

JOINT GAUSSIAN PROCESSES FOR INVERSE MODELING

Daniel Heestermans Svendsen[†], Luca Martino[†], Manuel Campos-Taberner[‡] and Gustau Camps-Valls[†]

[†]Image Processing Laboratory (IPL), Universitat de València, Spain

[‡]Faculty of Physics, Universitat de València, Spain

ABSTRACT

Solving inverse problems is central in geosciences and remote sensing. Very often a *mechanistic* physical model of the system exists that solves the forward problem. Inverting the implied radiative transfer model (RTM) equations numerically implies, however, challenging and computationally demanding problems. Statistical models tackle the inverse problem and predict the biophysical parameter of interest from radiance data, exploiting either *in situ* data or simulated data from an RTM. We introduce a novel nonlinear and nonparametric statistical inversion model which incorporates both real observations and RTM-simulated data. The proposed Joint Gaussian Process (JGP) provides a solid framework for exploiting the regularities between the two types of data, in order to perform inverse modeling. Advantages of the JGP method over competing strategies are shown on both a simple toy example and in leaf area index (LAI) retrieval from Landsat data combined with simulated data generated by the PROSAIL model.

Index Terms— Forward, inverse, modeling, Vegetation monitoring, Gaussian process regression, kernel methods

1. INTRODUCTION

Synoptikos, σύνοψις
“Affording a general view of a whole.”

Solving inverse problems is a recurrent topic of research in Physics in general, and in geosciences and remote sensing in particular. After all, Science is about making inferences about physical parameters from sensory data. In general, mechanistic models implement the laws of Physics and allow us to compute the data values given a model [1]. This is known as the *forward* problem. In the *inverse* problem, the aim is to reconstruct the model from a set of measurements.

The inverse problem is at the core of remote sensing and geosciences. A very relevant problem is that of estimating vegetation properties from remote sensing data. Accurate inverse models help determine the phenological stage and health status (e.g., development, productivity, stress) of crops and forests [2], which has important societal, environmental

and economical implications. Leaf chlorophyll content (*Chl*), leaf area index (LAI), and fractional vegetation cover (FVC) are among the most important vegetation parameters [3, 4].

Methods for model inversion and parameter retrieval can be roughly separated in three main families: statistical, physical and hybrid methods [5]. *Statistical inversion* predicts a biogeophysical parameter of interest using a training dataset of input-output data pairs coming from concurrent measurements of the parameter of interest (e.g. leaf area index -LAI-) and the corresponding observations (e.g. reflectances). Statistical methods typically outperform other approaches, but ground truth measurements are necessary. *Physical inversion* reverses RTMs by searching for similar spectra in look-up-tables (LUTs) and assigning the most closest parameter. This requires selecting an appropriate cost function, and generating a rich, representative LUT from the RTM. The use of RTMs to generate data sets is a common practice, and especially convenient because acquisition campaigns are very costly (in terms of time, money, and human resources) and usually limited in terms of parameter combinations. Finally, *hybrid inversion* exploits the input-output data generated by RTM simulations and train statistical regression models to invert the RTM model. Hybrid models combine the flexibility and scalability of machine learning while respecting the physics encoded in the RTMs. Currently, kernel machines in general [6], and Bayesian non-parametric approaches such as Gaussian Process (GP) regression [7] in particular, are among the preferred regression models [8, 9].

While hybrid inversion is practical when no *in situ* data is available, it seems intuitive to let predictions be guided by actual measurements whenever they are present. Likewise, when only very few real *in situ* measurements are available, it is sensible to incorporate simulated data from RTMs to properly ground the models. Therefore, in this paper we extend the hybrid inversion, proposing a statistical method which performs nonlinear and nonparametric inversion blending both real and simulated data. The Joint Gaussian Process (JGP) proposed in §2 exploits the regularities between them, and provides a solid framework for incorporating physical knowledge into a GP. We give empirical evidence of performance in §4 on LAI retrievals from Landsat observations and PROSAIL simulated data assimilation. We conclude in §5 with some remarks and an outline of future work.

The research was funded by the European Research Council (ERC) under the ERC-CoG-2014 SEDAL project (grant agreement 647423), and the Spanish Ministry of Economy and Competitiveness (MINECO) through the project TIN2015-64210-R.

2. JOINT GAUSSIAN PROCESSES

2.1. Notation: Forward and Inverse Modeling

Notationally, a forward model describing the system is expressed as $\mathbf{x} = g(\mathbf{y}, \boldsymbol{\omega})$, where \mathbf{x} is a measurement obtained by the satellite (e.g. radiance); the vector \mathbf{y} represents the state of the biophysical variables of the Earth (which we desire to infer or predict and is often referred to as *outputs* in inverse modeling approach); $\boldsymbol{\omega}$ contains a set of controllable conditions (e.g. wavelengths, viewing direction, time, Sun position, and polarization); and $g(\cdot)$ is a function which relates \mathbf{y} with \mathbf{x} . Such a function g is typically considered to be nonlinear, smooth and continuous. Our goal is to obtain an inverse model $f(\cdot) = g^{-1}(\cdot)$ parametrized by $\boldsymbol{\theta}$, which approximates the biophysical variables \mathbf{y} using given the data received by the satellite \mathbf{x} , i.e. $\hat{\mathbf{y}} = f(\mathbf{x}, \boldsymbol{\theta})$.

2.2. Gaussian Process (GP) Regression

GPs are state-of-the-art tools for regression and function approximation. Let us consider a set of n pairs of observations or measurements $\{\mathbf{x}_i, y_i\}_{i=1}^n$, perturbed by an additive independent noise. More specifically, we assume the following model,

$$y_i = f(\mathbf{x}_i) + e_i, \quad e_i \sim \mathcal{N}(0, \sigma^2), \quad (1)$$

where $f(\mathbf{x})$ is an unknown latent function and $\mathbf{x} \in \mathbb{R}^d$. In a GP approach, we assume $f(\mathbf{x}) \sim \mathcal{GP}(\mathbf{0}, k(\mathbf{x}, \mathbf{x}'))$ where $k(\mathbf{x}, \mathbf{x}')$ is a covariance (kernel) function, for instance, $k(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / (2\lambda^2))$ [7].¹ Then, the hyper-parameters to be tuned of the standard GP model are $\boldsymbol{\theta} = [\lambda, \sigma]$ where λ determines the width of the kernel function and σ is the standard deviation of the additive Gaussian noise e_i . The goal is to learn the latent function $f(\mathbf{x})$ given the received data points $\mathcal{D}_n = \{\mathbf{x}_i, y_i\}_{i=1}^n$. Considering a test location \mathbf{x}_* and the noise-free output $f_* = f(\mathbf{x}_*)$, the GP predictive density is

$$\begin{aligned} p(f_* | \mathbf{x}_*, \mathcal{D}_n) &\sim \mathcal{N}(\mu_{\text{GP}}(\mathbf{x}_*), \sigma_{\text{GP}}^2(\mathbf{x}_*)) \\ \mu_{\text{GP}}(\mathbf{x}_*) &= \mathbf{k}_*^\top (\mathbf{K}_{nn} + \sigma^2 \mathbf{I}_n)^{-1} \mathbf{y} = \mathbf{k}_*^\top \boldsymbol{\alpha} \\ \sigma_{\text{GP}}^2(\mathbf{x}_*) &= k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^\top (\mathbf{K}_{nn} + \sigma^2 \mathbf{I}_n)^{-1} \mathbf{k}_*, \end{aligned}$$

where $\mathbf{k}_* = [k(\mathbf{x}_*, \mathbf{x}_1), \dots, k(\mathbf{x}_*, \mathbf{x}_n)]^\top$ is a vector of dimension $n \times 1$, \mathbf{K}_{nn} is a $n \times n$ kernel matrix with entries $\mathbf{K}_{i,j} := k(\mathbf{x}_i, \mathbf{x}_j)$, and $\boldsymbol{\alpha} = (\mathbf{K}_{nn} + \sigma^2 \mathbf{I}_n)^{-1} \mathbf{y}$ is the solution weight vector.

2.3. Joint Gaussian Process (JGP) Regression

In this section, we assume that the dataset \mathcal{D}_n is formed by two disjoint sets: one set of r real data pairs, $\mathcal{D}_r = \{(\mathbf{x}_i, y_i)\}_{i=1}^r$, and one set of s RTM-simulated pairs $\mathcal{D}_s = \{(y_j, \mathbf{x}_j)\}_{j=1}^s$, so that $n = r + s$ and $\mathcal{D}_n = \mathcal{D}_r \cup \mathcal{D}_s$. In matrix

¹Essentially, a GP is a stochastic process whose marginals are distributed as a multivariate Gaussian densities.

form, we have $\mathbf{X}_r \in \mathbb{R}^{r \times d}$, $\mathbf{y}_r \in \mathbb{R}^{1 \times r}$, $\mathbf{X}_s \in \mathbb{R}^{s \times d}$, $\mathbf{y}_s \in \mathbb{R}^{1 \times s}$, containing all the inputs and outputs of \mathcal{D}_r and \mathcal{D}_s , respectively. Moreover, the $n \times 1$ vector $\mathbf{y} = [\mathbf{y}_r, \sqrt{\gamma} \mathbf{y}_s]^\top$ contains all the n outputs, where the $\gamma > 0$ parameter accounts for the importance of the two sources of information relative to each other.

Let us now define a mapping function to a Hilbert space \mathcal{H} of dimension H , i.e. a transformation $\phi(\mathbf{x}) : \mathbb{R}^d \rightarrow \mathcal{H}$ which defines a $1 \times H$ vector. Abusing of the notation, the mapped data matrices $\Phi_r = \phi(\mathbf{X}_r)$ of dimension $r \times H$, $\Phi_s = \phi(\mathbf{X}_s)$ of dimension $s \times H$, and $\Phi_n := [\Phi_r^\top, \sqrt{\gamma} \Phi_s^\top]^\top$ of dimension $n \times H$. We assume that the latent function in Eq. (1) has the form $f(\mathbf{x}) = \Phi_n \mathbf{w}$, where the parameter vector \mathbf{w} is an $H \times 1$ unknown and it must be inferred. Then the likelihood function is $p(\mathbf{y} | \Phi_n, \mathbf{w}) \sim \mathcal{N}(\Phi_n \mathbf{w}, \sigma^2 \mathbf{I}_n)$ where \mathbf{I}_n is an $n \times n$ identity matrix. We consider a Gaussian prior of zero mean and the $H \times H$ covariance matrix Σ_w over the parameter vector, $\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \Sigma_w)$. Thus, the posterior density is $p(\mathbf{w} | \Phi_n, \mathbf{y}) \propto p(\mathbf{y} | \Phi_n, \mathbf{w}) p(\mathbf{w}) \sim \mathcal{N}(\bar{\mathbf{w}}, \mathbf{A}^{-1})$, where $\bar{\mathbf{w}} = \mathbf{A}^{-1} \Phi_n^\top \mathbf{y}$ and $\mathbf{A}^{-1} = (\sigma^{-2} \Phi_n^\top \Phi_n + \Sigma_w^{-1})^{-1}$. The predictive distribution of f_* at \mathbf{x}_* is

$$p(f_* | \mathbf{x}_*, \mathcal{D}_n) \sim \mathcal{N}(\sigma^{-2} \phi(\mathbf{x}_*) \mathbf{A}^{-1} \Phi_n^\top \mathbf{y}, \phi(\mathbf{x}_*) \mathbf{A}^{-1} \phi(\mathbf{x}_*)^\top).$$

This primal expression however, is not convenient whenever we ignore the dimensionality H of the space \mathcal{H} or when it is higher than the number of data points n [6]. Fortunately, after some algebraic manipulation, and defining the reproducing kernel function $k(\mathbf{x}_i, \mathbf{x}_j) := \phi(\mathbf{x}_i) \Sigma_w \phi(\mathbf{x}_j)^\top$ [6, 7], we can obtain a dual expression of the predictive distribution of JGP (similarly to the standard GP),

$$\begin{aligned} p(f_* | \mathbf{x}_*, \mathcal{D}_n) &\sim \mathcal{N}(\mu_{\text{JGP}}(\mathbf{x}_*), \sigma_{\text{JGP}}^2(\mathbf{x}_*)) \\ \mu_{\text{JGP}}(\mathbf{x}_*) &= \mathbf{k}_*^\top \tilde{\mathbf{K}}_{nn}^{-1} \mathbf{y} = \mathbf{k}_*^\top \boldsymbol{\beta} \\ \sigma_{\text{JGP}}^2(\mathbf{x}_*) &= k(\mathbf{x}_*, \mathbf{x}_*) - \mathbf{k}_*^\top \tilde{\mathbf{K}}_{nn}^{-1} \mathbf{k}_*, \end{aligned}$$

where

$$\tilde{\mathbf{K}}_{nn}^{-1} = (\mathbf{K}_{nr} \mathbf{K}_{rn} + \gamma \mathbf{K}_{ns} \mathbf{K}_{sn} + \sigma^2 \mathbf{K}_{nn})^{-1} [\mathbf{K}_{nr} \quad \gamma \mathbf{K}_{ns}],$$

while the subscripts of the kernel matrices indicate their sizes and the samples involved in their calculation, with $\mathbf{K}_{ij} = \Phi_i \Sigma_w \Phi_j^\top$, $i, j \in \{r, s, n\}$. Note that when $\gamma = 0$ the standard GP is obtained, otherwise γ acts as an extra regularization term accounting for the relative importance of the real and the simulated data points. The hyperparameters of the JGP are $\boldsymbol{\theta} = [\lambda, \gamma, \sigma]$, which can be selected by maximizing the marginal likelihood of the observations as usual in the GP framework, or by standard cross-validation procedures.

2.4. An illustrative example

Let us exemplify the core idea of JGP in a toy example. Imagine that we are measuring the position of a mass on a perfect spring, subject to constant friction. Now, any physicist knows very well the solution to the damped harmonic oscillator problem and is thus able to simulate data from it which may then be used in a JGP framework. In remote sensing ap-

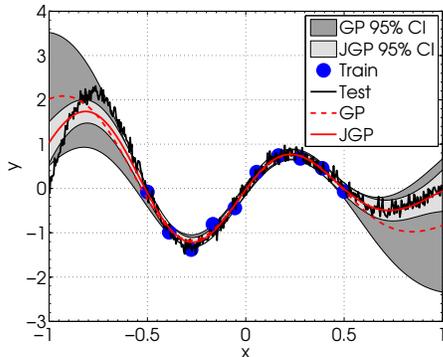


Fig. 1: Illustration of the predictive mean, μ_* , and variance, σ_* of GP and JGP in a toy example. Training data is confined to the region $[-0.5, +0.5]$ while the actual range of the input variable and the corresponding target variable is larger.

plications, real data come from *in situ* terrestrial campaigns, while a wealth of simulated data are available through runs of *radiative transfer models*. Figure 1 shows the above explained scenario, where a regular GP has been trained on the real, noisy data, and a JGP is trained on both the real and simulated data. Real data is typically available in a restricted interval, so the predictive mean of the ordinary GP fails to capture what happens outside this interval. The predictive variance increases whenever no observations are available, reflecting the added uncertainty. On the other hand, the JGP is more certain in its predictions, and follows the out-of-the-sample data better.

Unfortunately, radiative transfer models are much more complex than the one considered here, and not completely reliable either. Even so, it seems probable that the physical intuition encoded in the simulated training data would make the model more robust to generalization, especially in the regions where data is scarce. We see in §4 that this is indeed the case.

3. DATA COLLECTION

Non-destructive real LAI data were acquired in the framework of the local ERMES field activities within rice fields in Spain, Italy and Greece (see Fig. 2). The field campaigns were conducted during the 2014, 2015 and 2016 European rice seasons. During each season, the temporal frequency of the campaigns was approximately 10 day starting from the very beginning of rice emergence (early-June) up to the maximum rice green LAI development (mid-August). The same sampling scheme was used over each field following the guidelines and recommendations of the Validation of Land European Remote sensing Instruments (VALERI) protocol. LAI measurements were acquired using a dedicated smartphone app (PocketLAI) which uses both smartphone’s accelerometer and camera to acquire images at 57.5° below the canopy and computes LAI through an internal segmentation algorithm [10, 11]. The center of the ESU was geo-located for later matching and associate the mean LAI estimate with the corresponding satellite spectra. In this study we used Landsat

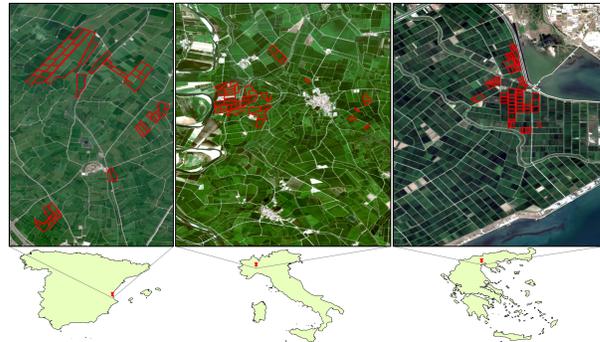


Fig. 2: Study areas and parcels where ground measurements were acquired.

8 surface reflectance data over each area corresponding to the dates of measurements’ acquisition.

On the other hand, the simulated data set (2000 pairs of Landsat 8 spectra and LAI) was obtained running the PROSAIL radiative transfer model in forward mode. PROSAIL simulates leaf reflectance for the optical spectrum from 400 to 2500 nm with a 1 nm spectral resolution and as a function of biochemistry and structure of the canopy, its leaves, the background soil reflectance and the sun-view geometry. The leaf and canopy variables as well as the soil brightness parameter, were randomly generated following a PROSAIL site-specific parameterization [12] in order to constrain the behavior of the model to Mediterranean rice areas (see Table 1).

Table 1: Distribution of the canopy, leaf and soil parameters within the PROSAIL RTM.

	Parameter	Min	Max	Mode	Std	Type
Canopy	LAI (m^2/m^2)	0	10	3.5	4.5	Gaussian
	ALA ($^\circ$)	30	80	60	20	Gaussian
	Hotspot	0.1	0.5	0.2	0.2	Gaussian
	vCover	0.5	1	1	0.2	Trunc. Gaussian
	N	1.2	2.2	1.5	0.3	Gaussian
Leaf	C_{ab} ($\mu\text{g}\cdot\text{cm}^{-2}$)	20	90	45	30	Gaussian
	C_{dm} ($\text{g}\cdot\text{cm}^{-2}$)	0.003	0.011	0.005	0.005	Gaussian
	C_{wREL}	0.6	0.8	-	-	Uniform
Soil	β_s	0.3	1.2	0.9	0.25	Gaussian

4. EXPERIMENTAL RESULTS

We assessed the effect of blending the real *in situ* measurements and the simulated data from PROSAIL for different amounts of data. The JGP model is compared to a regular GP for LAI-prediction using the real and simulated data detailed above. The gain in accuracy was measured as the reduction in root mean square error (RMSE gain [%] = $100 \times (\text{RMSE}_{\text{GP}} - \text{RMSE}_{\text{JGP}}) / \text{RMSE}_{\text{GP}}$). We evaluated the 9 datasets generated from the campaigns and simulations for different countries (SP, GR, IT) and years (2014, 2015, 2016). For each dataset, one half is used for training, while the other half is used for fitting the hyperparameters. Due to the scarcity of real data points for testing, the respective RMSEs were computed on

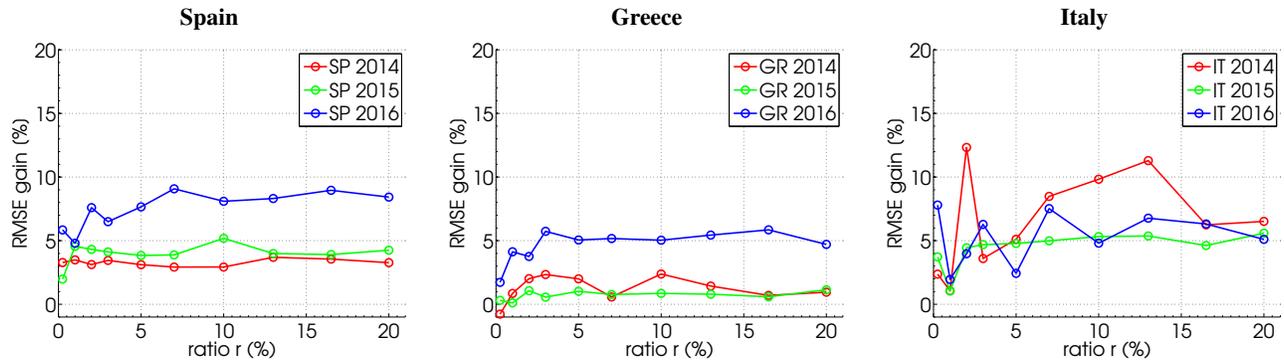


Fig. 3: Obtained accuracy gains in RMSE of JGP over GP for the different sites, campaign dates and simulated-to-real data ratios.

the same validation data used for hyperparameter tuning for both models.

Figure 3 shows the effect of the ratio between simulated and real data points $r = n_s/n_r$ on the RMSE gain. When no simulated data is used, the JGP model reduces to the canonical GP, but when introducing even a small amount of PROSAIL-data, i.e. $r > 0$, a noticeable gain is achieved. In the case of the data from the field campaigns in Spain and Greece, the gain appears rather stable (between 5-10%), indicating that little simulated data is needed for an increase in accuracy. The results obtained from the models trained on data from Italy, however, show no clear optimal level of simulated data. Seeing as the computational complexity is cubic in the number of data points, it seems the best strategy to include modest amounts of simulated data. However, as Fig. 3-Italy suggests a significant increase in accuracy (+10%) might occur as a result of including large amounts of simulated data to fill in the representation space in regions with scarce in situ sampling.

5. CONCLUSIONS

This paper presented a novel method based on Gaussian Process to perform inverse modeling. The model allows for the combination of *in situ* data and simulated data from an RTM to perform parameter retrieval. The formulation only incorporates one additional trade-off parameter. Important gains in accuracy are obtained in general for the estimation of LAI from Landsat and PROSAIL simulated data. The model exploits the space coverage of RTMs in regions where real data scarcity hampers performance, while at the same time incorporates the wealth of information provided by real data. Future work is tied to study the capabilities of the model for transportability across space and time.

6. REFERENCES

- [1] R. Snieder and J. Trampert, *Inverse Problems in Geophysics*, pp. 119–190, Springer Vienna, Vienna, 1999.
- [2] T. Hilker, N. C. Coops, M. A. Wulder, T. A. Black, and R. D. Guy, “The use of remote sensing in light use efficiency based models of gross primary production: A review of current status and future requirements,” *Science of the Total Environment*, vol. 404, no. 2-3, pp. 411–423, 2008.
- [3] R. H. Whittaker and P. L. Marks, “Methods of assessing terrestrial productivity,” *Primary Productivity of the Biosphere*, pp. 55–118, 1975.
- [4] H. K. Lichtenthaler, “Chlorophylls and carotenoids: Pigments of photosynthetic biomembranes,” *Methods Enzymol.*, vol. 148, pp. 350–382, 1987.
- [5] G. Camps-Valls, D. Tuia, L. Gmez-Chova, S. Jimnez, and J. Malo, Eds., *Remote Sensing Image Processing*, Morgan & Claypool Publishers, LaPorte, CO, USA, Sept 2011.
- [6] G. Camps-Valls and L. Bruzzone, *Kernel Methods for Remote Sensing Data Analysis*, John Wiley and Sons, 2009.
- [7] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, The MIT Press, New York, 2006.
- [8] J. Verrelst, L. Alonso, G. Camps-Valls, J. Delegido, and J. Moreno, “Retrieval of vegetation biophysical parameters using gaussian process techniques,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 5 PART 2, pp. 1832–1843, 2012, cited By 26.
- [9] G. Camps-Valls, J. Verrelst, J. Muoz-Mar, V. Laparra, F. Mateo-Jiménez, and J. Gomez-Dans, “A survey on gaussian processes for earth observation data analysis,” *IEEE Geoscience and Remote Sensing Magazine*, , no. 6, June 2016.
- [10] M. Campos-Taberner, F.J. Garcia-Haro, A. Moreno, M.A. Gilabert, S. Sanchez-Ruiz, B. Martinez, and G. Camps-Valls, “Mapping leaf area index with a smartphone and gaussian processes,” *Geoscience and Remote Sensing Letters, IEEE*, vol. 12, no. 12, pp. 2501–2505, Dec 2015.
- [11] M. Campos-Taberner, F.J. García-Haro, R. Confalonieri, B. Martínez, Á. Moreno, S. Sánchez-Ruiz, M.A. Gilabert, F. Camacho, M. Boschetti, and L. Busetto, “Multitemporal monitoring of plant area index in the valencia rice district with pocketlai,” *Remote Sensing*, vol. 8, no. 3, pp. 202, 2016.
- [12] M. Campos-Taberner, F.J. García-Haro, G. Camps-Valls, G. Grau-Muedra, F. Nutini, A. Crema, and M. Boschetti, “Multitemporal and multiresolution leaf area index retrieval for operational local rice crop monitoring,” *Remote Sensing of Environment*, vol. 187, pp. 102 – 118, 2016.