



HAL
open science

Joint Supervised Classification and Reconstruction of Irregularly Sampled Satellite Image Times Series

Alexandre Constantin, Mathieu Fauvel, Stéphane Girard

► **To cite this version:**

Alexandre Constantin, Mathieu Fauvel, Stéphane Girard. Joint Supervised Classification and Reconstruction of Irregularly Sampled Satellite Image Times Series. 2020. hal-02997573v2

HAL Id: hal-02997573

<https://inria.hal.science/hal-02997573v2>

Preprint submitted on 3 Mar 2021 (v2), last revised 21 Apr 2021 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Joint Supervised Classification and Reconstruction of Irregularly Sampled Satellite Image Time Series

Alexandre Constantin, *Student Member, IEEE*, Mathieu Fauvel, *Senior Member, IEEE*, and Stéphane Girard

Abstract—Recent satellite missions have led to a huge amount of earth observation data, most of them being freely available. In such a context, satellite image time series have been used to study land use and land cover information. However, optical time series, like Sentinel-2 or Landsat ones, are provided with an irregular time sampling for different spatial locations, and images may contain clouds and shadows. Thus, pre-processing techniques are usually required to properly classify such data. The proposed approach is able to deal with irregular temporal sampling and missing data directly in the classification process. It is based on Gaussian processes and allows to perform jointly the classification of the pixel labels as well as the reconstruction of the pixel time series. The method complexity scales linearly with the number of pixels, making it amenable in large scale scenarios. Experimental classification and reconstruction results show that the method does not compete yet with state of the art classifiers but yields reconstructions that are robust with respect to the presence of undetected clouds or shadows, and does not require any temporal preprocessing.

Index Terms—Satellite Image Time Series (SITS), Sentinel-2, classification, reconstruction, irregular sampling, Gaussian processes, Earth Observation (EO), remote sensing.

I. INTRODUCTION

IN the last decade, the successful launching of two satellites Sentinel-2 A and B offers an unique opportunity to record, analyze and predict the evolution of the Earth's land surface. Sentinel-2 mission provides a high resolution multispectral (13 spectral bands at 10m, 20m or 60m per pixel) acquisition with a 5 day revisit cycle [1]. It is planned for long-term operational observations (more than 15 years) and optical Sentinel-2 satellite image time-series (SITS) are available to users under a free and open data policy. Such a mission provides several terabytes of worldwide data per day [2].

Thanks to their spectral content and frequent update, SITS have found applications in various scientific field: water management [3], [4], agricultural systems mapping [5]–[7], urban area analysis [8], [9] or ecological monitoring [10], [11]. However, such abundance of images raises new challenges in terms of large scale multi-resolution SITS processing [12]. Issues related to *Big Earth Data* were obviously explored in the remote sensing community in order to calibrate and distribute the images seamlessly [13]–[17].

This work is supported by the French National Research Agency in the framework of the Investissements d'Avenir program (ANR-15-IDEX-02) and by the Centre National d'Etudes Spatiales (CNES).

A. Constantin and S. Girard are with Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, 38000 Grenoble, France (e-mail: alexandre.constantin@inria.fr; stephane.girard@inria.fr).

M. Fauvel is with CESBIO, Université de Toulouse, CNES/CNRS/INRAe/IRD/UPS, 31000 Toulouse, France (e-mail: mathieu.fauvel@inrae.fr).

Image and signal processing issues were also investigated specifically to SITS properties. Spatial differences in terms of resolution were tackled using super resolution strategies to resample all the spectral bands at the same finest spatial resolution (10m) [18]–[20]. Fusion with other sensors, such as Landsat [21] or Sentinel-1 [22], was also considered to complement Sentinel-2 satellite images.

Classification of land cover or land use using machine learning techniques has emerged as a major application of SITS. Deep learning was intensively studied either using only one or few temporal acquisitions with convolutional neural networks (CNN) [23], [24], or using the full temporal stack of acquisitions using recurrent neural networks (RNN) [25], [26]. Combination of RNN applied on SITS with CNN applied on SPOT-6 image was also investigated for land cover mapping over Reunion Island [27]. Temporal convolution was investigated in [28]. Yet, due to the huge amount of data to be processed when large geographical area or large temporal domain (e.g. 1 year of acquisition) are considered, lighter machine learning techniques, such as Random Forest (RF), were proposed as well, with very good results in terms of classification accuracy [29], [30]. Similarly, Dersken *et al.* have shown that, when the number of spectro-temporal measurements per pixel is high, conventional hand-crafted features and RF perform as well as CNN in terms of classification accuracy but with a drastically reduced processing time [31].

One major issue arising when dealing with large SITS, *i.e.* using several tracks or tiles¹, is related to the irregular temporal sampling. Indeed, each Sentinel-2 track is acquired on different dates, and using images from different tracks results in pixels with different acquisition times. For instance, Figure 1 shows the acquisition dates for three Sentinel-2 tiles located over three parts of France. Even though some dates are similar, the total number of acquisitions is different from one tile to another. Furthermore, clouds and shadows, occurring at random, are another cause of irregular temporal sampling: Dates with shadow/cloud are usually considered as missing values and are not taken into account to build the feature vector of the corresponding pixels. Hence, even pixels belonging to the same tile can be represented by feature vectors of different sizes [30].

Conventional machine learning algorithms take as input a time-series of finite and constant size. Random Forest, Support Vector Machine or RNN use a vector representation of input features, *i.e.*, a time-series, or pixel, is a collection of features

¹Sentinel-2 products are available as a collection of elementary tiles of size 100×100 km, see <https://sentinel.esa.int/web/sentinel/missions/sentinel-2/data-products>.

stacked into a p -dimensional vector and each pixel has the same number of features. Similarly, CNN use a constant size patch representation of the image. Therefore, a pre-processing step, sometimes called *missing information reconstruction*, is usually applied on the SITS to recover pixels of same size [32]. Among the various existing techniques, temporal filters, such as the Savitzky-Golay filter [33] or the best index slope extraction (BISE) [34], and parametric or non parametric curve fitting [35]–[37] are popular due to their simplicity and efficiency. At large scale (entire metropolitan France territory), linear temporal interpolation has shown to perform very well compared to spline interpolation [30]. Linear interpolation may also be combined with non parametric smoothers [38].

Methods taking account both spatial and temporal information have provided interesting results, see for instance non-local filters [39] or deep learning based approaches [40]–[42]. However, due to their high computational cost, they are hardly applicable in national/continental scale settings [31]. Furthermore, they are also known as “black-box” and hardly interpretable.

Recently, Gaussian Processes (GP) have gained attention in the remote sensing community [43]. GP mixed Bayesian and kernel methods to build statistical learning machines, for classification or regression problems. They have proved to be accurate and interpretable through their hyperparameters (mean and covariance functions) [44]–[47]. In this work, a Gaussian Process model is introduced, relying on a linear projection of the mean function on a well-chosen basis, conditionally to the classes of interest. It is able to deal with irregularly sampled pixels in the learning and classification steps as well as reconstructing missing data.

The contribution of this work is three-fold:

- 1) The definition of a GP model that handles irregular temporal sampling (Section II), without resampling all the data on a common temporal grid.
- 2) Jointly to the classification, the proposed model allows for the reconstruction of pixels on any temporal grid, together with confidence intervals (Section III). Furthermore, the mean and covariance functions associated with each class are available for analysis.
- 3) The reconstruction used the GP model with optimal parameters (from the likelihood point of view) and takes into account the class membership. Rather than relying only on the reconstruction error, as with conventional missing data reconstruction techniques, the proposed method used the classification model likelihood to control the reconstruction and its subsequent smoothing level.

The remainder of this paper is organized as follows. The proposed GP model and its estimation are described in Section II. The associated classification and reconstruction methods are derived in Section III. Section IV details the SITS datasets while the experimental set-up is presented in Section V. Experimental classification and reconstruction results are provided in Sections VI and VII respectively. Finally, Section VIII concludes this paper and opens discussion on future work.

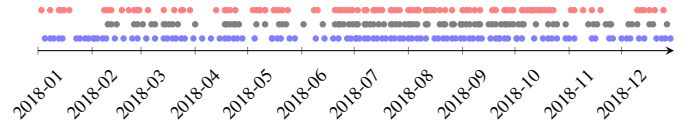


Fig. 1. Acquisition dates for tile T31TDN (red), T31TCJ (black) and T31TGK (blue) in 2018. The location of the tiles is given in Fig. 5. Large temporal gaps between two dates correspond to very cloudy acquisitions: Images tagged as composed of more than 90% of cloudy pixels are not processed by the French data provider.

TABLE I
SYMBOLS AND NOTATIONS

Symbol	Meaning
\sim	Distributed according to (a probability distribution).
\mathcal{GP}	Gaussian Process: $f \sim \mathcal{GP}(m, K)$, the function f is distributed as a GP with mean function m and covariance operator K . The function is indexed in time by t .
$\mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	d -variate Gaussian (Normal) distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$.
$f_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}^d$	Density associated with $\mathcal{N}_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.
$y x$	Random variable y conditionally on x ,
$p(y x)$	Probability distribution of $y x$.
\mathcal{S}	A set of SITS or pixels.
n	The total number of SITS available in \mathcal{S} .
p	Number of wavelenghts of a SITS.
b	Band number, in $\{1, \dots, p\}$.
\mathcal{T}	Temporal window where SITS are observed.
Y	p - dimensional process: $\mathcal{T} \rightarrow \mathbb{R}^p$.
$Y_{i,b}$	i th pixel from wavelenght b .
T_i	Number of temporal acquisitions associated with pixel i .
\mathcal{C}	Number of classes.
c	Class value, in $\{1, \dots, \mathcal{C}\}$.
z_i	Discrete random variable associated with pixel i and representing its class membership.
π_c	Prior probability of class c .
$\boldsymbol{\alpha}, \boldsymbol{\theta}$	Model parameters.
J	Number of basis functions <i>i.e.</i> dimension of $\boldsymbol{\alpha}$.
$ K $	Determinant of the matrix K .

II. IRREGULARLY SAMPLED GAUSSIAN PROCESSES MODEL

In the following $\mathcal{S} = \{(y_i, z_i)\}_{i=1}^n$ denotes a set of n independent and identically distributed (i.i.d.) random multivariate and irregularly sampled pixels from a SITS, with their associated class labels, see Table I for an overview of the main notations. A pixel y is modeled by a random vector Y containing p random square integrable processes $\mathcal{T} \rightarrow \mathbb{R}^p$, where $\mathcal{T} = [t_{\min}, t_{\max}]$ is the time interval where the SITS are observed:

$$Y : \mathbb{R} \rightarrow \mathbb{R}^p$$

$$t \mapsto [Y_1(t), \dots, Y_b(t), \dots, Y_p(t)]^\top.$$

This property is denoted by $Y \in L_2^p(\mathcal{T})$. The associated class z is modeled by a discrete random variable Z with possible values in $\{1, \dots, \mathcal{C}\}$. In the context of this work, p is the number of wavelenghts and/or spectral indices, $t_{\min} = 0$ and $t_{\max} = 365$ (in day-of-year unit).

A (univariate) Gaussian process (GP) f is a stochastic process such that any finite-dimensional marginal follows a multivariate Gaussian distribution [48]. It is specified by its

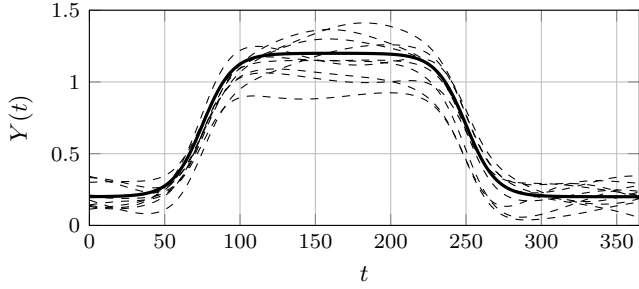


Fig. 2. Simulated univariate Gaussian processes with squared exponential covariance function (length scale parameter set to 50). The continuous black line is the mean function and the dashed lines are the corresponding 10 realizations.

mean function m and covariance function K :

$$\begin{aligned} m(t_1) &= \mathbb{E}[f(t_1)], \\ K(t_1, t_2) &= \mathbb{E}[(f(t_1) - m(t_1))(f(t_2) - m(t_2))], \end{aligned}$$

with $t_1 \in \mathcal{T}$, $t_2 \in \mathcal{T}$ two given times where the SITS are observed, and we note $f \sim \mathcal{GP}(m, K)$. The covariance function K usually encodes *a priori* knowledge about the phenomena to be modeled. Figure 2 shows several realizations of a GP with a squared exponential covariance function (see Section II-B2).

In the following, a multivariate Gaussian process model is proposed to cope with the multivariate nature of SITS and their irregular temporal sampling.

A. Mixture of Independent Multivariate Gaussian Processes Model

Definition 1 (Mixture of Independent Multivariate Gaussian Processes): The proposed model, namely Mixture of Independent Multivariate Gaussian Process (MIMGP), relies on two main assumptions:

- A1 Each process Y_b , conditionally to $Z = c$, follows a Gaussian Process: $Y_b|Z = c \sim \mathcal{GP}(m_{b,c}, K_{b,c})$;
- A2 Conditionally to $Z = c$, all components Y_b of Y are independent,

where $b \in \{1, \dots, p\}$ is the spectral band and $c \in \{1, \dots, \mathcal{C}\}$ is the class. The first property states that each band of a pixel from a given class is a realization of a univariate GP with class and band specific mean and covariance functions. The second property is introduced for computational reasons, and is discussed in Section VI-B.

Let \mathbf{y}_i be an observed pixel i at times $\{t_1^i, \dots, t_{T_i}^i\}$. Its b th coordinate is represented by a vector of dimension T_i :

$$\mathbf{y}_{i,b} = [Y_b^i(t_1^i), \dots, Y_b^i(t_{T_i}^i)]^\top.$$

By definition, conditionally to $Z = c$, this vector follows a T_i -variate Gaussian distribution

$$\mathbf{y}_{i,b}|Z_i = c \sim \mathcal{N}_{T_i}(\boldsymbol{\mu}_{i,b,c}, \boldsymbol{\Sigma}_{b,c}^i), \quad (1)$$

where \mathcal{N}_{T_i} is the Gaussian distribution on \mathbb{R}^{T_i} with vector mean $\boldsymbol{\mu}_{i,b,c} = [m_{b,c}(t_1^i), \dots, m_{b,c}(t_{T_i}^i)]^\top$ and covariance matrix $(\boldsymbol{\Sigma}_{b,c}^i)_{\ell, \ell'} = K_{b,c}(t_\ell^i, t_{\ell'}^i)$.

Finally, in view of the independence assumption on the p wavelengths, \mathbf{y}_i follows a product of p Gaussian densities

$$\mathbf{y}_i|Z_i = c \sim \prod_{b=1}^p \mathcal{N}_{T_i}(\boldsymbol{\mu}_{i,b,c}, \boldsymbol{\Sigma}_{b,c}^i). \quad (2)$$

B. Mean and covariance functions

The unknown parameters of the model are the mean functions $m_{b,c}$ and the covariance operators $K_{b,c}$, for each band $b \in \{1, \dots, p\}$ and each class $c \in \{1, \dots, \mathcal{C}\}$. To deal with irregularly sampled observations, these parameters should be defined (and estimated) for any time t in the interval \mathcal{T} . To this end, projection techniques and parametric models are adopted respectively for the mean and the covariance functions. Parametric modeling of the latter ones is usual with GP, see for instance [48, Chapter 4]. However, in our specific problem, using a projection method for the mean function allows to cope nicely with the irregular temporal sampling.

1) *Mean function:* Let $\{\varphi_j\}_{j=1}^J$ be a subset of basis functions of $L_2(\mathcal{T})$. Then, $m_{b,c}$ can be written as

$$m_{b,c}(t) = \sum_{j=1}^J \alpha_{b,c,j} \varphi_j(t) + \varepsilon_{b,c}(t), \quad t \in \mathcal{T},$$

with $\alpha_{b,c,j}$ the projection coefficient of $m_{b,c}$ on φ_j and where $\varepsilon_{b,c}$ is an approximation error term.

Any functional basis can be used, *e.g.*, Fourier, exponential, splines, *...*, see for instance [49, Chapter 5]. In case of the Fourier basis, $\alpha_{b,c,j}$ represents the amplitude of the corresponding frequency in the Fourier expansion. In the following, the basis is assumed to be fixed while J , the number of basis functions, is an hyperparameter.

Introducing $\boldsymbol{\alpha}_{b,c} = [\alpha_{b,c,1}, \dots, \alpha_{b,c,J}]^\top \in \mathbb{R}^J$ and \mathbf{B}^i the $T_i \times J$ design matrix associated with pixel \mathbf{y}_i defined by $(\mathbf{B}^i)_{\ell,j} = \varphi_j(t_\ell^i)$ with $\ell \in \{1, \dots, T_i\}$, the mean vector in (1) can be written as

$$\boldsymbol{\mu}_{i,b,c} = \mathbf{B}^i \boldsymbol{\alpha}_{b,c}. \quad (3)$$

2) *Covariance function:* The covariance function is modeled using conventional functions issued from the GP literature [48, Chapter 4]. A typical one would be a squared exponential covariance function with an additional colored noise [50] covariance function:

$$K_{b,c}(t, s) = \gamma_{b,c}^2 \exp\left\{-\frac{(t-s)^2}{2h_{b,c}^2}\right\} + \sigma_{b,c}^2(t)\delta_{t,s}. \quad (4)$$

Introducing $\boldsymbol{\theta}_{b,c} = \{\gamma_{b,c}^2, h_{b,c}^2, \sigma_{b,c}^2\}$, the covariance matrix in (1) is denoted in the following by

$$\boldsymbol{\Sigma}_{b,c}^i = \boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c}). \quad (5)$$

Parameters $\boldsymbol{\alpha}_{b,c}$ and $\boldsymbol{\theta}_{b,c}$ are estimated by maximizing the marginal log-likelihood, as explained in the next section.

C. Estimation

By plugging (3) and (5) in (2), it follows that

$$\mathbf{y}_i | Z_i = c \sim \prod_{b=1}^p \mathcal{N}_{T_i}(\mathbf{B}^i \boldsymbol{\alpha}_{b,c}, \boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})). \quad (6)$$

Such expression is a consequence of the use of non-zero mean GPs [48, Section 2.7]. The associated negative marginal log-likelihood is given by

$$\ell(\boldsymbol{\alpha}, \boldsymbol{\theta}) = \sum_{b,c=1}^{p,C} \ell_{b,c}(\boldsymbol{\alpha}_{b,c}, \boldsymbol{\theta}_{b,c}),$$

with

$$\begin{aligned} \ell_{b,c}(\boldsymbol{\alpha}_{b,c}, \boldsymbol{\theta}_{b,c}) = & \sum_{i|Z_i=c} \left[\log |\boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})| \right. \\ & \left. + (\mathbf{y}_{i,b} - \mathbf{B}^i \boldsymbol{\alpha}_{b,c})^\top \boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})^{-1} (\mathbf{y}_{i,b} - \mathbf{B}^i \boldsymbol{\alpha}_{b,c}) \right] + \kappa, \end{aligned}$$

where κ is a constant independent of the parameters. Since independence is assumed between each spectral component, the optimization ends up with $p \times C$ independent optimization problems:

$$(\hat{\boldsymbol{\alpha}}_{b,c}, \hat{\boldsymbol{\theta}}_{b,c}) = \arg \min_{\boldsymbol{\alpha}_{b,c}, \boldsymbol{\theta}_{b,c}} \ell_{b,c}(\boldsymbol{\alpha}_{b,c}, \boldsymbol{\theta}_{b,c}). \quad (7)$$

Marginal likelihood optimization is common practice in GP regression, while for binary classification, Laplace approximation is usually employed (and more computationally demanding) [48, Chapter 2 and 3]. Therefore, to the best of our knowledge, the use of marginal likelihood for multiclass classification is novel in this context.

Each sub-problem (7) is solved by gradient descent as conventionally done in likelihood optimization [48, Chapter 5]. The algorithm is based on an alternate optimization with respect to $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$ (see Algorithm 1). At iteration (k), the update rule for $\boldsymbol{\alpha}$ is given by

$$\boldsymbol{\alpha}_{b,c}^{(k+1)} = \left[\sum_{i|Z_i=c} \mathbf{B}^{i\top} \boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c}^{(k)})^{-1} \mathbf{B}^i \right]^{-1} \left[\sum_{i|Z_i=c} \mathbf{B}^{i\top} \boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c}^{(k)})^{-1} \mathbf{y}_{i,b} \right]. \quad (8)$$

If the design matrix has no redundancies, it can be shown that the matrix in the left hand side of (8) is indeed non singular provided that the number of basis functions J is smaller or equal to the total number of unique observations in the training set. A proof is given in Section I of the supplementary materials.

For $\boldsymbol{\theta}$, there is no close-form expression and a gradient step is required. The gradient is computed using the following partial derivative w.r.t. each coordinate θ_m of $\boldsymbol{\theta}$:

$$\frac{\partial}{\partial \theta_m} \ell_{b,c} = \sum_{i|Z_i=c} \text{tr} \left(\left(\boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c}^{(k)})^{-1} - \boldsymbol{\beta}_i \boldsymbol{\beta}_i^\top \right) \frac{\partial \boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c}^{(k)})}{\partial \theta_m} \right) \quad (9)$$

Algorithm 1: Estimation of model parameters.

Input : $\mathcal{S}, \boldsymbol{\alpha}^0, \boldsymbol{\theta}^0$
Output: $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\theta}}$

- 1 **for** $c=1$ **to** C **do**
- 2 **for** $b=1$ **to** p **do**
- 3 **repeat**
- 4 Update $\boldsymbol{\alpha}_{b,c}$ using (8);
- 5 Do a gradient step w.r.t. $\boldsymbol{\theta}_{b,c}$ using (9);
- 6 **until** $\ell_{b,c}(\boldsymbol{\alpha}_{b,c}, \boldsymbol{\theta}_{b,c})$ has converged;

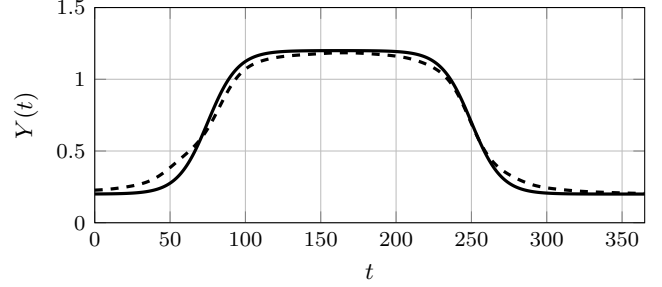


Fig. 3. The continuous line is the mean function associated with the GP displayed in Figure 2 and the dashed line is the estimated one using Algorithm 1 and $J = 20$ Gaussian functions (see Section V-A for details on bases).

with $\boldsymbol{\beta}_i = \boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c}^{(k)})^{-1} (\mathbf{y}_i - \mathbf{B}^i \boldsymbol{\alpha}_{b,c}^{(k+1)})$. Such an optimization procedure applied on simulated data from Figure 2 (*i.e.*, $C = 1$ and $p = 1$) leads to the estimation of the mean function displayed on Figure 3.

As a final remark, the $p \times C$ optimization problems can be solved in parallel since parameters are not shared between classes and spectral components.

D. Numerical Complexity

Conventional GPs have complexity that scales in $\mathcal{O}(n^3)$, making them unsuitable for very large scale problems. In contrast, the proposed method has a reduced complexity, that scales in $\mathcal{O}(n(T_M^3 + J^3))$ where T_M is the maximal length of observed time-series ($T_M \geq T_i, \forall i \in 1, \dots, n$). The first term comes from the inversion of $\boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})$ and the second term comes from the update rule (8). In our application, T_M and J are much smaller than n , typically by several orders of magnitude. Furthermore, MIMGP allows for the classification of any irregularly new time-series as well as the reconstruction of the observed time-series using reconstruction techniques, as described in the following.

III. CLASSIFICATION AND RECONSTRUCTION OF MISSING VALUES

Once the parameters are estimated, it is possible to classify a new pixel without any temporal resampling, as for training. In addition, MIMGP is also able to reconstruct missing values on any temporal scheme.

In the following, \mathbf{y}_j denotes a new SITS observed at T_j times denoted by $\{t_1, \dots, t_{T_j}\}$ which may not have been observed in the training set.

A. Classification of a new time-series

The *a posteriori* probability $\mathbb{P}(Z_j = c | \mathbf{y}_j)$ to belong to a class c given \mathbf{y}_j is computed using the product of Gaussian densities given in (6) and Bayes' rule:

$$\begin{aligned} \mathbb{P}(Z_j = c | \mathbf{y}_j) &\propto \pi_c \mathbb{P}(\mathbf{y}_j | Z_j = c) \\ &\propto \pi_c \prod_{b=1}^p f_{T_j}(\mathbf{y}_{j,b}; \mathbf{B}^j \boldsymbol{\alpha}_{b,c}, \boldsymbol{\Sigma}(\boldsymbol{\theta}_{b,c})), \end{aligned} \quad (10)$$

where $f_d(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ is the d -variate Gaussian density with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ and π_c is the *prior* probability $\mathbb{P}(Z = c)$ of the class $c \in \{1, \dots, C\}$.

In practice, π_c is estimated by its empirical counterpart $\hat{\pi}_c = n_c/n$ where n_c is the number of pixels assigned to class c in the set \mathcal{S} . Parameters $\hat{\boldsymbol{\alpha}}_{b,c}$ and $\hat{\boldsymbol{\theta}}_{b,c}$ are estimated thanks to Algorithm 1. The new time-series \mathbf{y}_j is then assigned to the class of maximum posterior probability (MAP rule):

$$\hat{z}_j = \max_c \mathbb{P}(Z_j = c | \mathbf{y}_j).$$

B. Time-series reconstruction

Time-series reconstruction is achieved using conditional expectations of Gaussian distributions. Two cases are considered: Either the class membership of the considered pixel is known, or the class membership is estimated using the posterior probability.

1) *Reconstruction when class membership is known:* Let us write $Y_b^j(t^*)$ the unobserved value at wavelength b and time t^* . The following reconstruction rule, based on the conditional expectation, is considered:

$$\hat{Y}_{b,c}^j(t^*) := \mathbb{E}[Y_b^j(t^*) | \mathbf{y}_j, Z_j = c].$$

Using properties of Gaussian distributions [51, p.63] and replacing the unknown quantities by their estimated counterparts yield (see Appendix A for a proof):

$$\begin{aligned} \hat{Y}_{b,c}^j(t^*) &= \mathbf{b}^* \hat{\boldsymbol{\alpha}}_{b,c} \\ &+ \mathbf{k}(t^*, t_{1:T_j}^j | \hat{\boldsymbol{\theta}}_{b,c})^\top \boldsymbol{\Sigma}^j(\hat{\boldsymbol{\theta}}_{b,c})^{-1} (\mathbf{y}_{j,b} - \mathbf{B}^j \hat{\boldsymbol{\alpha}}_{b,c}) \end{aligned} \quad (11)$$

and

$$\begin{aligned} \mathbb{V}[\hat{Y}_{b,c}^j(t^*)] &= K(t^*, t^* | \hat{\boldsymbol{\theta}}_{b,c}) \\ &- \mathbf{k}(t^*, t_{1:T_j}^j | \hat{\boldsymbol{\theta}}_{b,c})^\top \boldsymbol{\Sigma}^j(\hat{\boldsymbol{\theta}}_{b,c})^{-1} \mathbf{k}(t^*, t_{1:T_j}^j | \hat{\boldsymbol{\theta}}_{b,c}), \end{aligned} \quad (12)$$

where $\mathbf{k}(t^*, t_{1:T_j} | \hat{\boldsymbol{\theta}}_{b,c}) = [K(t^*, t_1 | \hat{\boldsymbol{\theta}}_{b,c}), \dots, K(t^*, t_{T_j} | \hat{\boldsymbol{\theta}}_{b,c})]^\top$, $\mathbf{b}^* = [\varphi_1(t^*), \dots, \varphi_J(t^*)]$ contains the evaluation of the basis functions at time t^* and \mathbf{B}^j is the $T_j \times J$ design matrix associated with \mathbf{y}_j .

Equations (11) and (12) provide respectively the reconstructed value and the variance of the reconstruction. They are given here for a single time t^* but similar formulas can be derived for multiple times $t_1^*, \dots, t_{T^*}^*$. Interestingly, Equation (11) shows that the reconstruction is given by $\mathbf{b}^* \hat{\boldsymbol{\alpha}}_{b,c}$, the estimated mean of the GP, corrected by a value proportional to the error made at the acquisition time of pixel j . A similar remark holds for the variance: It can be interpreted as the estimated variance of the GP corrected by the variance of the process observed at the acquisition times of pixel j .

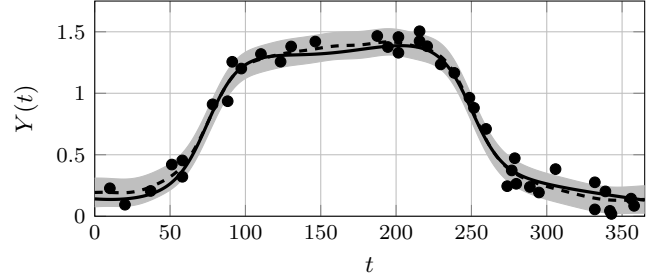


Fig. 4. Time-series reconstruction. The black continuous line is one (continuous) realization of the GP from Figure 2. The black dots are the observed noisy acquisitions at several (discrete) times. The dashed line is the reconstructed time-series. The gray region displays a pointwise confidence region for the reconstruction: $\hat{Y}_b^j(t^*) \pm \sqrt{\mathbb{V}[\hat{Y}_b^j(t^*)]}$.

2) *Reconstruction when class membership is unknown:* The reconstruction of $\hat{Y}_b^j(t^*)$ is done as the average of the previous reconstructions $\hat{Y}_{b,c}^j(t^*)$ in each class (see (11)) weighted by the posterior probabilities estimated with (10):

$$\hat{Y}_b^j(t^*) = \sum_{c=1}^c \mathbb{P}(Z_j = c | \mathbf{y}_j) \hat{Y}_{b,c}^j(t^*). \quad (13)$$

A similar formula holds for the variance with an additional between-classes variance term:

$$\begin{aligned} \mathbb{V}[\hat{Y}_b^j(t^*)] &= \sum_{c=1}^c \mathbb{P}(Z_j = c | \mathbf{y}_j) \mathbb{V}[\hat{Y}_{b,c}^j(t^*)] \\ &+ \sum_{c=1}^c \mathbb{P}(Z_j = c | \mathbf{y}_j) \left[\hat{Y}_{b,c}^j(t^*)^2 - \hat{Y}_b^j(t^*)^2 \right], \end{aligned}$$

where $\hat{Y}_{b,c}^j(t^*)$ and $\hat{Y}_b^j(t^*)$ are the reconstructions at t^* respectively when the class is known (11) and when the class is unknown (13). See Appendix A for a proof.

IV. SENTINEL-2 SATELLITE IMAGE TIME-SERIES DATASETS

Three Sentinel-2 tiles of level 2A over the French metropolitan territory were downloaded from the Theia Land Data Center². All available acquisitions between January 2018 and December 2018 for the two orbits of satellites Sentinel-2A and 2B were used. Figure 5 shows the location of the three tiles. They correspond to different climatic regions [30], with varying meteorological and topographical conditions.

Surface reflectance time-series were produced using the MAJA (Multi-sensor atmospheric correction and cloud screening-ATCOR Joint Algorithm) processing chain developed by the CNES-CESBIO and DLR [52]. It involves orthorectification, atmospheric correction, clouds and shadows detection. Spectral bands at 10m/pixel and 20m/pixel were used, for a total of 10 spectral bands. Bands at 20m/pixel were spatially up-sampled to 10m/pixel using the Orfeo Tool-Box³ [53]. Figure 6 displays the distribution of the number of clear dates (dates not tagged clouds, shadows or no-data

²<http://www.theia-land.fr/en/presentation/products>.

³Using "SuperImpose" application, see https://www.orefo-toolbox.org/CookBook/Applications/app_Superimpose.html.

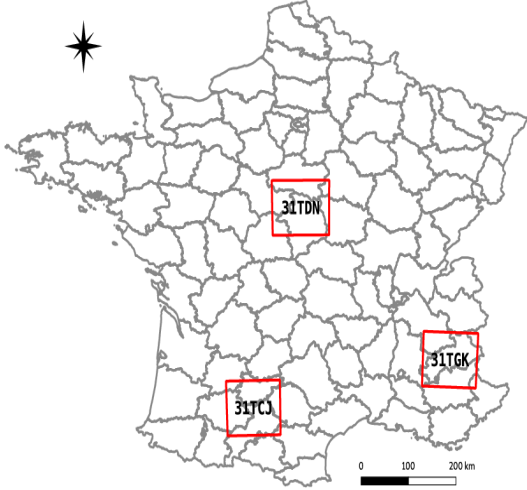


Fig. 5. Sentinel-2 tiles used. The label inside the square is the tile name.

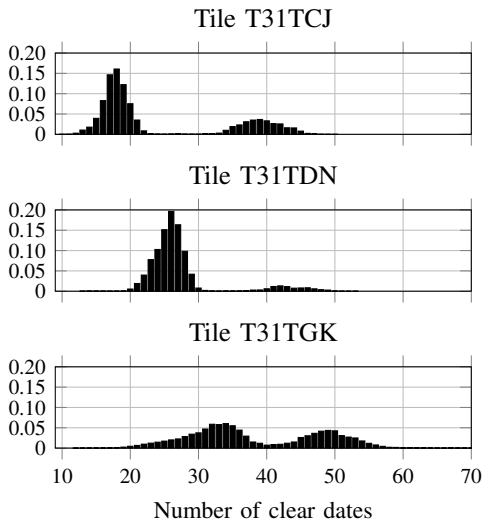


Fig. 6. Proportion of pixels as functions of the number of clear dates.

in the raw time-series): It appears that the number of clear acquisitions per pixel significantly varies depending on the tile.

Furthermore, the data from each tile were re-sampled and gap-filled (missing information due to clouds or shadows were reconstructed using linear interpolation) onto the same set of dates (every 5 days, starting from 2018-01-01 and ending 2018-12-27) as in [30]. Figure 11 shows the raw data and the re-sampled data for one random pixel from each tile in the same class. Hence, for each pixel location, three temporal informations are available after the pre-processing:

- 1) The raw multispectral time-series with irregular acquisition dates;
- 2) The mask time-series, indicating for each acquisition date the presence/absence of clouds/shadows. In our model, the presence of clouds/shadows implies that the corresponding raw spectral values are considered as missing values;

- 3) The re-sampled multispectral time-series with regular acquisition dates.

The reference data arise from the work of Inglada *et al.* [30]. They were extracted from freely available data source. Seventeen land cover classes were defined, ranging from artificial areas to vegetation and water bodies. Table II shows the exhaustive list of classes.

TABLE II
LAND COVER CLASSES AND NUMBER OF EXTRACTED SITS IN EACH TILE

Class	T31TCJ	T31TDN	T31TGK
Artificial areas			
Continuous urban fabric	10,000	8,292	959
Discontinuous urban fabric	10,000	10,000	10,000
Industrial or commercial units	10,000	10,000	10,000
Road surfaces	10,000	9,906	3,664
Agricultural areas			
Winter crops	30,000	30,000	15,975
Summer crops	40,000	40,000	24,912
Forest and semi-natural areas			
Meadow	10,000	10,000	10,000
Orchards	10,000	2,775	10,000
Vines	10,000	8,719	153
Broad-leaved forest	10,000	10,000	10,000
Coniferous forest	9,957	10,000	10,000
Natural grasslands	9,939	3,022	10,000
Woody moorlands	9,972	10,000	10,000
Open spaces with little or no vegetation			
Bare rock	0	0	10,000
Beaches, dunes and sand plains	0	5,355	10,000
Glaciers and perpetual snow	0	0	10,000
Water bodies	10,000	10,000	10,000
Total	189,868	178,069	165,663

The reference data are provided as a set of spatial polygons overlapping the 3 tiles. Figure 7 shows an extract of these polygons. The training and validation set were constructed by stratifying pixels according to the polygons membership information: Pixels from one polygon fully belong to either the training or the validation set. Depending on the number of available referenced pixels per class, 10,000 (or less) pixels were extracted for the training and validation set, except for winter and summer crops, for which 30,000 and 40,000 pixels were extracted, respectively. Ten independent train/validation sets were generated for statistical validation. Table II shows the average number of pixels per class and per tile⁴.

V. EXPERIMENTAL SET-UP

The model parameters are $\alpha_{b,c}$ and $\theta_{b,c}$ for each class c and each wavelength b . They are estimated using the training set of pixels. MIMGP has also some hyperparameters that are set before the learning step: The design matrix \mathbf{B} with the associated number of basis functions J and the family of parametric kernels for the covariance operator. These settings are common to each class and each wavelength.

⁴Since a stratification w.r.t polygons is done, and polygons size can vary in terms of area, the number of pixels may vary slightly from one experiment to another. The model was trained on a total average of 178,000 pixels and was evaluated on 178,000 pixels.

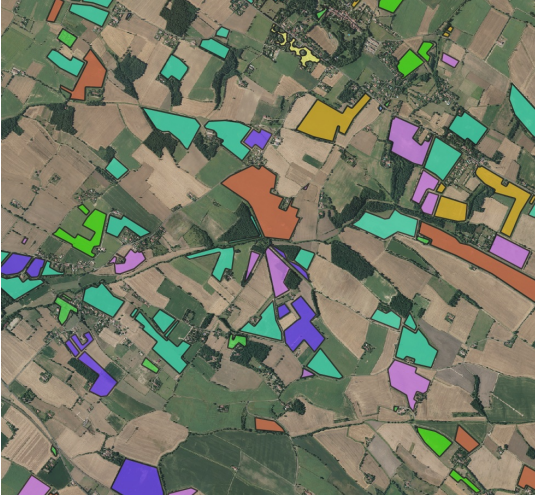


Fig. 7. Extract of the reference data. Each color represents a land cover class. The background image is an aerial orthophoto from the National Institute of Geographic and Forest Information (IGN) (<https://geoservices.ign.fr/documentation/diffusion/telechargement-donnees-libres.html>).

A. Functional bases

Four functional bases have been investigated: Two of them are local bases (non-zero only on a subinterval of $[0, \mathcal{T}]$), the *Gaussian* and *BSplines* bases, while the remaining ones are global bases, the well-known *Fourier* and *Polynomial* bases. An user-defined hyperparameter, denoted by J^* , is used to select the number of basis functions, as explained in the next paragraph. The choice of a basis setting corresponds to an assumption on the temporal behavior of the time-series: For instance, a *Fourier* basis can represent a periodic signal in the time domain.

- *Fourier*: The number of basis functions is $J = 2J^* + 1$ and

$$m(t) = \alpha_0 + \sum_{j=1}^{J^*} \left[\alpha_j \cos(2\pi j \frac{t}{\mathcal{T}}) + \alpha_{j+J^*} \sin(2\pi(j+J^*) \frac{t}{\mathcal{T}}) \right].$$

- *Polynomial*: $m(t) = \alpha_0 + \sum_{j=1}^{J^*} \alpha_j t^j$, with $J = J^* + 1$.
- *Gaussian*: $m(t) = \sum_{j=1}^{J^*} \alpha_j \exp\left(-\frac{(t-t_j)^2}{d_j^2}\right)$, with $J = J^*$.

The $t_j, j \in \{1, \dots, J^*\}$ can be equidistant in \mathcal{T} or chosen as quantiles of the distribution of the clean dates. The hyperparameter d_j is set such that $d_j^2 = 8|t_{j+1} - t_j|$ to ensure a sufficient overlap between two consecutive local exponential functions.

- *Cubic splines*: $m(t) = \sum_{j=0}^{J^*} \alpha_j S_j(t)$, with $J = J^* + 1$. S_j is the j^{th} bicubic spline on $[0, \mathcal{T}]$. The knots t_j are chosen as in the Gaussian case.

The invertibility condition associated with (8) implies that J should be smaller than 92, 100 and 103 for tiles T31TCJ, T31TDN and T31TGK, respectively.

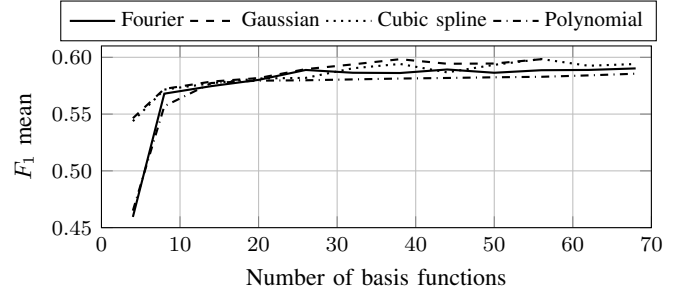


Fig. 8. Averaged “mean F_1 scores” of the MIMGP model with the 4 bases as a function of the basis dimension. The results are reported for the tile T31TCJ.

B. Covariance function

Any convex combination of positive semi-definite kernels is a valid covariance function [48, Chapter 4]. In this work, the squared exponential covariance function added with a colored noise covariance function, as given in (4), is used.

The parameters $\theta_{b,c} = \{\gamma_{b,c}^2, h_{b,c}^2, \sigma_{b,c}^2\}$, $\forall \{b, c\} \in [1, p] \times [1, C]$ are learned as described in Algorithm 1. The estimated covariance parameters provide some insights about the observed processes. In particular, the length-scale $h_{b,c}$ is related to the temporal behavior of the reflectance. For a given band b and conditionally to class c , the longer $h_{b,c}$ grows, the more correlated two distant dates are.

VI. SUPERVISED CLASSIFICATION

The classification accuracy is assessed by the F_1 score, computed as the harmonic mean of the precision and recall for each class. We report the “mean F_1 score”, which is the mean of F_1 scores computed on all the classes of the dataset.

First, the influence of the basis functions on the classification score is investigated. Then, comparison with other classifiers is reported and discussed. Convergence and model parameter analyses are provided in Section II of the supplementary materials.

A. Influence of the basis functions

The influence of the basis functions and its dimension are investigated. Figure 8 represents the averaged “mean F_1 scores” obtained on the 10 independent train/validation data sets for the tile T31TCJ. From the figure, it appears that the dimension of the basis should be larger than 9 for this tile to allow for a reasonable classification score, independently of the basis itself. Furthermore, MIMGP is robust w.r.t. basis selection since similar accuracy are obtained for all bases.

However, numerical instability has been observed for very large number of basis functions, in particular for the exponential basis. For instance, when the number of functions is greater than 50, consecutive local exponential bases overlap too much and the invertibility conditions are violated (see Section I of the supplementary materials). In practice, the Fourier basis is the most stable and was used to compare our method to other classifiers.

B. Comparison with other classifiers

The performances of MIMGP are compared to three other methods: Quadratic Discriminant Analysis (QDA) which involves a similar Gaussian assumption on regularly sampled data, a linear Support Vector Machine (SVM) fitted with a Stochastic Gradient Descent (SGD) [54], and Random Forests (RF) [55] which have shown state of the art results in large scale pixel-wise classification of SITS [31].

RF is used with 100 trees and a maximum depth of tree set to 25. MIMGP model is implemented with the Fourier basis and a RBF kernel with an additive colored noise which represents a total of 32 parameters for θ .

TABLE III
AVERAGED “MEAN F_1 SCORE” (MEAN(%) \pm STANDARD DEVIATION) COMPUTED ON 3 TILES. MIMGP MODEL IS PARAMETRIZED BY A FOURIER BASIS WITH 19 PARAMETERS FOR T31TCJ, AND 41 FOR T31TDN AND T31TGK TILES.

	QDA	SVM	RF	MIMGP
T31TCJ	36.2 \pm 2.5	70.1 \pm 2.1	72.2 \pm 1.6	57.6 \pm 2.0
T31TDN	30.5 \pm 2.8	74.8 \pm 1.9	77.6 \pm 1.5	65.1 \pm 1.0
T31TGK	38.9 \pm 2.1	62.1 \pm 1.6	63.9 \pm 1.6	45.5 \pm 2.7

Table III summarizes the results for the four methods⁵. We can see that some significant improvements are achieved compared to the standard QDA classifier, even though MIMGP imposes independence between wavelengths. The QDA classifier suffers from numerical instability due to the high number of spectro-temporal features. However, MIMGP is not as accurate as the RF and the SVM classifiers applied on the reconstructed time-series.

Figure 9 shows three extracts of the classification map for tile T31TCJ⁶. Visually, significant differences can be observed both in terms of homogeneity and in terms of pixel-wise prediction. For instance, for the airport area, SVM predicts wrongly a lot of *natural grasslands* while MIMGP and RF predict correctly *meadow* for vegetation between runways. Overall, MIMGP and SVM seem to predict more homogeneous maps than RF. This can be clearly seen in the second row, where the city center classification maps obtained with RF contain salt and pepper noise.

Results provided in this section show that MIMGP is not at the level of state-of-the art methods such as RF or SVM. One critical point is the spectral independence assumption that is not true in practice and used here for computational purposes. Removing this constraint would help to improve the classification accuracy.

Yet, the computation of MIMGP does not require any temporal resampling processes and directly handles the raw time-series in the training and prediction steps. In comparison with standard GPs for classification, its complexity is linear w.r.t to the number of samples and cubic w.r.t to the number of temporal acquisitions, thus allowing large scale processing. Furthermore, it can also be used to reconstruct pixel time-series, as discussed in the next section.

⁵Per class F_1 scores are provided in the supplementary materials.

⁶Full classification map and extracts for the three tiles are provided in the supplementary materials.

VII. TIME-SERIES RECONSTRUCTION

Following Section III-B, MIMGP is able to reconstruct missing values conditionally to a known, or predicted, class. Class mean functions and covariance operators are also obtained after the training step (for clarity, mean functions and covariance operators are provided in Section III of the supplementary materials). In the following, the quality of the reconstruction is compared with a state-of-the-art method: the Whittaker smoother [37], [38].

To assess the quality of the reconstruction, the validation set was used to ensure that reconstructed pixels were not seen during the training phase. We randomly remove one clear date from each pixel and reconstruct it with the trained MIMGP model and with the Whittaker smoother for which the regularization parameter is set using the V-curve technique for each individual pixel [38], [56]. The quality of the reconstruction is estimated using the normalized Mean Absolute Error (nMAE) computed as:

$$\text{nMAE} = \frac{\sum_{i=1}^{n_v} |Y_b^i(t^*) - \hat{Y}_b^i(t^*)|}{\sum_{i=1}^{n_v} |Y_b^i(t^*) - \bar{Y}_b|}$$

with $\bar{Y}_b = \frac{1}{n_v} \sum_{i=1}^{n_v} Y_b^i(t^*)$, n_v is the number of validation pixels. Table IV summarizes the results obtained on the three tiles and for the 10 independent data sets. The scatter plots associated with nMAE results are provided in the supplementary materials.

From the table, MIMGP outperforms the Whittaker smoother in terms of nMAE for each band for tiles T31TDN and T31TGK. For tile T31TCJ, Whittaker is better for 5 bands. It is important to note that MIMGP provides better nMAE and does not need any additional fitting, apart the training step, in contrast to the Whittaker smoother that requires to be fitted for each pixel, as well its regularization parameter.

Figure 10 shows a part of the tile T31TGK reconstructed for the band 4 (red) on August 6, 2018. For this simulation, all the clear dates are used to reconstruct the time series for both methods. Interestingly, for the Whittaker smoother, the clouds mask is visible indicating a non continuous reconstruction of the dynamic. This artifact is not visible for the proposed method.

Figure 11 illustrates the red wavelength (R) for two pixels of the same class (summer crops). They were selected to enlighten the robustness of the proposed method to an inaccurate cloud mask file. The first pixel has an undetected cloud for the third temporal acquisition (blue point in the upper Figure 11). A clear drop of the reflectance can be seen in the gap-filled infra-red (dashed red line). The reconstructed reflectance for MIMGP (black line) and for Whittaker smoother (black dotted line) do not exhibit such a drop in the reflectance. Similar comments can be done for the second pixel of the lower Figure 11 which has a undetected saturation (pixel value equal to zero). Other wavelengths are reported in the supplementary materials.

VIII. CONCLUSION

This work introduces a novel approach to jointly classify and reconstruct irregularly sampled multidimensional time-

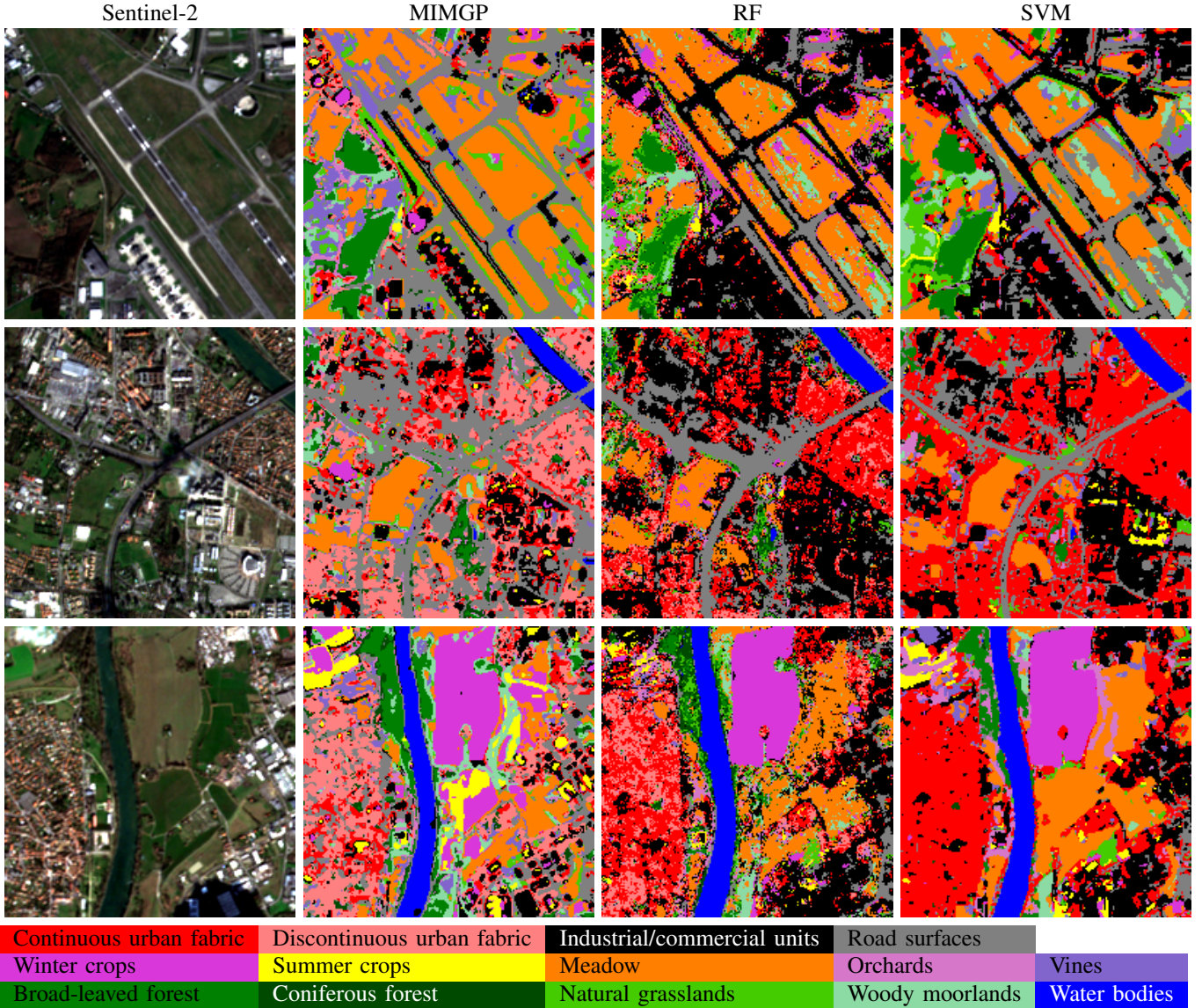


Fig. 9. Thematic maps for 3 sites from tile T31TCJ.

series. The proposed model, namely MIMGP, only involves a small number of parameters and is scalable to large datasets. The performances of the method were illustrated on a full-year of Sentinel-2 satellite image time-series dataset from 2018 involving a high number of temporal acquisitions. A by-product of MIMGP is also to infer a confidence interval on the reconstruction.

The reconstruction has shown a good behavior on noisy pixels, while the model does not compete with state of the art classifiers such as Random Forest or SVM. One limitation comes from the independence hypothesis between spectral wavelengths, which is not true in practice. Therefore, our future work will be dedicated to the definition of multivariate Gaussian Processes able to take into account the correlation between wavelengths.

Another work in progress concerns the estimation of the mean function. MIMGP is able to handle the irregular sampling caused by multiple tiles, but the mean function is assumed to be spatially stationary. Rather than using a linear

combination of fixed basis functions, a more complex approximating function depending on the geographical location could be used.

APPENDIX A TIME-SERIES RECONSTRUCTION

Let us first consider the case where the class of the missing value is known to be c . In such a situation, the reconstruction can be achieved using the conditional expectation:

$$\hat{Y}_{b,c}^j(t^*) := \mathbb{E}[Y_b^j(t^*) | \mathbf{y}_j, Z_j = c].$$

The independence assumption A2 then yields

$$\mathbb{E}[Y_b^j(t^*) | \mathbf{y}_j, Z_j = c] = \mathbb{E}[Y_b^j(t^*) | \mathbf{y}_{j,b}, Z_j = c].$$

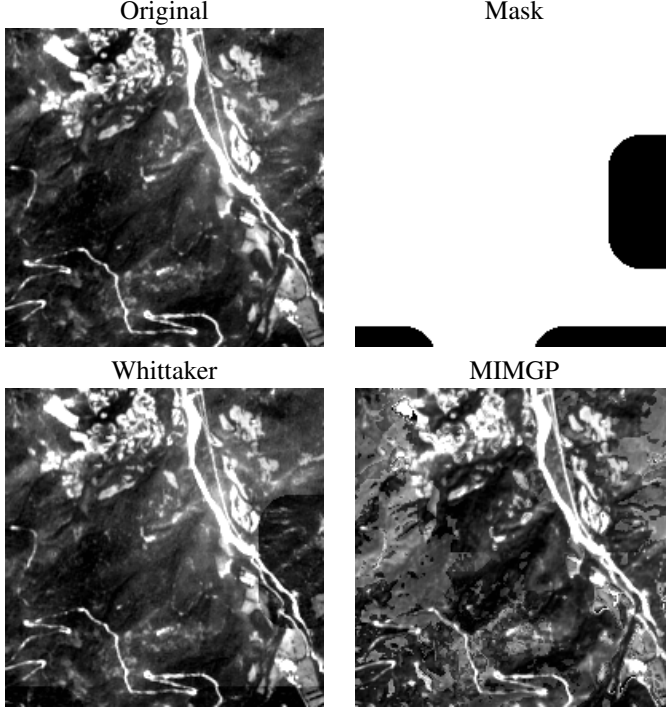


Fig. 10. Image reconstruction. The top left panel is the original spectral band 4 (red) on August 6, 2018. The top right panel corresponds to the mask: white color for a clear pixel and black color for pixel with clouds/saturation as detected by the pre-processing chain. Bottom left and right are the reconstruction with the Whittaker smoother and the MIMGP, respectively.

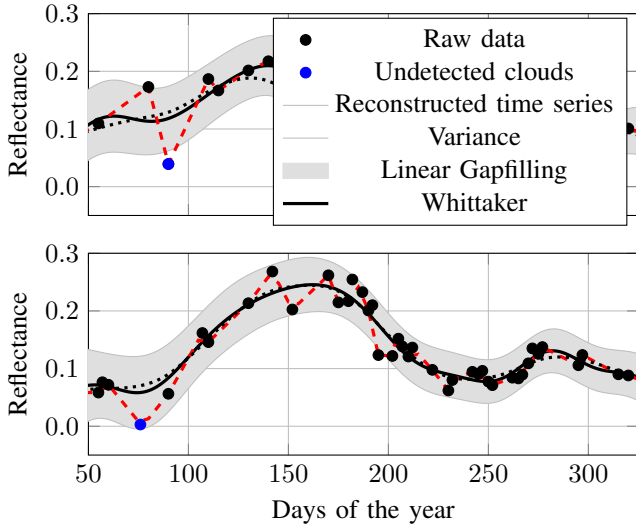


Fig. 11. Time-series reconstruction. Top and bottom figures represent the red reflectance for two different pixels. Dot points are the original values from the irregularly sampled time-series. Blue points are non-detected clouds and saturation by the satellite data preprocessing chain, for the top and bottom figure, respectively. The black continuous line is the conditional reconstruction of the signal and the gray region corresponds to the 95% confidence interval. Each day of the year was reconstructed with MIMGP. The red dashed line represents the linear interpolation taking into account all the dates tagged “clear”, the black dotted line represents the smoothed signal using Whittaker smoother.

TABLE IV
AVERAGED NORMALIZED MAE (MEAN (%) \pm STANDARD DEVIATION)
FOR EACH WAVELENGTH FOR THE DIFFERENT TILES WITH THE A FOURIER
BASIS OF DIMENSION $J = 19$.

S2 Band	Method	T31TCJ	T31TDN	T31TGK
Band 2	Whittaker	47.1 \pm 0.5	123 \pm 1.5	40.1 \pm 1.2
	MIMGP	46.9 \pm 2.4	46.5 \pm 1.3	36.3 \pm 1.4
Band 3	Whittaker	45.2 \pm 0.5	114 \pm 1.2	42.3 \pm 1.3
	MIMGP	43.9 \pm 3.7	44.0 \pm 1.6	39.6 \pm 1.4
Band 4	Whittaker	37.7 \pm 0.4	94.2 \pm 1.0	43.5 \pm 1.3
	MIMGP	42.7 \pm 4.4	45.2 \pm 2.2	41.3 \pm 1.5
Band 5	Whittaker	40.0 \pm 0.5	99.9 \pm 1.0	44.7 \pm 1.3
	MIMGP	45.8 \pm 0.4	40.9 \pm 2.0	41.3 \pm 2.2
Band 6	Whittaker	32.6 \pm 0.4	83.2 \pm 1.0	51.1 \pm 1.3
	MIMGP	38.6 \pm 1.8	46.4 \pm 0.9	46.9 \pm 2.6
Band 7	Whittaker	30.0 \pm 0.3	73.5 \pm 0.9	51.1 \pm 1.2
	MIMGP	24.7 \pm 2.7	45.1 \pm 0.9	41.8 \pm 1.1
Band 8	Whittaker	32.6 \pm 0.4	71.8 \pm 0.9	52.9 \pm 1.2
	MIMGP	33.1 \pm 1.9	44.6 \pm 2.3	42.9 \pm 1.1
Band 8A	Whittaker	29.4 \pm 0.3	69.3 \pm 0.9	51.9 \pm 1.2
	MIMGP	33.2 \pm 3.4	43.8 \pm 1.5	42.6 \pm 1.2
Band 11	Whittaker	36.1 \pm 0.5	53.1 \pm 0.7	46.9 \pm 0.6
	MIMGP	37.0 \pm 2.0	38.3 \pm 1.7	43.0 \pm 1.3
Band 12	Whittaker	34.1 \pm 0.4	46.0 \pm 0.6	51.3 \pm 0.7
	MIMGP	39.5 \pm 2.6	38.7 \pm 3.3	46.6 \pm 2.6

Besides, conditionally on $Z_i = c$, one has

$$\begin{pmatrix} Y_b^j(t^*) \\ \mathbf{y}_{j,b} \end{pmatrix} \sim \mathcal{N}_{T_j+1} \left(\begin{bmatrix} \mathbf{b}^* \boldsymbol{\alpha}_{b,c} \\ \mathbf{B}^j \boldsymbol{\alpha}_{b,c} \end{bmatrix}, \begin{bmatrix} K(t^*, t^* | \boldsymbol{\theta}_{b,c}) & \mathbf{k}(t^*, t_{1:T_j}^j | \boldsymbol{\theta}_{b,c})^\top \\ \mathbf{k}(t^*, t_{1:T_j}^j | \boldsymbol{\theta}_{b,c}) & \boldsymbol{\Sigma}^j(\boldsymbol{\theta}_{b,c}) \end{bmatrix} \right),$$

where $\mathbf{k}(t^*, t_{1:T_j} | \boldsymbol{\theta}_{b,c}) = [K(t^*, t_1 | \boldsymbol{\theta}_{b,c}), \dots, K(t^*, t_{T_j} | \boldsymbol{\theta}_{b,c})]^\top$, $\mathbf{b}^* = [\varphi_1(t^*), \dots, \varphi_J(t^*)]$ and \mathbf{B}^j is the $T_j \times J$ design matrix associated with the time-series \mathbf{y}_j such that $(\mathbf{B}^j)_{\ell,k} = \varphi_k(t_\ell^j)$ for all $(\ell, k) \in \{1, \dots, T_j\} \times \{1, \dots, J\}$. From classical properties of conditional Gaussian distributions (see for instance [51, p.63]), it follows that

$$\begin{aligned} \mathbb{E}(Y_b^j(t^*) | \mathbf{y}_{j,b}; Z_j = c) &= \mathbf{b}^* \boldsymbol{\alpha}_{b,c}, \\ &+ \mathbf{k}(t^*, t_{1:T_j}^j | \boldsymbol{\theta}_{b,c}) \boldsymbol{\Sigma}^j(\boldsymbol{\theta}_{b,c})^{-1} (\mathbf{y}_{i,b} - \mathbf{B}^j \boldsymbol{\alpha}_{b,c}), \end{aligned}$$

$$\begin{aligned} \mathbb{V}(Y_b^j(t^*) | \mathbf{y}_{j,b}; Z_j = c) &= K(t^*, t^* | \boldsymbol{\theta}_{b,c}) \\ &- \mathbf{k}(\boldsymbol{\theta}_{b,c} | t^*, t_{1:T_j}^j) \boldsymbol{\Sigma}^j(\boldsymbol{\theta}_{b,c})^{-1} \mathbf{k}(t^*, t_{1:T_j}^j | \boldsymbol{\theta}_{b,c})^\top. \end{aligned}$$

Finally, replacing the unknown quantities by their estimated counterparts yields the desired results.

Second, when the class is unknown, the reconstruction rule is given by

$$\hat{Y}_b^j(t^*) = \mathbb{E}[Y_b^j(t^*) | \mathbf{y}_j].$$

Since

$$\mathbb{E}[Y_b^j(t^*) | \mathbf{y}_j] = \sum_{c=1}^c \mathbb{P}(Z_j = c | \mathbf{y}_j) \mathbb{E}[Y_b^j(t^*) | \mathbf{y}_j, Z_j = c],$$

it straightforwardly follows that

$$\hat{Y}_b^j(t^*) = \sum_{c=1}^c \mathbb{P}(Z_j = c | \mathbf{y}_j) \hat{Y}_{b,c}^j(t^*),$$

where the reconstructed value $\hat{Y}_{b,c}^j(t^*)$ is provided in (11) while the posterior probabilities are given by (10). The variance is obtained by a similar calculation:

$$\begin{aligned} \mathbb{V}(Y_b^j(t^*)) &= \mathbb{E}[Y_b^j(t^*)^2 | \mathbf{y}_j] - \mathbb{E}[Y_b^j(t^*) | \mathbf{y}_j]^2, \\ &= \sum_{c=1}^c \mathbb{P}(Z_j = c | \mathbf{y}_j) \mathbb{E}[Y_b^j(t^*)^2 | \mathbf{y}_j, Z_j = c] \\ &\quad - \hat{Y}_b^j(t^*)^2. \end{aligned}$$

Besides, remarking that

$$\begin{aligned} \mathbb{E}[Y_b^j(t^*) | \mathbf{y}_j, Z_j = c] &= \hat{Y}_{b,c}^j(t^*), \\ \mathbb{E}[Y_b^j(t^*)^2 | \mathbf{y}_j, Z_j = c] &= \mathbb{V}(Y_{b,c}^j(t^*) | \mathbf{y}_j, Z_j = c) \\ &\quad + \mathbb{E}[Y_{b,c}^j(t^*) | \mathbf{y}_j, Z_j = c]^2, \end{aligned}$$

it follows that

$$\begin{aligned} \mathbb{V}(\hat{Y}_b^j(t^*)) &= \sum_{c=1}^c \mathbb{P}(Z_j = c | \mathbf{y}_j) \mathbb{V}(\hat{Y}_{b,c}^j(t^*) | \mathbf{y}_j, Z_j = c) \\ &\quad + \sum_{c=1}^c \mathbb{P}(Z_j = c | \mathbf{y}_j) \hat{Y}_{b,c}^j(t^*)^2 - \hat{Y}_b^j(t^*)^2, \end{aligned}$$

and the result is proved. Let us highlight that these derivations were conducted when a single time t^* is considered. Similar calculations can be achieved when reconstructed simultaneously several values. In such a case, this estimation procedure provides the covariance matrix of the reconstructed values at each time.

ACKNOWLEDGMENT

The authors thank the reviewers for their many constructive and helpful comments. The authors would also like to thank S. Iovleff for his support and advices during the construction of the model. The authors would also like to thank Y. Tanguy for his help when using the CNES computational resources to run the experiments presented in this paper.

REFERENCES

- [1] M. Drusch, U. Del Bello, S. Carlier, O. Colin, V. Fernandez, F. Gascon, B. Hoersch, C. Isola, P. Laberinti, P. Martimort, A. Meygret, F. Spoto, O. Sy, F. Marchese, and P. Bargellini, "Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services," *Remote Sensing of Environment*, vol. 120, pp. 25–36, May 2012.
- [2] P. Soille, A. Burger, D. D. Marchi, P. Kempeneers, D. Rodriguez, V. Syrris, and V. Vasilev, "A versatile data-intensive computing platform for information retrieval from big geospatial data," *Future Generation Computer Systems*, vol. 81, pp. 30–40, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167739X1730078X>
- [3] M. Pereira-Sandoval, A. Ruiz-Verdu, C. Tenjo, J. Delegido, P. Urrego, R. Pena, E. Vicente, J. Soria, J. Soria, and J. Moreno, "Calibration and Validation of Algorithms for the Estimation of Chlorophyll-A in Inland waters with Sentinel-2," in *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 7 2018, p. nil. [Online]. Available: <https://doi.org/10.1109/igarss.2018.8517371>
- [4] E. Batur and D. Maktav, "Assessment of Surface Water Quality by Using Satellite Images Fusion Based on PCA Method in the Lake Gala, Turkey," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 5, p. 29832989, May 2019. [Online]. Available: <http://dx.doi.org/10.1109/TGRS.2018.2879024>
- [5] A. Moeni Rad, D. Ashourloo, H. Salehi Shahrabi, and H. Nematollahi, "Developing an Automatic Phenology-Based Algorithm for Rice Detection Using Sentinel-2 Time-Series Data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 5, p. 14711481, May 2019. [Online]. Available: <http://dx.doi.org/10.1109/JSTARS.2019.2906684>
- [6] S. Feng, J. Zhao, T. Liu, H. Zhang, Z. Zhang, and X. Guo, "Crop Type Identification and Mapping Using Machine Learning Algorithms and Sentinel-2 Time Series Data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 9, p. 32953306, Sep 2019. [Online]. Available: <http://dx.doi.org/10.1109/JSTARS.2019.2922469>
- [7] J. Useya and S. Chen, "Comparative Performance Evaluation of Pixel-Level and Decision-Level Data Fusion of Landsat 8 OLI, Landsat 7 ETM+ and Sentinel-2 MSI for Crop Ensemble Classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 11, p. 44414451, Nov 2018. [Online]. Available: <http://dx.doi.org/10.1109/JSTARS.2018.2870650>
- [8] G. C. Iannelli and P. Gamba, "Urban Extent Extraction Combining Sentinel Data in the Optical and Microwave Range," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 7, p. 22092216, Jul 2019. [Online]. Available: <http://dx.doi.org/10.1109/JSTARS.2019.2920678>
- [9] J. Haas and Y. Ban, "Urban Land Cover and Ecosystem Service Changes based on Sentinel-2A MSI and Landsat TM Data," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 2, p. 485497, Feb 2018. [Online]. Available: <http://dx.doi.org/10.1109/JSTARS.2017.2786468>
- [10] C. Li, H. Wulf, B. Schmid, J.-S. He, and M. E. Schaepman, "Estimating Plant Traits of Alpine Grasslands on the Qinghai-Tibetan Plateau Using Remote Sensing," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 7, p. 22632275, Jul 2018. [Online]. Available: <http://dx.doi.org/10.1109/JSTARS.2018.2824901>
- [11] M. Fauvel, M. Lopes, T. Dubo, J. Rivers-Moore, P.-L. Frison, N. Gross, and A. Ouin, "Prediction of plant diversity in grasslands using Sentinel-1 and -2 satellite image time series," *Remote Sensing of Environment*, vol. 237, p. 111536, Feb 2020. [Online]. Available: <http://dx.doi.org/10.1016/j.rse.2019.111536>
- [12] M. Chi, A. Plaza, J. A. Benediktsson, Z. Sun, J. Shen, and Y. Zhu, "Big Data for Remote Sensing: Challenges and Opportunities," *Proceedings of the IEEE*, vol. 104, no. 11, p. 22072219, Nov 2016. [Online]. Available: <http://dx.doi.org/10.1109/JPROC.2016.2598228>
- [13] C. Revel, V. Lonjou, S. Marcq, C. Desjardins, B. Fougne, C. Coppolani-Delle Luche, N. Guillemot, A.-S. Lacamp, E. Lourme, C. Miquel, and et al., "Sentinel-2A and 2B absolute calibration monitoring," *European Journal of Remote Sensing*, vol. 52, no. 1, p. 122137, Jan 2019. [Online]. Available: <http://dx.doi.org/10.1080/22797254.2018.1562311>
- [14] M. Sudmanns, D. Tiede, S. Lang, H. Bergstedt, G. Trost, H. Augustin, A. Baraldi, and T. Blaschke, "Big Earth data: disruptive changes in Earth Observation data management and analysis?" *International Journal of Digital Earth*, p. 119, Mar 2019. [Online]. Available: <http://dx.doi.org/10.1080/17538947.2019.1585976>
- [15] P. Kempeneers and P. Soille, "Optimizing Sentinel-2 Image Selection in a Big Data Context," *Big Earth Data*, vol. 1, no. 1-2, pp. 145–158, 2017. [Online]. Available: <https://doi.org/10.1080/20964471.2017.1407489>
- [16] S. Skakun, E. Vermote, J.-C. Roger, and C. Justice, "Multispectral Misregistration of Sentinel-2A Images: Analysis and Implications for Potential Applications," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 12, pp. 2408–2412, 2017. [Online]. Available: <https://doi.org/10.1109/lgrs.2017.2766448>
- [17] L. Baetens, C. Desjardins, and O. Hagolle, "Validation of Copernicus Sentinel-2 Cloud Masks Obtained from MAJA, Sen2Cor, and FMask Processors Using Reference Cloud Masks Generated with a Supervised Active Learning Procedure," *Remote Sensing*, vol. 11, no. 4, p. 433, Feb 2019. [Online]. Available: <http://dx.doi.org/10.3390/rs11040433>
- [18] J. Wang, B. Huang, H. K. Zhang, and P. Ma, "Sentinel-2A Image Fusion Using a Machine Learning Approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, p. 95899601, Dec 2019. [Online]. Available: <http://dx.doi.org/10.1109/TGRS.2019.2927766>
- [19] C. Paris, J. Bioucas-Dias, and L. Bruzzone, "A Novel Sharpening Approach for Superresolving Multiresolution Optical Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 3, p.

- 15451560, Mar 2019. [Online]. Available: <http://dx.doi.org/10.1109/TGRS.2018.2867284>
- [20] M. O. Ulfarsson, F. Palsson, M. Dalla Mura, and J. R. Sveinsson, "Sentinel-2 Sharpening Using a Reduced-Rank Method," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, p. 64086420, Sep 2019. [Online]. Available: <http://dx.doi.org/10.1109/TGRS.2019.2906048>
- [21] Z. Shao, J. Cai, P. Fu, L. Hu, and T. Liu, "Deep learning-based fusion of Landsat-8 and Sentinel-2 images for a harmonized surface reflectance product," *Remote Sensing of Environment*, vol. 235, p. 111425, Dec 2019. [Online]. Available: <http://dx.doi.org/10.1016/j.rse.2019.111425>
- [22] R. Fernandez-Beltran, J. M. Haut, M. E. Paoletti, J. Plaza, A. Plaza, and F. Pla, "Multimodal Probabilistic Latent Semantic Analysis for Sentinel-1 and Sentinel-2 Image Fusion," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 9, p. 13471351, Sep 2018. [Online]. Available: <http://dx.doi.org/10.1109/LGRS.2018.2843886>
- [23] Y. Shendryk, Y. Rist, C. Ticehurst, and P. Thorburn, "Deep learning for multi-modal classification of cloud, shadow and land cover scenes in planetoscope and Sentinel-2 imagery," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 157, p. 124136, Nov 2019. [Online]. Available: <http://dx.doi.org/10.1016/j.isprsjprs.2019.08.018>
- [24] V. Syrris, P. Hasenohr, B. Delipetrev, A. Kotsev, P. Kempeneers, and P. Soille, "Evaluation of the Potential of Convolutional Neural Networks and Random Forests for Multi-Class Segmentation of Sentinel-2 Imagery," *Remote Sensing*, vol. 11, no. 8, p. 907, Apr 2019. [Online]. Available: <http://dx.doi.org/10.3390/rs11080907>
- [25] C. Qiu, L. Mou, M. Schmitt, and X. X. Zhu, "Local climate zone-based urban land cover classification from multi-seasonal Sentinel-2 images with a recurrent residual network," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 154, p. 151162, Aug 2019. [Online]. Available: <http://dx.doi.org/10.1016/j.isprsjprs.2019.05.004>
- [26] M. Ruwurm and M. Krner, "Multi-Temporal Land Cover Classification With Long Short-Term Memory Neural Networks," *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-1/W1, p. 551558, May 2017. [Online]. Available: <http://dx.doi.org/10.5194/isprs-archives-XLII-1-W1-551-2017>
- [27] P. Benedetti, D. Ienco, R. Gaetano, K. Ose, R. G. Pensa, and S. Dupuy, " M^3 Fusion: A Deep Learning Architecture for Multiscale Multimodal Multitemporal Satellite Data Fusion," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 12, p. 49394949, Dec 2018. [Online]. Available: <http://dx.doi.org/10.1109/JSTARS.2018.2876357>
- [28] C. Pelletier, G. I. Webb, and F. Petitjean, "Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series," *Remote Sensing*, vol. 11, no. 5, p. 523, Jan. 2019. [Online]. Available: <https://www.mdpi.com/2072-4292/11/5/523>
- [29] Y. J. Eudes Gbodjo, D. Ienco, and L. Leroux, "Toward SpatioSpectral Analysis of Sentinel-2 Time Series Data for Land Cover Mapping," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 2, p. 307311, Feb 2020. [Online]. Available: <http://dx.doi.org/10.1109/LGRS.2019.2917788>
- [30] J. Inglada, A. Vincent, M. Arias, B. Tardy, D. Morin, and I. Rodes, "Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series," *Remote Sensing*, vol. 9, no. 1, 2017.
- [31] D. Derksen, J. Inglada, and J. Michel, "Geometry Aware Evaluation of Handcrafted Superpixel-Based Features and Convolutional Neural Networks for Land Cover Mapping Using Satellite Imagery," *Remote Sensing*, vol. 12, no. 3, 2020. [Online]. Available: <https://www.mdpi.com/2072-4292/12/3/513>
- [32] H. Shen, X. Li, Q. Cheng, C. Zeng, G. Yang, H. Li, and L. Zhang, "Missing Information Reconstruction of Remote Sensing Data: A Technical Review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 3, no. 3, pp. 61–85, Sep. 2015.
- [33] J. Luo, K. Ying, and J. Bai, "SavitzkyGolay smoothing and differentiation filter for even number data," *Signal Processing*, vol. 85, no. 7, p. 14291434, Jul 2005. [Online]. Available: <http://dx.doi.org/10.1016/j.sigpro.2005.02.002>
- [34] N. VIOVY, O. ARINO, and A. S. BELWARD, "The Best Index Slope Extraction (BISE): A method for reducing noise in NDVI time-series," *International Journal of Remote Sensing*, vol. 13, no. 8, p. 15851590, May 1992. [Online]. Available: <http://dx.doi.org/10.1080/01431169208904212>
- [35] P. Beck, C. Atzberger, K. Hogda, B. Johansen, and A. Skidmore, "Improved monitoring of vegetation dynamics at very high latitudes: a new method using MODIS NDVI," *Remote Sensing of Environment* 100 (2006) 3, vol. 100(3), 01 2006.
- [36] P. Jonsson and L. Eklundh, "Seasonality extraction by function fitting to time-series of satellite sensor data," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 40, pp. 1824 – 1832, 09 2002.
- [37] P. H. C. Eilers, "A Perfect Smoother," *Analytical Chemistry*, vol. 75, no. 14, pp. 3631–3636, Jul. 2003, publisher: American Chemical Society. [Online]. Available: <https://doi.org/10.1021/ac034173t>
- [38] F. Vuolo, W.-T. Ng, and C. Atzberger, "Smoothing and gap-filling of high resolution multi-spectral time series: Example of Landsat data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 57, pp. 202–213, May 2017, tex.ids: vuoloSmoothingGapfillingHigh2017. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0303243416302100>
- [39] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR05)*, 2005. [Online]. Available: <http://dx.doi.org/10.1109/CVPR.2005.38>
- [40] Q. Zhang, Q. Yuan, C. Zeng, X. Li, and Y. Wei, "Missing Data Reconstruction in Remote Sensing Image With a Unified SpatialTemporalSpectral Deep Convolutional Neural Network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 8, pp. 4274–4288, Aug. 2018.
- [41] Q. Zhang, Q. Yuan, J. Li, Z. Li, H. Shen, and L. Zhang, "Thick cloud and cloud shadow removal in multitemporal imagery using progressively spatio-temporal patch group deep learning," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 162, pp. 148 – 160, 2020. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0924271620300423>
- [42] R. Cresson, D. Ienco, R. Gaetano, K. Ose, and D. H. Tong Minh, "Optical image gap filling using deep convolutional autoencoder from optical and radar images," *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, Jul 2019. [Online]. Available: <http://dx.doi.org/10.1109/IGARSS.2019.8900353>
- [43] G. Camps-Valls, J. Verrelst, J. Munoz-Mari, V. Laparra, F. Mateo-Jimenez, and J. Gomez-Dans, "A Survey on Gaussian Processes for Earth-Observation Data Analysis: A Comprehensive Investigation," *IEEE Geoscience and Remote Sensing Magazine*, vol. 4, no. 2, p. 5878, Jun 2016. [Online]. Available: <http://dx.doi.org/10.1109/MGRS.2015.2510084>
- [44] 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, p. 18321843, May 2012. [Online]. Available: <http://dx.doi.org/10.1109/TGRS.2011.2168962>
- [45] U. B. Gewali, S. T. Monteiro, and E. Saber, "Gaussian Processes for Vegetation Parameter Estimation from Hyperspectral Data with Limited Ground Truth," *Remote Sensing*, vol. 11, no. 13, p. 1614, Jul 2019. [Online]. Available: <http://dx.doi.org/10.3390/rs11131614>
- [46] P. Morales-Alvarez, A. Perez-Suay, R. Molina, and G. Camps-Valls, "Remote Sensing Image Classification With Large-Scale Gaussian Processes," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, p. 11031114, Feb 2018. [Online]. Available: <http://dx.doi.org/10.1109/TGRS.2017.2758922>
- [47] L. Pipia, J. Muoz-Mar, E. Amin, S. Belda, G. Camps-Valls, and J. Verrelst, "Fusing optical and SAR time series for LAI gap filling with multioutput Gaussian processes," *Remote Sensing of Environment*, vol. 235, p. 111452, Dec. 2019. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0034425719304717>
- [48] C. K. Williams and C. E. Rasmussen, "Gaussian Processes for Machine Learning," *the MIT Press*, vol. 2, no. 3, p. 4, 2006.
- [49] H. Trevor, T. Robert, and F. JH, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY: Springer, 2009.
- [50] Q. V. Le, A. J. Smola, and S. Canu, "Heteroscedastic Gaussian process regression," in *ICML 2005: Proceedings of the 22nd international conference on Machine learning*, New York, NY, USA, 2005, pp. 489–496. [Online]. Available: <https://doi.org/10.1145/1102351.1102413>
- [51] M. Bilodeau and D. Brenner, *Theory of Multivariate Statistics*. Springer Science & Business Media, 2008.
- [52] L. Baetens, C. Desjardins, and O. Hagolle, "Validation of Copernicus Sentinel-2 Cloud Masks Obtained from MAJA, Sen2Cor, and FMask Processors Using Reference Cloud Masks Generated with a Supervised Active Learning Procedure," *Remote Sensing*, vol. 11, no. 4, 2019. [Online]. Available: <http://www.mdpi.com/2072-4292/11/4/433>
- [53] O. D. Team, "Orfeo ToolBox 7.1," Mar. 2020. [Online]. Available: <https://doi.org/10.5281/zenodo.3715021>

- [54] T. Zhang, "Solving Large Scale Linear Prediction Problems Using Stochastic Gradient Descent Algorithms," in *ICML 2004: Proceedings of the Twenty-First International Conference on Machine Learning*. Omnipress, 2004, pp. 919–926.
- [55] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
- [56] G. Frasso and P. H. C. Eilers, "L- and V-curves for optimal smoothing," *Statistical Modelling*, vol. 15, pp. 91–111, Feb. 2015, publisher: SAGE Publications India. [Online]. Available: <https://doi.org/10.1177/1471082X14549288>