



POLITECNICO
MILANO 1863



cmcc
Centro Euro-Mediterraneo
sui Cambiamenti Climatici



WorldTensor and TerraNova

Open data and foundation models for the coupled human–Earth system

Open-Earth-Monitor Global Workshop 2026 · 7 October 2026

Carlos Rodriguez-Pardo, PhD · with Massimo Tavoni

Politecnico di Milano / EIEE / CMCC

Earth foundation models

	Human systems	Time	Uncertainty	Oceans	Global
Weather & climate foundation models GraphCast, Pangu-Weather, GenCast, Aurora, ClimaX					
Geospatial embeddings SatCLIP, GeoCLIP, AlphaEarth, Copernicus-Embed					
Population Dynamics FM environmental + human-activity signals					

✓ represented ~ partial or limited (e.g. ensembles only) ✗ absent

No model learns the physical Earth and the societies embedded in it together, through time, with uncertainty, at global scale.

Why these gaps matter

3



Couple climate and society

Impact and policy questions need emissions, land use, infrastructure and institutions next to the physical fields.



Map socioeconomic variables

No shared representation to gap-fill or downscale sparse national statistics



Know when the model is wrong

Point estimates fail silently, and fail most where observations are scarcest.



Cover most of the planet

Image-trained encoders carry weak signal over water, which is about 70% of the Earth's surface.

Why? One reason is a data problem

4



Time

Daily observations, monthly fields, multi-year socioeconomic snapshots.



Resolution

From 0.05° satellite products to 0.1° inventories to one value per country.



Domains

Climate, land, ocean, emissions, energy, population, hazards, conflict: separate communities.



Geometry

Rasters, point registries, lines, polygons and country tables.



Formats

GeoTIFF, NetCDF, CSV, shapefiles; different metadata, units, names, projections, licences.

Before any model can learn the coupled system, every source has to be regridded, rasterised, annualised and described the same way.

WorldTensor: our answer to the data problem

A harmonised, ML-ready dataset of the coupled human–Earth system

5

757

variable families
658 annual + 99 static

14

domains, climate
to conflict

0.25°

global grid,
ERA5-aligned

1900–2025

annual data

52,823

NetCDF files,
~46 GB

CC BY 4.0

data + code

14 domains

Climate

Extremes

Air quality

Emissions

Land use

Vegetation

Hydrology

Cryosphere

Ocean

Agriculture

Energy

Human systems

Hazards & conflict

Static context



**Download the
dataset**

Zenodo
doi:10.5281/zenodo.19047618

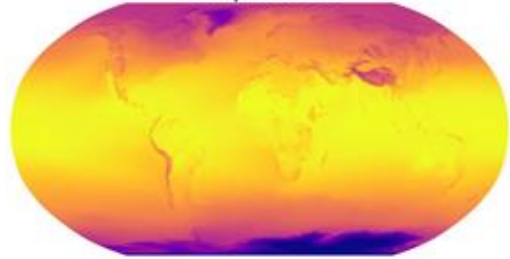
Published in Scientific Data (2026)

Code: github.com/crp94/worldtensor-pipeline

Some of the 757 families, 2015

6

Temperature



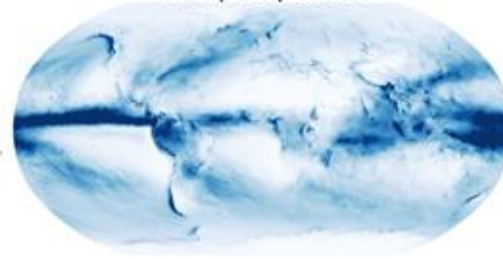
Methane emissions



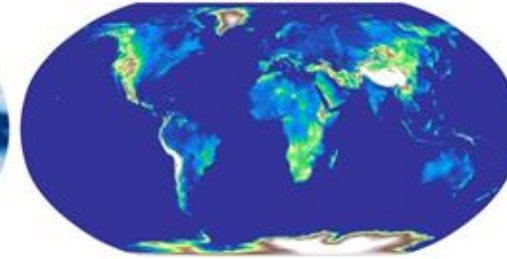
NDVI



Total precipitation



Surface elevation



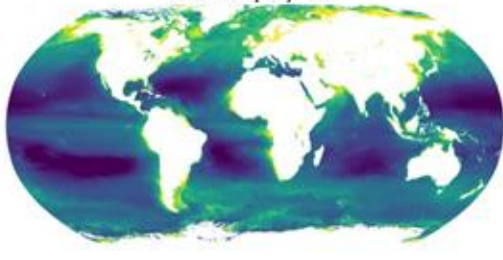
C3 annual crops



GDP (PPP)



Chlorophyll-a



Sea-ice cover



SPEI-12 drought



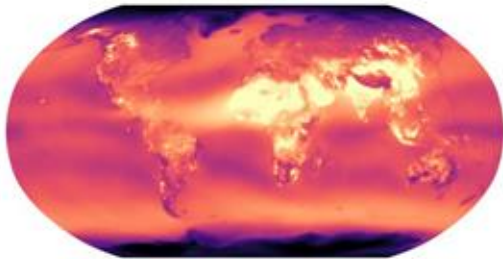
Soil texture: loam



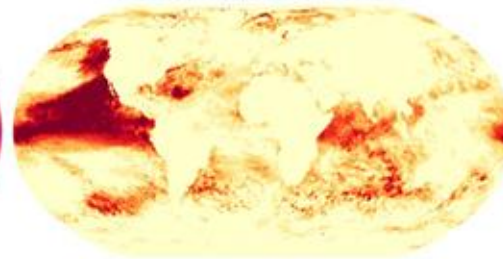
Secondary forest



PM2.5



Marine heatwave months



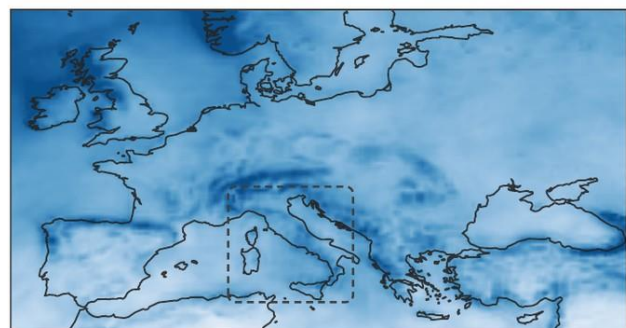
Installed power capacity



TerraNova: one model, two geometries

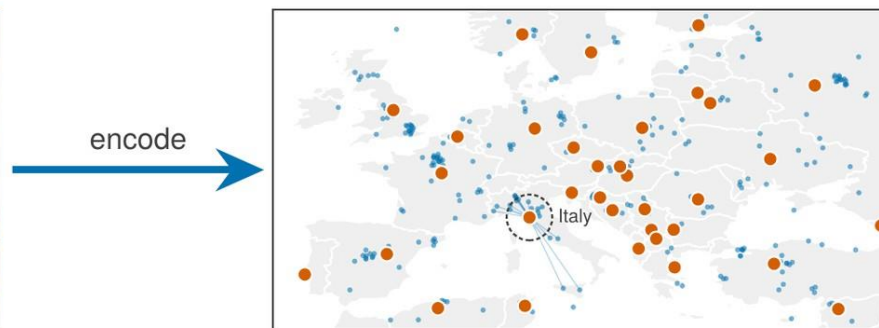
7

The main challenge is geometric, not just a matter of more data



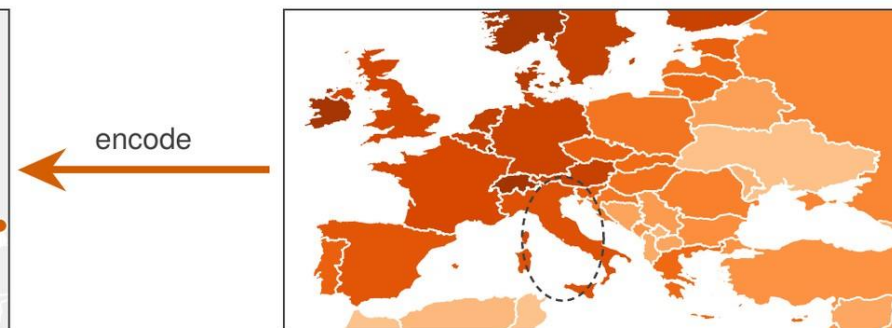
Earth-system fields

Continuous in space (x, t) — 0.25° grid, annual



Two views of one reality

The same place as a field and as a country



Societal indicators

Discrete administrative units (c, t) — annual

Fields

The physical Earth is measured continuously in space and time, (x, t) , and ignores borders.

Countries

Societies are reported for administrative units, (c, t) : one value per country per year.

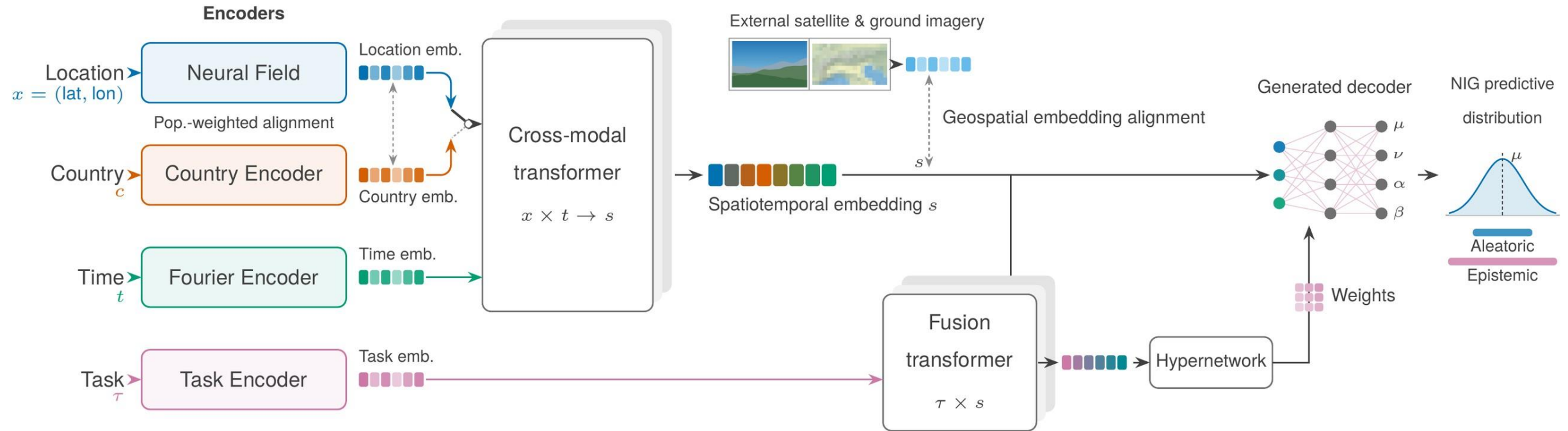
Usual fixes are lossy

Averaging fields over borders hides within-country gradients

TerraNova keeps each record in its native geometry: 512 WorldTensor fields + 512 national indicators, one latent space.

Architecture, briefly

8



Encoders

Location (gated neural field $\times$ hash grid), country, time (Fourier), task.



Fusion

Cross-modal transformers build one spatiotemporal state for both geometries.

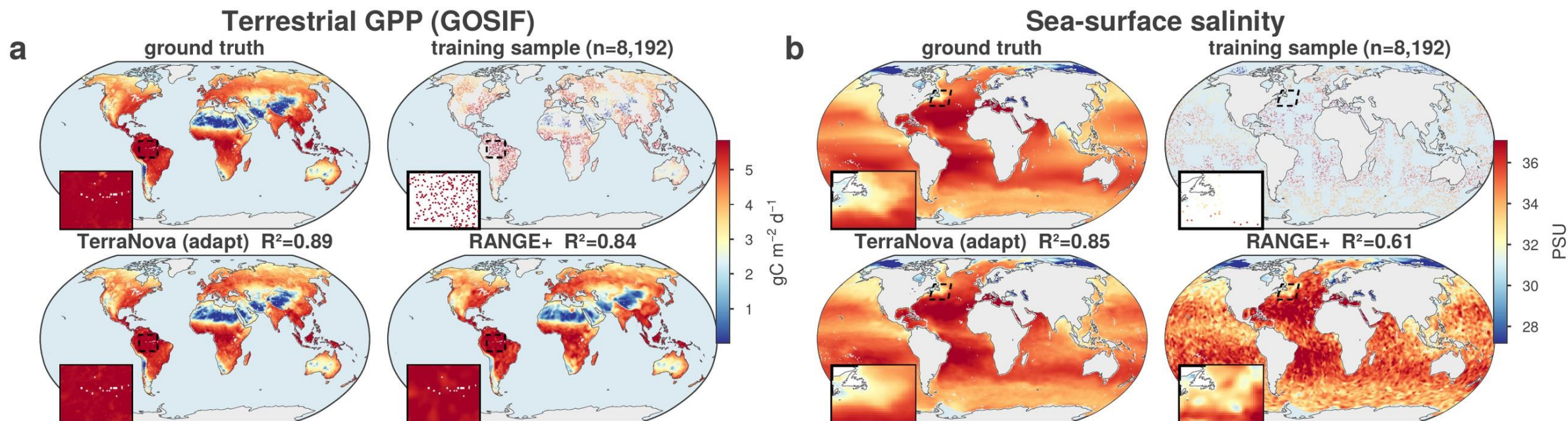


Generated decoder

A hypernetwork builds a small decoder per query; its head returns a distribution.

Competitive with purpose-built geospatial embeddings

9



0.88 vs 0.77

mean held-out R^2 vs the strongest baseline (RANGE+) at 8,192 labels

0.88 vs 0.62

on ocean targets, where image-trained encoders have little signal

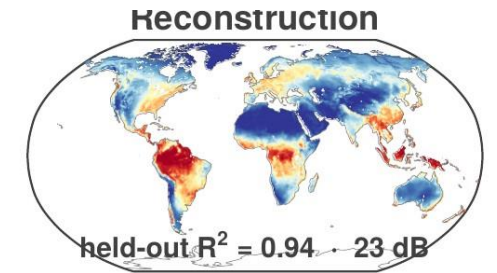
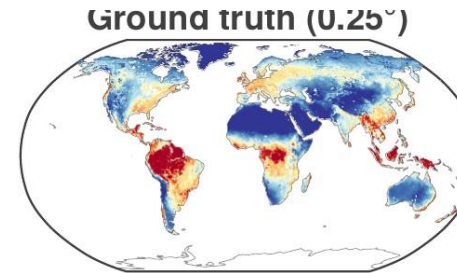
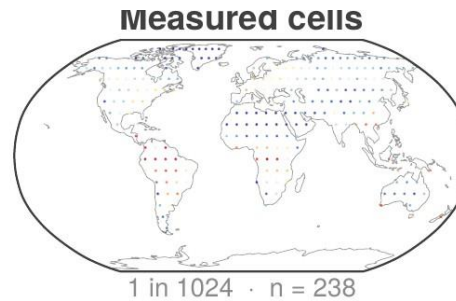
11 / 11

targets at or above the best baseline; RANGE+ still leads at 256 labels

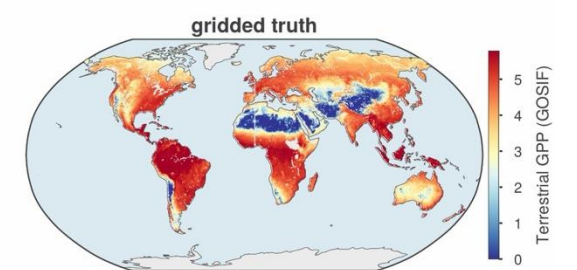
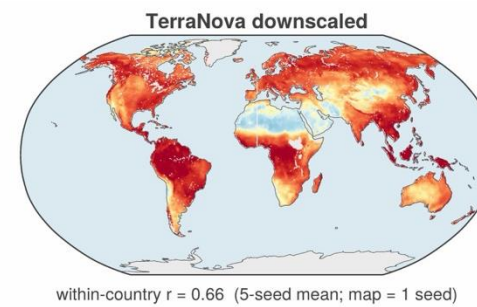
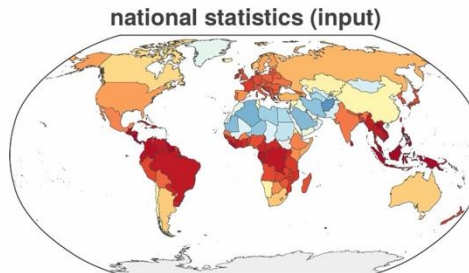
Beyond the benchmark

10

Sparse → dense fields



National → gridded



Conclusions, limitations and next steps



Conclusions

- Current Earth models miss the human side, time, uncertainty or the ocean.
- WorldTensor harmonises 757 families on one grid: open and published.
- TerraNova learns fields and countries in one space, with uncertainty, and is competitive with geospatial embeddings.



Limitations

- **Time dynamics:** time is an input, not a dynamical model.
- **Causality:** learned associations only. No causal effects, feedbacks or policy scenarios



Next steps

- Conditioned dynamics
- Networks between countries: trade and migration.
- New data modalities like language



TerraNova

project page · arXiv:2607.29527



WorldTensor

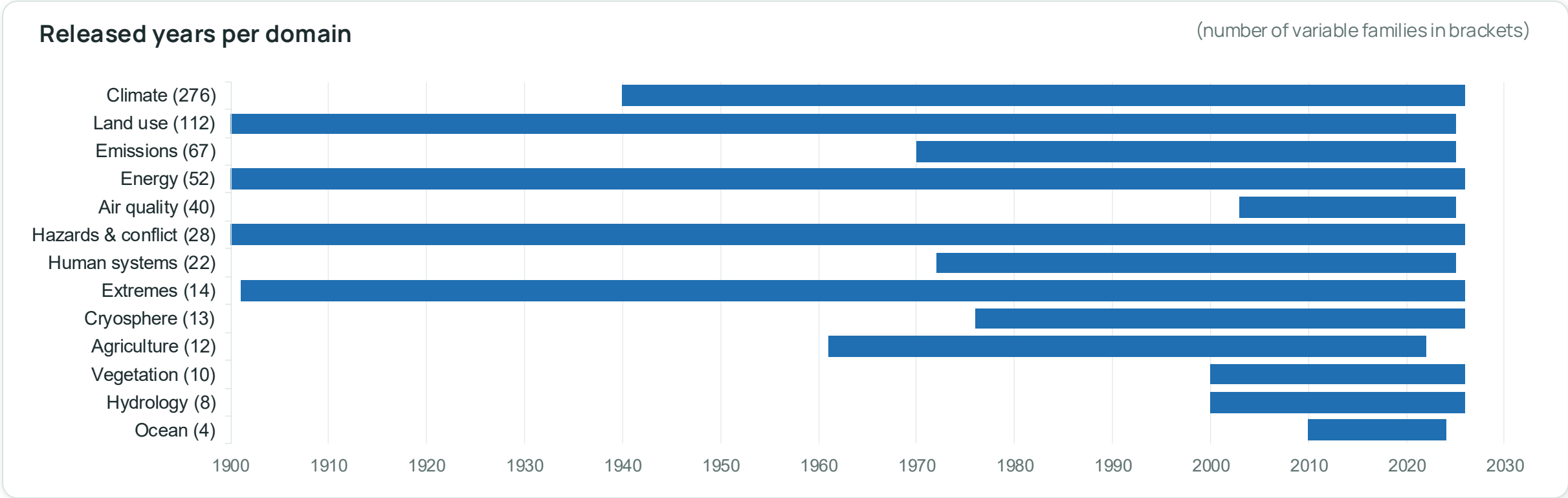
dataset · Zenodo

Thank you! Questions?

carlos.rodriguezpardo.jimenez@gmail.com

Supported by the ERC, grant agreement 101044703 (EUNICE).

WorldTensor through time

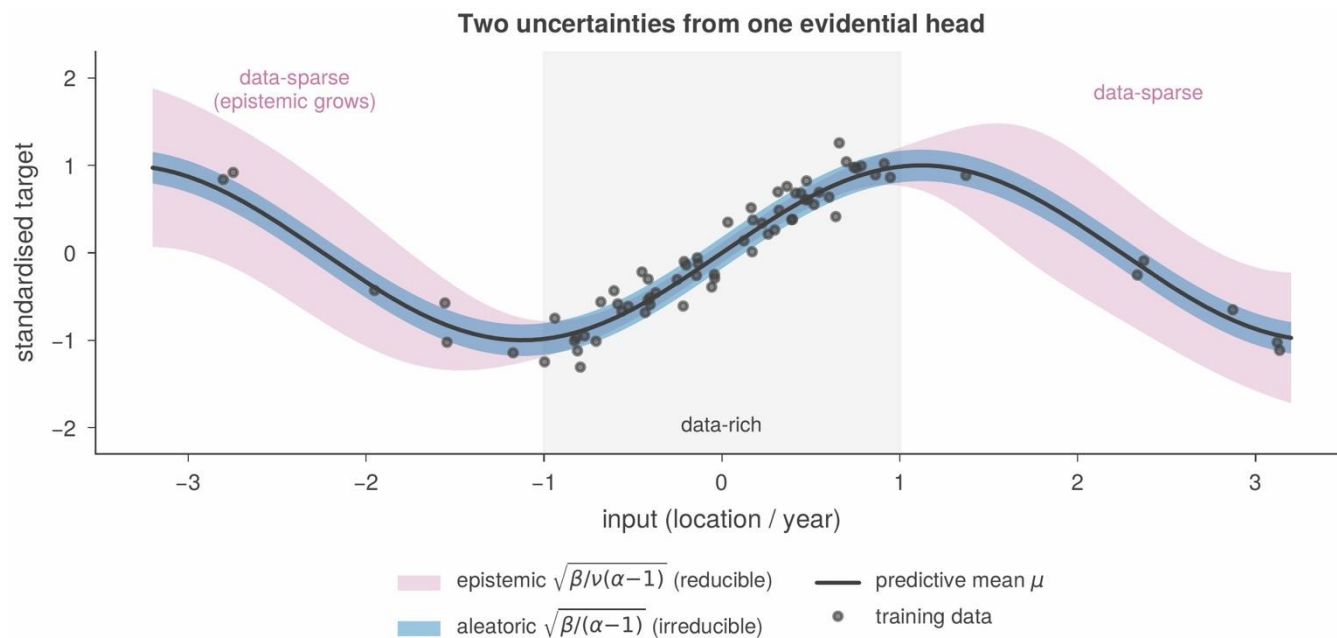


Variable families with data in each year, 1900 → 2025



Uncertainty quantification

Every query returns a predictive distribution, not a point



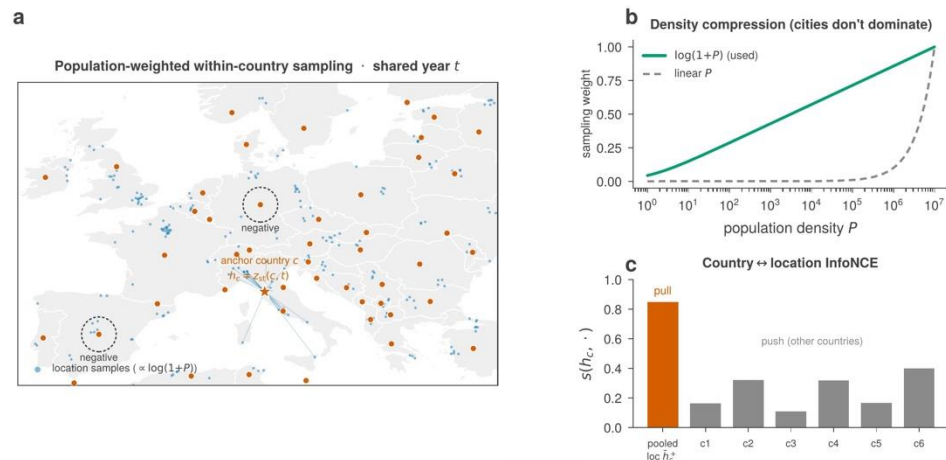
$$\text{Var}[y] = \beta/(\alpha-1) + \beta/(v(\alpha-1))$$

aleatoric proxy (data noise) + epistemic proxy (model ignorance)

- **Evidential head:** Normal-Inverse-Gamma parameters in one forward pass; Student-t predictive.
- **Calibrated:** raw intervals are too wide on new variables; one conformal constant per task restores 90% coverage.
- **Informative:** widths grow as measurements thin out and as queries leave the observed years.
- **Used in training:** the most uncertain tasks are resampled more often.

Coupling the geometries: two alignment losses

15

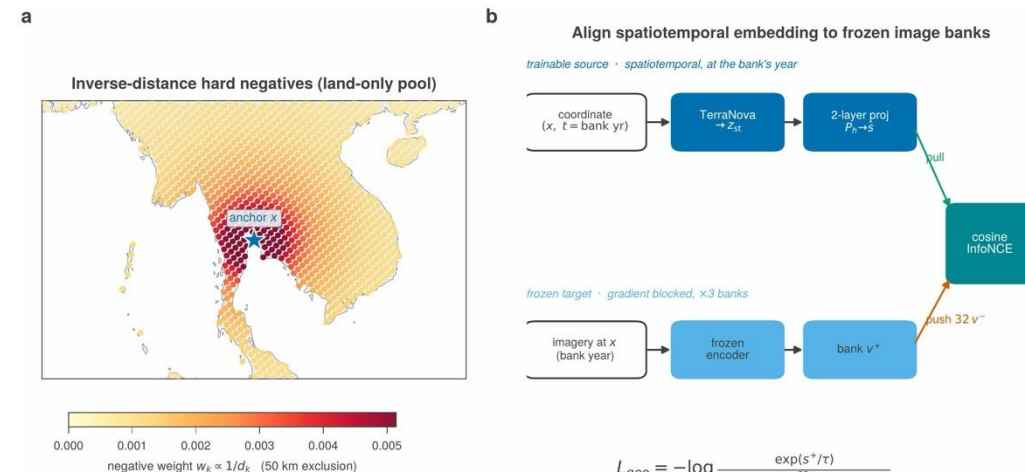


$$L_{\text{ctry}} = -\log \frac{\exp(s(h_c, \tilde{h}_c^+)/\tau)}{\exp(s(h_c, \tilde{h}_c^+)/\tau) + \sum_{j \neq c} \exp(s(h_c, \tilde{h}_j)/\tau)}$$

all embeddings spatiotemporal at one sampled year t : $h_c = z_{\text{st}}(c, t)$ (anchor); $\tilde{h}_c^+ =$ pooled in-country locations (positive); $\tilde{h}_j =$ other countries ($K=16$); $s = \text{cosine}$, $\tau = 0.07$

Population-weighted country–location alignment

Each country's embedding is pulled towards the mean embedding of locations inside its territory, sampled $\propto \log(1 + \text{population})$ so cities count but do not dominate.



$$L_{\text{geo}} = -\log \frac{\exp(s^+/\tau)}{\exp(s^+/\tau) + \sum_{k=1}^{32} w_k \exp(s_k^-/\tau)}$$

$$s^+ = \hat{s}^\top \hat{v}^+, \quad s_k^- = \hat{s}^\top \hat{v}_k^-, \quad w_k \propto 1/d_k, \quad \tau = 0.07$$

Alignment with geospatial embeddings

Locations are aligned with frozen SatCLIP, GeoCLIP and Copernicus-Embed vectors at the years their imagery was taken, with nearby hard negatives.

Ablation: removing the first erases cross-geometry abilities (retrieval falls to chance); removing the second costs R^2 on land but not on ocean, where its teachers have no data.