



**HAL**  
open science

# Supervised Classification of Irregularly Sampled Sattelite Image Times Series

Alexandre Constantin, Mathieu Fauvel, Stéphane Girard

► **To cite this version:**

Alexandre Constantin, Mathieu Fauvel, Stéphane Girard. Supervised Classification of Irregularly Sampled Sattelite Image Times Series. 2020. hal-02997573v1

**HAL Id: hal-02997573**

**<https://inria.hal.science/hal-02997573v1>**

Preprint submitted on 10 Nov 2020 (v1), 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.

# Supervised Classification of Irregularly Sampled Satellite Image Time Series

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

**Abstract**—Recent satellite missions launched in the past decades have led to a huge amount of earth observation data, most of them being freely available. In such 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 contain clouds and shadows. Thus specific pre-processing techniques are required to properly classify such data. The proposed approach in the paper is able to deal with irregular temporal sampling and missing data directly in the classification process. It is based on Gaussian Process and allows to perform jointly the classification of the pixel labels as well as the reconstruction of the pixel time series. Its complexity scales linearly with the number of pixels, it is amenable in a large scale scenario. Experimental classification and reconstruction show that the method does not compete yet with state-of-the-art classifier but allows for a robust reconstruction and does not need any temporal preprocessing.

**Index Terms**—Satellite Image Time Series (SITS), Sentinel-2, classification, 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 a 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 5 days systematic revisit time [1]. It is planned for long-term operational observations (more than 15 years) and 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, optical Sentinel SITS have found application in various scientific fields: water management [3], [4], agricultural systems mapping [5]–[8], urban area analysis [9], [10] or ecological monitoring [11], [12]. However, such abundance of images raises new challenges in terms of large scale multi-resolution SITS processing [13]. Issues related to *Big Earth Data* were obviously explored in the remote sensing community in order to calibrate and distribute the images seamlessly [14]–[18].

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 Sentinel-2 SITS properties. Spatial differences in terms of resolution were tackled using super-resolution strategies in order to resample all the spectral bands at the same finest spatial resolution (10m) [19]–[21]. Fusion with other sensors, such as Landsat [22] or Sentinel-1 [23], was also considered to complement Sentinel-2 SITS.

Classification of land cover or land use using machine learning techniques has emerged as a major application of Sentinel-2 SITS. Deep learning was intensively studied either using only one or few temporal acquisitions with convolutional neural network (CNN) [24], [25], or using the full temporal stack of acquisitions using recurrent neural network (RNN) [26], [27]. Combination of RNN applied on Sentinel-2 SITS with CNN applied on SPOT-6 image was also investigated for land cover mapping over Reunion Island [28]. Temporal convolution was investigated in [29]. Yet, due to the huge amount of data to be processed when large geographical areas or large temporal domains (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 [30], [31]. 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 [32].

One major issue arising when dealing with large Sentinel-2 SITS, *i.e.* using several tracks or tiles<sup>1</sup>, 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 result in pixels of various sizes, even though they belong to the same tile [31].

Conventional machine learning algorithms take as input a sample of finite and constant size. Random Forest, Support Vector Machine or RNN use a vector representation of input features, *i.e.*, a sample is a collection of features stacked into a  $p$ -dimensional vector and each sample has the same

<sup>1</sup>Sentinel-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>

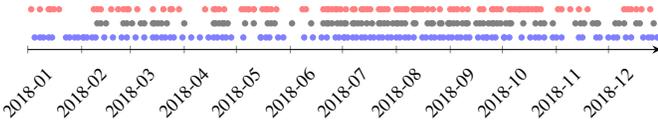


Fig. 1. Acquisition dates for tile T31TDN (red), T31TCJ (black) and T31TGK (blue) for the year 2018. The location of the tiles is given in Fig. 5. Large temporal gaps between two dates corresponds to very cloudy acquisitions: Images tagged as composed of more than 90% of cloudy pixels are not processed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

number of features. Similarly, CNN use a constant size patch representation of the image. Therefore, a pre-processing step, called *missing information reconstruction* is usually applied on the SITS in order to recover samples of same size [33]. Among the various existing techniques, temporal filter, such as the Savitzky-Golay filter [34] or the best index slope extraction (BISE) [35], and parametric curve fitting [36], [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 [31].

Methods taking account both of the spatial and temporal information have provided interesting results, see for instance non-local filters [38] or deep learning based approaches [39]–[41]. However, due to their high computational cost, they are hardly applicable in national/continental scale settings [32]. Furthermore, they are also known as “black-box” and hardly interpretable.

Recently, Gaussian Processes (GP) have gained attention in the remote sensing community [42]. GP mixed Bayesian and Kernel methods to build statistical learning machines, for classification or regression problems, that shown to be accurate, and interpretable through their hyperparameters (mean and covariance functions) [43]–[45]. In this work, it is proposed to define a Gaussian Process classifier that handles irregularly samples in the learning and prediction steps. It relies on a linear projection of the mean function on a well-chosen basis, conditionally to the classes of interest.

The contribution of this work is three-fold:

- 1) The definition of a GP that handles irregular temporal sampling (Section II), without the need to project all the data on a common temporal grid.
- 2) In addition to the classification, the proposed model allows for the reconstruction of the sample on any temporal grid, with confidence intervals (Section III). Furthermore, the mean and covariance functions for each class are available for analysis.
- 3) The reconstruction is optimal for the classification error, i.e., for the final objective. That is not the case for interpolation methods that are usually based on a reconstruction error.

The remainder of this paper is organized as follows. Sections II and III present the proposed GP model. Section IV details the Sentinel-2 SITS datasets and section V discuss the hyperparameters model’s set-up. Finally experimental results are given in sections VI and VII. Section VIII concludes this paper and opens discussion on future work.

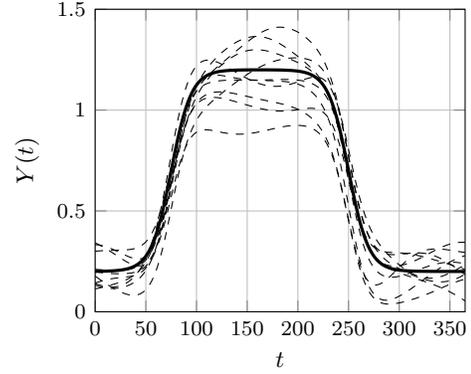


Fig. 2. Simulated univariate Gaussian processes with squared exponential covariance function (length scale parameter set to 50). The continuous black line is the mean function and the dashed lines are the corresponding 10 realizations.

## II. IRREGULARLY SAMPLED GAUSSIAN PROCESSES MODEL

In the following  $\mathcal{S} = \{(\mathbf{y}_i, z_i)\}_{i=1}^n$  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. A pixel  $\mathbf{y}$  is modeled by a random vector  $Y$  containing  $p$  random square integrable processes  $\mathcal{T} \rightarrow \mathbb{R}^p$ , where  $\mathcal{T} = [0, T]$  is the temporal 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 wavelengths and/or spectral indices and  $T$  is equal to 365 days (1 year).

A (univariate) Gaussian process (GP)  $f$  is a stochastic process such that any finite-dimensional marginal follows a multivariate Gaussian distribution [46]. It is specified by its mean function  $m$  and covariance function  $K$ :

$$m(t) = \mathbb{E}[f(t)],$$

$$K(t, s) = \mathbb{E}[(f(t) - m(t))(f(s) - m(s))],$$

with  $t \in \mathcal{T}$  and  $s \in \mathcal{T}$ . It is denoted as  $f \sim \mathcal{GP}(m, K)$ . 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  $b \in \{1, \dots, p\}$ ,  $c \in \{1, \dots, C\}$ . 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 function. The second property is introduced for computational reasons.

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 size  $T_i$ :

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

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  $c \in \{1, \dots, C\}$ . In order 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 modelling of the latter one is conventional with GP, see for instance [46, 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

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

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

Any functional basis can be used, *e.g.*, Fourier, exponential, splines,  $\dots$ , see for instance [47, 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 [46, Chapter 4]. A typical one would be a squared exponential covariance function with an additional colored noise [48] 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 [46, 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  $\kappa$  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 conventional in GP regression, while for binary classification, Laplace approximation is usually employed (and more computationally demanding) [46, 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 [46, Chapter 5]. The algorithm is based on an alternate optimization of  $\boldsymbol{\alpha}$  and  $\boldsymbol{\theta}$  (see Algorithm 1). At iteration ( $k$ ), the update rule for  $\boldsymbol{\alpha}$  is given by

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

---

**Algorithm 1:** Estimation of model parameters.
 

---

**Input :**  $\mathcal{S}, \alpha^0, \theta^0$   
**Output:**  $\hat{\alpha}, \hat{\theta}$   
**1** for  $c=1$  to  $C$  do  
**2**     for  $b=1$  to  $p$  do  
**3**         repeat  
**4**             Update  $\alpha_{b,c}$  using (8);  
**5**             Do a gradient step w.r.t.  $\theta_{b,c}$  using (9);  
**6**             until  $\ell_{b,c}(\alpha_{b,c}, \theta_{b,c})$  has converged;

---

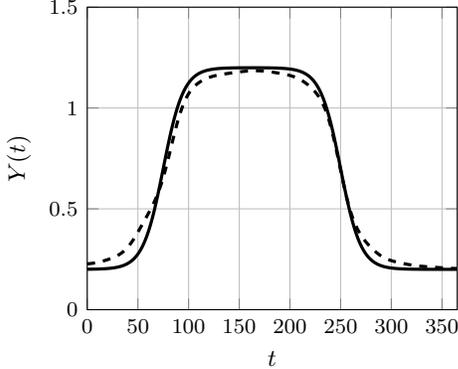


Fig. 3. The continuous line is the true mean function and the dashed line is the estimated one of GP shown in Figure 2 using Algorithm 1 and  $J = 20$  Gaussian functions (see Section V-A for details on bases).

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 material.

For  $\theta$ , 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  $\theta_m$  of  $\theta$ :

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

with  $\beta_i = \Sigma^i(\theta_{b,c}^{(k)})^{-1} (\mathbf{y}_i - \mathbf{B}^i \alpha_{b,c}^{(k+1)})$ .

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.

Such 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.

#### D. Numerical Complexity

Conventional GPs have complexity that scales in  $\mathcal{O}(n^3)$ , making them unsuitable for very large scale problems. 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  $\Sigma^i(\theta_{b,c})$  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, the proposed model allows for the classification of any irregularly new time series and also the reconstruction of the observed time series using imputation techniques, as described in the following.

### III. CLASSIFICATION OF SATELLITE IMAGE TIME-SERIES AND MISSING VALUES

Once the parameters are estimated, it is therefore possible to classify a new pixel time series without any temporal resampling, as for training. In addition, the proposed model is also able to reconstruct a pixel time series 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:

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

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

In practice,  $\pi_c$  is estimated by its empirical counterpart  $\hat{\pi}_c = n_c/n$  where  $n_c$  is the number of samples assigned to class  $c$  in the sample  $\mathcal{S}$ . Parameters  $\alpha_{b,c}$  and  $\theta_{b,c}$  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. Imputation of missing values

Imputation of missing values 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) *Imputation when class membership is known:* Let us write  $Y_b^j(t^*)$  the unobserved value at wavelength  $b$  and time  $t^*$ . The following imputation 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 [49, p.63] and replacing the unknown quantities by their estimated counterparts yield (see Appendix A for a proof):

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

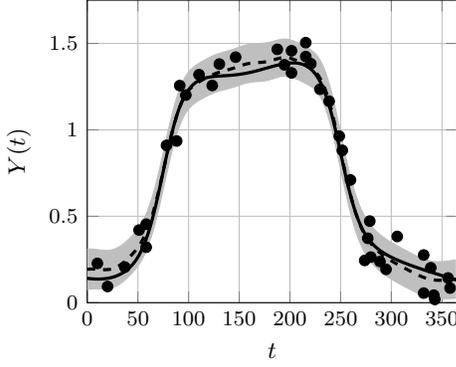


Fig. 4. Imputation of missing values. The black continuous line is one (continuous) realization of the GP from Figure 2. The black dots are the observed noisy acquisitions at several (discrete) times. The dashed line is the imputed time-series. The gray region displays a pointwise confidence region for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

and

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

where  $\mathbf{k}(t^*, t_{1:T_j} | \hat{\theta}_{b,c}) = [K(t^*, t_1 | \hat{\theta}_{b,c}), \dots, K(t^*, t_{T_j} | \hat{\theta}_{b,c})]^\top$ ,  $\mathbf{b}^* = [\varphi_1(t^*), \dots, \varphi_J(t^*)]$  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 imputed value and the variance of the imputation. They are given here for a single imputation time  $t^*$  but similar formulas can be derived for multiple imputation times  $t_1^*, \dots, t_{T^*}^*$ . Interestingly, equation (11) shows that the reconstruction is given by  $\mathbf{b}^* \hat{\alpha}_{b,c}$ , the estimated mean of the GP, corrected by a value proportional to the error made at the acquisition time of sample  $j$ . 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 sample  $j$ .

2) *Imputation when class membership is unknown:* The imputation of  $\hat{Y}_b^j(t^*)$  is done as the average of the previous imputations  $\hat{Y}_{b,c}^j(t^*)$  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 imputations at  $t^*$  respectively when the class is known (11) and when the class is unknown (13). See Appendix A for a proof.

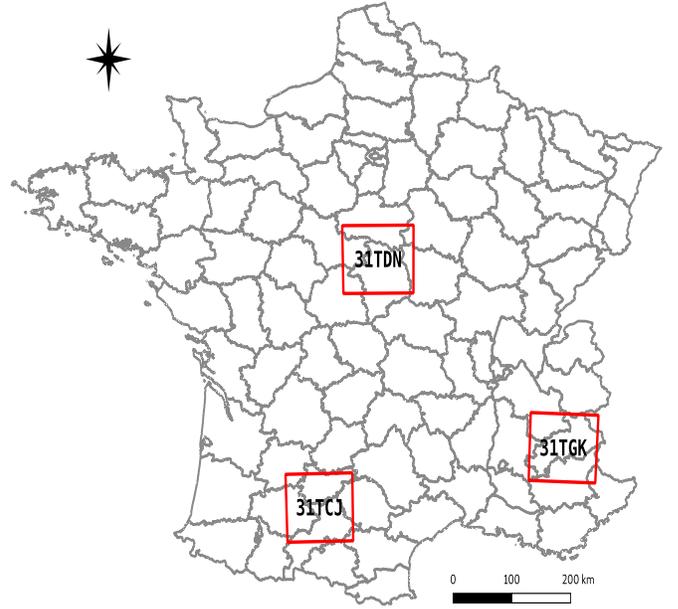


Fig. 5. Sentinel-2 tiles used. The label inside the square is the tile name.

#### 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 Landa Data Center<sup>2</sup>. All available acquisitions between January 2018 and December 2018 for the two orbits of satellites Sentinel-2A and 2B were used<sup>3</sup>. Figure 5 shows the location of the three tiles. They correspond to different climatic regions [31], 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 [50]. 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 up-sampled to 10m/pixel using the Orfeo ToolBox [51]. Figure 6 displays the distribution of the number of clear dates (dates not tagged clouds, shadows or no-data 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 [31]. Figure 10 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:

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

<sup>3</sup>Images tagged as composed of more than 90% of cloudy pixels by the MAJA processing chain are not processed by the data center.

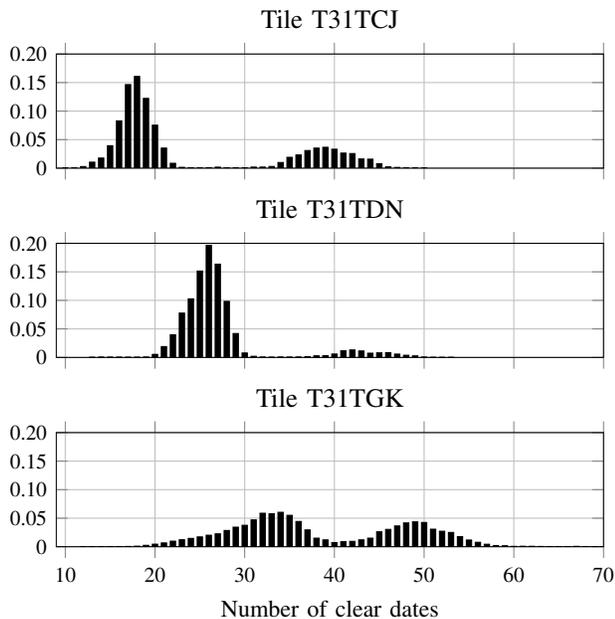


Fig. 6. Proportion of pixels as functions of the number of clear dates.

- 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.* [31]. They were extracted from freely available data source. Twenty-two land cover classes were defined, ranging from artificial areas to vegetation and water bodies. Table I shows the exhaustive list of classes.

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. After extraction, the model was trained on a total average of 178,000 pixels and was evaluated on 178,000 pixels. Finally, 10 independent train/validation sets were generated for statistical validation. Table I shows the average number of pixels per class and per tile<sup>4</sup>.

## V. EXPERIMENTAL SET-UP

The model parameters are  $\alpha_{b,c}$  and  $\theta_{b,c}$  for each class  $c$  and each feature - or wavelength -  $b$ . They are learned

<sup>4</sup>Since 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.

TABLE I  
LAND COVER CLASSES NUMBER OF SITS IN EXTRACTED DATASET FOR 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
<b>Total</b>	<b>189,868</b>	<b>178,069</b>	<b>165,663</b>

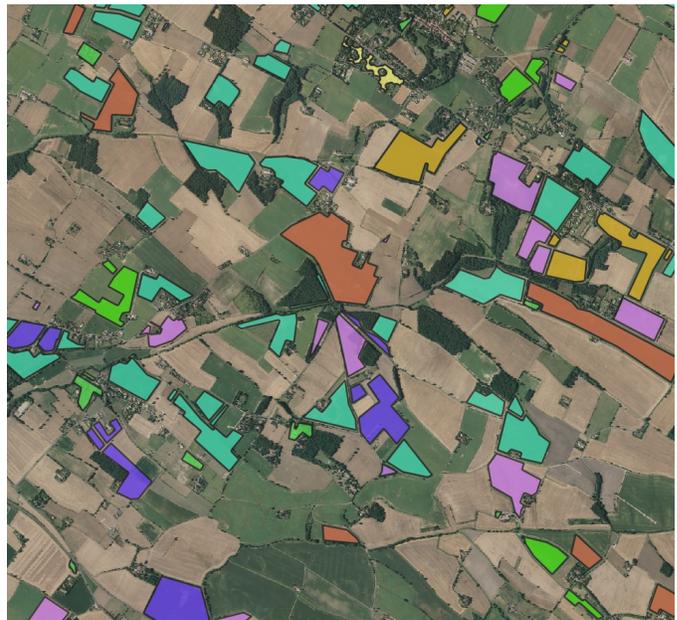


Fig. 7. Extract of the reference data. Each color represents a land cover class. The background image is an aerial orthophoto from the National Institute of Geographic and Forest Information (IGN) (<https://geoservices.ign.fr/documentation/diffusion/telechargement-donnees-libres.html>).

using the training samples. The proposed model has also some hyperparameters that are set before the learning step: The design matrix  $\mathbf{B}$  with the associated number of basis  $J$  and the family of parametric kernels for the covariance operator. These settings are common for each class and each wavelength.

### A. Functional bases

Four functional bases have been investigated: Two bases were locals (non-zero only a finite interval of  $[0, \mathcal{T}]$ ), the *Gaussian* and *BSplines* bases, while two bases were globals, 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  $J$  is equal to  $2J^* + 1$  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  $[0, \mathcal{T}]$  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  $j^{\text{th}}$  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.

### B. Covariance function

Any convex combination of positive semi-definite kernel functions is a valid covariance function [46, 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 similar between two distant dates the reflectances are.

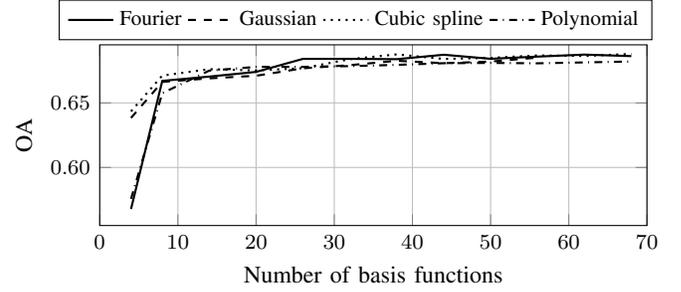


Fig. 8. Average Overall Accuracy (OA) scores of the proposed model with the 4 bases as a function of the basis size. The results are reported for the tile T31TCJ.

## VI. SUPERVISED CLASSIFICATION

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

### A. Influence of the basis functions

The influence of the basis functions and its size are investigated. Figure 8 represents the average Overall Accuracy (OA) scores averaged on the 10 independent train/validation data sets. It appears that the basis needs to be large enough to allow for a reasonable classification score, independently of the basis itself. Furthermore, the proposed model is robust w.r.t. basis selection since, when the size of the basis is high enough, the classification score does not vary significantly.

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 material). In practice, the Fourier basis is the most stable one and was used to compare our method to other classifiers.

### B. Comparison with other Classifiers

The performances of the proposed model 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) [52], and Random Forests (RF) [53] which has shown state of the art results in large scale pixel-wise classification of SITS [32].

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

Table II summarizes the results for the four methods. We can see that some significant improvements are achieved compared to the standard QDA classifier, even though MIMGP imposes independence between wavelengths. However, the proposed classifier is not as accurate as the RF and the SVM classifiers applied on the reconstructed time-series.

TABLE II  
OVERALL ACCURACY (MEAN(%)  $\pm$  STANDARD DEVIATION) FOR THE DIFFERENT TILES. THE PROPOSED MODEL IS PARAMETRIZED BY A FOURIER BASIS WITH 19 PARAMETERS FOR T31TCJ, AND 41 FOR T31TDN AND T31TGK TILES.

	QDA	SVM	RF	MIMGP
T31TCJ	39.6 $\pm$ 3.4	77.8 $\pm$ 0.9	78.9 $\pm$ 0.9	67.4 $\pm$ 1.7
T31TDN	28.5 $\pm$ 3.0	82.1 $\pm$ 0.8	84.2 $\pm$ 0.9	74.4 $\pm$ 0.8
T31TGK	48.2 $\pm$ 3.9	72.3 $\pm$ 1.7	73.4 $\pm$ 1.8	52.4 $\pm$ 3.0

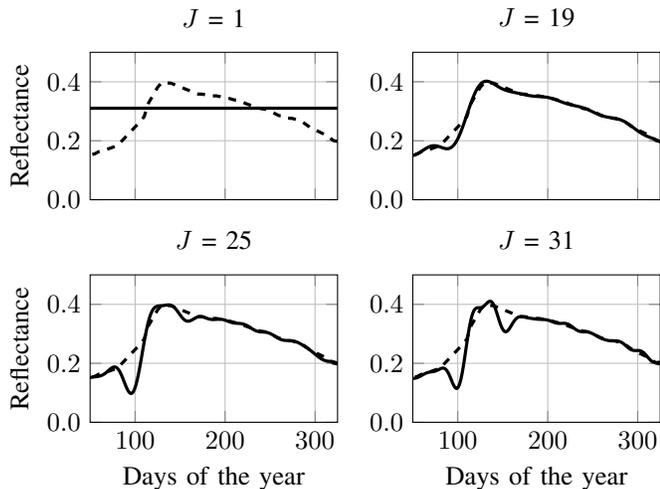


Fig. 9. Reconstructed mean near-infrared (IR) time series from broad-leaved forest class with parameters  $\alpha$  (continuous line) trained on the raw data, for different sizes of basis. The dashed line corresponds to the gap-filled data mean from QDA on the resampled grid.

Though, the computation of MIMGP does not require any temporal resampling processes and directly handles the raw time series. The computational cost given in Section II-D is only cubic with respect to the number of temporal acquisitions, which remains low for satellite image time-series. Conventional classification GPs, scaling cubically w.r.t. the number of samples, would be hardly usable in our large scale context.

## VII. TIME SERIES RECONSTRUCTION

Once the parameters are estimated during the training process, the proposed model can impute values at any time in the observed temporal interval  $\mathcal{T}$ , as described in Section III-B.

Figure 9 represents the estimated mean function in the infra-red for class “broad-leaved forest” for different sizes of the Fourier basis<sup>5</sup>. For  $J > 25$ , some oscillations appear in the reconstruction, due to a possible overfit.  $J = 19$  seems to be more appropriate, even though, in terms of classification,  $J = 25$  yields a slightly better accuracy score. In practice, a compromise should be found between a good reconstruction and a better classification score, depending on the final objective.

Figure 10 displays the infra-red bands for two pixels of the same class (summer crops). They were selected to enlighten

<sup>5</sup>Analysis of the optimized covariance operators is given in the supplementary material, section III

the robustness of the proposed method to an inaccurate cloud mask file. The MIMGP model was used with  $J = 19$  and a Fourier basis. The first pixel has an undetected cloud for the third temporal acquisition (blue point in the upper Figure 10). A clear drop of the reflectance can be seen in the gap-filled infra-red (dashed red line). The imputed reflectance (black line) does not exhibit such a drop in the reflectance. Similar comments can be done for the second pixel of the lower Figure 10 that has a undetected saturation (pixel value equal to zero).

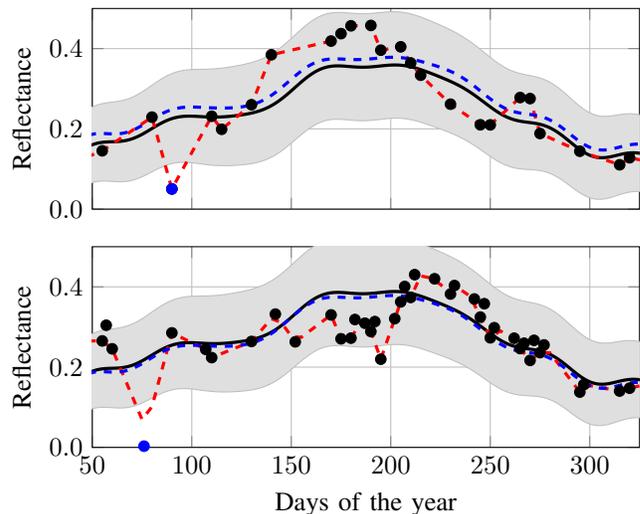


Fig. 10. Imputation of missing values. The top and bottom figures represent the near-infra red reflectance for two different pixels. Dot points are the original values from the irregularly sampled time-series. Blue points are non-detected clouds and saturation by the satellite data preprocessing chain, for the top and bottom figure, respectively. The black continuous line is the conditional imputation of the signal and the gray region corresponds to the 95% confidence interval. Each day of the year was imputed by the proposed model. The red dashed line represents the linear interpolation taking into account all the dates tagged “clear”. The blue dashed line represents the GP mean function learned from the associated ground truth class.

## VIII. CONCLUSION

This work introduces a novel approach to jointly classify and reconstruct time-series. The proposed model is able to handle irregularly sampled time-series. It 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 SITS dataset from 2018 involving a high number of temporal acquisitions. A by product of the proposed method is also to infer a confidence interval on the imputation.

The imputation 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. Rather than using a linear combination of fixed basis functions, a more complex approximating function can be used.

APPENDIX A  
IMPUTATION OF MISSING VALUES

Let us first consider the case where the class of the missing value is known to be  $c$ . In such a situation, the imputation 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

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

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 [49, p.63]), it follows that

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

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

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

Second, when the class is unknown, the imputation 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 imputed 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:

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

Besides, remarking that

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

it follows that

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

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

ACKNOWLEDGMENT

The authors would 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. n. pil. [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] M. S. Boori, K. Choudhary, R. Paringer, A. K. Sharma, A. Kupriyanov, and S. Corgne, "Monitoring crop phenology using ndvi time series from sentinel 2 satellite data," *2019 5th International Conference on Frontiers of Signal Processing (ICFSP)*, Sep 2019. [Online]. Available: <http://dx.doi.org/10.1109/ICFSP48124.2019.8938078>
- [6] A. Moeini 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>
- [7] 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>

- [8] 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>
- [9] 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>
- [10] 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>
- [11] 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>
- [12] 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>
- [13] 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>
- [14] 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>
- [15] 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>
- [16] 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>
- [17] 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/GRS.2017.2766448>
- [18] 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>
- [19] 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>
- [20] 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>
- [21] 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>
- [22] 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>
- [23] 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>
- [24] Y. Shendryk, Y. Rist, C. Ticehurst, and P. Thorburn, "Deep learning for multi-modal classification of cloud, shadow and land cover scenes in planetscope 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>
- [25] V. Syrris, P. Hasenohr, B. Delipetre, 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>
- [26] 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>
- [27] 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>
- [28] P. Benedetti, D. Ienco, R. Gaetano, K. Ose, R. G. Pensa, and S. Dupuy, "M<sup>3</sup>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>
- [29] 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>
- [30] 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>
- [31] 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.
- [32] 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>
- [33] 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.
- [34] 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>
- [35] 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>
- [36] 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.
- [37] 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.
- [38] 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>
- [39] Q. Zhang, Q. Yuan, C. Zeng, X. Li, and Y. Wei, "Missing Data Reconstruction in Remote Sensing Image With a Unified Spatial-Temporal Spectral Deep Convolutional Neural Network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 8, pp. 4274–4288, Aug. 2018.
- [40] 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>
- [41] 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>
- [42] 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>
- [43] 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>
- [44] U. B. Gwali, 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>
- [45] 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>
- [46] C. K. Williams and C. E. Rasmussen, "Gaussian Processes for Machine Learning," *the MIT Press*, vol. 2, no. 3, p. 4, 2006.
- [47] H. Trevor, T. Robert, and F. JH, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY: Springer, 2009.
- [48] 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>
- [49] M. Bilodeau and D. Brenner, *Theory of Multivariate Statistics*. Springer Science & Business Media, 2008.
- [50] 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>
- [51] O. D. Team, "Orfeo ToolBox 7.1," Mar. 2020. [Online]. Available: <https://doi.org/10.5281/zenodo.3715021>
- [52] 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.
- [53] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001.