Earth Sciences Department

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

ECMWF | 50
Workshop on ancillary data for land surface and Earth system modelling
Bonn, Germany | 9-10 April 2025

Reconstruction of high-resolution historical Land Cover (LC) & Leaf Area Index (LAI) with Machine Learning

Amirpasha Mozaffari[1], Enric Millán Iglesias[1,2], Amanda Duarte[1], Vinayak Huggannavar[1], Stefano Materia[1], Lina Teckentrup[1], Iria Ayan[1], Etienne Tourigny[1], Markus Donat[1]

amirpasha.mozaffari@bsc.es
@amirpasha.bsky.social

1 Barcelona Supercomputing Center (BSC), Barcelona, Spain
2 Polytechnic University of Catalonia (UPC), Barcelona, Spain

Context

CERISE

The scope of CERISE (2023 - 2026) is to enhance the quality of the Copernicus Climate Change Service (C3S) reanalysis and seasonal forecast portfolio, with a focus on land-atmosphere coupling, and contributes to :

- Next generation of ECMWF reanalysis (ERA6 — Land)
- Seasonal Forecast (SEAS6)

Why dynamic LC & LAI?

- **Improved representation of land-atmosphere interactions:** ensures more realistic exchanges of water, energy, and carbon between land and atmosphere
- **Enhances forecasts and projections:** improves near-surface weather and climate predictions for better risk management, adaptation, and mitigation

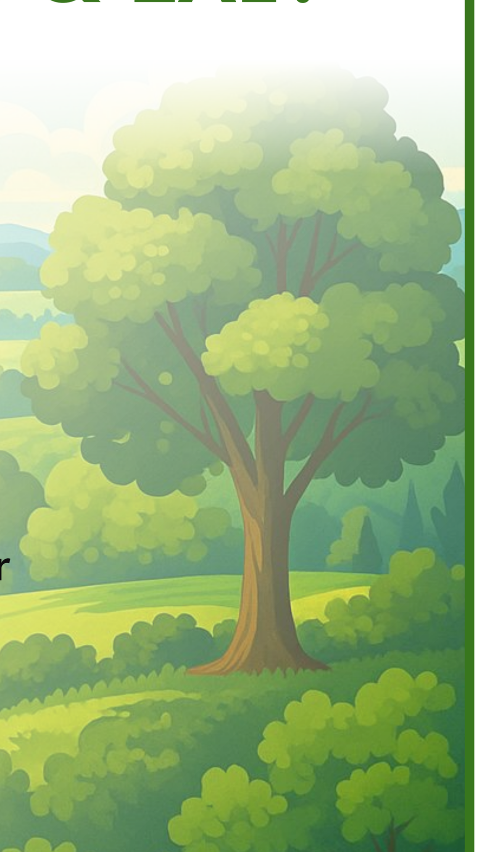


Figure is generated by ChatGPT

Datasets

- **ESA CCI Land Cover (LC)** yearly maps 300m resolution from 1992 - 2020 that includes 38 classes of LC in discrete values.
- **CONFESS Leaf Area Index (LAI)** 5-days maps at 1km resolution from 1980 - 2020 that has continuous values between (0 - 7).
- **LUH2h Land (LU)**Use yearly maps 0.25° x 0.25° resolution from 950 - 2014 that has fractional values for 8 classes in continuous values (0 -1).
- **HILDA+ LU** yearly maps at 1km resolution from 1899 - 2020 that has 7 classes in discrete values (0 -77)

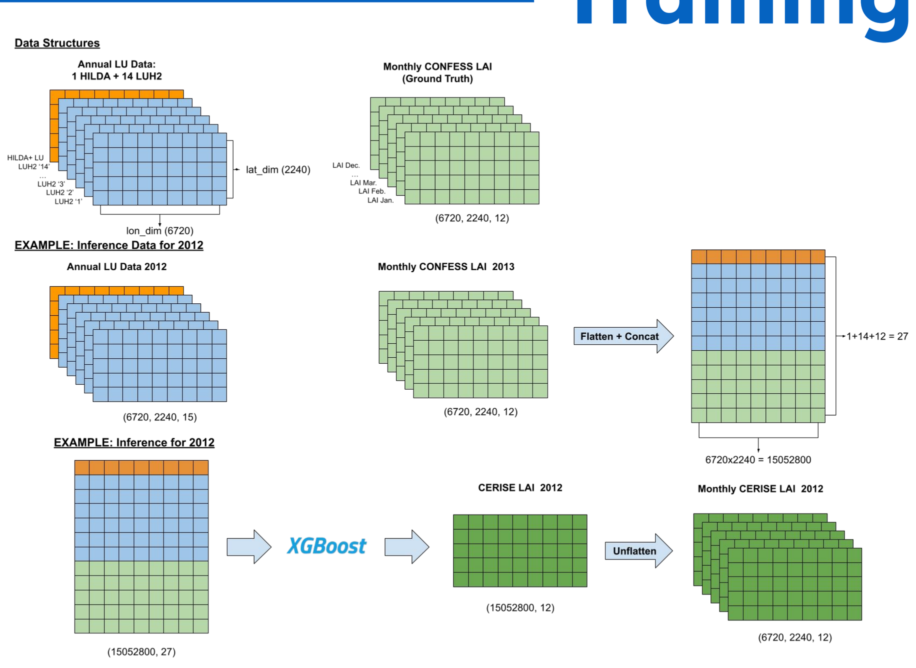
Pipeline

1. Re-gridding and downscaling to target grid and resolution
2. Split the global model to independent regional models
3. Training and inference on GPU
4. Post processing to stitch the models to generate global output

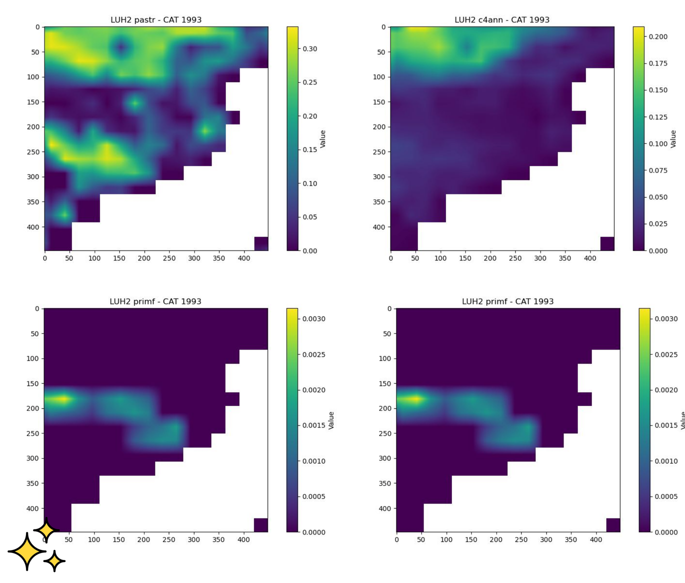
Training

On a yearly basis, the model supports autoregressive prediction using Random Forest (RF) and Extreme Gradient Boosting (XGB).

XGBoost learn



Results

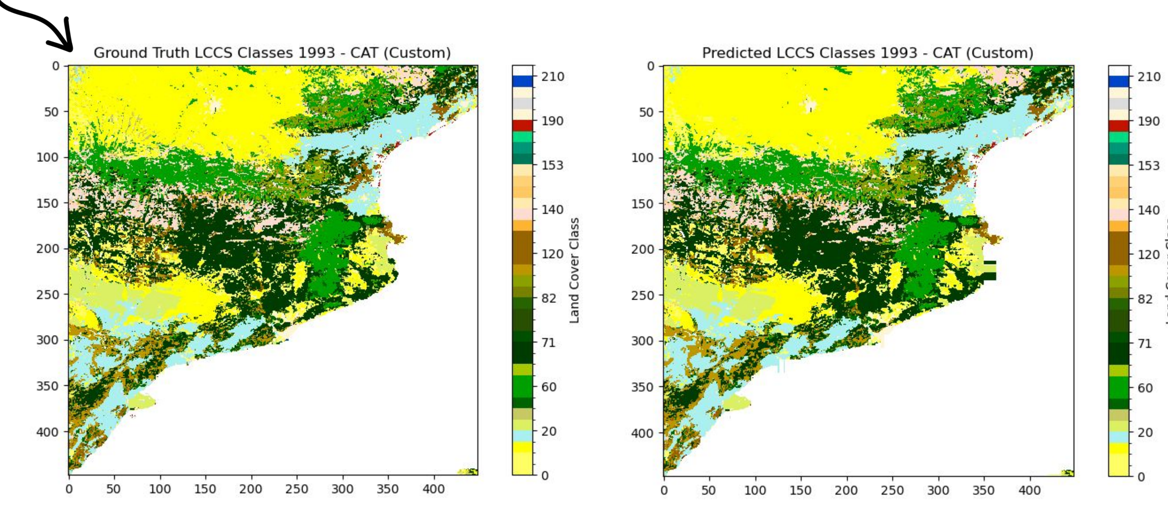


Some example of input data.

Trained on 20 years data on a 400 x 400 km window, with xgb , showed 92% accuracy and 89% precision on classification task.

Historical reconstruction of LC

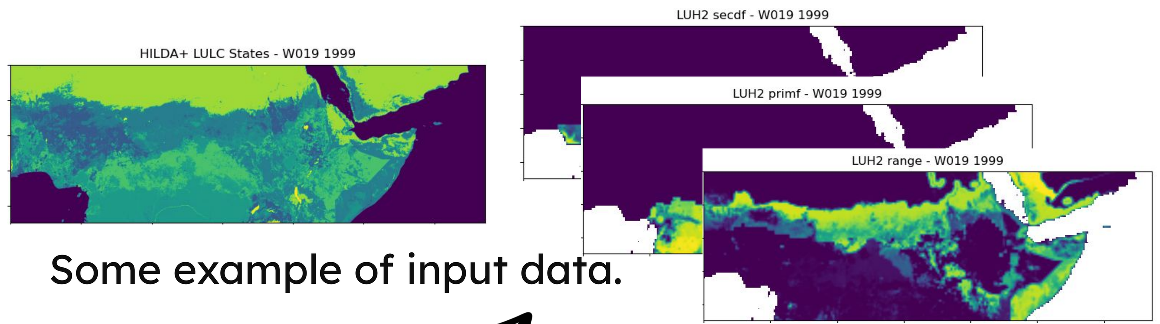
The historical coarse resolution LU data (LUH2h & Hilda+) are used to reconstruct the LC in 1 km resolution for region of Catalonia as a proof of concept.



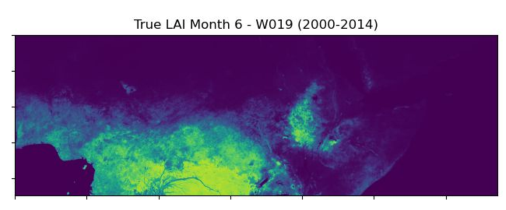
Left) ground truth LC & right) predicted LC for 1993

Historical reconstruction of LAI

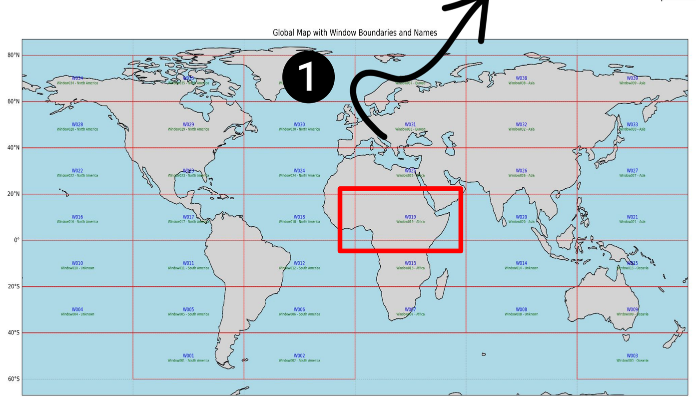
The historical coarse resolution LU data (LUH2h & Hilda+) are used to reconstruct the monthly LAI in 1 km resolution for in independent tiles of 20°x60° (≈ 2226 km x 6679 km). In total, 39 regional models were created. Each model is trained independently on 14 years of data with xgb and prediction are stored as Zarr store. The average misfit reduction of models are around 90%.



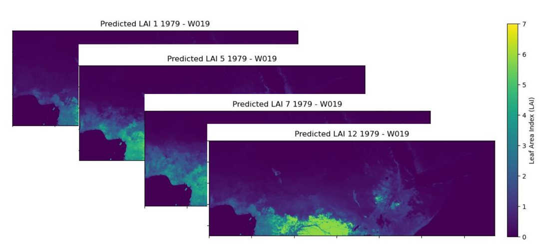
Some example of input data.



Example of ground truth data.



1

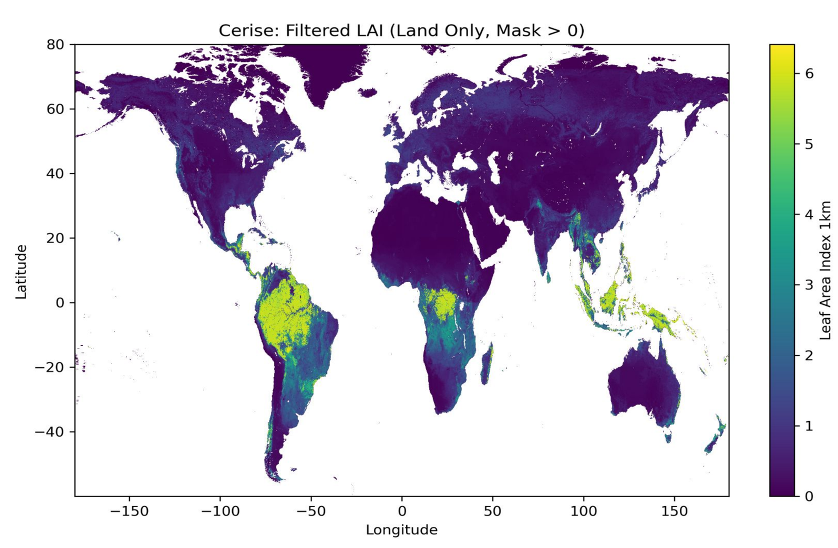


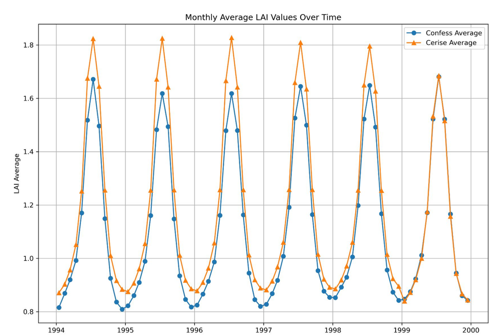
2

Predicted LAI for a single region.

3

The regional models are stitched together to create a global LAI maps in the preprocessing step. Early results show agreement with ground truth on global scale.





What is next?

- **Incorporating more diverse data**: climate, topography, soil type, anthropogenic footprint, and more
- **Leveraging advanced ML** architectures to fully exploit the richness of this data
- **Developing emulators** for LC, LU, and LAI tailored to IFS & ICON (weather-climate models)

We will be in Vienna! Come and talk to us!

Here!

Session AS3.44
Wed, 30 Apr, 14:00–18:00 Hall X5.

TerraDT CONCERTO EGU General Assembly

The CERISE project (grant agreement No101082139) is funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Commission. Neither the European Union nor the granting authority can be held responsible for them.

Funded by the European Union