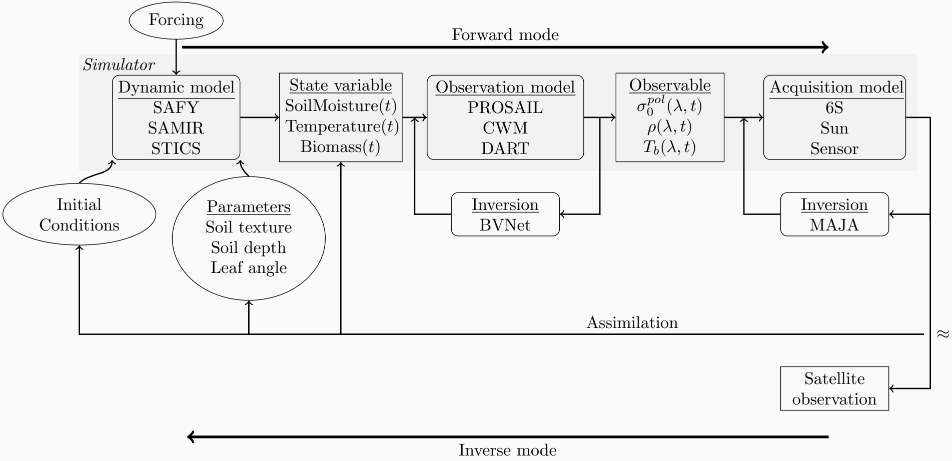
RELEO - REpresentation Learning for Earth Observation

J. Inglada, N. Dobigeon, M. Fauvel, S. Valero, T. Oberlin, J. Michel, S. Gürol

CESBIO (CNES/CNRS/INRAe/IRD/UPS), INPT - IRIT, ISAE, CERFACS, Toulouse, FRANCE

Land surface monitoring with satellites

Land surface monitoring with satellites



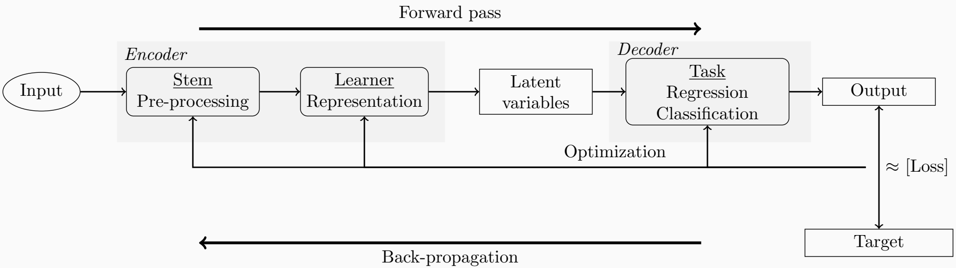
Machine learning

The interest of connectionist approaches (i.e. Deep Learning)

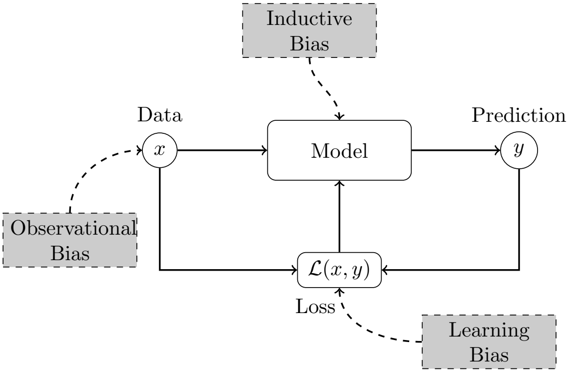
- Modular approach (architecture based on specialized building blocks)
- Straightforward scaling (parallelism, incremental learning)
- Bayesian point of view $\rightarrow$ uncertainty estimation

The interest of connectionist approaches (i.e. Deep Learning)

- Modular approach (architecture based on specialized building blocks)
- Straightforward scaling (parallelism, incremental learning)
- Bayesian point of view → uncertainty estimation



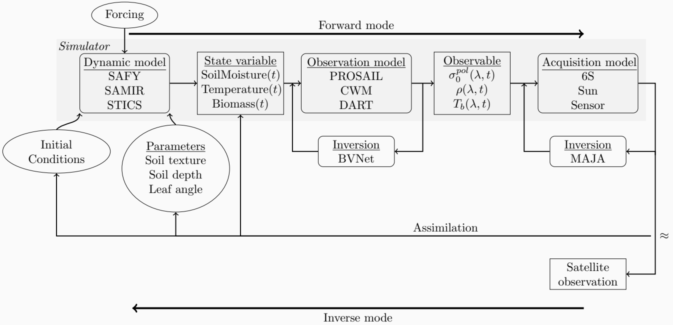
Introducing prior domain knowledge through biases

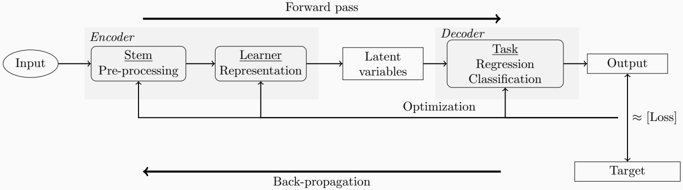


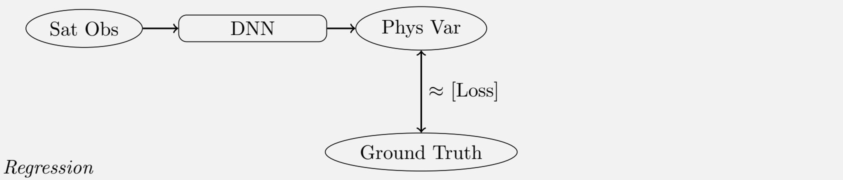
- Observational: choosing the data source for the problem.
- Inductive: structure of the model (convolution, recurrence, attention).
- Learning: MSE, NLL, Perceptual, etc.

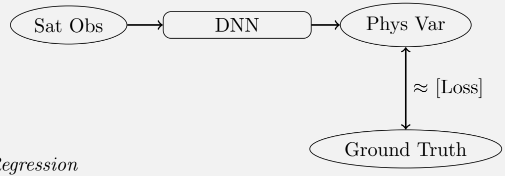
ML and assimilation

Assimilation vs back-propagation

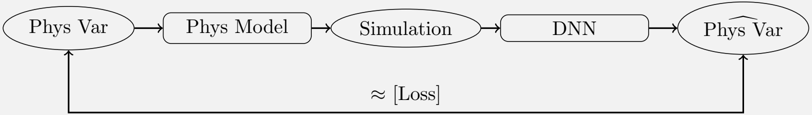





Sat Obs
DNN
Phys Var
≈ [Loss]
Ground Truth
Regression

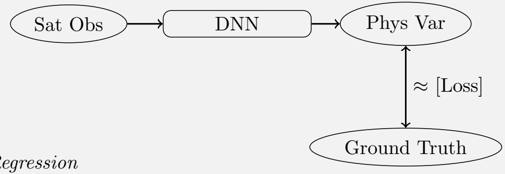


Regression

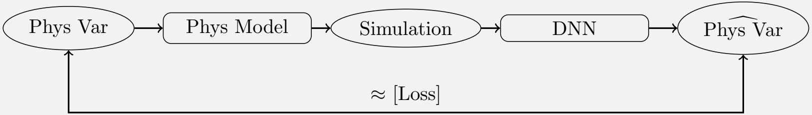


Model Inversion

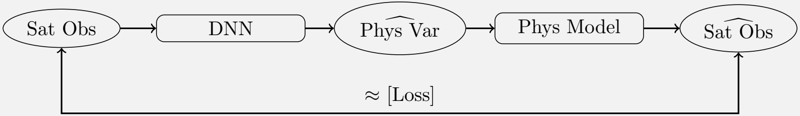
Physically aware ML in EO



Regression

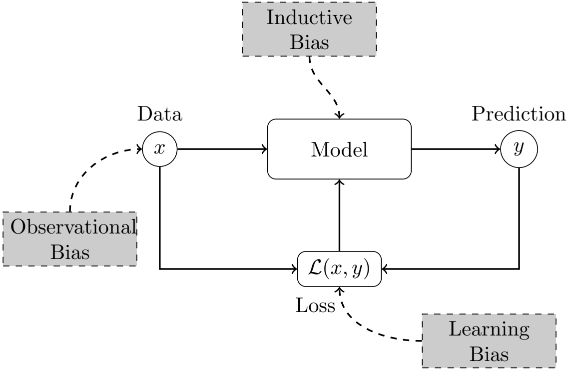


Model Inversion

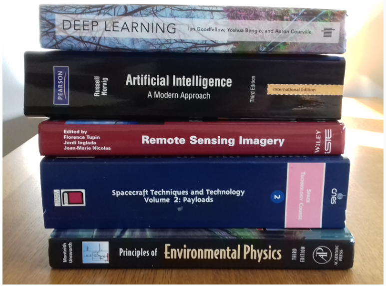


ML Assimilation

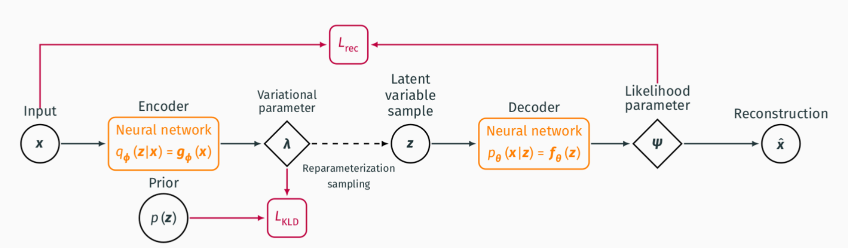
Physically aware DL - Using inductive and learning biases



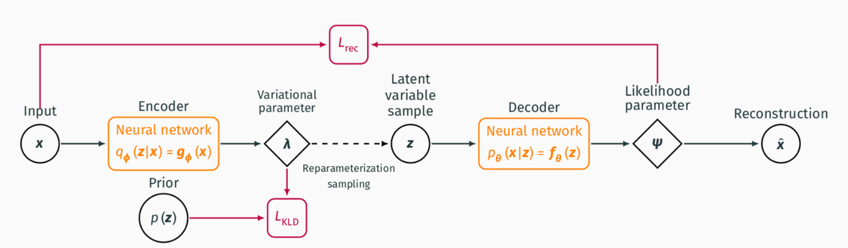
- Inductive: physical (differentiable) models in the generative process, constraints on latent variables (distributions).
- Learning: reconstruction penalties (sensor model), statistical model for outputs (noise model).



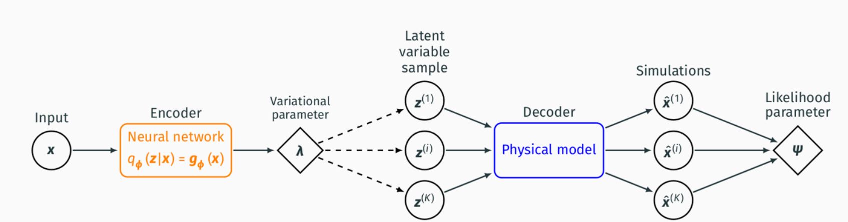
Amortized Variational Inference



Amortized Variational Inference



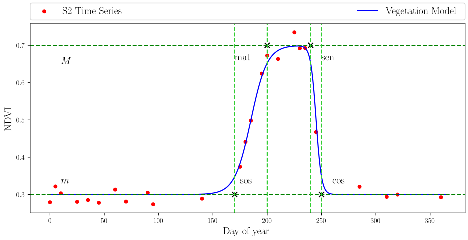
Physical model as likelihood: *hic sunt dracones…*



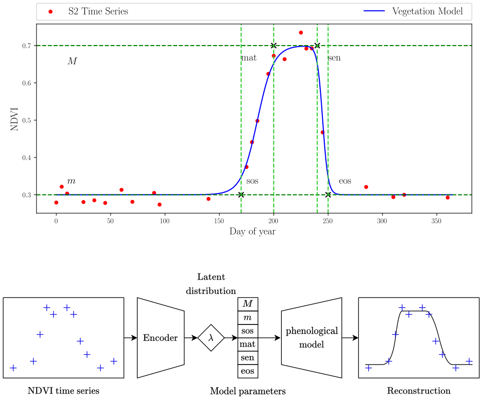
Examples

Pheno-VAE (Zérah et al. 2023)

Pixel scale vegetation phenology retrieval from satellite image time series

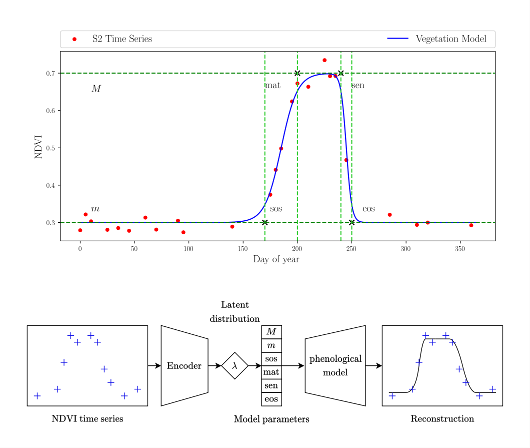


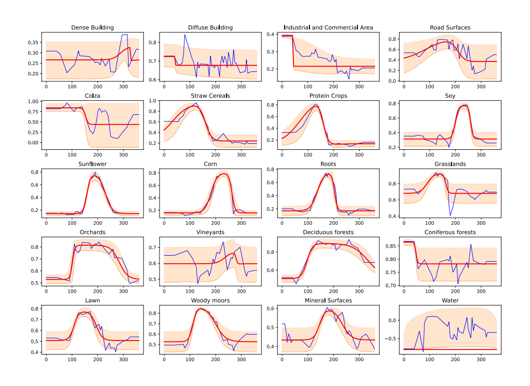
Pixel scale vegetation phenology retrieval from satellite image time series

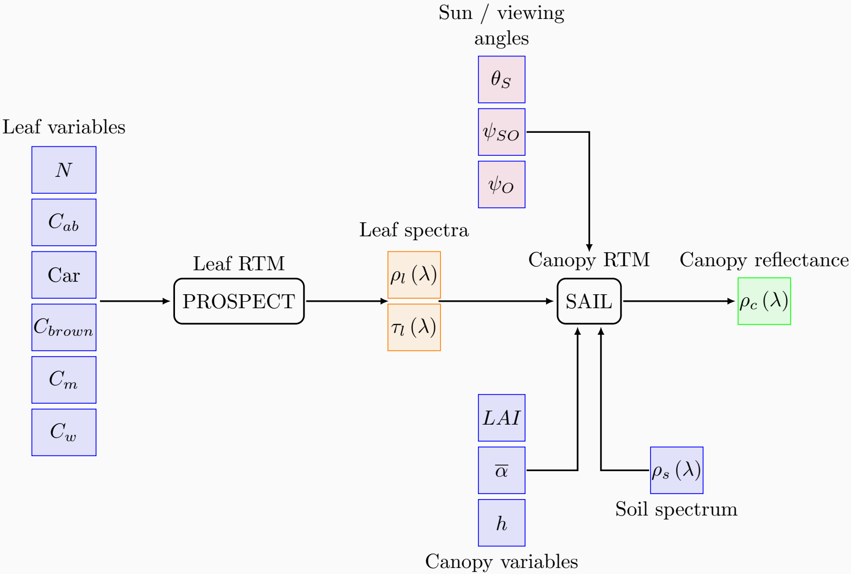


Pheno-VAE (Zérah et al. 2023)

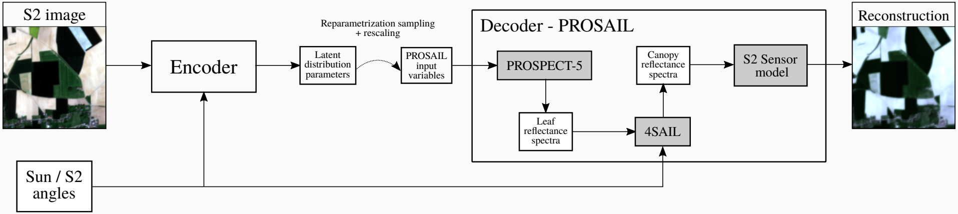
Pixel scale vegetation phenology retrieval from satellite image time series





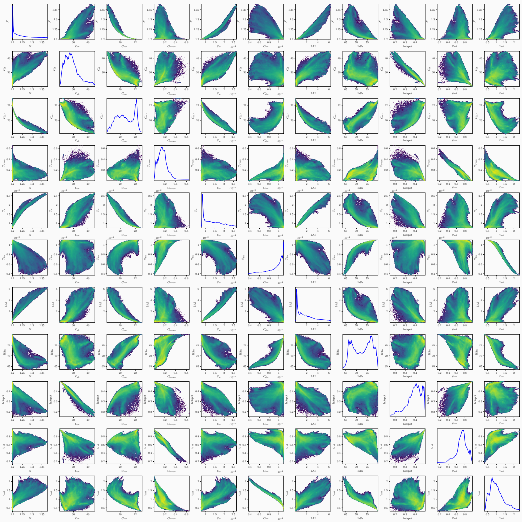

Sun / viewing angles
θ_S
ψ_{SO}
ψ_O
Leaf variables
N
C_{ab}
Car
C_{brown}
C_m
C_w
Leaf RTM
PROSPECT
Leaf spectra
$\rho_l(\lambda)$
$\tau_l(\lambda)$
Canopy RTM
SAIL
Canopy reflectance
$\rho_c(\lambda)$
LAI
$\overline{\alpha}$
h
Canopy variables
$\rho_s(\lambda)$
Soil spectrum

Pixel scale leaf and canopy parameter retrieval from satellite image time series

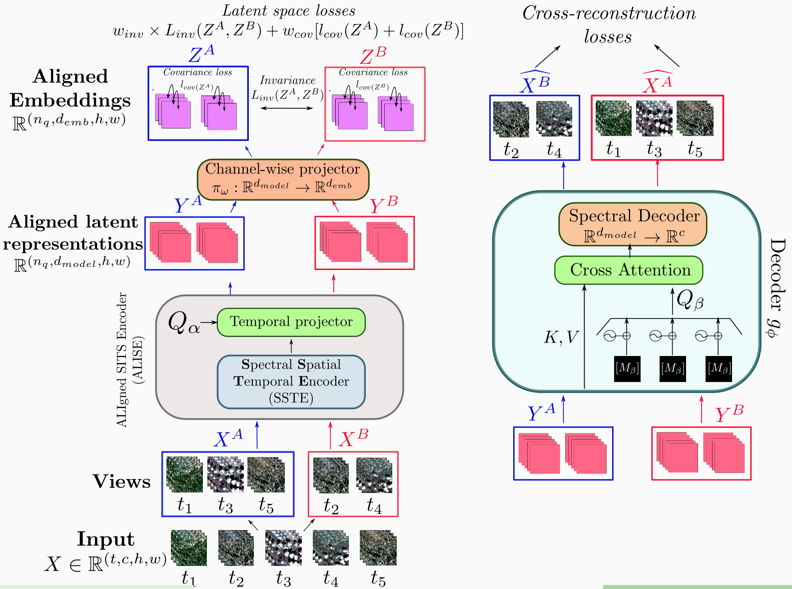


PROSAIL-VAE (Zérah et al. 2024)

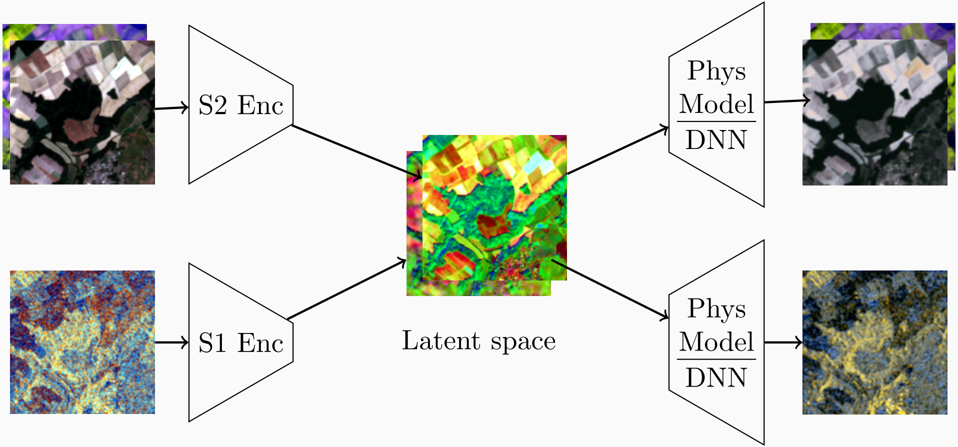
Beyond parameter retrieval



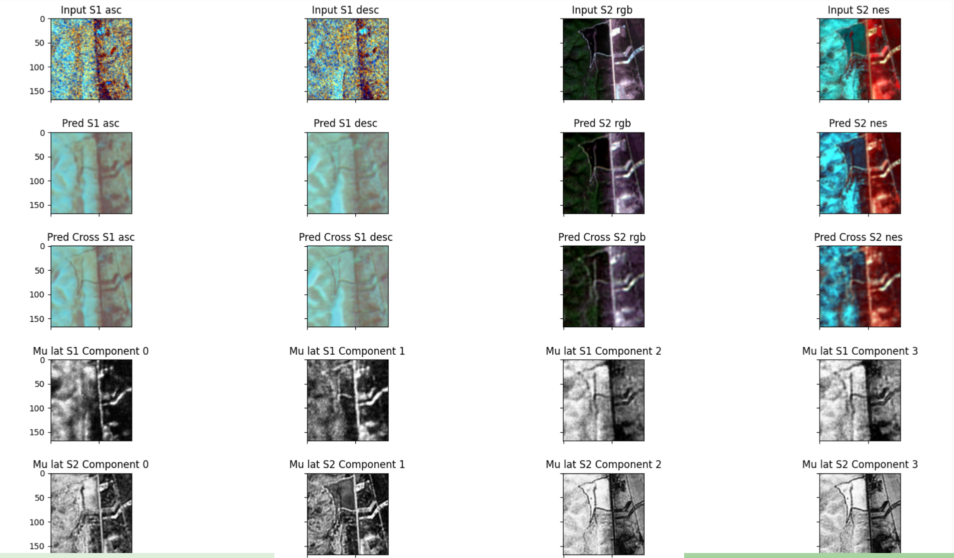
Generic multi-temporal representations: U-BARN, ALISE (Dumeur et al. 2024)



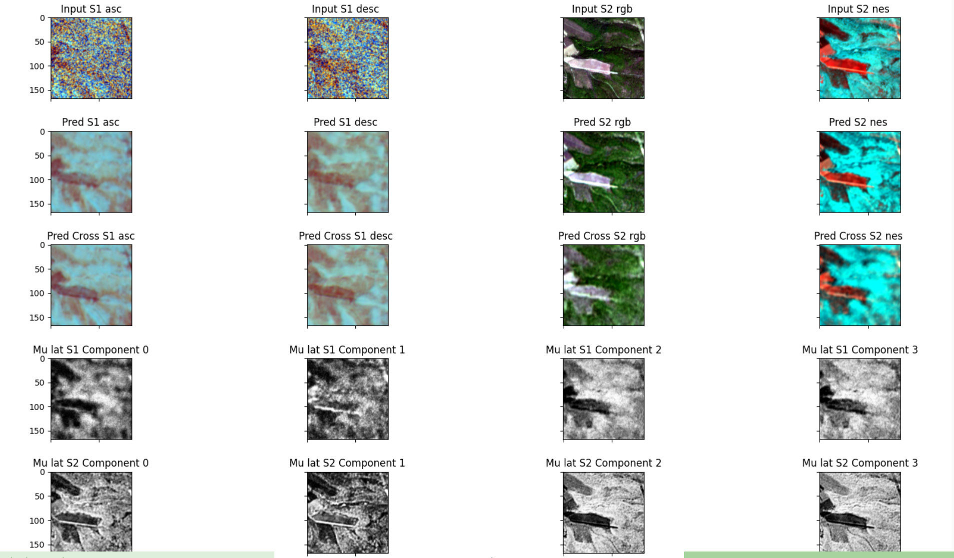
Find a physically plausible, common latent space for multi-modal observations



Find a physically plausible, common latent space for multi-modal observations

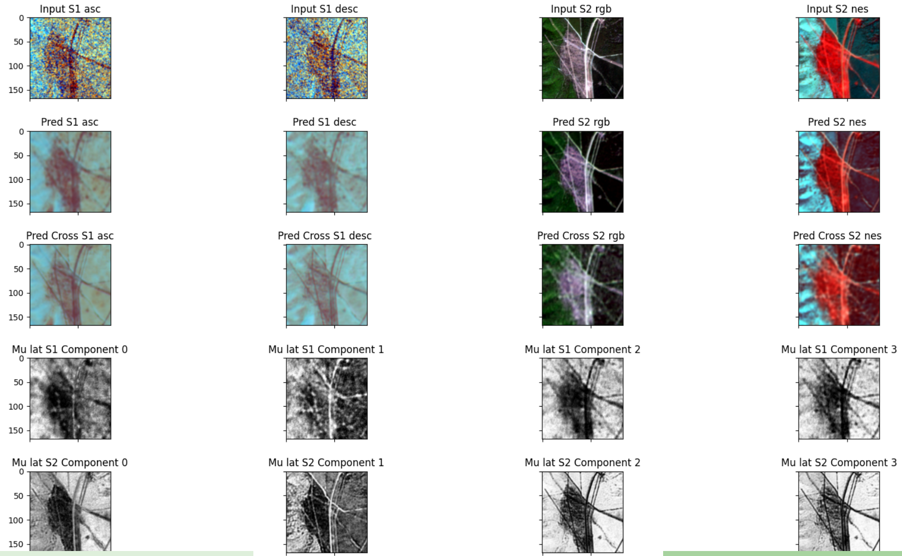


Find a physically plausible, common latent space for multi-modal observations



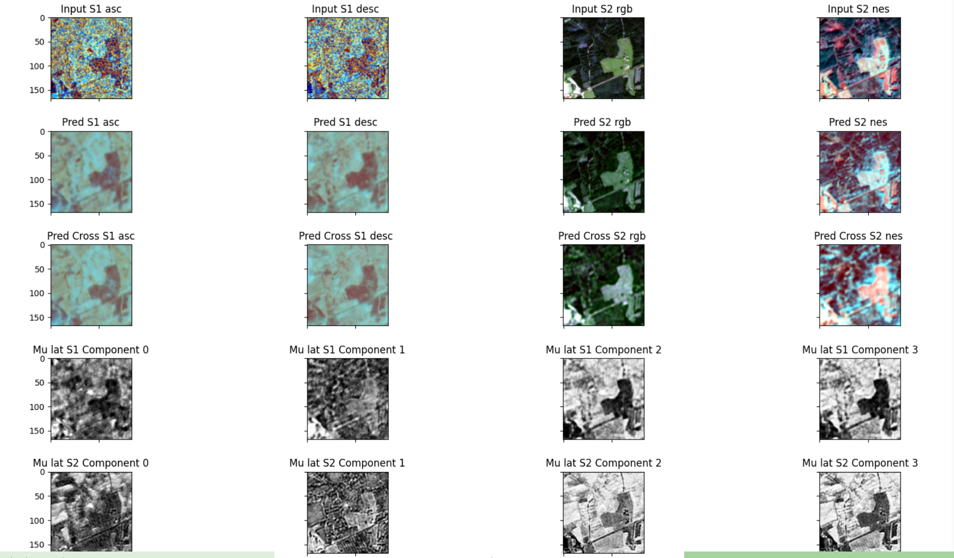
Multi-Modal Data Cube

Find a physically plausible, common latent space for multi-modal observations

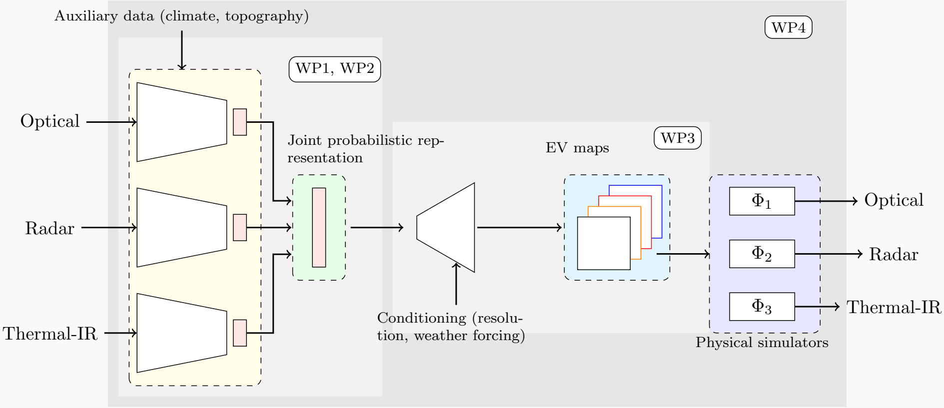


Multi-Modal Data Cube

Find a physically plausible, common latent space for multi-modal observations



Putting it all together: Large Representation Models for EO - RELEO


Auxiliary data (climate, topography)
WP4
WP1, WP2
Optical
Radar
Thermal-IR
Joint probabilistic representation
EV maps
WP3
Φ_1
Φ_2
Φ_3
Optical
Radar
Thermal-IR
Conditioning (resolution, weather forcing)
Physical simulators

Existing works

- Dumeur, I., Valero, S., & Inglada, J. (2024). Paving the way toward foundation models for irregular and unaligned satellite image time series. CoRR, http://arxiv.org/abs/2407.08448v1b.
- Dumeur, I., Valero, S., & Inglada, J. (2024). Self-supervised spatio-temporal representation learning of satellite image time series. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, (), 1–18. http://dx.doi.org/10.1109/jstars.2024.3358066
- Yoël Zérah, Valero, S., & Inglada, J. (2024). Physics-constrained deep learning for biophysical parameter retrieval from sentinel-2 images: inversion of the prosail model. Remote Sensing of Environment, 312(), 114309. http://dx.doi.org/10.1016/j.rse.2024.114309
- Zérah, Yoël, Valero, S., & Inglada, J. (2023). Physics-driven probabilistic deep learning for the inversion of physical models with application to phenological parameter retrieval from satellite times series. IEEE Transactions on Geoscience and Remote Sensing, 61(), 1–23. http://dx.doi.org/10.1109/tgrs.2023.3284992
- Bellet, V., Fauvel, M., Inglada, J., & Michel, J. (2023). End-to-end learning for land cover classification using irregular and unaligned sits by combining attention-based interpolation with sparse variational gaussian processes. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, (), 1–16. http://dx.doi.org/10.1109/jstars.2023.3343921
- Baudoux, L., Inglada, J., & Mallet, C. (2022). Multi-nomenclature, multi-resolution joint translation: an application to land-cover mapping. International Journal of Geographical Information Science, 37(2), 403–437. http://dx.doi.org/10.1080/13658816.2022.2120996
- Michel, J., Vinasco-Salinas, J., Inglada, J., & Hagolle, O. (2022). Sen2venμs, a dataset for the training of Sentinel-2 super-resolution algorithms. Data, 7(7), 96. http://dx.doi.org/10.3390/data7070096
- Inglada, J., Michel, J., & Hagolle, O. (2022). Assessment of the usefulness of spectral bands for the next generation of Sentinel-2 satellites by reconstruction of missing bands. Remote Sensing, 14(10), 2503. http://dx.doi.org/10.3390/rs14102503

Work Packages

WP1 - Representation of heterogeneous EO data

- Learning joint representations of multi-modal RS data that are heterogeneous in time, space and features

Work Packages

WP1 - Representation of heterogeneous EO data

- Learning joint representations of multi-modal RS data that are heterogeneous in time, space and features

WP2 - Modeling uncertainties with generative models

- Learning scalable generative models for representing RS data and their uncertainties while accounting for multi-modal complex distributions

Work Packages

WP1 - Representation of heterogeneous EO data

- Learning joint representations of multi-modal RS data that are heterogeneous in time, space and features

WP2 - Modeling uncertainties with generative models

- Learning scalable generative models for representing RS data and their uncertainties while accounting for multi-modal complex distributions

WP3 - Probabilistic predictions of Essential Variables

- Decoding generic probabilistic embeddings of RS observations into EVs maps with associated uncertainties at user-defined spatial and temporal scales

Work Packages

WP1 - Representation of heterogeneous EO data

- Learning joint representations of multi-modal RS data that are heterogeneous in time, space and features

WP2 - Modeling uncertainties with generative models

- Learning scalable generative models for representing RS data and their uncertainties while accounting for multi-modal complex distributions

WP3 - Probabilistic predictions of Essential Variables

- Decoding generic probabilistic embeddings of RS observations into EVs maps with associated uncertainties at user-defined spatial and temporal scales

WP4 - A multi-task self-supervised learning framework

- Designing a self-supervised multi-task framework for end-to-end training

Work Packages

WP1 - Representation of heterogeneous EO data

- Learning joint representations of multi-modal RS data that are heterogeneous in time, space and features

WP2 - Modeling uncertainties with generative models

- Learning scalable generative models for representing RS data and their uncertainties while accounting for multi-modal complex distributions

WP3 - Probabilistic predictions of Essential Variables

- Decoding generic probabilistic embeddings of RS observations into EVs maps with associated uncertainties at user-defined spatial and temporal scales

WP4 - A multi-task self-supervised learning framework

- Designing a self-supervised multi-task framework for end-to-end training

WP5 - Integration

- Integrating research from WP1-4 into an open-source scalable system to support applications in WP6

Work Packages

WP1 - Representation of heterogeneous EO data

- Learning joint representations of multi-modal RS data that are heterogeneous in time, space and features

WP2 - Modeling uncertainties with generative models

- Learning scalable generative models for representing RS data and their uncertainties while accounting for multi-modal complex distributions

WP3 - Probabilistic predictions of Essential Variables

- Decoding generic probabilistic embeddings of RS observations into EVs maps with associated uncertainties at user-defined spatial and temporal scales

WP4 - A multi-task self-supervised learning framework

- Designing a self-supervised multi-task framework for end-to-end training

WP5 - Integration

- Integrating research from WP1-4 into an open-source scalable system to support applications in WP6

WP6 - Assessment of pre-trained LEOM for RS applications

- Assessing the usefulness of the proposed framework through selected representative use-cases

When we say multi-modal...

- OK, we don't have video, audio, tweets, kitties, ...

When we say multi-modal...

- OK, we don't have video, audio, tweets, kitties, ... but we have: visible, infrared, thermal, microwave, lidar ... in 4D (x, y, t, λ)

What we have learned so far

When we say multi-modal...

- OK, we don't have video, audio, tweets, kitties, ... but we have: visible, infrared, thermal, microwave, lidar ... in 4D (x, y, t, λ)

We have plenty of open data for pre-training, but building good pre-training datasets is not easy

- Use all available spectral bands, with appropriate levels of corrections, look for temporal co-occurrences
- Add auxliary data: weather, climate, topography, etc.

When we say multi-modal...

- OK, we don't have video, audio, tweets, kitties, ... but we have: visible, infrared, thermal, microwave, lidar ... in 4D (x, y, t, λ)

We have plenty of open data for pre-training, but building good pre-training datasets is not easy

- Use all available spectral bands, with appropriate levels of corrections, look for temporal co-occurrences
- Add auxliary data: weather, climate, topography, etc.

Latent spaces are not $\mathcal{N}(0, I)$

- Physical magnitudes may be bounded, have assymetric distributions, be correlated, ...

What we have learned so far

When we say multi-modal...

- OK, we don't have video, audio, tweets, kitties, ... but we have: visible, infrared, thermal, microwave, lidar ... in 4D (x, y, t, λ)

We have plenty of open data for pre-training, but building good pre-training datasets is not easy

- Use all available spectral bands, with appropriate levels of corrections, look for temporal co-occurrences
- Add auxliary data: weather, climate, topography, etc.

Latent spaces are not $\mathcal{N}(0, I)$

- Physical magnitudes may be bounded, have assymetric distributions, be correlated, ...

We need fully differentiable physical simulators...

When we say multi-modal...

- OK, we don't have video, audio, tweets, kitties, ... but we have: visible, infrared, thermal, microwave, lidar ... in 4D (x, y, t, λ)

We have plenty of open data for pre-training, but building good pre-training datasets is not easy

- Use all available spectral bands, with appropriate levels of corrections, look for temporal co-occurrences
- Add auxliary data: weather, climate, topography, etc.

Latent spaces are not $\mathcal{N}(0, I)$

- Physical magnitudes may be bounded, have assymetric distributions, be correlated, ...

We need fully differentiable physical simulators... and all models are wrong!

What we have learned so far

When we say multi-modal...

- OK, we don't have video, audio, tweets, kitties, ... but we have: visible, infrared, thermal, microwave, lidar ... in 4D (x, y, t, λ)

We have plenty of open data for pre-training, but building good pre-training datasets is not easy

- Use all available spectral bands, with appropriate levels of corrections, look for temporal co-occurrences
- Add auxliary data: weather, climate, topography, etc.

Latent spaces are not $\mathcal{N}(0, I)$

- Physical magnitudes may be bounded, have assymetric distributions, be correlated, ...

We need fully differentiable physical simulators... and all models are wrong!

Pre-training strategies from NLP or CV may not work...

- Pretext tasks need the right amount of difficulty to be pertinent.

What we have learned so far

When we say multi-modal...

- OK, we don't have video, audio, tweets, kitties, ... but we have: visible, infrared, thermal, microwave, lidar ... in 4D (x, y, t, λ)

We have plenty of open data for pre-training, but building good pre-training datasets is not easy

- Use all available spectral bands, with appropriate levels of corrections, look for temporal co-occurrences
- Add auxliary data: weather, climate, topography, etc.

Latent spaces are not $\mathcal{N}(0, I)$

- Physical magnitudes may be bounded, have assymetric distributions, be correlated, ...

We need fully differentiable physical simulators... and all models are wrong!

Pre-training strategies from NLP or CV may not work...

- Pretext tasks need the right amount of difficulty to be pertinent.

Most published models in EO are not validated on meaningful downstream tasks

- We want to accurately map the continental biosphere at high resolution and with uncertainty estimation.
- Hallucination is not an option 😵‍💫

Acknowledgements

- Contributions from
  - Yoël Zerah's PhD
  - Iris Dumeur's PhD
  - Ekaterina Kalinicheva's post-doc
  - Kevin De Sousa's PhD

Let’s do science FAST with LLMs!

- LLMs can summarize papers
- LLMs can review papers
- LLMs can write papers
- LLMs can suggest scientific questions

Let's do science FAST with LLMs!

- LLMs can summarize papers
- LLMs can review papers
- LLMs can write papers
- LLMs can suggest scientific questions

Let’s do science FAST with LLMs!

- LLMs can summarize papers
- LLMs can review papers
- LLMs can write papers
- LLMs can suggest scientific questions

If a Tree Falls in the Forest, and There’s No One Around to Hear It, Does It Make a Sound?