Remote Sensing of Environment 280 (2022) 113199

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Quantifying uncertainty in high resolution biophysical variable retrieval with machine learning

Laura Martínez-Ferrer [a,*], Álvaro Moreno-Martínez [a], Manuel Campos-Taberner [b], Francisco Javier García-Haro [b], Jordi Muñoz-Marí [a], Steven W. Running [c], John Kimball [c], Nicholas Clinton [d], Gustau Camps-Valls [a]

[a] Image Processing Laboratory (IPL), Universitat de València, València, Spain
[b] Environmental Remote Sensing Group, Universitat de València, València, Spain
[c] Numerical Terradynamic Simulation Group (NTSG), University of Montana, USA
[d] Google, Inc., Mountain View, CA, USA

ARTICLE INFO

Edited by Jing M. Chen

Keywords:
MODIS
Landsat
Downscaling
Biophysical parameter estimation
Uncertainty
Neural networks

ABSTRACT

The estimation of biophysical variables is at the core of remote sensing science, allowing a close monitoring of crops and forests. Deriving temporally resolved and spatially explicit maps of parameters of interest has been the subject of intense research. However, deriving products from optical sensors is typically hampered by cloud contamination and the trade-off between spatial and temporal resolutions. In this work we rely on the HIghly Scalable Temporal Adaptive Reflectance Fusion Model (HISTARFM) algorithm to generate long gap-free time series of Landsat surface reflectance data by fusing MODIS and Landsat reflectances. An artificial neural network is trained on PROSAIL inversion to predict monthly biophysical variables at 30 m spatial resolution with associated, realistic uncertainty bars. We emphasize the need for a more thorough analysis of uncertainty, and propose a general and scalable approach to combine both epistemic and aleatoric uncertainties by exploiting Monte Carlo (MC) dropout techniques from the trained artificial network and the propagation of HISTARFM uncertainties through the model, respectively. A model recalibration was performed in order to provide reliable uncertainties. We provide new high resolution products of several key variables to quantify the terrestrial biosphere: Leaf Area Index (LAI), Fraction of Absorbed Photosynthetically Active Radiation (FAPAR), Canopy Water Content (CWC) and Fractional Vegetation Cover (FVC) are at 30 m Landsat spatial resolution and over large continental areas. Two study areas are considered: the large heterogeneous but moderately cloud covered contiguous United States, and the homogeneous but largely cloud covered Amazonia. The produced vegetation products largely agree with the test dataset ($R = 0.90$, RMSE $= 0.80\ m^2/m^2$ and ME $= 0.12\ m^2/m^2$ for LAI, and $R = 0.98$, RMSE $= 0.07$ and ME $= 0.01$ for FAPAR) providing low error and high accuracy. Additionally, the validation considers a thorough comparison with operational and largely validated medium resolution products, such as the Moderate-Resolution Imaging Spectroradiometer (MODIS) and Copernicus Global Land Service. Our products presented a good agreement and consistency with both MODIS ($R = 0.84$ and $R = 0.85$ for LAI and FAPAR, respectively) and Copernicus ($R = 0.92$ and $R = 0.91$ for LAI and FAPAR, respectively). To foster a wider adoption and reproducibility of the methodology we provide an application in GEE and source code at:https://github.com/IPL-UV/ee_BioNet/

1. Introduction

Quantifying vegetation biochemistry, structure and functioning globally is key to study and understand global change, biodiversity and agriculture. Remote sensing science serves these purposes by deriving products with estimates of biophysical variables from satellite sensors and models. The derived products may constitute an excellent timely tool for monitoring of crops and forests in space and time. Deriving maps of biophysical variables of interest in terms of sufficient spatial and temporal resolution has been the subject of intense research and a

* Corresponding author.
E-mail address: laura.martinez-ferrer@uv.es (L. Martínez-Ferrer).

https://doi.org/10.1016/j.rse.2022.113199
Received 7 September 2021; Received in revised form 22 July 2022; Accepted 25 July 2022
Available online 9 August 2022

central problem in Earth monitoring of the terrestrial biosphere (Lillesand et al., 2008; Liang, 2004, 2008; Camps-Valls et al., 2011; Verger et al., 2011).

The Global Climate Observing System (GCOS) identified 50 Essential Climate Variables (ECVs) (GCOS, 2011; Spence and Townshend, 1996) considered to be feasible for global climate observation. Among these ECVs, the Leaf Area Index (LAI) and the Fraction of Absorbed Photosynthetically Active Radiation (FAPAR) are the most widely used biophysical vegetation variables to study the land surface, being routinely estimated at global and regional scales using remote sensing data. LAI is a quantitative measure of the amount of live green leaf material present in the canopy, it is defined as half the total area of green elements per unit horizontal ground area (Chen and Black, 1992) and accounts for the amount of green vegetation that absorbs or scatters solar radiation. FAPAR accounts for the fraction of solar radiation absorbed by plants, it constitutes an indicator of the vegetation health status and thereby its primary productivity (Asner, 1998; GCOS, 2011). In addition, there are other related variables of interest as the Fraction of Vegetation Cover (FVC), which can be applied in cropland areas (Verger et al., 2009; Sun et al., 2021), and the Canopy Water Content (CWC), used for measuring canopy status in droughts (Martin et al., 2018). FVC represents the green vegetation fraction that covers a unit area of horizontal soil, and can be estimated from the gap fraction at nadir view (Bonham, 2013). FVC does not depend on variables such as the geometry of illumination, thus being a good alternative to vegetation indices for monitoring Earth's green vegetation (Baret et al., 2013; García-Haro et al., 2018). Canopy status, water stress and health can be monitored by estimating the amount of water in the leaves per unit ground area which is provided by the CWC. CWC is usually computed as the product of leaf water content C_w and LAI (García-Haro et al., 2020; Clevers et al., 2010; Knyazikhin et al., 1998b). Closely monitoring all these four variables serves to characterize the structure and functioning of vegetation, which are key inputs for a broad variety of biosphere and land applications, from climate, forestry and agriculture, to environmental and natural hazards management (Peng et al., 2019; Haboudane et al., 2004).

The literature of variable estimation in remote sensing is vast, and can be categorized under different approaches and methodologies. While sample estimators exploit a sample to produce the estimates (see e.g. Weiss and Baret (2016)), point estimators use one single measurement and a set to produce such estimates (see e.g. Clerici et al. (2010)). Statistical and machine learning methods, both parametric and nonparametric methods have been vastly used (Verrelst et al., 2015). Recent methods like the kernel distribution regression (Adsuara et al., 2019) and invertible neural networks (Ardizzone et al., 2019) are nonparametric models that can work as both sample and point estimators. The parametric regression methods aim at explicitly parameterizing the relationship between spectral bands and the biophysical parameters (Glenn et al., 2008). This is mainly performed using vegetation indices regressed with the biophysical parameters using a regression function. The field of vegetation indices is very active and has recently developed unifying frameworks to deal with nonlinear relations (Camps-Valls et al., 2021a). The non-parametric methods make use of machine learning regression techniques to establish a mapping function from input variables represented in the spectral bands and the output biophysical parameters through a training phase (Verrelst et al., 2012). The third and most traditional approach to parameter estimation is to invert a radiative transfer model (RTM) (Knyazikhin et al., 1998a), but this has shown to be a complex ill-posed inverse problem. The RTM inversion has found application in local campaigns, but they typically require site-specific information (e.g., leaf and canopy information) for proper model parameterization (Campos-Taberner et al., 2016), which is not always available.

The field of variable estimation has considered the combined use of model simulations that are inverted with machine learning regression algorithms (Camps-Valls et al., 2011). This approach reports flexibility and computational efficiency. Within such a scheme, one uses physically constrained calibration data for training a machine learning (ML) regression method. This approach has been widely adopted recently to estimate biophysical parameters with all kinds of machine learning algorithms, such as Gaussian processes (Camps-Valls et al., 2016; Verrelst et al., 2015, 2012), random forests (Campos-Taberner et al., 2018), and neural networks (Baret et al., 2007). In this paper, we propose machine learning parameter estimation approach based on RTM inversion with neural networks.

Deriving global products fulfilling GCOS requirements for spatial and temporal resolution (daily products at 50 m for adaptation studies and 200-500 m for modelling studies) from optical sensors faces two important challenges: (1) the important trade-off between spatial and temporal resolution, often calls for fusion of multisensor data to exploit multimodalities and improve resolution; and (2) variable estimation is severely hampered by cloud contamination, which leads to scenarios with limited observations. To address both issues, in this paper we will rely on the HIghly Scalable Temporal Adaptive Reflectance Fusion Model (HISTARFM) algorithm, which combines multispectral images of different sensors (e.g. MODIS and Landsat) to reduce noise and produce monthly gap free high resolution (30 m) reflectance data to be used for parameter retrieval Moreno-Martínez et al., 2020). The algorithm is based on an unbiased Kalman filter, it is implemented in the Google Earth Engine (GEE) cloud computing platform (Gorelick et al., 2017), and yields reflectance data and associated uncertainty estimates.

Having access to input uncertainties is a key aspect of our proposal, and addresses an important limitation with most of the state of the art remote sensing data products, especially at high spatial resolution: quantifying the uncertainty of the estimates. We propose an holistic treatment of uncertainty, focusing both on *epistemic uncertainty* (model uncertainty) which refers to the confidence a model has about its predictions and is related to the choice of parameters and *aleatoric uncertainty* (data uncertainty) that stems from noise in the input data, priors or errors in RTM and is a consequence of unaccounted factors that introduce variability in the inputs or targets (Caers, 2011). While *epistemic uncertainty* has been usually computed as an output of RTM inversion approaches (Verrelst et al., 2012; Camps-Valls et al., 2016; García-Haro et al., 2018), *aleatoric uncertainty* is rarely estimated in these models or empirical procedures are used to do so (Knyazikhin et al., 1998b). In this paper, we capitalize on a more rigorous approach for the uncertainty propagation of the inputs to capture the influence of the combined uncertainties of the input bands used for the estimation of biophysical variables and produce realistic uncertainty estimates. Having access to high resolution reflectance data with associated data uncertainties is crucial in this context, as it will allow us to propagate them through our neural network retrieval model.

Neural networks have shown an excellent trade-off between epistemic uncertainty and efficiency at the product (test) time, and have outperformed traditional models in Earth environmental monitoring (Reichstein et al., 2019; Camps-Valls et al., 2021b). However, standard neural networks cannot deal with uncertainty quantification. A full Bayesian treatment of uncertainty for model inversion is computationally very costly and often impractical in real applications, but still useful for validation processes. We propose to use *dropout* for neural networks, which prevents overfitting during the training phase (Srivastava et al., 2014) and, when applied in the test phase, is called *Monte Carlo (MC) dropout*. This approach can be interpreted, under certain assumptions, as a Bayesian approximation to quantify model uncertainty (Gal and Ghahramani, 2016).

In this work we argue and hypothesize that the problem of high-resolution long-term and gap-free parameter estimation can be resolved with new machine learning models that fuse multisource data to alleviate the gaps in time series, operate in a cloud computing platform to provide products at high spatial resolution, and can estimate and propagate uncertainties. Our goal is thus to build a processing chain for efficient derivation of high spatial resolution products of several key variables to quantify the terrestrial biosphere (LAI, FAPAR, FVC and

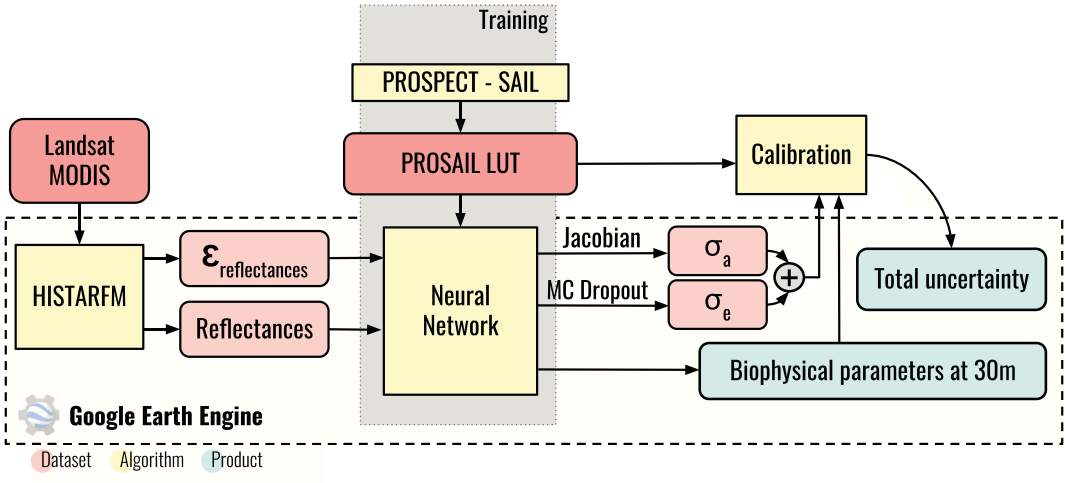

Fig. 1. Processing chain proposed in this paper. We exploit high-resolution cloud-free data derived from the HISTARFM. The reflectances are used for the neural network to produce high-resolution (30 m) estimates of biophysical parameters (LAI, FAPAR, FVC, CWC). The neural network is trained inverting PROSAIL. The HISTARFM also provides high-resolution reflectance uncertainties, which are propagated through the neural network using the network Jacobians, and yield the aleatoric uncertainty σ_a. Applying a MC dropout strategy to the network we derive the epistemic uncertainty σ_e. Both uncertainties are summed up and, together with the ground truth and variable estimates, are the inputs for a model calibration step that will provide a recalibrated total uncertainty estimates. The methodology yield both the total high-resolution uncertainty maps accompanying the high-resolution biophysical parameter products.

CWC) at 30 m Landsat scale and over large continental areas. We used neural networks as nonparametric regression models using physically-based calibration data from the PROSAIL RTM (Berger et al., 2018) model runs. Special attention is given to the quantification of the components of uncertainty (both epistemic and aleatoric) of the products. We illustrate resulting monthly maps in large complicated areas with spatial heterogeneity and moderate-to-high cloud cover, like the Contiguous US (CONUS) and Amazonia. Biophysical variables estimates are compared with other operational products (e.g. MODIS and Copernicus) (Knyazikhin, 1999; Baret et al., 2013) in the representative BELMANIP sites database (Baret et al., 2006) and a model calibration analysis is performed in order to provide reliable uncertainty estimates.

2. Material and methods

The proposed methodology is based on a simulated reflectance dataset from the PROSAIL model. A non-linear regression predictor is built between the vegetation variables of interest and simulated reflectances in Landsat wavelengths. Training a neural network with dropout to invert the PROSAIL database and then using gap-free data at Landsat resolution derived from the HISTARFM algorithm, allows us to obtain estimates at 30 m of all considered variables, and the associated data and model uncertainty maps applying MC dropout and uncertainty propagation. Working with petabytes of datasets ($8x10^9$ pixels per image) is a challenging task and has storing, accessing and computing requirements that can be managed by using High Performance Computing (HPC). Fig. 1 shows a schematic of the approach, which is further detailed in the following sections.

2.1. High-resolution gap-filled data

We begin with the use of high-resolution gap-filled surface reflectance data derived from the fusion of Landsat and MODIS with the HISTARFM algorithm presented in Moreno-Martínez et al. (2020). HISTARFM provides expected value and one standard deviation prediction intervals of reflectance over land for bands required as input to the network predictor (Moreno-Martínez et al., 2020). This approach allows to combine two estimators operating synergistically to filter out random noise and reduce the bias of Landsat spectral reflectances. The first estimator is an optimal interpolator that produces estimates of Landsat reflectance values for a given time by combining a Landsat climatology, pre-computed from the available Landsat record, and a fusion of MODIS and Landsat reflectances obtained from the respective satellite overpasses closest to the time of interest. The second estimator is a Kalman filter that corrects the bias of the reflectance produced by the first estimator. HISTARFM was implemented and designed to work on a per-pixel basis to allow for scalability in cloud computing environments such as GEE.

The direct validation of HISTARFM at a continental scale over a thousand locations representing major vegetation types over the contiguous US proved the reliability of the method, with relative mean errors (rME) below 1.5% for all bands and with relative mean absolute errors (rMAE) and relative root mean squared errors (rRMSE) being low or moderate (Moreno-Martínez et al., 2020). In addition, HISTARFM also provides uncertainty estimates along with the predicted reflectance as ancillary data. These predicted uncertainties were also validated, and showed high consistency with the errors over the validation data set. The validation of these spatially and temporally explicit realistic uncertainties is especially relevant for the present study, as it provides a solid foundation for the reliable use of HISTARFM in error propagation tasks. Adding high quality uncertainty estimates in our model inputs will help to provide more realistic uncertainty estimates at the end of the proposed processing stream.

2.2. Inversion of PROSAIL

We used neural networks for inverting the PROSAIL radiative transfer model. PROSAIL couples two RTMS: PROSPECT (Feret et al., 2008) and SAIL (Verhoef, 1984). PROSAIL simulates the scene bidirectional reflectance taking into account the leaf biophysical properties using PROSPECT, and the canopy architecture, soil background, and hot spot using SAIL, for a given set of illumination and sensor (observation) geometry conditions. In this study we used the PROSPECT-5B, and 4SAILH versions (see Appendix A for further details).

We created a database of spectral simulations representative of a wide range of vegetation conditions. The simulations were obtained after the PROSAIL execution in forward mode, which requires the set up of the following input parameters:

1) Within PROSPECT: A set of leaf optical properties, given by the mesophyll structural parameter (N), leaf chlorophyll (Chl), dry matter (Cm), water (Cw), carotenoid (Car) and brown pigment (Cbr) contents. It is worth mentioning that we used the PROSPECT-5B model, which allows to deal with Chl and Car contents separately.
2) Within SAIL: A set of canopy level and geometry characteristics, determined by leaf area index (LAI), the average leaf angle inclination (ALA), the hot-spot parameter (Hotspot), the solar zenith angle (θ_s), view zenith angle (θ_v), and the relative azimuth angle between both angles ($\Delta\Theta$).

We considered PROSAIL for simulating Landsat-5 spectra for the sake of consistency with the high-resolution data coming from HISTARFM which is subsequently ingested for production and uncertainty propagation. Reflectances were simulated for each wavelength and filtered in accordance with the spectral response of the Landsat-5 channels (see Fig. A.1 in Appendix A). Hence, the resulting data base mimics six Landsat-5 Thematic Mapper (TM) spectral bands with wavelengths ranging from 0.45μm to 2.35μm (blue, green, red, NIR, SWIR-1, and SWIR-2 channels) and took into account the orbit characteristics of Landsat to simulate satellite angular sampling.

The parameter distributions were obtained from previous studies (García-Haro et al., 2018, 2020; Campos-Taberner et al., 2018) to be representative of realistic cases. In particular, the leaf parameters were set up by exploiting the information of a global scale database of leaf properties (TRY) (Kattge et al., 2020) to optimize our prior information on RTM modelling to better constrain the retrievals (see Table A.1 in Appendix A) (Combal et al., 2003). We have adopted a random Latin hypercube sampling design allowed to populate more evenly the canopy realization space (Mckay et al., 2000). Altogether 3303 cases of surface reflectances and associated biophysical variables were simulated.

2.3. Neural networks for biophysical parameter retrieval

Artificial neural networks are powerful nonlinear and non-parametric regression methods and widely used in Earth sciences, from remote sensing to general-purpose geoscientific problems (Goodfellow et al., 2016; Reichstein et al., 2019; Camps-Valls et al., 2021b). Their application for biophysical parameter retrieval has been widespread and has shown uncertainties comparable with existing approaches (Gong, 1999; Fang and Liang, 2003; Danson et al., 2003; Walthall et al., 2004; Combal et al., 2003). Neural networks are well suited for model inversion and parameter estimation as they are able to model complex relationships between land surface reflectance and the biophysical parameters, which are known to be highly non-linear by nature.

Feedforward neural networks are formed by a hierarchical structure of basic operation units (or neurons) distributed in layers (Bishop et al., 1995). Neural networks, once trained, can be readily applied to test data and produce output predictions quickly. They are easy to implement, scale and run in high-performance platforms like GEE, where only a set of weights are stored and prediction only involves some matrix multiplications, summations and data transformations (such as non-linear functions). A network with a sufficient number of hidden layers and units is capable of approximating any bounded, continuous function to an arbitrary level of precision over a finite domain so they are regarded as universal function approximations (Bishop et al., 1995). Current training algorithms and regularization strategies allow us for complexity control and avoid overfitting (Srivastava et al., 2014; Montavon et al., 2012). In addition, as we will see later, neural networks allow for efficient uncertainty propagation and, thanks to recent developments in machine learning, account for model uncertainty too (Gal and Ghahramani, 2016).

Let us fix the notation, where for simplicity we assume a shallow, one hidden layer network. The n input features $\mathbf{x} \in \mathbb{R}^n$ to a neuron are weighted times $\mathbf{w} \in \mathbb{R}^n$, then translated with a bias term b, and finally passed through an activation (or link) nonlinear function f:

$$y = g\left(\sum_{j=1}^{m} v_j f\left(\sum_{i=1}^{n} w_{ij}x^i + b_j\right) + c\right), \tag{1}$$

where the input feature vector is $\mathbf{x} = [x^1, x^2, \ldots, x^n]$ and x^i is the ith input feature, n is the number of variables/features, w_{ij} is the weight connecting the ith input with the jth node in the hidden layer, b_i is the bias term of the ith node, m is the number of nodes in the hidden layer, $f(\cdot)$ is the transfer function of the neurons in the hidden layer, v_i are the weights between the ith hidden node and the output node, c is the bias of the output node, and $g(\cdot)$ is the transfer function of the outputs node. Typically the transfer functions f are sigmoid functions, $f(z) = \tanh(z)$, while the output link g is chosen to be a linear link, $g(z) = z$ in regression settings.

Training the network weights and biases can be done with many different algorithms. We used the Adam algorithm which is a very efficient method, providing state-of-the-art performance (Kingma and Ba, 2014). The network architecture was selected by cross-validation and finally yielded a rather simple network with $m = 4$ neurons and one single hidden layer, which worked well for all variables. The network was trained with early stopping on cross-validation: 85% of the data for training ($N = 2807$) and 15% for validation ($N = 496$). We used the least squares error as a loss function, a batch size of 64 and 1000 epochs. Dropout was used during training (Srivastava et al., 2014); it is customary to use a dropout factor of 0.2, but we fixed it to $\gamma = 0.01$ due to the small number of nodes in the network as proposed in Piotrowski et al. (2020). In order to quantify model uncertainty, MC dropout is applied in the test phase (Gal and Ghahramani, 2016). We sampled binary variables for every input example and for every network unit in each layer. Each binary variable takes values 1 with probability $1 - \gamma$. Then a unit is dropped for a given input if its corresponding binary variable takes value 0 drawn from a Bernoulli distribution. Due to the requirement for reproducibility in GEE, the prediction module is composed of the most probable models, that in this case were five models; the ones where one neuron is dropped (four models) plus the one where all neurons are active.[1] The neural network weights obtained in this process for all biophysical parameter models were transferred to the GEE platform, where the prediction module was implemented therein for production. We also computed the differences in the estimates between the full model and the prediction module finally considered in this work, concluding that the differences were negligible (<1%) while computation efficiency was greatly improved. Therefore, the final prediction for point $\mathbf{x}_*$ is the weighted mean of the predictions $y^i_* = f(\mathbf{x}_*; \theta^i)$ of the $i = 1, \ldots, M$ submodels parameterized with model weights $\theta := \{\mathbf{w}, \mathbf{v}, \mathbf{b}, \mathbf{c}\}$ obtained for each case of dropping neurons, and multiplied by the probability p_i that this model takes place, $\mu_* = \sum_{i=1}^{M} y^i_* p_i$.

2.4. Uncertainty quantification and propagation

Understanding what a model does not know is a critical part of many machine learning systems, and has strong implications in real-life applications. Good uncertainty estimates quantify when one can trust a model's predictions. There are two main types of uncertainty: epistemic uncertainty and aleatoric uncertainty (Pearce et al., 2020; Caers, 2011):

- *Epistemic uncertainty*. We represent this uncertainty by placing a distribution over the neural network weights $\boldsymbol{\theta}$. This distribution depends on the training dataset, $\mathscr{D} := \{(\mathbf{x}_i, y_i) | i = 1, \ldots, N\}$.

[1] The probabilities of other cases occurring (two dropped neurons, three dropped neurons, all neurons dropped and their respective commutations) were not considered as the probability of these cases was lower than 10^{-4}.

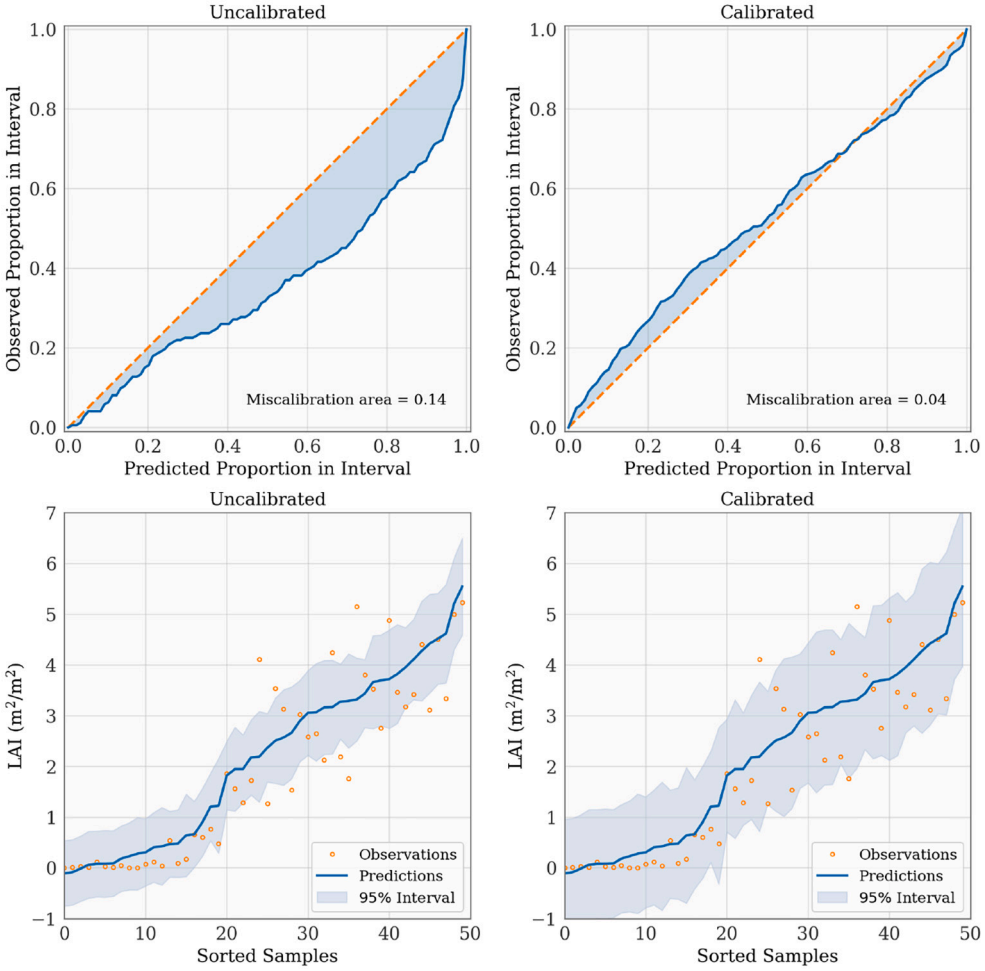

Fig. 2. LAI calibration curves (top) and prediction with uncertainties (bottom) using uncalibrated (left) and calibrated model (right). Only 50 observations form the calibration independent set are represented for illustrative purposes.

Table 1
Metrics obtained in calibration: mean absolute calibration error (MACE[m^2/m^2]), root mean squared calibration error (RMSCE[m^2/m^2]) and miscalibration area (MA).

	MACE	RMSCE	MA
Uncalibrated	0.14	0.16	0.14
Calibrated	0.04	0.05	0.04

Table 2
Results for BioNet methodology: mean error ME; root-mean-square error RMSE; Pearson's correlation coefficient ρ, Spearman's correlation coefficient ρ_S, mutual information MI, and distance correlation D_{corr} over an independent test set.

Variable	ME	RMSE	ρ	ρ_S	MI	D_{corr}
LAI (m^2/m^2)	−0.12	0.80	0.90	0.92	0.83	0.92
FVC	−0.01	0.07	0.98	0.97	1.73	0.98
FAPAR	−0.01	0.07	0.98	0.96	1.64	0.98
CWC (g/cm^2)	0.00	0.02	0.84	0.89	0.62	0.88

Therefore, the weight distribution after training can be written as $p(\boldsymbol{\theta}|\mathscr{D})$, which is intractable except in trivial cases. In order to approximate this distribution we followed a $p \sim \text{Bern}(\boldsymbol{\theta};\gamma)$, where γ is the dropout rate on the weights. Following this assumption, the model (epistemic) uncertainty for prediction μ is given by the variance as follows:

$$\sigma_e = \sum_{i=1}^{M} p_i \left(y_*^i - \mu_*\right)^2, \tag{2}$$

where y_*^i are the predictions for $\mathbf{x}_*$ of the ith submodel obtained in the MC dropout process.

- *Aleatoric uncertainty.* HISTARFM provides not only high resolution reflectances but associated uncertainty estimates as generated by the bias-aware Kalman filter algorithm (Moreno-Martínez et al., 2020). We propagate these data uncertainties through the nonlinear function implemented by the neural networks before. Estimating the propagation of the uncertainties for arbitrarily complex nonlinear functions can be a computationally costly problem. Denoting the arbitrary nonlinear function implemented by a network as $\mathbf{F}$, a distortion in the input space $\Delta\mathbf{x}$ propagates through $\mathbf{F}$ giving rise to a distortion in the output $\mathbf{F}(\mathbf{x}+\Delta\mathbf{x})$, which can be approximated with the Taylor's expansion of the function where $\partial\mathbf{F}/\partial\mathbf{x}_i$ denotes the partial derivative of $\mathbf{F}$ with respect to the ith component, evaluated at the mean values of all components of $\mathbf{x}$. In matrix notation, the truncated linear approximation reduces to $\mathbf{F} \approx \mathbf{F}^0 + \mathbf{J}$ where $\mathbf{J}$ is the Jacobian matrix. Therefore, to propagate the input uncertainties, we used the neural network Jacobian which contains the first derivatives of the network errors with respect to the weights and biases, which is a quite standard approach in remote sensing (Aires et al., 2004). Moreover, Jacobians provide information on the relevance of

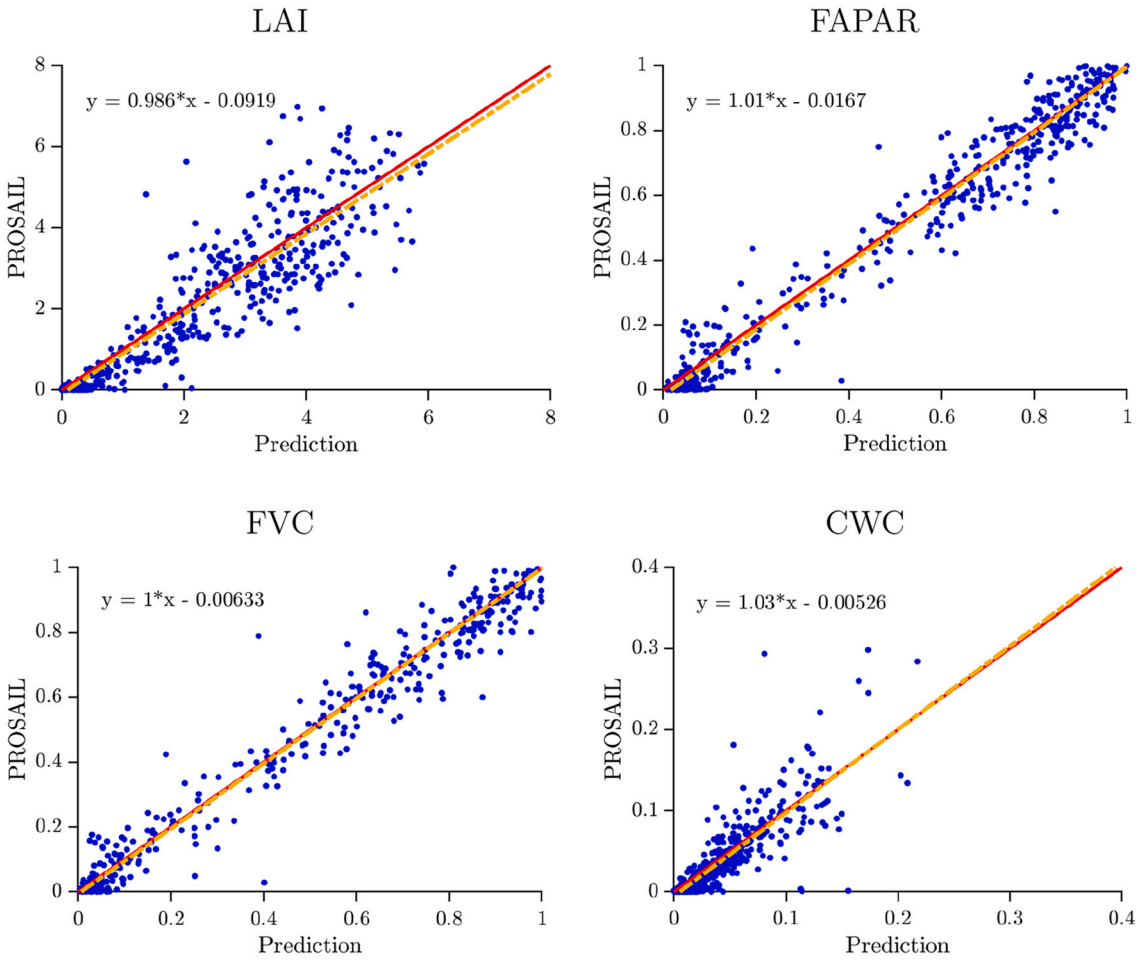

Fig. 3. Scatter plots over an independent test set for all considered variables. The fit line is shown in yellow and the bisector in red. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

network inputs and can therefore be used to select the most significant inputs.

Deriving the output of a network **F** wrt the input features in (1), and using the chain rule, we obtain

$$J_i := \frac{\partial y}{\partial x_i} = \sum_{j=1}^{m} v_{ij}\left(1 - f^2\left(a_j\right)\right) w_{ij}, \tag{3}$$

where $a_j = \sum_i^n w_{ij} x_i + b_j$ and we assumed a liner output layer $g(z) = z$. This expression depends on the neural network weights and biases $\boldsymbol{\theta}$ (which are fixed after training), and allows us to estimate the propagation of the uncertainty through the network. Hence we compute the aleatoric uncertainty as follows:

$$\sigma_a = \sqrt{\sum_{i=1}^{n} J_i^2 e^2}, \tag{4}$$

where J_i are the Jacobians of the network and e are the uncertainties of the inputs, which depend on the band and the time step as proved by the HISTARFM algorithm. The combination of the two types of uncertainties can be done in many ways. However, according (Loquercio et al., 2020), which follows the definition of the expected value and covariance, approximates the posterior over the network weights as a Bernoulli distribution and finally approximates the integral by Monte Carlo methods, the total uncertainty can be generated by summing the two components of data and model uncertainty:

$$\sigma_T = \sigma_e + \sigma_a. \tag{5}$$

In our experiments we report both uncertainty components as well as the total uncertainty.

2.5. Model calibration

Assessing the reliability of uncertainty estimates implies the quantification of how well the derived predictive confidence represents the actual probability. Therefore, in order to make use of uncertainty quantification methods, one has to be sure that the network is well calibrated (Guo et al., 2017). Calibration curves and calibration metrics permit to assess the calibration analysis by determining the agreement between predictions and observations in different percentiles of the predicted values. Fig. 2 shows the calibration curves and the prediction plots for the output LAI. It can be seen that when the model is uncalibrated we are reporting too little uncertainty so that the predictor is overconfident. In order to calibrate it, we computed a scaling factor which uniformly recalibrates predicted standard deviations by minimizing (see Coleman and Li (1996); Dennis (1977)) the metric of interest that in this case was mean absolute calibration error (see Fig. 2 bottom). The method searches the standard deviation scale factor (τ) which produces the best recalibration, and the updated standard deviation is simply rewritten as $\tau^*\sigma_T$. Table 1 shows how all the metric values decrease when a recalibration is performed. Calibration analysis for the rest of biophysical variables can be found in Appendix B. Even though FAPAR model does not need recalibration, we decided to sistematically

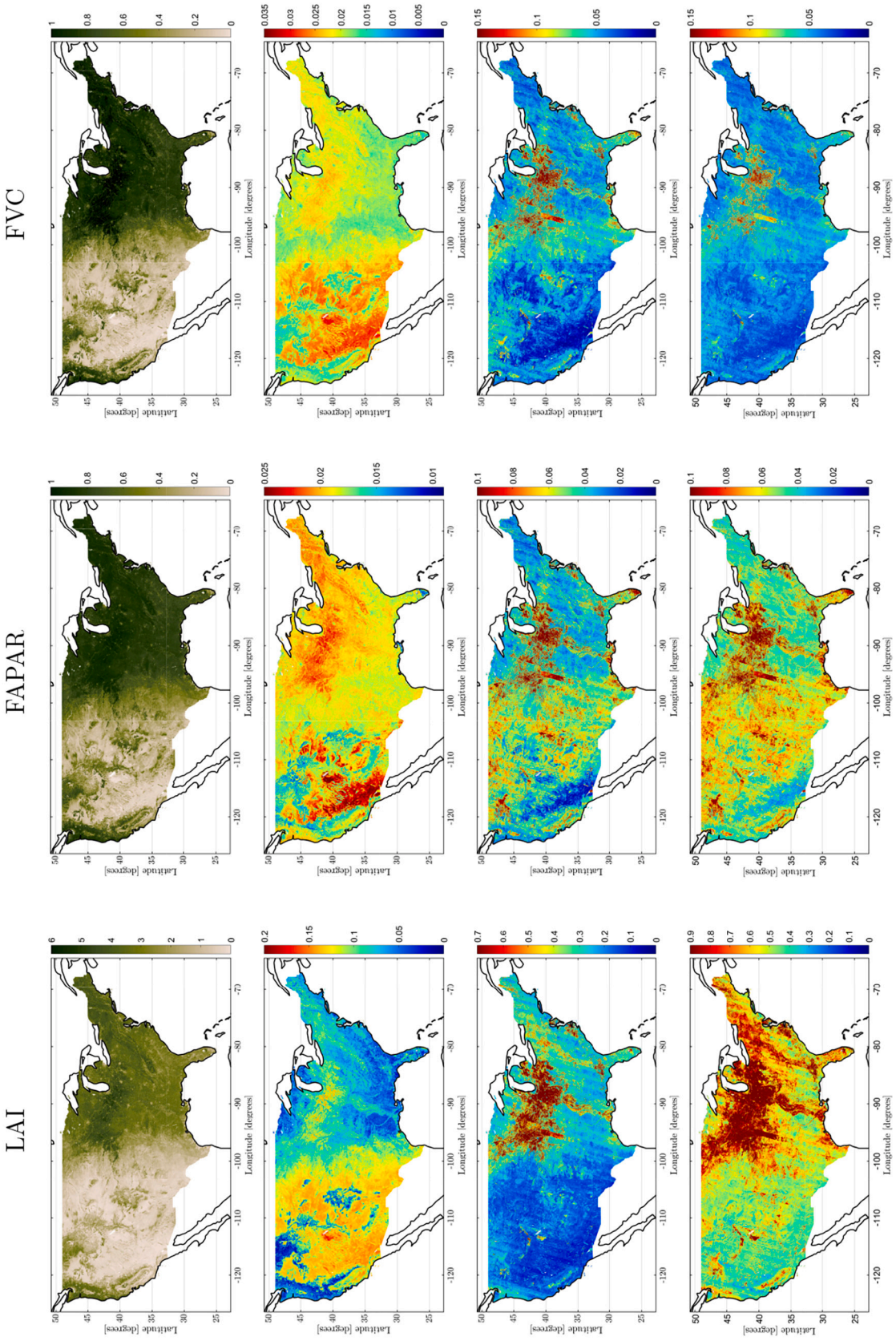

Fig. 4. BioNet results of prediction and uncertainties for LAI, FAPAR and FVC. For illustrative purposes only August 2016 are shown.

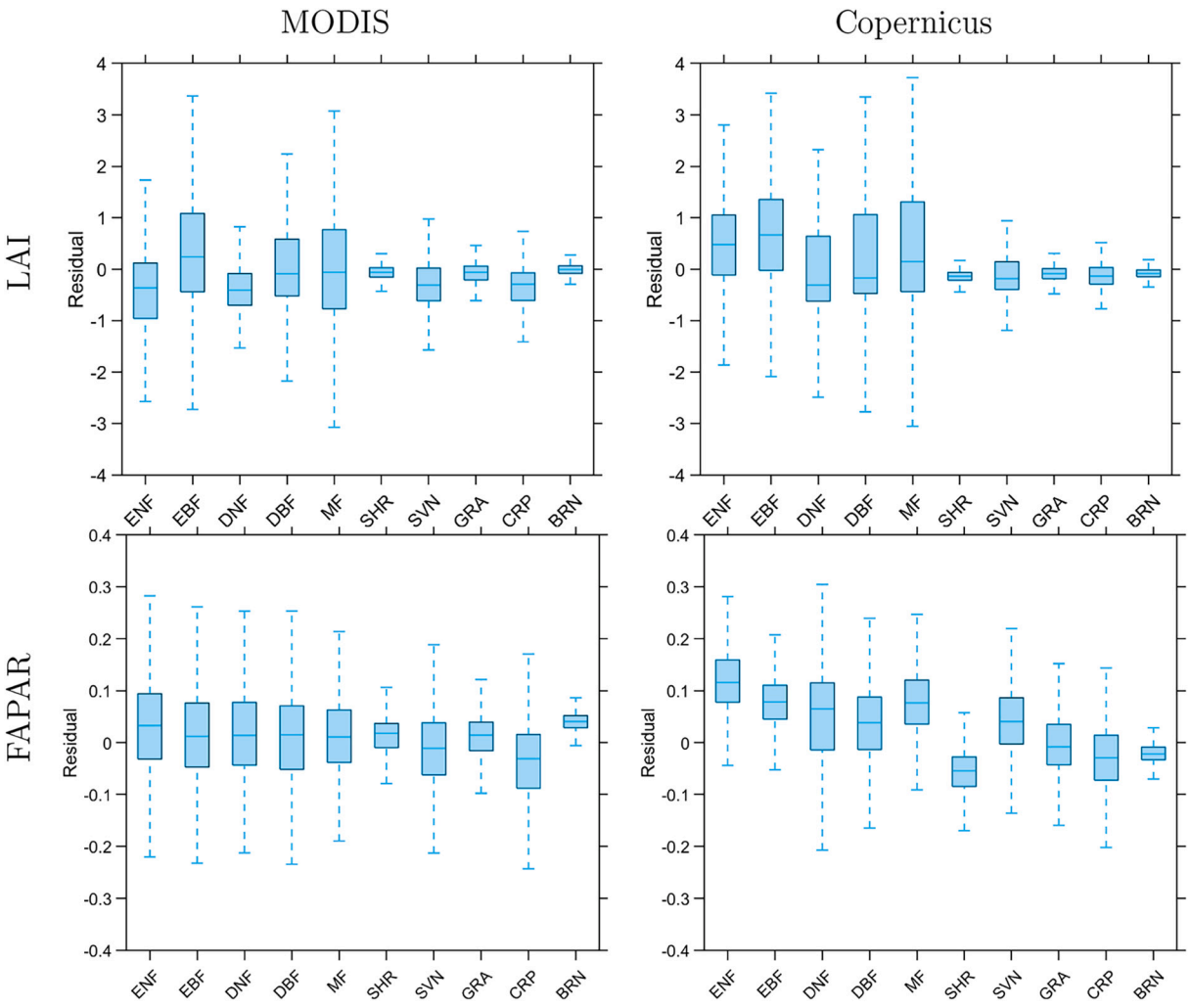

Fig. 5. Boxplot representing the differences between BioNet and MODIS/Copernicus for the classes named in Tables C.3 for 2016.

apply it to all the variables. Note that this just implies a multiplication times a scalar in the production (test) phase, which is computationally efficient.

3. Results

We evaluate the proposed retrieval algorithm (hereafter named BioNet) performance with a quantitative and qualitative analysis. We provide accuracy, fit and bias scores for the calibrated neural network training, as well as a thorough validation of the derived product estimates in time and space. We also compare the proposed methodology with MODIS (LAI, FAPAR) and Copernicus (LAI, FVC, FAPAR) operational products. The spatial inspection of the products include two main regions. We mainly focused our study area in the Contiguous US (CONUS), a large and highly heterogeneous and representative area with a wide range of climatic regions and vegetation types. We complement this study in the more challenging area of Amazonia, where the heavy cloud contamination imposes difficulties in any biophysical variable estimation processing chain. Finally, we carried out a site level comparison with MODIS and Copernicus operational products in the Belmanip site network over long temporal records to check the temporal consistency among them.

3.1. Quantitative analysis

Table 2 and Fig. 3 illustrate the statistical performance and the scatterplots of BioNet over an independent out-of-sample test set consisting of $N \simeq 500$ observations for the four biophysical variables considered: LAI, FAPAR, FVC and CWC. Performance is evaluated with several standard measures of association: mean error (ME) reflecting the bias of the estimates, the root-mean-square-error (RMSE) capturing accuracy, the Pearson's correlation coefficient for ρ linear goodness-of-fit. In addition, we also use nonlinear measures of fit like the Spearman's correlation ρ_S (Hollander et al., 2013), mutual information MI (Cover and Thomas, 2006) and distance correlation D_{corr} (Székely et al., 2007). Both linear and nonlinear measures of fit reflect good-to-outstanding results for all variables, with high correlations and shared information, low error levels and biases. Results for LAI, FVC and FAPAR are numerically similar and better than for CWC, which is a more challenging problem as reported elsewhere (García-Haro et al., 2020). Statistics are consistent with results previously reported in the literature (Kang et al., 2021; Djamai et al., 2019; García-Haro et al., 2018). Fig. 3 confirms the low bias and very good fit of the models in general (small intercept and close to one slopes), with some slight underestimation for LAI and the impact of outliers in CWC.

3.2. Qualitative analysis

Here we illustrate the performance of our methodology in two challenging cases involving heterogeneous covers and high missing data and noise levels.

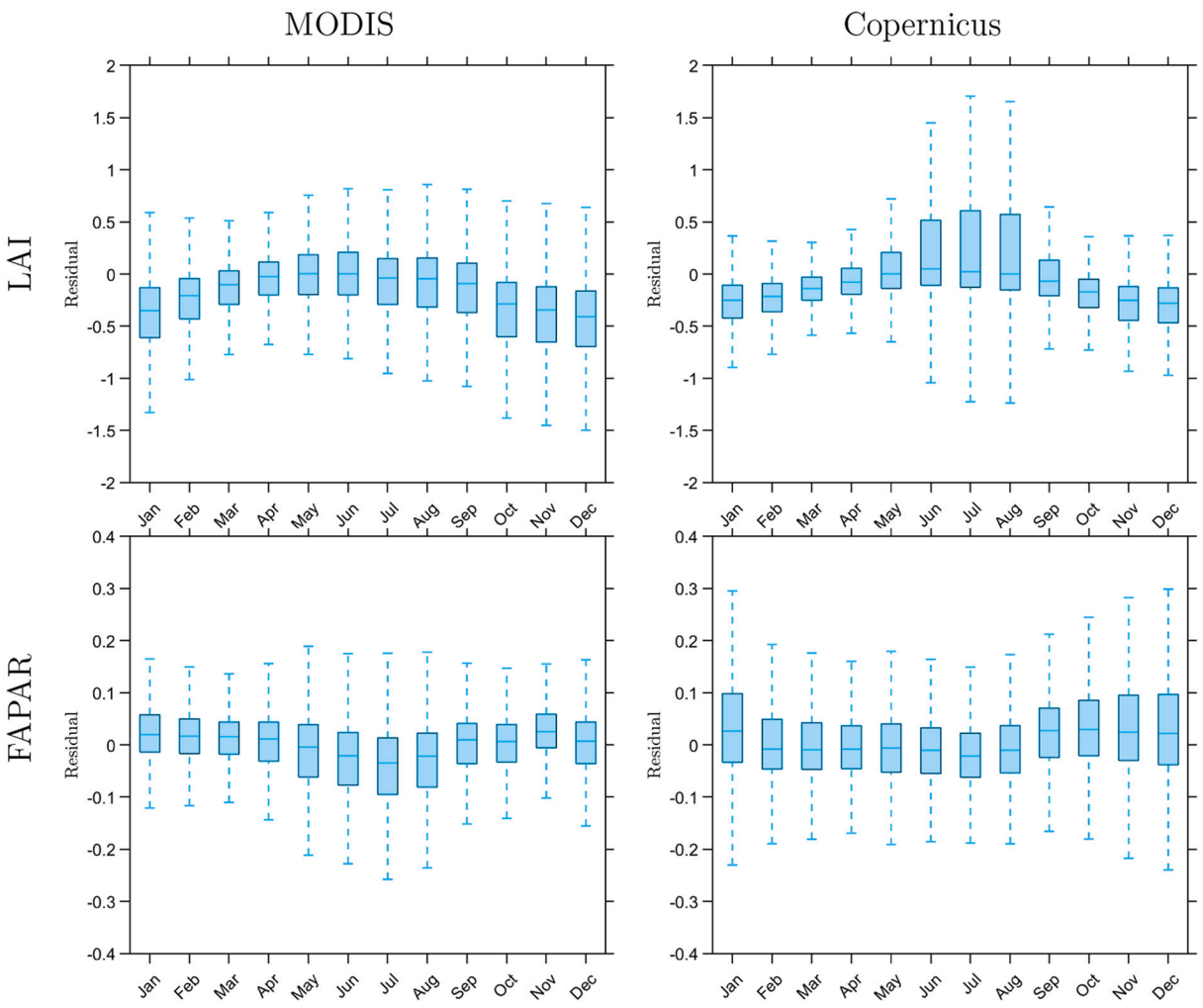

Fig. 6. Boxplot representing the differences between BioNet and MODIS/Copernicus over the months of 2016.

3.2.1. Heterogeneous CONUS area

The extent and heterogeneity of the CONUS area gives us the opportunity to assess how the approach performs in different biome types. The time series of this study area comprises 10 years of monthly products derived at high spatial resolution (30 m) from Landsat between 2009 and 2019. For illustrative purposes, Fig. 4 shows an example of the results corresponding to August 2016, both the predictions, the two sources of uncertainty and the total uncertainty.

The spatial patterns of predictions have similar behaviour obtaining high values in the East of the CONUS corresponding with areas of croplands and deciduous and mixed forests, and lower values for the western part corresponding with grasslands and shrublands. According to the uncertainty maps, it is remarkable that for the aleatoric one there is a substantial similarity for the higher values, since nearly all of them are located in the cropland area of the country conforming a common spatial pattern for the specific data shown in the figure. Grasslands, pastures, and crops have a significantly faster growth (and phenological cycles) than forests, yielding higher variance at the considered temporal resolution (monthly time steps). This increase in variance should not be, in fact, attributed to noise exclusively but also related with faster changes in vegetation canopy which occur within a monthly time step (Moreno-Martínez et al., 2020).

In order to evaluate the similarities between BioNet and the operational products from MODIS and Copernicus we computed the differences for each class (acronyms in Appendix CTable C.3) and for each month. Results are shown in Figs. 5 and 6 for 2016 for illustrative purposes. The mean values for each biophysical product computed with BioNet and the mean of the MODIS products were calculated within the extent of the Copernicus pixels (1000 m pixel size) to match their spatial resolution for a more proper comparison.

The comparisons with MODIS/Copernicus have been made at pixel level but this type of comparison introduces considerable errors due to: spatial differences and scale effects. They cause increased dispersion in our cross-product comparisons and are expected to be more pronounced in areas domitated by sub-pixel heterogeneity. Fig. 5 shows that the differences between BioNet and MODIS or Copernicus products are land cover class specific. For LAI estimation, BioNet underestimates over EBF for MODIS and over ENF, EBF and MF for Copernicus. No significant differences with MODIS are observed for the other land cover classes (DBF, MF, SHR, GRA and ENF), while the main discrepancies with Copernicus are observed over GRA and BRN. There is more variability for classes ENF, EBF, DNF, DBF and MF corresponding with the different types of forest in the CONUS. A possible explanation could be that our methodology is based on a turbid model so that our predictions are closer to an *effective* LAI, whereas both MODIS and Copernicus correspond to *actual* LAI. This becomes more visible in biomes with high vegetation content. Regarding FAPAR, higher discrepancies are observed with Copernicus in forests (ENF, EBF, DNF, DBF and MF) and Savanas (SVN) where the values of BioNet FAPAR estimates are lower. Fig. 6 reveals that there is practically no variation in the central months of the year between BioNet LAI estimates and MODIS and Copernicus estimates. Interestingly, a pattern can be seen that generates an upward

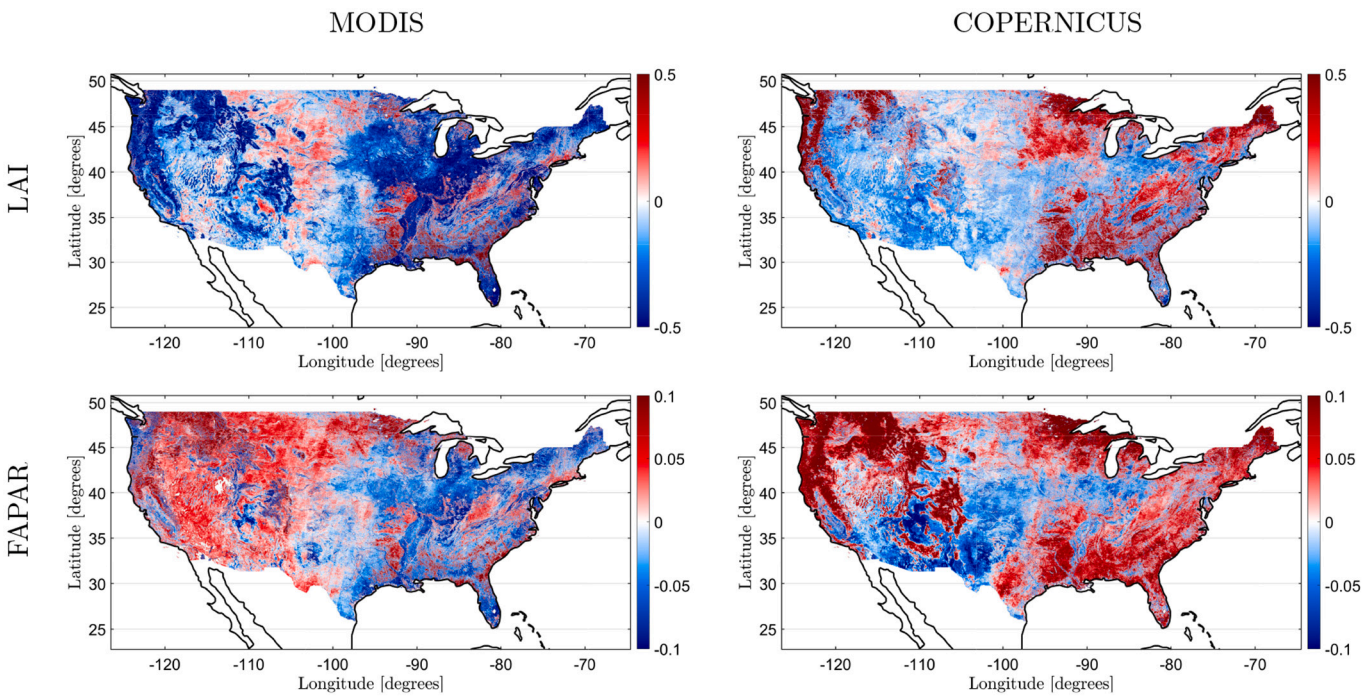

Fig. 7. Maps representing the relative differences between BioNet and MODIS/Copernicus.

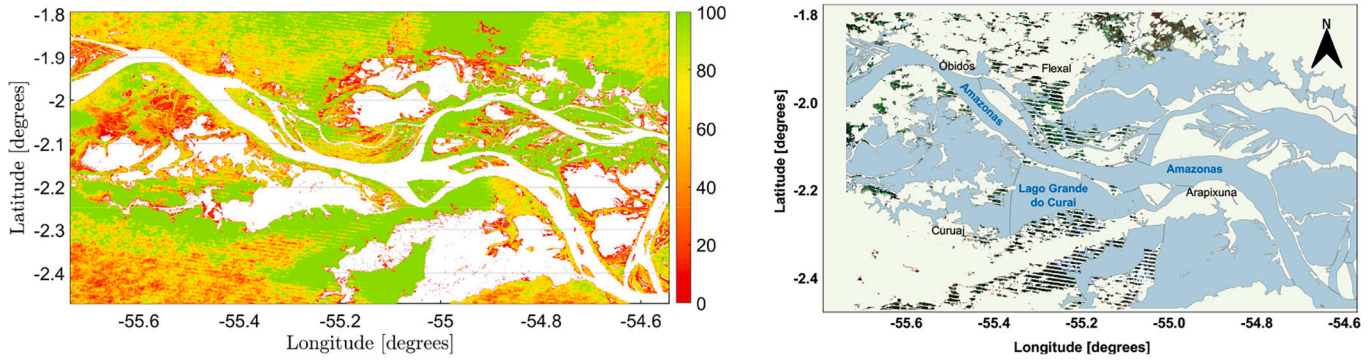

Fig. 8. Map of the available data (%) on the left and the original Landsat images for the same area.

curve towards the middle months of the year, which might be related with the standard vegetation phenology so that, when the vegetation is at its maximum, BioNet cannot reach the higher values because it reflects the *effective* LAI rather than the *actual* LAI. For FAPAR the more representative discrepancies are with Copernicus at the end of the year.

Fig. 7 represents the relative differences between BioNet and the MODIS and Copernicus products for LAI and FAPAR during 2016. It is remarkable that in general the differences do not exceed 0.5 for LAI and 0.15 for FAPAR. However, some forest biomes yield higher differences. This might be explained because MODIS is biome-dependent and therefore Copernicus too (indirectly), as it is trained with a combination of MODIS (for intermediate-dense vegetation) and CYCLOPES (sparse vegetation).

3.2.2. *High missing data region - Amazonia*

Let us now show the performance of our proposed scheme over the challenging Amazonia region, where the large abundance of clouds hamper the retrieval of parameters (Liu et al., 2019; López-Puigdollers et al., 2021), see Fig. 8. The use of the HISTARFM model allows us to alleviate this problem by providing high resolution reflectance data of the region, and hence allows us to derive parameter and uncertainty estimates.

Fig. 9 shows an example of performance of BioNet for LAI. The difficulty presented in this area translates into higher values of aleatoric uncertainty, caused by the elevated noise input level. This case study highlights the importance of propagating the input uncertainties through the model as they are the main source of uncertainty in the predictions and need to be accounted accordingly. While also representing both components of the total parameter uncertainty, which together provide more comprehensive and meaningful estimates.

3.3. *Site-level comparison with other operational products*

An inter-comparison of LAI and FAPAR with the available LAI/FAPAR product (MCD15A3H) on GEE and the Copernicus product (GEOV1) from the Copernicus Land Service was performed. For FVC we used just GEOV1 as it is the only operational product available. The inter-comparison was conducted over a selection of sites extracted from the BELMANIP-2.1 (Benchmark Land Multiuse Analysis and Intercomparison of Products) for the study area. These sites were selected for representing the global variability of vegetation, making them suitable for global inter-comparison of land biophysical products (Baret et al.,

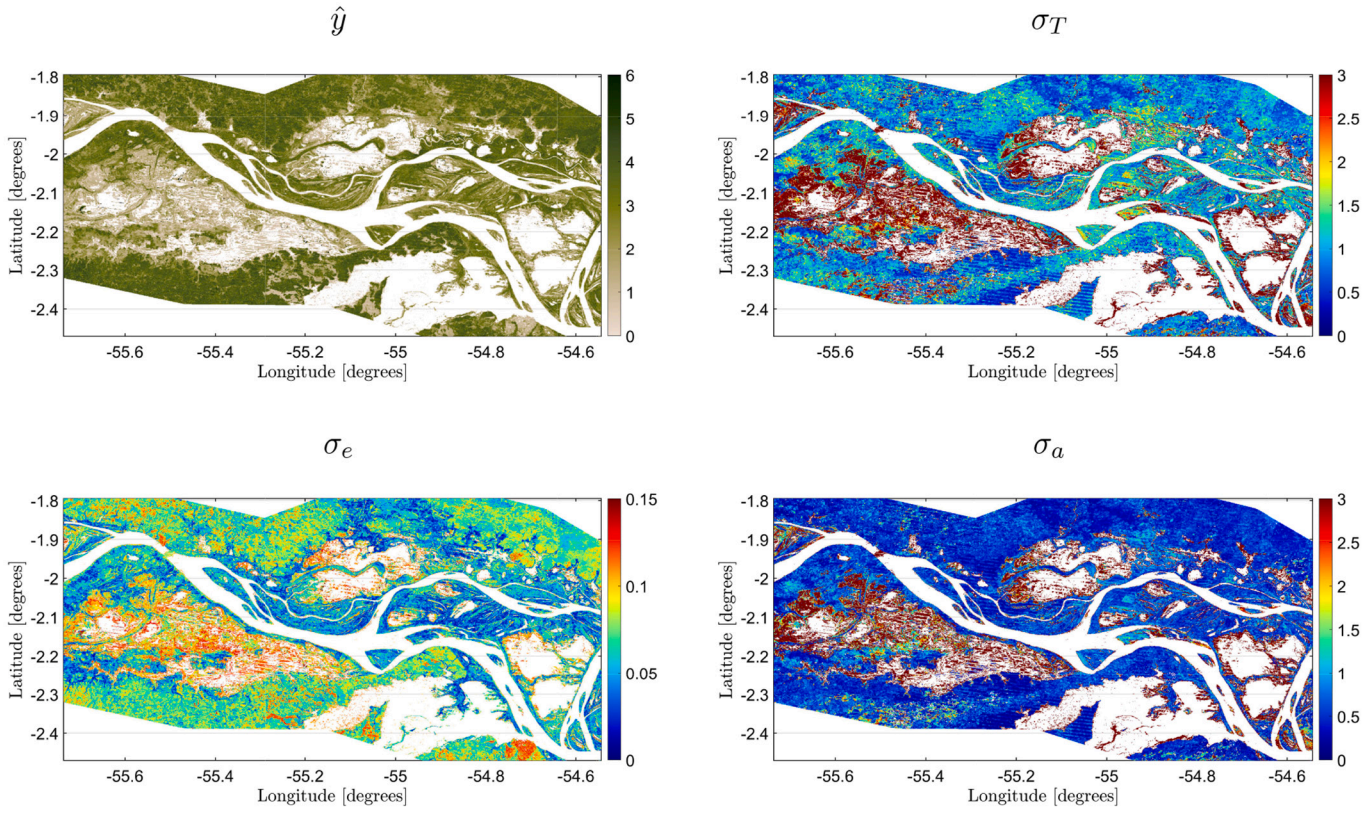

Fig. 9. Results of LAI estimation and over the Amazonia region: Prediction $\hat{y}$, total uncertainty σ_T, epistemic σ_e and aleatoric σ_a, respectively.

Table 3
Results for the inter-comparison between BioNet methodology and MODIS/Copernicus (mean error, ME, root-mean-square error, RMSE, Pearson's correlation coefficient ρ) Spearman's correlation coefficient ρ_S, mutual information MI, and distance correlation D_{corr}.

	LAI(m^2/m^2)						FAPAR					
	ME	RMSE	ρ	ρ_S	MI	D_{corr}	ME	RMSE	ρ	ρ_S	MI	D_{corr}
MODIS	0.23	0.75	0.84	0.82	0.60	0.85	0.015	0.14	0.85	0.86	0.65	0.87
Copernicus	0.02	0.69	0.92	0.84	0.92	0.91	0.0007	0.13	0.91	0.89	0.87	0.91

2006). The sites are aimed to be representative of the different planet biomes over an 10×10 km^2 area, mostly flat, and with minimum fractions of urban area and permanent water bodies. There are 27 sites in the CONUS that we used for validation. Reasonable results were obtained when calculating linear and non-linear measures of association (see Table 3) between BioNet and MODIS/Copernicus, with high correlations and low error values and biases. Interestingly, the mutual information is higher with MODIS in both cases for LAI and FAPAR.

Fig. 10 illustrates the full time series of the considered 10 years ($2009 - 2019$) over two illustrative sites: 1) site "Texas" with lat/lon coordinates $(29.99, -104.19)$; and 2) site "Arkansas" with coordinates lat/lon (35.79, 93.49). In terms of prediction/estimation, BioNet behaviour is similar to the official products with no big changes or discrepancies. Site "Arkansas" corresponds to an evergreen forest area and thus there is a large amount of vegetation mass which translates into high LAI values (sometimes so high that they can only be obtained with *actual* LAI). This is more evident during the summer season (when the forest cover is at its peak) and BioNet estimates are below both MODIS and GEOV1. As described previously, BioNet provides the effective values of LAI this could explain the considerable differences in the "Arkansas" site where MODIS and GEOV1 LAI values are higher than those predicted by the BioNet method. Site "Texas" is a shrubland area, dominated by perennial trees and shrubs, and hence LAI, FAPAR and FVC values are lower. This figure illustrates the consistency of the methodology as it performs with no noticeable discrepancies in two very different phenologies. The total uncertainty bars provided can be very useful not only to assess model and product robustness, but also to mask out eventual anomalous estimates by attributing them to either model, data or both uncertainties together.

4. Conclusions

This paper proposed a new methodology for the estimation of biophysical variables such as LAI, FAPAR, FVC and CWC at high spatial resolution over large areas, along with realistic uncertainties composed of both aleatoric and epistemic contributions. The retrieval algorithm is a neural network trained to invert a dataset derived from representative PROSAIL simulations, and applied to high resolution (30 m) gap-filled reflectance data derived using the HISTARFM algorithm. This approach is not incidental; it not only allowed us to obtain prediction gap-free monthly maps of biophysical parameters of interest at 30 m and over large areas, but also (and very importantly) it allowed us to deal with data and model uncertainties effectively. First, because we can propagate data uncertainties through the neural network in a rigorous way, and second because by taking the model uncertainty into account through MC dropout, the two types of uncertainties can be combined. Calibration curves facilitate assessing the reliability of uncertainty estimates. This analysis showed that for LAI MC dropout tends to

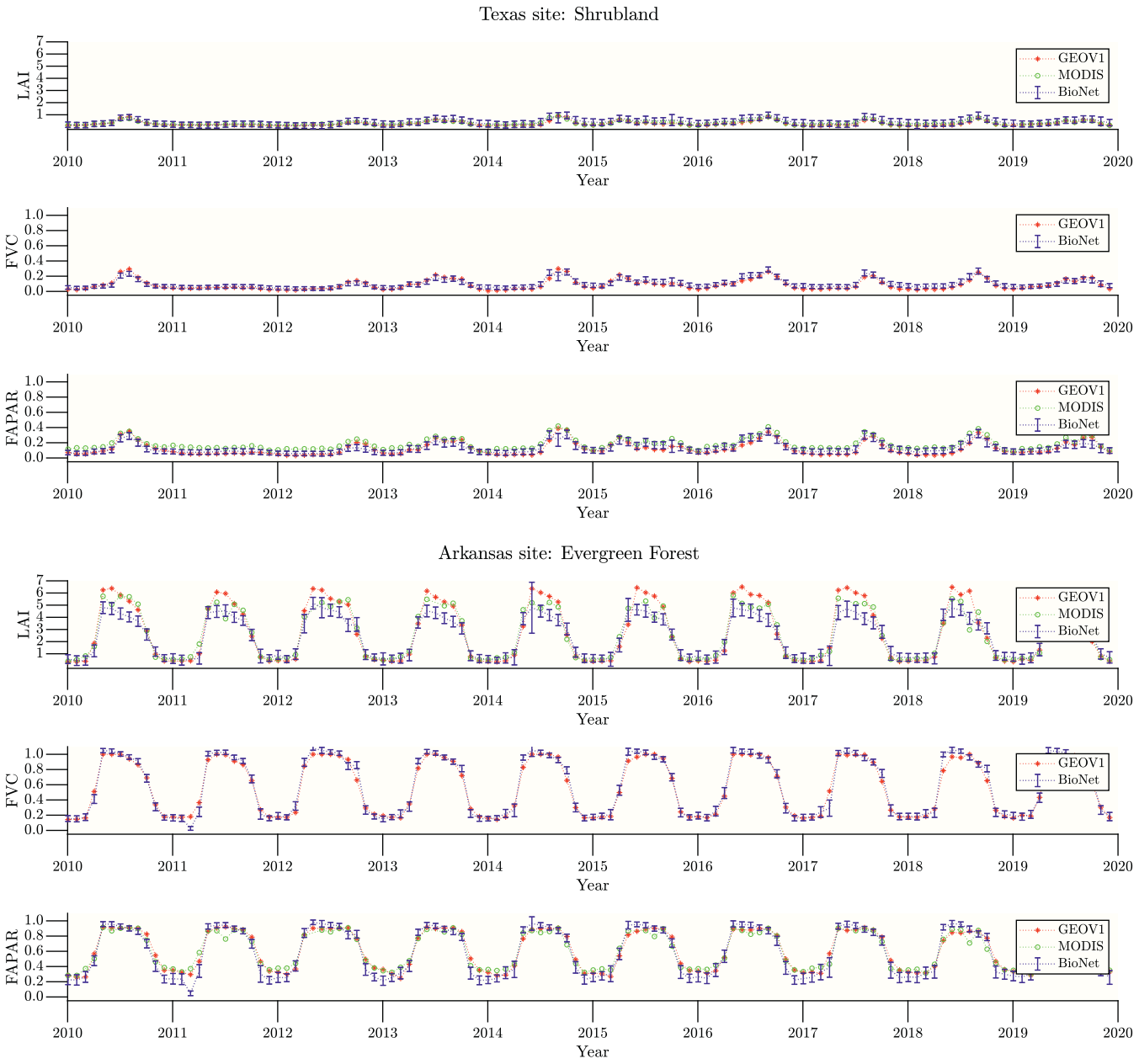

Fig. 10. 10 year-time series comparison from sites located in Texas and Arkansas with vegetation type Shrubland and Evergreen Forest respectively.

underestimate the uncertainties which results in an overconfident model. By recalibrating our trained neural network, we were able to accompany the prediction maps with its corresponding reliable uncertainty maps, which can be very useful to understand model performance, mask out errors and learn about potential sampling biases.

We argue that the generation of global products at this unprecedented high spatio-temporal accuracy requires large computing resources and efficient implementations, which was possible by using the GEE free platform. Other high-performance computing platforms could be eventually used. The huge amount of data ($8x10^9$ pixels per image) was exploited efficiently through the GEE platform, which offers distinct advantages mainly related to storage capacity and processing speed. Scaling and deployment would have been simply impossible without it. The methodology was applied mainly over large, heterogeneous and representative CONUS area, but the performance over areas where the data loss is around 60% like the Amazonia, demonstrates the versatility of the methodology.

We performed an indirect site-level model intercomparison over Belmanip network located in CONUS between BioNet and the available operational products (Copernicus and MODIS) for LAI, FAPAR and FVC. The obtained results with our methodology showed spatial consistency and highlighted the general robustness of the method. In densely vegetated areas we observed some discrepancies between products, which could be attributed to the definition of the considered variables: e. g. BioNet LAI estimates are closer to an *effective* LAI rather than *actual* LAI, which are the MODIS/Copernicus estimates.

Our study provides the framework and lends support for future studies to assess the uncertainty of its predictions, and opens the door to undertake gap free global studies for crop monitoring, carbon fluxes or high resolution phenologies.

CRediT authorship contribution statement

Laura Martínez-Ferrer: Conceptualization, Methodology, Software,

Writing – original draft. **Álvaro Moreno-Martínez:** Methodology, Supervision, Writing – review & editing. **Manuel Campos-Taberner:** Validation, Writing – review & editing. **Francisco Javier García-Haro:** Validation, Writing – review & editing. **Jordi Muñoz-Marí:** Software, Writing – review & editing. **Steven W. Running:** Writing – review & editing. **John Kimball:** Writing – review & editing. **Nicholas Clinton:** Resources, Software. **Gustau Camps-Valls:** Methodology, Supervision, Writing – review & editing.

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.

Data availability

I have shared the link to my code

Acknowledgements

This research was financially supported by the ERC under the ERC-SyG-2019 USMILE Project (Grant Agreement 855187). The authors want to thank the Google Earth Engine developers for their support and technical advice.

Appendix A. PROSAIL simulation details

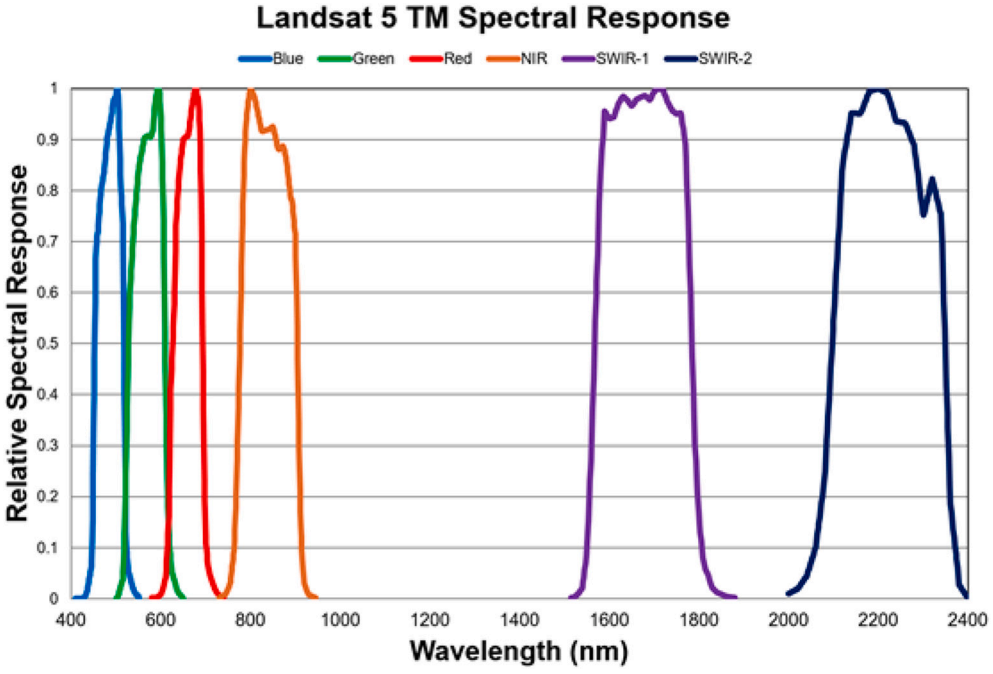

Fig. A.1. Spectral response functions of Landsat-5 bands (blue, green, red, nir, SWIR-1, and SWIR-2 channels). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table A.1
Ranges and distributions of the PROSAIL parameters adopted in the EPS retrieval chain. (*) KDE refers to kernel density estimation method, which does not provide any parameters being a non parametric model of the marginal distributions. (**) A 5% of spectra representative of pure background (vCover = 0) were included to account for bare areas. KDE refers to kernel density estimation method, which does not provide any parameters being a non parametric model of the marginal distributions.

Parameter		Min	Max	Mode	Std	Type
Leaf	N	1.2	2.2	1.5	0.3	Gaussian
	C_{ab} (μg·cm^{-2})	–	–	–	–	KDE*
	C_{ar} (μg·cm^{-2})	0.6	16	5	7	Gaussian
	C_{dm} (g·cm^{-2})	–	–	–	–	KDE*
	C_w	–	–	–	–	KDE*
	C_{bp}	0	0	0	0	–
Canopy	LAI (m^2/m^2)	0	8	3.5	4	Gaussian
	ALA (°)	35	80	62	12	Gaussian
	Hotspot	0.1	0.5	0.2	0.2	Gaussian
	vCover	0.3	1	0.99	0.2	Truncated Gaussian**
Soil	β_s	0.1	1	0.8	0.6	Gaussian

Appendix B. Calibration

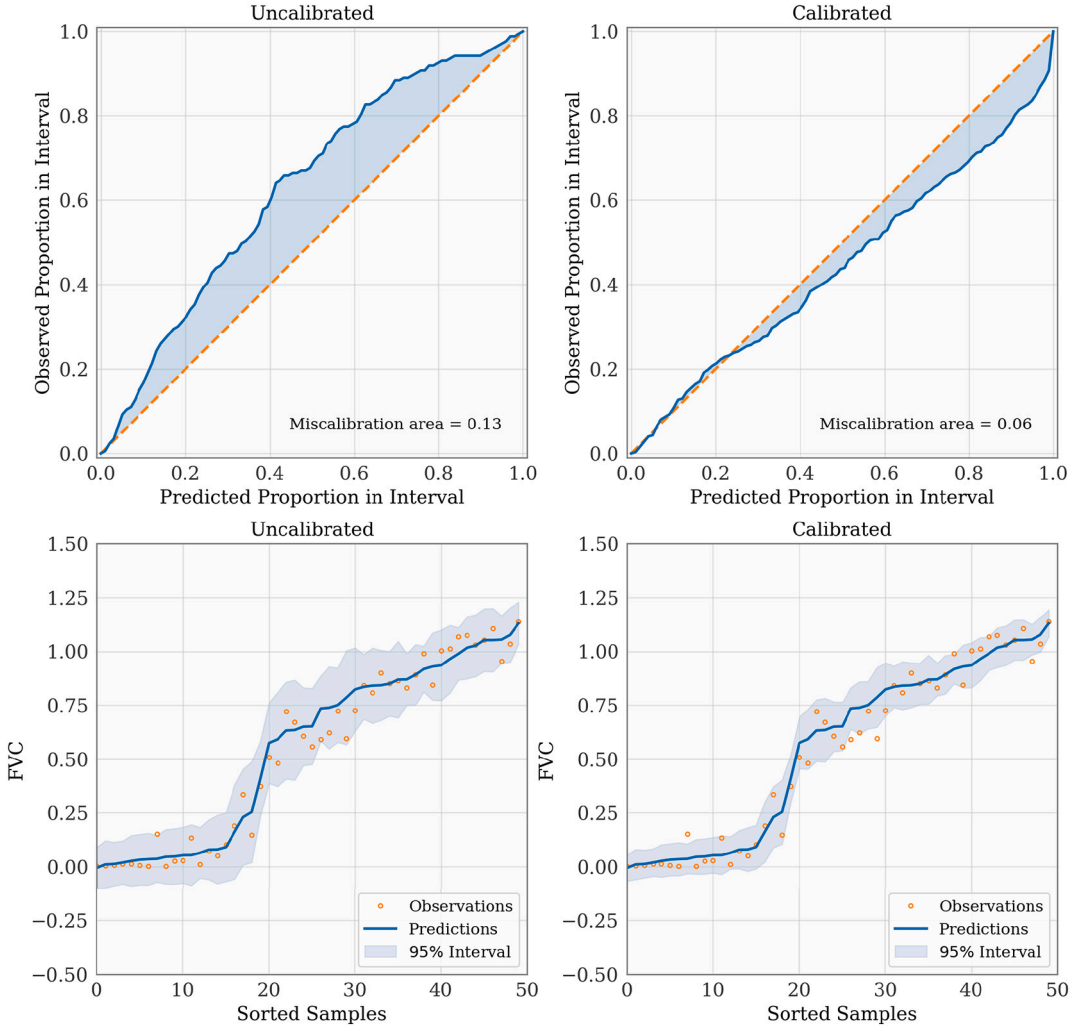

Fig. B.2. FVC calibration curves (top) and prediction with uncertainties (bottom) using uncalibrated and calibrated model.

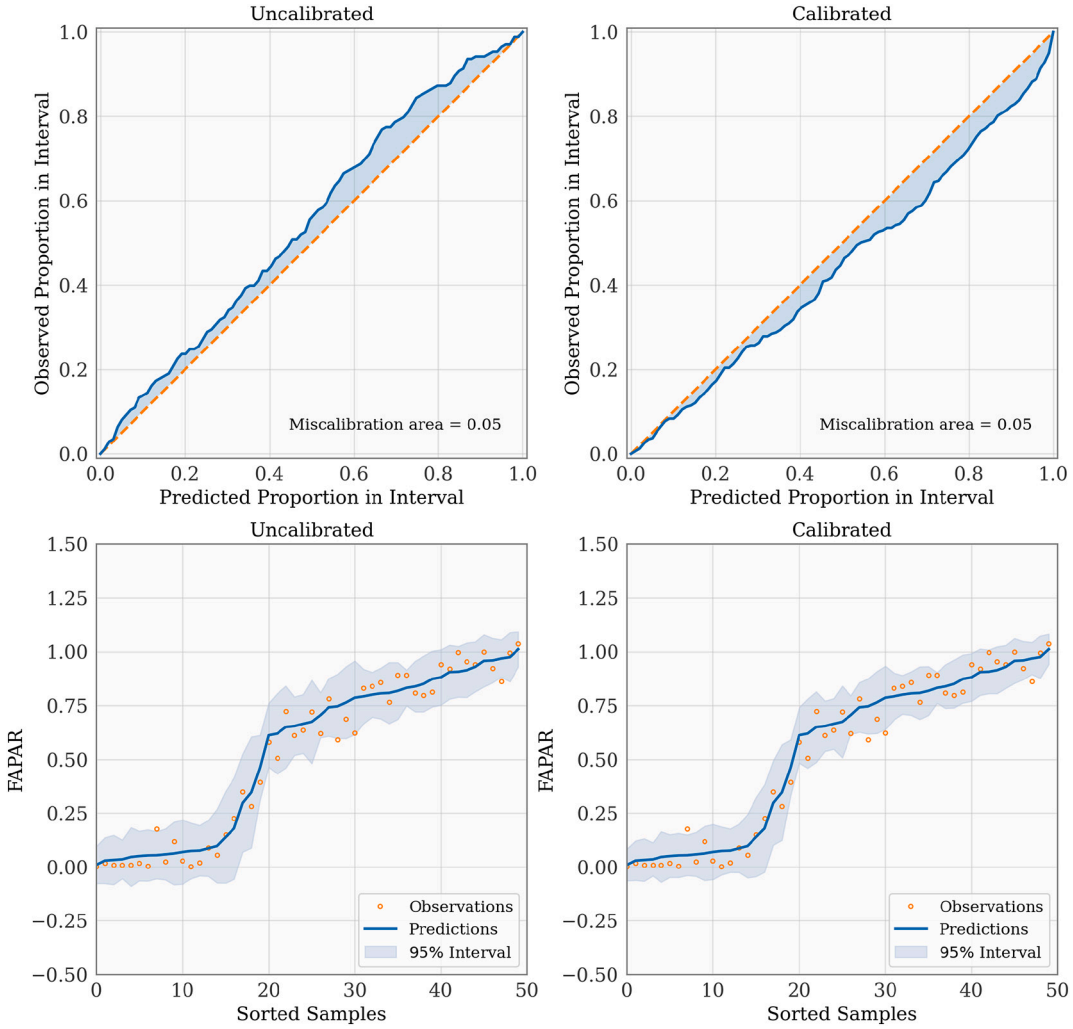

Fig. B.3. FAPAR calibration curves (top) and prediction with uncertainties (bottom) using uncalibrated and calibrated model.

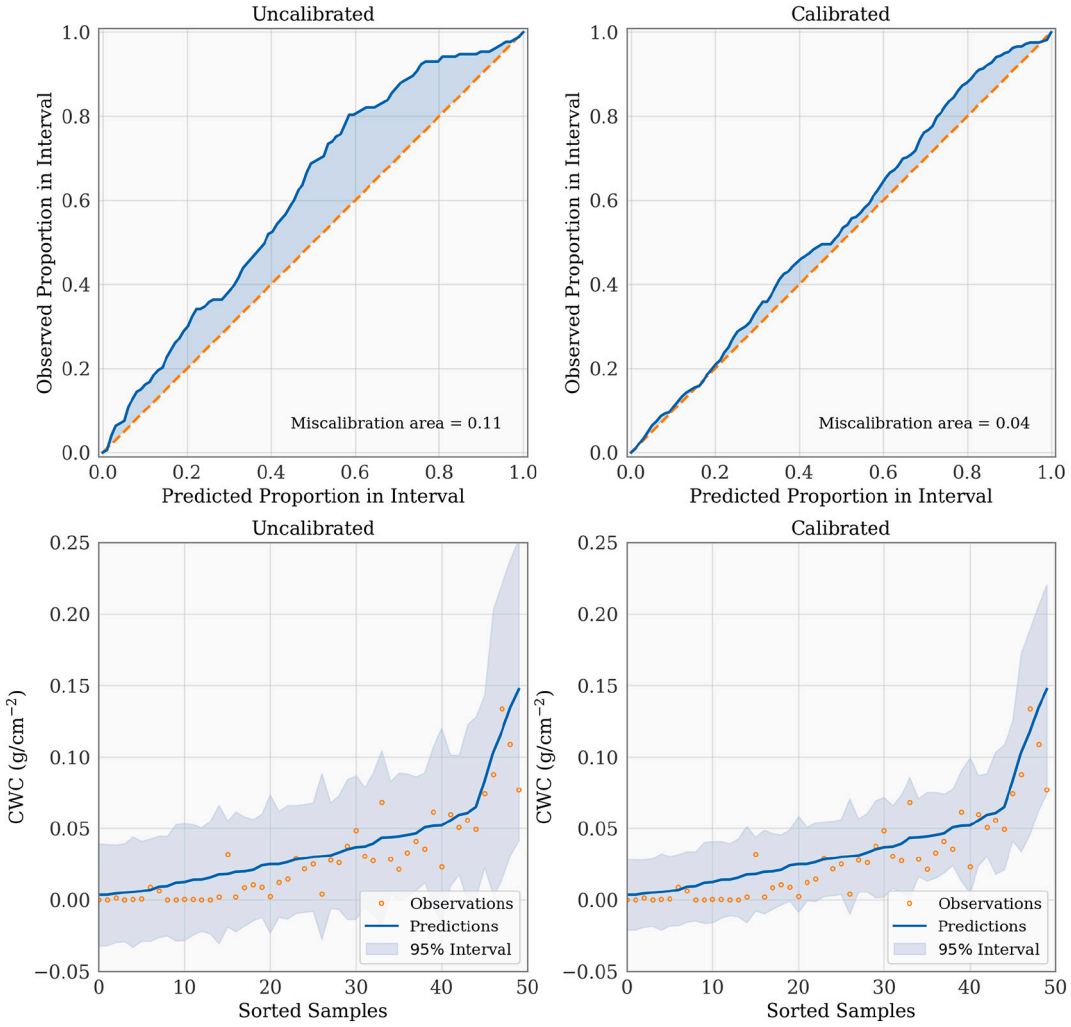

Fig. B.4. CWC calibration curves (top) and prediction with uncertainties (bottom) using uncalibrated and calibrated model.

Table B.2
Metrics obtained in calibration: mean absolute calibration error MACE, root mean squared calibration error RMSCE and miscalibration area MA.

	FVC			FAPAR			CWC(g/cm^2)		
	MACE	RMSCE	MA	MACE	RMSCE	MA	MACE	RMSCE	MA
Uncalibrated	0.13	0.15	0.13	0.05	0.06	0.05	0.11	0.13	0.11
Calibrated	0.06	0.07	0.06	0.05	0.06	0.05	0.04	0.05	0.04

Appendix C. Table of acronyms

Table C.3
Classes from the MODIS Land cover (MCD12Q1).

Acronym	Description
ENF	Evergreen Needleleaf Forests
EBF	Evergreen Broadleaf Forests
DNF	Deciduous Needleleaf Forests
DBF	Deciduous Broadleaf Forests
MF	Mixed Forests
SHR	Shrublands
SVN	Savannas
GRA	Grasslands
CRP	Croplands
BRN	Barren

Appendix D. Maps of results for CWC

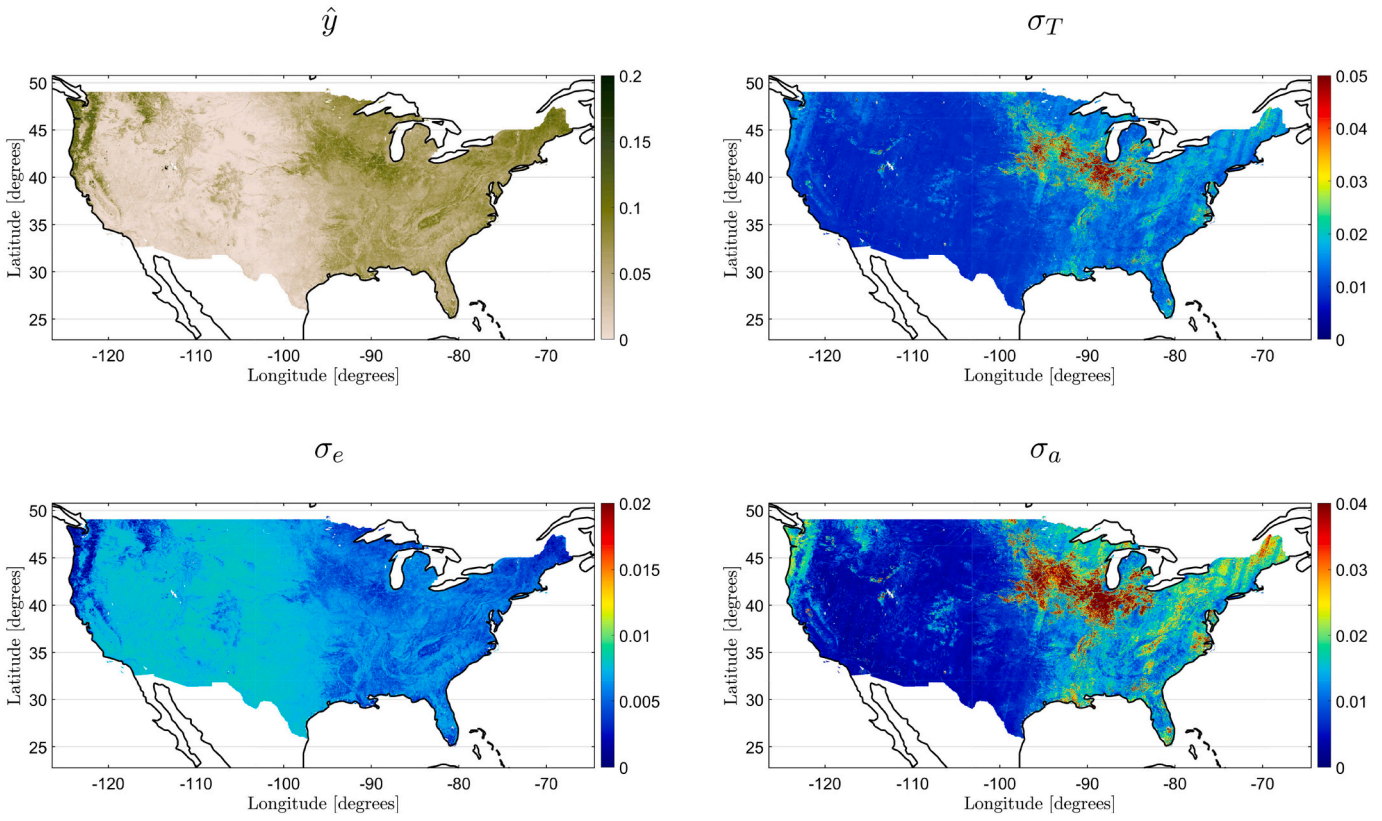

Fig. D.5. Results of CWC estimation: Prediction $\hat{y}$, total uncertainty σ_T, epistemic σ_e and aleatoric σ_a, respectively.

References

Adsuara, J.E., Pérez-Suay, A., Muñoz-Marí, J., Mateo-Sanchis, A., Piles, M., Camps-Valls, G., 2019. Nonlinear distribution regression for remote sensing applications. IEEE Trans. Geosci. Remote Sens. 57, 10025–10035.

Aires, F., Prigent, C., Rossow, W.B., 2004. Neural network uncertainty assessment using Bayesian statistics: a remote sensing application. Neural Comput. 16, 2415–2458.

Ardizzone, L., Kruse, J., Wirkert, S., Rahner, D., Pellegrini, E.W., Klessen, R.S., Maier-Hein, L., Rother, C., Köthe, U., 2019. Analyzing inverse problems with invertible neural networks. In: International Conference on Learning Representations.

Asner, G.P., 1998. Biophysical and biochemical sources of variability in canopy reflectance. Remote Sens. Environ. 64, 234–253.

Baret, F., Morissette, J., Fernandes, R., Champeaux, J., Myneni, R., Chen, J., Plummer, S., Weiss, M., Bacour, C., Garrigues, S., Nickeson, J., 2006. Evaluation of the representativeness of networks of sites for the global validation and intercomparison of land biophysical products: proposition of the CEOS-BELMANIP. IEEE Trans. Geosci. Remote Sens. 44, 1794–1803.

Baret, F., Hagolle, O., Geiger, B., Bicheron, P., Miras, B., Huc, M., Berthelot, B., Niño, F., Weiss, M., Samain, O., Roujean, J.L., Leroy, M., 2007. LAI, fAPAR and fCover CYCLOPES global products derived from VEGETATION: part 1: principles of the algorithm. Remote Sens. Environ. 110, 275–286.

Baret, F., Weiss, M., Lacaze, R., Camacho, F., Makhmara, H., Pacholcyzk, P., Smets, B., 2013. GEOV1: LAI and FAPAR essential climate variables and FCOVER global time series capitalizing over existing products. Part1: principles of development and production. Remote Sens. Environ. 137, 299–309.

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.

Bishop, C.M., et al., 1995. Neural Networks for Pattern Recognition. Oxford university press.

Bonham, C.D., 2013. Measurements for Terrestrial Vegetation. John Wiley & Sons.

Caers, J., 2011. Modeling Uncertainty in the Earth Sciences. John Wiley & Sons.

Campos-Taberner, M., García-Haro, F., Camps-Valls, G., Grau-Muedra, G., Nutini, F., Crema, A., Boschetti, M., 2016. Multitemporal and multiresolution leaf area index retrieval for operational local rice crop monitoring. Remote Sens. Environ. 187, 102–118.

Campos-Taberner, M., Moreno-Martínez, A., García-Haro, F.J., Camps-Valls, G., Robinson, N.P., Kattge, J., Running, S.W., 2018. Global estimation of biophysical variables from Google Earth Engine platform. Remote Sens. 10, 1167.

Camps-Valls, G., Tuia, D., Gómez-Chova, L., Jiménez, S., Malo, J. (Eds.), 2011. Remote Sensing Image Processing. Morgan & Claypool Publishers, LaPorte, CO, USA.

Camps-Valls, G., Verrelst, J., Muñoz-Marí, J., Laparra, V., Mateo-Jimenez, F., Gómez-Dans, J., 2016. A survey on Gaussian processes for earth observation data analysis: a comprehensive investigation. IEEE Geosci. Remote Sens. Mag. 4, 58–78.

Camps-Valls, G., Campos-Taberner, M., Moreno-Martínez, A., Walther, S., Duveiller, G., Cescatti, A., Mahecha, M.D., Muñoz-Marí, J., García-Haro, F.J., Guanter, L., Jung, M., Gamon, J.A., Reichstein, M., Running, S.W., 2021a. A unified vegetation index for quantifying the terrestrial biosphere. Sci. Adv. 7.

Camps-Valls, G., Tuia, D., Zhu, X., Reichstein, M., 2021b. Deep Learning for the Earth Sciences: A Comprehensive Approach to Remote Sensing, Climate Science and Geosciences. Wiley & Sons.

Chen, J.M., Black, T.A., 1992. Defining leaf area index for non-flat leaves. Plant Cell Environ. 15, 421–429.

Clerici, M., Vossbeck, M., Pinty, B., Kaminski, T., Taberner, M., Lavergne, T., Andredakis, I., 2010. Consolidating the two-stream inversion package (jrc-tip) to retrieve land surface parameters from albedo products. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 3, 286–295.

Clevers, J., Kooistra, L., Schaepman, M., 2010. Estimating canopy water content using hyperspectral remote sensing data. Int. J. Appl. Earth Obs. Geoinf. 12, 119–125.

Coleman, T.F., Li, Y., 1996. An interior trust region approach for nonlinear minimization subject to bounds. SIAM J. Optim. 6, 418–445.

Combal, B., Baret, F., Weiss, M., Trubuil, A., Mace, D., Pragnere, A., Myneni, R., Knyazikhin, Y., Wang, L., 2003. Retrieval of canopy biophysical variables from bidirectional reflectance: using prior information to solve the ill-posed inverse problem. Remote Sens. Environ. 84, 1–15.

Cover, T.M., Thomas, J.A., 2006. Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing). Wiley-Interscience, USA.

Danson, F., Rowland, C., Baret, F., 2003. Training a neural network with a canopy reflectance model to estimate crop leaf area index. Int. J. Remote Sens. 24, 4891–4905.

Dennis, J.J., 1977. Nonlinear least-squares. In: State of the Art in Numerical Analysis, pp. 269–312.

Djamai, N., Fernandes, R., Weiss, M., McNairn, H., Goïta, K., 2019. Validation of the sentinel simplified level 2 product prototype processor (SL2P) for mapping cropland

biophysical variables using Sentinel-2/MSI and Landsat-8/OLI data. Remote Sens. Environ. 225, 416–430.
Fang, H., Liang, S., 2003. Retrieving leaf area index with a neural network method: simulation and validation. IEEE Trans. Geosci. Remote Sens. 41, 2052–2062.
Feret, J.-B., François, C., Asner, G.P., Gitelson, A.A., Martin, R.E., Bidel, L.P., 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.
Gal, Y., Ghahramani, Z., 2016. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In: International Conference on Machine Learning. Proceedings of Machine Learning Research, pp. 1050–1059.
García-Haro, F.J., Campos-Taberner, M., Muñoz-Marí, J., Laparra, V., Camacho, F., Sānchez-Zapero, J., Camps-Valls, G., 2018. Derivation of global vegetation biophysical parameters from EUMETSAT polar system. ISPRS J. Photogramm. Remote Sens. 139, 57–74.
García-Haro, F.J., Campos-Taberner, M., Moreno, A., Tagesson, H.T., Camacho, F., Martínez, B., Sánchez, S., Piles, M., Camps-Valls, G., Yebra, M., Gilabert, M.A., 2020. A global canopy water content product from AVHRR/Metop. ISPRS J. Photogramm. Remote Sens. 162, 77–93.
GCOS, 2011. Systematic Observation Requirements for Satellite-Based Products for Climate, 2011.
Glenn, E.P., Huete, A.R., Nagler, P.L., Nelson, S.G., 2008. Relationship between remotely-sensed vegetation indices, canopy attributes and plant physiological processes: what vegetation indices can and cannot tell us about the landscape. Sensors 8, 2136–2160.
Gong, P., 1999. Inverting a canopy reflectance model using a neural network. Int. J. Remote Sens. 20, 111–122.
Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.
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.
Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q., 2017. On calibration of modern neural networks. In: Precup, D., Teh, Y.W. (Eds.), Proceedings of the 34th International Conference on Machine Learning, pp. 1321–1330. PMLR volume 70 of Proceedings of Machine Learning Research.
Haboudane, D., Miller, J.R., Pattey, E., Zarco-Tejada, P.J., Strachan, I.B., 2004. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: modeling and validation in the context of precision agriculture. Remote Sens. Environ. 90, 337–352.
Hollander, M., Wolfe, D.A., Chicken, E., 2013. Nonparametric Statistical Methods, vol. 751. John Wiley & Sons.
Kang, Y., Ozdogan, M., Gao, F., Anderson, M.C., White, W.A., Yang, Y., Yang, Y., Erickson, T.A., 2021. A data-driven approach to estimate leaf area index for Landsat images over the contiguous US. Remote Sens. Environ. 258, 112383.
Kattge, J., Bönisch, G., Díaz, S., Lavorel, S., Prentice, I.C., Leadley, P., Tautenhahn, S., Werner, G.D., Aakala, T., Abedi, M., et al., 2020. TRY plant trait database–enhanced coverage and open access. Glob. Chang. Biol. 26, 119–188.
Kingma, D.P., Ba, J., 2014. Adam: A Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980, pp. 1–15.
Knyazikhin, Y., 1999. MODIS leaf area index (LAI), and fraction of photosynthetically active radiation absorbed by vegetation FPAR. In: Technical Report GSFC/NASA.
Knyazikhin, Y., Kranigk, J., Myneni, R.B., Panfyorov, O., Gravenhorst, G., 1998a. Influence of small-scale structure on radiative tranfer and photosynthesis in vegetation canopies. J. Geophys. Res. 103, 6133–6144.
Knyazikhin, Y., Martonchik, J.V., Diner, D.J., Myneni, R.B., Verstraete, M., Pinty, B., Gobron, N., 1998b. Estimation of vegetation canopy leaf area index and fraction of absorbed photosynthetically active radiation from atmosphere-corrected MISR data. J. Geophys. Res.-Atmos. 103, 32239–32256.
Liang, S., 2004. Quantitative Remote Sensing of Land Surfaces. John Wiley & Sons, New York.
Liang, S., 2008. Advances in Land Remote Sensing: System, Modeling, Inversion and Applications. Springer Verlag, Germany.
Lillesand, T.M., Kiefer, R.W., Chipman, J., 2008. Remote Sensing and Image Interpretation. John Wiley & Sons, New York.
Liu, C.-C., Zhang, Y.-C., Chen, P.-Y., Lai, C.-C., Chen, Y.-H., Cheng, J.-H., Ko, M.-H., 2019. Clouds classification from Sentinel-2 imagery with deep residual learning and semantic image segmentation. Remote Sens. 11.
López-Puigdollers, D., Mateo-García, G., Gómez-Chova, L., 2021. Benchmarking deep learning models for cloud detection in Landsat-8 and Sentinel-2 images. Remote Sens. 13.
Loquercio, A., Segu, M., Scaramuzza, D., 2020. A general framework for uncertainty estimation in deep learning. IEEE Robot. Automat. Lett. 5, 3153–3160.
Martin, R.E., Asner, G.P., Francis, E., Ambrose, A., Baxter, W., Das, A.J., Vaughn, N.R., Paz-Kagan, T., Dawson, T., Nydick, K., Stephenson, N.L., 2018. Remote measurement of canopy water content in giant sequoias (Sequoiadendron giganteum) during drought. For. Ecol. Manag. 419-420, 279–290.
Mckay, M.D., Beckman, R.J., Conover, W.J., 2000. A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 42, 55–61.
Montavon, G., Orr, G., Müller, K.-R., 2012. Neural Networks: Tricks of the Trade, vol. 7700. Springer.
Moreno-Martínez, Á., Izquierdo-Verdiguier, E., Maneta, M.P., Camps-Valls, G., Robinson, N., Muñoz-Marí, J., Sedano, F., Clinton, N., Running, S.W., 2020. Multispectral high resolution sensor fusion for smoothing and gap-filling in the cloud. Remote Sens. Environ. 247, 111901.
Pearce, T., Leibfried, F., Brintrup, A., 2020. Uncertainty in neural networks: Approximately bayesian ensembling. In: International Conference on Artificial Intelligence and Statistics, pp. 234–244. Proceedings of Machine Learning Research.
Peng, J., Muller, J.-P., Blessing, S., Giering, R., Danne, O., Gobron, N., Kharbouche, S., Ludwig, R., Müller, B., Leng, G., You, Q., Duan, Z., Dadson, S., 2019. Can we use satellite-based FAPAR to detect drought? Sensors 19.
Piotrowski, A.P., Napiorkowski, J.J., Piotrowska, A.E., 2020. Impact of deep learning-based dropout on shallow neural networks applied to stream temperature modelling. Earth Sci. Rev. 201, 103076.
Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven earth system science. Nature 566, 195–204.
Spence, T., Townshend, J., 1996. The Global Climate Observing System (GCOS). In: Long-Term Climate Monitoring by the Global Climate Observing System. Springer, pp. 1–4.
Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. 15, 1929–1958.
Sun, Y., Lu, L., Liu, Y., 2021. Inversion of the leaf area index of rice fields using vegetation isoline patterns considering the fraction of vegetation cover. Int. J. Remote Sens. 42, 1688–1712.
Székely, G.J., Rizzo, M.L., Bakirov, N.K., 2007. Measuring and testing dependence by correlation of distances. Ann. Stat. 35, 2769–2794.
Verger, A., Martínez, B., Coca, F.C., García-Haro, F.J., 2009. Accuracy assessment of fraction of vegetation cover and leaf area index estimates from pragmatic methods in a cropland area. Int. J. Remote Sens. 30, 2685–2704.
Verger, A.B., Baret, F., Camacho, F., 2011. Optimal modalities for radiative transfer-neural network estimation of canopy biophysical characteristics: evaluation over an agricultural area with CHRIS/PROBA observations. Remote Sens. Environ. 115, 415–426.
Verhoef, W., 1984. Light scattering by leaf layers with application to canopy reflectance modeling: the SAIL model. Remote Sens. Environ. 16, 125–141.
Verrelst, J., Muñoz, J., Alonso, L., Delegido, J., Rivera, J., 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.
Verrelst, J., Camps-Valls, G., Muñoz-Marí, J., Rivera, J.P., Veroustraete, F., Clevers, J.G., Moreno, J., 2015. Optical remote sensing and the retrieval of terrestrial vegetation bio-geophysical properties – a review. ISPRS J. Photogramm. Remote Sens. 108, 273–290.
Walthall, C., Dulaney, W., Anderson, M., Norman, J., Fang, H., Liang, S., 2004. A comparison of empirical and neural network approaches for estimating corn and soybean leaf area index from Landsat ETM+ imagery. Remote Sens. Environ. 92, 465–474.
Weiss, M., Baret, F., 2016. S2ToolBox Level 2 Products: LAI, FAPAR, FCOVER.