH: Schorfheide-Chorin (D, I, N), US: United States (E, J, O), RFR: Random Forest regression, SVR: Support Vector regression, XGB: Extreme Gradient Boosting regression.

4. Discussion

Remote sensing has gained traction in its application for monitoring ecosystem functioning, with spatial (and temporal) scalability often highlighted as a key advantage. When developing scalable models, model transferability should be an integral part of performance assessment; however, systematic evaluation often lags behind the ambition for scalability within and across ecosystems. In this study, we demonstrated the importance of assessing model transferability when evaluating the performance of remotely sensed grassland biomass estimates and related model selection considerations. Our model comparisons indicated that in most cases physically-based models exhibited the highest transferability when applied to unseen grassland sites. However, no single model consistently outperformed others when trained or optimized with data from multiple sites. These results underscore the challenges in developing scalable models, highlight the importance and possible trade-offs of appropriate model selection, and shed light on the discrepancies between epistemic uncertainty and predictive accuracy.

4.1. Accuracy of local models

All model types performed similarly when predicting biomass locally with accuracies comparable to those of other studies conducted in alpine and semi-natural grasslands in Switzerland and Germany (Raab et al., 2020; Schweiger et al., 2015a). Generally, model accuracy decreased with increasing biomass range and (to a lesser extent) spectral variability (see also Dehghan-Shoar et al., 2023, Fig. 1, and Appendix A Section A.18). Overall, empirical models performed best at the site level, presumably due to their high flexibility, allowing them to incorporate site-specific relationships in the data without necessarily relying on physical principles.

For the physical models, we found a large discrepancy between the in situ and modeled data for the US site caused by the employed LUT parametrization. Although the models recognized the underlying relationship between biomass and spectral information to a certain extent (mean $R^2 = 0.30$), a high RRMSE of 0.71 resulted probably due to a too narrow range of values for LMA in the LUT since testing an alternative value range with a higher upper bound for LMA resulted in a mean R^2 of 0.51 and mean RRMSE of 0.33 (Appendix A Section A.19). An evaluation of the agreement between the PROSAIL simulations and Sentinel-2 spectra of the different study sites as an indicator of the respective PROSAIL forward mode fit quality can be found in Appendix A Section A.20. However, broadening the trait ranges increased ill-posedness (Combal et al., 2003; Verrelst et al., 2014) and deteriorated model performance for the other sites; therefore, a site-specific parametrization of multiple narrowly defined LUTs would be necessary to obtain comparable results to the empirical models for all sites.

The hybrid models performed similarly compared to the empirical models. Selecting training samples with AL led to substantial improvements in performance as stated in AL theory (Verrelst et al., 2016). In our case, the use of AL for selecting the most informative optimization samples was essential because we did not parameterize the LUT a priori or excluded biophysically unrealistic simulations as done by others (see, e.g., Campos-Taberner et al., 2018). Moreover, the hybrid models did not display the aforementioned imbalance between R^2 and RRMSE for the US site although having been trained on a suboptimally configured LUT, presumably because the employed GPR kernel was able to scale the predictions by the magnitude of the employed constant kernel resulting in a lower RRMSE compared to the physically-based models while keeping R^2 comparatively high (scikit-learn Developers, 2023).

The fact that local models in grasslands sometimes underperform compared to croplands is largely due to the specific particularities of

Table 5
Mean difference between predicted and measured grassland biomass and epistemic uncertainty ±1 standard deviation and coefficient of determination (R^2) for the 10 local (in italic) and local transferred models. For the local models, 200 samples (10 predictions of 20 samples each) and for the transferred models, 1000 samples (10 predictions of 100 samples each) were available. If the percentage of solutions of a given physically-based model was 0.01%, only 1 solution was used to derive the predicted biomass value and no epistemic uncertainty could be calculated for the corresponding predicted samples (see Appendix A Section A.5). The absolute difference was calculated by taking the absolute value after subtracting the measured biomass value from the predicted biomass value for each predicted sample. For the empirical and hybrid models, the epistemic uncertainty for a predicted sample corresponds to the predicted standard deviation of the Gaussian process regression (GPR) models. For the physically-based models, the epistemic uncertainty corresponds to the standard deviation of all selected solutions. CH: Switzerland, ALB: Schwäbische Alb, HAI: Hainich-Dün, SCH: Schorfheide-Chorin, US: United States.

		Prediction site														
		CH			ALB			HAI			SCH			US		
Model type	Training/ optimization site	Absolute difference [g m^{-2}]	Epistemic uncertainty [g m^{-2}]	R^2	Absolute difference [g m^{-2}]	Epistemic uncertainty [g m^{-2}]	R^2	Absolute difference [g m^{-2}]	Epistemic uncertainty [g m^{-2}]	R^2	Absolute difference [g m^{-2}]	Epistemic uncertainty [g m^{-2}]	R^2	Absolute difference [g m^{-2}]	Epistemic uncertainty [g m^{-2}]	R^2
Empirical	CH	*83.69 ± 74.12*	*94.67 ± 21.25*	*0.08*	98.15 ± 71.22	108.78 ± 11.78	0.18	119.89 ± 72.50	109.00 ± 10.10	0.19	101.10 ± 82.35	111.90 ± 8.79	0.02	94.41 ± 81.49	116.21 ± 3.83	0.03
	ALB	90.79 ± 77.02	82.00 ± 3.57	0.10	*55.08 ± 41.92*	*64.64 ± 14.89*	*0.01*	53.80 ± 48.66	70.18 ± 12.51	0.05	96.30 ± 90.96	69.07 ± 11.85	0.00	123.80 ± 84.85	82.86 ± 1.06	0.00
	HAI	96.51 ± 83.30	87.96 ± 7.87	0.03	58.07 ± 44.78	65.96 ± 16.17	0.02	*41.89 ± 43.50*	*57.10 ± 17.38*	*0.03*	102.05 ± 95.57	65.59 ± 15.87	0.07	146.93 ± 86.94	89.98 ± 6.35	0.00
	SCH	100.43 ± 66.40	123.63 ± 5.15	0.00	100.29 ± 62.11	111.75 ± 14.43	0.05	105.34 ± 57.31	111.95 ± 13.54	0.13	*86.17 ± 76.45*	*110.94 ± 14.80*	*0.02*	69.60 ± 64.11	123.97 ± 5.05	0.00
	US	120.43 ± 73.51	96.99 ± 7.78	0.00	129.78 ± 73.50	96.99 ± 7.78	0.00	156.39 ± 65.59	96.98 ± 7.78	0.00	104.73 ± 69.81	96.99 ± 7.78	0.00	*54.01 ± 48.06*	*66.35 ± 19.96*	*0.01*
Physically-based	CH	*72.56 ± 74.38*	*72.64 ± 28.22*	*0.02*	64.22 ± 37.31	90.64 ± 25.47	0.00	68.23 ± 41.77	87.73 ± 24.39	0.01	80.90 ± 80.60	95.65 ± 21.07	0.04	182.43 ± 73.30	40.22 ± 11.08	0.20
	ALB	76.45 ± 72.35	71.98 ± 24.64	0.01	*63.10 ± 37.78*	*88.68 ± 25.09*	*0.00*	64.28 ± 41.40	84.10 ± 24.24	0.02	81.28 ± 83.70	92.55 ± 20.37	0.03	182.09 ± 73.19	38.86 ± 10.85	0.21
	HAI	76.45 ± 72.35	71.98 ± 24.64	0.01	60.97 ± 36.65	86.64 ± 25.29	0.00	*59.18 ± 34.42*	*81.59 ± 25.56*	*0.00*	81.28 ± 83.70	92.55 ± 20.37	0.03	182.09 ± 73.19	38.86 ± 10.85	0.21
	SCH	79.70 ± 67.66	95.62 ± 24.64	0.01	76.02 ± 43.42	103.37 ± 27.38	0.04	87.85 ± 44.65	101.54 ± 24.53	0.05	*76.14 ± 71.71*	*109.97 ± 19.35*	*0.02*	170.48 ± 70.77	51.96 ± 15.66	0.17
	US	77.93 ± 69.49	85.86 ± 26.66	0.01	74.45 ± 42.24	100.68 ± 28.68	0.03	83.05 ± 46.07	97.41 ± 27.84	0.07	77.86 ± 71.21	107.35 ± 23.30	0.03	*162.16 ± 64.37*	*52.71 ± 8.87*	*0.31*
Hybrid	CH	*85.15 ± 79.45*	*168.42 ± 64.24*	*0.02*	100.74 ± 69.41	116.83 ± 57.08	0.10	113.54 ± 67.88	84.32 ± 47.26	0.00	109.92 ± 92.14	102.34 ± 55.32	0.02	151.07 ± 108.91	152.27 ± 53.92	0.19
	ALB	96.76 ± 82.75	127.65 ± 34.26	0.03	*49.07 ± 43.37*	*100.59 ± 38.06*	*0.00*	52.49 ± 50.49	76.32 ± 33.13	0.04	96.64 ± 86.68	84.99 ± 36.44	0.06	137.02 ± 92.86	129.28 ± 28.89	0.25
	HAI	94.79 ± 82.74	130.98 ± 27.94	0.03	52.71 ± 49.66	101.47 ± 30.03	0.00	*41.62 ± 42.70*	*83.56 ± 29.03*	*0.01*	106.33 ± 98.54	89.19 ± 31.22	0.07	124.46 ± 93.54	135.26 ± 23.55	0.17
	SCH	111.38 ± 80.29	138.79 ± 41.65	0.00	93.55 ± 52.99	109.98 ± 42.53	0.00	92.91 ± 59.13	87.59 ± 43.11	0.05	*79.79 ± 68.74*	*100.39 ± 44.62*	*0.08*	92.30 ± 77.51	144.42 ± 34.61	0.17
	US	135.55 ± 98.87	137.77 ± 47.69	0.01	135.62 ± 86.92	120.75 ± 38.63	0.04	163.14 ± 82.09	107.63 ± 33.42	0.00	118.59 ± 85.61	112.79 ± 36.61	0.00	*58.76 ± 53.06*	*135.36 ± 38.89*	*0.04*

Table 6
Mean difference between predicted and measured grassland biomass and epistemic uncertainty ±1 standard deviation and coefficient of determination (R^2) for global models. For each model, 100 samples (1 prediction of 100 samples each) were available. The absolute difference was calculated by taking the absolute value after subtracting the measured biomass value from the predicted biomass value for each predicted sample. For the empirical and hybrid models, the epistemic uncertainty for a predicted sample corresponds to the predicted standard deviation of the Gaussian process regression (GPR) models. For the physically-based models, the epistemic uncertainty corresponds to the standard deviation of all selected solutions. CH: Switzerland, ALB: Schwäbische Alb, HAI: Hainich-Dün, SCH: Schorfheide-Chorin, US: United States.

	Prediction site															
	CH			ALB			HAI			SCH			US			
Model type	Absolute difference [g m^{-2}]	Epistemic uncertainty [g m^{-2}]	R^2	Absolute difference [g m^{-2}]	Epistemic uncertainty [g m^{-2}]	R^2	Absolute difference [g m^{-2}]	Epistemic uncertainty [g m^{-2}]	R^2	Absolute difference [g m^{-2}]	Epistemic uncertainty [g m^{-2}]	R^2	Absolute difference [g m^{-2}]	Epistemic uncertainty [g m^{-2}]	R^2
Empirical	89.36 ± 65.55	112.11 ± 0.00	0.00	83.81 ± 55.91	115.07 ± 0.00	0.00	105.16 ± 56.70	111.4 ± 0.00	0.00	92.12 ± 90.44	107.64 ± 0.00	0.00	102.31 ± 81.66	108.8 ± 0.00	0.00
Physically-based	81.37 ± 71.47	83.49 ± 27.37	0.00	74.45 ± 42.24	100.68 ± 28.68	0.03	83.05 ± 46.07	97.41 ± 27.84	0.07	78.35 ± 79.36	94.48 ± 21.91	0.01	182.99 ± 73.17	39.74 ± 11.05	0.20
Hybrid	98.72 ± 75.78	138.74 ± 25.66	0.01	62.54 ± 40.14	138.67 ± 51.71	0.03	57.72 ± 45.84	88.45 ± 24.39	0.14	97.38 ± 92.43	106.99 ± 34.61	0.11	101.97 ± 90.03	132.54 ± 7.63	0.38

grassland systems. Grasslands often contain a complex mixture of photosynthetic and non-photosynthetic vegetation, with litter potentially contributing a substantial fraction of total biomass (Schweiger et al., 2015a). Contrasting absorption and scattering properties of litter compared to green biomass can reduce the sensitivity of reflectance-based predictors to total biomass. High species and trait diversity of grasslands is reflected in heterogeneous canopy architectures that violate the homogeneity assumptions underlying radiative transfer models such as PROSAIL (Rossi et al., 2020). In addition, strong fine-scale spatial heterogeneity driven by grazing by wild and domestic herbivores further confounds the link between plot-scale measurements and satellite-based observations.

4.2. Transferability of local models

Several factors possibly hampered model transferability. First, the transferability of empirical and hybrid models remained limited to combinations of training or optimization and validation sites sharing similar environmental and management characteristics such as ALB and HAI. This is consistent with the findings of Muro et al. (2022) who, among the German study sites, found the prediction of SCH to be the most challenging due to an elevated level of soil organic content and different management compared to ALB and HAI (Busch et al., 2018). At the other extreme, the US site, which was poorly predicted by transferred models, was sampled during a very dry summer, and the Sentinel-2 spectra showed a substantial increase in reflectance in the short-wave infrared region, likely due to low water content confounding the overall spectra (Jacquemoud et al., 2009; Mesonet https://www.mesonet.org/weather/rainfall/monthly-rainfall-table?ref=1210, 2024; Appendix A Section A.21). The summer drought may have also led to early senescence, resulting in a potential underestimation of LAI in PROSAIL as the latter is predominantly equipped for modeling green vegetation and struggles to capture NPV (Amin et al., 2021; Delegido et al., 2015; Schiefer et al., 2021; Verrelst et al., 2023). Second, differences in the abundance of C3 and C4 grasses among sites may have further limited model transferability from European sites to the US exhibiting a substantially higher cover fraction of C4 grasses (Kothari and Schweiger, 2022; Shoko et al., 2016; Appendix A Section A.22). Third, the derivation of biomass values in physically-based and hybrid approaches via LUT-based LMA and LAI multiplication resulted in numerous parameter combinations associated with the same biomass values, increasing the prevalent ill-posedness of the model. Fourth, plot definition and field measurements were handled differently among the study sites and field campaigns. For example, the plots being monitored by the Biodiversity Exploratories were selected to be representative of an area of 2500 m^2 and biomass was cut for an area of 2 m^2 (Hinderling et al., 2023), whereas the plots in Switzerland were representative for an area of 36 m^2 and biomass harvesting was limited to 1 m^2 (Schweiger et al., 2015b). While such discrepancies between monitoring programs are likely to introduce uncertainties, they are difficult to mitigate when working with already existing field data. Moreover, the in situ data collection might not necessarily have been optimized for the pixel grid of the utilized remote sensing data, e.g., 20 × 20 m grid for Sentinel-2, leading to the problem of diminished representativeness (Hauser et al., 2021b; Schweiger, 2020) such as in the case of mixed pixels. Even though standardization of field campaigns could potentially benefit the transferability of model trained with remote sensing data, it might be hardly feasible due to ecological or policy constraints and the legitimate interest of existing monitoring programs to ensure temporally consistent measurements.

Nevertheless, physically-based models outperformed empirical and hybrid models when predicting novel study sites in an agnostic scenario which is consistent with the claimed transferability of physically-based models and the results of previous studies conducted at leaf level (Féret et al., 2019; Verrelst et al., 2015; Wang et al., 2023). For the hybrid models, we expected a comparatively high transferability due to the

physical foundation of the RTM-simulated training data. In addition, previous studies have shown the general ability of machine learning regression models to serve as RTM emulators, i.e., being able to accurately grasp the physical principles of RTMs during model training (Rivera et al., 2015; Verrelst et al., 2017). However, our results showed that the transferability of the hybrid models was weakest in most cases, potentially because the empirical features of hybrid models overrode the physical foundation of the training data by fitting data-specific single-trait relationships between the AL-selected training set and the optimization data. To increase the transferability of hybrid models, it has been suggested to increase the initial training set size (Berger et al., 2021b; Tagliabue et al., 2022), enforcing the inclusion of more general training data prior to the employment of AL. However, no general increase in transferability was apparent in our results even when using different initial training set sizes (Appendix A Section A.23).

4.3. Accuracy of global models

In this study, global models were defined as models that were trained (empirical) or optimized (physically-based and hybrid) using field data from multiple sites and applied to an unseen prediction site. The reasoning behind this model set-up was that model training or optimization with heterogeneous field data might increase model transferability by allowing the model to learn from a larger pool of diverse field data motivated by previous studies employing similar approaches (Muro et al., 2022) and the emergence of global plant trait products derived from remote sensing data using physically-based or machine learning models (Campos-Taberner et al., 2018; Kovács et al., 2023; Moreno-Martinez et al., 2020) relying on accurate predictions for unsampled areas. The results of this third model comparison were not conclusive as to which model performed best at this task.

The global physically-based models achieved lower accuracy than the local models for all sites, indicating a strong influence of the selected combinations of cost function and percentage of solutions, which differed strongly from those of the local models (Appendix A Sections A.5 and A.8). This seems to contradict the observed higher transferability of the transferred physically-based models and the eventual conclusion that physically-based models are always the best choice if no validation data is available (see Section 4.2).

One major challenge for global empirical and hybrid models was that different confounding relationships between biomass and spectral information are possible for various study sites. This was reflected for the empirical and hybrid models by the increased mean RRMSE of the US site, whose specific conditions could not be adequately learned by the models without additional contextual information (in the case of the global physically-based models, the reason for the high mean RRMSE was most likely the parametrization of the LUT as discussed in 4.1). Nevertheless, comparatively good results were obtained by the empirical models for the ALB site ($R^2 = 0.49$, RRMSE = 0.49, MBE = 28.87 g/m^2) and by the hybrid models for the ALB ($R^2 = 0.43$, RRMSE = 0.53, MBE = 37.30 g/m^2) and HAI sites ($R^2 = 0.42$, RRMSE = 0.64, MBE = 34.41 g/m^2) which is consistent with the results of Muro et al. (2022) who reported a lower transferability of models applied to the SCH site due to confounding factors such as soil organic content and management practices (Busch et al., 2018). Identifying the exact factors contributing to the divergent performance is challenging within the multi-site training set-up and limits definitive conclusions to be drawn on model selection and suitability.

4.4. Prediction uncertainty as a measure of model applicability to unseen data

Value and acceptance of remote sensing products increase with a quantitative specification of uncertainty (Woodcock, 2002); increased attention to uncertainty is also called for in the context of machine learning applications (Meyer and Pebesma, 2020). Motivated by the finding that epistemic uncertainty of GPR models might help to identify reliable and unreliable predictions for croplands (Verrelst et al., 2013b; Verrelst et al., 2013a; Verrelst et al., 2012), we tested if epistemic uncertainty could be used as proxy for model transferability across heterogeneous grasslands by comparing the absolute difference between measured and predicted biomass with the epistemic uncertainty associated with each prediction and found low correspondence between the two (as shown in Section 3.4).

In the case of GPR, epistemic uncertainty is a direct model output in the form of the predictive SD. According to Rasmussen and Williams (2006), the epistemic uncertainty of GPR models is determined by the similarity between a data point to be predicted and the training data as well as the properties of the optimized kernel, e.g., the magnitude of the constant kernel and the lengthscale of the RBF kernel. Accordingly, it is not surprising that the empirical models for the mutual prediction of the spectrally more similar ALB and HAI sites showed lower mean epistemic uncertainty than the US models, whose training data exhibited different spectral properties, e.g., the high reflectance in the short-wave infrared region for the US site. The fact that the AL-selected training data of the hybrid models did not differ much among the study sites (Appendix A Section A.24) led to a consistently lower epistemic uncertainty for the German sites, for which the spectral properties of the field data were more similar to those of the AL-selected training data. Correspondingly, a high mean epistemic uncertainty resulted for the CH and US sites, since these sites exhibited larger spectral heterogeneity (as shown in Appendix A Section A.18) or contained spectra that were not fully covered by the current parametrization of the PROSAIL LUT, respectively. Thus, a high epistemic uncertainty indicated that the data points to be predicted are not optimally covered by the training data, while a low epistemic uncertainty on the contrary did not necessarily indicate a reliable prediction as suggested by the missing correspondence between low absolute differences and associated epistemic uncertainty. Hence, using the epistemic uncertainty as a measure of model applicability to unseen data might be valid only in the case of comparable relationships between biomass and spectral information – a condition violated in our study. Nevertheless, the GPR uncertainty still serves as valuable information to optimize field sampling efforts as outlined in Verrelst et al. (2012).

In the case of physically-based models, the SD of the solutions considered in the LUT inversion process is a comparatively simple approach to express the variability of the candidate solutions and thus to quantify the diversity of the possible solutions. This diversity was relatively consistent per prediction site, e.g., mean epistemic uncertainty was between 38.86 and 51.96 g/m^2 for predictions of the US site, suggesting a subordinate influence of the cost functions and percentages of solutions used for different LUT inversions. Rather, the epistemic uncertainty appeared to depend primarily on the spectral properties of the predicted data. For example, the US site had by far the lowest epistemic uncertainty, which was not consistent with the comparatively high mean RRMSE. In fact, a more accurate prediction of the US site would have required a modification of the LUT parameterization (as shown in Appendix A Section A.19), which is obscured by relying solely on the prediction uncertainty. Therefore, in our study, the epistemic uncertainty expressed as diversity of the possible solutions did not allow a direct conclusion on the applicability of the models to unseen data.

4.5. Possible improvements

Regarding the performance of local models, various opportunities for improvement are conceivable depending on the model type. First, the use of alternative algorithms or cost functions should be considered. For empirical models, alternative algorithms such as Artificial Neural Networks (ANNs) as employed by Ali et al. (2017) and Muro et al. (2022) could be tested to determine further improvements in model accuracy. For the physically-based models, the cost functions discussed by Rivera et al. (2013) were tested but since the choice of the optimal cost functions varied between study sites, the inclusion of additional cost

functions such as genetic algorithms (Fang et al., 2003) could benefit the accuracy of the LUT inversion. Second, subsetting the LUT using correlations between PROSAIL input parameters could reduce solution space (Campos-Taberner et al., 2018). For this, in situ data of several PROSAIL input parameters are necessary whose correlations could be exploited to prevent unrealistic parameter combinations being included in the LUT (Combal et al., 2003) as done by Schiefer et al. (2021). Ideally, the in situ data should be collected for each study site, since the correlation between the parameters is subject to environmental gradients such as climatic and soil factors (Joswig et al., 2021), laying the foundation for a locally informed scenario. In such a scenario, further variables besides the aforementioned LMA (see Section 4.1) could have played a key role in increasing the representativeness of the simulated LUT for site-specific conditions. On the one hand, Berger et al. (2021a) and Verrelst et al. (2023) have shown that NPV can be modeled by coupling PROSPECT-PRO (Féret et al., 2021), the successor model to PROSPECT-D, with 4SAIL by combining carbon-based constituents (CBC) and LAI. The approximation of NPV would allow the local ratio of green and brown vegetation to be better represented. On the other hand, the default spectra for dry and wet soil of the prosail R package could be replaced by locally measured spectra or spectra contained in a spectral library in order to take their spatial variability into account. Using existing soil spectral libraries (Safanelli et al., 2025; Viscarra Rossel et al., 2016), the default spectra used for the PROSAIL simulation could be replaced by spectra tailored to local soil properties. Third, the inclusion of additional predictor variables could improve model accuracy and transferability, for example additional biomass-relevant variables such as vegetation height could lead to improved predictions. However, sampling of additional variables is time- and labor-intensive and deriving them from remote sensing data is challenging, highlighting the need for reliable, accurate, and transferable remote sensing models. For example, the mean average error (MAE) of the estimated canopy height from spaceborne laser altimeters amounts to at least 2 m (Liu et al., 2021) and the use of Sentinel-1 SAR data has led to mixed results, ranging from improving grassland biomass estimation to providing little or no added value when combined with optical data (Muro et al., 2022; Raab et al., 2020; Wang et al., 2019). For hybrid models, it is possible to employ further advanced models such as multi-output GPR models (MOGPs) if additional in situ measured PROSAIL parameters are available, which allow to predict multiple output variables while preserving the correlations among the input variables (Liu et al., 2018). The feasibility of employing MOGPs in the context of biophysical variable estimation has already been demonstrated by Pipia et al. (2019) and Caballero et al. (2023) who predicted LAI and vegetation water content (VWC), respectively, by combining Sentinel-1 and Sentinel-2 data.

To improve the local models' transferability as well as the global models' accuracy, it is essential to disentangle the site-specific confounded relationships between biomass and spectral information, effectively shifting from agnostic to locally informed biomass models. First, additional predictor variables could help to distinguish these relationships, including vegetation height, climate data, land use intensity as continuous proxy for management information (Blüthgen et al., 2012; Rossi et al., 2024), evapotranspiration, and the quantification of NPV. Second, a stratification based on climatic priors for subsequent LUT optimization would be useful in the context of diverse study sites and resonates with studies making the case for a stronger embedding of remote sensing data with ancillary data (Aguirre-Gutiérrez et al., 2021; Cavender-Bares et al., 2022; Moreno-Martinez et al., 2020; Verrelst et al., 2023). Although a remote sensing-only solution may have been desirable, our results reaffirm, through a new perspective, the trade-off between a LUT parametrization covering the ecological conditions of all sites and a degradation of the LUT inversion accuracy. Such a stratification would allow, in the first step, the data points to be projected into different environmental regimes based on climate data and in the second step, to use LUTs subsetted based on correlations between PROSAIL parameters and with optimized parameter ranges for the physically-based or hybrid models. While coarse-resolution climate data for the first step are freely available over large spatial scales (e.g., ERA5-Land; Muñoz Sabater, 2019), the second step still places high demand on the availability of in situ data.

5. Conclusion

In this study, we compared different models for biomass estimation from multispectral spaceborne remote sensing data across heterogeneous grasslands to assess model assumptions and facilitate model choice for specific applications. In three model comparisons, we investigated the accuracy and transferability of empirical, physically-based, and hybrid models across five study sites regarding (1) their local applicability, (2) their spatial transferability, and (3) the opportunity to compile field data from multiple study sites to increase transferability. Our results showed that:

1) on the local level, all models performed similarly well in terms of RRMSE and R^2. Further, we found that in the context of hybrid models, employing AL to identify the most informative training samples was required.
2) when transferring local models to a different study site, the physically-based models led to the most promising results for most combinations of training (empirical models) or optimization (physically-based and hybrid models) and validation sites. Moreover, we observed a trade-off between LUT specificity and generality, impeding the universal application of a single physically-based model. The transferability of empirical and hybrid models was limited to combinations of sites sharing similar ecological and/or spectral conditions.
3) when compiling the field data of four study sites to predict the remaining one, no model clearly outperformed the others. Differences in model performance remained challenging to explain, highlighting the need to further explore the possibilities and characterize the trade-offs of developing models applicable on a large spatial scale and across ecological gradients.
4) common epistemic uncertainty implementations were not necessarily reliable measures of model applicability to unseen data in the case of varying relationships between biomass and spectral information across study sites. While high epistemic uncertainty indicates suboptimal coverage of the data points to be predicted by the training data and can therefore guide future sampling efforts, low epistemic uncertainty does not necessarily indicate high prediction accuracy.
5) model transferability needs to be thoroughly tested when developing remote sensing applications whose intended applicability goes beyond the local scale.

Possibilities for improving local model accuracy and transferability include the testing of alternative machine and deep learning algorithms and cost functions, hierarchical subsetting of LUTs based on ecological priors, and the use of multi-output models to preserve correlations among predictor variables, incorporating additional predictors such as NPV or climate data, and exploring multi-sensor approaches. To fully exploit the spatio-temporal potential of satellite observations, evaluating and improving model transferability should be a priority; this requires datasets that enable rigorous testing across sites and conditions, along with theory-driven foundations to build models capable of robust prediction when applied to new unseen data.

CRediT authorship contribution statement

Jan Schweizer: Writing – original draft, Visualization, Software, Methodology, Investigation, Data curation, Conceptualization. **Leon T. Hauser:** Writing – original draft, Supervision, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization. **Hamed**

Gholizadeh: Writing – review & editing, Data curation. **Anna K. Schweiger:** Writing – review & editing, Data curation. **Christian Rossi:** Writing – original draft, Supervision, Methodology, Investigation, Funding acquisition, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was conducted under the 'Biodiv-Watch' project, funded by the ESA Initial Support for Innovation (EISI) [Grant Number 4000138090], and with the financial and operational support of the Swiss National Park (SNP). Hamed Gholizadeh was supported by a NASA NIP award [80NSSC21K0941]. This work is based on data obtained within the DFG Priority Program 1374 'Infrastructure-Biodiversity-Exploratories'. We thank the staff of the three exploratories, the BE office and the BExIS team for their work in maintaining the plot and project infrastructure, and Markus Fischer, the late Elisabeth Kalko, Eduard Linsenmair, Dominik Hessenmöller, Jens Nieschulze, Daniel Prati, Ingo Schöning, François Buscot, Ernst-Detlef Schulze and Wolfgang W. Weisser for their role in setting up the Biodiversity Exploratories project. Field work permits were issued by the responsible state environmental offices of Baden-Württemberg, Thüringen, and Brandenburg (according to § 72 BbgNatSchG). The authors thank Antonia Ludwig for her insightful comments on an earlier version of the manuscript and Matteo Delucchi of the consulting team of the Applied Statistics group at the Department of Mathematical Modeling and Machine Learning, University of Zurich. The authors would like to thank the anonymous reviewers for their valuable comments and suggestions.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2026.115294.

Data availability

This work is based on data elaborated by the BExIS and Botany core projects of the Biodiversity Exploratories program (DFG Priority Program 1374). The datasets are publicly available in the Biodiversity Exploratories Information System (http://doi.org/10.17616/R32P9Q). The datasets are listed in the references section. Field data for Switzerland and Oklahoma are available upon request. Python codes and synthetic mock data are available upon request.

References

Aguirre-Gutiérrez, J., Rifai, S., Shenkin, A., Oliveras, I., Bentley, L.P., Svátek, M., Girardin, C.A.J., Both, S., Riutta, T., Berenguer, E., Kissling, W.D., Bauman, D., Raab, N., Moore, S., Farfan-Rios, W., Figueiredo, A.E.S., Reis, S.M., Ndong, J.E., Ondo, F.E., N'ssi Bengone, N., Mihindou, V., Moraes de Seixas, M.M., Adu-Bredu, S., Abernethy, K., Asner, G.P., Barlow, J., Burslem, D.F.R.P., Coomes, D.A., Cernusak, L. A., Dargie, G.C., Enquist, B.J., Ewers, R.M., Ferreira, J., Jeffery, K.J., Joly, C.A., Lewis, S.L., Marimon-Junior, B.H., Martin, R.E., Morandi, P.S., Phillips, O.L., Quesada, C.A., Salinas, N., Schwantes Marimon, B., Silman, M., Teh, Y.A., White, L.J. T., Malhi, Y., 2021. Pantropical modelling of canopy functional traits using Sentinel-2 remote sensing data. Remote Sens. Environ. 252, 112122. https://doi.org/10.1016/j.rse.2020.112122.

Ali, I., Cawkwell, F., Dwyer, E., Barrett, B., Green, S., 2016. Satellite remote sensing of grasslands: from observation to management. J. Plant Ecol. 9, 649–671. https://doi.org/10.1093/jpe/rtw005.

Ali, I., Cawkwell, F., Dwyer, E., Green, S., 2017. Modeling managed grassland biomass estimation by using multitemporal remote sensing data—a machine learning approach. IEEE J Sel Top Appl Earth Obs Remote Sens 10, 3254–3264. https://doi.org/10.1109/JSTARS.2016.2561618.

Amin, E., Verrelst, J., Rivera-Caicedo, J.P., Pipia, L., Ruiz-Verdú, A., Moreno, J., 2021. Prototyping Sentinel-2 green LAI and brown LAI products for cropland monitoring. Remote Sens. Environ. 255, 112168. https://doi.org/10.1016/j.rse.2020.112168.

Anderson, R.H., Fuhlendorf, S.D., Engle, D.M., 2006. Soil nitrogen availability in tallgrass prairie under the fire–grazing interaction. Rangel. Ecol. Manage. 59, 625–631. https://doi.org/10.2111/05-088R2.1.

Bai, Y., Cotrufo, M.F., 2022. Grassland soil carbon sequestration: current understanding, challenges, and solutions. Science 1979 (377), 603–608. https://doi.org/10.1126/science.abo2380.

Bailey, D.W., Gross, J.E., Laca, E.A., Rittenhouse, L.R., Coughenour, M.B., Swift, D.M., Sims, P.L., 1996. Mechanisms that result in large herbivore grazing distribution patterns. J. Range Manage. 49, 386. https://doi.org/10.2307/4002919.

Bardgett, R.D., Bullock, J.M., Lavorel, S., Manning, P., Schaffner, U., Ostle, N., Chomel, M., Durigan, G.L., Fry, E., Johnson, D., Lavallee, J.M., Le Provost, G., Luo, S., Png, K., Sankaran, M., Hou, X., Zhou, H., Ma, L., Ren, W., Li, X., Ding, Y., Li, Y., Shi, H., 2021. Combatting global grassland degradation. Nat Rev Earth Environ 2, 720–735. https://doi.org/10.1038/s43017-021-00207-2.

Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood, E.F., 2018. Present and future Köppen-Geiger climate classification maps at 1-km resolution. Sci. Data 5, 180214. https://doi.org/10.1038/sdata.2018.214.

Bengtsson, J., Bullock, J.M., Egoh, B., Everson, C., Everson, T., O'Connor, T., O'Farrell, P. J., Smith, H.G., Lindborg, R., 2019. Grasslands—more important for ecosystem services than you might think. Ecosphere 10, e02582. https://doi.org/10.1002/ecs2.2582.

Berger, K., Atzberger, C., Danner, M., D'Urso, G., Mauser, W., Vuolo, F., Hank, T., 2018. Evaluation of the PROSAIL model capabilities for future hyperspectral model environments: a review study. Remote Sens. 10, 85. https://doi.org/10.3390/rs10010085.

Berger, K., Verrelst, J., Féret, J.-B., Hank, T., Wocher, M., Mauser, W., Camps-Valls, G., 2020. Retrieval of aboveground crop nitrogen content with a hybrid machine learning method. International Journal of Applied Earth Observation and Geoinformation 92, 102174. https://doi.org/10.1016/j.jag.2020.102174.

Berger, K., Hank, T., Halabuk, A., Rivera-Caicedo, J.P., Wocher, M., Mojses, M., Gerhátová, K., Tagliabue, G., Dolz, M.M., Venteo, A.B.P., Verrelst, J., 2021a. Assessing non-photosynthetic cropland biomass from Spaceborne hyperspectral imagery. Remote Sens. 13, 4711. https://doi.org/10.3390/rs13224711.

Berger, K., Rivera Caicedo, J.P., Martino, L., Wocher, M., Hank, T., Verrelst, J., 2021b. A survey of active learning for quantifying vegetation traits from terrestrial earth observation data. Remote Sens. 13, 287. https://doi.org/10.3390/rs13020287.

Binh, N.A., Hauser, L.T., Viet Hoa, P., Thi Phuong Thao, G., An, N.N., Nhut, H.S., Phuong, T.A., Verrelst, J., 2022. Quantifying mangrove leaf area index from Sentinel-2 imagery using hybrid models and active learning. Int J Remote Sens 43, 5636–5657. https://doi.org/10.1080/01431161.2021.2024912.

Blüthgen, N., Dormann, C.F., Prati, D., Klaus, V.H., Kleinebecker, T., Hölzel, N., Alt, F., Boch, S., Gockel, S., Hemp, A., Müller, J., Nieschulze, J., Renner, S.C., Schöning, I., Schumacher, U., Socher, S.A., Wells, K., Birkhofer, K., Buscot, F., Oelmann, Y., Rothenwöhrer, C., Scherber, C., Tscharntke, T., Weiner, C.N., Fischer, M., Kalko, E.K. V., Linsenmair, K.E., Schulze, E.-D., Weisser, W.W., 2012. A quantitative index of land-use intensity in grasslands: integrating mowing, grazing and fertilization. Basic Appl. Ecol. 13, 207–220. https://doi.org/10.1016/j.baae.2012.04.001.

Bolliger, R., Prati, D., Fischer, M., 2020. Vegetation Records for Grassland EPs, 2008 - 2020. Version 2. Biodiversity Exploratories Information System. Dataset.. https://www.bexis.uni-jena.de/ddm/data/Showdata/27386. Dataset ID = 27386.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Brown, L.A., Ogutu, B.O., Dash, J., 2019. Estimating Forest leaf area index and canopy chlorophyll content with Sentinel-2: An evaluation of two hybrid retrieval algorithms. Remote Sens. 11, 1752. https://doi.org/10.3390/rs11151752.

Busch, V., Klaus, V.H., Penone, C., Schäfer, D., Boch, S., Prati, D., Müller, J., Socher, S.A., Niinemets, Ü., Peñuelas, J., Hölzel, N., Fischer, M., Kleinebecker, T., 2018. Nutrient stoichiometry and land use rather than species richness determine plant functional diversity. Ecol. Evol. 8, 601–616. https://doi.org/10.1002/ece3.3609.

Caballero, G., Pezzola, A., Winschel, C., Sanchez Angonova, P., Casella, A., Orden, L., Salinero-Delgado, M., Reyes-Muñoz, P., Berger, K., Delegido, J., Verrelst, J., 2023. Synergy of Sentinel-1 and Sentinel-2 time series for cloud-free vegetation water content mapping with multi-output Gaussian processes. Remote Sens. 15, 1822. https://doi.org/10.3390/rs15071822.

Campos-Taberner, M., Moreno-Martínez, Á., García-Haro, F., Camps-Valls, G., Robinson, N., Kattge, J., Running, S., 2018. Global estimation of biophysical Variables from Google earth engine platform. Remote Sens. 10, 1167. https://doi.org/10.3390/rs10081167.

Cavender-Bares, J., Schneider, F.D., Santos, M.J., Armstrong, A., Carnaval, A., Dahlin, K. M., Fatoyinbo, L., Hurtt, G.C., Schimel, D., Townsend, P.A., Ustin, S.L., Wang, Z., Wilson, A.M., 2022. Integrating remote sensing with ecology and evolution to advance biodiversity conservation. Nat Ecol Evol 6, 506–519. https://doi.org/10.1038/s41559-022-01702-5.

Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining 13-17-August-2016, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

Combal, B., Baret, F., Weiss, M., Trubuil, A., Macé, D., Pragnère, A., Myneni, R., Knyazikhin, Y., Wang, L., 2003. Retrieval of canopy biophysical variables from bidirectional reflectance. Remote Sens. Environ. 84, 1–15. https://doi.org/10.1016/S0034-4257(02)00035-4.

Darvishzadeh, R., Atzberger, C., Skidmore, A., Schlerf, M., 2011. Mapping grassland leaf area index with airborne hyperspectral imagery: a comparison study of statistical

approaches and inversion of radiative transfer models. ISPRS Journal of Photogrammetry and Remote Sensing 66, 894–906. https://doi.org/10.1016/J.ISPRSJPRS.2011.09.013.
de Sá, N.C., Baratchi, M., Hauser, L.T., van Bodegom, P., 2021. Exploring the impact of noise on hybrid inversion of PROSAIL RTM on Sentinel-2 data. Remote Sens. 13, 648. https://doi.org/10.3390/rs13040648.
De Vroey, M., de Vendictis, L., Zavagli, M., Bontemps, S., Heymans, D., Radoux, J., Koetz, B., Defourny, P., 2022. Mowing detection using Sentinel-1 and Sentinel-2 time series for large scale grassland monitoring. Remote Sens. Environ. 280, 113145. https://doi.org/10.1016/j.rse.2022.113145.
Dehghan-Shoar, M.H., Orsi, A.A., Pullanagari, R.R., Yule, I.J., 2023. A hybrid model to predict nitrogen concentration in heterogeneous grassland using field spectroscopy. Remote Sens. Environ. 285, 113385. https://doi.org/10.1016/j.rse.2022.113385.
Delegido, J., Verrelst, J., Rivera, J.P., Ruiz-Verdú, A., Moreno, J., 2015. Brown and green LAI mapping through spectral indices. Int. J. Appl. Earth Obs. Geoinf. 35, 350–358. https://doi.org/10.1016/j.jag.2014.10.001.
Dixon, A.P., Faber-Langendoen, D., Josse, C., Morrison, J., Loucks, C.J., 2014. Distribution mapping of world grassland types. J. Biogeogr. 41, 2003–2019. https://doi.org/10.1111/jbi.12381.
Erb, K.-H., Kastner, T., Plutzar, C., Bais, A.L.S., Carvalhais, N., Fetzel, T., Gingrich, S., Haberl, H., Lauk, C., Niedertscheider, M., Pongratz, J., Thurner, M., Luyssaert, S., 2018. Unexpectedly large impact of forest management and grazing on global vegetation biomass. Nature 553, 73–76. https://doi.org/10.1038/nature25138.
Fang, H., Liang, S., Kuusk, A., 2003. Retrieving leaf area index using a genetic algorithm with a canopy radiative transfer model. Remote Sens. Environ. 85, 257–270. https://doi.org/10.1016/S0034-4257(03)00005-1.
Féret, J., de Boissieu, F., 2022. PROSAIL: PROSAIL Leaf and Canopy Radiative Transfer Model and Inversion Routines. R package version 1.1.1 [WWW document]. URL. https://gitlab.com/jbferet/prosail.
Féret, J.-B., François, C., Asner, G.P., Gitelson, A.A., Martin, R.E., Bidel, L.P.R., Ustin, S. L., le Maire, G., Jacquemoud, S., 2008. PROSPECT-4 and 5: advances in the leaf optical properties model separating photosynthetic pigments. Remote Sens. Environ. 112, 3030–3043. https://doi.org/10.1016/j.rse.2008.02.012.
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. https://doi.org/10.1016/j.rse.2017.03.004.
Féret, J.-B., le Maire, G., Jay, S., Berveiller, D., Bendoula, R., Hmimina, G., Cheraiet, A., Oliveira, J.C., Ponzoni, F.J., Solanki, T., de Boissieu, F., Chave, J., Nouvellon, Y., Porcar-Castell, A., Proisy, C., Soudani, K., Gastellu-Etchegorry, J.-P., Lefèvre-Fonollosa, M.-J., 2019. Estimating leaf mass per area and equivalent water thickness based on leaf optical properties: potential and limitations of physical modeling and machine learning. Remote Sens. Environ. 231, 110959. https://doi.org/10.1016/j.rse.2018.11.002.
Féret, J.-B., Berger, K., de Boissieu, F., Malenovský, Z., 2021. PROSPECT-PRO for estimating content of nitrogen-containing leaf proteins and other carbon-based constituents. Remote Sens. Environ. 252, 112173. https://doi.org/10.1016/j.rse.2020.112173.
Fischer, M., Bossdorf, O., Gockel, S., Hänsel, F., Hemp, A., Hessenmöller, D., Korte, G., Nieschulze, J., Pfeiffer, S., Prati, D., Renner, S., Schöning, I., Schumacher, U., Wells, K., Buscot, F., Kalko, E.K.V., Linsenmair, K.E., Schulze, E.-D., Weisser, W.W., 2010. Implementing large-scale and long-term functional biodiversity research: the biodiversity Exploratories. Basic Appl. Ecol. 11, 473–485. https://doi.org/10.1016/j.baae.2010.07.009.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 29, 1189–1232. https://doi.org/10.1214/AOS/1013203451.
Fuhlendorf, S.D., Engle, D.M., 2001. Restoring heterogeneity on rangelands: ecosystem management based on evolutionary grazing patterns. Bioscience 51, 625–632. https://doi.org/10.1641/0006-3568(2001)051[0625:RHOREM]2.0.CO;2.
Fuhlendorf, S.D., Engle, D.M., 2004. Application of the fire-grazing interaction to restore a shifting mosaic on tallgrass prairie. J. Appl. Ecol. 41, 604–614. https://doi.org/10.1111/j.0021-8901.2004.00937.x.
Gascon, F., Bouzinac, C., Thépaut, O., Jung, M., Francesconi, B., Louis, J., Lonjou, V., Lafrance, B., Massera, S., Gaudel-Vacaresse, A., Languille, F., Alhammoud, B., Viallefont, F., Pflug, B., Bieniarz, J., Clerc, S., Pessiot, L., Trémas, T., Cadau, E., De Bonis, R., Isola, C., Martimort, P., Fernandez, V., 2017. Copernicus sentinel-2A calibration and products validation status. Remote Sens. 9, 584. https://doi.org/10.3390/rs9060584.
Gholizadeh, H., Friedman, M.S., McMillan, N.A., Hammond, W.M., Hassani, K., Sams, A. V., Charles, M.D., Garrett, D.R., Joshi, O., Hamilton, R.G., Fuhlendorf, S.D., Trowbridge, A.M., Adams, H.D., 2022. Mapping invasive alien species in grassland ecosystems using airborne imaging spectroscopy and remotely observable vegetation functional traits. Remote Sens. Environ. 271, 112887. https://doi.org/10.1016/j.rse.2022.112887.
Gholizadeh, H., Rakotoarivony, M.N.A., Hassani, K., Johnson, K.G., Hamilton, R.G., Fuhlendorf, S.D., Schneider, F.D., Bachelot, B., 2024. Advancing our understanding of plant diversity-biological invasion relationships using imaging spectroscopy. Remote Sens. Environ. 304, 114028. https://doi.org/10.1016/j.rse.2024.114028.
Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031.
Guerini Filho, M., Kuplich, T.M., De Quadros, F.L.F., 2020. Estimating natural grassland biomass by vegetation indices using sentinel 2 remote sensing data. Int. J. Remote Sens. 41, 2861–2876. https://doi.org/10.1080/01431161.2019.1697004.
Habel, J.C., Dengler, J., Janišová, M., Török, P., Wellstein, C., Wiezik, M., 2013. European grassland ecosystems: threatened hotspots of biodiversity. Biodivers. Conserv. 22, 2131–2138. https://doi.org/10.1007/s10531-013-0537-x.
Hamilton, R.G., 2007. Restoring heterogeneity on the tallgrass prairie preserve: Applying the fire-grazing interaction model. In: Masters, R.E., Galley, K.E.M. (Eds.), Proceedings of the 23rd Tall Timbers Fire Ecology Conference: Fire in Grassland and Shrubland Ecosystems. Tall Timbers Research Station, Tallahassee, FL, pp. 163–169.
Hauser, L.T., Féret, J.-B., An Binh, N., van der Windt, N., Sil, Â.F., Timmermans, J., Soudzilovskaia, N.A., van Bodegom, P.M., 2021a. Towards scalable estimation of plant functional diversity from Sentinel-2: in-situ validation in a heterogeneous (semi-)natural landscape. Remote Sens. Environ. 262, 112505. https://doi.org/10.1016/j.rse.2021.112505.
Hauser, L.T., Timmermans, J., van der Windt, N., Sil, Â.F., César de Sá, N., Soudzilovskaia, N.A., van Bodegom, P.M., 2021b. Explaining discrepancies between spectral and in-situ plant diversity in multispectral satellite earth observation. Remote Sens. Environ. 265, 112684. https://doi.org/10.1016/j.rse.2021.112684.
He, L., Li, A., Yin, G., Nan, X., Bian, J., 2019. Retrieval of grassland aboveground biomass through inversion of the PROSAIL model with MODIS imagery. Remote Sensing 11, 1597. https://doi.org/10.3390/RS11131597.
Helfenstein, I.S., Schneider, F.D., Schaepman, M.E., Morsdorf, F., 2022. Assessing biodiversity from space: impact of spatial and spectral resolution on trait-based functional diversity. Remote Sens. Environ. 275, 113024. https://doi.org/10.1016/j.rse.2022.113024.
Hinderling, J., Penone, C., Fischer, M., Prati, D., Hoelzel, N., 2023. Biomass data for grassland EPs, 2009 - 2022. Version 8. Biodiversity Exploratories Information System. Dataset. https://www.bexis.uni-jena.de/ddm/data/Showdata/31448. Dataset ID = 31448.
Hobohm, C., Bruchmann, I., 2009. Endemische Gefäßpflanzen und ihre Habitate in Europa - Plädoyer für den Schutz der Grasland-Ökosysteme. Berichte der Reinhold-Tüxen-Gesellschaft 21, 142–161.
Mesonet, 2024. Monthly Rainfall Table: Norman (NRMN) Monthly Actual Oklahoma Mesonet Rainfall Table (in inches). https://www.mesonet.org/weather/rainfall/monthly-rainfall-table?ref=1210 (accessed 1.4.24).
Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83, 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2.
Jacquemoud, S., Baret, F., 1990. PROSPECT: a model of leaf optical properties spectra. Remote Sens. Environ. 34, 75–91. https://doi.org/10.1016/0034-4257(90)90100-Z.
Jacquemoud, S., Verhoef, W., Baret, F., Bacour, C., Zarco-Tejada, P.J., Asner, G.P., François, C., Ustin, S.L., 2009. PROSPECT + SAIL models: a review of use for vegetation characterization. Remote Sens. Environ. 113, S56–S66. https://doi.org/10.1016/J.RSE.2008.01.026.
Jehle, M., Hueni, A., Damm, A., D'Odorico, P., Weyermann, J., Kneubiihler, M., Schlapfer, D., Schaepman, M.E., Meuleman, K., 2010. APEX - current status, performance and validation concept. In: 2010 IEEE Sensors. IEEE, pp. 533–537. https://doi.org/10.1109/ICSENS.2010.5690122.
Joswig, J.S., Wirth, C., Schuman, M.C., Kattge, J., Reu, B., Wright, I.J., Sippel, S.D., Rüger, N., Richter, R., Schaepman, M.E., van Bodegom, P.M., Cornelissen, J.H.C., Díaz, S., Hattingh, W.N., Kramer, K., Lens, F., Niinemets, Ü., Reich, P.B., Reichstein, M., Römermann, C., Schrodt, F., Anand, M., Bahn, M., Byun, C., Campetella, G., Cerabolini, B.E.L., Craine, J.M., Gonzalez-Melo, A., Gutiérrez, A.G., He, T., Higuchi, P., Jactel, H., Kraft, N.J.B., Minden, V., Onipchenko, V., Peñuelas, J., Pillar, V.D., Sosinski, Ê., Soudzilovskaia, N.A., Weiher, E., Mahecha, M.D., 2021. Climatic and soil factors explain the two-dimensional spectrum of global plant trait variation. Nat Ecol Evol 6, 36–50. https://doi.org/10.1038/s41559-021-01616-8.
Kothari, S., Schweiger, A.K., 2022. Plant spectra as integrative measures of plant phenotypes. J. Ecol. 110, 2536–2554. https://doi.org/10.1111/1365-2745.13972.
Kovács, D.D., Reyes-Muñoz, P., Salinero-Delgado, M., Mészáros, V.I., Berger, K., Verrelst, J., 2023. Cloud-free global maps of essential vegetation traits processed from the TOA Sentinel-3 catalogue in Google earth engine. Remote Sens. 15, 3404. https://doi.org/10.3390/rs15133404.
Lawrence, D.M., Fisher, R.A., Koven, C.D., Oleson, K.W., Swenson, S.C., Bonan, G., Collier, N., Ghimire, B., van Kampenhout, L., Kennedy, D., Kluzek, E., Lawrence, P.J., Li, F., Li, H., Lombardozzi, D., Riley, W.J., Sacks, W.J., Shi, M., Vertenstein, M., Wieder, W.R., Xu, C., Ali, A.A., Badger, A.M., Bisht, G., van den Broeke, M., Brunke, M.A., Burns, S.P., Buzan, J., Clark, M., Craig, A., Dahlin, K., Drewniak, B., Fisher, J.B., Flanner, M., Fox, A.M., Gentine, P., Hoffman, F., Keppel-Aleks, G., Knox, R., Kumar, S., Lenaerts, J., Leung, L.R., Lipscomb, W.H., Lu, Y., Pandey, A., Pelletier, J.D., Perket, J., Randerson, J.T., Ricciuto, D.M., Sanderson, B.M., Slater, A., Subin, Z.M., Tang, J., Thomas, R.Q., Val Martin, M., Zeng, X., 2019. The community land model version 5: description of new features, benchmarking, and impact of forcing uncertainty. J Adv Model Earth Syst 11, 4245–4287. https://doi.org/10.1029/2018MS001583.
Lemaire, G., Hodgson, J., Chabbi, A. (Eds.), 2011. Grassland Productivity and Ecosystem Services. CAB International, Wallingford.
Li, F., Zeng, Y., Luo, J., Ma, R., Wu, B., 2016. Modeling grassland aboveground biomass using a pure vegetation index. Ecol. Indic. 62, 279–288. https://doi.org/10.1016/j.ecolind.2015.11.005.
Li, C., Zhou, L., Xu, W., 2021. Estimating aboveground biomass using Sentinel-2 MSI data and ensemble algorithms for grassland in the Shengjin Lake wetland, China. Remote Sensing 13, 1595. https://doi.org/10.3390/RS13081595.
Limb, R.F., Hickman, K.R., Engle, D.M., Norland, J.E., Fuhlendorf, S.D., 2007. Digital photography: reduced investigator variation in visual obstruction measurements for southern tallgrass prairie. Rangel. Ecol. Manage. 60, 548–552. https://doi.org/10.2111/1551-5028(2007)60[548:DPRIVI]2.0.CO;2.
Liu, H., Cai, J., Ong, Y.-S., 2018. Remarks on multi-output Gaussian process regression. Knowl Based Syst 144, 102–121. https://doi.org/10.1016/j.knosys.2017.12.034.

Liu, A., Cheng, X., Chen, Z., 2021. Performance evaluation of GEDI and ICESat-2 laser altimeter data for terrain and canopy height retrievals. Remote Sens. Environ. 264, 112571. https://doi.org/10.1016/j.rse.2021.112571.

Locherer, M., Hank, T., Danner, M., Mauser, W., 2015. Retrieval of seasonal leaf area index from simulated EnMAP data through optimized LUT-based inversion of the PROSAIL model. Remote Sens. 7, 10321–10346. https://doi.org/10.3390/rs70810321.

Martínez-Ferrer, L., Moreno-Martínez, Á., Campos-Taberner, M., García-Haro, F.J., Muñoz-Marí, J., Running, S.W., Kimball, J., Clinton, N., Camps-Valls, G., 2022. Quantifying uncertainty in high resolution biophysical variable retrieval with machine learning. Remote Sens. Environ. 280, 113199. https://doi.org/10.1016/J.RSE.2022.113199.

Masek, J.G., Wulder, M.A., Markham, B., McCorkel, J., Crawford, C.J., Storey, J., Jenstrom, D.T., 2020. Landsat 9: empowering open science and applications through continuity. Remote Sens. Environ. 248, 111968. https://doi.org/10.1016/j.rse.2020.111968.

Meyer, H., Pebesma, E., 2020. Predicting into unknown space? Estimating the area of applicability of spatial prediction models. Methods Ecol. Evol. 12, 1620–1633. https://doi.org/10.1111/2041-210X.13650.

Moreno-Martinez, A., Camps-Valls, G., Kattge, J., Robinson, N., Reichstein, M., van Bodegom, P., Kramer, K., Cornelissen, J.H.C., Reich, P., Bahn, M., Niinemets, U., Peñuelas, J., Craine, J., Cerabolini, B.E.L., Minden, V., Laughlin, D.C., Sack, L., Allred, B., Baraloto, C., Byun, C., Soudzilovskaia, N.A., Running, S.W., 2020. A methodology to derive global maps of leaf traits using remote sensing and climate data. Remote Sens. Environ. 218, 69–88. https://doi.org/10.1016/j.rse.2018.09.006.

Müller-Wilm, U., Louis, J., Richter, R., Gascon, F., Niezette, M., 2013. Sentinel-2 level 2A prototype processor: architecture, algorithms and first results. In: Proceedings of the ESA Living Planet Symposium. Edinburgh, UK, pp. 9–13.

Muñoz Sabater, J., 2019. ERA5-land hourly data from 1950 to present. In: Copernicus Climate Change Service (C3S). Climate Data Store. https://doi.org/10.24381/cds.e2161bac.

Muro, J., Linstädter, A., Magdon, P., Wöllauer, S., Männer, F.A., Schwarz, L.-M., Ghazaryan, G., Schultz, J., Malenovský, Z., Dubovyk, O., 2022. Predicting plant biomass and species richness in temperate grasslands across regions, time, and land management with remote sensing and deep learning. Remote Sens. Environ. 282, 113262. https://doi.org/10.1016/j.rse.2022.113262.

O'Mara, F.P., 2012. The role of grasslands in food security and climate change. Ann. Bot. 110, 1263–1270. https://doi.org/10.1093/aob/mcs209.

Ostrowski, A., Nieschulze, J., Schulze, E.D., König-Ries, B., 2020. Coordinates and Inventory Overview of all Grid Plots (GPs). Version 6. Biodiversity Exploratories Information System. Dataset. https://www.bexis.uni-jena.de (Dataset ID = 20907).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Müller, A., Nothman, J., Louppe, G., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2012. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Pipia, L., Muñoz-Marí, J., Amin, E., Belda, S., Camps-Valls, G., Verrelst, J., 2019. Fusing optical and SAR time series for LAI gap filling with multioutput Gaussian processes. Remote Sens. Environ. 235, 111452. https://doi.org/10.1016/j.rse.2019.111452.

Poortinga, A., Tenneson, K., Shapiro, A., Nquyen, Q., San Aung, K., Chishtie, F., Saah, D., 2019. Mapping plantations in Myanmar by fusing Landsat-8, Sentinel-2 and Sentinel-1 data along with systematic error quantification. Remote Sens. 11, 831. https://doi.org/10.3390/rs11070831.

Punalekar, S.M., Verhoef, A., Quaife, T.L., Humphries, D., Bermingham, L., Reynolds, C. K., 2018. Application of sentinel-2A data for pasture biomass monitoring using a physically based radiative transfer model. Remote Sens. Environ. 218, 207–220. https://doi.org/10.1016/j.rse.2018.09.028.

Quan, X., He, B., Yebra, M., Yin, C., Liao, Z., Zhang, X., Li, X., 2017. A radiative transfer model-based method for the estimation of grassland aboveground biomass. International Journal of Applied Earth Observation and Geoinformation 54, 159–168. https://doi.org/10.1016/J.JAG.2016.10.002.

R Core Team, 2021. R: A Language and Environment for Statistical Computing. Vienna.

Raab, C., Riesch, F., Tonn, B., Barrett, B., Meißner, M., Balkenhol, N., Isselstein, J., 2020. Target-oriented habitat and wildlife management: estimating forage quantity and quality of semi-natural grasslands with Sentinel-1 and Sentinel-2 data. Remote Sens Ecol Conserv 6, 381–398. https://doi.org/10.1002/rse2.149.

Ranghetti, M., Boschetti, M., Ranghetti, L., Tagliabue, G., Panigada, C., Gianinetto, M., Verrelst, J., Candiani, G., 2022. Assessment of maize nitrogen uptake from PRISMA hyperspectral data through hybrid modelling. Eur J Remote Sens. https://doi.org/10.1080/22797254.2022.2117650.

Rasmussen, C.E., Williams, C.K.I., 2006. Gaussian Processes for Machine Learning. MIT Press, Cambridge, MA.

Reinermann, S., Asam, S., Kuenzer, C., 2020. Remote sensing of grassland production and management—a review. Remote Sens. 12, 1949. https://doi.org/10.3390/rs12121949.

Rempfler, T., Rossi, C., Schweizer, J., Peters, W., Signer, C., Filli, F., Jenny, H., Hackländer, K., Buchmann, S., Anderwald, P., 2024. Remote sensing reveals the role of forage quality and quantity for summer habitat use in red deer. Mov Ecol 12, 80. https://doi.org/10.1186/s40462-024-00521-6.

Reuter, H.I., Nelson, A., Jarvis, A., 2007. An evaluation of void-filling interpolation methods for SRTM data. Int. J. Geogr. Inf. Sci. 21, 983–1008. https://doi.org/10.1080/13658810601169899.

Richter, R., Schläpfer, D., 2002. Geo-atmospheric processing of airborne imaging spectrometry data. Part 2: atmospheric/topographic correction. Int J Remote Sens 23, 2631–2649. https://doi.org/10.1080/01431160110115834.

Richter, K., Atzberger, C., Hank, T.B., Mauser, W., 2012. Derivation of biophysical variables from earth observation data: validation and statistical measures. J. Appl. Remote. Sens. 6, 063557–063561. https://doi.org/10.1117/1.JRS.6.063557.

Rivera, J., Verrelst, J., Leonenko, G., Moreno, J., 2013. Multiple cost functions and regularization options for improved retrieval of leaf chlorophyll content and LAI through inversion of the PROSAIL model. Remote Sens. 5, 3280–3304. https://doi.org/10.3390/rs5073280.

Rivera, J., Verrelst, J., Gómez-Dans, J., Muñoz-Marí, J., Moreno, J., Camps-Valls, G., 2015. An emulator toolbox to approximate radiative transfer models with statistical learning. Remote Sens. 7, 9347–9370. https://doi.org/10.3390/rs70709347.

Rossi, C., Kneubühler, M., Schütz, M., Schaepman, M.E., Haller, R.M., Risch, A.C., 2020. From local to regional: functional diversity in differently managed alpine grasslands. Remote Sens. Environ. 236, 111415. https://doi.org/10.1016/j.rse.2019.111415.

Rossi, C., McMillan, N.A., Schweizer, J.M., Gholizadeh, H., Groen, M., Ioannidis, N., Hauser, L.T., 2024. Parcel level temporal variance of remotely sensed spectral reflectance predicts plant diversity. Environ. Res. Lett. 19, 074023. https://doi.org/10.1088/1748-9326/ad545a.

Roudier, P., 2021. Introduction to Conditioned Latin Hypercube Sampling With the clhs Package [WWW document]. URL. https://CRAN.R-project.org/package=clhs.

Roy, D.P., Wulder, M.A., Loveland, T.R., Allen, R.G., Anderson, M.C., Helder, D., Irons, J. R., Johnson, D.M., Kennedy, R., Scambos, T.A., Schaaf, C.B., Schott, J.R., Sheng, Y., Vermote, E.F., Belward, A.S., Bindschadler, R., Cohen, W.B., Gao, F., Hipple, J.D., Hostert, P., Huntington, J., Justice, C.O., Kilic, A., Kovalskyy, V., Lee, Z.P., Lymburner, L., Masek, J.G., McCorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R.H., Zhu, Z., 2014. Landsat-8: science and product vision for terrestrial global change research. Remote Sens. Environ. 145, 154–172. https://doi.org/10.1016/j.rse.2014.02.001.

Safanelli, J.L., Hengl, T., Parente, L.L., Minarik, R., Bloom, D.E., Todd-Brown, K., Gholizadeh, A., de Mendes, W.S., Sanderman, J., 2025. Open soil spectral library (OSSL): building reproducible soil calibration models through open development and community engagement. PloS One 20, e0296545. https://doi.org/10.1371/journal.pone.0296545.

Schaepman, M.E., Jehle, M., Hueni, A., D'Odorico, P., Damm, A., Weyermann, J., Schneider, F.D., Laurent, V., Popp, C., Seidel, F.C., Lenhard, K., Gege, P., Küchler, C., Brazile, J., Kohler, P., De Vos, L., Meuleman, K., Meynart, R., Schläpfer, D., Kneubühler, M., Itten, K.I., 2015. Advanced radiometry measurements and earth science applications with the airborne prism experiment (APEX). Remote Sens. Environ. 158, 207–219. https://doi.org/10.1016/j.rse.2014.11.014.

Schiefer, F., Schmidtlein, S., Kattenborn, T., 2021. The retrieval of plant functional traits from canopy spectra through RTM-inversions and statistical models are both critically affected by plant phenology. Ecol. Indic. 121, 107062. https://doi.org/10.1016/j.ecolind.2020.107062.

Schläpfer, D., Richter, R., 2002. Geo-atmospheric processing of airborne imaging spectrometry data. Part 1: parametric orthorectification. Int J Remote Sens 23, 2609–2630. https://doi.org/10.1080/01431160110115825.

Schweiger, A.K., 2020. Spectral field campaigns: Planning and data collection. In: Remote Sensing of Plant Biodiversity. Springer International Publishing, Cham, pp. 385–423. https://doi.org/10.1007/978-3-030-33157-3_15.

Schweiger, A.K., Risch, A.C., Damm, A., Kneubühler, M., Haller, R., Schaepman, M.E., Schütz, M., 2015a. Using imaging spectroscopy to predict above-ground plant biomass in alpine grasslands grazed by large ungulates. J. Veg. Sci. 26, 175–190. https://doi.org/10.1111/jvs.12214.

Schweiger, A.K., Schütz, M., Anderwald, P., Schaepman, M.E., Kneubühler, M., Haller, R., Risch, A.C., 2015b. Foraging ecology of three sympatric ungulate species – Behavioural and resource maps indicate differences between chamois, ibex and red deer. Mov Ecol 3, 6. https://doi.org/10.1186/s40462-015-0033-x.

Schweiger, A.K., Schütz, M., Risch, A.C., Kneubühler, M., Haller, R., Schaepman, M.E., 2017. How to predict plant functional types using imaging spectroscopy: linking vegetation community traits, plant functional types and spectral response. Methods Ecol Evol 8, 86–95. https://doi.org/10.1111/2041-210X.12642.

Schwieder, M., Buddeberg, M., Kowalski, K., Pfoch, K., Bartsch, J., Bach, H., Pickert, J., Hostert, P., 2020. Estimating grassland parameters from Sentinel-2: a model comparison study. PFG – journal of photogrammetry. Remote Sensing and Geoinformation Science 88, 379–390. https://doi.org/10.1007/s41064-020-00120-1.

scikit-learn Developers, 2023. ConstantKernel [WWW Document]. In: https://scikit-learn.org/stable/modules/generated/sklearn.gaussian_process.kernels.ConstantKernel.html (accessed 10.12.23).

Sherrill, C.W., 2019. Analyzing Sericea lespedeza (*lespedeza cuneata*) Management Practices and the Importance of Forbs in the Diet of Cattle and bison on Tallgrass Prairie (Master's Thesis). Oklahoma State University, Stillwater, OK.

Shoko, C., Mutanga, O., Dube, T., 2016. Progress in the remote sensing of C3 and C4 grass species aboveground biomass over time and space. ISPRS Journal of Photogrammetry and Remote Sensing 120, 13–24. https://doi.org/10.1016/J.ISPRSJPRS.2016.08.001.

Socher, S.A., Prati, D., Boch, S., Müller, J., Klaus, V.H., Hölzel, N., Fischer, M., 2012. Direct and productivity-mediated indirect effects of fertilization, mowing and grazing on grassland species richness. J. Ecol. 100, 1391–1399. https://doi.org/10.1111/j.1365-2745.2012.02020.x.

Stevens, A., Ramirez-Lopez, L., 2022. An Introduction to the Prospectr Package [WWW document]. URL: https://CRAN.R-project.org/package=prospectr.

Tagliabue, G., Boschetti, M., Bramati, G., Candiani, G., Colombo, R., Nutini, F., Pompilio, L., Rivera-Caicedo, J.P., Rossi, M., Rossini, M., Verrelst, J., Panigada, C., 2022. Hybrid retrieval of crop traits from multi-temporal PRISMA hyperspectral imagery. ISPRS J. Photogramm. Remote Sens. 187, 362–377. https://doi.org/10.1016/j.isprsjprs.2022.03.014.

The Nature Conservancy, 2023. Joseph H. Williams Tallgrass Prairie Preserve [WWW Document]. https://www.nature.org/en-us/get-involved/how-to-help/places-we-protect/tallgrass-prairie-preserve/ (accessed 3.29.23).

Van Vooren, L., Reubens, B., Broekx, S., Reheul, D., Verheyen, K., 2018. Assessing the impact of grassland management extensification in temperate areas on multiple ecosystem services and biodiversity. Agric. Ecosyst. Environ. 267, 201–212. https://doi.org/10.1016/j.agee.2018.08.016.

Vapnik, V., Golowich, S.E., Smola, A., 1997. Support vector method for function approximation, regression estimation, and signal processing. Adv Neural Inf Process Syst 281–287.

Verhoef, W., Jia, L., Xiao, Q., Su, Z., 2007. Unified optical-thermal four-stream radiative transfer theory for homogeneous vegetation canopies. IEEE Trans. Geosci. Remote Sens. 45, 1808–1822. https://doi.org/10.1109/TGRS.2007.895844.

Verrelst, J., Muñoz, J., Alonso, L., Delegido, J., Rivera, J.P., Camps-Valls, G., Moreno, J., 2012. Machine learning regression algorithms for biophysical parameter retrieval: opportunities for Sentinel-2 and -3. Remote Sens. Environ. 118, 127–139. https://doi.org/10.1016/j.rse.2011.11.002.

Verrelst, J., Alonso, L., Rivera Caicedo, J.P., Moreno, J., Camps-Valls, G., 2013a. Gaussian process retrieval of chlorophyll content from imaging spectroscopy data. IEEE J Sel Top Appl Earth Obs Remote Sens 6, 867–874. https://doi.org/10.1109/JSTARS.2012.2222356.

Verrelst, J., Rivera, J.P., Moreno, J., Camps-Valls, G., 2013b. Gaussian processes uncertainty estimates in experimental Sentinel-2 LAI and leaf chlorophyll content retrieval. ISPRS Journal of Photogrammetry and Remote Sensing 86, 157–167. https://doi.org/10.1016/j.isprsjprs.2013.09.012.

Verrelst, J., Rivera, J.P., Leonenko, G., Alonso, L., Moreno, J., 2014. Optimizing LUT-based RTM inversion for semiautomatic mapping of crop biophysical parameters from Sentinel-2 and -3 data: role of cost functions. IEEE Trans. Geosci. Remote Sens. 52, 257–269. https://doi.org/10.1109/TGRS.2013.2238242.

Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J.P., Veroustraete, F., Clevers, J.G. P.W., Moreno, J., 2015. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – a review. ISPRS Journal of Photogrammetry and Remote Sensing 108, 273–290. https://doi.org/10.1016/j.isprsjprs.2015.05.005.

Verrelst, J., Dethier, S., Rivera, J.P., Munoz-Mari, J., Camps-Valls, G., Moreno, J., 2016. Active learning methods for efficient hybrid biophysical variable retrieval. IEEE Geosci. Remote Sens. Lett. 13, 1012–1016. https://doi.org/10.1109/LGRS.2016.2560799.

Verrelst, J., Rivera Caicedo, J., Muñoz-Marí, J., Camps-Valls, G., Moreno, J., 2017. SCOPE-based emulators for fast generation of synthetic canopy reflectance and Sun-induced fluorescence spectra. Remote Sens. 9, 927. https://doi.org/10.3390/rs9090927.

Verrelst, J., Rivera-Caicedo, J.P., Reyes-Muñoz, P., Morata, M., Amin, E., Tagliabue, G., Panigada, C., Hank, T., Berger, K., 2021. Mapping landscape canopy nitrogen content from space using PRISMA data. ISPRS Journal of Photogrammetry and Remote Sensing 178, 382–395. https://doi.org/10.1016/j.isprsjprs.2021.06.017.

Verrelst, J., Halabuk, A., Atzberger, C., Hank, T., Steinhauser, S., Berger, K., 2023. A comprehensive survey on quantifying non-photosynthetic vegetation cover and biomass from imaging spectroscopy. Ecol. Indic. 155, 110911. https://doi.org/10.1016/j.ecolind.2023.110911.

Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Brown, D.J., Demattê, J.A.M., Shepherd, K.D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aïchi, H., Barthès, B.G., Bartholomeus, H.M., Bayer, A.D., Bernoux, M., Böttcher, K., Brodský, L., Du, C.W., Chappell, A., Fouad, Y., Genot, V., Gomez, C., Grunwald, S., Gubler, A., Guerrero, C., Hedley, C.B., Knadel, M., Morrás, H.J.M., Nocita, M., Ramirez-Lopez, L., Roudier, P., Campos, E.M.R., Sanborn, P., Sellitto, V.M., Sudduth, K.A., Rawlins, B.G., Walter, C., Winowiecki, L.A., Hong, S.Y., Ji, W., 2016. A global spectral library to characterize the world's soil. Earth Sci. Rev. 155, 198–230. https://doi.org/10.1016/j.earscirev.2016.01.012.

Wang, G., Liu, S., Liu, T., Fu, Z., Yu, J., Xue, B., 2019. Modelling above-ground biomass based on vegetation indexes: a modified approach for biomass estimation in semi-arid grasslands. Int J Remote Sens 40, 3835–3854. https://doi.org/10.1080/01431161.2018.1553319.

Wang, Z., Féret, J.-B., Liu, N., Sun, Z., Yang, L., Geng, S., Zhang, H., Chlus, A., Kruger, E. L., Townsend, P.A., 2023. Generality of leaf spectroscopic models for predicting key foliar functional traits across continents: a comparison between physically- and empirically-based approaches. Remote Sens. Environ. 293, 113614. https://doi.org/10.1016/j.rse.2023.113614.

Wellstein, C., Chelli, S., Campetella, G., Bartha, S., Galiè, M., Spada, F., Canullo, R., 2013. Intraspecific phenotypic variability of plant functional traits in contrasting mountain grasslands habitats. Biodivers. Conserv. 22, 2353–2374. https://doi.org/10.1007/s10531-013-0484-6.

White, R., Murray, S., Rohweder, M., 2000. Pilot Analysis of Global Ecosystems: Grassland Ecosystems. WRI Publications, Hapden Station, Baltimore, MD.

Wocher, M., Berger, K., Verrelst, J., Hank, T., 2022. Retrieval of carbon content and biomass from hyperspectral imagery over cultivated areas. ISPRS Journal of Photogrammetry and Remote Sensing 193, 104–114. https://doi.org/10.1016/j.isprsjprs.2022.09.003.

Woodcock, C.E., 2002. Uncertainty in remote sensing. In: Foody, G.M., Atkinson, P.M. (Eds.), Uncertainty in Remote Sensing and GIS. Wiley, Chichester, pp. 19–24. https://doi.org/10.1002/0470035269.ch2.

Xu, D., Guo, X., Li, Z., Yang, X., Yin, H., 2014. Measuring the dead component of mixed grassland with Landsat imagery. Remote Sens. Environ. 142, 33–43. https://doi.org/10.1016/j.rse.2013.11.017.

Yuan, H., Ma, R., Atzberger, C., Li, F., Loiselle, S., Luo, J., 2015. Estimating Forest fAPAR from multispectral Landsat-8 data using the invertible forest reflectance model INFORM. Remote Sens. 7, 7425–7446. https://doi.org/10.3390/rs70607425.

Yuhas, R., Goetz, A., Boardman, J., 1992. Discrimination among semi-arid landscape endmembers using the spectral angle mapper (SAM) algorithm. In: Summaries of the Third Annual JPL Airborne Geoscience Workshop. Volume 1: AVIRIS Workshop. JPL, pp. 147–149.

Zeng, Y., Hao, D., Park, T., Zhu, P., Huete, A., Myneni, R., Knyazikhin, Y., Qi, J., Nemani, R.R., Li, F., Huang, J., Gao, Y., Li, B., Ji, F., Köhler, P., Frankenberg, C., Berry, J.A., Chen, M., 2023. Structural complexity biases vegetation greenness measures. Nat Ecol Evol 7, 1790–1798. https://doi.org/10.1038/s41559-023-02187-6.

Zhang, F., Bennett, J.A., Zhang, B., Zhao, T., Bai, K., Zhao, M., Han, G., 2023. Responses of grassland productivity to mowing intensity and precipitation variability in a temperate steppe. Oecologia 201, 259–268. https://doi.org/10.1007/s00442-022-05305-6.

Zhao, Y., Liu, Z., Wu, J., 2020. Grassland ecosystem services: a systematic review of research advances and future directions. Landsc. Ecol. 35, 793–814. https://doi.org/10.1007/s10980-020-00980-3.

Zhou, K., Liu, Z., Qiao, Y., Xiang, T., Loy, C.C., 2022. Domain generalization: a survey. IEEE Trans. Pattern Anal. Mach. Intell. 1–20. https://doi.org/10.1109/TPAMI.2022.3195549.

Zupanc, A., 2017. Improving Cloud Detection with Machine Learning. https://medium.com/sentinel-hub/improving-cloud-detection-with-machine-learning-c09dc5d7cf13 (accessed 9.28.21).