

# Physics informed deep generic embeddings for multi-modal satellite image time series

## 1 Introduction

Recent advances in deep generative models have shown impressive results in terms of multi-modal image translation (from SAR to optical, for instance). If image translation between modalities is possible, that means that common factors are involved in the image formation process. These common factors can be seen as a common latent space for different image modalities. If typically, in deep learning, the variables in this latent space have to be disentangled to be interpretable, when physical models of the image formation are available (radiative transfer models, sensor models, etc.) they can be used to constrain the generation of the latent space to produce interpretable factors.

To build this latent space, an encoder-decoder architecture is usually implemented. Typically, the encoder and the decoder are deep neural networks which are trained using a pre-text task, as for instance input data reconstruction. At Cesbio, we have been working on this type of architecture where we replace the decoder by a physical model which generates a reconstruction of the data using information about the image formation (the static characteristics of the observed surfaces, but also their dynamics). In this framework, the latent variables generated by the encoder (a neural network) are the input parameters of the physical model and are interpretable by construction: they may correspond to vegetation parameters, soil moisture, surface temperature, etc.

For this kind of architecture to work, theoretical contributions have been made in terms of loss functions, choice of the prior and conditional distributions for the latent variables, etc. Further work is ongoing in terms of encoder architecture, since the different dimensions of the data (spectral, spatial and temporal) have to be taken into account in an appropriate way. Convolutional, and attention layers are used for the spatial and temporal dimensions respectively. The encoder has also to take into account the irregular temporal sampling (presence of clouds, orbit overlaps) and the different resolutions within and between sensor modalities.

The final element in the overall picture has to do with the latent factors that are not common between modalities and that are not taken into account by the physical models used as decoders. Indeed, embedding data of different modalities into a common latent space discards information that can't be retrieved by all the sensors. Furthermore, if the latent variables are bound to the input of physical models, information captured by the sensor corresponding to other phenomena, may also be removed by the encoder. In multi-modal settings, one way to deal with these issues is to grow the latent space to contain common and exclusive variables for the different modalities. These constitute the exclusive factors of each modality and, some of them, can be taken into account by the physical models cited above. The remaining factors, which are not explained by the models, correspond to the aspects of the physical reality observed by the sensors and which are not taken into account by the models. Indeed, a model is a simplification of the reality.

## **2 Focus of the PhD**

In this PhD work, we are interested in modeling and interpreting the exclusive latent variables for each sensor modality (optical and SAR) that are not used by the available physical models. The research will be focused on 2 main axes:

1. Building neural architectures for residual modeling.
2. Data-driven discovery of governing equations for the residuals.

The first axis will aim at obtaining an estimation of the residuals using statistical properties (disentanglement with respect to the physical variables) and structure in the data so that meaningful information can be separated from noise.

The second axis will exploit recent advances in sparse regression for non-linear dynamic systems identification. Using these techniques, it is possible to incorporate partial knowledge of the physics, such as symmetries, constraints, and conservation laws and produce analytic expressions which are interpretable. This can lead to the discovery of meaningful variables relevant for downstream applications related to climate studies and ecosystem monitoring.

The coupling of these 2 axes will constitute a major contribution to the efficient exploitation of novel satellite image time series for Earth observation.

### 3 Work environment

The PhD will take place at Cesbio<sup>1</sup> in Toulouse. The PhD candidate will be integrated into the *Observation Systems* team and more precisely, within the AI unit. For more information on our AI activities, the videos and slides of the *ds@cb* seminars<sup>2</sup> are a good entry point.

The team works on CNES' (the French Space Agency) high performance computing (HPC) infrastructure (250 nodes with 8000 CPU, 53 GPU) which also hosts a full mirror of all Sentinel-1 and Sentinel-2 data.

### 4 Applications

The candidate should have a strong background in several of the following subjects:

- Machine learning
- Physics
- Applied mathematics
- Scientific computing

Send Curriculum Vitae, motivation letter and recommendation letters to [jordi.inglada@cesbio.eu](mailto:jordi.inglada@cesbio.eu) before February 28 2023.

---

<sup>1</sup><https://www.cesbio.cnrs.fr/homepage/>

<sup>2</sup><https://src.koda.cnrs.fr/activites-ia-cesbio/ds-cb/-/blob/master/README.md>