

# PhD Position: Very large scale learning for remote sensing image time series classification

ANR MAESTRIA

Mathieu Fauvel<sup>\*</sup>

Clément Mallet<sup>◇</sup>

<sup>\*</sup> CESBIO, INRA, Toulouse, France

<sup>◇</sup> LaSTIG, IGN-Université Paris Est, Paris, France

## Context

The MAESTRIA project (Multi-modal Earth obServation Image Analysis) aims to solve the methodological challenges related to the fully automatic analysis of the massive amount of images acquired by Earth Observation platforms. MAESTRIA targets to generate land-cover and land-use descriptions at country scale at many spatial resolutions and sets of classes. The ultimate goal is to provide a continuum of spatially and semantically consistent products, that are relevant for many end-users and applications. Both public policies at local or national levels and scientific models will benefit from such kinds of products for climate modelling, urban planning, crop monitoring or impact assessment of surface changes.

The output of the MAESTRIA project will be two-fold: (i) methods that leverage current challenges in Earth Observation image analysis; (ii) a large range of precise and up-to-date land-cover maps available over very large scales from 2m to 100m. Both will be made freely available so as to stimulate research and commercial services built upon such maps.

The current PhD position integrates in and is funded by the MAESTRIA project.

## PhD Objectives

The PhD work is dedicated to the development of an efficient large scale learning method that can: (i) use large amount of multi-modal features; (ii) deal with noisy measurement and/or noisy labels; (iii) extract relevant information from the labelled classes.

Two main tasks are considered. First, the design of semi-supervised learning strategies that rely both on unlabelled and labelled samples will be addressed to alleviate the needs of large/well-balanced and accurate training set. It will be based on an unsupervised description of the data followed by a supervised step to link the statistical description of the data set to the semantic thematic needs. These two steps require to solve *very large* learning problems. Consequently, the second task of the work package concerns the development of an efficient large-scale strategy adapted to our particular field of research.

## Application

**Requirements** The candidate must have a solid background at least in one of the following items

- Statistical signal and image processing,
- Machine learning,
- Scientific programming (Python, C/C++).

A good knowledge of English is required.

**Contact** Candidates should send an e-mail to [mathieu.fauvel@inra.fr](mailto:mathieu.fauvel@inra.fr) and [clement.mallet@ign.fr](mailto:clement.mallet@ign.fr) containing:

- Full CV,
- Motivation letter,
- Contact information for 2 references, and/or recommendation letter.

Application is open until the position is fulfilled.

## Practical details

The PhD grant is about 1700 € per month, for 36 months. The recruit will be located in the CESBIO lab, in Toulouse and will interact with people involved in the project (IGN, CNES & INRA). French is not mandatory.

**CESBIO** Research at CESBIO aims to develop knowledge on continental biosphere dynamics and functioning at various temporal and spatial scales and as such participates in the specification of space missions and the processing of remotely sensed data. CESBIO is or has been PI for 2 ESA satellite missions (SMOS, the Soil Moisture and Ocean Salinity satellite, and BIOMASS, a P-band SAR system to be launched in 2020) and for the French-Israeli Venus satellite (2-day revisit, 10 m resolution, optical sensor for vegetation monitoring, launched in 2017). CESBIO has developed the IOTA2 processing chain for the operational production of land-cover maps at the national French scale. It has therefore a strong experience in upscaling learning and classification processes. CESBIO has been committed over the last two years in collecting feedback, tailoring IOTA2 outputs for various end-users, and disseminating it for several research institutes in France.

## References

- [BB12] C. Bouveyron and C. Brunet. Theoretical and practical considerations on the convergence properties of the Fisher-EM algorithm. *Journal of Multivariate Analysis*, 109:29–41, 2012.
- [BCN16] L. Bottou, F. E. Curtis, and J. Nocedal. Optimization methods for large-scale machine learning. *CoRR*, 1606.04838, 2016.
- [BG09] C. Bouveyron and S. Girard. Robust supervised classification with mixture models: Learning from data with uncertain labels. *Pattern Recognition*, 42(11):2649 – 2658, 2009.
- [BHS13] A. R. Brodtkorb, T. R. Hagen, and M. L. Sætra. Graphics processing unit (GPU) programming strategies and trends in GPU computing. *Journal of Parallel and Distributed Computing*, 73(1):4 – 13, 2013.
- [CZ05] O. Chapelle and A. Zien. Semi-supervised classification by low density separation. In *AISTATS*, pages 57–64, 2005.
- [FBG15] M. Fauvel, C. Bouveyron, and S. Girard. Parsimonious gaussian process models for the classification of hyperspectral remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 12(12):2423–2427, 2015.
- [HBRN07] L. Honglak, T. Battle, R. Raina, and A. Y. Ng. Efficient sparse coding algorithms. In *NIPS*, pages 801–808. 2007.
- [IVJCCV15] E. Izquierdo-Verdiguier, R. Jenssen, L. Gómez-Chova, and G. Camps-Valls. Spectral clustering with the probabilistic cluster kernel. *Neurocomputing*, 149:1299–1304, 2015.
- [LZCL13] Y. Li, K. Zhao, X. Chu, and J. Liu. Speeding up k-means algorithm by GPUs. *Journal of Computer and System Sciences*, 79(2):216 – 229, 2013.
- [Vid11] R. Vidal. Subspace clustering. *IEEE Signal Processing Magazine*, 28(2):52–68, 2011.