

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

Valentine Bellet, Mathieu M. Fauvel, Jordi Inglada

#### ▶ To cite this version:

Valentine Bellet, Mathieu M. Fauvel, Jordi Inglada. Land cover mapping with Gaussian Processes at the country scale using sparse and variational approaches. Living Planet Symposium 2022, May 2022, Bonn, Germany. hal-03720074

#### HAL Id: hal-03720074 https://hal.science/hal-03720074

Submitted on 11 Jul 2022

**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.

# living planet symposium BONN



## LAND COVER MAPPING WITH GAUSSIAN PROCESSES AT THE COUNTRY SCALE USING SPARSE AND





VARIATIONAL APPROACHES

V. Bellet<sup>1,2</sup>, M. Fauvel<sup>2</sup>, J. Inglada<sup>2</sup>

<sup>1</sup> Université Fédérale Toulouse Midi-Pyrénées <sup>2</sup>CESBIO, Université de Toulouse, CNES/CNRS/INRAe/IRD/UPS



**Massive Earth Observation Data** 

Land cover classification:

#### Data generation:

Maximum number of pixels for each class



## Large Scale Gaussian Process Classification

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



#### **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

#### **Classification configuration:**



**Global configuration:** learn an unique model in the full area

#### **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: dataset A, second line: dataset B)

#### **Stratification configuration**

 $\lambda t$ -GP  $\phi \lambda t$ -GPSC  $\phi \lambda t$ -GPPC  $\lambda t$ -RF  $\phi \lambda t$ -RF  $\lambda t$ -MLP  $\phi \lambda t$ -MLP

#### **Global configuration**

 $\lambda t$ -GP  $\phi \lambda t$ -GPSC  $\phi \lambda t$ -GPPC  $\lambda t$ -RF  $\phi \lambda t$ -RF  $\lambda t$ -MLP  $\phi \lambda t$ -MLP

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



**Stratification configuration:** learn a model for each ecoclimatic region

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

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

**Stochastic Variational SGP** 

SOFTMAX  $\mathbf{A} \in \mathbb{R}^{C imes L}$  $(g_1)$  $g_l \sim \mathcal{GP}(\mu_l, k_l)$ 

**Multi-output GP Classification** 

## Sparse GP

**Spectro-temporal and spatial covariates:** 





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

Spectro-temporal kernel

Spatio-spectro-temporal kernel

## **Experimental setup**

Study area:



#### SITS and reference data

• 27 Sentinel-2 tiles

ba	$\begin{array}{c} 77.2 \pm 0.3 \\ 77.5 \pm 0.3 \end{array}$	$78.2 \pm 0.4 \\ 76.1 \pm 0.4$	$78.6 \pm 0.3 \\ 76.7 \pm 0.3$	$\begin{array}{c} 75.6 \pm 0.1 \\ 75.5 \pm 0.2 \end{array}$	$\begin{array}{c} 76.8 \pm 0.1 \\ 76.5 \pm 0.1 \end{array}$	$\begin{array}{c} 75.5 \pm 0.3 \\ 77.1 \pm 0.3 \end{array}$	$\begin{array}{c} 76.0 \pm 0.2 \\ 77.8 \pm 0.3 \end{array}$	oa	$\begin{array}{c} 76.6 \pm 0.6 \\ 77.2 \pm 0.2 \end{array}$	$\begin{array}{c} 79.3 \pm 0.5 \\ 79.2 \pm 0.4 \end{array}$	$\begin{array}{c} 79.7 \pm 0.6 \\ 79.6 \pm 0.4 \end{array}$	$\begin{array}{c} 75.3 \pm 0.1 \\ 75.5 \pm 0.1 \end{array}$	$\begin{array}{c} 76.2 \pm 0.1 \\ 76.6 \pm 0.1 \end{array}$	$\begin{array}{c} 77.7 \pm 0.1 \\ 77.4 \pm 0.6 \end{array}$	$\begin{array}{c} 78.6 \pm 0.1 \\ 78.2 \pm 0.8 \end{array}$
core	$\begin{array}{c} 78.2 \pm 0.4 \\ 78.5 \pm 0.4 \end{array}$	$79.1 \pm 0.4 \\ 76.9 \pm 0.4$	$\begin{array}{c} 79.6 \pm 0.3 \\ 77.4 \pm 0.3 \end{array}$	$\begin{array}{c} 76.8 \pm 0.1 \\ 76.7 \pm 0.2 \end{array}$	$\begin{array}{c} 78.0 \pm 0.1 \\ 77.7 \pm 0.2 \end{array}$	$\begin{array}{c} 76.5\pm0.4\\ 78.1\pm0.4\end{array}$	$\begin{array}{c} 77.1 \pm 0.3 \\ 78.8 \pm 0.3 \end{array}$	fscore	$\begin{array}{c} 77.5 \pm 0.6 \\ 78.3 \pm 0.3 \end{array}$	$\begin{array}{c} 80.3 \pm 0.5 \\ 79.9 \pm 0.5 \end{array}$	$80.7 \pm 0.6$ $80.3 \pm 0.6$	$\begin{array}{c} 76.4 \pm 0.2 \\ 76.7 \pm 0.2 \end{array}$	$\begin{array}{c} 77.4\pm0.2\\ 77.8\pm0.2\end{array}$	$\begin{array}{c} 78.6\pm0.1\\ 78.4\pm0.6\end{array}$	$\begin{array}{c} 79.5 \pm 0.1 \\ 79.2 \pm 0.7 \end{array}$

#### **Qualitative results:** (with stratification configuration using *iota2* chain with the model $\lambda t$ -GP)





## Perspectives

#### • Dimensionality reduction

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

#### • 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

Color	Name	Area (km <sup>2</sup> )
	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 0 3 0
	Rice	77
	Tubers / roots	49

• acquisitions of level-2A from January to December 2018 linearly resampled (interval of 10 days) (total of 37 virtual dates) • 13 spectral features  $\lambda t$ : 10 spectral-bands (10m ground sampling distance) + 3 spectral indices (NDVI, NDWI, brightness) • 2 spatial features  $\phi$ : geographic coordinates (latitude, longitude) in meters in Lambert 93 projection

• 23 land cover classes ranging from artificial areas to vegetation and water bodies

 $\bullet \sim 5 \text{ TB}$ 

Color	Name	Area (km <sup>2</sup> )		
	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		

## 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.

