cmcc
Centro Euro-Mediterraneo sui Cambiamenti Climatici

www.cmcc.it

Reconstruction of historical land surface boundary conditions with ML

Amirpasha Mozaffari
Barcelona Supercomputing Center (BSC)

AI4C Workshop
Session 2: Climate Predictions and Extremes

June 30th – July 2nd 2025 Como, Italy

Introduction

Amir**pasha** Mozaffari (he/him)
Postdoc at the Barcelona Supercomputing Center (BSC)
Juan de Cierva fellow

Artificial Intelligence team in Computational Earth Sciences (CES)
Earth sciences department

Earth Sciences Department | BSC Barcelona Supercomputing Center, Centro Nacional de Supercomputación

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES | Cofinanciado por la Unión Europea | AGENCIA ESTATAL DE INVESTIGACIÓN

Team

Ai4Land

Amanda Cardoso Duarte
Enric Millán Iglesias
Amirpasha Mozaffari

Land Surface & Vegetation Modeling

Markus Donat
Stefano Materia
Etienne Tourigny
Lina Teckentrup
Vinayak Huggannavar

Projects

(2023 - 2026) Historical reconstruction of Leaf Area Index (LAI) 1925 - 2020

(2025 - 2028) Historical reconstruction & future prediction of LU/LC & LAI 1850 - 2100 for weather and climate

(2025 - 2028) Historical reconstruction, future prediction & emulator of LU/LC & LAI for Digital Twins

(2025 - 2029) Fine-tuning of foundation model for LU/LC & LAI emulation

cmcc

Overview

Just to be on the same page, here are **some terminology:**

- **Land Use (LU):**
  Human activities or management practices that alter or maintain land, such as agriculture, forestry, or urban development.

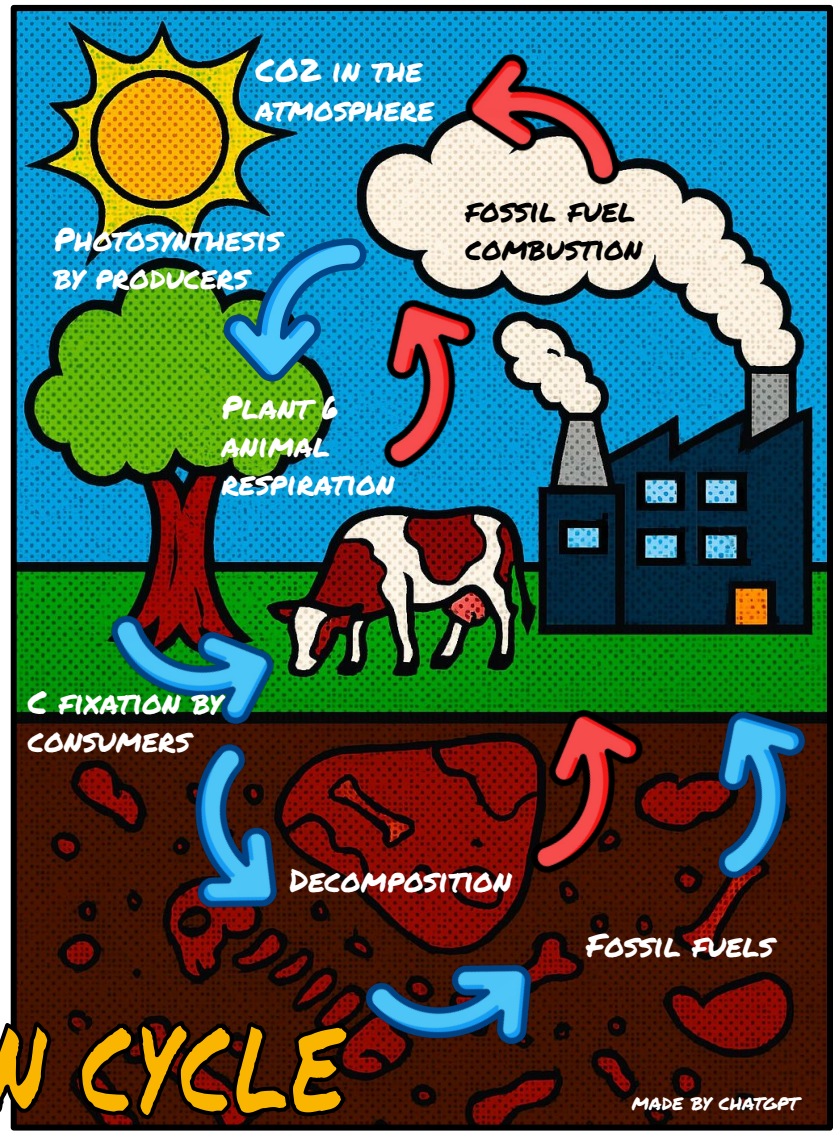
- **Land Cover (LC):**
  The physical surface characteristics of Earth's land, including vegetation, water bodies, and urban areas.

- **Leaf Area Index (LAI):**
  A dimensionless measure of leaf canopy coverage, representing the total leaf area per unit ground surface area.

The global carbon cycle regulates climate, and natural land carbon sinks take up 25% of fossil fuel emissions!

But…

- The future strength of the land carbon sink is very uncertain (due to many factors).
- Getting the land carbon sink right is important for reducing uncertainty in climate projections and, consequently, any mitigation plan.
- In our work, we try to reduce this uncertainty using available data and ML.

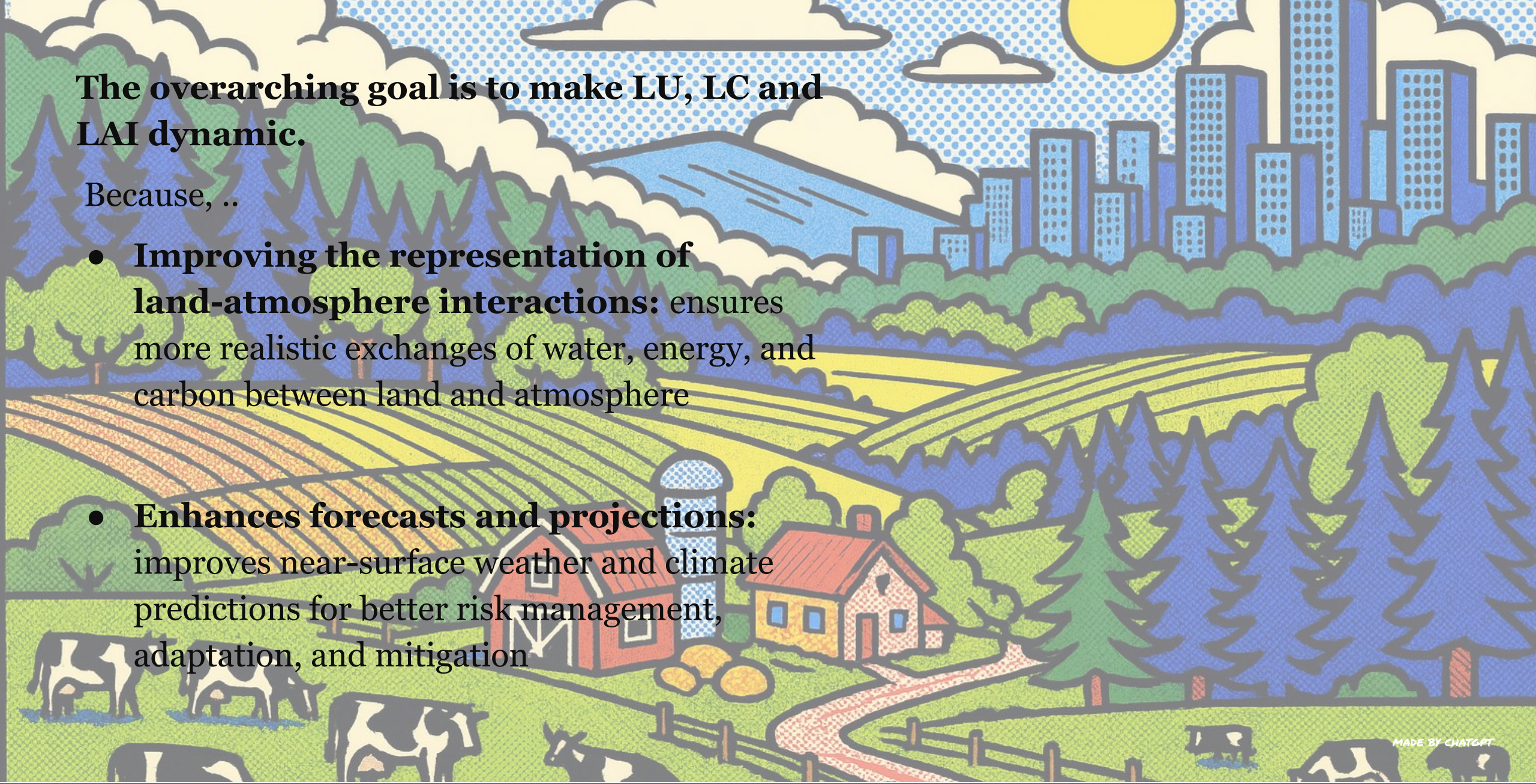


The overarching goal is to make LU, LC and LAI dynamic.

Because, ..

- **Improving the 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



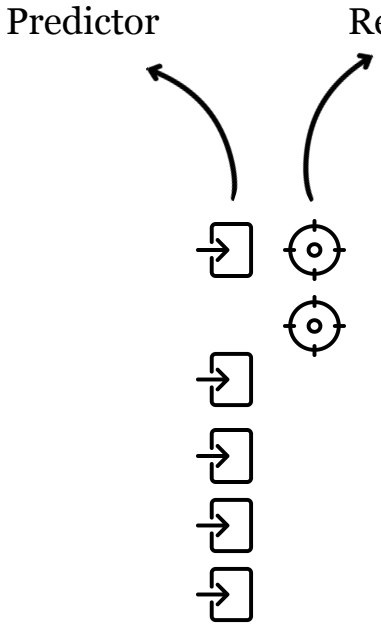
Results

- **AI-based reconstruction of land-surface boundary conditions is a capable tool.**

- **What we tested**
  - LC ⇄ LU mapping – infer land cover (LC) from land use (LU)
  - LAI from LU – derive leaf-area index (LAI) directly from LU
  - LAI anomalies – predict LAI departures using climate-anomaly + LU + LC inputs

- **Early takeaways**
  - Simple models already capture short-term trends well
  - Climatology remains the baseline to beat for LAI
  - Rigorous validation of historical reconstructions is still the biggest challenge

Technical aspects



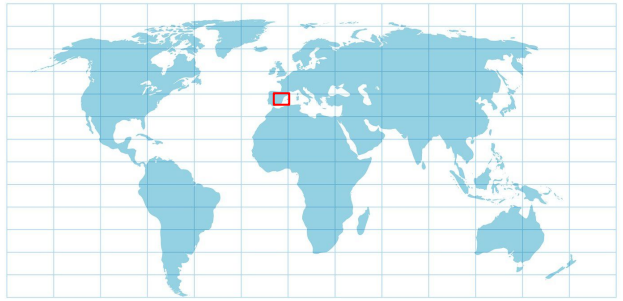
Datasets:

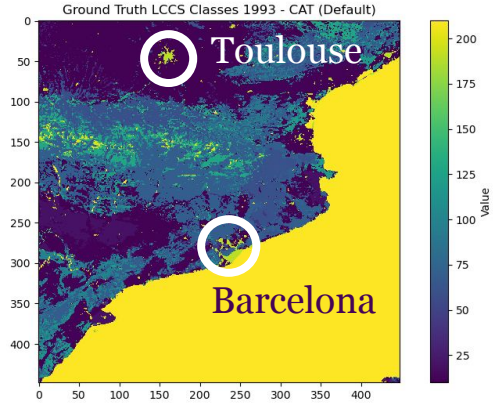
- **ESA CCI Land Cover** yearly maps 300m resolution from 1992 to 2020
- **CONFESS LAI** 5-days maps at 1 km resolution from 1980 to 2020
- **LUH2h Land Use** yearly maps 0.25° × 0.25° resolution from 950 to 2015
- **HILDA+** Land use yearly maps at 1 km resolution from 1899 to 2020
- **ERA5 - Land** Monthly reanalysis dataset at 9 km resolution from 1950 - present
- **CO2 Concentration** Monthly Average Mauna Loa CO2 from, 1957 - present

Regional Land Cover reconstruction:

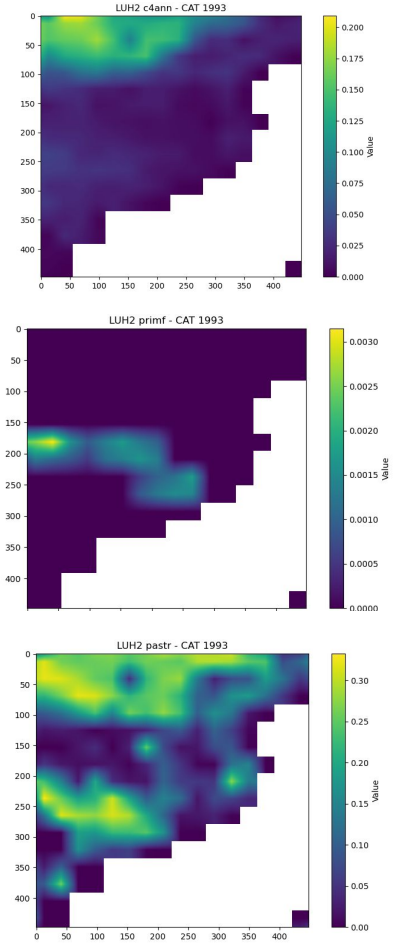
Trained on 20 years, tested on 3 year

- Features:
  - LU: LUH2h
  - LU: HILDA+ as water mask
- Target:
  - LC: ESA CCI LC
- Models: RF, XGB
- No auto-regression





Target : LC

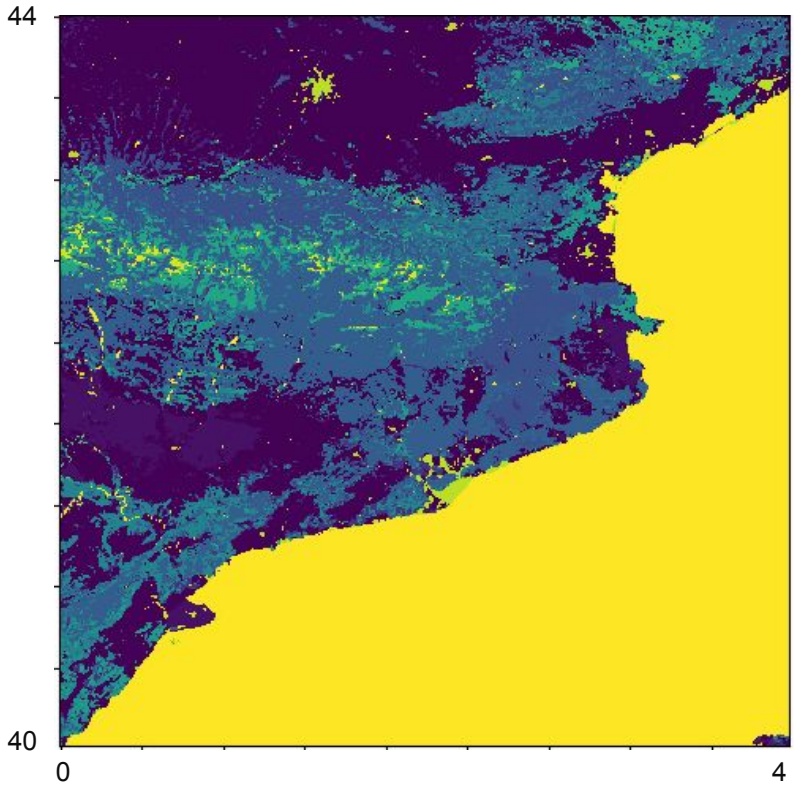


Input: LU

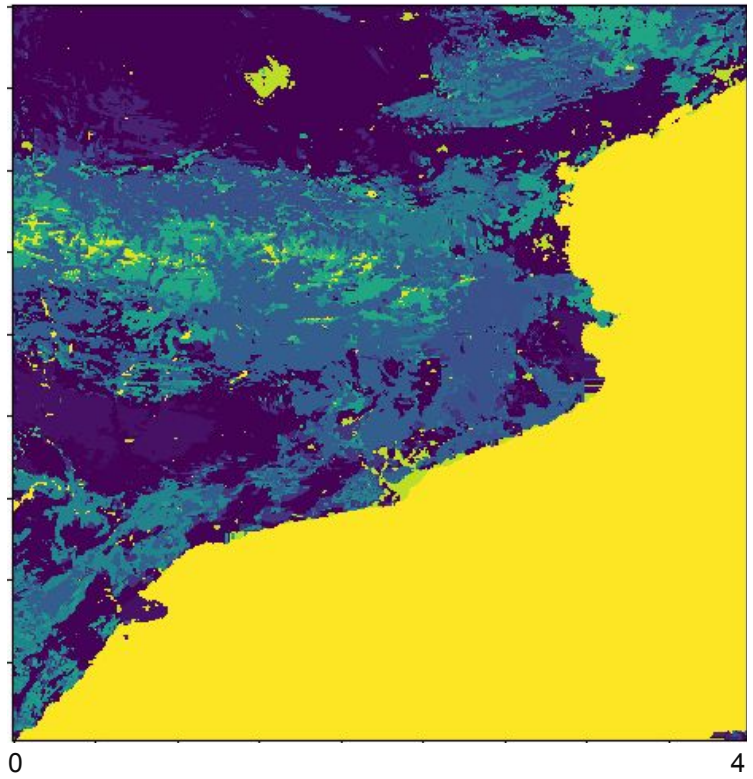
Regional Land Cover reconstruction:

92% accuracy

89% precision



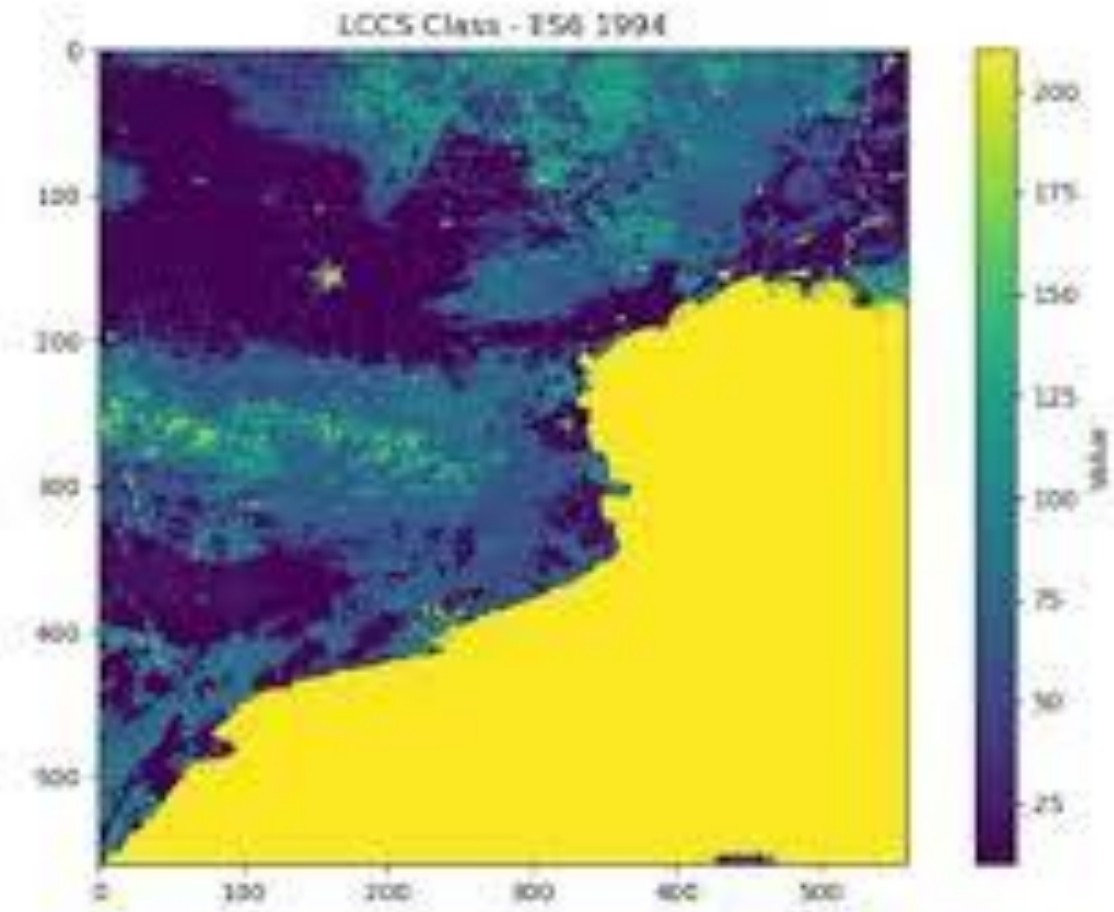
Ground truth LC 1993 (Catalonia, ES)



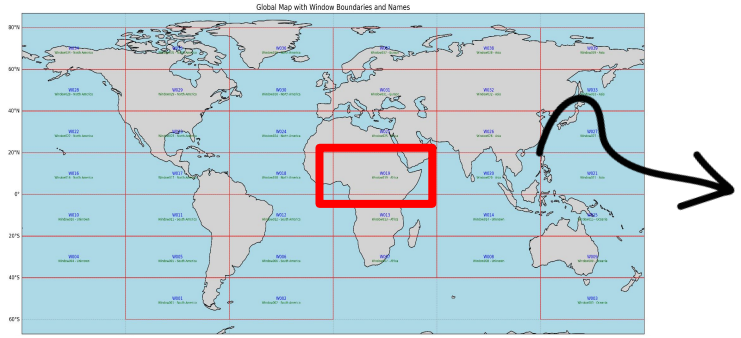
Predicted LC 1993 (Catalonia, ES)

Regional LC reconstruction:

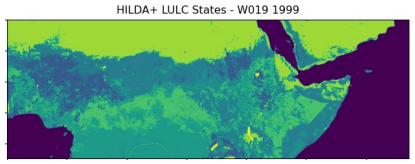
Inference without autoregression

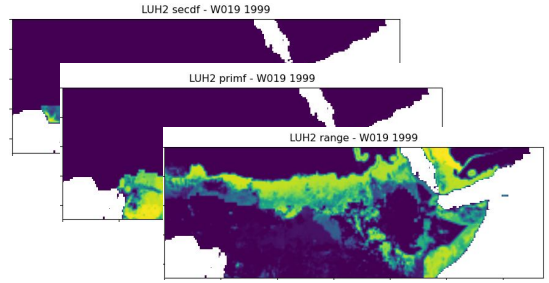


Global Leaf Area Index reconstruction

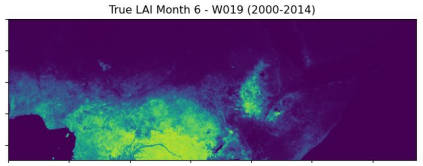


Some example of input data.

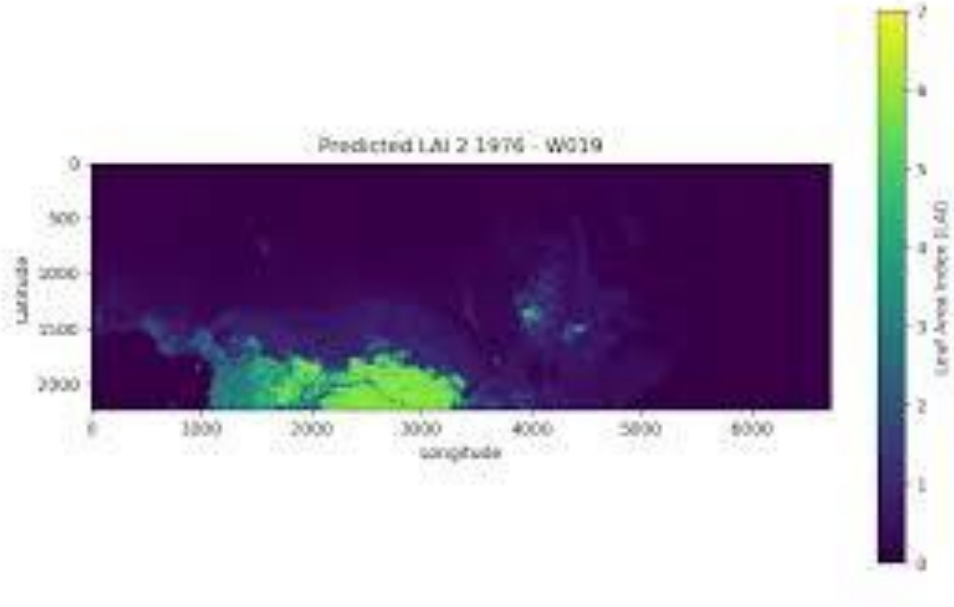




Example of ground truth data.

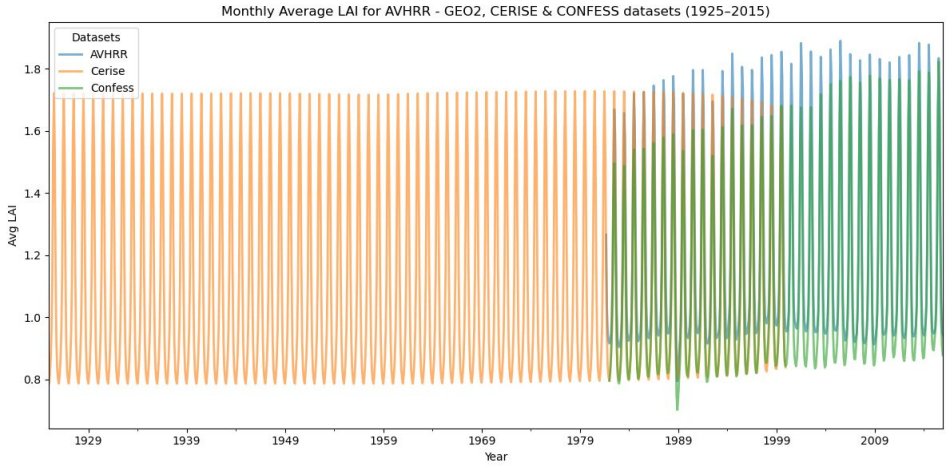


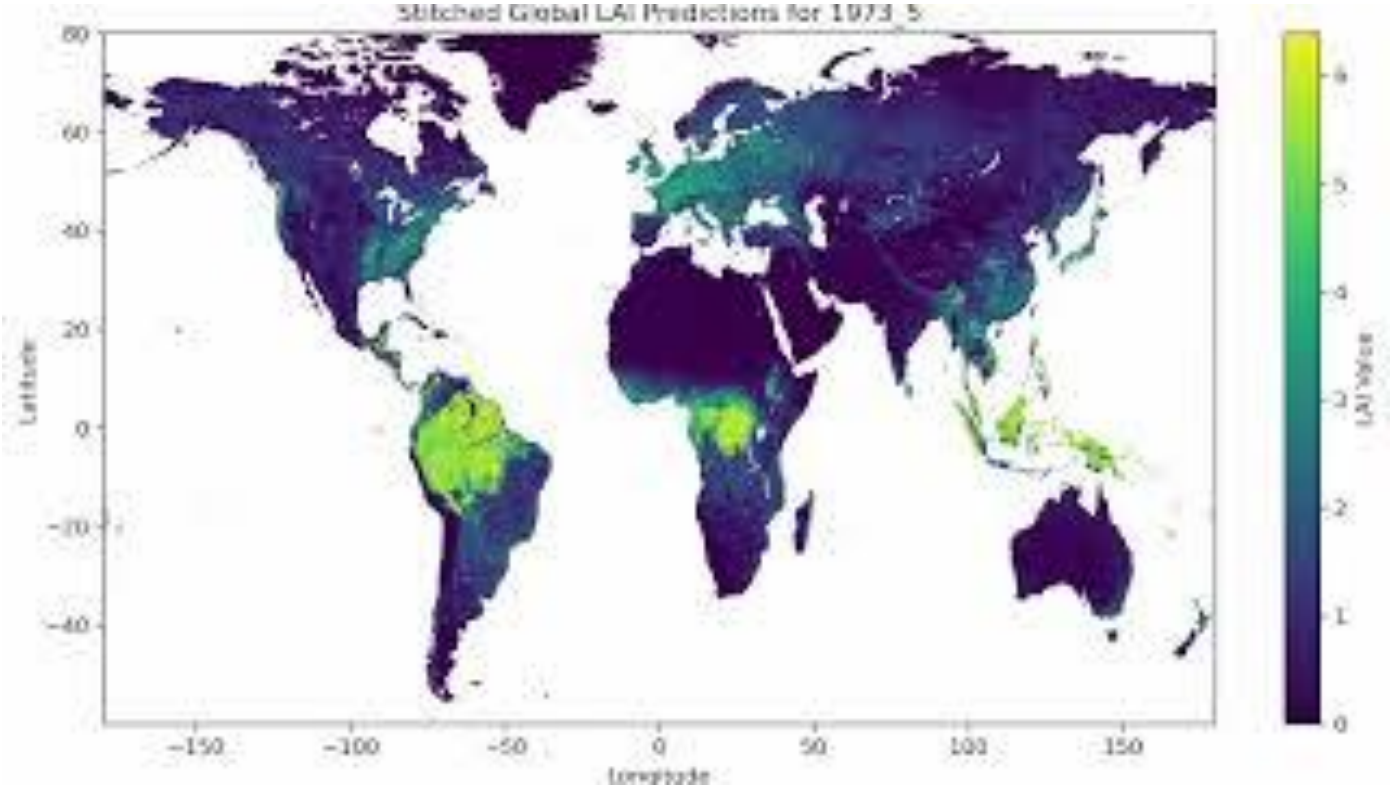
- Global model made of 39 independent regional models
- Each 20° by 60° (≈ 2226 km by 6679 km)
- Trained & validated on 14 years, tested on 1 year
- Features:
  - LU: LUH2h
  - LU: HILDA+ in-land water mask
- Target:
  - LAI: CONFESS
- Model : XGB
- The average misfit reduction of 90%



Global Leaf Area Index reconstruction

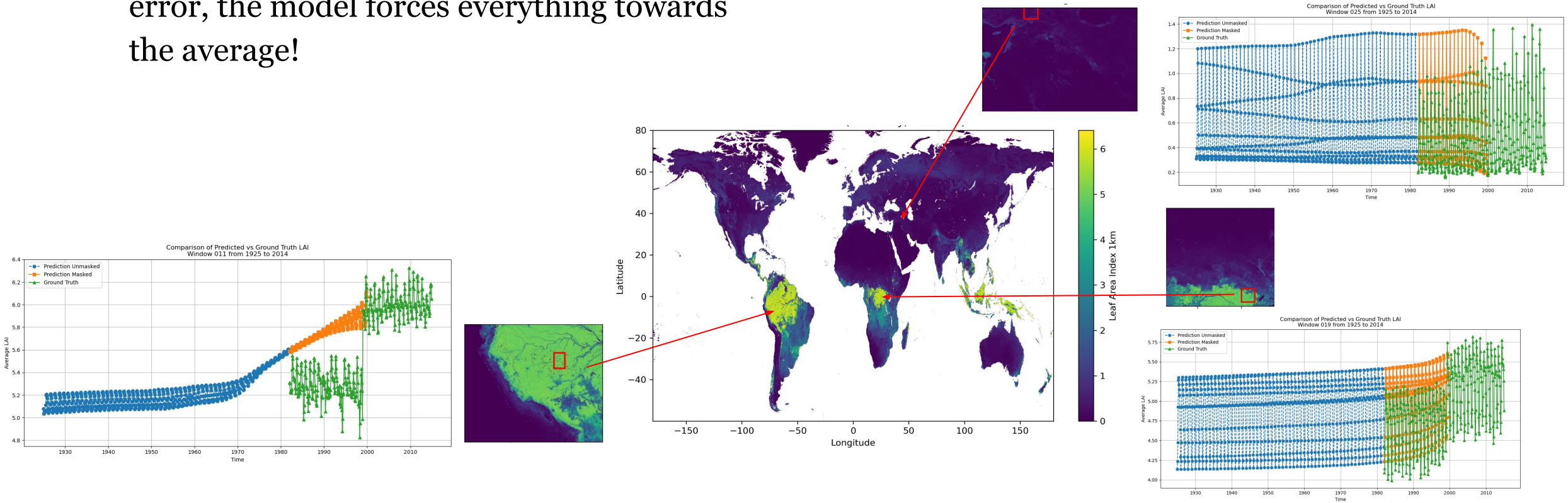
The regional models are stitched and merged to create a global LAI dataset that capture the climatology and reflecting the LU change





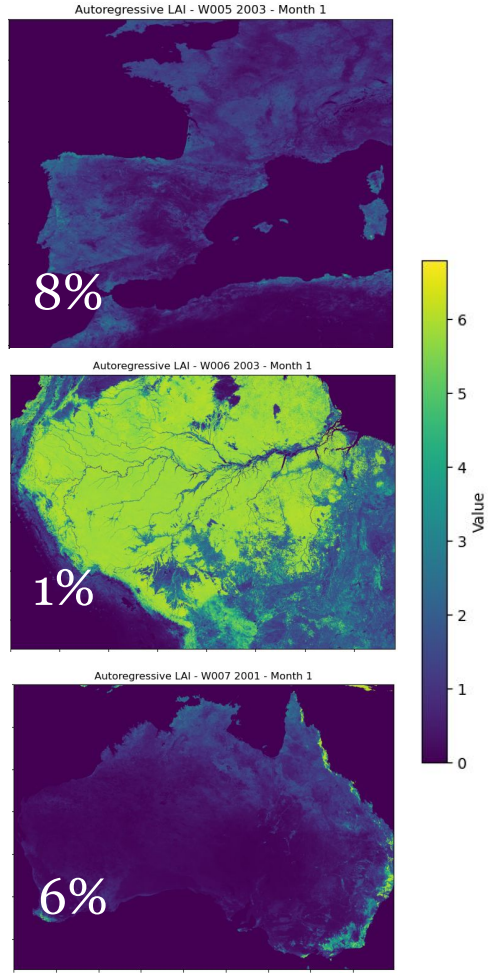
Global Leaf Area Index reconstruction

As it is aimed at minimizing the average error, the model forces everything towards the average!



Regional Leaf Area Index reconstruction

- Regional model to reconstruct LAI anomalies
- Trained 12 years & validated on 2 years, tested on 1 year
- Features:
  - LU: LUH2hl
  - LU: HILDA+
  - LC: ESA CCI
  - ERA5-L (monthly precipitation, solar radiation, soil moisture & temperature)
  - CO2
- Target:
  - CONFESS LAI
- Models: XGB, RNN, GRU, LSTM, RF



Improvement of LAI prediction compared to the baseline climatology

Regional Leaf Area Index reconstruction

Where to test next?

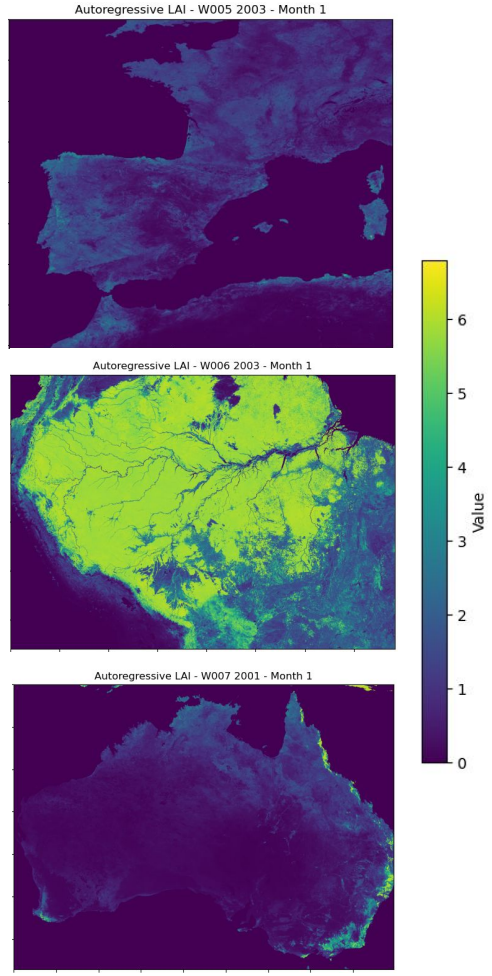
Benchmark both ends of the spectrum of climatology skills:

- Best-case: evergreen, humid tropics (Amazon / Congo) – climatology already strong
- Worst-case: semi-arid & disturbance-prone biomes (Sahel, savannas, fire/insect zones) – climatology weakest

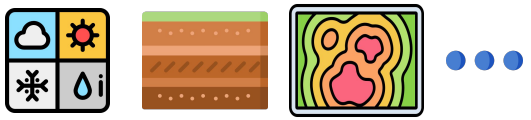
Stress-test in extreme years

- Drought | Heat-wave | Compound event

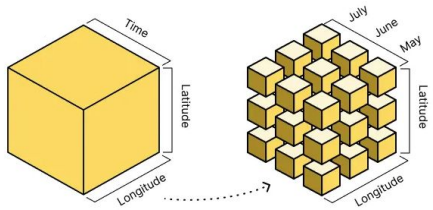
Metric: accuracy gain over climatology in each biome-event combo



Future developments



More relevant data! Climate and weather, topography, soil type, anthropogenic footprint,…



Exploring more efficient data pipeline and structure (e.g., Analysis-Ready Cloud-Optimized (ARCO)..) for training on large data

Exploring the task-based trained models & fine-tuned FMs as emulators (LC, LU & LAI) for climate & weather models (IFS & ICON)

Q&A

[10 min] *Open discussion*

www.cmcc.it