Designing complex variational distributions in variational autoencoders for biophysical variables estimations with Earth observation data

M. Fauvel (INRAe-CESBIO) RELEO

October 13, 2025

1 Context

This offer is part of the RELEO (REpresentation Learning for Earth Observation) of ANITI-2, the follow-on of the Interdisciplinary Artificial Intelligence Institute in the frame of the French ANR "AI Clusters".

Over the last ten years, Earth Observation (EO) has made large advances in terms of spatial and temporal resolutions, data availability and open policies for end-users. The increasing availability of complementary imaging sensors allows land ecosystems state variables and processes to be observed at different spatio-temporal scales. Big EO data can thus enable the design of new land monitoring systems providing critical information in order to guide climate change monitoring, mitigation and adaptation.

Variational Autoencoders (VAEs) are widely used modern generative models, providing a probabilistic framework for learning latent representations of complex data. Using such framework, the PhD work of Z. Zerah at CESBIO lab has proposed to retrieve leaf and canopy structural variables directly from Sentinel-2 imagery and a radiative transfer model (Y. Zérah. *Physics-based deep representation learning of vegetation using optical satellite image time series.* Université de Toulouse). The approach shows good agreement with in-situ data.

The standard formulation of VAEs relies on independent latent feature. While computationally convenient, this simplifying assumption often limits the model's ability to represent complex, multi-modal, or correlated latent structures. Recent research has demonstrated that richer variational families (using normalizing flows, hierarchical priors, etc ...) can significantly improve expressivity and reconstruction quality. However, defining and training these more complex distributions remains challenging, both theoretically and computationally.

2 Work-plan

The goal of this internship is to design, implement, and evaluate more expressive variational distributions for VAEs, going beyond the standard independent assumption. The main objectives are:

- Theoretical exploration of alternative variational families (e.g., truncated multivariate distribution, flow-based, ...).
- Implementation of one or several candidate architectures within the VAE framework developed at CESBIO.
- Evaluation on benchmark datasets from the CESBIO lab in terms of reconstruction quality, generative performance, and latent space structure.

- Contribution to the RELEO project.

3 Candidate background

The candidate must have a solid background in at least one of the following subjects:

- Statistical signal and image processing,
- Computer Science, Applied Mathematics, Machine learning or related field
- Remote sensing data processing.

A good knowledge of English and scientific programming in Python is required. Experience with deep learning frameworks will be appreciated.

4 Additional information

- Location: CESBIO-lab, Toulouse
- Period & Duration: From February, for 6 months
- Grant: about 700 €/month
- Contact: Send email to mathieu.fauvel@inrae.fr