Supplementary Materials
Supervised Classification of Irregularly Sampled Sattelite Image Times Series

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

I. Estimation of the parameters

The identifiability of the model depends on the ability to compute $\boldsymbol{\alpha}_{b,c}$. In (1) below, we remind that the optimal parameter is computed using the inverse of the matrix $\mathbf{G}_c$, defined as the sum of terms involving the design and precision matrices:

$$\boldsymbol{\alpha}_{b,c} = \mathbf{G}_c^{-1}\left[\sum_{i|Z_i=c} \mathbf{B}^{i^\top}\boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})^{-1}\mathbf{y}_{i,b}\right], \qquad \text{where } \mathbf{G}_c = \sum_{i|Z_i=c} \mathbf{B}^{i^\top}\boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})^{-1}\mathbf{B}^i. \tag{1}$$

The goal this Section is to establish a necessary and sufficient condition for $\mathbf{G}_c$ to be non-singular. Let T^c be the total number of unique temporal acquisitions for a given class c. Let us also introduce the $T_i \times T^c$ matrix $\mathbf{N}_i$ (with $T_i \leq T_c$ for all i such as $z_i = c$) and the global design $T^c \times J$ matrix $\mathbf{B}_c$ such that $\mathbf{B}^i = \mathbf{N}_i\mathbf{B}_c$. The matrix $\mathbf{N}_i$ is composed of ones and zeros to select the observed samples in signal i from the global design matrix. Then, the matrix $\mathbf{G}_c$ can be rewritten as

$$\mathbf{G}_c = \sum_{i|Z_i=c} \mathbf{B}_c^\top\mathbf{N}_i^\top\boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})^{-1}\mathbf{N}_i\mathbf{B}_c = \mathbf{B}_c^\top\left[\sum_{i|Z_i=c} \mathbf{N}_i^\top\boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})^{-1}\mathbf{N}_i\right]\mathbf{B}_c = \mathbf{B}_c^\top\mathbf{M}_c\mathbf{B}_c, \tag{2}$$

where we set $\mathbf{M}_c = \sum_{i|Z_i=c} \mathbf{N}_i^\top\boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})^{-1}\mathbf{N}_i$. In view of (2), $\mathbf{G}_c$ is non-singular if and only if $\mathbf{M}_c$ is non-singular and $\mathbf{B}_c$ is injective.

1) Let us first prove that $\mathbf{M}_c$ is non-singular. To this end, consider $\mathbf{v}^* \in \mathbb{R}^{T^c}$ such that $\mathbf{M}_c\mathbf{v}^* = \mathbf{0}$. This implies that $\mathbf{v}^{*\top}\mathbf{M}_c\mathbf{v}^* = 0$. From the definition of $\mathbf{M}_c$, and since $\boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})^{-1}$ is a positive definite matrix, $\mathbf{v}^{*\top}\mathbf{M}_c\mathbf{v}^*$ is a sum of non-negative terms. Consequently, this entails that $(\mathbf{N}_i\mathbf{v})^{*\top}\boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})^{-1}(\mathbf{N}_i\mathbf{v}^*) = 0$ for all i such that $z_i = c$. Using again the positive definiteness of $\boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})^{-1}$ yields $\mathbf{N}_i\mathbf{v}^* = \mathbf{0}$ for all i such that $z_i = c$ or equivalently $\mathbf{N}\mathbf{v}^* = \mathbf{0}$ where $\mathbf{N} := \left[\mathbf{N}_1^\top, \ldots, \mathbf{N}_{n_c}^\top\right]^\top$ and n_c is the number of samples in class c. Up to a lines permutation, $\mathbf{N}$ can be rewritten as $\mathbf{N} = \left[\mathbf{I}_{T^c}, \tilde{\mathbf{N}}\right]^\top$ where $\mathbf{I}_{T^c}$ denotes the $T^c \times T^c$ identity matrix. Then, $\mathbf{N}\mathbf{v}^* = \mathbf{0}$ implies $\mathbf{v}^* = \mathbf{0}$. Hence, the result.
2) Second, $\mathbf{B}_c$ is injective if and only if $\mathbf{B}_c$ has full rank, *i.e.* $\text{rank}(\mathbf{B}_c) \geq J$.

As a conclusion, a necessary and sufficient condition for the existence of the inverse in (1) is $\text{rank}(\mathbf{B}_c) \geq J$. This condition involves constraints both on the orthogonal basis and on the temporal acquisition points. It can be checked numerically on each experiment. Finally, let us note that, in all situations, $T^c \geq J$ is a necessary condition.

II. Numerical implementation and Convergence

The model has been implemented with Python and inherits from Scikit Learn the Gaussian Process and Kernel classes [1]. The optimization problem is solved using L-BGFS-B subroutine [2] that allows additionnal constraints on bounds (i.e., positivity or minimal/maximal values) for the covariance function parameters.

Minimization of (7) in the main document is known to converge to a local minima, which depends on initial values of the hyerparameters. Figure 1 shows the evolution of the parameters h of the RBF kernel for the covariance function used in the experiments with a Fourier basis functions of size 19 as well as the marginal log-likelihood w.r.t the number of iterations.

III. Covariance function

Figure 2 presents the RBF kernel value computed as $\exp\left\{-\frac{(t-180)^2}{2h^2}\right\}$, $t \in \{1, \ldots 365\}$ for the Infrared wavelength from three different classes. From the figure, it can be seen that the continuous urban fabric has a higher temporal correlation than Broad-Leaved forests and Water bodies.

IV. Imputation of Missing Values

This Section shows temporal reconstructions from additional classes. Figure 3 and 4 presents the reconstructed time-series with an undetected cloud for the former and for another class for the later.

This work is supported by the French National Research Agency in the framework of the Investissements dAvenir program (ANR-15-IDEX-02) and by the Centre National dEtudes 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).

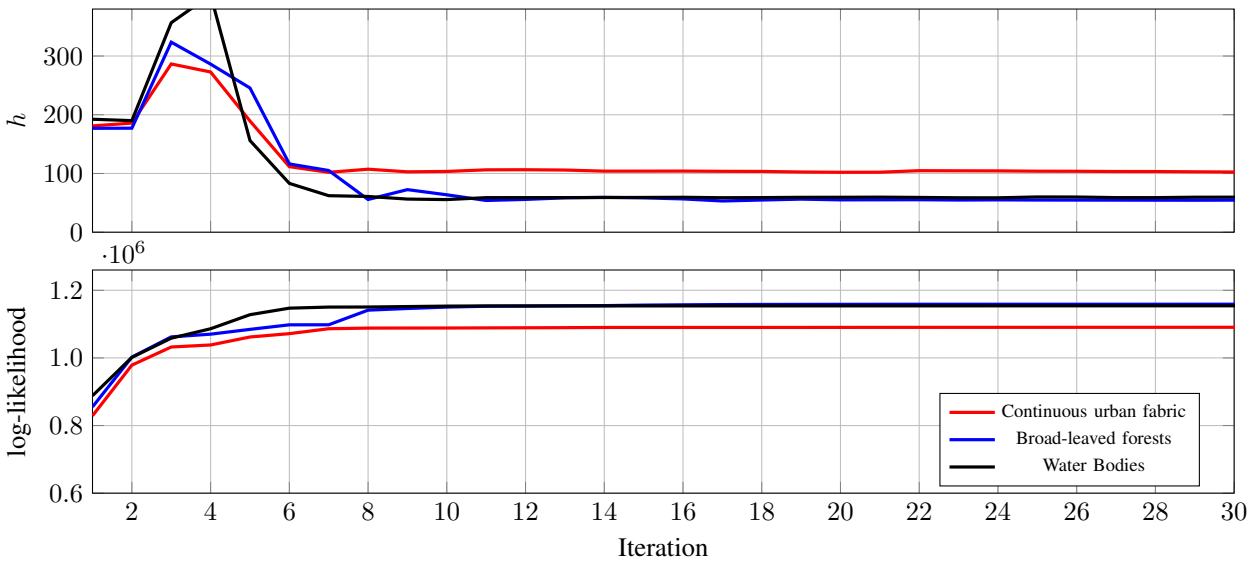

Fig. 1. Evolution of the length-scale $h \in \boldsymbol{\theta}$, given in (4) in the main document, and associated log-likelihood for the IR wavelength and three different classes.

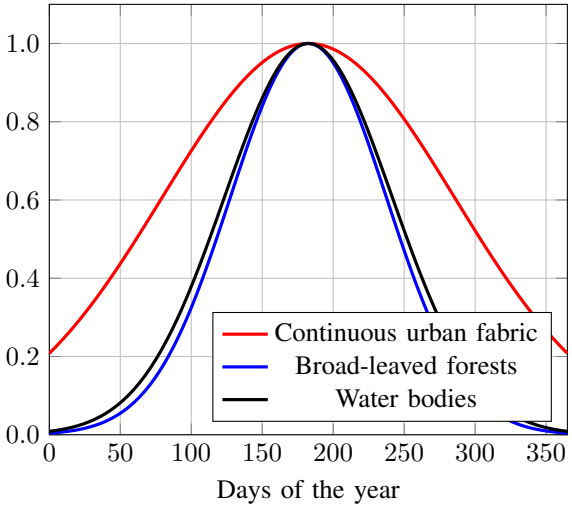

Fig. 2. Radial basis functions for the IR wavelength from artificial (continuous urban fabric), semi-natural (broad-leaved forests) and water areas.

REFERENCES

[1] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[2] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization," *ACM Transactions on Mathematical Software (TOMS)*, vol. 23, no. 4, pp. 550–560, 1997.

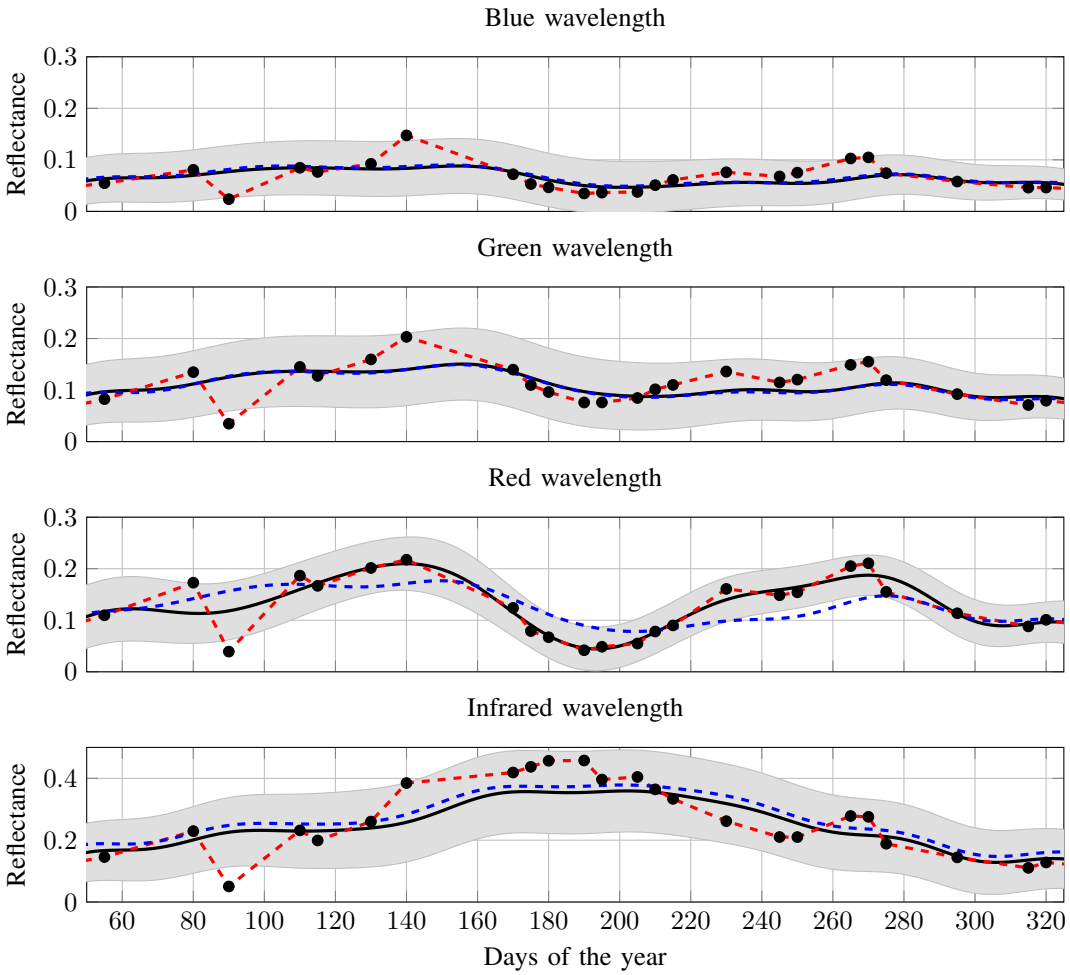

Fig. 3. Blue, Green, Red and Infrared wavelengths reconstruction for the Summer crops class from Figure 10 in the main document. The red dashed line represents the linear interpolation. The blue dashed line represents the GP mean function.

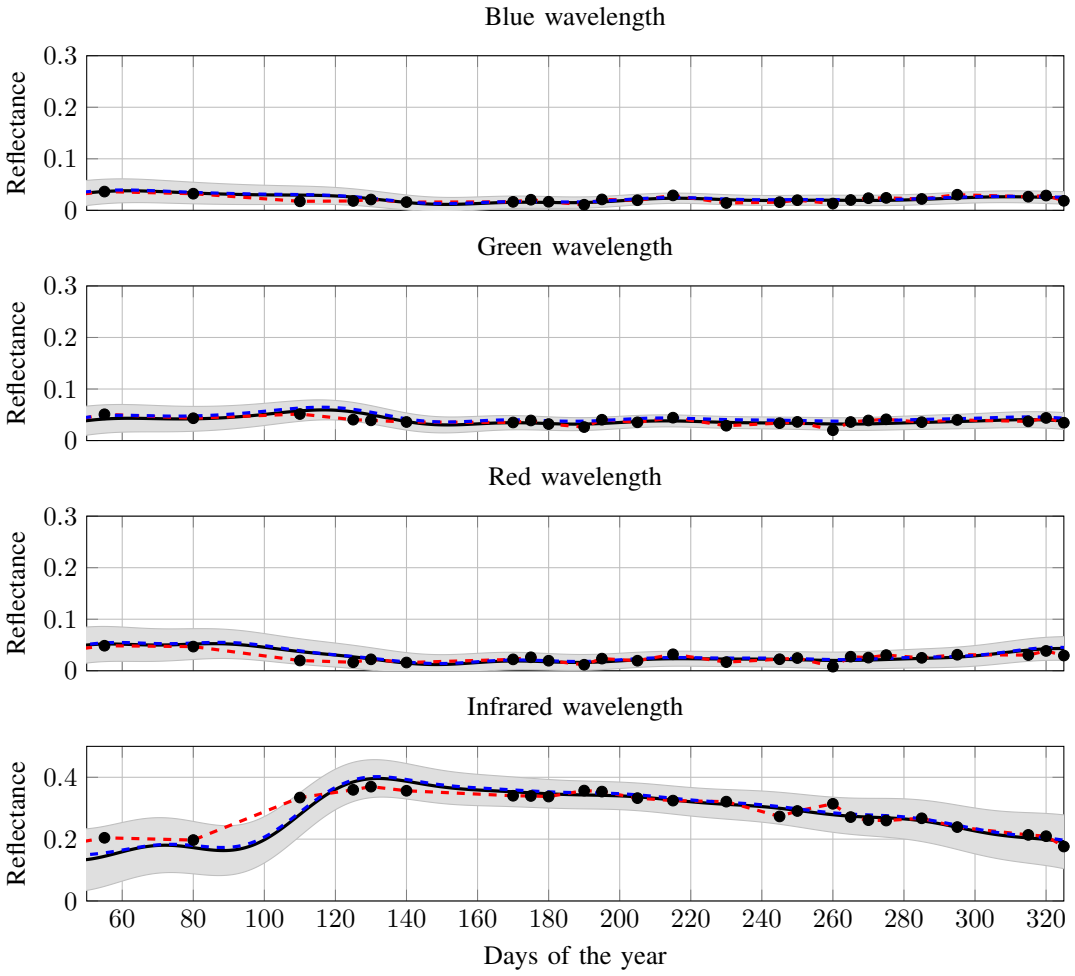

Fig. 4. Blue, Green, Red and Infrared wavelengths reconstruction for the broad-leaved forests class.