Physics-constrained deep learning for biophysical parameter retrieval from Sentinel-2 images: inversion of the PROSAIL model

Yoël Zérah, Silvia Valero, Jordi Inglada

Physics-constrained deep learning for biophysical parameter retrieval from Sentinel-2 images: inversion of the PROSAIL model

Yoël Zérah[1,*], Silvia Valero[1], Jordi Inglada[1]

[a]CESBIO, CNES/CNRS/INRAE/IRD/UT3 - Université de Toulouse, 18 avenue Edouard Belin, 31401, Toulouse CEDEX 9, France

Abstract

In this era of global warming, the regular and accurate mapping of vegetation conditions is essential for monitoring ecosystems, climate sustainability and biodiversity. In this context, this work proposes a physics-guided data-driven methodology to invert Radiative Transfer Models for the retrieval of vegetation biophysical variables. A hybrid paradigm is proposed by incorporating the physical model to be inverted into the design of a neural network architecture, which is trained by exploiting unlabeled satellite images. In this study, we show how the proposed strategy allows the simultaneous probabilistic inversion of all input PROSAIL model parameters by exploiting Sentinel-2 images. The interest of the proposed self-supervised learning strategy is corroborated by showing the limitations of existing simulation-trained machine learning algorithms. Results are assessed on LAI and CCC in-situ measurements collected on 4 different field campaigns over three European tests sites. Prediction accuracies are compared with performances reached by the well-established Simplified Level 2 Product Prototype Processor (SL2P). Obtained overall accuracies corroborate that the proposed methodology achieves performances equivalent to or better than the state-of-the-art methods.

Keywords: Sentinel-2, Radiative transfer modeling, PROSAIL, Model inversion, Simulation, Bayesian physics-guided learning

1. Introduction

Climate change and human activities affect ecosystems at an unprecedented magnitude. Monitoring vegetation traits at large scale and precise resolution is essential to maintain crop yield in a context of global population growth (Weiss et al., 2020), as it could help both to adapt agriculture practices and to react to increasingly frequent extreme events that threaten to disrupt food supply (Raj et al., 2022; Balasundram et al., 2023).

The regular and accurate mapping of vegetation biophysical variables (BV) on a global scale is thus becoming increasingly important in environmental monitoring. The international community has confirmed a strong consensus on the need of high-resolution BV maps at sub-weekly temporal frequency, and the Copernicus Land Monitoring Service identified it as a priority. These variables are essential to understand the vital role of biodiversity and ecosystem services. They include variables such as the leaf area index (LAI) and the canopy chlorophyll content (CCC), which are respectively defined as half the leaf surface area per unit ground area, and the "total amount of chlorophyll *a* and *b* pigments per unit ground area" (Gitelson et al., 2005). The LAI is a critical attribute of canopy involved in processes such as photosynthesis (Duncan, 1971), land evaporation (Kergoat et al., 2002; Wang et al., 2014), precipitation interception (Boussetta et al., 2013), and can be used to estimate crop yield (Mokhtari et al., 2018). The CCC, which is highlighted as an essential climate variable (ECV) by the Global Climat Observing System (GCOS) (GCOS, 2011), has an important role to quantify vegetation health, carbon and water fluxes, and productivity (Ali et al., 2020).

Quantifying vegetation BV from satellite imagery is a well-known problem in remote sensing (Zheng and Moskal, 2009). As reviewed in Verrelst et al. (2019); Fang et al. (2019), retrieval methods can be broadly classified into three main categories: (i) parametric (ii) non-parametric and (iii) physically-based model inversion methods.

Parametric methods are model fitting strategies assuming that it exists an explicit relationship between spectral observations and BV. Traditionally, they build explicit parameterized functions relating a limited number of spectral bands (or vegetation indices) with the variables of interest. Despite being a widely-used strategy, these methods suffer several disadvantages. One of the main limitations is that their robustness depends on the properties of the required in-situ data (i.e., number, quality and representativeness) used to fit simple mathematical functions. Besides, they do not exploit full-spectrum information and mostly combine data features in a linear way. Therefore, these methods are not reliable given their poor transferability performances to other sites with different types of vegetation (Heiskanen et al., 2012).

Non-parametric methods do not assume any information about fitting functions, spectral band relationships, correlation between variables or data distributions. In this category, most of the existing works propose machine learning regression methodologies. These algorithms build a regression model through a learning phase based on a high number of training samples. The most popular algorithms are decision trees (Houborg and McCabe, 2018; Srinet et al., 2019), artificial neural network (Weiss and Baret, 2016; Danson et al., 2003) and kernel methods (Tuia et al., 2011; Panigrahi and Das, 2021; Durbha et al., 2007). These methodologies have important strengths : (i) their ability to capture non-linear relationships of full data spectrum without explicitly knowing the underlying data distribution, (ii) to cope well with datasets showing redundancy and high noise levels and (iii) efficient data processing times once the model is trained.

Despite the multiple advantages offered by these methods, they have important weaknesses for being suitable candidates for operational retrieval processing. A well-known limitation of these black-box methodologies is that the accurate models need to be trained with high number of labeled in-situ samples, which is difficult or even impossible. Besides, training samples should represent the real world data as much as possible since algorithms do not generalize well if training and testing data have deviating statistics.

To avoid performance degradations due to label scarcity and diverse data distributions, physically-based retrieval methods propose the inversion of radiative transfer models (RTMs). These physical models allow to simulate canopy reflectances from biophysical variables. The inversion of the well-known RTM PROSAIL (Jacquemoud et al., 2009; Zhu et al., 2018; Darvishzadeh et al., 2012) is the most established approach. Considering a high number of combinations of RTM input values, these hybrid methods generate simulated spectral reflectance scenarios. These simulations enable the creation of look-up-tables (LUT) containing a large set of simulated vegetation reflectance spectra, (Duan et al., 2014; Hauser et al., 2021). For these approaches, the inversion problem is solved by finding the close match between the simulated spectra in the LUT database and the observed spectrum. Different similarity cost functions are proposed to identify the most similar sample or the set of samples (Rivera et al., 2013; Marie Weiss et al., 2000; Verrelst et al., 2015). Unfortunately, these ap-

⋆This work was supported by the Natural Intelligence Toulouse Institute (ANITI) from the Université Fédérale Toulouse Midi-Pyrenées under Grant Agreement ANITI ANR-19-P3IA-0004, by the ANR-JCJC DeepChange project under Grant Agreement 20-CE23-0003, and by the Centre national d'études spatiales (CNES) under Grant Agreement n° 51/19560.

*Corresponding author

Email addresses: yoel.zerah@univ-toulouse.fr (Yoël Zérah), silvia.valero@iut-tlse3.fr (Silvia Valero), jordi.inglada@cesbio.eu (Jordi Inglada)

proaches are computationally expensive due to the iterative calls of LUT entries on a per-pixel basis making the retrieval of variables unfeasible for large areas.

The most recent works propose to combine the advantages of physically-based approaches with non-parametric methods (Verrelst et al., 2015; Svendsen et al., 2018). These hybrid inversion strategies train machine learning regression algorithms by using artificial data generated by RTM simulations. The problem of training data availability is thus avoided and the model inversion process is computationally efficient. Different machine learning algorithms are proposed by hybrid methodologies aiming the inversion of PROSAIL : artificial neural networks (ANNs) (Baret et al., 2007), random forests (RF) (Campos-Taberner et al., 2018), support vector machine regression (Durbha et al., 2007), Gaussian processes (GPs) Camps-Valls et al. (2016); Svendsen et al. (2018), etc. The most well-known method is the biophysical variable neural network (BVNET) (Marie Weiss et al., 2002), which proposes the inversion of the RTM by training an ANN on PROSAIL simulated data. The satisfactory results obtained by BVNET explain why this hybrid approach is implemented on the SL2P algorithm of the Sentinel Application Platform (SNAP) Biophysical Processor (BP) (Weiss and Baret, 2016) and Sen2-Agri ToolBox (Defourny et al., 2019).

Despite hybrid strategies are very promising, they do not alleviate some of the main issues of the ill-posedness inversion problem (Combal et al., 2003; Atzberger, 2004). The main shortcoming is the lack of realism of physical simulations which are not representative of real-world data. As a result, poor generalization performances and inaccurate predictions can be obtained by synthetic unbalanced and biased training datasets. Besides, the choice of the imperfect RTM to be inverted (epistemic uncertainty) also affects the performances of these methods.

These limitations are recently highlighted by several studies (Kamenova and Dimitrov, 2021; Xie et al., 2019) where the poor performance of SL2P over heterogeneous canopies such as forests are reported (Xie et al., 2019; Brown et al., 2021b). In it, the low SL2P accuracies are justified by the inability of PROSAIL to model non-agricultural vegetation types of reflectances.

The main challenge of hybrid methodologies is that the generation of accurate simulations reflecting the characteristics of real-world vegetation spectra requires a lot of prior knowledge from the physical process. Simulation input modeling is a hard complex task for experts which requires: (i) the identification and the choice of the input data distribution, (ii) the selection of the interval range of the data and, (iii) the recognition of the correlation or dependency between variables.

Despite the large number of existing works, the modeling of the ensemble of BV to generate RTM simulations is still limited. In practice, the distribution of BV is available only for some vegetation types in a few geographical spatial and temporal conditions. Furthermore, correlations between BV are not well established and most of the time the dependency between variables is arbitrarily imposed.

Recently, some solutions to reduce the impact of real observations and simulated match-mismatch distributions have been proposed. The work in Gallo et al. (2023) proposes to train a neural network in a self-supervised manner to build a feature extraction encoder from hyperspectral data. Then, the pre-trained encoder is fine-tuned to perform a supervised regression retrieval BV task by using a simulated training data-set of moderate size.

The work in Svendsen et al. (2021) proposes a different hybrid strategy to combine machine learning algorithms and physical models for the inversion of PROSAIL. In it, two different variational inference strategies relying on Monte Carlo expectation maximization (MCEM) and variational autoencoder (VAE) are proposed to retrieve three BV from Landsat-8 data. The first strength of this work is that both methods allow to measure the uncertainty of the predictions by estimating BV probability distributions. It must be noticed that the quantification of the uncertainties associated to the estimated BV is a well-known challenge in the literature Brown et al. (2021b). Only a few solutions based on GP regression algorithms are presented by some works (Verrelst et al., 2013; Estévez et al., 2021).

Another advantage presented in Svendsen et al. (2021) is that a physics-based guidance is incorporated in to the training process avoiding simulation-assisted learning. Unfortunately, the benefits of training without synthetic simulations is not corroborated in the study since only simulated data are considered for training and testing purposes. Another limitation of the work is that PROSAIL is used as a non-differentiable decoder into a VAE architecture design. As a consequence, an end-to-end learning procedure can not be performed. Besides, the computational burden is a serious practical impediment for sampling intensive methods such MCEM or Markov Chain Monte Carlo (MCMC) (Wang et al., 2022), which can not be deployed for large scale operational variable retrieval.

Under this framework, this work aims to propose a physics-driven probabilistic deep learning method for the inversion of PROSAIL. Based on the work presented in Zérah et al. (2023), our methodology proposes to incorporate a differentiable PROSAIL model in a deep learning architecture. Then, an end-to-end training strategy is presented for the simultaneously retrieval of all the input PROSAIL parameters. There are some works that propose the simultaneous estimation of a few BV (Tuia et al., 2011; Atzberger, 2004; Darvishzadeh et al., 2008; Sehgal et al., 2016; Zhu et al., 2023). However, it must be noticed that most of the proposed methods focus on a single variable at once (typically the LAI and other related indices, or the chlorophyll content) and operational methods such as SL2P train independent ANN for predicting LAI, CCC and canopy water content (CWC).

Besides the multi-parameter inversion purpose, the proposed methodology, denoted as PROSAIL-VAE, presents other important contributions: (i) to assist the training of advanced regression algorithms by using real Sentinel-2 data (without simulations), (ii) to incorporate prior information about physical variables to constraint the retrieval estimation and (iii) to exploit spatial context information in the retrieval process.

Different experiments are carried out to corroborate the interest of the proposed PROSAIL-VAE. First, the limitations of a simulation-based learning strategy are analyzed by considering different synthetic training data scenarios. The experiments evaluate how different choices (prior BV distributions, correlations between parameters and RTM) used to generate the simulations can limit the transferability of the regression algorithms to real-world situations. Secondly, different hybrid PROSAIL-VAE architecture designs and training cost functions are evaluated and compared with the well-established SL2P processor.

The article is organized as follows. Section 2 presents the proposed methodology. Section 3 describes the three types of data-sets considered in this work (synthetic PROSAIL simulations, Sentinel-2 satellite observations and in-situ data measurements). Section 4 presents the above described experiments and the obtained results are discussed in Section 5. Then, Section 6 contains concluding remarks.

2. Methodology

2.1. PROSAIL

The combined PROSPECT leaf optical properties model (Jacquemoud and Baret, 1990; Feret et al., 2008; Feret and de Boissieu, 2023) and SAIL canopy bidirectional reflectance model (Verhoef, 1984), known as the PROSAIL RTM (Jacquemoud et al., 2009) that simulates canopy reflectances from 400 to 2500 nm. PROSAIL links the chemical properties of leaves with canopy structural properties (e.g. the LAI) to simulate canopy reflectance for a given observation geometry (with the relative directions of the Sun and sensor from the observed surface). Scattering by Arbitrary Inclined Leaves (SAIL) is a turbid medium model, that assumes that the canopy is made of parallel transparent layers between the atmosphere and the soil. The model is well-suited for dense canopies with small (or even infinitesimal) leaves such as grasslands, agricultural crops, and some forests.

The first version of PROSPECT simulates leaf spectra at 5 nm resolution from three parameters: a leaf structure parameter, a water content and a leaf pigment content (Jacquemoud and Baret, 1990). The next early PROSPECT versions improve the spectral resolution to 1 nm, and introduce the spectral contribution of the dry matter contained in leaf cell wall molecules (Jacquemoud et al., 2000). All photosynthetic pigments in PROSPECT-4 are assumed to be chlorophyll, whereas PROSPECT-5 differentiates chlorophylls from carotenoids (Feret et al., 2008). PROSPECT-D adds the contribution of anthocynanins (Féret et al., 2017, 2019), and PROSPECT-PRO incorporates nitrogen-based proteins and carbon-based constituents (Féret et al., 2021).

The PROSAIL version used in our study is composed of PROSPECT-5 and 4SAIL (Verhoef et al., 2007), which is a numerically optimized version of SAIL. The corresponding BV are detailed in Table Section 2.1. A differentiable PROSAIL version has been implemented to allow the propagation of gradients through a hybrid deep learning architecture incorporating the physical model. Specifically, the leaf inclination distribution function (LIDF) (Wang et al., 2007) in SAIL considered is the Campbell's ellipsoidal LIDF (Campbell, 1990) controlled by the mean leaf angle α, instead of other distributions such as Verhoef's LIDF (Verhoef).

Table 1: PROSAIL input parameters

Model	Input	Description	Unit
PROSPECT-5	N	Leaf structure parameter	-
	C_{ab}	Chlorophyll $a+b$ content	$\mu g\,cm^{-2}$
	C_w	Water equivalent thickness	cm
	C_c	Carotenoid concentration	$g\,cm^{-2}$
	C_m	Dry matter content	$\mu g\,cm^{-2}$
	C_b	Brown pigments content	-
4SAIL	LAI	Leaf Area Index	-
	α	Mean leaf angle	deg
	h	Hotspot parameter	-
	ρ_S	Wet soil factor	-
	r_S	Soil brightness factor	-
Geometry	θ_S	Solar zenith angle	deg
	θ_O	Observer zenith angle	deg
	ψ_{SO}	Relative azimuth angle	deg

Considering as a starting point the code available on (Domenzain et al., 2019), an implementation of PROSAIL in Pytorch compatible with automatic differentiation and allowing the generation of simulations in parallelized batches has been developed. Finally, the simulated PROSAIL reflectance values are converted to Sentinel-2 bands by applying the band-specific response functions of the sensor (see Appendix Appendix A for more details).

2.2. Variational Autoencoder

Auto-Encoders (AE) are symmetric encoder-decoder neural network models trained in a self-supervised manner to infer embedding representations from input data. Deterministic embeddings are outputted by the encoder and the decoder part uses them to match the reconstruction with the input data.

Introduced by (Kingma and Welling, 2014), Variational Auto-Encoders (VAE) incorporate Bayesian theory with classical AE architectures to solve stochastic inference problems. Assuming that the input data $\mathbf{x}$ is generated from an ensemble of probabilistic embeddings $\mathbf{z}$ (called *latent* variables), the VAE's goal is to retrieve the posterior distributions $p(\mathbf{z}|\mathbf{x})$. Because the joint distribution of the input data $p(\mathbf{x})$ is untractable, variational inference is used to retrieve $p(\mathbf{z}|\mathbf{x})$, by approximating them with distributions $q(\mathbf{z}|\mathbf{x})$.

The encoder (resp. the decoder) corresponds to $q_\phi(\mathbf{z}|\mathbf{x})$, (resp. $p_\theta(\mathbf{x}|\mathbf{z})$), with ϕ (resp. θ) the parameters to be optimized. Both $q_\phi(\mathbf{z}|\mathbf{x})$ and $p_\theta(\mathbf{x}|\mathbf{z})$ are commonly chosen as isotropic Gaussian distributions.

For a given input $\mathbf{x}$, the encoder learns the distribution parameters $\mu_\phi(\mathbf{z})$ and $\sigma_\phi(\mathbf{z})$. The distributions are sampled from the latent space and injected through the decoder which outputs the distribution parameters $\mu_\theta(\mathbf{x})$ and $\sigma_\theta(\mathbf{x})$, from which reconstructions can be sampled. Latent variables are sampled by considering the *reparametrization trick*, which enables the back-propagation during the training process minimizing the following objective function:

$$\underbrace{-\mathbb{E}_{z\sim q_\phi(\mathbf{z}|\mathbf{x})}\left[p_\theta(\mathbf{x}|\mathbf{z})\right]}_{\text{Reconstruction term}} + \underbrace{\mathbb{KL}\left[q_\phi(\mathbf{z}|\mathbf{x})\|p(\mathbf{z})\right]}_{\text{Regularization term}}. \quad (1)$$

The loss reconstruction term incites the VAE to learn reconstructions matching the input data $\mathbf{x}$. This term is known as the expected negative log-likelihood and, under the Gaussian assumption, it can be written as:

$$\mathbb{E}_{z\sim q_\phi(\mathbf{z}|\mathbf{x})}\left[p_\theta(\mathbf{x}|\mathbf{z})\right] = \frac{1}{2}\sum_i\left[\log 2\pi\sigma^2_{\theta,i}(\mathbf{z}) + \frac{(\mathbf{x}_i-\mu_{\theta,i}(\mathbf{z}))^2}{\sigma^2_{\theta,i}(\mathbf{z})}\right], \quad (2)$$

with i indexing the feature dimension of the input data.

The regularization term is the Kullback-Leibler divergence (KLD) between the approximate posteriors $q_\phi(\mathbf{z}|\mathbf{x})$ and the prior distributions $p(\mathbf{z})$, that are usually chosen as standard Gaussian distributions. This term encourages the matching of latent and prior distributions.

2.3. PROSAIL-VAE

PROSAIL-VAE is defined as a VAE in which PROSAIL is integrated as a physics-based deterministic differentiable decoder (see Fig. 1). In this architecture design, the encoder takes Sentinel-2 (S2) reflectances and observation angles as input, and outputs the parameters of the marginal distribution of each PROSAIL variable (i.e. the latent distributions correspond to PROSAIL input parameter distributions). Once trained, the encoder is thus a fast, probabilistic inverser of PROSAIL.

During the training phase, samples from latent BV distributions are drawn and forwarded to the PROSAIL model. The physical decoder then simulates the corresponding canopy spectra and the resulting reflectances are converted to Sentinel-2 bands by applying the band-specific response functions of the sensor. Although only the encoder is trained during the end-to-end learning process, our differentiable implementation of PROSAIL allows gradient back-propagation to be performed through the whole architecture (from the loss term, upstream towards the encoder inputs).

As mentioned above, the distributions inferred by the encoder correspond to the physical PROSAIL input variables. Therefore, the learned variables are constrained to ensure accurate and realistic PROSAIL simulations. This is achieved by forcing latent variable distributions to be within a specified range. Specifically, sampled values are constrained to BV definition domains by modeling their distributions as two-sided truncated Normals (TN). The specific range (see Table 2) is imposed by re-scaling the output encoder distributions, which are previously constrained to the range $[0, 1]$.

PROSAIL-VAE is optimized by minimizing the VAE loss function described in Eq. 1. In our case, the KLD loss term encourages the inferred TN distributions to match prior latent distributions corresponding to the ensemble of BV. Uniform distributions over their definition ranges (see Table 2) are considered as BV priors. The work in (Zérah et al., 2023) gives more information about the choice and sampling of meaningful latent distributions, the KLD distance between TN and uniform distributions and the computation of the decoder's distribution with Monte Carlo (MC) sampling.

To balance reconstruction and KLD loss terms, the incorporation of a β_{KL} weighting coefficient (see Eq. 3) is proposed (Higgins et al., 2017). The influence and the impact of KLD loss terms is studied in Section 4 through different scenarios by considering : (i) $\beta_{KL} = 0$, (ii) a KLD loss term computed on all PROSAIL BV and (iii) the selection of a subset of the variables (LAI and C_{ab}) for KLD term computation.

$$\mathcal{L}_{\text{PROSAIL-VAE}} = \mathcal{L}_{rec} + \beta_{KL}\mathcal{L}_{KL}. \quad (3)$$

The two different PROSAIL-VAE encoder architectures described in Fig. 2 are proposed. Both are based on a residual network backbone, with 3 residual connection blocks, and they infer BV for each pixel of the input images. The first architecture is a pixel-wise multi-layer perceptron (MLP) which handles independently each image pixel. The second proposed encoder is defined as a convolutional neural network (CNN) and it captures the spatial context information.

2.4. Supervised regression strategies

Besides the proposed self-supervised PROSAIL-VAE, two simulation-assisted learning regression methods are also studied for the inversion of PROSAIL. The first strategy is the well-known BVNET, which is integrated as a SL2P tool in SNAP's BP (Weiss and Baret, 2016). Secondly, a multiple probabilistic supervised regression approach is presented by considering the proposed pixel-wise probabilistic encoder of Section 2.3 (see Fig. 2).

2.4.1. BVNET

BVNET is a very simple two-layered neural network architecture predicting single biophysical variables (LAI, CCC, CWC, fraction of vegetation cover (F-COVER) or fraction of absorbed photosynthetically active radiation (FAPAR)). Composed by 66 weights, the network is trained by using a classical mean squared error (MSE) loss function. Satellite reflectances and observation angles are considered as input data.

In this study, the performance of BVNET is assessed by two different configurations. Firstly, the transferability of BVNET is evaluated by training several models using different simulated data-sets scenarios (Section 3.1). Secondly, the performance of predictions obtained by the pre-trained BVNET available in SL2P is studied in Section 4.3. The studied pre-trained model was originally trained by using PROSPECT-3+SAILH simulations and it predicts LAI and CCC variables.

Table 2: Range of the PROSAIL input variables in PROSAIL-VAE

Variable	N	C_{ab}	C_w	C_c	C_m	C_b	LAI	α	h	ρ_S	r_S
min	1.2	20.0	0.0075	5	0.003	0	0	30	0.0	0	0.3
max	1.8	90.0	0.0750	23	0.011	2	10	80	0.5	1	3.5

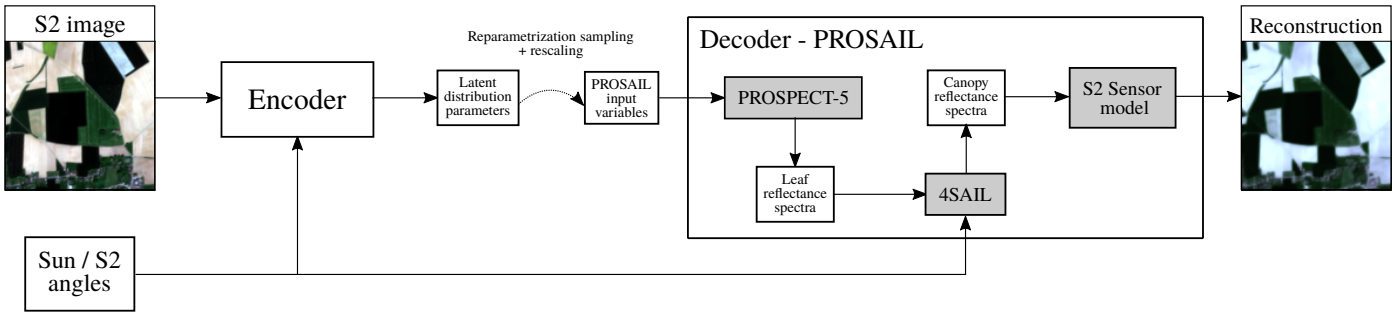


Figure 1: Description of the proposed PROSAIL-VAE methodology

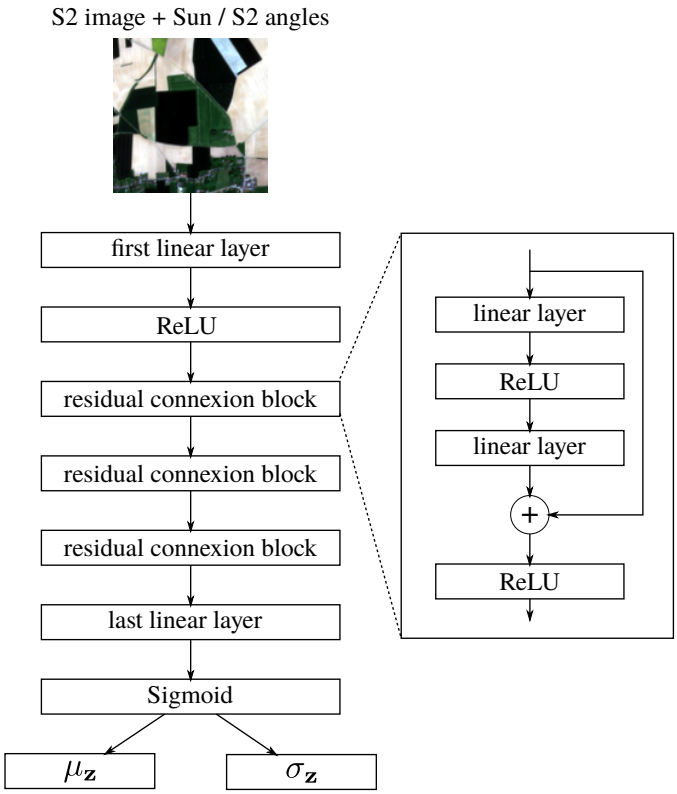


Figure 2: Backbone of the encoder of PROSAIL-VAE, with a first input layer followed by 3 blocks of 2 layers with skip connections, and a last layer that outputs the parameters of the distribution of PROSAIL variables. For the pixel-wise version of the encoder, all linear layers are dense layers of size 32. For the spatial version, all linear layers are 2D convolutional layers with size 32 and stride 1. The first layer has a filter size of 3×3, whereas the size of the rest of the layers is 1×1.

2.4.2. Multiple probabilistic supervised regression (MPSR)

The neural network architecture of the pixel-wise PROSAIL-VAE encoder (see Section 2.3 and Fig. 2) is proposed to solve the supervised regression problem. It must be noticed that the pixel-wise encoder is considered because simulated training data-sets composed of realistic images cannot be generated by PROSAIL. Compared to BVNET, this strategy has several advantages: (i) more complex relationships can be discovered by increasing the complexity of the neural network and (ii) a multiple probabilistic prediction of BV can be performed by training a single model. In this case, the supervised training is carried out by using the Negative Log-Likelihood of Truncated Gaussians described in (Zérah et al., 2023). Thus, the predicted distribution parameters of all BV are compared with the simulated data-set values at each training step.

2.5. Evaluation metrics

Assuming a set of n reference measurements $(x_i)_{i\in\{1,\dots n\}}$ and their corresponding predictions $(\hat{x}_i)_{i\in\{1,\dots n\}}$, the root mean squared error (RMSE) is used to evaluate the accuracy of the predictive models:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2}$$

Probabilistic predictions are also evaluated by studying the uncertainties associated to the prediction intervals of each reference measurement. The prediction interval $\left[l_{\hat{x}_i}, u_{\hat{x}_i}\right]$ estimates in what range a future individual observation will fall, with probability α. In this study, α is set to 0.95 (corresponding to a 2σ rule) and the quantitative evaluation is carried out by:

- the mean prediction interval width (MPIW), that quantifies the confidence of the model in its estimation.

$$\text{MPIW} = \frac{1}{n}\sum_{i=1}^{n}\left(u_{\hat{x}_i} - l_{\hat{x}_i}\right)$$

- the prediction interval coverage probability (PICP) that corresponds to the ratio between the number of measurements that are inside their prediction intervals, over the total number of measurements.

3. Data-sets

PROSAIL simulations, Sentinel-2 satellite data and in-situ measurements are used in our experiments. Simulations correspond to a set of data pairs constructed from PROSAIL input variables and corresponding outputs. The supervised BVNET and MPSR strategies described in Section 2.4 are trained with this synthetic data. The Sentinel-2 satellite data is used to train the proposed self-supervised PROSAIL-VAE methodologies. Finally, the in-situ data-set, which integrates measurements from three different field campaigns, is considered to assess the predictions of the trained models.

3.1. Simulated data-sets

A simulated data-set is composed of a set of S2 reflectances derived from the PROSAIL simulations with their corresponding input parameters. Different design decisions are considered for the simulation input modeling process: (i) the choice of the input data distributions and their interval ranges, (ii) the correlation or dependency between variables and (iii) the selection of PROSAIL model version. As a result, several synthetic data-sets are built and used in our experiments.

1. *Ranges and distributions of PROSAIL input parameters.* The description of the input PROSAIL parameters used for the simulation process are described in Table 3. These distributions are mostly identical to that of the training data-set of SL2P (Weiss and Baret, 2016). Only few differences can be found for some variables since PROSAIL model versions chosen in this study differ from the one used to pre-train the SL2P toolbox (see Section 2.4.1). For instance, instead of using a reference soil spectra database, the soil is made up of the linear combination of two spectra, dry and wet, respectively in our PROSAIL simulations. The wet soil spectrum is controlled by a wet soil factor coefficient ρ_S, whose distribution is set to uniform over its definition range. Equivalent water thickness C_w distribution parameters are derived from relative water content ($C_{w,\mathrm{rel}}$) specifications previously defined in Weiss and Baret (2016) by considering the relation $C_{w,\mathrm{rel}} = \frac{C_w}{1-C_m}$. Also, the observation angles are simulated with S2 orbital characteristics, by uniformly drawing random locations between -56° and 83° latitudes, like performed in Weiss and Baret (2016).
2. *Correlation between input PROSAIL variables and LAI.* Two procedures are studied to transform the independently sampled input PROSAIL parameters V into variables $V^\star$ correlated with LAI. Both define the correlated variable $V^\star$ as an affine transformation of the LAI. The first strategy referred as type 1 is proposed in Inglada (2017) and it is described by Eq. 4. It uses two parameters: the mode μ_V of the PROSAIL variable distribution, and a scaling constant C_{LAI} (see Table 3).

$$V^\star(\mathrm{LAI}) = \mu_V + (V - \mu_V)\left(1 - \frac{\min(\mathrm{LAI}, C_{\mathrm{LAI}})}{C_{\mathrm{LAI}}}\right) \quad (4)$$

   The second procedure presented in Weiss and Baret (2016) and referred as type 2 is denoted by Eq. 5. This method uses 4 auxiliary variables $V_{l,0}$, $V_{u,0}$, $V_{l,M}$ and $V_{u,M}$[1] for each PROSAIL input variable (see Table 3). $V_{l,0}$ and $V_{u,0}$ are the lower and upper bounds of the interval of the independently sampled variable V, and $V_{l,M}$ and $V_{u,M}$ are the lower and upper bounds of the correlated variable $V^\star$ at $\mathrm{LAI} = \mathrm{LAI}_{\max} = 15$.

$$V^\star(\mathrm{LAI}) = \frac{(V - V_{l,0})\,(v_u(\mathrm{LAI}) - v_l(\mathrm{LAI}))}{V_{u,0} - V_{l,0}} + V_{l,M}(\mathrm{LAI}) \quad (5)$$

   with the functions

$$v_l(\mathrm{LAI}) = V_{l,0} + \mathrm{LAI}\,(V_{l,M} - V_{l,0})$$
$$v_u(\mathrm{LAI}) = V_{u,0} + \mathrm{LAI}\,(V_{u,M} - V_{u,0}).$$

3. *PROSAIL models.* Two versions are evaluated by combining 4SAIL with PROSPECT-5 (or PROSPECT-D) versions. The PROSPECT-D model is created by adding a new parameter (anthocyanins) to PROSPECT-5 model. As a result, this model is able to to simulate leaf optical properties through a complete lifecycle, including juvenile and senescent stages, as well as environmental stresses.

3.2. Sentinel-2 patch data-set

To train PROSAIL-VAE, a data-set of S2 image patches is built. The fourteen western Europe S2 Military Grid Reference System (MGRS) tiles shown in Fig. 3 and freely available in the THEIA catalogue[2] are considered in this data-set. They are orthorectified, terrain-flattened and atmospherically corrected with the MAJA processor (Rouquié et al., 2017; Hagolle et al., 2017). Only spectral bands acquired at 10 and 20 m spatial resolutions are considered and 20 m bands are upsampled using cubic interpolation to 10 m. For each pixel,

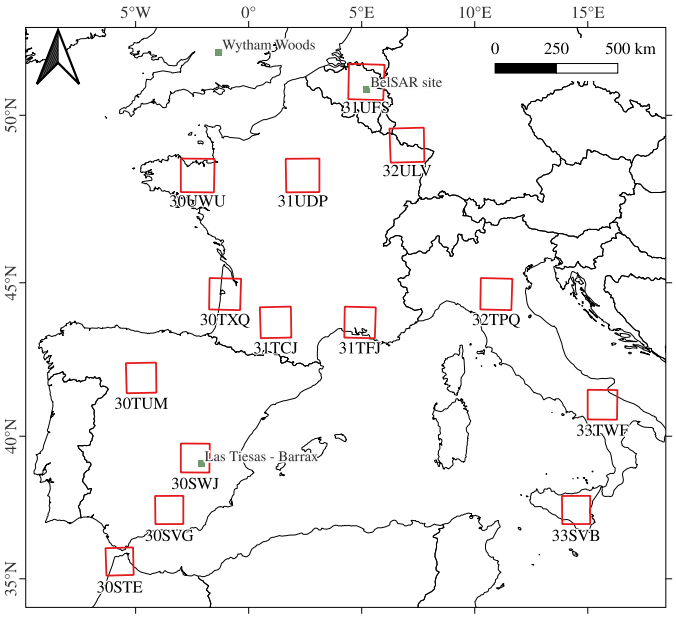


Figure 3: The selected 14 MGRS tiles. They include the location of 2 of the 3 the test sites where in-situ measurements were collected (Las Tiesas-Barrax and the Belsar Site).

the three angles defining its observation geometry are also taken into account: the satellite and solar zenith angles at the Earth surface and the relative azimuth angle between the sun-Earth and Earth-satellite directions. For each tile, 1–3 different regions of interest (ROIs), of size 5120 × 5120 m describing croplands and forest areas are selected (see Table 4). The resulting ROIs do not overlap test sites where in-situ measurements described in Section 3.3 were collected. For each tile, multiple acquisitions with dates ranging between January 2016 and December 2019 are considered (see Fig. 4). Within chosen S2 acquisitions, when the selected ROIs are observed (the cloud coverage is below 3%), the corresponding 512 × 512 pixels patches are extracted.

As depicted in Fig. 5, each 512 × 512 ROI patch is spatially split into 16 disjoint patches of size 128 × 128 10 m pixels: 14 for training, 1 for validation, and 1 for testing. Any 128 × 128 patch with invalid pixels (due to clouds) is discarded. The training data-set is used for PROSAIL-VAE model training, the validation data-set is used to monitor the loss during training and ensure that the model does not over-fit. The testing data-set is used to assess the performances of the trained models[3]. Specifically, the testing data is used to assess reconstructions and parameter inference on unseen samples (see Section 4 and Appendix Fig. B.22). The number of patches for each data-set is described in Table 4.

3.3. In-situ data collection

The data-set is composed of direct measurements collected in different field campaigns under the framework of fiducial reference measurements for vegetation (FRM4Veg) and BelSAR projects. The complete data-set contains 211 LAI and 121 CCC reference measurements. This data serves as a reliable reference for quantitatively evaluating the accuracy of the biophysical variable predictions.

3.3.1. *fiducial reference measurements for vegetation (FRM4Veg)*

FRM4VEG is a European Space Agency (ESA) managed project focused on establishing the protocols required for traceable in-situ measurements of vegetation-related parameters, to support the validation of Copernicus products (Origo et al., 2020). In this project, different field campaigns have been performed over two test sites covering agricultural crops (Las Tiesas-Barrax,

[1] Indices notations are: u for "upper bound", l for "lower bound", 0 for "at LAI=0", M for "at maximum LAI value".

[2] https://www.theia-land.fr/en/product/sentinel-2-surface-reflectance/

[3] The testing data-set doesn't intervene in either model parameters nor hyper-parameter tuning.

Table 3: Sampling distributions of PROSAIL parameters to generate the simulated data-sets.

Variable V	Distribution	Range		Distribution parameters		Co-distribution type 1 parameter	Co-distribution type 2 parameters	
		$V_{l,0}$ (min)	$V_{u,0}$ (max)	μ_V	σ_V	C_{LAI}	$V_{l,M}$	$V_{u,M}$
N	TN	1.2	2.2	1.5	0.3	10	1.3	1.8
C_{ab}	TN	20	90	45	30	10	45	90
C_w	TN	0.0075	0.0750	0.025	0.020	10	0.015	0.055
C_c	TN	5	23	11	5	-	-	-
C_m	TN	0.003	0.011	0.005	0.005	10	0.005	0.0110
C_b	TN	0.0	2.0	0.0	0.3	10	0.0	0.2
LAI	TN	0	15	2	3	-	-	-
α	TN	30	80	60	20	10	55	65
h	TN	0.10	0.50	0.25	0.50	-	-	-
ρ_S	Uniform	0	1	-	-	-	-	-
r_S	TN	0.3	3.5	1.2	2.0	10	0.5	1.20

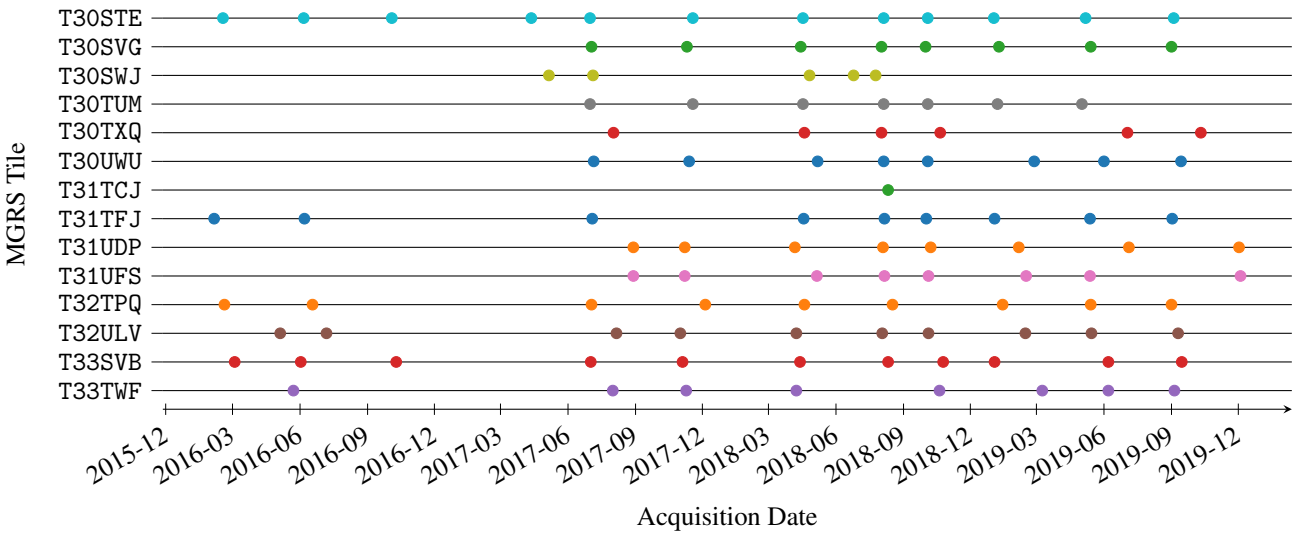


Figure 4: Dates of image acquisitions for each MGRS tile

Table 4: Patches and pixel content of training, validation and testing S2 image data-sets for PROSAIL-VAE.

S2 Tile	Acquisitions	ROIs	Training patches (32 x 32)	Training pixels	Validation patches (32 x 32)	Validation pixels	Testing patches (128 x 128)	Testing pixels	Total pixels
T30STE	13	1	2688	2752512	192	196608	12	196608	3145728
T30SVG	8	1	1568	1605632	128	131072	8	131072	1867776
T30SWJ	5	3	896	917504	64	65536	4	65536	1048576
T30TUM	7	1	1568	1605632	112	114688	7	114688	1835008
T30TXQ	6	1	1344	1376256	96	98304	6	98304	1572864
T30UWU	8	1	1792	1835008	112	114688	8	131072	2080768
T31TCJ	1	1	224	229376	16	16384	1	16384	262144
T31TFJ	9	1	2016	2064384	144	147456	9	147456	2359296
T31UDP	8	2	1792	1835008	128	131072	8	131072	2097152
T31UFS	8	2	1792	1835008	128	131072	8	131072	2097152
T32TPQ	9	1	1808	1851392	144	147456	9	147456	2146304
T32ULV	10	1	2240	2293760	160	163840	10	163840	2621440
T33SVB	11	1	2464	2523136	176	180224	10	163840	2867200
T33TWF	8	3	1792	1835008	128	131072	8	131072	2097152
Total			23984	24559616	1728	1769472	108	1769472	28098560

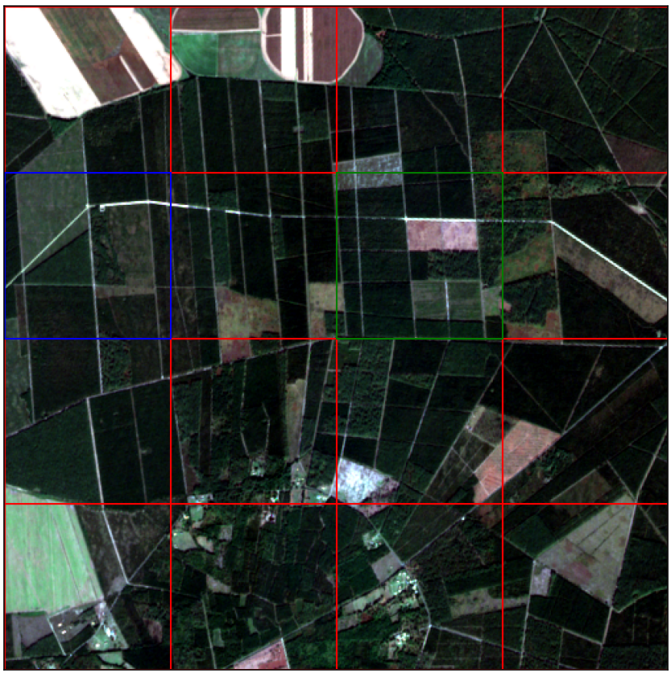


Figure 5: Example of ROI contained in T30TXQ tile (French Aquitaine) on 2019-10-11. The splitting result patches are shown in red (training), blue (validation) and green (testing).

Spain) and deciduous broadleaf forest (Wytham Woods, UK). Besides LAI, CCC and bare soil measurements, their associated uncertainties are also available for both test sites.

Considering an elementary sampling unit (ESU) of 20×20m, about 12 to 15 LAI individual measurements were performed using digital hemispheric photography (DHP). Leaf chlorophyll content (LCC) measurements were made on 13 points per ESU with a Konica Minolta SPAD-502 chlorophyll meter. Considering 3 leaves per point with 6 replicates per leaf, 234 measurements were thus performed for each ESU. The relative values provided by the SPAD-502 were converted to absolute units using calibration functions specific to each vegetation type (Origo et al., 2020). Finally, CCC measurements were obtained by applying CCC = LCC × LAI. Although the measurement of non-destructive chlorophyll can lead to imprecise and unreliable results (Zhang et al., 2022), it must be noted that measurements provided by the FRM4Veg campaigns were performed with rigorous and high standard protocols considering important number of repetitions and uncertainty estimations (Brown et al., 2021a).

In our study, the measurements collected in 2018 and 2021 over Barrax test site are used (see Fig. 6). As proposed in Brown et al. (2021b), alfalfa measurements are not considered because these crops had been thinned prior to the Sentinel acquisitions, but after the in situ measurements were made. By considering the dates of in-situ measurements, satellite images acquired on (2018-05-16, 2018-06-13, 2018-07-22) are considered for the Barrax test site.

In the case of Wytham test site, only data from 2018 is considered due to the lack of clear satellite image acquisitions over the summer of 2021. For this study area, S2 images acquired on 2018-06-29 and 2018-07-06 are used.

3.3.2. BelSAR

In the framework of the BelSAR project (Bouchat et al., 2023; Orban et al., 2021; Bouchat et al., 2022), field measurements and airborne SAR bistatic acquisitions were collected over a test site in Belgium, near the town of Gembloux during the summer of 2018. This project had the objective of assessing the interest of SAR bistatic acquisitions for vegetation and soil humidity monitoring. It also wanted to validate the capabilities of active-passive satellite configurations by ensuring the performances of L-band SAR bistatic and multistatic imagery.

In the BelSAR campaign, measurements were collected over 10 maize and 10 winter wheat fields larger than 1 hectare (ha) in size (see Fig. 8).

The BelSAR project provides plant area index (PAI) measures for wheat parcels and green area index (GAI) for maize fields. Considering that PAI and GAI are similar to LAI Fang et al. (2019), both measurements are interpreted as LAI in our study. For each field, 3 measurements were made at each date. Accordingly, the average of the measurements computed at each parcel for each

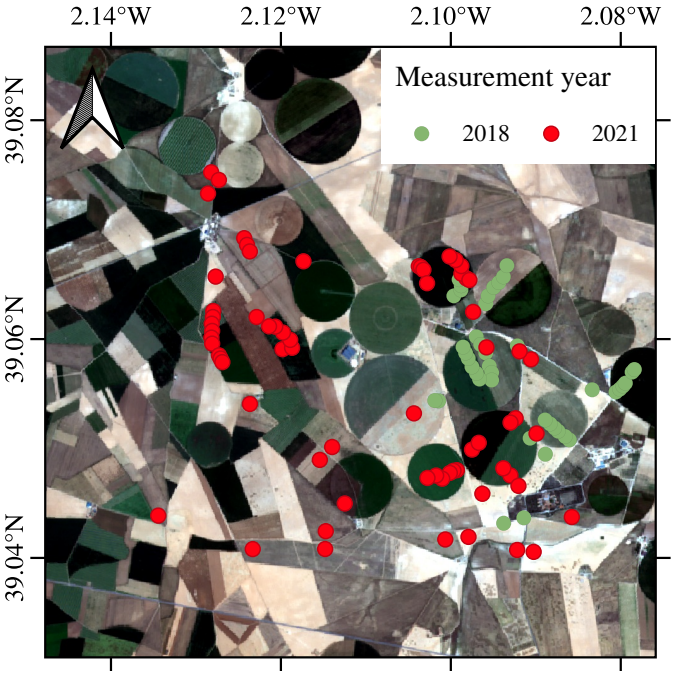


Figure 6: In-situ measurements of 2018 and 2021 FRM4Veg in Las Tiesas - Barrax test site (S2 image of 2018-06-13).

Figure 7: In-situ measurements collected over FRM4Veg Wytham area in 2018.

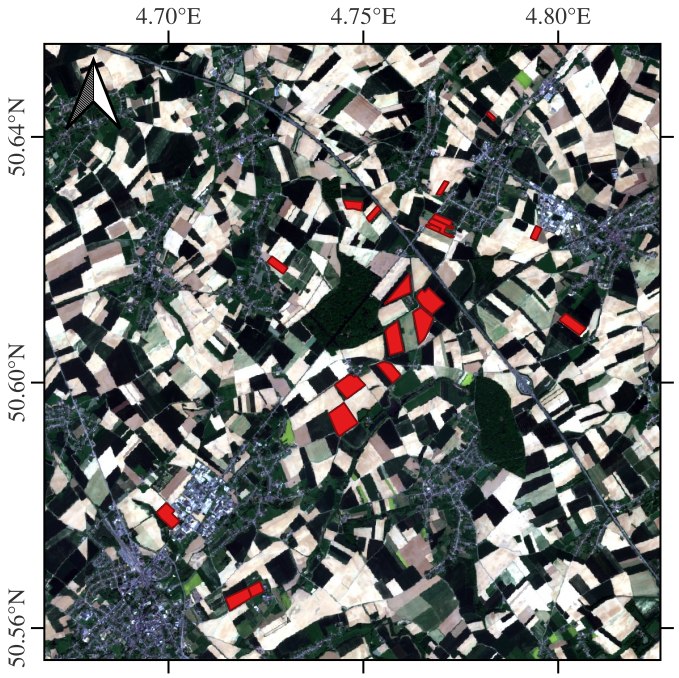


Figure 8: Field parcels of BelSAR test site over a S2 image acquired on 2018-05-08).

date is considered as reference. Following the same idea, the standard deviation at parcel level is interpreted as an uncertainty measurement.

A timeline of the BelSAR measurement dates and available S2 images is shown in Fig. 9. It should be noted that the measurements of 2018-08-29 were excluded from our study, as no valid S2 images were available within 24 days before or after. There are three or four acquisitions for each parcel, as field measurements are not carried out for each maize or wheat parcel for each measurement date.

4. Experiments

4.1. Experimental setup

First, experiments investigate how the simulation process described in Section 3.1 affects the performances of the well-established BVNET algorithm. Second, different hybrid PROSAIL-VAE architecture designs and training cost functions are evaluated and compared with the regression strategies described in Section 2.4.

The quantitative evaluation is carried out by using the in-situ data described in Section 3.3. The predictions are temporally interpolated because satellite and in-situ measurements are usually not acquired at the same dates. The weighted interpolation considers the two most temporally closest available S2 images. The predicted uncertainty intervals output by PROSAIL-VAE and MPSR are also interpolated.

The high-performance computing (HPC) platform of CNES's Data Processing Centre with the following hardware is used for the experiments :

- CPU model: Intel(R) Xeon(R) CPU E5-2698 v4
- GPU model: NVIDIA Tesla V100-SXM2-32GB
- RAM: 64 GB.

To greatly reduce the time-cost of simulations, the resolution of 1 nm over the PROSAIL spectral range of [400, 2500] nm is down-sampled. This task is performed by a moving average filter which moves a sliding fixed size window with a specified stride over the simulated spectra. A filtering stride size equals to the window length is used to avoid zero overlap. Some experiments have corroborated that a size equal to 7 ensures a negligible error with respect the non-down-sampled version. Specifically, the simulation error is at least an order of magnitude lower than the uncertainty induced by MACCS-ATCOR joint algorithm (MAJA) atmospheric corrections (RMSE $\approx 10^{-2}$) (Colin et al., 2023; Doxani et al., 2023).

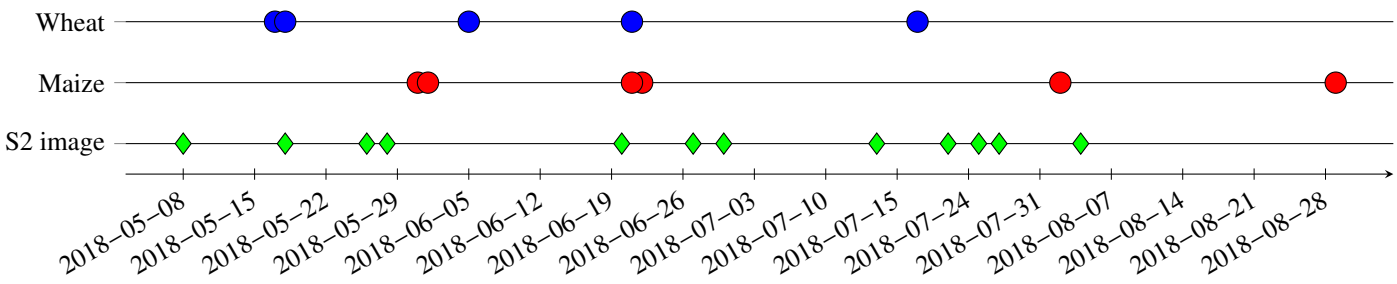


Figure 9: Timeline of measurement dates for maize and wheat parcels of BelSAR campaign, and available S2 images.

4.2. Impact of the simulation process on BVNET performances

The performances of several BVNET models trained using different simulation scenarios are evaluated. The LAI and CCC predictive accuracies of BVNET are assessed over a simulated and the in-situ datasets. Within SL2P, the trained BVNET has 2 linear layers with optimized weights, and performs normalization of input data, and de-normalization of the output prediction, with constants computed from its training data-set. In our experiments, we use the same normalization and de-normalization constants, that we retrieved from SL2P, and only the BVNET linear layer weights are optimized during the training process.

4.2.1. Learning under training and testing distribution mismatch

This experiment studies how a mismatch between the training and testing LAI distributions affects the BVNET model performances. A set of training data-sets are generated using simulations provided by the PROSAIL model with PROSPECT-5. Except for LAI, the distribution parameters specified in Table 3 are considered for the generation of all the data-sets. For LAI, each training data-set is generated using a TN with range $[0, 15]$, with parameters $\mu_{train} \in (0, 1, 2, 3, 4), \sigma_{train} \in (0.5, 1, 2, 3, 4)$. Each of the simulated training data-sets contains 40000 samples and it is used to independently train different BVNET models. For each training data-set, 10 BVNET models are trained to account for variability due to randomness in trainings.

The RMSE is used to assess the LAI predictive performances of the trained BVNET models. First, a synthetic testing data-set with 40000 samples is used to assess performances across all models. The generation of this data-set follows the same training data simulation procedure considering a LAI TN distribution with $\mu_{test} = 2, \sigma_{test} = 3$. A second validation is performed by considering the in-situ measurements of Section 3.3. Results obtained for both testing data-sets are shown in Fig. 10. This figure shows how the discrepancy between training and testing LAI distributions impacts the BVNET predictive performances. The similarity measure between the training and testing LAI TN distributions is computed with the KLD metric. The upper row in Fig. 10 shows the LAI RMSE obtained on the simulated testing data-set. The lower row shows the LAI RMSE obtained on the in-situ measurements. The RMSE is shown as a function of the log-scale KLD between the LAI testing (p_{test}) and training (p_{train}) distributions. The dots correspond to the average RMSE of the 10 BVNET models trained with a given data-set, whereas the vertical lines are the standard deviations (stds) of the RMSE. The horizontal red line shows the performances of the well-established SL2P. The results displayed in both rows corroborate that BVNET performances decrease when the mismatch between training and testing distributions increases. Furthermore, the behavior is the same for simulations and for real data, showing that the experiment is reflecting what may occur with real-world applications. This also shows that the distributions used to generate the training data of SL2P are very well chosen since optimal performances are reached on in-situ measurements.

It can be noted that there is a BVNET trained with a data-set whose KLD is high (orange dot with KLD= 8.75), which has a similar performance on in-situ data than BVNET with lower KLD. Despite this last satisfactory result obtained on the in-situ data, it is unlikely that this BVNET model would generalize well because the LAI distribution of the training data-set is very narrow ($\sigma = 0.5$). As a matter of fact, the large std of RMSE on in-situ data confirms this hypothesis.

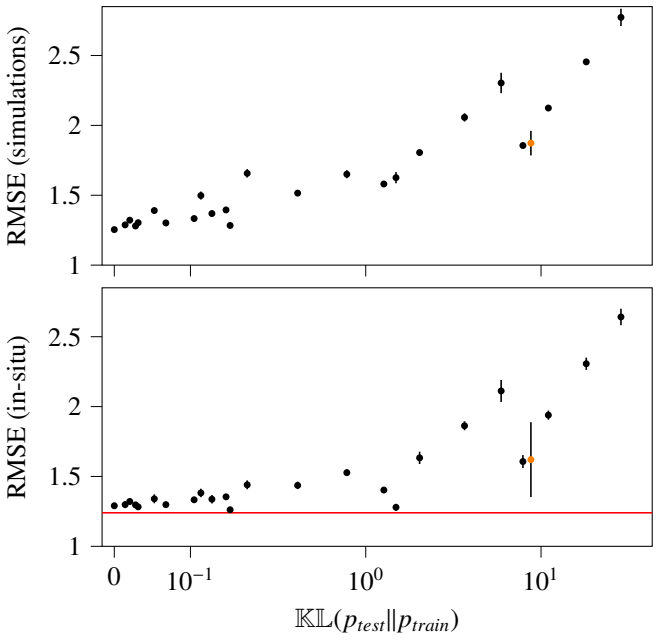


Figure 10: LAI RMSE (lower is better) shown as a function of the log-scale $\mathbb{KL}$ divergence between the LAI testing (p_{test}) and training (p_{train}) distributions. The dots are the average RMSE of the 10 BVNET models trained with a given data-set, whereas the vertical lines are the standard deviation of the RMSE. The horizontal red line shows the performances of the well-established SL2P. Upper row: simulated testing data-set. Lower row: in-situ measurements.

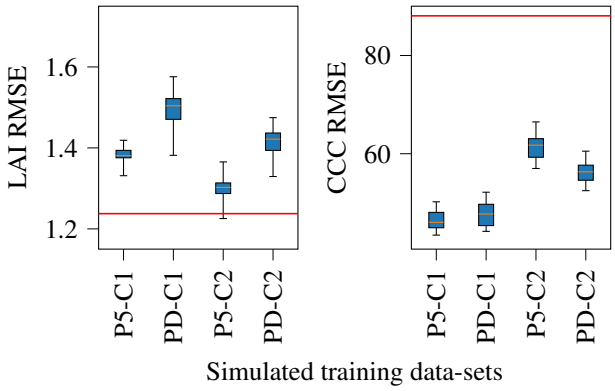


Figure 11: LAI and CCC RMSE (lower is better) obtained by BVNET models trained by using the 4 different data-sets. The data-sets naming convention is P{V}-C{i}, with V being the PROSPECT version and i being the co-distribution type used to generate the data-set samples. The horizontal red line indicates SL2P's accuracy performances.

4.2.2. *The effect of PROSAIL version and variable co-distributions*

The following experiments evaluate how the BVNET performances are impacted by the choice of (i) the PROSAIL model version and (ii) the procedure used to incorporate correlations between PROSAIL input parameters. To evaluate these effects, the simulation modeling designs described in Section 3.1 are used to generate four different training data-sets, with the distributions parameters of Table 3. The resulting data-sets, with 40000 samples each, are used to train BVNET models to predict LAI or CCC variables (i.e. eight training configurations). For each training configuration, 20 BVNET models are independently trained.

The training data-sets are referred as P{V}-C{i}, with V denoting the PROSPECT model version (5 or D). To ensure that input variables for both studied models are strictly identical, the anthocyanin content parameter required by PROSPECT-D is set to $0.0\,\mu g\,cm^{-2}$. Concerning C{i}, it denotes the procedure used to incorporate correlations between LAI and the rest of PROSAIL variables. We denote the co-distribution types 1 and 2 which follow Eq. 4 and Eq. 5 respectively.

A quantitative evaluation is performed using the previously described in-situ LAI and CCC validation data (see Section 3.3). Fig. 11 shows the obtained results which corroborate that the different simulation engineering designs impact the BVNET performances. Unfortunately, concluding which configuration to choose from the results is not straightforward. For instance, co-distribution type 2 seems to improve LAI predictions whereas it deteriorates the predictive performances of CCC.

Using PROSPECT-5 instead of PROSPECT-D seems to increase LAI prediction accuracy no matter the used co-distribution. PROSPECT-5 obtains slightly better performances for CCC than PROSPECT-D for the co-distribution type 1. Conversely, the predictive accuracies of these models decrease when the co-distribution type 2 is used. The difference in performance observed here for chlorophyll between PROSPECT versions in training data-sets is corroborated in Hauser et al. (2021). The impact of simulation modeling design on predictive performances explains why results obtained by our trained BVNET models are slightly different from the ones reached by the BVNET in SL2P. Note that the simulated data-base used to train SL2P's BVNET is generated by PROSPECT-3 model (this model doesn't differentiate carotenoid from chlorophyll pigments). Besides, different strategies are used to characterize soil spectra required by the SAIL model.

Another important remark of the obtained results is that the best performances of LAI and CCC are not reached by the same training data-set. This confirms that the generation of optimal simulations to assist supervised regression algorithms requires extensive efforts and it can lead to poor transferability performances.

4.3. *Evaluation of the proposed PROSAIL-VAE methodology*

Different PROSAIL-VAE model configurations are trained and evaluated here. The studied configurations summarized in Table 5 aim to investigate: (i) spatial or pixel-wise encoder architectures (Fig. 2), (ii) different prior KL regularization terms and (iii) the impact of the β_{KL} parameter balancing the training loss terms (Eq. 3). This last experiment is performed by using three different β_{KL} values (0,1,2). The configuration $\beta_{KL} = 0$ is denoted as PROSAIL-VAE-NP and it indicates that no prior is considered. For all configurations, uniform prior distributions $p(z)$ are considered for all PROSAIL input parameters. To investigate the effects of different KL regularization terms, the following scenarios are studied:

- a single LAI prior (PROSAIL-VAE-L)
- the use of LAI and CAB priors (PROSAIL-VAE-LC)
- all PROSAIL input variables priors (PROSAIL-VAE-AV)

The proposed PROSAIL-VAE configurations are trained on the S2 image data-set presented in Section 3.2 (see Table 4). By using the classical Adam optimizer (kin, 2015), a two-phase optimization strategy is used for each configuration. First, 10 different models are trained with different seeds, with a learning rate (LR) equal to 10^{-3} for 10 epochs. Secondly, the model performing best on S2 validation data is selected for the second optimization step and further trained for 300 epochs. During this second training phase, a cyclical LR scheduler is used by considering a maximum and minimum bound values equal to $LR_{max} = 5 \times 10^{-4}$ and $LR_{min} = 10^{-8}$. The scheduler reduces the LR by a factor of 10 when the validation loss hits a plateau during 5 consecutive epochs.

The prediction performances of the different trained PROSAIL-VAE models are compared with the MPSR supervised strategy described in Section 2.4.2 and the well-known pre-trained SL2P processor. The training of MPSR is carried out by using a simulated data-set containing 2×10^5 samples. This data-set is generated with PROSPECT-5 + 4SAIL, following the procedure detailed in Section 3, with the BV sampled from the distributions provided in Table 3, and with correlations introduced by co-distribution type 2. For the supervised training stage, similarly to PROSAIL-VAE, different models are initialized and trained for a few epochs, and the best is selected for further training. In this second training phase, the number of epochs is set to 5000 and the LR is scheduled identically to that of PROSAIL-VAE.

The quantitative assessment of the results is performed using the LAI and CCC in-situ measurements of Section 3.3. The metrics described in Section 2.5 are used to evaluate the TN distributions associated to the input PROSAIL parameters predicted by PROSAIL-VAE models and the MPSR strategy.

The estimate of the LAI is taken as the expectation μ_{LAI} of the LAI TN distribution produced by the encoder of PROSAIL-VAE, and is directly compared with the in-situ measurement. A simple prediction interval is computed as $[\mu_{LAI} - n\sigma_{LAI}, \mu_{LAI} + n\sigma_{LAI}]$, with σ_{LAI} the standard deviation of the TN distribution and $n = 2$. Concerning CCC, the predicted value is obtained by $\mathrm{CCC} = \mu_{\mathrm{LAI}} \times \mu_{C_{ab}}$[4].

The SL2P evaluation is performed by taking directly LAI and CCC predictions. For this approach, uncertainty evaluation can not be performed.

4.3.1. *Quantitative assessment using in-situ data-sets*

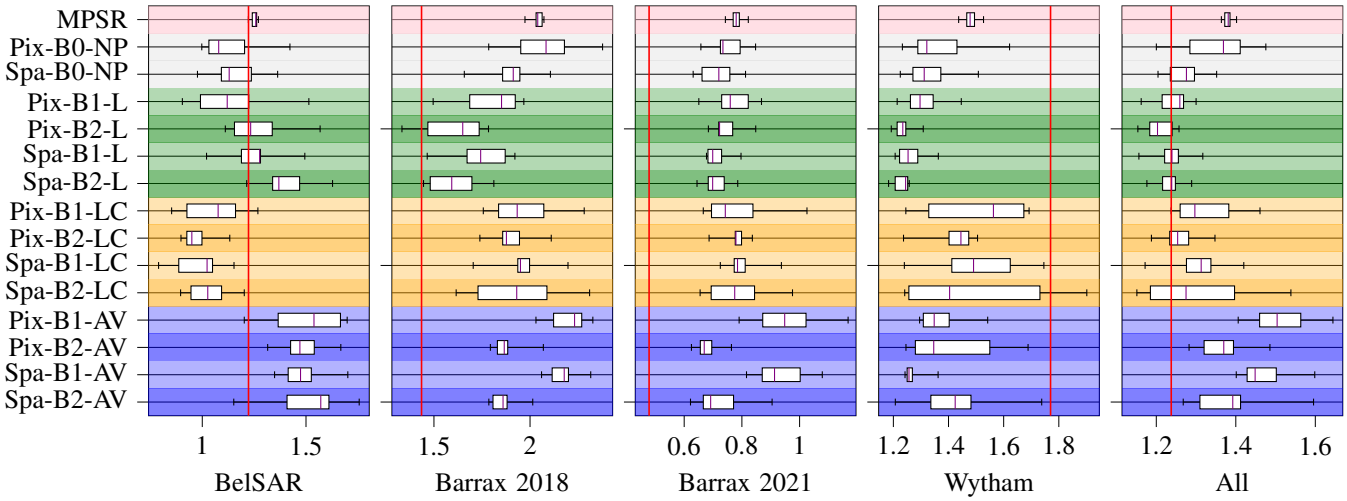
Fig. 12 shows the LAI results, whereas CCC performances are shown in Fig. 13 for the different test sites.

Some differences can be observed by comparing the results obtained by the studied methods on the different test sites. For instance, PROSAIL-VAE models always outperform SL2P on the Wytham site. In contrast, SL2P seems to perform a little better on Barrax's 2021 campaign. However, the SL2P improvement on this last test site is less significant since good RMSE and uncertainty metrics are also obtained by the rest of methods.

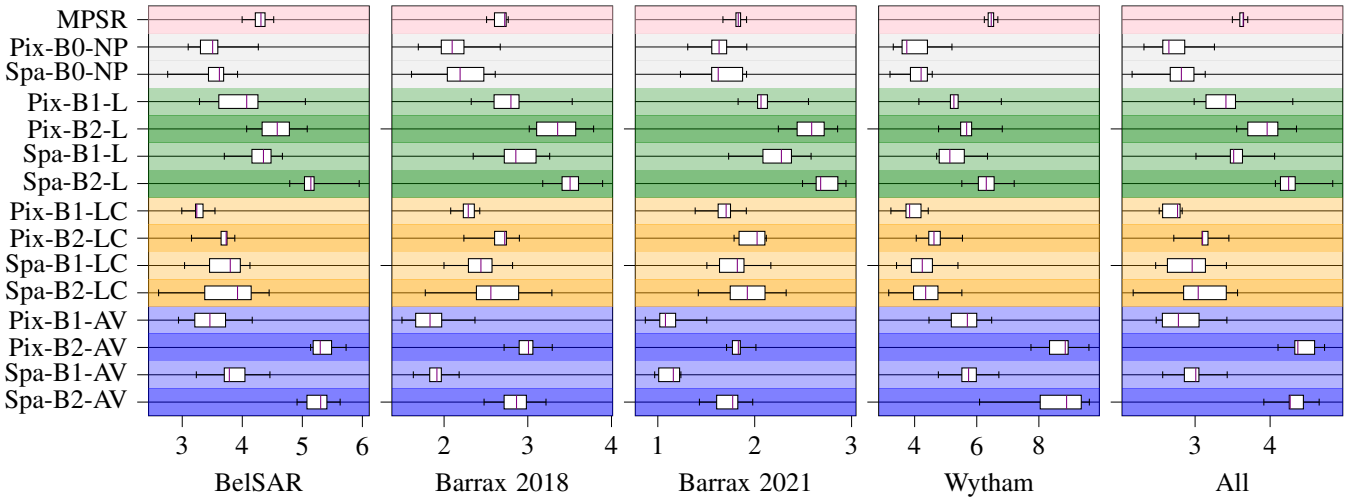
On the BelSAR campaign, best results are mostly obtained by PROSAIL-VAE models. In this case, some differences can be highlighted by comparing PROSAIL-VAE strategies proposing different prior KLD regularization terms. The PROSAIL-VAE-AV configuration forcing the prior on all input PROSAIL parameters obtains the worst results. In contrast, PROSAIL-VAE-LC reaches the best performances on BelSAR campaign by only considering LAI and C_{ab} priors. The performance of PROSAIL-VAE-L configuration is corroborated by results on the Barrax 2018 campaign.

Despite the differences between results obtained on the different sites, most of PROSAIL-VAE models achieve better overall RMSE performances than

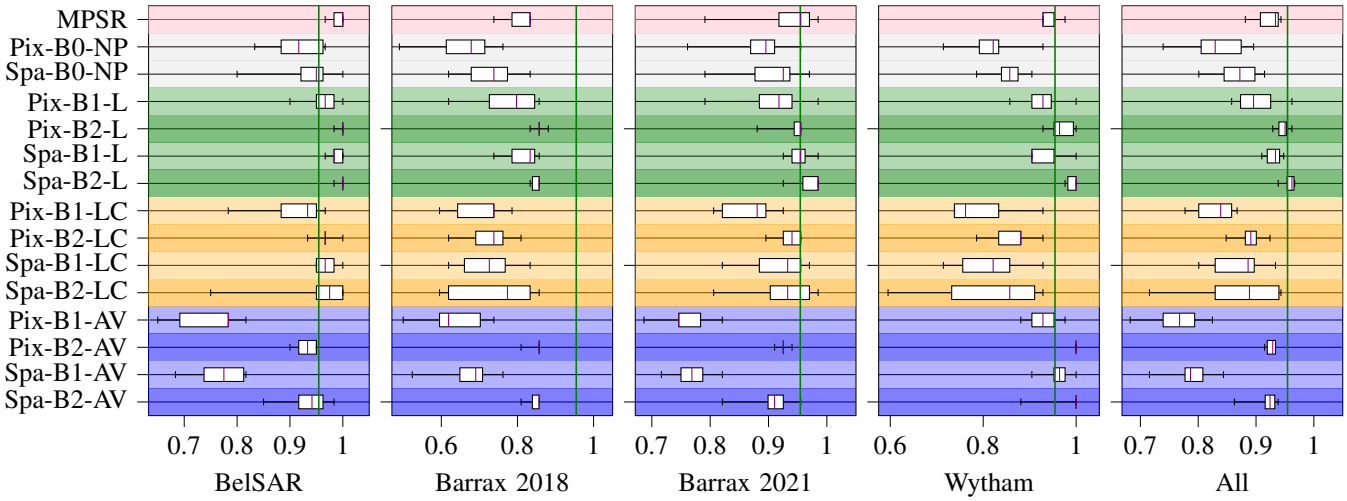
[4]The estimated uncertainty is derived from the variance of the product of LAI and C_{ab}, by assuming that they are not correlated (Goodman, 1960): $\mathrm{var}\,(\mathrm{LAI} \times C_{ab}) = \left(\mathrm{var}\,(\mathrm{LAI}) + \mu^2_{\mathrm{LAI}}\right)\left(\mathrm{var}\,(C_{ab}) + \mu^2_{C_{ab}}\right) - \mu^2_{\mathrm{LAI}}\mu^2_{C_{ab}}$



(a) RMSE of regression experiments obtained on in-situ data-sets. The vertical red line corresponds to SL2P results for each test site.

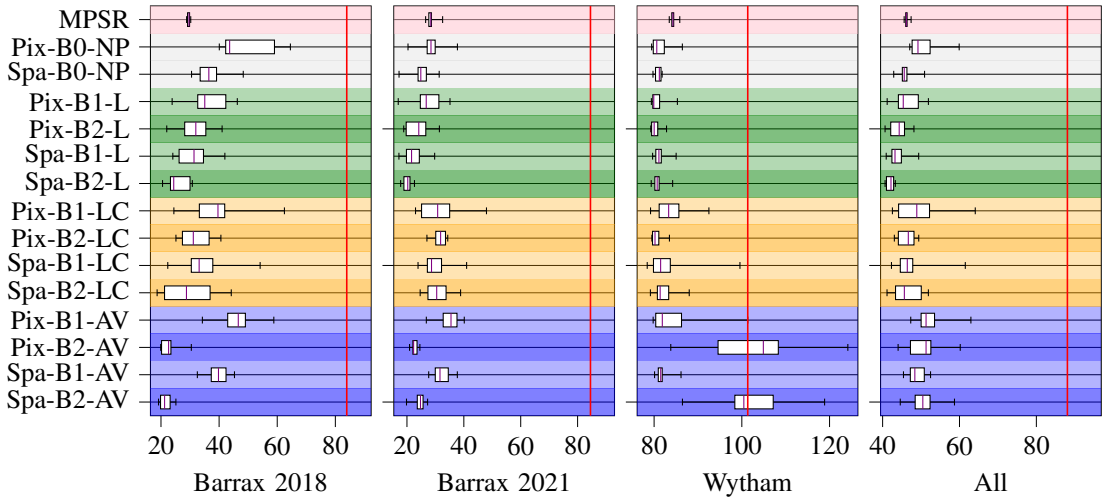


(b) MPIW of regression experiments on in-situ data-sets

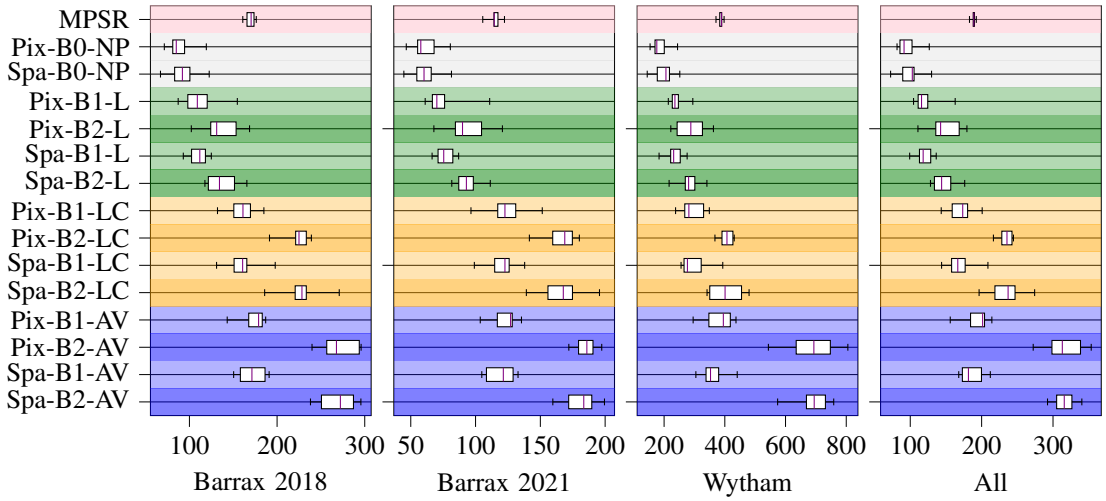


(c) PICP of regression experiments on in-situ data-sets. The vertical green line is the ratio of values that lie within a 2-σ interval estimate in a normal distribution (≈ 0.95)

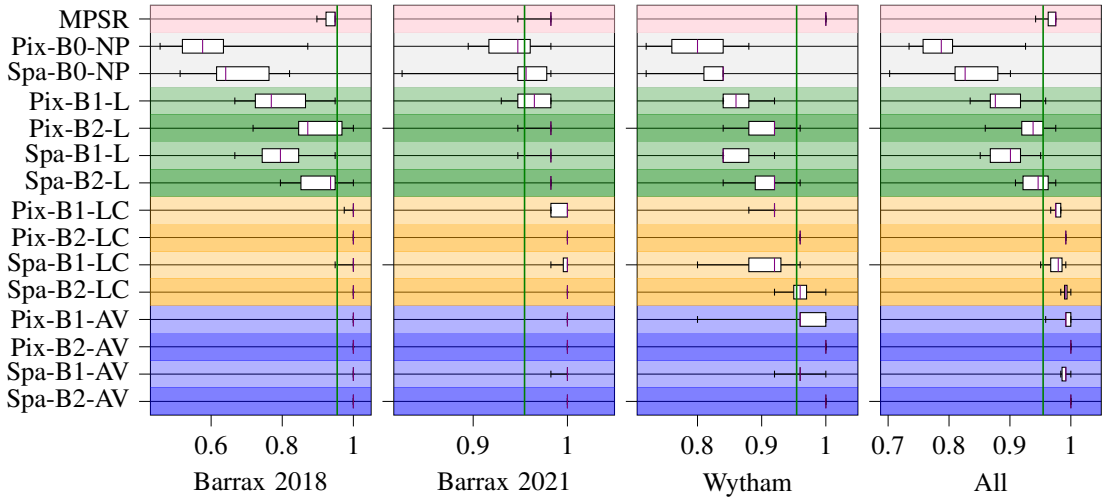
Figure 12: Boxplots of LAI metrics obtained on in-situ data-sets with PROSAIL-VAE models and MPSR. For each configuration, 10 models are trained and attained min and max values are displayed by boxplot whiskers. The box sites are the 25^{th} and 75^{th} centiles, and the purple line inside corresponds to the median.



(a) RMSE of regression experiments obtained on in-situ datasets. The vertical red line corresponds to SL2P results for each test site.



(b) MPIW of regression experiments on ground validation data-sets



(c) PICP of regression experiments on in-situ data-sets, the vertical green line is the ratio of values that lie within a 2-σ interval (considering a normal distribution) (≈ 0.95)

Figure 13: Boxplots of CCC metrics obtained on in-situ data-sets with PROSAIL-VAE models and MPSR. For each configuration, 10 models are trained and attained min and max values are displayed by boxplot whiskers. The box sites are the 25^{th} and 75^{th} centiles, and the purple line inside corresponds to the median.

Table 5: Studied PROSAIL-VAE configurations which depend on the encoder architecture, β_{KL} value and the used variable priors. The configuration acronyms consider : *pix* for "pixel-wise", *spa* for "spatial", *NP* for "no prior", *L* for "LAI", *LC* for "LAI and C_{ab}", *AV* for "all variables".

Configuration acronym	Prior type	Variable prior	Encoder architecture	β_{KL}
Pix-B0-NP	None	None (PROSAIL-VAE-NP)	pixel-wise	0
Spa-B0-NP			spatial	0
Pix-B1-L	Uniform	LAI (PROSAIL-VAE-L)	pixel-wise	1
Pix-B2-L				2
Spa-B1-L			spatial	1
Spa-B2-L				2
Pix-B1-LC		LAI, C_{ab} (PROSAIL-VAE-LC)	pixel-wise	1
Pix-B2-LC				2
Spa-B1-LC			spatial	1
Spa-B2-LC				2
Pix-B1-AV		All variables (PROSAIL-VAE-AV)	pixel-wise	1
Pix-B2-AV				2
Spa-B1-AV			Spatial	1
Spa-B2-AV				2

SL2P as shown in the last column of 12a. Comparing the two supervised strategies, the MPSR shows slightly worse LAI performance than SL2P.

The assessment of CCC predictions is corroborated by results shown in 13a. For this variable, PROSAIL-VAE models always outperform SL2P in all in-situ data-sets. Except for PROSAIL-VAE-AV, the different PROSAIL-VAE configurations obtain similar results. Results obtained in Wytham shows how the use of a high number of priors decreases the accuracy of the results, which is exacerbated by increasing β_{KL}. Results obtained by MPSR are slightly worse than those reached by PROSAIL-VAE models, however, its performance is better than SL2P.

Prediction uncertainties can be evaluated by MPIW and PICP metrics obtained by LAI (12b and 12c) and CCC (13b and 13c) predictions.

The MPSR predicts wider distributions than PROSAIL-VAE-NP for both BV (see 13b and 12b). The MPIW values obtained for the LAI and CCC show that adding KLD regularization terms ($\beta_{KL} > 0$) increases the variance of LAI and CCC distributions. In general, the width of prediction intervals improves by increasing β_{KL} and the narrowest intervals are reached by PROSAIL-VAE-NP. On the LAI, the MPIW is further increased with PROSAIL-VAE-L and with PROSAIL-VAE-AV with $\beta_{KL} = 2$.

The PICP depends on both the estimation error being low and the prediction intervals being large enough. Compared to PROSAIL-VAE-NP, PROSAIL-VAE-L has a similar LAI RMSE, but wider prediction intervals (larger MPIW), therefore the PICP is increased. For the CCC, the RMSE varies relatively little between PROSAIL-VAE models. The PICP increases as the MPIW increases.

PICP results obtained on Barrax 2021 show that LAI predictions of all PROSAIL-VAE models reach 2σ target (except PROSAIL-VAE-AV with $\beta_{KL} = 1$). However, overall results show that the LAI prediction intervals tend to uncertainty underestimation. This can be corroborated by PICP results shown in the last column of 13b, which range from 0.65 to 0.9.

Concerning CCC, PROSAIL-VAE-NP obtains the lowest PICP results, whereas slightly better metrics are obtained by PROSAIL-VAE-L and MPSR. The overall PICP metrics obtained by CCC are close to 2σ target for PROSAIL-VAE-LC and PROSAIL-VAE-AV.

Results show how, despite low CCC RMSE results are obtained by PROSAIL-VAE-LC and PROSAIL-VAE-AV in Barrax 2021 (under $40\,\mu g\,cm^{-2}$), accurate predictions are contained on their prediction intervals.

All the presented results show that there is no notable difference between results obtained by pixel-wise and spatial encoder architectures. This can be attributed to the fact that only the first layer of the CNN uses filters larger than 1 pixel in order to preserve the input data resolution.

4.3.2. Comparison between PROSAIL-VAE and SL2P

The performances of the best PROSAIL-VAE model are compared here against the pre-trained SL2P. Considering overall results from Section 4.3.1 on situ-data (see Section 3.3), the lowest RMSE for LAI and CCC are obtained by PROSAIL-VAE-L and PROSAIL-VAE-LC configurations. Both models reach PICP results close to the target for the CCC variable. In terms of MPIW, PROSAIL-VAE-L obtains overall CCC results ($\approx$ 100–170 $\mu g\,cm^{-2}$) better than PROSAIL-VAE-LC ($\approx$ 150–250 $\mu g\,cm^{-2}$). However, the large prediction intervals of this last configuration leads to PICP close to 1 suggesting that the uncertainty is overestimated. Accordingly, PROSAIL-VAE-L model associated to the Pix-B2-L results is chosen as the best model to be compared against SL2P. Additional results assessing the good training performances of PROSAIL-VAE-L are available in Appendices B and C.

The first performance comparison between the selected Pix-B2-L PROSAIL-VAE and SL2P is shown in Table 6. Comparing results obtained for LAI and CCC predictions, it can be observed that PROSAIL-VAE obtains the best overall RMSE metrics.

Fig. 14 allows to corroborate these results by showing the individual predictions of both methods. These results show that similar LAI predictions are obtained by PROSAIL-VAE and SL2P. For instance, LAI and CCC predictions for the alfalfa crop are underestimated by both methods. The same behavior can be observed by the LAI predictions of the poppy class on Barrax 2108. The underestimation seems slightly lower for PROSAIL-VAE.

Both methods predict a limited range of LAI values for the Wytham site. However, PROSAIL-VAE seems to perform better than SL2P, which can be explained by the heterogeneous forest canopies LAI underestimation problem of SL2P reported on (Xie et al., 2019; Brown et al., 2021b).

PROSAIL-VAE slightly overestimates the prediction of low LAI values.

The temporal evolution of LAI predictions of both methods is also studied in Fig. 15. This last figure displays the LAI time series predictions obtained over a maize parcel belonging to the BelSAR site. Both predicted time series show a well-defined summer crop phenology curve, that can be fitted with a double-logistic model (Zeng et al., 2020; Zérah et al., 2023). As observed, similar predictions for both PROSAIL-VAE and SL2P closely match the maize in-situ measurements.

Fig. 16 allows to assess how closely the SL2P and PROSAIL-VAE predictions are correlated. It shows the LAI, CCC and CWC variables predicted on the S2 testing data-set (see Section 3.2) by both methods. The results show that a strong correlation is observed between LAI predictions, whereas a different behavior is obtained for CCC and CWC variables.

SL2P tends to predict higher CCC values than PROSAIL-VAE which saturates at $250\,\mu g\,cm^{-2}$. In this case, the in-situ validation showed that SL2P tends to overestimate. The same behavior is observed by CWC where PROSAIL-VAE predictions saturate at at 0.17 cm. To assess these results, more in-situ measurements would be required to corroborate the quality of these two predicted variables. A visual comparison of the results can be found in Fig. D.24 in Appendix Appendix Appendix D, where predicted biophysical variable maps are shown.

Despite PROSAIL-VAE performing the inversion of all input PROSAIL

Table 6: LAI and CCC prediction performance on in-situ data-sets for SL2P and PROSAIL-VAE.

Method	BV	LAI					CCC			
	Metric	BelSAR	Barrax (2018)	Barrax (2021)	Wytham	All	Barrax (2018)	Barrax (2021)	Wytham	All
SL2P	RMSE	**1.22**	1.43	**0.48**	1.77	1.24	83.92	84.53	101.35	88.08
PROSAIL-VAE	RMSE	1.30	**1.42**	0.72	**1.21**	**1.16**	**27.60**	**20.51**	**80.78**	**42.33**
	MPIW	4.74	3.74	2.72	5.45	4.04	140.53	94.02	235.20	138.18
	PICP	1.0	0.88	0.96	0.95	0.95	0.95	0.98	0.84	0.94

parameters, not all predicted variables can be compared with SL2P results. Therefore, a visual evaluation is proposed in Fig. 17 to assess the set of PROSAIL-VAE predictions. This figure shows the histograms of expectation and std of PROSAIL variables predicted on the testing S2 image data-set. The histograms of the PROSAIL variables expectations are compared to the distribution of variables (see Table 3) used to generate training data-sets for BVNET (see Section 4.2) and MPSR.

Fig. 17 shows that the leaf parameter index N is likely to be poorly estimated. The histogram of the expectation of the leaf parameter N is concentrated on the lower bound of its definition interval at $N = 1.2$, which is associated with monocotyledon vegetation (Féret et al., 2021). Unfortunately, it is well-known that vegetation with N values significantly higher than 1.3 occurs in real scenarios, and should be present in the images of the testing data-set.

Besides N, looking at the histograms of expected values of C_{ab}, C_c, C_b, C_w, α, it can be observed how predictions are not occurring over the full range of their definition intervals described in Table 3. For these variables, it is uncertain whether their predictions are flawed, or if they reflect the vegetation observed in the data-set. The low predicted C_{ab} values corroborate the CCC saturation effect observed in Fig. 16. In spite of this, PROSAIL-VAE obtains good CCC and LAI results for in-situ data-sets which hints at accurate C_{ab} predictions.

In general, the distributions used in the literature (see Table 3) are very different from the variable distributions predicted by PROSAIL-VAE. The closest match is observed by LAI and C_b variables. For C_m, predicted values saturate on the upper bound (at C_m=0.011 $\mu g^2\,cm^{-1}$) of its definition interval, suggesting that it is too tight. Similarly C_w saturates on its lower bound (at C_m=0.0075 cm), which may be lowered to 0 cm.

For a further evaluation, the scatter-plots between the expectation of all pairs of PROSAIL variables predicted by PROSAIL-VAE on the S2 testing data-set are available in Appendix Fig. E.25. These figures allow us to study the correlations between predicted variables. Results show how arbitrary co-distributions traditionally used in the literature (see Section 3.1) may not describe real relationships between variables. In particular, most of the predicted variables do not exhibit a simple linear relationship with LAI.

The results also corroborate that there exists a correlation between LAI and hot-spot parameters predicted by PROSAIL-VAE. Fig. 18 shows the scatter plot between these variables, which are inferred for a wheat crop parcel of the BelSAR site at different 2018 dates. As observed, the hot-spot parameter decreases when LAI values are greater than one. This result follows the theory about the hot-spot parameter of the SAIL model (Verhoef) suggesting that $h \propto 1/LAI$ for a tall wheat plant with constant leaf size.

5. Discussion

The experiments have shown the potential of the proposed hybrid methodology incorporating a physical model into the design of a trainable neural network architecture. Obtained results have corroborated that it is possible to train a neural network in a self-supervised manner for the simultaneously retrieval of all the input PROSAIL parameters. Accurate performances are obtained by our proposed models, which have not been trained with simulations generated by the physical model that is inverted. To highlight the advantages of avoiding simulation modeling design, first experiments have evaluated the impact of training simulations on predictive supervised regression algorithm performances. LAI and CCC prediction results obtained by different BVNET models trained with different pre-simulated data-sets have been compared. Results evaluated on simulated and the in-situ data have corroborated that accuracies were impacted by

(i) the physical model used to generate the simulations,

(ii) the choice of the LAI data distribution and

(iii) the strategy proposed to incorporate dependencies between the LAI and the rest of variables.

The second family of experiments have evaluated the performances of the different proposed PROSAIL-VAE configurations. The obtained results have been validated with in-situ measurements and compared with predictions obtained by the pre-trained SL2P. Overall results have shown that PROSAIL-VAE achieves similar or better performances than the SL2P pre-trained model.

Among the several PROSAIL-VAE scenarios, a strategy has been investigated to incorporate spatial information at local scale in the prediction process. It must be noticed that most of existing regression algorithms inverting PROSAIL are purely pixel-wise. Unfortunately, obtained results have not allowed us to corroborate the interest of using a spatial encoder architecture. These results can be explained by the fact that the convolutional part of the encoder was limited to 3×3 filters on the first layer. Also, the reconstruction loss metric does not inject spatial context information during the learning process. New reconstruction loss functions or new priors on input model variables considering spatial dependencies could be proposed. For instance, the use of a perceptual loss penalizing the texture of reconstructions during the training process may be a next step towards this goal.

Experiments have also studied the incorporation of prior knowledge in the physics-based guidance learning process. Compared to the existing simulation-assisted regression methods, it must be noticed that PROSAIL-VAE requires little prior knowledge about the distribution of input model parameters. For instance, the configuration $\beta_{KL} = 0$ has shown that accurate results can be obtained by only setting information about the input PROSAIL parameter's value ranges (only upper and lower bounds, not their distributions).

The incorporation of knowledge about the prior distributions of PROSAIL parameters has also been studied by investigating configurations with different prior $\mathbb{KL}$ regularization terms. The studied PROSAIL-VAE configurations have shown how information about prior distributions can be injected in the loss term to guide the model training.

Experiments have corroborated that the selection of priors to be used in the KLD regularization term is not trivial. Adding some few priors describing well-known variables such as LAI can improve the prediction accuracies. However, low prediction accuracies can be reached by configurations incorporating less well-known parameters. It must be remarked that only uniform prior distributions have been considered in our experiments. To improve the results, more informative and specific priors about PROSAIL input parameters could be incorporated in the KLD regularization term.

Experiments have shown how multiple probabilistic predictions can be obtained by training a single model. Visual results have shown that some PROSAIL parameters seem to be better predicted than others. This could be explained by the principle of the ill-posed inverse problem or by the importance of each input variable in PROSAIL model. In general, although a visual evaluation has been carried out, in-situ measurements related to less studied parameters such as carotenoids, brown pigments or dry mater content are completely necessary to quantitatively assess the performance of predicted variables. Besides, more in-situ data is necessary, in both quantity and variety (vegetation types, location, season) to further validate the proposed hybrid methodology.

The best PROSAIL-VAE configuration has been deeply investigated and compared with SL2P. In our work, the choice of the best model has been proposed by comparing the performances obtained on in-situ measurements. In operational contexts, this selection criterion can lead to an over-fitting risk since testing data should be never used for setting hyper-parameters or choosing the

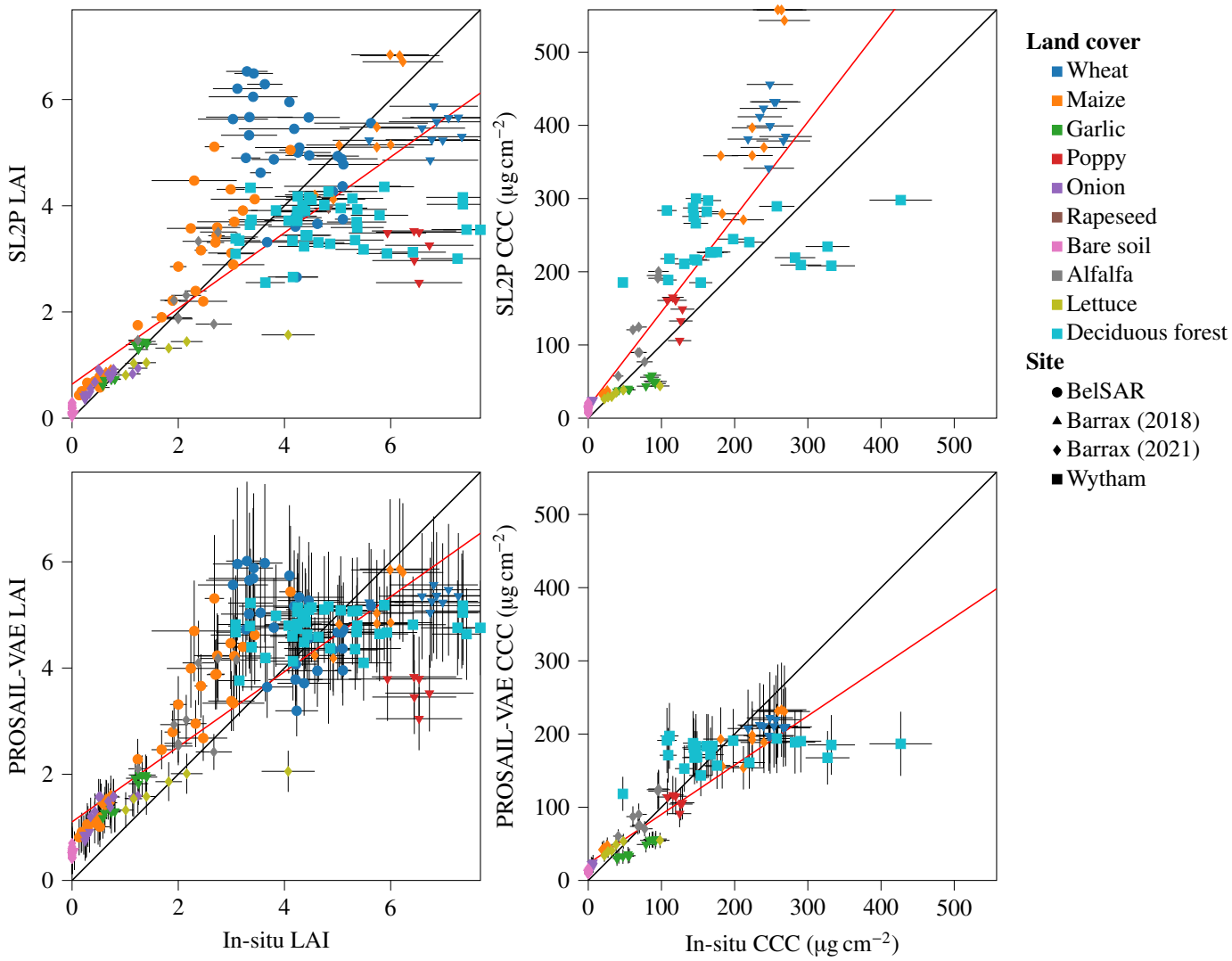


Figure 14: Scatter plots of LAI and CCC predictions from SL2P and PROSAIL-VAE versus in-situ test sites measurements. For each data point, the horizontal black lines correspond to the predicted uncertainty measures. In contrast, the vertical black lines indicate PROSAIL-VAE 2σ prediction intervals, derived from the inferred PROSAIL variable distributions.

best configuration. Some experiments (outside the scope of the paper) have shown that it does not exist a correlation between the training reconstruction loss values and prediction accuracies obtained on LAI and CCC in-situ measurements. In fact, reconstruction is a training proxy task and the selection model criteria must be related to their inversion performances. Therefore, reliable solutions based on cross-validation techniques could be proposed to compare trained models if a large number of in-situ samples would exist.

Accuracy prediction improvement could be also obtained by changing the SAIL version considered in our PROSAIL-VAE implementation. The obtained results have shown that a positive bias can exist for low LAI values, which may be explained by the insufficient capacity of the model to simulate realistic soil spectra. Instead of combining only two reference soil spectra, the use of a soil spectral library could improve the prediction of low LAI values where the ground is visible. In the same direction, using other PROSPECT model versions or other RTM could improve the performances of our hybrid methodology. An important remark is that the change of the physical-based decoder to be inverted do not require any additional tuning task, which would be the case for simulation-based approaches as BVNET.

6. Conclusion

This work has presented a new data-driven paradigm to invert radiative transfer models for the retrieval of biophysical variables. The proposed hybrid methodology has proposed to incorporate the PROSAIL model into the design of a trainable neural network architecture, which has been trained in a self-supervised manner by directly exploiting unlabeled S2 images. As a result, the presented strategy has allowed the simultaneously probabilistic prediction of all the input PROSAIL parameters.

The main advantage of the proposed method is that physics-based guidance has been incorporated in to the training process avoiding simulation-assisted learning. Training without simulations avoid unrealistic failure scenarios that can limit the transferability of the algorithms to real-world situations. Besides, simulation engineering process is a very hard task requiring a lot of prior knowledge about the physical process and the dependencies between the involved variables. Different experiments have corroborated the interest of the proposed hybrid methodology by showing the limitations of simulation-trained machine learning algorithms.

Results obtained by PROSAIL-VAE have been validated with LAI and CCC in-situ measurements collected on 4 different field campaigns over three European tests sites. The performances of the proposed method have been compared with the state-of-the art SL2P. The different experiments have corrobo-

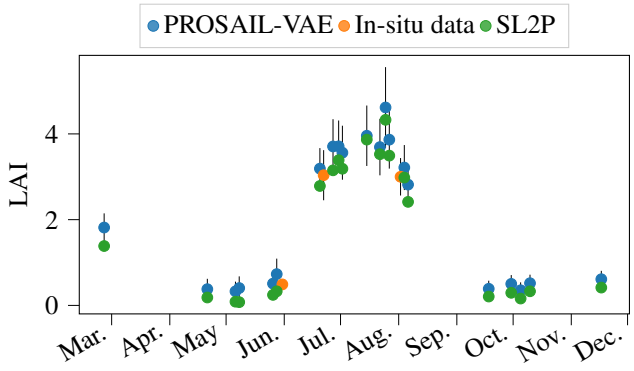


Figure 15: LAI time series predictions obtained over a maize parcel belonging to the BelSAR site. PROSAIL-VAE and SL2P predictions are obtained by considering non-cloudy S2 available images acquired on 2018.

rated that the presented strategy can reach more accurate predictions (specially for CCC variable).

Another important advantage of the proposed methodology is that it is agnostic to the physical-based decoder to be inverted. Therefore, other canopy reflectance models could be plugged into our architecture and be trained in the same self-supervised manner. Future work may therefore focus on using different RTM to perform the simultaneous inversion of several physical models sharing input parameters.

In an attempt to enable reproducible research, the implementation of the methods presented in this paper is available at https://src.koda.cnrs.fr/yoel.zerah.1/prosailvae.git.

Acknowledgments

The authors would like to thank CNES for the provision of its high performance computing (HPC) infrastructure to run the experiments presented in this paper and the associated help. We also thank the ANR-18-CE23-0023 MAESTRIA and ANR-JCJC-20-CE23-0003 DeepChange project for funding this research. This study has been undertaken using data from the "Fiducial Reference Measurements for Vegetation – Phase 2" (FRM4VEG – Phase 2) project, which was funded by the European Space Agency. We thank Jean Bouchat for providing the BelSAR measurement data, and his help in its use. We thank Marie Weiss for the fruitful discussions we had with her. We thank Iris Dumeur for her help in the retrieval of the Sentinel-2 data we used throughout this study.

References

, 2015. Adam: A Method for Stochastic Optimization. arXiv:1412.6980.

Ali, A.M., Darvishzadeh, R., Skidmore, A., Gara, T.W., O'Connor, B., Roeoesli, C., Heurich, M., Paganini, M., 2020. Comparing methods for mapping canopy chlorophyll content in a mixed mountain forest using sentinel-2 data. International Journal of Applied Earth Observation and Geoinformation 87, 102037. URL: https://www.sciencedirect.com/science/article/pii/S0303243419310967, doi:https://doi.org/10.1016/j.jag.2019.102037.

Atzberger, C., 2004. Object-based retrieval of biophysical canopy variables using artificial neural nets and radiative transfer models. Remote Sensing of Environment 93, 53–67. URL: https://www.sciencedirect.com/science/article/pii/S0034425704001956, doi:https://doi.org/10.1016/j.rse.2004.06.016.

Balasundram, S.K., Shamshiri, R.R., Sridhara, S., Rizan, N., 2023. The role of digital agriculture in mitigating climate change and ensuring food security: An overview. Sustainability 15. URL: https://www.mdpi.com/2071-1050/15/6/5325, doi:10.3390/su15065325.

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 Sensing of Environment 110, 275–286. URL: https://www.sciencedirect.com/science/article/pii/S0034425707000909, doi:https://doi.org/10.1016/j.rse.2007.02.018.

Bouchat, J., Tronquo, E., Orban, A., de Macedo, K.A.C., Verhoest, N.E.C., Defourny, P., 2023. The belsar dataset: Mono- and bistatic full-pol l-band sar for agriculture and hydrology. arXiv:2310.04423.

Bouchat, J., Tronquo, E., Orban, A., Verhoest, N.E.C., Defourny, P., 2022. Assessing the potential of fully polarimetric mono- and bistatic sar acquisitions in l-band for crop and soil monitoring. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 15, 3168–3178. doi:10.1109/JSTARS.2022.3162911.

Boussetta, S., Balsamo, G., Beljaars, A., Kral, T., Jarlan, L., 2013. Impact of a satellite-derived leaf area index monthly climatology in a global numerical weather prediction model. International Journal of Remote Sensing 34, 3520–3542. URL: https://doi.org/10.1080/01431161.2012.716543, doi:10.1080/01431161.2012.716543, arXiv:https://doi.org/10.1080/01431161.2012.716543.

Brown, L.A., Camacho, F., García-Santos, V., Origo, N., Fuster, B., Morris, H., Pastor-Guzman, J., Sánchez-Zapero, J., Morrone, R., Ryder, J., Nightingale, J., Boccia, V., Dash, J., 2021a. Fiducial reference measurements for vegetation bio-geophysical variables: An end-to-end uncertainty evaluation framework. Remote Sensing 13. URL: https://www.mdpi.com/2072-4292/13/16/3194, doi:10.3390/rs13163194.

Brown, L.A., Fernandes, R., Djamai, N., Meier, C., Gobron, N., Morris, H., Canisius, F., Bai, G., Lerebourg, C., Lanconelli, C., Clerici, M., Dash, J., 2021b. Validation of baseline and modified sentinel-2 level 2 prototype processor leaf area index retrievals over the united states. ISPRS Journal of Photogrammetry and Remote Sensing 175, 71–87. URL: https://www.sciencedirect.com/science/article/pii/S0924271621000617, doi:https://doi.org/10.1016/j.isprsjprs.2021.02.020.

Campbell, G., 1990. Derivation of an angle density function for canopies with ellipsoidal leaf angle distributions. Agricultural and forest meteorology 49, 173–176.

Campos-Taberner, M., Moreno-Martínez, , 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 Sensing 10. URL: https://www.mdpi.com/2072-4292/10/8/1167, doi:10.3390/rs10081167.

Camps-Valls, G., Verrelst, J., Munoz-Mari, J., Laparra, V., Mateo-Jimenez, F., Gomez-Dans, J., 2016. A survey on gaussian processes for earth-observation data analysis: A comprehensive investigation. IEEE Geoscience and Remote Sensing Magazine 4, 58–78. doi:10.1109/MGRS.2015.2510084.

Colin, J., Hagolle, O., Landier, L., Coustance, S., Kettig, P., Meygret, A., Osman, J., Vermote, E., 2023. Assessment of the performance of the atmospheric correction algorithm maja for sentinel-2 surface reflectance estimates. Remote Sensing 15. URL: https://www.mdpi.com/2072-4292/15/10/2665, doi:10.3390/rs15102665.

Combal, B., Baret, F., Weiss, M., Trubuil, A., Macé, D., Pragnère, A., Myneni, R., Knyazikhin, Y., Wang, L., 2003. Retrieval of

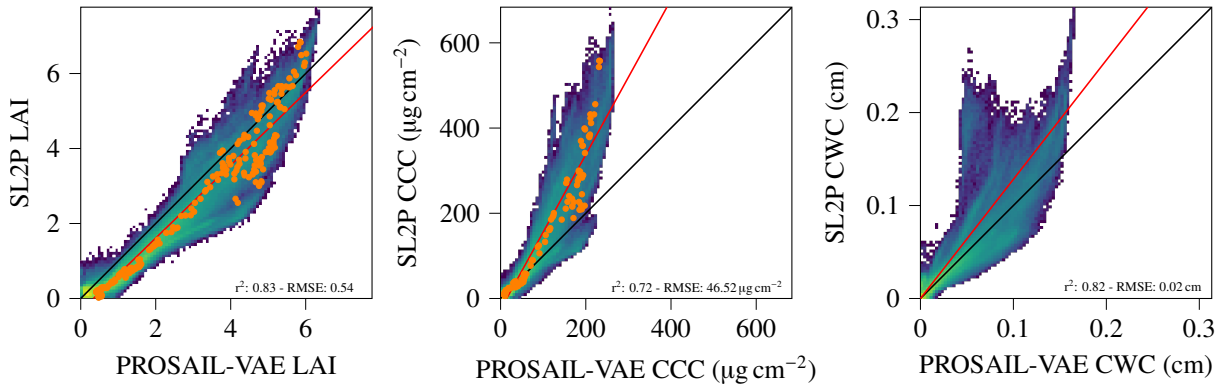


Figure 16: Scatter-plots measuring the correlation between PROSAIL-VAE and SL2P predictions. LAI, CCC, and CWC are computed on the testing S2 image data-set. The regression line is plotted in red and orange points correspond to predictions of in-situ measurements.

canopy biophysical variables from bidirectional reflectance: Using prior information to solve the ill-posed inverse problem. Remote Sensing of Environment 84, 1–15. URL: https://www.sciencedirect.com/science/article/pii/S0034425702000354, doi:https://doi.org/10.1016/S0034-4257(02)00035-4.

Danson, F., Rowland, C., Baret, F., 2003. Training a neural network with a canopy reflectance model to estimate crop leaf area index. International Journal of Remote Sensing 24, 4891–4905.

Darvishzadeh, R., Matkan, A.A., Ahangar, A.D., 2012. Inversion of a radiative transfer model for estimation of rice canopy chlorophyll content using a lookup-table approach. IEEE Journal of selected topics in applied earth observations and remote sensing 5, 1222–1230.

Darvishzadeh, R., Skidmore, A., Schlerf, M., Atzberger, C., 2008. Inversion of a radiative transfer model for estimating vegetation lai and chlorophyll in a heterogeneous grassland. Remote Sensing of Environment 112, 2592–2604. URL: https://www.sciencedirect.com/science/article/pii/S0034425707004968, doi:https://doi.org/10.1016/j.rse.2007.12.003. earth Observations for Terrestrial Biodiversity and Ecosystems Special Issue.

Defourny, P., Bontemps, S., Bellemans, N., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Nicola, L., Rabaute, T., Savinaud, M., Udroiu, C., Valero, S., Bégué, A., Dejoux, J.F., El Harti, A., Ezzahar, J., Kussul, N., Labbassi, K., Lebourgeois, V., Miao, Z., Newby, T., Nyamugama, A., Salh, N., Shelestov, A., Simonneaux, V., Traore, P.S., Traore, S.S., Koetz, B., 2019. Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the sen2-agri automated system in various cropping systems around the world. Remote Sensing of Environment 221, 551–568. URL: https://www.sciencedirect.com/science/article/pii/S0034425718305145, doi:https://doi.org/10.1016/j.rse.2018.11.007.

Domenzain, L.M., Gómez-Dans, J., Lewis, P.P., 2019. jgomezdans/prosail: Pip package bug fix release. URL: https://doi.org/10.5281/zenodo.2574925, doi:10.5281/zenodo.2574925.

Doxani, G., Vermote, E.F., Roger, J.C., Skakun, S., Gascon, F., Collison, A., De Keukelaere, L., Desjardins, C., Frantz, D., Hagolle, O., Kim, M., Louis, J., Pacifici, F., Pflug, B., Poilvé, H., Ramon, D., Richter, R., Yin, F., 2023. Atmospheric correction inter-comparison exercise, acix-ii land: An assessment of atmospheric correction processors for landsat 8 and sentinel-2 over land. Remote Sensing of Environment 285, 113412. URL: https://www.sciencedirect.com/science/article/pii/S0034425722005181, doi:https://doi.org/10.1016/j.rse.2022.113412.

Duan, S.B., Li, Z.L., Wu, H., Tang, B.H., Ma, L., Zhao, E., Li, C., 2014. Inversion of the prosail model to estimate leaf area index of maize, potato, and sunflower fields from unmanned aerial vehicle hyperspectral data. International Journal of Applied Earth Observation and Geoinformation 26, 12–20. URL: https://www.sciencedirect.com/science/article/pii/S0303243413000561, doi:https://doi.org/10.1016/j.jag.2013.05.007.

Duncan, W.G., 1971. Leaf angles, leaf area, and canopy photosynthesis1. Crop Science 11, cropsci1971.0011183X001100040006x. URL: https://acsess.onlinelibrary.wiley.com/doi/abs/10.2135/cropsci1971.0011183X001100040006x, doi:https://doi.org/10.2135/cropsci1971.0011183X001100040006x, arXiv:https://acsess.onlinelibrary.wiley.com/doi/pdf/10.2135/crops

Durbha, S.S., King, R.L., Younan, N.H., 2007. Support vector machines regression for retrieval of leaf area index from multiangle imaging spectroradiometer. Remote Sensing of Environment 107, 348–361. URL: https://www.sciencedirect.com/science/article/pii/S0034425706004159, doi:https://doi.org/10.1016/j.rse.2006.09.031. multi-angle Imaging SpectroRadiometer (MISR) Special Issue.

Estévez, J., Berger, K., Vicent, J., Rivera-Caicedo, J.P., Wocher, M., Verrelst, J., 2021. Top-of-atmosphere retrieval of multiple crop traits using variational heteroscedastic gaussian processes within a hybrid workflow. Remote Sensing 13. URL: https://www.mdpi.com/2072-4292/13/8/1589, doi:10.3390/rs13081589.

Fang, H., Baret, F., Plummer, S., Schaepman-Strub, G., 2019. An overview of global leaf area index (lai): Methods, products, validation, and applications. Reviews of Geophysics 57, 739–799. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018RG000608, doi:https://doi.org/10.1029/2018RG000608, arXiv:https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2018

Feret, J.B., de Boissieu, F., 2023. prospect: PROSPECT leaf radiative transfer model and inversion routines. URL: https://gitlab.com/jbferet/prospect. r package version 1.3.0.

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 Sensing of Environment 112, 3030–3043. URL: https://www.sciencedirect.com/science/article/pii/S0034425708000813, doi:https://doi.org/10.1016/j.rse.2008.02.012.

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 Sensing of Environment 252, 112173. URL: https://www.sciencedirect.com/science/article/pii/S0034425720305460, doi:https://doi.org/10.1016/j.rse.2020.112173.

Féret, J.B., Gitelson, A., Noble, S., Jacquemoud, S., 2017. Prospect-d: Towards modeling leaf optical properties through a complete lifecycle. Remote Sensing of Environment 193, 204–215. URL: https://www.sciencedirect.com/science/article/pii/S0034425717300962, doi: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., Ponzoni, F., 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 Sensing of Environment 231, 110959. URL: https://www.sciencedirect.com/science/article/pii/S0034425718305030, doi:https://doi.org/10.1016/j.rse.2018.11.002.

Gallo, I., Boschetti, M., Rehman, A.U., Candiani, G., 2023. Self-supervised

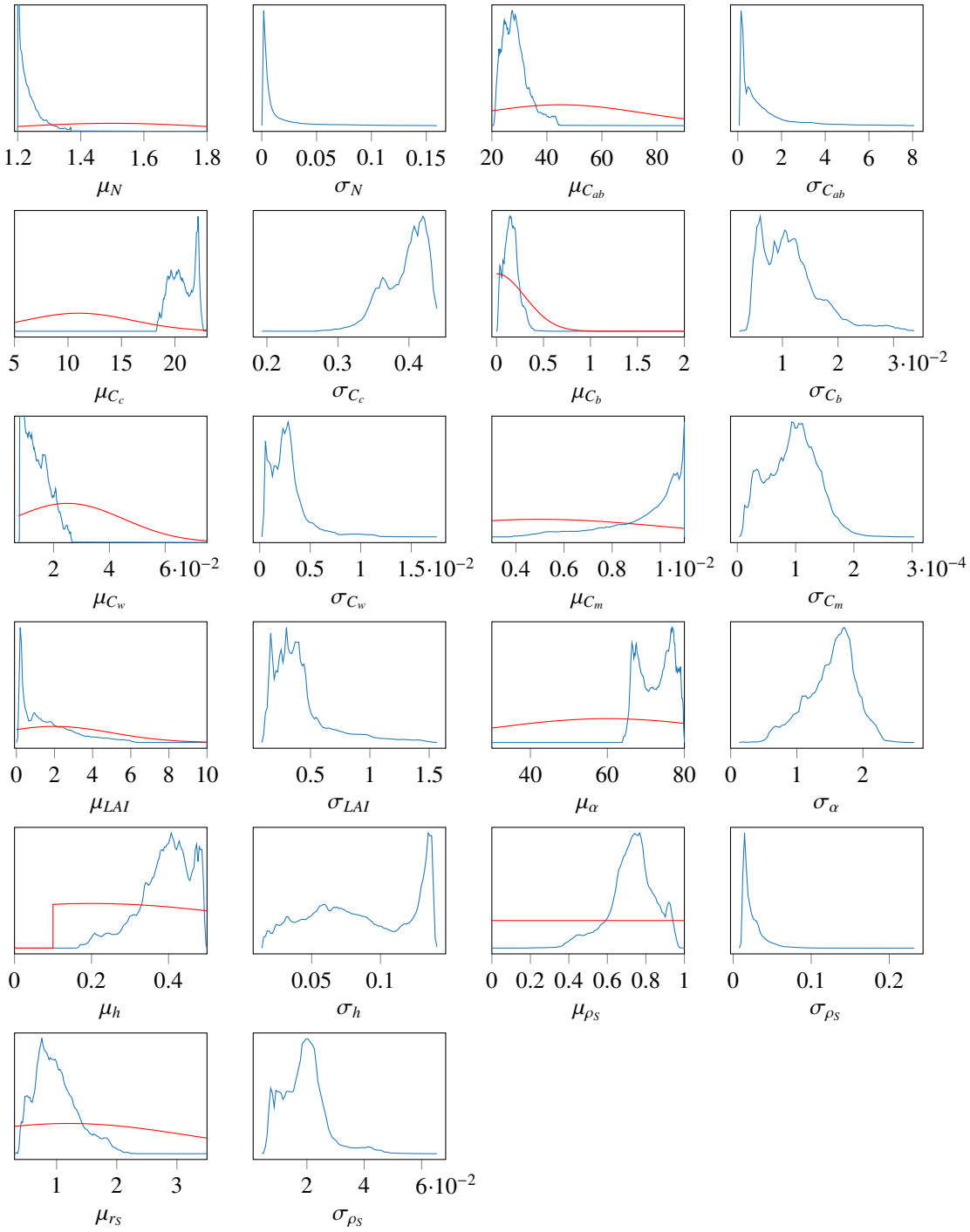


Figure 17: Blue: Histograms of PROSAIL variables (expected values and standard deviation of) inferred by PROSAIL-VAE on testing S2 data-set. Red: PROSAIL variables distributions well-established in the literature (see Table 3).

convolutional neural network learning in a hybrid approach framework to estimate chlorophyll and nitrogen content of maize from hyperspectral images. Remote Sensing 15, 4765. URL: http://dx.doi.org/10.3390/rs15194765, doi:10.3390/rs15194765.

GCOS, 2011. Systematic observation requirements for satellite- based products for climate 2011 update: Supplemental details to the satellite-based component of the "implementation plan for the global observing system for climate in support of the unfccc (2010 update).

Gitelson, A.A., Viña, A., Ciganda, V., Rundquist, D.C., Arkebauer, T.J., 2005. Remote estimation of canopy chlorophyll content in crops. Geophysical research letters 32.

Goodman, L.A., 1960. On the exact variance of products. Journal of the American Statistical Association 55, 708–713. URL: https://api.semanticscholar.org/CorpusID:120014324.

Hagolle, O., Huc, M., Desjardins, C., Auerand Rudolf Richter, S., 2017. Maja algorithm theoretical basis document. URL: https://doi.org/10.5281/zenodo.1209633, doi:10.5281/zenodo.1209633.

Hauser, L.T., Féret, J.B., An Binh, N., van der Windt, N., Ângelo F. Sil, Timmermans, J., Soudzilovskaia, N.A., van Bodegom, P.M., 2021. Towards scalable estimation of plant functional diversity from sentinel-2:

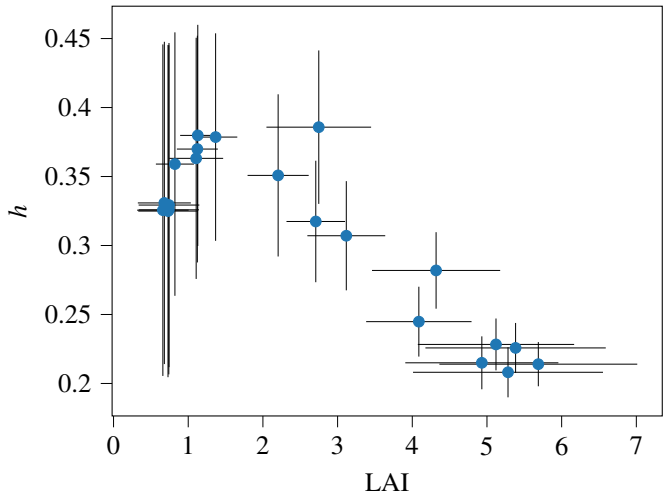


Figure 18: Scatter-plot of LAI versus hot-spot parameters predicted by a PROSAIL-VAE Pix-B2-L model. Blue dots are the expected value of LAI and hot-spot, and horizontal and vertical lines correspond are the std. The predictions are averaged over for a wheat crop parcel of the BelSAR campaign. The prediction are performed over the year 2018.

In-situ validation in a heterogeneous (semi-)natural landscape. Remote Sensing of Environment 262, 112505. URL: https://www.sciencedirect.com/science/article/pii/S0034425721002236, doi:https://doi.org/10.1016/j.rse.2021.112505.

Heiskanen, J., Rautiainen, M., Stenberg, P., Mõttus, M., Vesanto, V.H., Korhonen, L., Majasalmi, T., 2012. Seasonal variation in modis lai for a boreal forest area in finland. Remote Sensing of Environment 126, 104–115.

Higgins, I., Matthey, L., Pal, A., Burgess, C., Glorot, X., Botvinick, M., Mohamed, S., Lerchner, A., 2017. beta-VAE: Learning basic visual concepts with a constrained variational framework, in: International Conference on Learning Representations. URL: https://openreview.net/forum?id=Sy2fzU9gl.

Houborg, R., McCabe, M.F., 2018. A hybrid training approach for leaf area index estimation via cubist and random forests machine-learning. ISPRS Journal of Photogrammetry and Remote Sensing 135, 173–188. URL: https://www.sciencedirect.com/science/article/pii/S0924271617303209, doi:https://doi.org/10.1016/j.isprsjprs.2017.10.004.

Inglada, J., 2017. Sentinel-2 Agriculture Vegetation status DPM. Research Report. CESBIO. URL: https://hal.science/hal-02874654.

Jacquemoud, S., Bacour, C., Poilvé, H., Frangi, J.P., 2000. Comparison of four radiative transfer models to simulate plant canopies reflectance: Direct and inverse mode. Remote Sensing of Environment 74, 471–481. URL: https://www.sciencedirect.com/science/article/pii/S0034425700001395, doi:https://doi.org/10.1016/S0034-4257(00)00139-5.

Jacquemoud, S., Baret, F., 1990. Prospect: A model of leaf optical properties spectra. Remote Sensing of Environment 34, 75–91. URL: https://www.sciencedirect.com/science/article/pii/003442579090100Z, doi: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 Sensing of Environment 113, S56–S66. URL: https://www.sciencedirect.com/science/article/pii/S0034425709000765, doi:https://doi.org/10.1016/j.rse.2008.01.026. imaging Spectroscopy Special Issue.

Kamenova, I., Dimitrov, P., 2021. Evaluation of sentinel-2 vegetation indices for prediction of lai, fapar and fcover of winter wheat in bulgaria. European Journal of Remote Sensing 54, 89–108. URL: https://doi.org/10.1080/22797254.2020.1839359, doi:10.1080/22797254.2020.1839359, arXiv:https://doi.org/10.1080/22797254.2020.1839359.

Kergoat, L., Lafont, S., Douville, H., Berthelot, B., Dedieu, G., Planton, S., Royer, J.F., 2002. Impact of doubled co2 on global-scale leaf area index and evapotranspiration: Conflicting stomatal conductance and lai responses. Journal of Geophysical Research: Atmospheres 107, ACL 30–1–ACL 30–16. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2001JD001245, doi:https://doi.org/10.1029/2001JD001245, arXiv:https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2001

Kingma, D.P., Welling, M., 2014. Auto-encoding variational bayes, in: Bengio, Y., LeCun, Y. (Eds.), 2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings.

Marie Weiss, Frédéric Baret, Marc Leroy, Olivier Hautecoeur, Cédric Bacour, Laurent Prévot, Nadine Bruguier, 2002. Validation of neural net techniques to estimate canopy biophysical variables from remote sensing data. Agronomie 22, 547–553. URL: https://doi.org/10.1051/agro:2002036, doi:10.1051/agro:2002036.

Marie Weiss, Frédéric Baret, Ranga B. Myneni, Agnès Pragnère, Youri Knyazikhin, 2000. Investigation of a model inversion technique to estimate canopy biophysical variables from spectral and directional reflectance data. Agronomie 20, 3–22. URL: https://doi.org/10.1051/agro:2000105, doi:10.1051/agro:2000105.

Mokhtari, A., Noory, H., Vazifedoust, M., 2018. Improving crop yield estimation by assimilating lai and inputting satellite-based surface incoming solar radiation into swap model. Agricultural and Forest Meteorology 250-251, 159–170. URL: https://www.sciencedirect.com/science/article/pii/S0168192317306706, doi:https://doi.org/10.1016/j.agrformet.2017.12.250.

Orban, A., Defrere, D., Barbier, C., 2021. Belsar : the first belgian airborne campaign for l-band, full polarimetric bistatic and interferometric sar acquisitions over an agricultural site in belgium, in: EUSAR 2021; 13th European Conference on Synthetic Aperture Radar, pp. 1–4.

Origo, N., Gorroño, J., Ryder, J., Nightingale, J., Bialek, A., 2020. Fiducial reference measurements for validation of sentinel-2 and proba-v surface reflectance products. Remote Sensing of Environment 241, 111690. URL: https://www.sciencedirect.com/science/article/pii/S0034425720300596, doi:https://doi.org/10.1016/j.rse.2020.111690.

Panigrahi, N., Das, B.S., 2021. Evaluation of regression algorithms for estimating leaf area index and canopy water content from water stressed rice canopy reflectance. Information Processing in Agriculture 8, 284–298. URL: https://www.sciencedirect.com/science/article/pii/S2214317320301876, doi:https://doi.org/10.1016/j.inpa.2020.06.002.

Raj, S., Roodbar, S., Brinkley, C., Wolfe, D.W., 2022. Food security and climate change: Differences in impacts and adaptation strategies for rural communities in the global south and north. Frontiers in Sustainable Food Systems 5. URL: https://www.frontiersin.org/articles/10.3389/fsufs.2021.691191, doi:10.3389/fsufs.2021.691191.

Rivera, J.P., 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 Sensing 5, 3280–3304. URL: https://www.mdpi.com/2072-4292/5/7/3280, doi:10.3390/rs5073280.

Rouquié, B., Hagolle, O., Bréon, F.M., Boucher, O., Desjardins, C., Rémy, S., 2017. Using copernicus atmosphere monitoring service products to constrain the aerosol type in the atmospheric correction processor maja. Remote Sensing 9, 1230.

Sehgal, V.K., Chakraborty, D., Sahoo, R.N., 2016. Inversion of radiative transfer model for retrieval of wheat biophysical parameters from broadband reflectance measurements. Information Processing in Agriculture 3, 107–118. URL: https://www.sciencedirect.com/science/article/pii/S2214317316300075, doi:https://doi.org/10.1016/j.inpa.2016.04.001.

Srinet, R., Nandy, S., Patel, N., 2019. Estimating leaf area index and light extinction coefficient using random forest regression algorithm in a tropical moist deciduous forest, india. Ecological Informatics 52, 94–102.

Svendsen, D.H., Hernández-Lobato, D., Martino, L., Laparra, V., Moreno-Martínez, Á., Camps-Valls, G., 2021. Inference over radiative transfer models using variational and expectation maximization methods. Machine Learning 112, 921–937. URL: http://dx.doi.org/10.1007/s10994-021-05999-4, doi:10.1007/s10994-021-05999-4.

Svendsen, D.H., Martino, L., Campos-Taberner, M., García-Haro, F.J., Camps-Valls, G., 2018. Joint gaussian processes for biophysical parameter retrieval. IEEE Transactions on Geoscience and Remote Sensing 56, 1718–1727. doi:10.1109/TGRS.2017.2767205.

Tuia, D., Verrelst, J., Alonso, L., Perez-Cruz, F., Camps-Valls, G., 2011. Multioutput support vector regression for remote sensing biophysical parameter estimation. IEEE Geoscience and Remote Sensing Letters 8, 804–808. doi:10.1109/LGRS.2011.2109934.

Tupin, F., Inglada, J., Nicolas, J.M., 2014. Remote Sensing Imagery. John Wiley & Sons.

Verhoef, W., . Theory of Radiative Transfer Models Applied in Optical Remote Sensing of Vegetation Canopies.

Verhoef, W., 1984. Light scattering by leaf layers with application to canopy reflectance modeling: The sail model. Remote Sensing of Environment 16, 125–141. URL: https://www.sciencedirect.com/science/article/pii/0034425784900579, doi:https://doi.org/10.1016/0034-4257(84)90057-9.

Verhoef, W., Jia, L., Xiao, Q., Su, Z., 2007. Unified optical-thermal four-stream radiative transfer theory for homogeneous vegetation canopies. IEEE Transactions on Geoscience and Remote Sensing 45, 1808–1822. doi:10.1109/TGRS.2007.895844.

Verrelst, J., Malenovský, Z., Tol, C., Camps-Valls, G., Gastellu-Etchegorry, J.P., Lewis, P., North, P., Moreno, J., 2019. Quantifying vegetation biophysical variables from imaging spectroscopy data: A review on retrieval methods. Surveys in Geophysics 40. doi:10.1007/s10712-018-9478-y.

Verrelst, J., Rivera, J.P., Moreno, J., Camps-Valls, G., 2013. Gaussian processes uncertainty estimates in experimental sentinel-2 lai and leaf chlorophyll content retrieval. ISPRS journal of photogrammetry and remote sensing 86, 157–167.

Verrelst, J., Rivera, J.P., Veroustraete, F., Muñoz-Marí, J., Clevers, J.G., Camps-Valls, G., Moreno, J., 2015. Experimental sentinel-2 lai estimation using parametric, non-parametric and physical retrieval methods – a comparison. ISPRS Journal of Photogrammetry and Remote Sensing 108, 260–272. URL: https://www.sciencedirect.com/science/article/pii/S0924271615001239, doi:https://doi.org/10.1016/j.isprsjprs.2015.04.013.

Wang, J., Lopez-Lozano, R., Weiss, M., Buis, S., Li, W., Liu, S., Baret, F., Zhang, J., 2022. Crop specific inversion of prosail to retrieve green area index (gai) from several decametric satellites using a bayesian framework. Remote Sensing of Environment 278, 113085. URL: https://www.sciencedirect.com/science/article/pii/S0034425722001997, doi:https://doi.org/10.1016/j.rse.2022.113085.

Wang, L., Good, S.P., Caylor, K.K., 2014. Global synthesis of vegetation control on evapotranspiration partitioning. Geophysical Research Letters 41, 6753–6757. URL: https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014GL061439, doi:https://doi.org/10.1002/2014GL061439, arXiv:https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1002/2014GL061439.

Wang, W.M., Li, Z.L., Su, H.B., 2007. Comparison of leaf angle distribution functions: Effects on extinction coefficient and fraction of sunlit foliage. Agricultural and Forest Meteorology 143, 106–122. URL: https://www.sciencedirect.com/science/article/pii/S016819230600390X, doi:https://doi.org/10.1016/j.agrformet.2006.12.003.

Weiss, M., Baret, F., 2016. Sentinel-2 toolbox level 2 biophysical product algorithms.

Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: A meta-review. Remote Sensing of Environment 236, 111402. URL: https://www.sciencedirect.com/science/article/pii/S0034425719304213, doi:https://doi.org/10.1016/j.rse.2019.111402.

Xie, Q., Dash, J., Huete, A., Jiang, A., Yin, G., Ding, Y., Peng, D., Hall, C.C., Brown, L., Shi, Y., Ye, H., Dong, Y., Huang, W., 2019. Retrieval of crop biophysical parameters from sentinel-2 remote sensing imagery. International Journal of Applied Earth Observation and Geoinformation 80, 187–195. URL: https://www.sciencedirect.com/science/article/pii/S0303243419301199, doi:https://doi.org/10.1016/j.jag.2019.04.019.

Zeng, L., Wardlow, B.D., Xiang, D., Hu, S., Li, D., 2020. A review of vegetation phenological metrics extraction using time-series, multispectral satellite data. Remote Sensing of Environment 237, 111511. URL: https://doi.org/10.1016/j.rse.2019.111511, doi:10.1016/j.rse.2019.111511.

Zhang, R., Yang, P., Liu, S., Wang, C., Liu, J., 2022. Evaluation of the methods for estimating leaf chlorophyll content with spad chlorophyll meters. Remote Sensing 14. URL: https://www.mdpi.com/2072-4292/14/20/5144, doi:10.3390/rs14205144.

Zheng, G., Moskal, L.M., 2009. Retrieving leaf area index (lai) using remote sensing: Theories, methods and sensors. Sensors 9, 2719–2745. URL: https://www.mdpi.com/1424-8220/9/4/2719, doi:10.3390/s90402719.

Zhu, X., Li, C., D.D.S., L.T., 2018. Look-up-table approach for leaf area index retrieval from remotely sensed data based on scale information. Optical Engineering 57, 033104. URL: https://doi.org/10.1117/1.OE.57.3.033104, doi:10.1117/1.OE.57.3.033104.

Zhu, X., Yang, Q., Chen, X., Ding, Z., 2023. An approach for joint estimation of grassland leaf area index and leaf chlorophyll content from uav hyperspectral data. Remote Sensing 15. URL: https://www.mdpi.com/2072-4292/15/10/2525, doi:10.3390/rs15102525.

Zérah, Y., Valero, S., Inglada, J., 2023. Physics-driven probabilistic deep learning for the inversion of physical models with application to phenological parameter retrieval from satellite times series. IEEE Transactions on Geoscience and Remote Sensing 61, 1–23. doi:10.1109/TGRS.2023.3284992.

Appendix A. Sentinel-2 bands simulation from PROSAIL canopy spectra

For each band i with spectral range $\left[\lambda_{\text{low},i}, \lambda_{\text{up},i}\right]$, the reflectance is derived from canopy spectra output $\rho(\lambda)$ by PROSAIL, the solar irradiance spectrum $E_s(\lambda)$ (see Fig. A.20), and the spectral response of the Sentinel-2 sensor $S_i(\lambda)$ (see Fig. A.19) (Tupin et al., 2014):

$$\rho_i = \frac{\int_{\lambda_{\text{low},i}}^{\lambda_{\text{up},i}} S_i(\lambda)\, E_S(\lambda)\, \rho(\lambda)\, d\lambda}{\int_{\lambda_{\text{low},i}}^{\lambda_{\text{up},i}} S_i(\lambda)\, E_S(\lambda)\, d\lambda}. \tag{A.1}$$

Appendix B. Band reconstruction capabilities of PROSAIL-VAE-L

Fig. B.21 shows the reconstruction performances obtained by the trained model on the S2 testing data-set described in Section 3.2. The scatter plots compare original S2 band reflectance values against reflectances reconstructed by the trained PROSAIL-VAE-L configuration. The results show that the reconstructions predicted by the trained encoder match original spectral S2 bands.

The assessment of the results can be also done with Fig. B.22, which illustrates visible and infra-red color compositions and their corresponding reconstructions. Visual results corroborate the accurate reconstruction of crops areas for both color compositions.

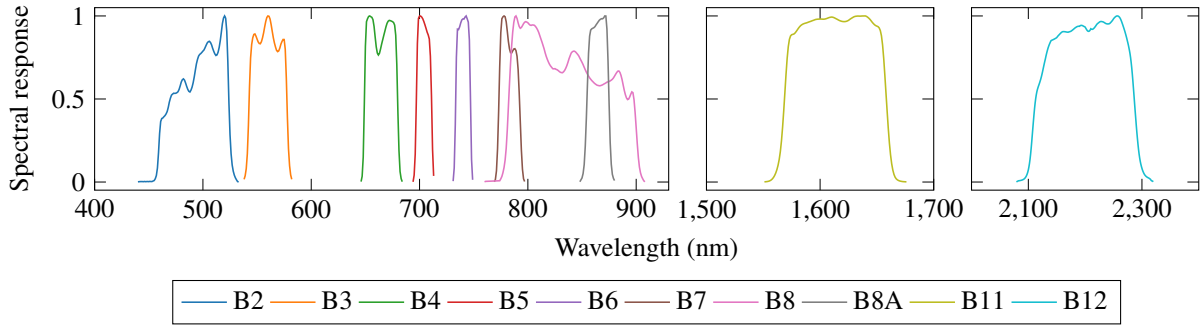


Figure A.19: Spectral response of the Sentinel-2A multi spectral instrument (MSI) sensor (only the bands used by PROSAIL-VAE are shown).

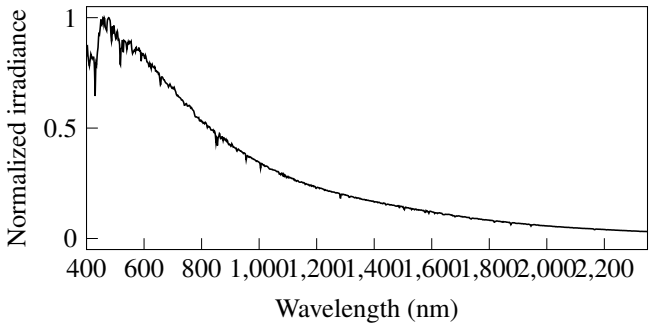


Figure A.20: Normalized solar irradiance spectra

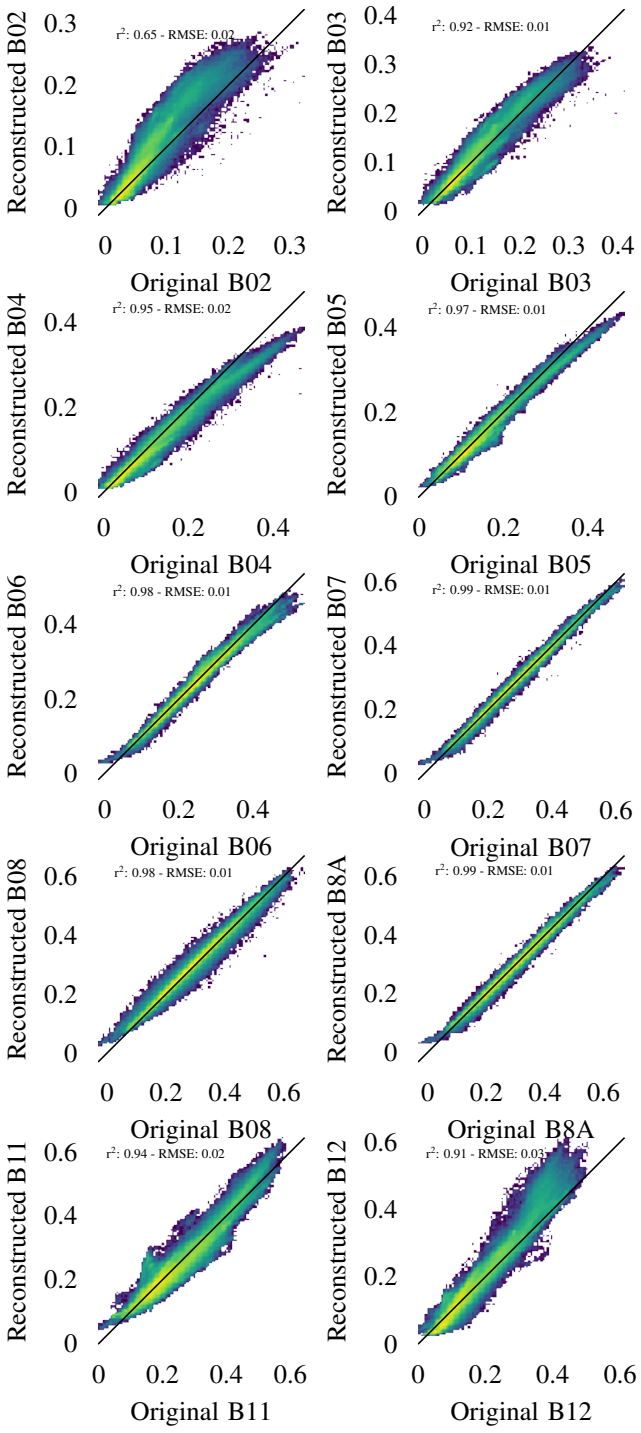


Figure B.21: Scatter plots comparing the original S2 band reflectance values against reflectances reconstructed by PROSAIL-VAE-L on S2 test data-set.

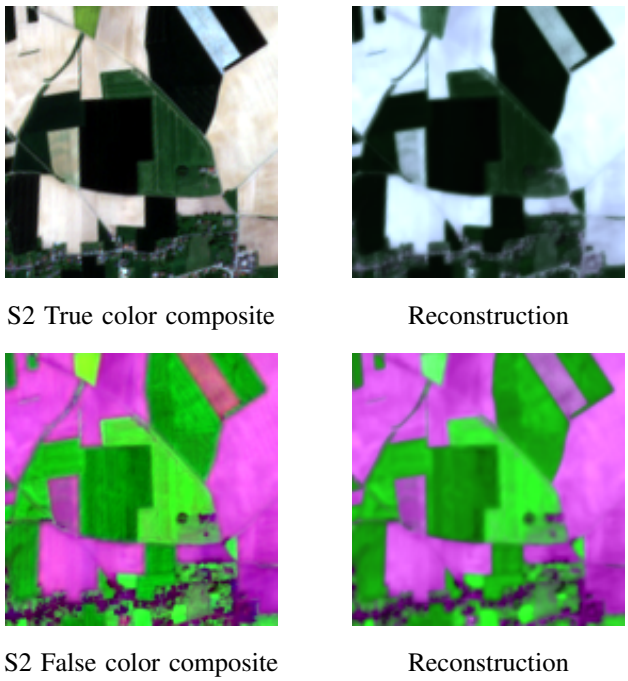


Figure B.22: S2 ROI image from test data-set acquired on 2023-05-13 and located at T31UFS tile (Southern Belgium). First column shows true and false color composites constructed by original S2 reflectance values whereas PROSAIL-VAE reconstruction results are displayed at the second column. The RGB true color composite corresponds to bands (B4, B3, B2). The RGB false color composite corresponds to bands (B11,B8,B5)

Appendix C. Biophysical variable prediction on an image

In Fig. C.23 are showed the BV expectations and std, besides the LAI, predicted by PROSAIL-VAE from the example patch of Fig. B.22. This S2 image patch contains both crop vegetation elements, and areas without vegetation: roads, buildings, bare soils. In areas devoid of vegetation, variables besides ρ_S and r_S are irrelevant, even though some of them exhibit high values (C_c, C_m, α, h). In particular, the soil wetness factor ρ_S is well correlated to the bare soil areas. For areas with vegetation, N, C_{ab}, C_b, C_w seem correlated to the density of vegetation. Overall, the predicted std are correlated to the expected values.

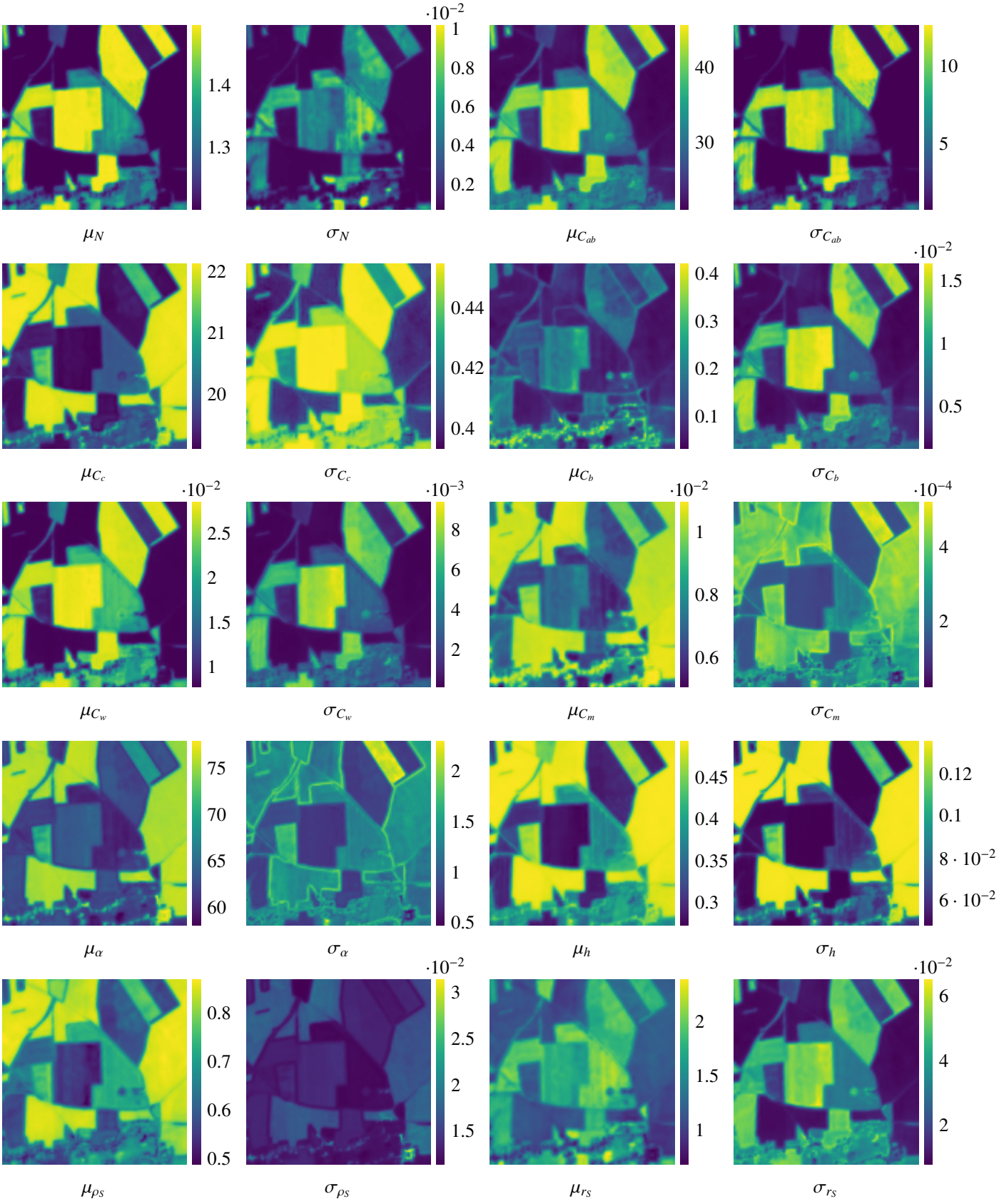


Figure C.23: Inference of BV by PROSAIL-VAE Pix-B2-L. BV are predicted at pixel level in a ROI of the MGRS tile T31UFS (Southern Belgium) on 2023-06-01.

Appendix D. Visual comparison between PROSAIL-VAE and SNAP results

In Fig. D.24, the predictions of SL2P and PROSAIL-VAE over a patch are compared. For both methods, the LAI, CCC and CWC are well correlated to the presence of vegetation (see Fig. B.22). The predictions made by PROSAIL-VAE looks sharper than the prediction by SL2P, and some structures, such as roads and the shapes of the parcels, are better outlined by PROSAIL-VAE. Within the parcels, the predictions of PROSAIL-VAE seem more homogeneous. The CCC and CWC tend to be predicted with higher values within the parcels. For LAI, CCC and CWC, the stds are correlated to high expected values.

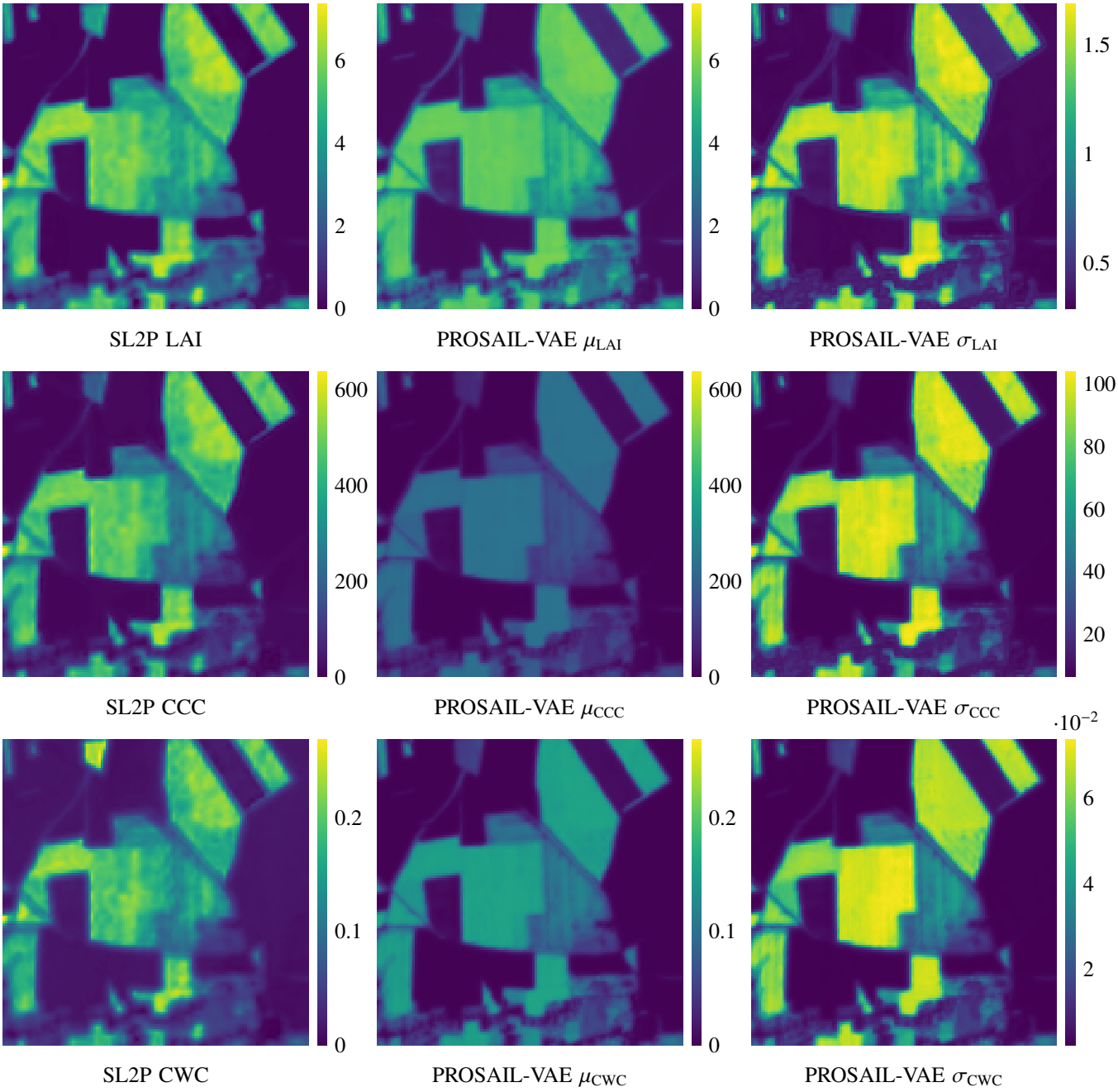


Figure D.24: Comparison of LAI, CCC and CWC results obtained by SL2P and a PROSAIL-VAE Pix-B2-L model. The image ROI is located at T31UFS tile (Southern Belgium) and acquired on 2018-06-01.

Appendix E. PROSAIL variables inferred co-distribution

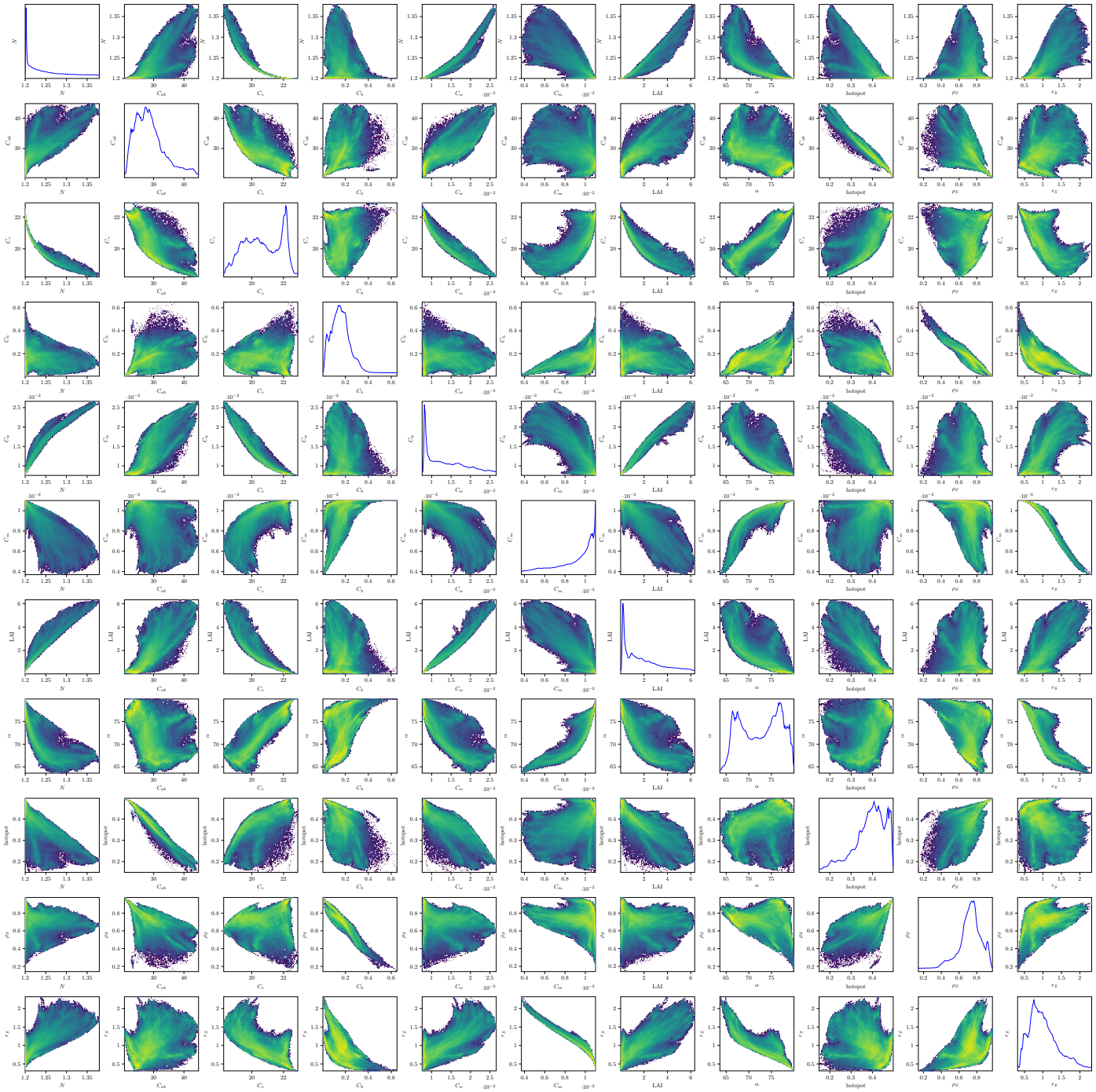


Figure E.25: Scatter-plot of the expected values of all pairs of variables of PROSAIL (PROSPECT-5 + 4SAIL), predicted by PROSAIL-VAE over the S2 testing data-set.

Appendix F. Acronyms

ANITI Natural Intelligence Toulouse Institute. 1
ANN artificial neural network. 2

BP Biophysical Processor. 2, 3
BV biophysical variables. 1–4, 10, 13, 14, 22, 23
BVNET biophysical variable neural network. 2–4, 8–10, 14, 15

CCC canopy chlorophyll content. 1, 2, 4, 5, 7, 9, 10, 12–17, 24, 25
CNN convolutional neural network. 3
CWC canopy water content. 2, 4, 13, 17, 24, 25

DHP digital hemispheric photography. 7

ECV essential climate variable. 1
ESU elementary sampling unit. 7

F-COVER fraction of vegetation cover. 4
FAPAR fraction of absorbed photosynthetically active radiation. 4
FRM4Veg fiducial reference measurements for vegetation. 5, 7, 8

GAI green area index. 7
GCOS Global Climat Observing System. 1
GP Gaussian process. 2

HPC high performance computing. 16

KLD Kullback-Leibler divergence. 3, 9, 10, 13, 14

LAI leaf area index. 1–7, 9–11, 13–17, 19, 22, 24, 25
LCC leaf chlorophyll content. 7
LIDF leaf inclination distribution function. 2
LR learning rate. 10
LUT look-up-tables. 1

MAJA MACCS-ATCOR joint algorithm. 8
MC Monte Carlo. 3
MCEM Monte Carlo expectation maximization. 2
MCMC Markov Chain Monte Carlo. 2
MGRS Military Grid Reference System. 5, 6, 23
MLP multi-layer perceptron. 3
MPIW mean prediction interval width. 4, 11–14
MPSR multiple probabilistic supervised regression. 4, 8, 10–14
MSE mean squared error. 4
MSI multi spectral instrument. 20

PAI plant area index. 7
PICP prediction interval coverage probability. 4, 11–14

RF random forests. 2
RMSE root mean squared error. 4, 8–14
ROI region of interest. 5, 7, 21, 23, 25
RTM radiative transfer model. 1, 2, 15, 16

S2 Sentinel-2. 3, 5, 7–10, 13, 14, 16–18, 21, 22, 26
SAIL Scattering by Arbitrary Inclined Leaves. 2, 10, 14, 15
SL2P Simplified Level 2 Product Prototype Processor. 1–5, 9–17, 24, 25
SNAP Sentinel Application Platform. 2, 3, 24
std standard deviation. 9, 14, 19, 22, 24

TN truncated Normal. 3, 6, 9, 10

VAE variational autoencoder. 2, 3

Appendix G. Notation

C_{ab} Chlorophyll a+b concentration. 3, 4, 6, 10, 13, 14, 22
C_c Carotenoid concentration. 3, 4, 6, 14, 22
C_b Brown pigments content. 3, 4, 6, 14, 22
C_m Dry matter content. 3–6, 14, 22
C_w Water equivalent thickness. 3–6, 14, 22
$C_{w,\mathbf{rel}}$ relative water content. 5

h Hotspot parameter. 3, 4, 6, 22

$\mathbb{KL}$ Kullback-Leibler divergence. 9, 14

α mean leaf inclination. 2–4, 6, 14, 22

N Leaf structure parameter. 3, 4, 6, 14, 22

θ_O Observer zenith angle. 3

ρ_S Wet soil factor. 3–6, 22

ψ_{SO} Relative azimuth angle. 3
r_S Soil brightness factor. 3, 4, 6, 22

θ_S Solar zenith angle. 3