Supervised Classification of Multidimensional and Irregularly Sampled Signals

Application to Earth Observation

Alexandre Constantin[1]

Mathieu Fauvel[2], Stéphane Girard[1], Serge Iovleff[3], Yannick Tanguy[4]

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

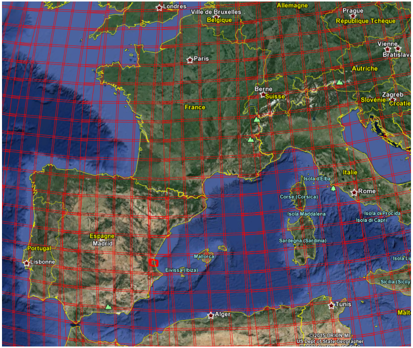
[2] CESBIO, Université de Toulouse, CNES/CNRS/IRD/UPS/INRA, Toulouse

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

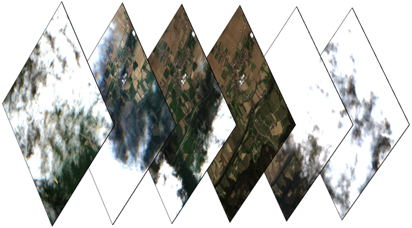
[4] CNES, DSO/SI/2A, Toulouse

Satellite Image Time Series - SITS

Sentinel 2



Sentinel 2 Tiles



Time Series[1].

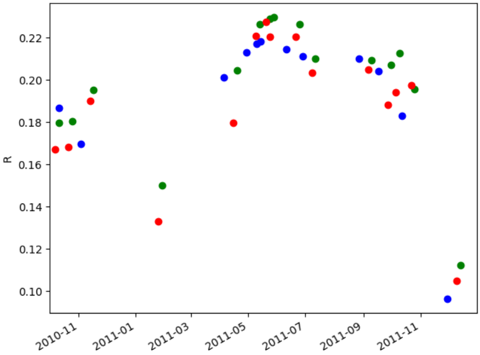
Width: 290km, altitude: 786km, ground resolution: 10m (R, G, B, near IR) - 20m (6 IR).

[1] From *Sentinel Hub Blog* 2019.

Satellite **I**mage **T**ime **S**eries - SITS

Temporal Features

- High temporal coverage
- Large amount of data
  (∼ 70TB for Sentinel-2)
- Noisy data
  (clouds/shadows/..)



3 sampled SITS (Broad-Leaved, 2010)

Supervised Classification

For each SITS we want to assign a class *c* in $\{1, \ldots, \mathcal{C}\}$ among

- Artificial Areas (urban fabrics, road surfaces, ..),
- Agricultural Areas (annual winter crops, vineyards, ..),
- Forests and Semi-natural Areas (dunes, glaciers, ..),
- Water Bodies (swamp, water, ..).

Supervised Classification

For each SITS we want to assign a class *c* in $\{1, \dots, \mathcal{C}\}$ among

- Artificial Areas (urban fabrics, road surfaces, ..),
- Agricultural Areas (annual winter crops, vineyards, ..),
- Forests and Semi-natural Areas (dunes, glaciers, ..),
- Water Bodies (swamp, water, ..).

Originality of our model

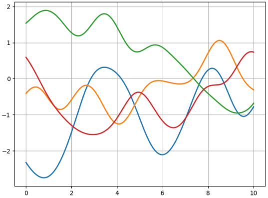
- Handle noisy data and **Irregular** sampling.
- SITS might be observed at any time.

Standard approaches use temporal resampling, such as [Inglada et al. 2017] and use SITS classifier.

Gaussian Processes

Definition

A **Gaussian process**[1] (GP) is a stochastic process such that any finite-dimensional marginal follows a multivariate Gaussian distribution.



4 samples from a GP.

We write $X \sim \mathcal{GP}(m, K)$, where

$$m(t) = \mathbb{E}[X_t]$$
$$K(t, t') = \mathbb{E}[(X_t - m(t))(X_{t'} - m(t'))]$$

X_t is real-valued and t, t' are times.

[1] [Williams and Rasmussen 2006]

Irregularly Sampled GP (IrSGP) Model

Y_b is a SITS and $b \in \{1, \dots, p\}$.
Our **main assumptions** are, conditionally to class Z (discrete)

1. It is a GP
$$Y_b | \, Z = c \sim \mathcal{GP}(m_{b,c}, K_{b,c})$$
2. $Y_1, \dots, Y_p$ are **independent.**

The ith sample Y^i associated with Z^i at times $\{t_1, \dots, t_{T_i}\}$, then

$$Y_b^i | \, Z^i = c \sim \mathcal{N}_{T_i}(\boldsymbol{\mu}^i_{b,c}, \boldsymbol{\Sigma}^i_{b,c})$$

where $\boldsymbol{\mu}^i_{b,c} = [m_{b,c}(t_1), \dots, m_{b,c}(t_{T_i})]$ and $(\boldsymbol{\Sigma}^i_{b,c})_{k,l} = K_{b,c}(t_k, t_l)$.

lrSGP Model

Estimation of the *b*th wavelength

- m is supposed to belong to a subspace of J basis functions $\{\varphi_j\}_{j\geq 1}$

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

Then $\boldsymbol{\mu}^i_{b,c} = \mathbf{B}^i \boldsymbol{\alpha}_{b,c}$ where

- $(\mathbf{B}^i)_{k,j} = \varphi_j(t^i_k)$ the design matrix;
- $\boldsymbol{\alpha}_{b,c} = [\alpha_{b,c,1}, \ldots, \alpha_{b,c,p}]^\top$ projection coefficients.

- K is chosen in a set of parametric kernels indexed by $\boldsymbol{\theta}$, we write $\boldsymbol{\Sigma}^i_{b,c} = \boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})$.

Finally

$$Y^i_b |\, Z^i = c \sim \mathcal{N}_{T_i}(\mathbf{B}^i \boldsymbol{\alpha}_{b,c}, \boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c}))$$

with parameters $\{\boldsymbol{\alpha}_{b,c}, \boldsymbol{\theta}_{b,c}\}$.

Optimizing hyperparameters

Maximizing Likelihood

Find $\{\hat{\alpha}_{b,c}, \hat{\theta}_{b,c}\}$ that maximizes the log-marginal likelihood.

$$LML(\alpha_{b,c}, \theta_{b,c}) \propto -\sum_{i|Z_i=c} \log \left| \mathbf{\Sigma}^i(\theta_{b,c}) \right| - \sum_{i|Z_i=c} (Y_b^i - \mathbf{B}^i \alpha_{b,c})^T \mathbf{\Sigma}^i(\theta_{b,c})^{-1} (Y_b^i - \mathbf{B}^i \alpha_{b,c})$$

Alternate[1] minimization of:

- α: Explicit formula in $\nabla_\alpha(LML) = 0$.
- θ: No explicit formulation: **Gradient descent** algorithm.

[1] More details can be found in [Constantin et al. 2019] (French).

Optimizing hyperparameters

Optimizing w.r.t. $\boldsymbol{\alpha}$ and $\boldsymbol{\theta}$

- $$\hat{\boldsymbol{\alpha}}_{b,c} = \left[\sum_{i | Z_i = c} (\mathbf{B}^i)^\top \boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})^{-1} \mathbf{B}^i \right]^{-1} \left[\sum_{i | Z_i = c} (\mathbf{B}^i)^\top \boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})^{-1} Y_b^i \right],$$

- $$\frac{\partial}{\partial \theta_k} LML_{b,c} = \sum_{i | Z_i = c} \operatorname{tr} \left(([\boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})]^{-1} - \boldsymbol{\beta}_{b,c}^i \boldsymbol{\beta}_{b,c}^{i\top}) \frac{\partial \boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})}{\partial \theta_k} \right),$$

  where $\boldsymbol{\beta}_{b,c}^i = \boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c})^{-1} (\mathbf{Y}_b^i - \mathbf{B}^i \hat{\boldsymbol{\alpha}}_{b,c})$.

We compute $p \times \mathcal{C}$ independent likelihood optimizations.

Classification

For a new signal $\tilde{Y}$, Bayes' rule yields to

$$P(Z = c \,|\, \tilde{Y}) \propto P(Z = c) \prod_{b=1}^{p} P(\tilde{Y}_b |\, Z = c)$$

- $P(Z = c)$ is estimated by the proportion of obs. in class c.
- $P(\tilde{Y}_b |\, Z = c)$ is the density of the Gaussian vector evaluated with parameters $\hat{\alpha}_{b,c}$ and $\hat{\theta}_{b,c}$ at times $[\tilde{t}_1, \ldots, \tilde{t}_{\tilde{T}}]$.
- **Select** the class with highest probability (MAP).

Imputation of Missing Values

Let t^* be the time where the value is missing, then $\left(Y_b^i(t^*), \mathbf{y}_{i,b}^\top \right)_{Z_i=c}^\top$ follows a Gaussian distribution

$$\begin{pmatrix} Y_b^i(t^*) \\ \mathbf{y}_{i,b} \end{pmatrix}_{Z_i=c} \sim \mathcal{N}_{T_i+1}\left(\begin{bmatrix} m_{b,c}(t^*) \\ \mathbf{B}^i \boldsymbol{\alpha}_{b,c} \end{bmatrix}, \begin{bmatrix} K^i(\boldsymbol{\theta}_{b,c}|t^*, t^*) & \mathbf{K}^i(\boldsymbol{\theta}_{b,c}|t_{1:T_i}, t^*) \\ \mathbf{K}^i(\boldsymbol{\theta}_{b,c}|t^*, t_{1:T_i}) & \boldsymbol{\Sigma}^i(\boldsymbol{\theta}_{b,c}) \end{bmatrix} \right).$$

where $\mathbf{K}^i(\boldsymbol{\theta}_{b,c}|t^*, t_{1:T_i}) = \{K(\boldsymbol{\theta}_{b,c}|t^*, t_s)\}_{\forall s \in T_i}^\top$.

Thanks to the properties of conditional expectations,

$$Y_b^i(t^*)|\mathbf{y}_{i,b}; Z_i = c \sim \mathcal{N}_1\left(\hat{Y}_{b,c}^i(t^*), \mathrm{cov}(\hat{Y}_{b,c}^i(t^*)) \right).$$

Imputation of Missing Values

Global and Conditional

When class is known to be c (Conditionally to $Z_i = c$)

$$\begin{cases} \hat{Y}^i_{b,c}(t^*) = & B^i(t^*)\hat{\alpha}_{b,c} \\ & + \mathbf{K}^i(\hat{\theta}_{b,c}|t^*, t_{1:T_i})^\top \mathbf{\Sigma}^i(\hat{\theta}_{b,c})^{-1}(\mathbf{y}_{i,b} - \mathbf{B}^i\hat{\alpha}_{b,c}) \\ \mathrm{cov}(\hat{Y}^i_{b,c}(t^*)) = & K^i(\hat{\theta}_{b,c}|t^*, t^*) \\ & - \mathbf{K}^i(\hat{\theta}_{b,c}|t^*, t_{1:T_i})\mathbf{\Sigma}^i(\hat{\theta}_{b,c})^{-1}\mathbf{K}^i(\hat{\theta}_{b,c}|t_{1:T_i}, t^*) \end{cases}$$

When class is unknown

$$\begin{cases} \hat{Y}^i_b(t^*) = \sum_{c=1}^{\mathcal{C}} P(Z_i = c|\mathbf{y}_{i,b})\hat{Y}^i_{b,c}(t^*), \\ \mathrm{cov}(\hat{Y}^i_b(t^*)) = \sum_{c=1}^{\mathcal{C}} P(Z_i = c|\mathbf{y}_{i,b})\mathrm{cov}(\hat{Y}^i_{b,c}(t^*)|\mathbf{y}_{i,b}, Z_i = c) \\ \qquad + \sum_{c=1}^{\mathcal{C}} P(Z_i = c|\mathbf{y}_{i,b})\hat{Y}^i_{b,c}(t^*)\hat{Y}^i_{b,c}(t^*)^\top - \hat{Y}^i_b(t^*)\hat{Y}^i_b(t^*)^\top. \end{cases}$$

Results

Formosat 2011

Formosat 2011 SITS: $\mathcal{C} = \mathbf{17}$ classes and $p = \mathbf{6}$ dimensions, including 4 wavelenghts and 2 spectral indices (*NDVI* and *Brightness*).

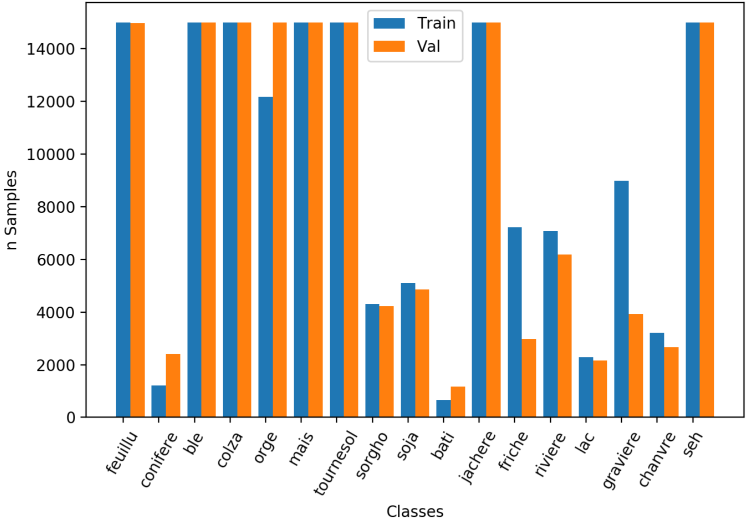
Groundtruth is divided into two spatially disjoint sets (train/val), with a maximum of 15k samples (per class).

The model is tested according to

- Estimation of the processes (m and K).
- The imputation of missing values.
- The classification scores.

Results

Formosat 2011 - Samples per Class



Model Settings

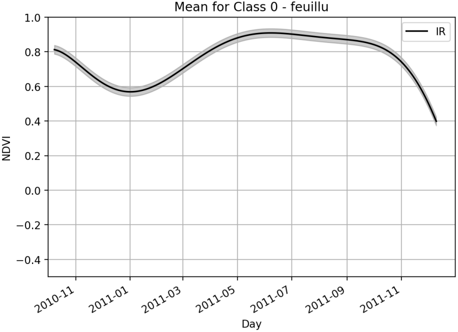
3 different basis functions:

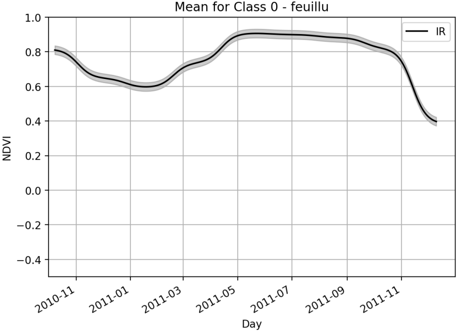
- *Fourier*: α represents the Fourier coefficients
$$m(t) = \alpha_0 + \sum_{j=1}^{J} \alpha_j \cos(2\pi * j * f_0 t) + \sum_{j=1}^{J} \alpha_j \sin(2\pi * j * f_0 t),$$
- *Exp*: equidistant Gaussian densities
$$m(t) = \sum_{j=0}^{J} \alpha_j \exp\left(-\frac{(t-t_j)^2}{d}\right),$$
- *Polynomial*: Polynomial function of t
$$m(t) = \alpha_0 + \sum_{j=1}^{J} \alpha_j t^j.$$

K is parametrized by $K(t, t' | \boldsymbol{\theta}_{b,c}) = \sigma_{b,c} \exp(-\frac{(t-t')^2}{h_{c,b}}) + \gamma_{b,c}\delta_{tt'}$.

Estimation[1] of the Process

Polynomial and exp-basis

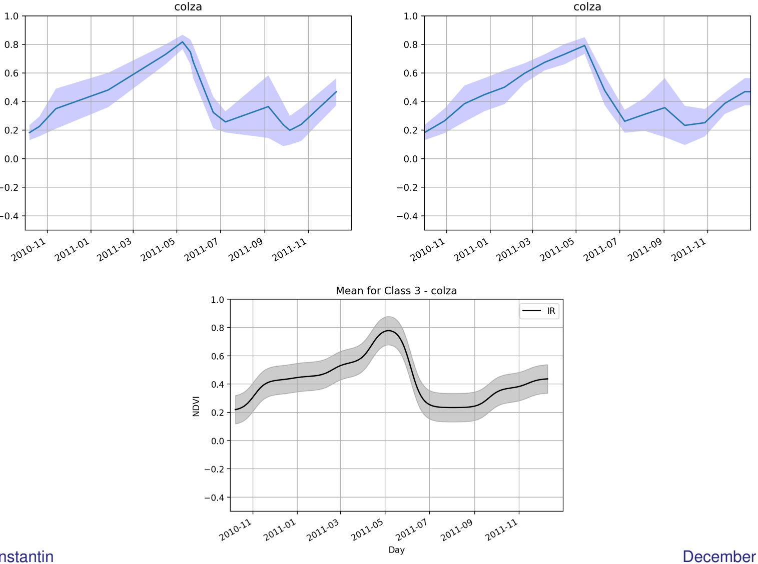




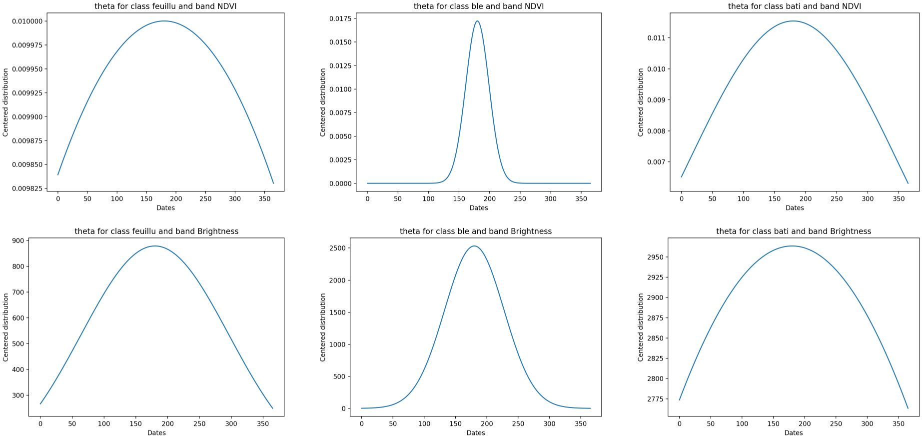
[1] The standard deviation is the square root of the diagonal from the covariance matrix.

Estimation of the Process - Colza with Exp-basis

Figure: Emp. distrib. of training set (Top) - GP Estimation (Bottom)

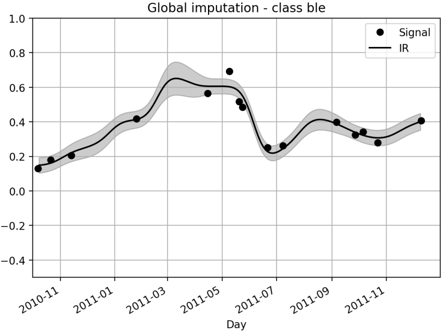


Estimation of Thetas

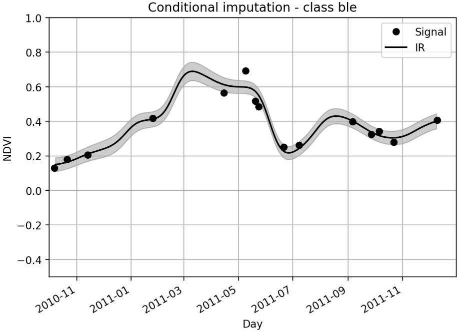


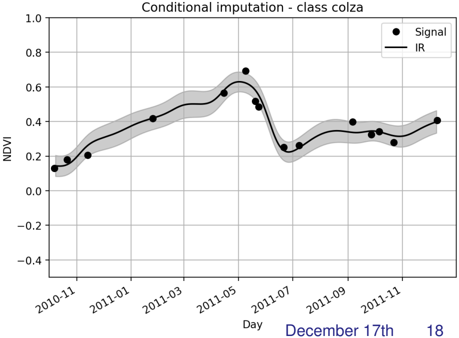
Reconstruction

Conditional imputation of missing values



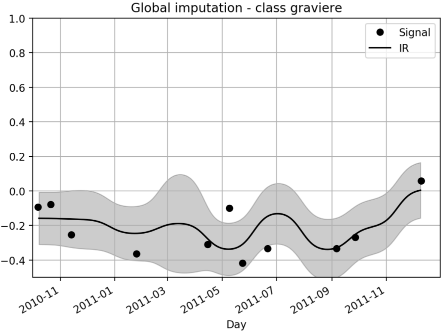
$$P(Z = \text{'wheat'}|Y^i) = 0.796$$
$$P(Z = \text{'colza'}|Y^i) = 0.204$$



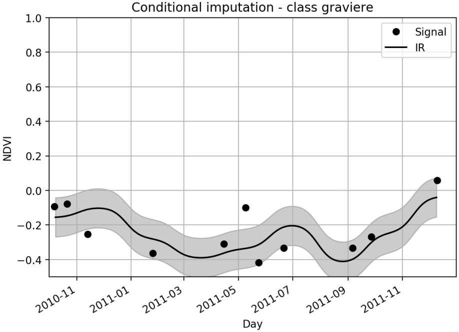


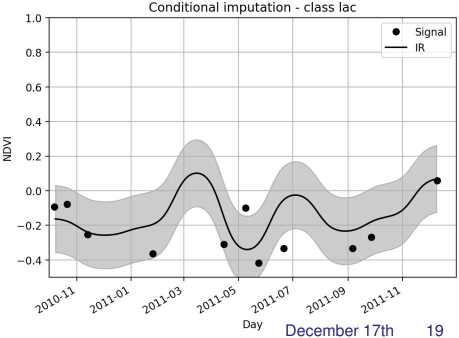
Reconstruction

Conditional imputation of missing values



$$P(Z = \text{'gravel'} | Y^i) = 0.592$$
$$P(Z = \text{'lake'} | Y^i) = 0.408$$





Classification Results

Table: Average scores among 10 independent runs.

	Gap-filled		
	QDA	RF	**lrSGP**
Acc	0.73	0.79	0.71
κ	0.70	0.77	0.68
$F_{1_{\text{mean}}}$	0.70	0.76	0.68

Settings:
QDA (Quadratic Discriminant Analysis), RF (Random Forest) maxdepth 25, lrSGP (exp basis - 11 centers).

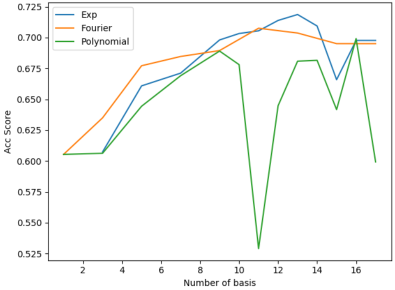
Comment:
lrSGP almost same performances as QDA and RF despite independence assumptions on the bands.

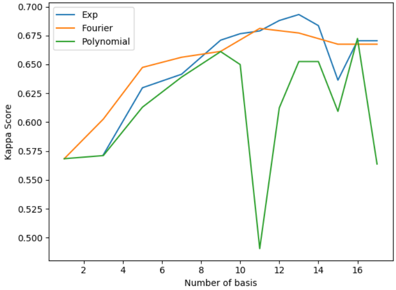
Confusion Matrix

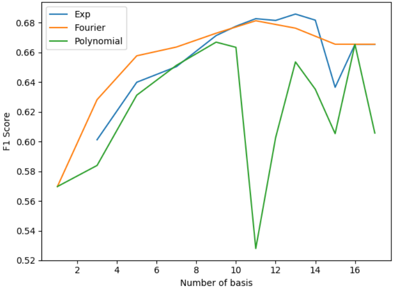
Confusion Matrix

Predicted \ Ground truth	feuillu	conifere	ble	colza	orge	mais	tournesol	sorgho	soja	bati	jachere	friche	riviere	lac	graviere	chanvre	seh
feuillu	0.89	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.38	0.00	0.00	0.00	0.00	0.00
conifere	0.02	0.94	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00
ble	0.00	0.00	0.68	0.07	0.14	0.00	0.00	0.03	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.02
colza	0.00	0.00	0.16	0.81	0.13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
orge	0.00	0.00	0.14	0.06	0.62	0.00	0.01	0.01	0.00	0.00	0.03	0.02	0.00	0.00	0.00	0.00	0.09
mais	0.00	0.00	0.00	0.00	0.00	0.91	0.04	0.04	0.08	0.00	0.00	0.00	0.00	0.00	0.00	0.11	0.00
tournesol	0.00	0.00	0.00	0.00	0.01	0.01	0.73	0.50	0.13	0.01	0.00	0.01	0.00	0.00	0.00	0.20	0.00
sorgho	0.00	0.00	0.00	0.00	0.00	0.01	0.08	0.41	0.40	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00
soja	0.00	0.00	0.00	0.00	0.00	0.06	0.10	0.00	0.37	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00
bati	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.93	0.00	0.00	0.00	0.00	0.00	0.00	0.01
jachere	0.04	0.01	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.72	0.18	0.00	0.00	0.00	0.00	0.21
friche	0.04	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.22	0.00	0.00	0.00	0.00	0.01
riviere	0.01	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.02	0.01	0.98	0.01	0.00	0.00	0.00
lac	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.01	0.58	0.08	0.00	0.00
graviere	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.42	0.92	0.00	0.00
chanvre	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.62	0.00
seh	0.00	0.00	0.02	0.05	0.07	0.00	0.01	0.01	0.00	0.01	0.17	0.15	0.00	0.00	0.00	0.00	0.63

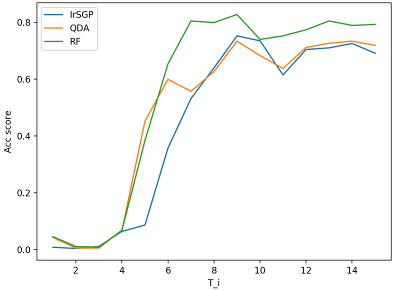
Classification - Sensitivity

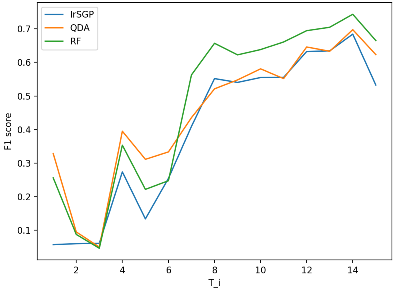


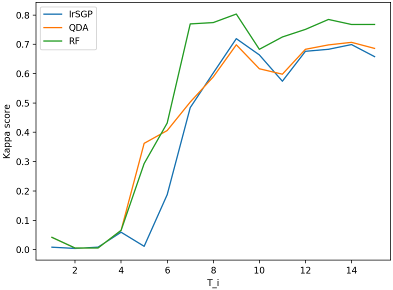




Classification - T_i available acquisitions







Conclusion

Able to handle irregularly sampled signals including

1. Modeling the classes using Gaussian processes.
2. Imputation of missing values.
3. Classify the timeseries.

Future work

1. Scale up to Sentinel-2 Dataset.
2. Consider the dependence between wavelengths.

References

Constantin, A., M. Fauvel, S. Girard, and S. Iovleff (Aug. 2019). *Classification de Signaux Multidimensionnels Irrégulièrement Échantillonnés*. GRETSI 2019 - 27e Colloque francophone de traitement du signal et des images.

Inglada, J., A. Vincent, M. Arias, B. Tardy, D. Morin, and I. Rodes (2017). "Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series". In: *Remote Sensing* 9.1. DOI: 10.3390/rs9010095.

Sentinel Hub Blog (2019). https://medium.com/sentinel-hub. Accessed: 2019-03-21.

Williams, C. K. and C. E. Rasmussen (2006). "Gaussian processes for machine learning". In: *the MIT Press*.