RÉPUBLIQUE FRANÇAISE
Liberté Égalité Fraternité

Land cover mapping with Gaussian Processes at the country scale using sparse and variational approaches

V. Bellet[1,2], M. Fauvel[2], J. Inglada[2]

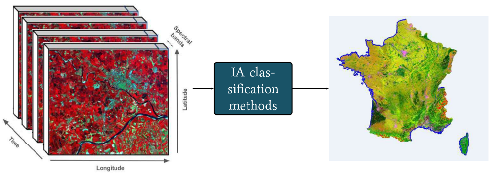
[1] Université Fédérale Toulouse Midi-Pyrénées

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

ANITI — CS GROUP — CESBIO — cnes CENTRE NATIONAL D'ÉTUDES SPATIALES

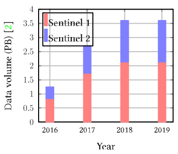
Massive Earth Observation Data

Land cover classification:

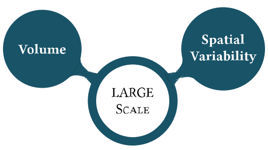


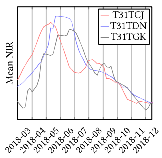
SITS — Land cover map (OSO 2019) [1]

Challenges:



Annual data volume





Winter crops

Large Scale Gaussian Process Classification

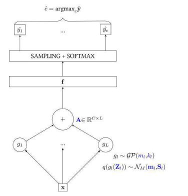
Multi-output Stochastic Variational Sparse Gaussian Process Classification [3]:

To overcome large scale issues [4], Gaussian Process $f \sim \mathcal{GP}(\mu, k)$ can be approximated using inducing points $\mathbf{Z} = \{\mathbf{z}_i\}_{i=1}^m$ with $m \ll n$ (i.e. n number of training inputs $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^n$). The complexity is reduced from $\mathcal{O}(n^3)$ to $\mathcal{O}(nm^2)$. (μ and k are modeled by some parametric functions with hyperparameters $\boldsymbol{\theta}$)

Sparse GP

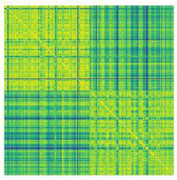
These inducing points $\mathbf{Z}$ can be learned with the hyperparameters $\boldsymbol{\theta}$ using the ELBO ε (stochastic optimization):
ε = data driven term + regularization term (Kullback-Leibler divergence)

Stochastic Variational SGP

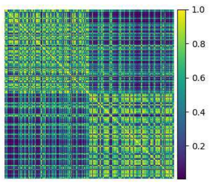


Multi-output GP Classification

Spectro-temporal and spatial covariates:



Spectro-temporal kernel



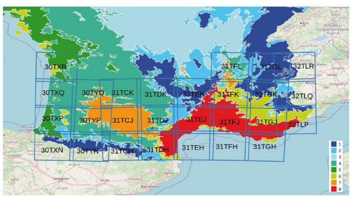
Spatio-spectro-temporal kernel

$$\mathbf{x} = \{\mathbf{x}_\phi, \mathbf{x}_{\lambda t}\}$$

1. λt-GP: $k_l(\mathbf{x}, \mathbf{x}') = k_{\lambda t}(\mathbf{x}_{\lambda t}, \mathbf{x}'_{\lambda t})$
2. $\phi\lambda t$-GPSC: $k_l(\mathbf{x}, \mathbf{x}') = \alpha_\phi^2 \times k_\phi(\mathbf{x}_\phi, \mathbf{x}'_\phi) + \alpha_{\lambda t}^2 \times k_{\lambda t}(\mathbf{x}_{\lambda t}, \mathbf{x}'_{\lambda t})$
3. $\phi\lambda t$-GPPC: $k_l(\mathbf{x}, \mathbf{x}') = k_\phi(\mathbf{x}_\phi, \mathbf{x}'_\phi) \times k_{\lambda t}(\mathbf{x}_{\lambda t}, \mathbf{x}'_{\lambda t})$

Experimental setup

Study area:



- 27 Sentinel-2 tiles
- acquisitions of level-2A from January to December 2018 linearly resampled (interval of 10 days) (total of 37 virtual dates)
- 13 spectral features λt: 10 spectral-bands (10m ground sampling distance) + 3 spectral indices (NDVI, NDWI, brightness)
- 2 spatial features ϕ: geographic coordinates (latitude, longitude) in meters in Lambert 93 projection
- 23 land cover classes ranging from artificial areas to vegetation and water bodies
- $\sim$ 5 TB

SITS and reference data

Color	Name	Area (km^2)
	Continuous urban fabric	104
	Discontinuous urban fabric	654
	Industrial and commercial units	564
	Road surfaces	62
	Rapeseed	297
	Straw cereals	564
	Protein crops	150
	Soy	470
	Sunflower	1 441
	Corn	1 030
	Rice	77
	Tubers / roots	49

Color	Name	Area (km^2)
	Grasslands	1 167
	Orchards and fruit growing	93
	Vineyards	523
	Broad-leaved forest	1 593
	Coniferous forest	4 934
	Natural grasslands	3 386
	Woody moorlands	1 713
	Natural mineral surfaces	1 680
	Beaches, dunes and sand plains	126
	Glaciers and perpetual snows	164
	Water bodies	14 567

Data generation:

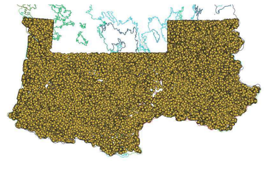
Classification dataset

- 3 distinct datasets: train, validation and test
- dataset produced for each ecoclimatic region (8 ecoclimatic regions)
- sampling repeated 11 times with different random seed
- preprocessing: feature scaling

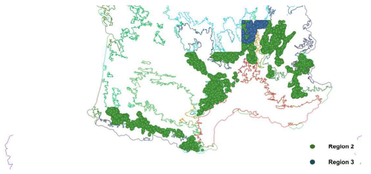
Maximum number of pixels for each class

	Region X	Global
DS-A (train-val)	4 000-1 000	32 000 -8 000
DS-B (train-val)	16 000-4 000	128 000-32 000
DS-A&DS-B (test)	10 000	80 000

Classification configuration:



Global configuration:
learn an unique model in the full area



Stratification configuration:
learn a model for each ecoclimatic region

Competitive methods:

1. Random Forest (RF): 100 trees with no maximum depth and the number of features considered for splitting at each leaf node is equal to the square root of the total number of features. (*Scikit-learn*)
2. Multilayer Perceptron (MLP): four hidden layers. The number of neurons in the first layer is the number of features divided by 2 and in the last three layers: the number of classes multiplied by 3. The activation function used is the ReLU. (*Pytorch*)

Results

Quantitative results: (first line: DS-A, second line: DS-B)

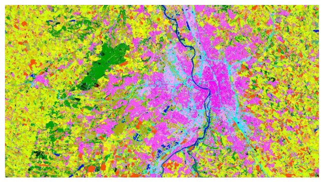
Stratification configuration

	λt-GP	$\phi\lambda t$-GPSC	$\phi\lambda t$-GPPC	λt-RF	$\phi\lambda t$-RF	λt-MLP	$\phi\lambda t$-MLP
oa	77.2 ± 0.3	78.2 ± 0.4	78.6 ± 0.3	75.6 ± 0.1	76.8 ± 0.1	75.5 ± 0.3	76.0 ± 0.2
	77.5 ± 0.3	76.1 ± 0.4	76.7 ± 0.3	75.5 ± 0.2	76.5 ± 0.1	77.1 ± 0.3	77.8 ± 0.3
fscore	78.2 ± 0.4	79.1 ± 0.4	79.6 ± 0.3	76.8 ± 0.1	78.0 ± 0.1	76.5 ± 0.4	77.1 ± 0.3
	78.5 ± 0.4	76.9 ± 0.4	77.4 ± 0.3	76.7 ± 0.2	77.7 ± 0.2	78.1 ± 0.4	78.8 ± 0.3

Global configuration

	λt-GP	$\phi\lambda t$-GPSC	$\phi\lambda t$-GPPC	λt-RF	$\phi\lambda t$-RF	λt-MLP	$\phi\lambda t$-MLP
oa	76.6 ± 0.6	79.3 ± 0.5	79.7 ± 0.6	75.3 ± 0.1	76.2 ± 0.1	77.7 ± 0.1	78.6 ± 0.1
	77.2 ± 0.2	79.2 ± 0.4	79.6 ± 0.4	75.5 ± 0.1	76.6 ± 0.1	77.4 ± 0.6	78.2 ± 0.8
fscore	77.5 ± 0.6	80.3 ± 0.5	80.7 ± 0.6	76.4 ± 0.2	77.4 ± 0.2	78.6 ± 0.1	79.5 ± 0.1
	78.3 ± 0.3	79.9 ± 0.5	80.3 ± 0.6	76.7 ± 0.2	77.8 ± 0.2	78.4 ± 0.6	79.2 ± 0.7

Qualitative results: (with stratification configuration using *iota2* chain with the model λt-GP)

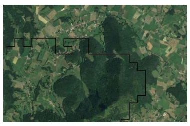


Perspectives

Dimensionality reduction:

Optimization of the inducing points $\mathbf{Z} = \{\mathbf{z}_i\}_{i=1}^m$ with $\mathbf{z}_i \in \mathbb{R}^{\phi + \lambda t}$ can be difficult for large dimension: feature reduction.

Spatial constraints in boundary zones:



Join optimization between models of 2 regions [5]:

$$\mathcal{L}_1 + \mathcal{L}_2 + \lambda R(f_1, f_2)$$

- $\mathcal{L}_1$: ELBO region 1
- $\mathcal{L}_2$: ELBO function region 2
- $f_1 = f_2$ in boundary zones

References

[1] Jordi Inglada et al. "Operational High Resolution Land Cover Map Production at the Country Scale Using Satellite Image Time Series". en. In: *Remote Sensing* 9.1 (Jan. 2017). Number: 1 Publisher: Multidisciplinary Digital Publishing Institute, p. 95. DOI: 10.3390/rs9010095. URL: https://www.mdpi.com/2072-4292/9/1/95 (visited on 09/30/2020).

[2] Pierre Soille et al. "A versatile data-intensive computing platform for information retrieval from big geospatial data". In: *Future Gener. Comput. Syst.* 81 (2018), pp. 30–40. URL: https://doi.org/10.1016/j.future.2017.11.007.

[3] Andrew G Wilson et al. "Stochastic Variational Deep Kernel Learning". In: *Advances in Neural Information Processing Systems*. Vol. 29. 2016.

[4] Gustau Camps-Valls et al. "A Survey on Gaussian Processes for Earth-Observation Data Analysis: A Comprehensive Investigation". In: *IEEE Geoscience and Remote Sensing Magazine* 4.2 (2016), pp. 58–78. DOI: 10.1109/MGRS.2015.2510084.

[5] Chiwoo Park and Daniel Apley. "Patchwork Kriging for Large-scale Gaussian Process Regression". en. In: (), p. 43.

Acknowledgements

This work is supported by the Natural Intelligence Toulouse Institute (ANITI) from Universite Federale Toulouse Midi-Pyrenees under grant agreement ANITI ANR-19-PI3A-0004 (this PhD is co-founded by CS-Group and by the Centre National d'Etudes Spatiales (CNES)). Special thanks to Benjamin Tardy for the support.
Pre-print paper: