Supervised classification of multidimensional and irregularly sampled signals.

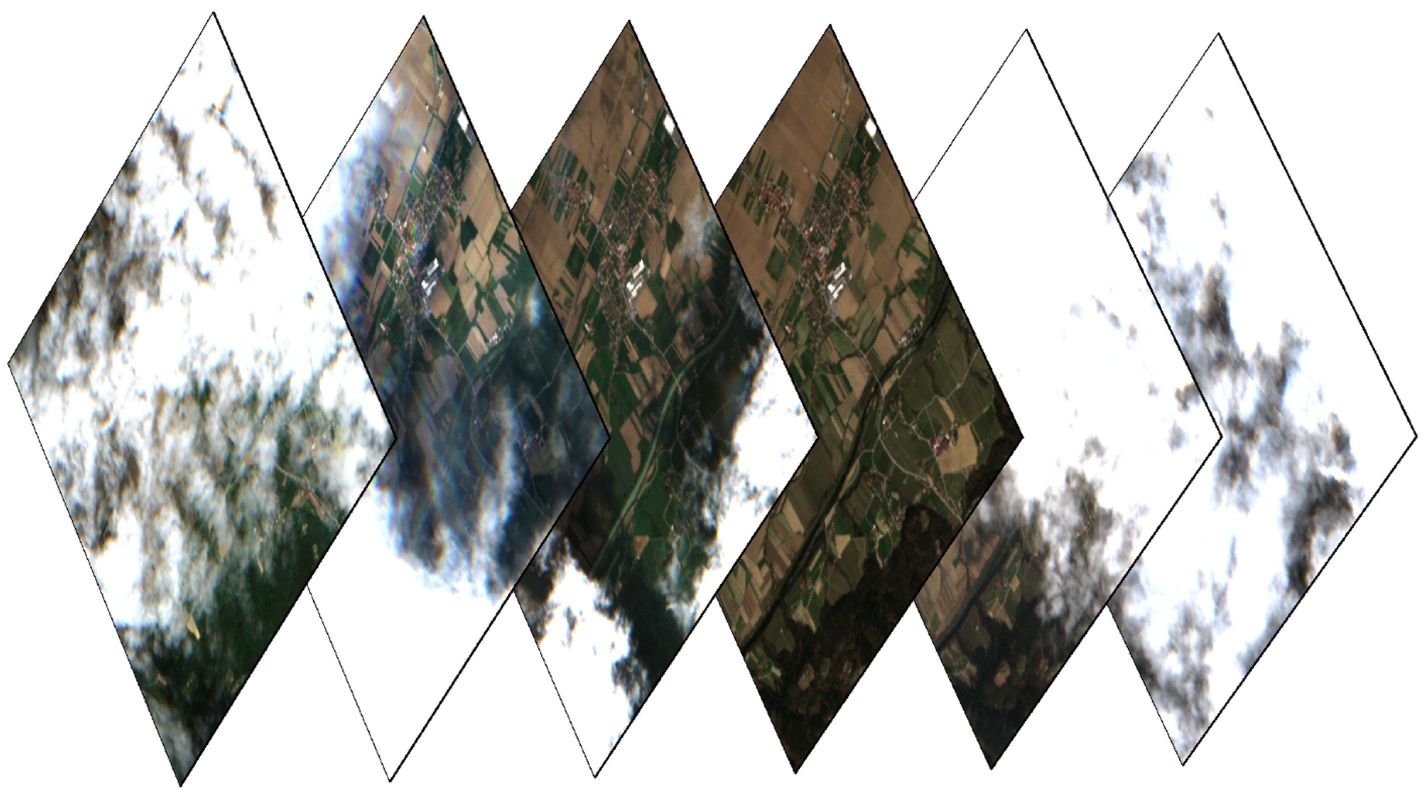
Alexandre Constantin[1], Mathieu Fauvel[2], Stéphane Girard[1] and Serge Iovleff[3]

[1]Université Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK, Grenoble, France

[2]CESBIO, Université de Toulouse, CNES/CNRS/IRD/UPS/INRA, Toulouse, France

[3]Laboratoire Paul Painlevé - Université Lille 1, CNRS, Inria, France

Introduction



Background:
Recent space missions, such as Copernicus Sentinel-2[a], provide high resolution Satellite Image Time Series (SITS) to study continental surfaces, with a very short revisit period (5 days for sentinel-2). In order to process such data, statistical models are regularly used [1, 2], which usually require a regular temporal sampling. However, for SITS, clouds and shadows (eg. figure from [3]), as well as the satellite orbite, an irregular temporal sampling is common.

Contribution:
A new statistical approach using Gaussian processes is proposed to classify irregularly sampled signals without temporal rescaling. Moreover, the model offers a theoretical framework to impute missing values such as cloudy pixels.

[a] `https://www.esa.int/Our_Activities/Observing_the_Earth/Copernicus/Sentinel-2`

Model

Gaussian Processes ($\mathcal{GP}$) model:
Let $\mathcal{S} = \{(\mathbf{y}_i, z_i)\}_{i=1}^n$ a set of multidimensional and irregularly sampled signals. A signal Y is modeled as a vector of p independent random processes $\mathcal{T} \to \mathbb{R}^p$, with $\mathcal{T} = [0, T]$. The associated label is modeled by a discrete random variable Z taking its values in $\{1, \dots, \mathcal{C}\}$. The model introduced here is based on two assumptions: 1) The coordinate processes Y_b, $b \in \{1, \dots, p\}$ of Y are independent, 2) Each process Y_b is, conditionally to $Z = c$, a Gaussian process. Then

$$Y_b(t)|Z = c \sim \mathcal{GP}(m_{b,c}(t), K_{b,c}(t,s)),$$

where $m_{b,c} : \mathcal{T} \to \mathbb{R}^p$ is a mean function, and $K_{b,c}$ a covariance kernel with hyperparameters $\boldsymbol{\theta}_{b,c}$. For example $\boldsymbol{\theta}_{b,c} = \{\gamma_{b,c}^2, h_{b,c}, \sigma_{b,c}^2\}$ with

$$K_{b,c}(t,s) = \gamma_{b,c}^2 k(t, s|h_{b,c}) + \sigma_{b,c}^2 \delta_{t,s}$$

An irregularly sampled noisy signal $\mathbf{y}_i$ is observed on T_i time stamps $\{t_1^i, \dots, t_{T_i}^i\} \in \mathcal{T}$ and its bth coordinate is represented by a vector in $\mathbb{R}^{T_i}$. We write $\mathbf{y}_{i,b} = [Y_b^i(t_1^i), \dots, Y_b^i(t_{T_i}^i)]^T$, with

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

There $\boldsymbol{\mu}_{i,b,c} = \mathbf{B}_b^i \boldsymbol{\alpha}_{b,c}$ is the sampled mean projected on a finite-dimensional space ($\mathbf{B}_b^i$ is the fixed design matrix, $\boldsymbol{\alpha}_{b,c}$ is the unknown vector of coordinates). $\boldsymbol{\Sigma}_{b,c}^i$ is the matrix kernel $K_{b,c}$ evaluations at $\{t_1^i, \dots, t_{T_i}^i\}$.

Estimation:

- $\boldsymbol{\alpha}_{b,c}$ and $\boldsymbol{\theta}_{b,c}$ are estimated by maximizing the log-likelihood,

$$-\frac{1}{2} \sum_{i|Z_i=c} \log\left|\boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})\right| + (\mathbf{y}_{i,b} - \mathbf{B}_b^i \boldsymbol{\alpha}_{b,c})^\top \boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})^{-1} (\mathbf{y}_{i,b} - \mathbf{B}_b^i \boldsymbol{\alpha}_{b,c}).$$

- $\boldsymbol{\alpha}_{b,c}$ is given by an explicit formula, while $\boldsymbol{\theta}_{b,c}$ is computed thanks to a gradient technique.

Classification and Imputation of missing values

The assigned class is given by the MAP rule from the posterior probability

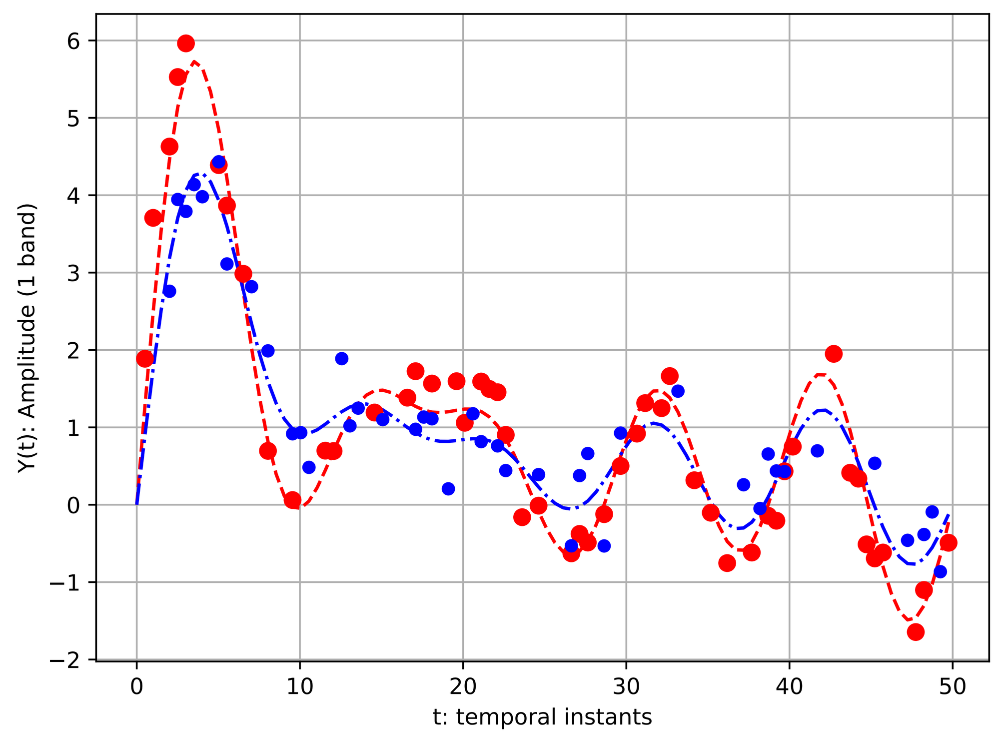
$$P(Z = c|\mathbf{y}_j) = \frac{\hat{\pi}_c \prod_{b=1}^p f_{T_j}\left(\mathbf{y}_j, \mathbf{B}_b^j \hat{\boldsymbol{\alpha}}_{b,c}, \boldsymbol{\Sigma}^j(\hat{\boldsymbol{\theta}}_{b,c})\right)}{\sum_{\ell=1}^K \hat{\pi}_\ell \prod_{b=1}^p f_{T_j}\left(\mathbf{y}_j, \mathbf{B}_b^j \hat{\boldsymbol{\alpha}}_{b,\ell}, \boldsymbol{\Sigma}^j(\hat{\boldsymbol{\theta}}_{b,\ell})\right)}.$$

When the class is known to be c, the missing value at t^* is estimated through the computation of conditional expectation.

$$\begin{cases} \hat{Y}_{b,c}^i(t^*) = B_b^i(t^*)\hat{\alpha}_{b,c} + K_{b,c}(t^*, t_{1:T_i})^\top \Sigma^i(\hat{\boldsymbol{\theta}}_{b,c})^{-1}(\mathbf{y}_{i,b} - B_b^i \hat{\alpha}_{b,c}) \\ \mathrm{var}(\hat{Y}_{b,c}^i(t^*)) = K_{b,c}(t^*, t^*) \\ \qquad - K_{b,c}(t^*, t_{1:T_i}) \Sigma^i(\hat{\boldsymbol{\theta}}_{b,c})^{-1} K_{b,c}(t_{1:T_i}, t^*) \end{cases}$$

We also generalized this imputation when the class is unknown.

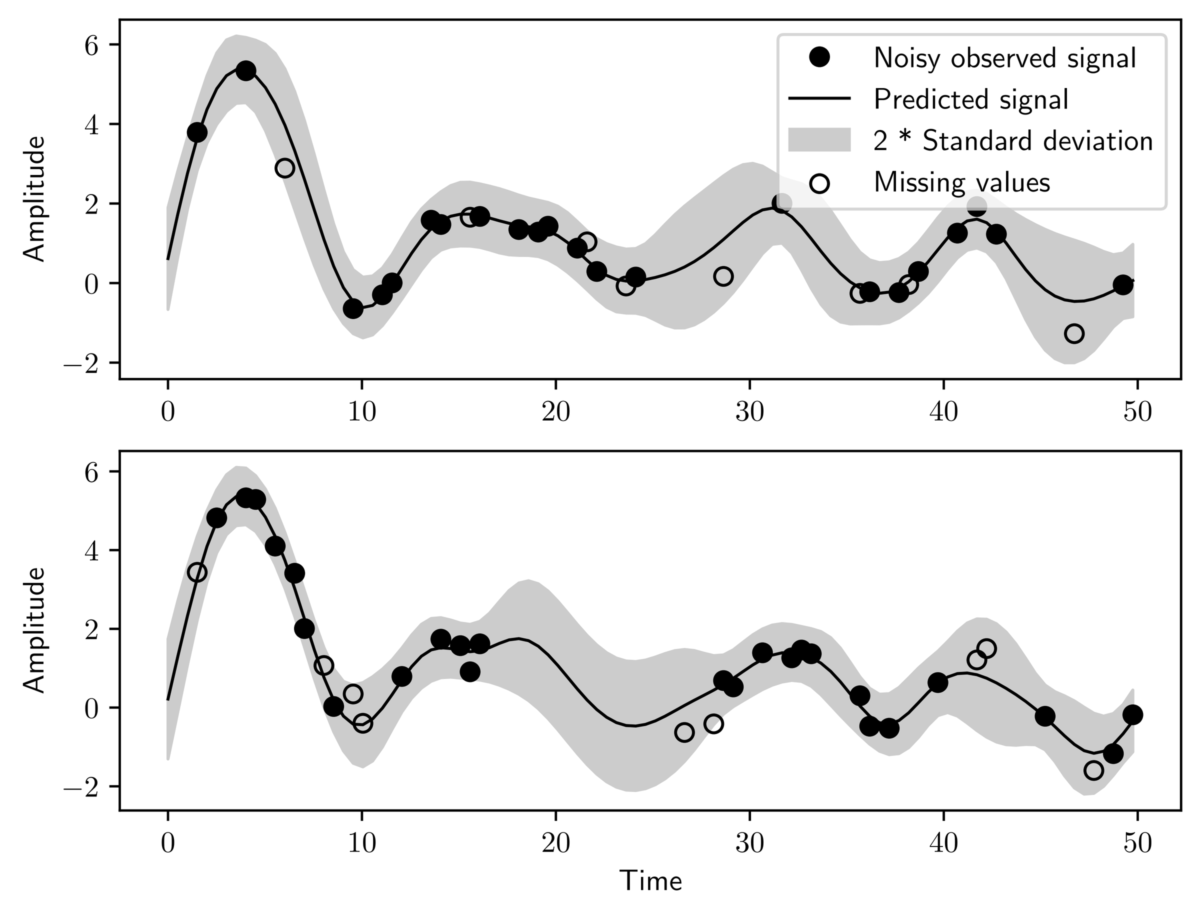
Validation (Synthetic data)



Example of two signals (dots) that belongs to two different classes

Classification rate based on average time samples

n_t	5	10	25	50	75
$\mathrm{Acc}_{\mathrm{exp}}$ (%)	52.8	52.9	74.3	93.9	94.2
$\mathrm{Acc}_{\mathrm{sin}}$ (%)	64.3	85.3	100	100	100



Imputation on two signals belonging to the same class.

Future work

We are now implementing the model for massive real data (Sentinel-2). We are also working on a new model when the bands are correlated.

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).

[1] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods.* Springer-Verlag, Berlin, Heidelberg, 1986.

[2] C. K. Williams and C. E. Rasmussen. *Gaussian processes for machine learning.* the MIT Press, 2006.

[3] Sentinel hub blog. `https://medium.com/sentinel-hub`. Accessed: 2019-03-21.