TerraNova: A Foundation Model for the Anthropocene

Carlos Rodriguez-Pardo carlos.rodriguezpardo.jimenez@gmail.com
Department of Management, Economics and Industrial Engineering, Politecnico di Milano
RFF-CMCC European Institute on Economics and the Environment (EIEE)
Euro-Mediterranean Center on Climate Change (CMCC)

Massimo Tavoni massimo.tavoni@polimi.it
Department of Management, Economics and Industrial Engineering, Politecnico di Milano
RFF-CMCC European Institute on Economics and the Environment (EIEE)
Euro-Mediterranean Center on Climate Change (CMCC)

Abstract

A defining problem of the Anthropocene is to model the physical Earth and human societies as one coupled system, yet no learned representation spans their observational breadth. We argue the obstacle is geometric: the physical Earth is measured as continuous fields that ignore political borders, whereas societies are reported for administrative units. Earth-system foundation models serve the first geometry; coupling it to the second has required lossy averaging over borders. We introduce *TerraNova*, a foundation model trained on 1,024 physical and societal records in their native geometries: 512 gridded Earth-system fields and 512 national indicators. Dedicated encoders represent location, country, time and task, cross-modal transformers fuse them into a shared spatiotemporal state, and a hypernetwork generates a per-query decoder whose evidential head returns a predictive distribution. Two contrastive objectives couple the representation: a population-weighted alignment between each country and coordinates in its territory, and one to pretrained geospatial embeddings carrying image-derived semantics. Read out through that decoder, the representation is competitive with purpose-built geospatial encoders while spanning axes they do not represent (time, oceans and uncertainty) and supporting country-level capabilities. The frozen backbone reconstructs dense fields from sparse observations and adapts to unseen variables in minutes on consumer hardware.

Keywords: foundation models, representation learning, climate change, Earth system, uncertainty quantification

Project page: https://carlosrodriguezpardo.es/projects/TerraNova/
The Supplementary Information (methods, ablations, extended results, cost) is hosted there.

1 Introduction

A key scientific problem of the Anthropocene (Crutzen, 2002) is to jointly model societies and the Earth system as the interconnected whole they form: greenhouse-gas emissions alter atmospheric composition; land transformation changes hydrology, albedo and carbon storage; infrastructure concentrates exposure; institutions shape vulnerability; and climate impacts affect economic development, migration and policy (Steffen et al., 2015; Richardson et al., 2023). These connections are already visible in observations: climate change has altered the distribution of economic growth and inequality across countries (Diffenbaugh and Burke, 2019), its effects are already visible in human development outcomes (Tavoni

et al., 2025), and estimates of its economic cost are large and widely dispersed (Moore et al., 2024). Planetary change and policy-relevant research requires data representations that span physical, ecological, economic and institutional systems, rather than treating their records as unrelated datasets (Ou et al., 2026).

Existing model families cover only a part of that system. Weather and climate models resolve physical dynamics through numerical approximations to physical law, and are compared through coordinated multi-model protocols (Eyring et al., 2016); economic and integrated-assessment models represent welfare and policy through regional aggregates, scenarios and damage relationships (Nordhaus, 2017; Riahi et al., 2017). Earth-system foundation models have improved the accuracy and speed of geophysical prediction, from global forecasting (Lam et al., 2023; Bi et al., 2023) to multi-variable and multi-domain adaptation (Nguyen et al., 2023; Bodnar et al., 2025), but they remain confined to the physical geometry. No single learned representation spans the physical Earth and the societies embedded in it, and to our knowledge none learns from both simultaneously at a global scale.

The physical Earth is measured as a *field*: climate, vegetation, or soil moisture vary continuously over space and time. Societies are observed within *boundaries*: their governance, health and development are reported yearly for countries and other units whose borders are political rather than physical. Italy is not a coordinate, and a grid cell in the Alps is not an economy, yet local environmental conditions and national institutions describe the same places at different scales. Standard processing resolves the mismatch by averaging fields over countries or rasterising administrative statistics, but both transformations are lossy: aggregation hides within-country gradients, whereas rasterisation implies precision that was never observed. We therefore formulate coupled environmental–societal modelling as a *multi-geometry representation-learning* problem (Figure 1).

Here we present *TerraNova*, a foundation model trained on 1,024 variables relevant to the Anthropocene, from physical Earth-system fields to national socioeconomic and institutional indicators, each in its native geometry. Purpose-built encoders represent location, country, time and task, cross-modal transformers fuse them into a shared spatiotemporal state, and a hypernetwork generates a per-query decoder whose evidential head returns a predictive distribution with aleatoric and epistemic uncertainty proxies. Two contrastive objectives couple the geometries and modalities: a population-weighted, time-aware objective aligns each country with coordinates sampled from its territory, establishing the field–boundary correspondence, while a second aligns TerraNova's location representations with pretrained geospatial embeddings, importing geographic semantics learned from visual information. Conceptually, TerraNova treats physical fields, administrative records and geospatial imagery as different observations of one latent reality. The shared statistical structure it learns supports transfer, spatial inference and hypothesis generation rather than causal or scenario analysis. We make the following contributions:

- To our knowledge, TerraNova is the first model to learn one representation across the observational breadth of the Anthropocene, trained on 1,024 variables in two geometries within a single backbone.
- That breadth is made possible by inter-geometry coupling. A country–location contrastive objective connects administrative units to their territories, while alignment to geospatial embeddings anchors the representation in image-derived semantics. The coupling supports

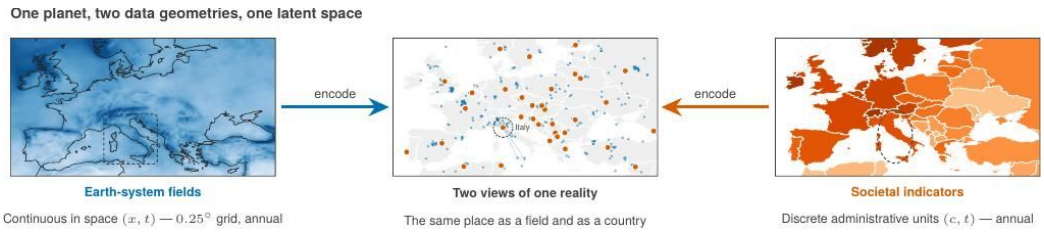


Figure 1: The physical Earth is observed as continuous fields, whereas societal indicators are reported for discrete administrative units. Both are observations of one shared latent reality.

capabilities neither geometry expresses alone, from country–territory retrieval to national-to-gridded downscaling.

- Read out through its task-conditioned decoder, the shared representation leads purpose-built geospatial embeddings on most static targets and label budgets, and extends to axes they do not represent: uncertainty, time and the oceans.
- Every query returns a predictive distribution, a hypernetwork-generated evidential head supplying aleatoric and epistemic uncertainty proxies alongside each prediction. Adaptation to new variables can be reached cheaply through lightweight adapters on a frozen backbone.

2 Related Work

Foundation models for the physical Earth system. Data-driven models now rival operational numerical weather prediction at a fraction of its inference cost. GraphCast learns medium-range forecasting on an icosahedral mesh (Lam et al., 2023); Pangu-Weather applies a three-dimensional transformer over pressure levels (Bi et al., 2023); and FourCastNet applies Fourier neural operators to forecasting (Pathak et al., 2022). ClimaX tokenises heterogeneous climate variables for a shared backbone (Nguyen et al., 2023), while Aurora adapts one pretrained model across weather, air quality, ocean waves and cyclones (Bodnar et al., 2025). GenCast adds probabilistic ensemble forecasting (Price et al., 2025), and Aardvark learns an end-to-end observation-to-forecast pipeline (Allen et al., 2025). These systems are designed for geophysical fields on grids or meshes. TerraNova does not aim to outperform them; instead it addresses a different and complementary problem.

Geospatial embeddings. Geospatial embedding foundation models learn representations of location from Earth observation data (Klemmer et al., 2025b). SatCLIP aligns coordinates with satellite imagery (Klemmer et al., 2025a), GeoCLIP with ground-level imagery (Vivanco Cepeda et al., 2023); Clay (Clay Foundation, 2024) learns multi-sensor Earth-observation features and Prithvi-EO-2.0 adds multi-temporal modelling with explicit location and time embeddings (Szwarcman et al., 2026); Copernicus-FM unifies Sentinel sensors through dynamic hypernetworks, including a global $0.25°$ embedding product (Wang et al., 2025). TerraMind extends this family to any-to-any generation across nine Earth-observation modalities Jakubik et al. (2025) and AlphaEarth learns an embedding field, released as global annual layers Brown et al. (2025). Related to our work, the Population Dynamics Foundation Model combines environmental signals such as weather and air quality with aggregated human-activity signals, only over the United States (Agarwal et al., 2024). Location encoders based on spherical harmonics provide resolution-free coordinate features (Rußwurm et al.,

2024). Existing models share three limitations: trained largely on imagery, they carry a weak signal over water, which covers most of the planet; they do not encode time; and they do not model uncertainty. We leverage this literature in two ways. As the closest comparison class for TerraNova's continuous geographic representation, these encoders are our evaluation baselines; and, as pretrained embeddings, we use them as alignment targets during training, anchoring TerraNova's location representation in semantics learned from imagery. TerraNova is thus time-dependent, uncertainty-aware, and defined at every location on the planet, and it connects these geospatial embeddings' view of place to environmental fields and country variables in one model.

Climate–economics integration. Integrated assessment models (IAMs) couple reduced-form climate and economic systems through scenarios, technologies and damage functions (Nordhaus, 2017; Barrage and Nordhaus, 2024; Riahi et al., 2017; O'Neill et al., 2014; Weyant, 2017); richer systems add detailed energy and land-use sectors (Stehfest et al., 2014), and the family has drawn a substantial critical literature of its own (Gambhir et al., 2019). Empirical climate economics instead estimates relationships from observations, including nonlinear temperature–growth effects (Burke et al., 2015; Hsiang et al., 2017), climate-driven inequality (Diffenbaugh and Burke, 2019), and impacts on human development (Tavoni et al., 2025). Closest to our output geometry, gridded socioeconomic products disaggregate national statistics such as GDP and HDI onto global grids (Kummu et al., 2018); these are valuable per-variable constructions, each built from a fixed covariate recipe, without a shared representation, any learned temporal model beyond the years tabulated, or per-cell uncertainty. TerraNova is not an alternative to the causal, normative or scenario-based frameworks above: instead, it provides a joint observational layer from which downstream empirical and policy models can draw, without assigning causal meaning to the learned associations.

Multi-modal representation learning. A rich literature aligns representations across modalities, notably between language and vision: CLIP aligns images and text (Radford et al., 2021); ImageBind connects several modalities through a shared anchor (Girdhar et al., 2023); and Perceiver IO maps heterogeneous inputs through a common array (Jaegle et al., 2022). The *Platonic Representation Hypothesis* argues that neural networks trained on different data and modalities are converging towards a shared statistical model of reality (Huh et al., 2024). TerraNova instantiates a vision of this idea for the Earth: environmental fields, administrative records and image-derived embeddings describe the same world, but differ in content, geometry, resolution, and availability. Rather than expecting compatibility to emerge from data, TerraNova couples these modalities through highly multimodal training (1,024 reconstruction tasks) and explicit auxiliary learning objectives.

Predictive uncertainty. Uncertainty quantification is key for making robust decisions. Two sources should be distinguished: *epistemic* uncertainty, which reflects limited knowledge and might be possibly reduced with more data, and *aleatoric* uncertainty, the irreducible randomness of a known data-generating process; the two carry different implications for whether to gather observations or to hedge against variability (Tavoni and Valente, 2022). Evidential deep learning places a higher-order prior over a Gaussian likelihood and produces a predictive distribution in a single forward pass (Amini et al., 2020), although subsequent

work identifies regularisation and identifiability limitations (Meinert et al., 2023). Earth-system science and integrated assessment quantify uncertainty through ensembles, parameter uncertainty, internal variability and alternative scenarios (Hawkins and Sutton, 2009, 2011; Chiani et al., 2025), and communicate the result through calibrated confidence and likelihood language (Mastrandrea et al., 2011). TerraNova targets task-conditional, per-query predictive uncertainty for heterogeneous variables: following Amini et al. (2020), it outputs the parameters of a Normal–Inverse–Gamma distribution, providing a total predictive distribution alongside aleatoric and epistemic uncertainty proxies.

3 Model design

In this section, we introduce each component of TerraNova's design; further detail is provided in the Supplementary Information[1]. Two principles run through the whole architecture: every encoder output is rescaled to a sphere of fixed radius, and every residual pathway is initialised as a near-identity transformation.

3.1 Model definition

TerraNova, illustrated in Figure 2, is a neural mapping over tasks, time and space,

$$f_\theta \colon (\tau, u, t) \longmapsto (\mu, \nu, \alpha, \beta), \tag{1}$$

in which τ indexes the prediction task, t is the observation year, and u is the geometry-dependent spatial input: a coordinate $u = x = (\lambda, \phi)$ represents a gridded field, while an administrative unit $u = c$ addresses an indicator reported at the national level. The four outputs parameterise a Normal–Inverse–Gamma predictive distribution over the standardised target. We assume the two data types share a statistical structure that can be decoded into accurate predictions, but we do not enforce a shared geometry: TerraNova preserves each native query geometry, routes both into a common spatiotemporal state, and couples the geometries through soft alignment objectives.

All encoder outputs are radius-normalised,

$$\mathrm{N}_d(\mathbf{v}) = \sqrt{d}\,\frac{\mathbf{v}}{\|\mathbf{v}\|_2}, \tag{2}$$

which keeps cosine similarities, residual gates and attention logits at a stable scale across encoders of very different internal magnitude, and prevents any modality from dominating the fused representation by norm alone. Embedding widths are deliberately small relative to what they encode (256 dimensions for location against a global 0.25° grid, 256 for task against 1,024 target variables, 64 for country, 32 for time, with a 256-dimensional shared spatiotemporal state feeding a 512-dimensional conditioning vector) so the model operates as an encoder–decoder that compresses the Earth system, enabling parameter efficiency and cheap transfer to unseen tasks (§4.4).

1. The Supplementary Information — the full methodological specification, the complete ablation programme, extended results and the computational-cost analysis — is available at the project page: https://carlosrodriguezpardo.es/projects/TerraNova/.

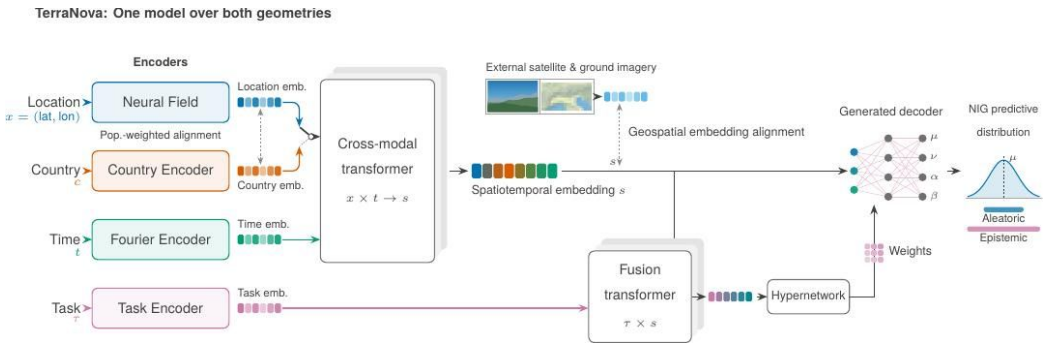


Figure 2: **TerraNova architecture.** Dedicated encoders for location, country, time and task feed geometry routing and temporal fusion; a task-conditioned hypernetwork then generates the local decoder, whose NIG head returns a predictive distribution.

3.2 Encoders

Four purpose-built encoders map location, country, time and task to unit-radius embeddings (Eq. 2).

Location. The location encoder serves two regimes at once: smooth fields (temperature, sea-level pressure) whose spatial spectrum concentrates at low frequencies, and sparse ones (population density, built surface, burned area) with heavy high-frequency content and sharp discontinuities at coastlines and city edges. A network biased towards smoothness would blur the latter; one with enough local capacity for the latter injects noise into the former. We therefore run two complementary encodings of the same coordinate in parallel and let the encoder route between them per latent dimension. A continuous branch (a residual U-Net implicit neural network (Rodriguez-Pardo et al., 2023; Sitzmann et al., 2020) with Gabor-wavelet activations (Saragadam et al., 2023)) provides a smoothness-biased prior; a multi-resolution hash grid (Müller et al., 2022) provides local high-frequency capacity. Then, a learned per-dimension gate blends them,

$$\mathbf{e}_\ell = \mathrm{N}_{256}\big(W_{\text{comb}}\big(\mathbf{g} \odot \mathbf{h}_{\text{wire}} + (\mathbf{1} - \mathbf{g}) \odot \mathbf{h}_{\text{hash}}\big)\big)\,, \tag{3}$$

so each of the 256 output dimensions draws from whichever branch carries the structure it needs. The gate is initialised with a bias towards the continuous branch to introduce another smoothness prior inspired by coarse-to-fine approaches. Coordinates pass through a fixed seam-free spherical lift of our own, which guarantees a non-degenerate longitude gradient at every orientation; see Rußwurm et al. (2024) for the case against naive latitude–longitude parameterisations.

Time. The time encoder maps the year, normalised over 1900–2035, to log-spaced Fourier features refined by a residual MLP (Tancik et al., 2020),

$$\gamma(t) = \big[\sin(2\pi \bar{t}\, s f_k),\ \cos(2\pi \bar{t}\, s f_k)\big]_{k=1}^{32}, \qquad \boldsymbol{\tau} = \mathrm{N}_{32}\big(\mathrm{MLP}_{\text{res}}(\gamma(t))\big), \tag{4}$$

with a learned scale s: low frequencies carry decadal trend, high frequencies year-to-year structure, and the residual MLP supplies the nonlinear refinement the basis alone cannot. A complementary learned transport operator $T(\boldsymbol{\tau}, \Delta)$ represents signed temporal displacement; initialised near the identity, it is shaped by the consistency objectives of §4.2.

Country and task. A country identifier carries no geometric structure to exploit, so the country encoder is an embedding table refined by a residual MLP. The task encoder is an embedding table alone: the task vector conditions decoding jointly with the query's spatiotemporal state (§3.3), and new rows allow unseen variables to be registered without disturbing trained tasks (§4.4). During training, *embedding dropout* is applied to every encoder output before radius normalisation, regularising the embedding's direction while preserving the scale seen by fusion and alignment, and encouraging each latent variable to be useful for every training objective by preventing specialisation.

3.3 Spatiotemporal state and conditional decoding

Geometry routing builds the shared spatial state $\mathbf{s} \in \mathbb{R}^{256}$: a coordinate query uses the location embedding directly, $\mathbf{s} = \mathbf{e}_\ell$, while a country query lifts the country embedding into the same space through a linear projector. Routing rather than mixing lets one fusion and decoding stack serve both geometries: after this point, every query is a 256-dimensional vector on a common bus, whatever its origin. Two cross-modal transformers (Vaswani et al., 2017) then condition this state.

The first transformer injects time information into spatial embeddings. Its output is applied as a gated residual on the projected spatial state,

$$\mathbf{z}_{st} = \mathrm{N}_{256}\big(\mathbf{s}' + g\,\boldsymbol{\delta}\big), \tag{5}$$

with the correction $\boldsymbol{\delta}$ zero-initialised and the gate g starting small: fusion begins at the spatial identity and learns temporal structure only as it earns influence. The second transformer lets the spatiotemporal state query the task token. It queries $\mathbf{z}_{st}$, so its output (the conditioning vector $\mathbf{c} \in \mathbb{R}^{512}$) is aware of when as well as where the query is observed.

A one-layer hypernetwork (Ha et al., 2017) maps $\mathbf{c}$ to the weights of a small decoder, which is then applied to $\mathbf{z}_{st}$: the decoder is generated per query, its computation shaped jointly by task and spatiotemporal context. This is a more expressive form of task conditioning than concatenation or feature-wise modulation (Rodriguez-Pardo et al., 2026), since each of the 1,024 tasks realises a different local decoding function, and makes adaptation to an unseen variable cheap.

3.4 Predictive distribution and uncertainty

The generated decoder outputs the four parameters of a Normal–Inverse–Gamma distribution,

$$\mu = a_\mu, \quad \nu = \mathrm{softplus}(a_\nu) + \tfrac{1}{2}, \quad \alpha = \mathrm{softplus}(a_\alpha) + 1 + 10^{-4}, \quad \beta = \mathrm{softplus}(a_\beta) + 10^{-4}, \tag{6}$$

where the offsets guarantee the parameter constraints and the $\nu \geq \frac{1}{2}$ floor in particular prevents an epistemic blow-up as $\nu \to 0$. Because targets are standardised, a predictive width beyond a few standard deviations is a priori implausible; differentiable one-sided caps exploit this to bound both variance components without clamping gradients. Marginalising the latent Gaussian mean and variance yields a Student-t predictive distribution with mean μ, whose total variance admits the conventional moment decomposition

$$\sigma^2_{\text{pred}} = \underbrace{\frac{\beta}{\alpha - 1}}_{\sigma^2_{\text{alea}}} + \underbrace{\frac{\beta}{\nu(\alpha - 1)}}_{\sigma^2_{\text{epi}}}. \tag{7}$$

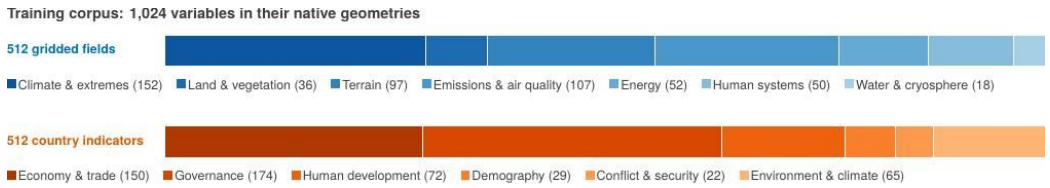


Figure 3: **Training corpus.** 1,024 variables in their native geometries: 512 gridded Earth-system fields and 512 country-level indicators.

The two terms are useful read-outs: the first tracks observation and data variability, the second uncertainty in the predicted mean; but they are not perfectly separately identifiable from the marginal likelihood without additional assumptions (Meinert et al., 2023). We therefore treat them as *proxies*: we evaluate calibration of the total predictive distribution, and use the decomposition as a diagnostic rather than claiming separate calibration of its components. The total uncertainty also steers training, through an active task-resampling schedule.

4 Training

4.1 Data and held-out evaluation

TerraNova is trained on two reconstruction datasets and three fixed alignment banks (Fig. 3). WorldTensor (Rodriguez-Pardo and Tavoni, 2026) contributes 512 fields on a common 0.25° grid, covering, among others: climate, extremes, land use, vegetation, hydrology, energy, and human systems. A companion dataset we assemble for this work, which we name *CountryTensor*, provides 512 national indicators selected from the Quality-of-Government ecosystem (Teorell et al., 2021) for coverage and relevance, indexed by ISO3 code and year across development, institutions, governance, inequality, demography, and conflict. Both are historical observations rather than simulated data. The three alignment banks hold frozen image-derived embeddings (SatCLIP (Klemmer et al., 2025a), GeoCLIP (Vivanco Cepeda et al., 2023) and Copernicus-Embed (Wang et al., 2025)) at 500,000 land coordinates generated by uniform Fibonacci sphere sampling Swinbank and Purser (2006). Targets are standardised per task, so one reconstruction objective serves heterogeneous units and scales and, as §3.4 exploits, unit marginal variance gives the evidential width caps a universal, task-comparable meaning. Further detail is given in the Supplementary Information.

4.2 Loss functions

Training interleaves one reconstruction family and three representation-level auxiliaries,

$$\mathcal{L} = \lambda_{\text{rec}}\mathcal{L}_{\text{rec}} + \lambda_{\text{tr}}\mathcal{L}_{\text{tr}} + \lambda_{\text{geo}}\mathcal{L}_{\text{geo}} + \lambda_{\text{cl}}\mathcal{L}_{\text{cl}}, \tag{8}$$

with fixed weights λ and a stochastically mixed step schedule (§4.3). The reconstruction term $\mathcal{L}_{\text{rec}}$ is the exact negative log-likelihood of the NIG marginal plus two corrections that keep the evidential split well-posed on standardised targets: an evidence regulariser, which penalises standardised rather than raw error and helps avoid the evidence collapse of the

original formulation (Amini et al., 2020; Meinert et al., 2023), and the soft predictive-width cap of §3.4.

Temporal transport ($\mathcal{L}_{\text{tr}}$) supervises the operator $T(\boldsymbol{\tau}, \Delta)$ simultaneously as a predictor and as a regulariser: a transported prediction must be accurate, transport must land on the true target-year embedding, zero displacement must be the identity, and semigroup consistency must be preserved. The consistency targets are stop-gradient, so these terms shape T without collapsing the time encoder itself; the objective regularises temporal geometry rather than replacing direct prediction.

The two alignment objectives create TerraNova's coupled views of place, and both use sampled-negative InfoNCE (van den Oord et al., 2018): for a positive pair $(\hat{\mathbf{s}}, \mathbf{v}^+)$ and negatives $\{\mathbf{v}_k^-\}$,

$$\mathcal{L}_{\text{align}} = -\log \frac{\exp\big(s(\hat{\mathbf{s}}, \mathbf{v}^+)/\tau_{\text{T}}\big)}{\exp\big(s(\hat{\mathbf{s}}, \mathbf{v}^+)/\tau_{\text{T}}\big) + \sum_k \exp\big(s(\hat{\mathbf{s}}, \mathbf{v}_k^-)/\tau_{\text{T}}\big)}, \tag{9}$$

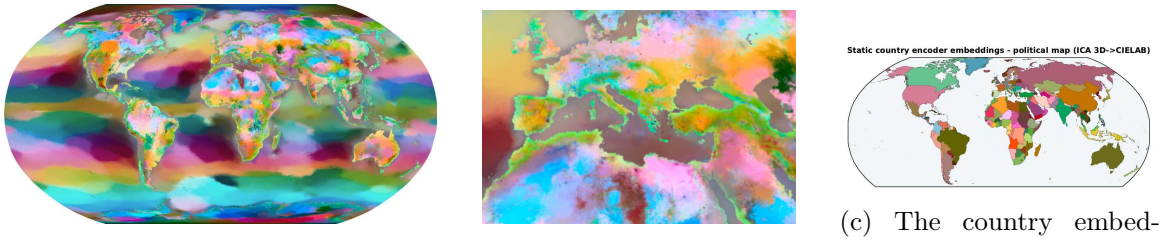
with s cosine similarity.

External geospatial alignment ($\mathcal{L}_{\text{geo}}$) projects the spatiotemporal state through a per-bank head into each embedding space and contrasts it with the bank vector at the same coordinate. The contrast is anchored in time: each bank is queried at the years its own training imagery was collected, so TerraNova's time-dependent state is always read in the bank's native temporal context; and biased towards hard negatives, drawn inverse-distance-weighted from nearby but distinct coordinates, which forces the representation to resolve fine spatial distinctions rather than global gradients alone. This objective distils image-derived geography without making imagery a target or inference-time input.

Country–location alignment ($\mathcal{L}_{\text{cl}}$) acts within TerraNova's own shared space: it aligns the country's spatiotemporal embedding with the mean embedding of coordinates sampled inside its territory, against other countries as negatives. Sampling density follows $\log(1 + P_y(\mathbf{x}))$ under the population for that year, so the territorial summary concentrates where people live (compressing the heavy-tailed population distribution so that dense cities do not monopolise it). This objective couples the two geometries directly; the external alignment provides the complementary image-derived anchor.

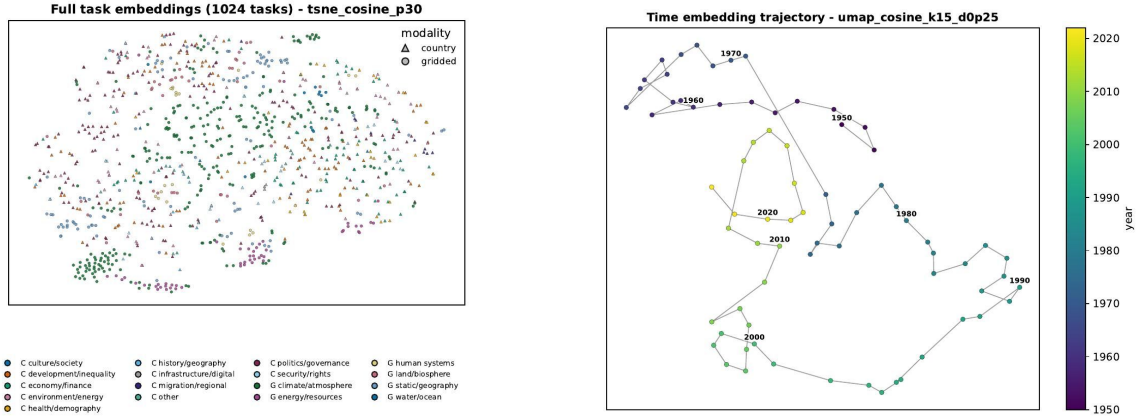
4.3 Optimisation and active task resampling

Each epoch interleaves fixed counts of reconstruction, transport and alignment steps in a globally shuffled schedule. Reconstruction sampling starts from the gridded/country mixture with per-domain floors and bounded task boosts, so the much larger gridded inventory does not crowd out the country indicators. The model trains for 128 epochs with AdamW (Loshchilov and Hutter, 2019) under a warmup–cosine schedule (Loshchilov and Hutter, 2017), gradient clipping, bf16-mixed precision, a batch size of 8,192, and an exponential moving average for evaluation. Every two epochs, *active resampling* re-evaluates every reconstruction task on held-out points and updates its sampling probability from a smoothed mean of the total predictive variance (Eq. 7): tasks the model is most uncertain about receive more training in the next interval. The multiplier is bounded so difficult tasks are boosted without starving the remainder.



(a) Location embedding. (b) A local zoom. (c) The country embeddings.

Figure 4: **(a)** The location embedding on a Robinson projection (PCA, then UMAP McInnes et al. (2018) to three dimensions, mapped into CIELAB Luo (2023)). **(b)** The same at a local zoom. **(c)** The static country embeddings (ICA to three dimensions, then CIELAB).



(a) Task embeddings (t-SNE). (b) Time embeddings (UMAP).

Figure 5: **(a)** A t-SNE (van der Maaten and Hinton, 2008) of the 1,024 task embeddings. **(b)** A UMAP of the year embeddings, 1950 to 2020.

4.4 Adaptation to unseen tasks

The task encoder reserves rows for variables absent from pretraining. For a new target $\tau^\star$, the native adaptation operator jointly optimises a fresh task row and rank-4 MiSS residuals (Kang and Yin, 2026) in the task–spatiotemporal fusion: the input factors are layer-specific while a single output factor is shared across layers and sliced to each map's width, for roughly 4.2×10^5 trainable parameters in total. All pretrained weights and trained task rows remain frozen.

5 Validation

In this section, we evaluate the model design and the learned representations. We test three properties: that the latent space is organised rather than memorised, that the coupling between the two geometries is real and separable into distinct contributions, and that the per-query uncertainty is trustworthy. Please refer to the Supplementary Information for extensive per-component ablations.

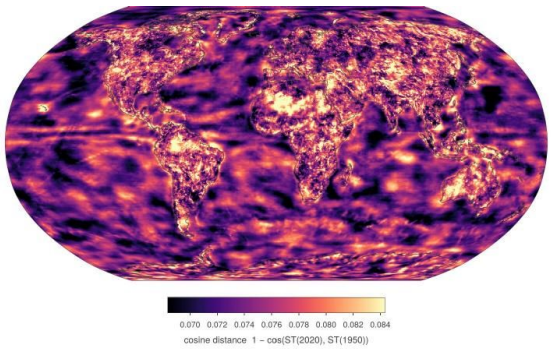


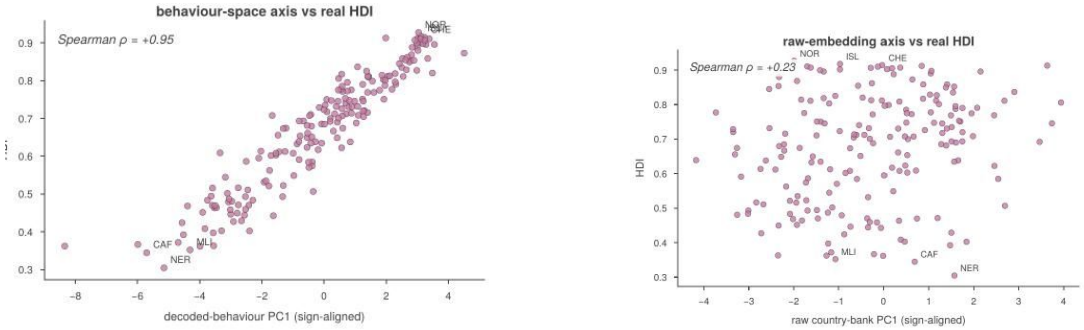
Figure 6: Cosine distance between each place's spatiotemporal representation in 2020 and in 1950.

5.1 Embedding analysis

The encoders recover structure that was never explicitly provided as a target. Projecting the location encoder's output over the globe (Fig. 4) recovers land topography, coastal areas, biome boundaries and coherent ocean basins, and a local zoom shows the same organisation at fine detail. The static country encoder, rendered as a political choropleth with each country coloured by a dimensionality reduction of its embedding, sorts by region. In their own spaces (Fig. 5) the task embeddings cluster by scientific domain while separating the gridded fields from the country indicators, and the time encoder's year embeddings trace a smooth, ordered trajectory from 1950 to 2020. The organisation is continuous: nearby places, regions, domains and years receive nearby codes, and the principal axes of variation align with real geography. The manifold has an intrinsic dimension (Rao et al., 2026) of 9.8–9.9 (FisherS estimator, across land and land–ocean pools).

The representation moves through time Because the state is spatiotemporal, the representation of a fixed place varies across years. The per-cell cosine distance $1 - \cos(\mathrm{ST}(2020), \mathrm{ST}(1950))$ between each place's representation in 2020 and in 1950 (Fig. 6) is largest over populated and developing areas and locations with rapidly changing climate; and smallest over the stable interiors of the oceans, so the change in the representation is concentrated where the observed world changed most, with no explicit change target.

An emergent development axis The embeddings also carry socioeconomic structure, and the construction that recovers it runs through the decoder rather than over the embeddings themselves. We take the learned embedding of each of the 246 countries the model was trained on, decode it through the model's own task-conditioned head into the values it predicts for eleven curated national indicators in 2015, and assemble one row per country. Each column is then standardised across countries, so that indicators on different units contribute comparably, and the Human Development Index and GDP per capita are dropped from the panel before any decomposition, leaving nine indicators: life expectancy, infant and under-five mortality, fertility, urbanisation, internet adoption, the Gini index, carbon footprint and liberal democracy. The leading principal component of that 246×9 matrix carries 54% of its variance and tracks the real Human Development Index at Spearman $|\rho| \approx 0.95$ and real log GDP per capita at ≈ 0.90 (Fig. 7a), neither of which entered the



(a) Decoded behaviour.

(b) Raw embedding.

Figure 7: **An emergent development axis.** The leading principal component of the country representation against the real Human Development Index, which enters neither decomposition. **(a)** From the *decoded* indicator panel. **(b)** From the *raw* embedding vectors.

decomposition. The sign of a principal component is arbitrary, and the figure orients the axis so that it increases with development. Three indicators that remain in the panel, life expectancy and the two mortality rates, bear on the longevity dimension HDI is itself built from, so we repeat the decomposition with them removed. Dropping life expectancy leaves the correlation at 0.95, and dropping the two mortality rates as well, which reduces the panel to the six remaining indicators of demography, infrastructure, inequality, environment and institutions, leaves it at 0.91. The panel contains no education variable at all, so HDI's third dimension is absent throughout. The same construction applied to the raw 64-dimensional country vectors gives a flat spectrum, with a leading component carrying 5% of the variance and correlating with HDI at only 0.23 (Fig. 7b). The development gradient is therefore carried by what the model *predicts* about each country, recovered from the joint structure it has learned across hundreds of national indicators, rather than by the leading direction of variation in the raw embedding.

Embedding arithmetic The learned spaces support vector operations analogous to word embeddings (Mikolov et al., 2013; Almeida and Xexéo, 2019) (Fig. 8). Linear directions carry meaning: projecting the country field onto a rich-minus-poor direction recovers the development axis (AUC 0.98, $\rho \approx 0.65$ against GDP per capita), and projecting the location field onto a coastal-minus-interior direction recovers a coastalness axis (AUC 0.92). Analogies compose in the country space: Cuba minus Russia plus the United States lands near the Philippines, the Falkland Islands and the Faroe Islands. Nearest-neighbour queries return sensible neighbours in all three spaces: the coordinates most similar to Nairobi are Addis Ababa, Bogotá and São Paulo; the cities whose representations co-evolved most with Reykjavík between 1980 and 2020 are Montevideo, Edinburgh and Dublin; and in task space the nearest neighbours of the Gini index are GNI per capita, GDP per capita and the Human Development Index, while those of mixed-forest cover are other vegetation classes. Across spaces, decoded income and life expectancy trace the empirical Preston curve ($\rho \approx 0.74$). These examples indicate that structure in these spaces is accessible through directions and

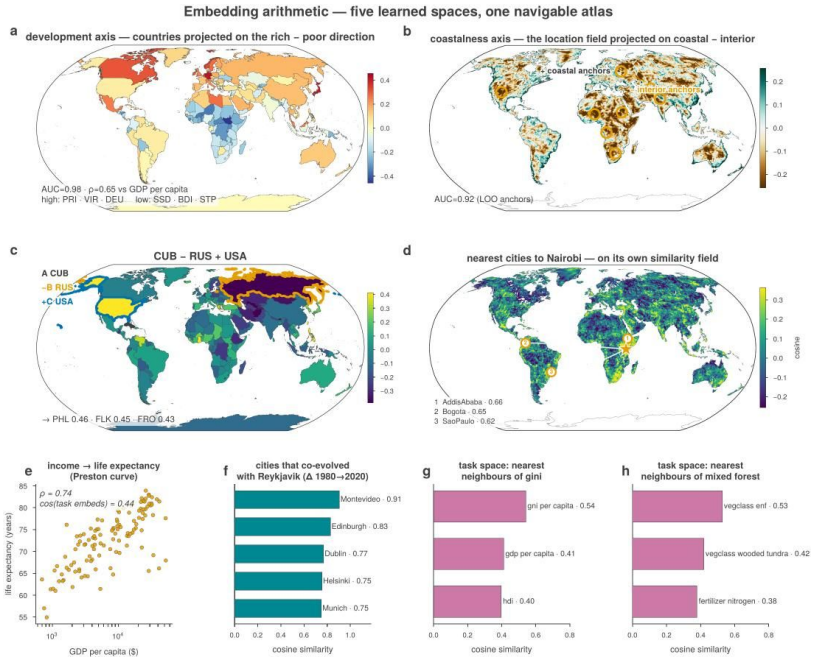


Figure 8: Embedding arithmetic across the model spaces: linear semantic axes, a composed country analogy, nearest-neighbour queries in each space, and a decoded relationship read across spaces.

offsets constructed by hand, which supports use of the representation for exploration as well as prediction.

5.2 Coupling between geometries

We next test whether the coupling between the two geometries, the population-weighted alignment that ties each country to the field over its territory (§4.2), is effective. We first test cross-geometry retrieval: we take the mean embedding of the coordinates inside a country and ask which of a held-out pool of 80 countries it names. TerraNova reaches recall@1 0.875, rising to 1.00 at recall@5 and 1.00 at recall@10. Since the query is a territorial summary rather than a single coordinate, what this measures is whether a country's own code sits where its territory sits. It is a question the geospatial baselines cannot pose at all, since each holds only one geometry, and passing it is evidence that the alignment placed the two geometries in one common latent space rather than leaving them in two unrelated ones. The coupling is soft throughout, encouraged rather than imposed.

To show that the two alignment objectives do separate work, we run an ablation study: a crossed pair of leave-one-out experiments, one dropping the country–location alignment from training, and one dropping the geospatial alignment, otherwise identical, so that each serves as the other's matched control (Fig. 9). Removing the internal alignment removes the cross-geometry capabilities and only those: coordinate-to-country retrieval collapses to chance (0.043) and national-to-gridded downscaling falls to the flat national-mean null, while reconstruction and calibration are untouched (RMSE 0.276 against 0.272, expected

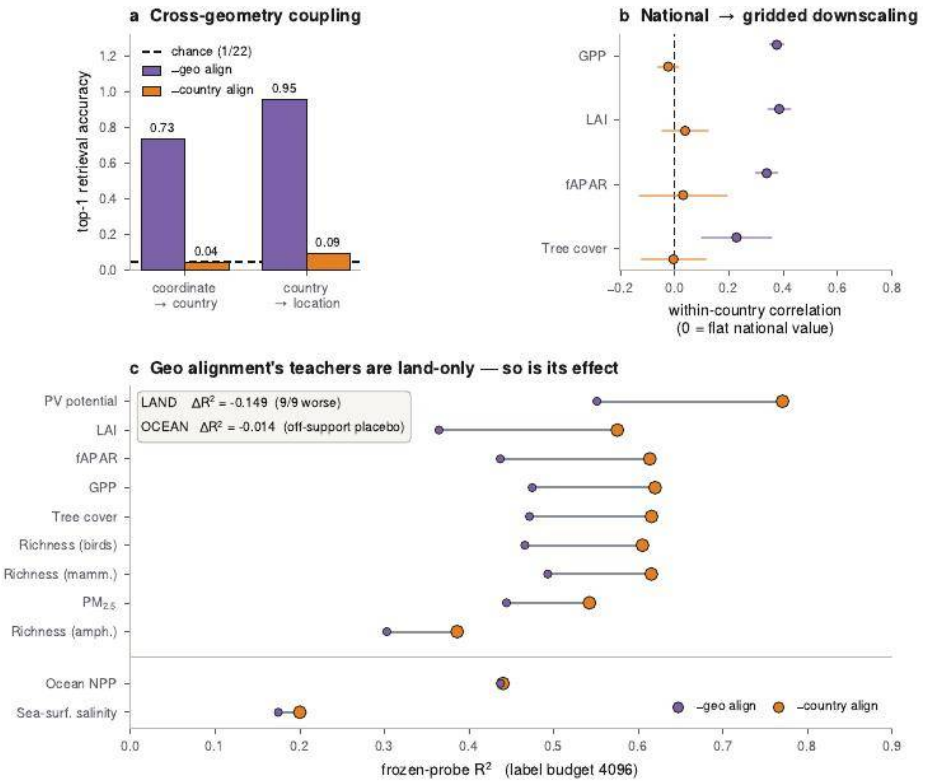


Figure 9: A crossed pair of leave-one-out models, identical except for which alignment objective each drops. **(a)** Cross-geometry retrieval in both directions, read from the native embeddings with no learned probe, over a pool of 22 countries. **(b)** National-to-gridded downscaling on four biosphere targets. **(c)** Frozen-probe reconstruction R^2, split into land and ocean targets.

calibration error 0.173 against 0.172); the two geometries are decoupled, not degraded. Removing the external alignment instead costs representation quality, and it does so on the support of the teacher banks it borrows from: -0.149 R^2 on all nine *land* targets and essentially nothing (-0.014) on the two *ocean* targets. Because the three teacher banks are land-only by construction, the ocean targets act as a support-matched placebo. One objective builds the common space; the other enriches it. This double dissociation is a methodological finding of the paper: two objectives drawing on different observations of one world contribute different, separable structure to the representation.

5.3 Uncertainties

Every TerraNova query returns a NIG predictive distribution rather than a point estimate, and how far that distribution can be trusted depends on which of two properties is asked for. Its ordering is informative and its scale is not. On unseen variables the raw evidential interval is systematically too wide, which we correct with one recalibration constant per task: a country-split conformal interval restores nominal 90% coverage on held-out years. Every interval we report is recalibrated. Mapping the predictive uncertainty per domain

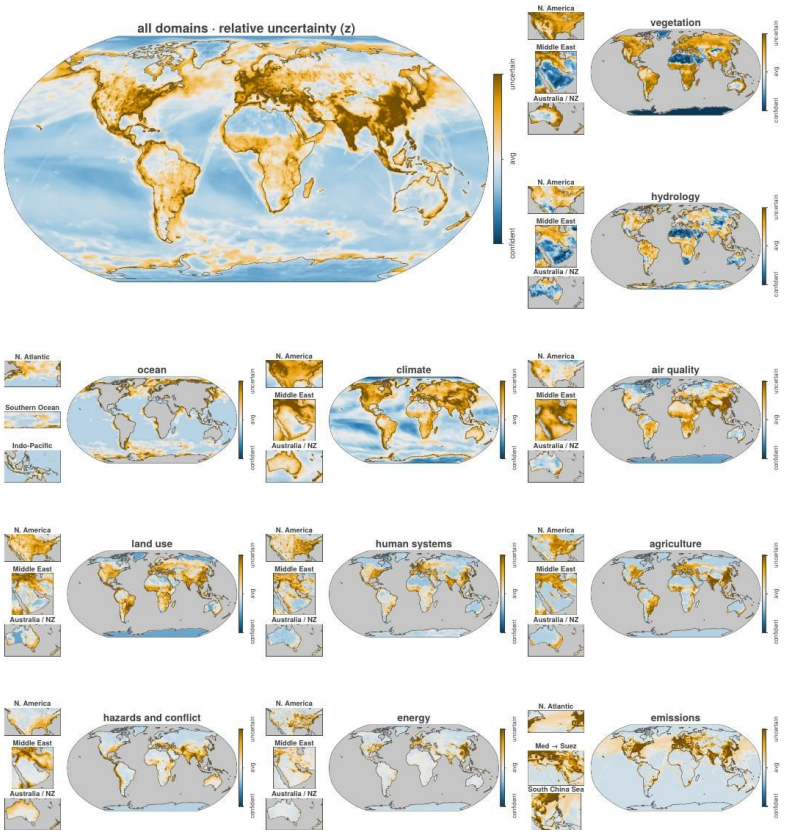


Figure 10: Per-domain maps of predictive uncertainty, each standardised within its own domain.

(Fig. 10), standardised within each domain so that domains on very different scales can be compared, gives an atlas of where the model is least certain. Its geography is one of difficulty rather than of coverage: the width is largest where a field is strongly structured, over China, western Europe and the eastern United States for emissions, over the Congo and Amazon basins for vegetation and over the Eurasian interior for climate, and it is smallest over more homogeneous places such as the Sahara, the Australian desert and the open subtropical gyres.

6 Results

In this section, we measure what TerraNova delivers on held-out data: static prediction against purpose-built geospatial encoders (§6.1), reconstruction and nowcasting of national records through time (§6.2), dense fields from sparse measurements (§6.3), national cross-sections from a few reporting countries (§6.4), national-to-gridded downscaling (§6.5) and what all of this costs to train and to run (§6.6). Splits are fixed across methods and label budgets, regression is scored by held-out R^2 and Pearson correlation, classification by accuracy, and predictive distributions by coverage and sharpness.

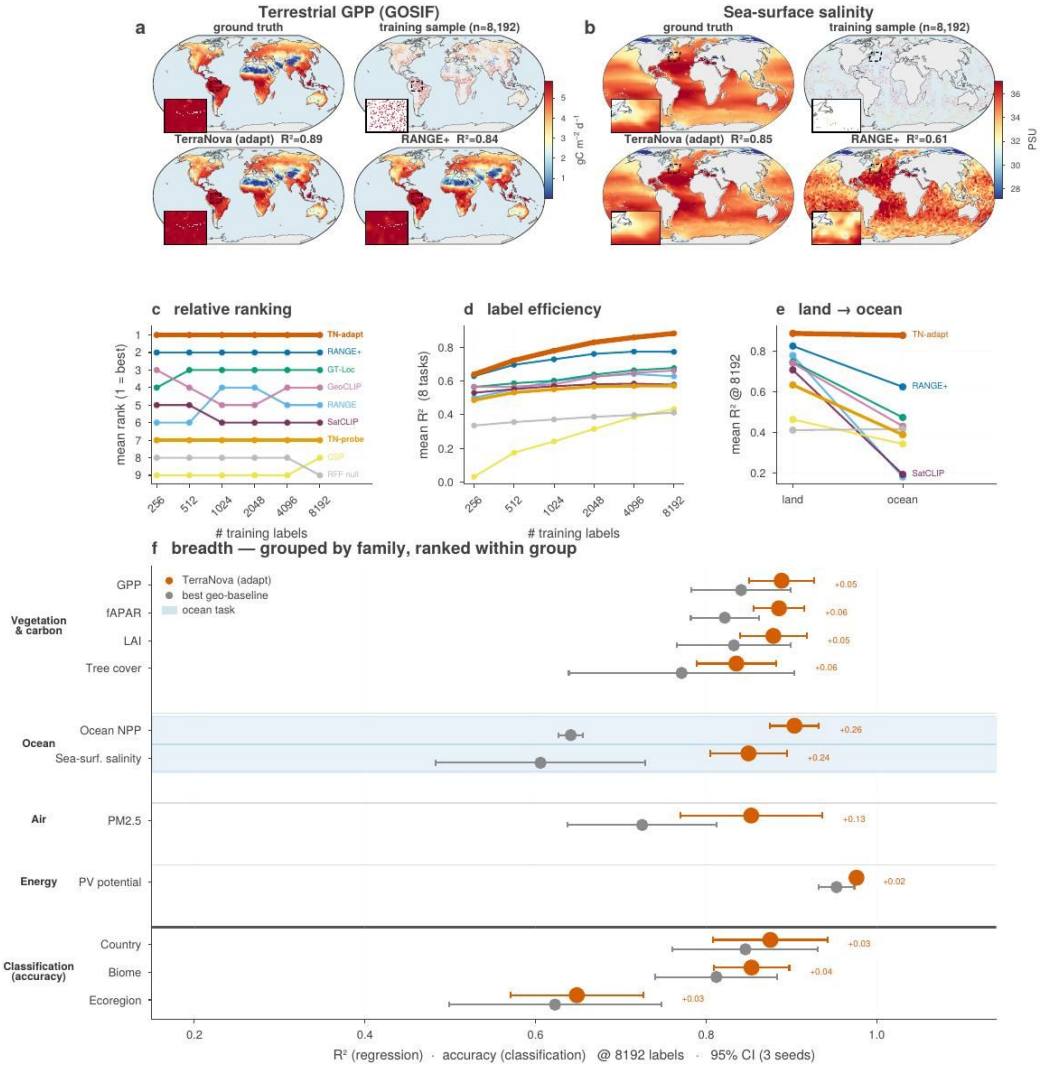


Figure 11: **Comparison with existing geospatial embeddings, on unseen data. a**,**b**, Two targets at 8,192 labels: truth, training sample, TerraNova's reconstruction and the strongest baseline. **c**, Mean rank against label budget, where TN-adapt is the decoder read-out and TN-probe the frozen-embedding read-out of the same model. **d**, Mean held-out R^2, and **e**, the same split into land and ocean. **f**, Per-target results against the best baseline for that target.

6.1 Comparison with existing geospatial embeddings

In this benchmark, we hold out eleven environmental and eco-geographic targets, eight scored by R^2 and three by accuracy, and compare nine read-outs at label budgets from 256 to 8,192 over three seeds (Fig. 11). Six models are published location encoders read out by a probe on their frozen embeddings, one is a random-Fourier-feature null, and two are TerraNova: a probe on its frozen embedding, which is the like-for-like comparison, and its own decoder with a fresh task row and a rank-4 MiSS adapter, the recommended reuse path. The adapted

read-out receives coordinates only and no time, so its margin cannot come from temporal information the baselines cannot express.

MiSS-adapted TerraNova reaches a mean held-out R^2 of 0.88 against 0.77 for the strongest baseline (RANGE+ Dhakal et al. (2025)), and it is at or above the best geospatial baseline on every target, by 0.02–0.26 (Fig. 11d,f). The ordering depends on the label budget only at the low end, where the adapter has little to fit: by mean rank RANGE+ leads at 256 labels, and from 512 upwards TerraNova is best at every budget (Fig. 11c). What delivers that lead is the cheaply-adaptable decoder rather than the raw geometry. Probed linearly, TerraNova's own embedding ranks seventh of the nine, and the margin appears only once a fresh task row and rank-4 MiSS residuals are fitted, at 416,000 trainable parameters per task. The margin is widest where the baselines have no training signal. On the two marine targets TerraNova reaches 0.88, while the image-trained encoders lose 0.12–0.60 R^2 between land and ocean and the best of them reaches 0.62 (Fig. 11e). Time is another axis these encoders cannot express, and we explore it next.

6.2 Temporal interpolation and nowcasting

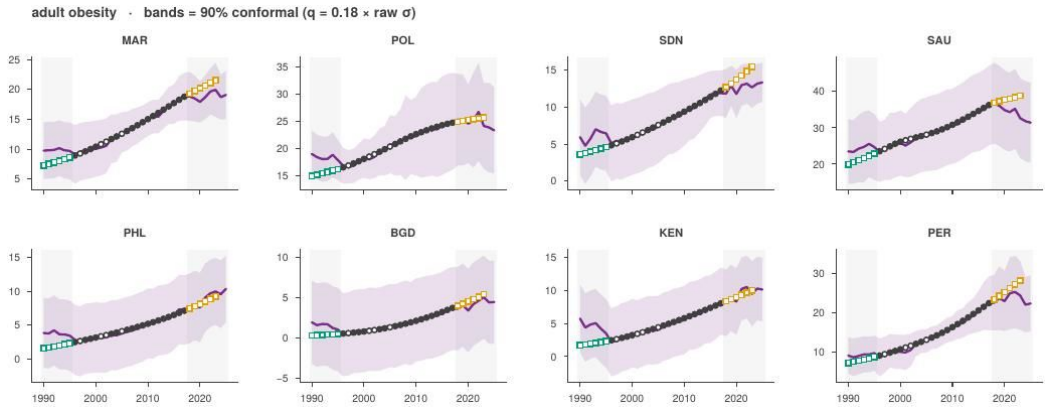


Figure 12: **Reconstructing and nowcasting.** Filled circles are training years, open circles held-out years inside the window and open squares the held-out blocks at either end; the line is the predicted mean and the band the conformal 90% interval.

In this experiment, we aim to assess the temporal capabilities of TerraNova. Each national record (unseen during training) is split by holding out its earliest and latest years, and the model is adapted on the years in between by fitting a fresh task row together with rank-4 MiSS residuals on the frozen backbone, then asked to predict both held-out ends (Fig. 12). Across the battery of 29 indicators the mean short-range anomaly correlation is 0.74 and the median 0.79, ranging from near-perfect on smooth series to near-zero on intrinsically noisy ones. Skill is highest at the training-window edge (0.79) and decays with lead time, backcasts holding 0.60 six years into the past and forecasts 0.49 six years ahead. A time-aware trend extrapolation is a closer reference: TerraNova leads it at the training-window edge, matches it throughout the backcast direction, and falls behind it in the forecast direction beyond about two years. The evidential interval over-covers on the held-out years, which we correct with country-split conformal recalibration. The same fit reconstructs three national cross-sections

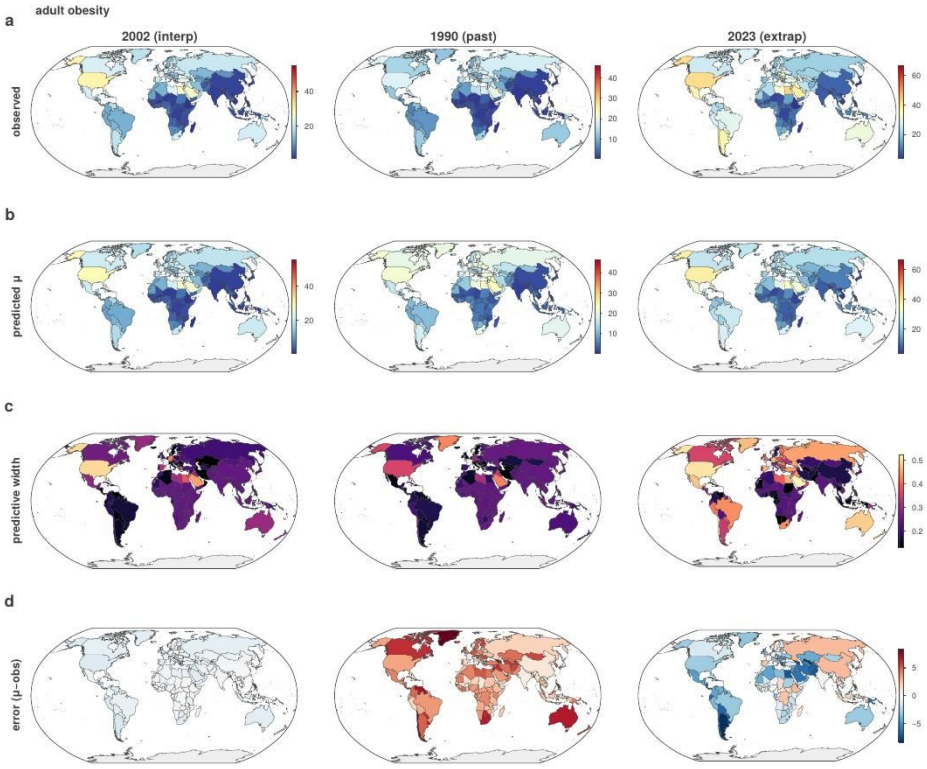


Figure 13: The fit of Fig. 12 read out over every country at three held-out years. Rows are **a**, the observation, **b**, the predicted mean, **c**, the predictive width and **d**, the error; columns are an interpolated year, a year before the record starts and a year after it ends.

from one adapter (Fig. 13). The between-country pattern is recovered almost exactly in an interpolated year (level correlation 0.999, mean absolute error 0.6 percentage points) and holds at both extrapolated ends (0.98, 3–4 points), and the predictive width is 26% wider at the forecast year than at the interpolated one. One frozen model serves indicators as different as scientific output, tuberculosis incidence, adult obesity and patent applications.

6.3 Gridded reconstruction from sparse measurements

We now assess whether TerraNova can reconstruct a dense 0.25° field from a sparse lattice of measured cells, for variables it never saw in training. Parameter-efficient adaptation of the frozen backbone holds up at measurement densities where classical interpolation degrades sharply. From 1/1024 of the cells, roughly one measurement every eight degrees, the target shown is recovered at a held-out R^2 of 0.94 and 23 dB, scored under the held-out new-region split of §4 (Fig. 14a). Pooled over the leakage-free targets, each scored with the best of six adaptation read-outs against the best of five interpolators, the two approaches are close where measurements are dense (0.97 against 0.95 at one cell in sixteen) and the margin widens monotonically as the lattice thins, reaching 0.74 against 0.56 at one cell in 4,096 (Fig. 14b). Unlike interpolation, the model attaches a per-cell predictive width to every value

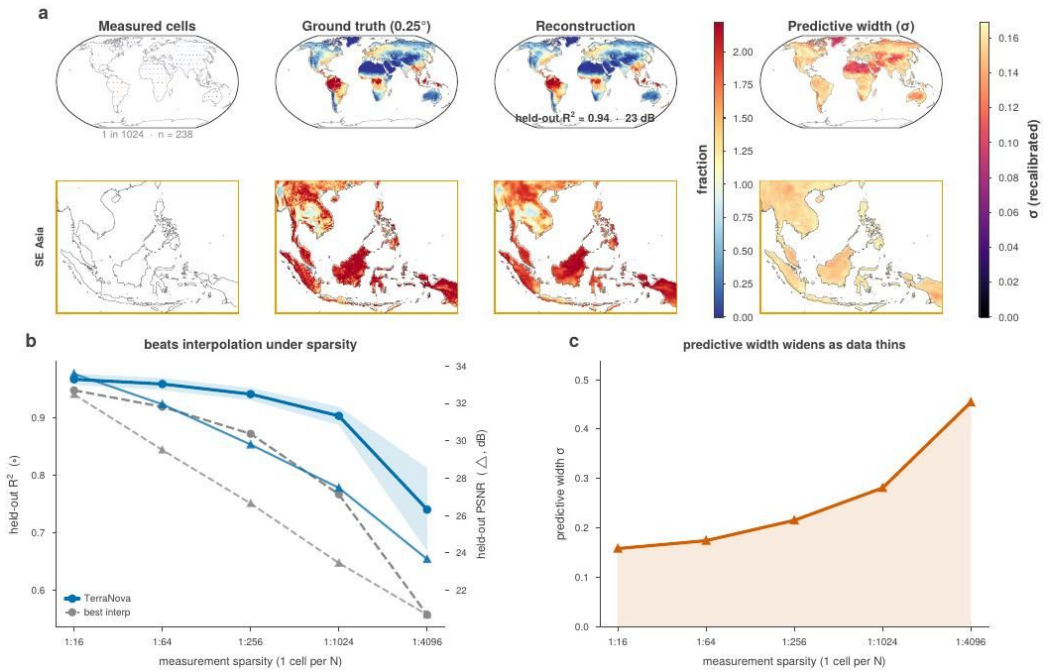


Figure 14: **a**, A dense 0.25° field reconstructed from a sparse lattice of measured cells (measured | truth | reconstruction | predictive width σ). **b**, Held-out R^2 (circles, left axis) and PSNR (triangles, right axis) against measurement sparsity, for the best adaptation read-out and the best classical interpolator. **c**, The predictive width as a fraction of the target's standard deviation.

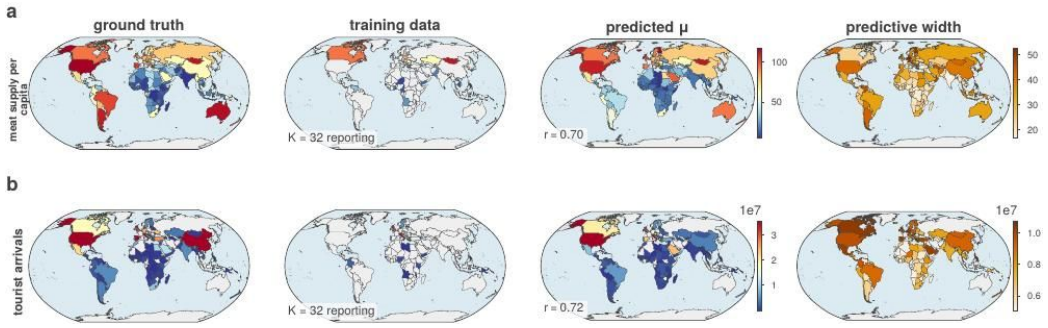


Figure 15: **Few nations in, all nations out.** Two indicators reconstructed from 32 reporting countries: truth, the revealed countries, the predicted national level and the predictive width. **a**, Meat supply per capita. **b**, Tourist arrivals. Maps show one seed; the text quotes three-seed means.

it fills in, widening from 0.16–0.45 of the target's standard deviation as the lattice thins (Fig. 14c), and it stays calibrated as it widens.

6.4 Country-level extrapolation

We use the same operator to reconstruct a national cross-section from a few reporting countries. For each held-out indicator we keep its best-covered year, hide 30% of countries,

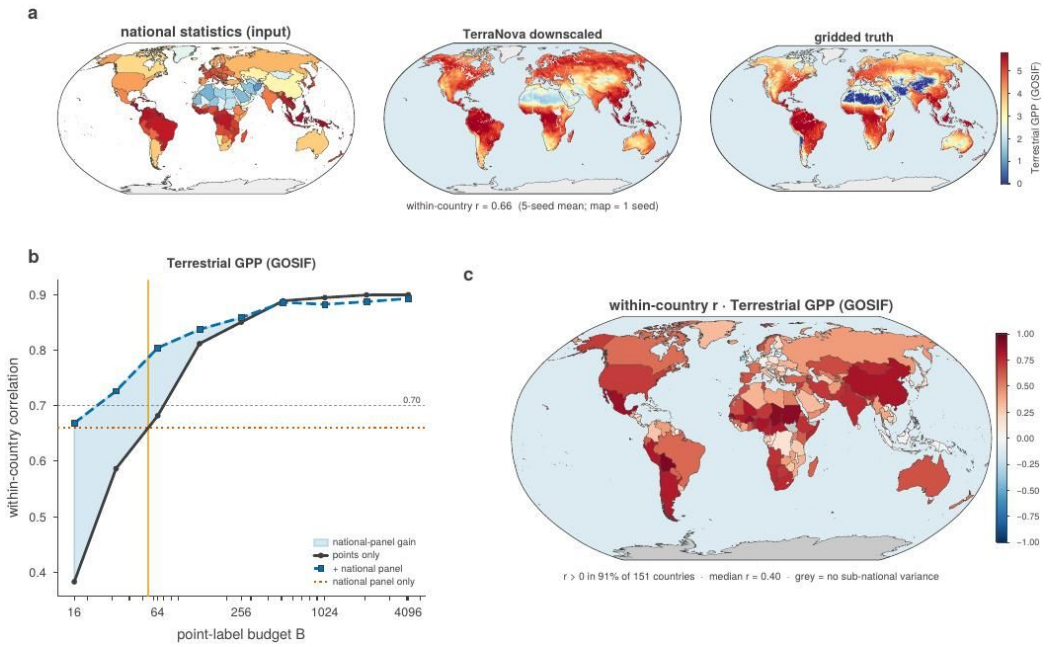


Figure 16: **From national panels to gridded fields. a**, Terrestrial GPP downscaled from national values: the national input, TerraNova's gridded reconstruction and the gridded truth. **b**, National panels against gridded point labels at matched budgets. **c**, Within-country correlation per country; grey marks countries with no sub-national variance.

reveal the value for K of the rest, fit a fresh task row with rank-4 MiSS residuals (Kang and Yin, 2026) on the country route, and predict the hidden national levels. At $K = 32$, roughly one country in seven, held-out level correlation reaches 0.65 for meat supply per capita and 0.71 for tourist arrivals (Fig. 15), and pooled over the 31 indicators of the battery it climbs across 0.17–0.51 as the reporting budget grows from four countries to 128. A frozen-embedding probe of the same cross-sections trails the adapted read-out at every budget. The natural baseline is spatial interpolation between the reporting countries, and the strongest of the four tried is inverse-distance weighting over country centroids, at 0.13–0.47 over the same budget range: a handful of points spread over the globe is too little to fit a Gaussian-process length scale. The learned representation and spatial interpolation carry different information about a national cross-section. What separates them is spatial smoothness: TerraNova's margin over interpolation anti-correlates with how well interpolation itself does ($r = -0.57$ across indicators), so the representation is far ahead where a country's value is not predictable from its neighbours, as for tourist arrivals (0.71 against 0.22), and behind where it is. Their held-out errors decorrelate as coverage grows, from 0.81 at four reporting countries to 0.61 at 128. Neither arm dominates: TerraNova leads at four, sixteen, sixty-four and 128 revealed countries, interpolation leads at eight and thirty-two, and per indicator at $K = 32$ TerraNova is ahead on 7 of 31. We therefore report the reconstruction as a capability of the coupled representation and the combination as the way to use it, not as a claim on spatial statistics' own ground.

Table 1: Every read-out on the frozen backbone at a 300-label budget on one laptop GPU. The shaded row is the method used throughout the paper and the last row repeats it on the machine's CPU; energy is omitted where it cannot be measured.

		gridded field			national indicator	
Read-out	Trainable	Fit (s)	Energy (kJ)	Skill	Fit (s)	Skill
Linear probe	256	0.3	n/a	0.68 ± 0.03	< 0.1	0.46 ± 0.12
Task row only	256	50.2	3.1	0.83 ± 0.02	46.2	0.15 ± 0.12
VeRA Kopiczko et al. (2024)	101,324	55.6	3.2	0.90 ± 0.05	49.6	0.46 ± 0.24
MiSS Kang and Yin (2026)	416,000	53.3	3.3	0.95 ± 0.00	49.5	0.60 ± 0.12
LoRA Hu et al. (2022)	803,072	53.4	3.3	0.94 ± 0.02	49.2	0.62 ± 0.14
Final-block fine-tuning	25,717,508	56.6	3.3	0.96 ± 0.00	52.0	0.66 ± 0.12
MiSS on CPU	416,000	599.7	n/a	0.95 ± 0.01	538.1	0.60 ± 0.18

6.5 Gridded field estimation from country-level values

The coupling between each country and the field over its territory can be read in the other direction: a quantity known only at the national level can be downscaled into a gridded field. Adapting the same MiSS operator, TerraNova recovers sub-national structure on four biosphere targets (GPP, fAPAR, LAI and tree cover) at seed-mean within-country correlations of 0.51–0.66 against a label-shuffle null of at most 0.10 (Fig. 16a). It is not carried by a few large countries: across the four targets the correlation is positive in 91–96% of the 149–154 countries scored, with medians of 0.30–0.40 (Fig. 16c). Measured in labels, the national panel is worth about 55 gridded point labels for this target, and its advantage closes once a few hundred points are available (Fig. 16b). We report this as the spatial disaggregation the coupled representation makes possible, not as a validated product. Note that this sort of downscaling is not always meaningful (e.g. governance variables may not be downscaled in a intelligible way).

6.6 Computational cost

We now report the computational cost of training the model once and of using it afterwards. Adaptation and inference are measured on a consumer laptop, with a mobile RTX 5070 Laptop with 8 GB of memory, and on the same machine's CPU. Wall-clock is a median over three seeds and skill a three-seed mean with its standard deviation; energy is integrated from device power sampling with the idle baseline subtracted.

Training. One production run of 128 epochs reached 4.23 million optimiser steps and 34 billion training samples on a single H100 PCIe, at 3.6 steps per second end to end. Training takes about 327 GPU-hours, 128 kWh and 42 $kgCO_2e$, estimated using CodeCarbon Courty et al. (2024).

Adaptation. Attaching a new variable takes seconds to minutes on a laptop (Tables 1 and 2a). At a given label budget every gradient read-out costs the same to within a few seconds, although they span 256 to 25.7 million trainable parameters: the backward pass through the frozen backbone dominates. On gridded fields the linear probe reaches 0.68

Table 2: **(a)** Fit wall-clock on the gridded route against budget K; the shaded row is the operator used throughout the paper and the italic row repeats it on CPU. **(b)** Inference throughput against batch size B. A single query is omitted, since it measures dispatch latency rather than the model.

(a) Fit wall-clock (s), gridded route					
Read-out	K=30	100	300	1,000	3,000
Linear probe	0.3	0.3	0.3	0.3	0.4
Task row only	14.2	22.7	50.2	147.3	148.3
VeRA Kopiczko et al. (2024)	17.9	25.4	55.6	163.8	174.3
MiSS Kang and Yin (2026)	16.9	24.7	53.3	161.9	163.6
LoRA Hu et al. (2022)	16.4	24.4	53.4	157.3	158.2
Final-block fine-tuning	20.2	28.2	56.6	158.1	157.3
MiSS on CPU	104.4	246.8	599.7	1,428.8	1,460.2
(b) Inference throughput (queries s^{-1})					
Route and device	B=8	64	512	4,096	16,384
Coordinate, GPU	856	5,034	9,770	10,000	9,092
Country, GPU	1,199	6,153	10,708	10,229	10,152
Coordinate, CPU	150	484	732	778	765
Country, CPU	139	396	625	868	875

against 0.95 for the adapted decoder, so most of the added skill costs almost nothing. On national indicators the ordering differs: the probe reaches 0.46 and beats optimising the task row alone, which never exceeds 0.25 at any budget, and only the weight-space adapters clear it, at 0.60 to 0.66. The same fit on the CPU costs 6.2 to 11.2 times the wall-clock depending on budget and geometry, 11.2 times on the gridded task at this budget, for skill inside the seed spread of the GPU fit. MiSS is the option used throughout the paper because of its efficiency: it matches LoRA on both geometries with half the trainable parameters and a tighter seed spread on the gridded route, and gives up 0.005 and 0.056 against final-block fine-tuning while training 1.6% as many parameters.

Inference. A forward pass costs 713 MFLOP for a coordinate query and 625 MFLOP for a country query, of which the task–spatiotemporal fusion accounts for 57% and the temporal fusion a further quarter, against 366 million parameters in the model. Throughput saturates on the GPU by a batch of 512 at about 10,000 queries per second, and on the CPU by a batch of 4,096 at about 800, a twelve-fold gap (Table 2b). Single-query latency is dominated by dispatch rather than by the model, at 9.6 ms against 102 μs per query inside a batch of 512. The largest batch tested, 16,384 queries, peaked at 3.1 GiB of allocated tensor memory.

7 Discussion

Training one model on 1,024 variables in two geometries gives the representation properties that neither a gridded geophysical model nor a country-level panel has on its own.

The sparse-reconstruction result is the sharpest of them. A model that has seen a thousand variables over the whole planet *knows* what a plausible field looks like before it sees a single measurement of a new one, so a handful of observations is enough to place that field in a structure the model already holds. The same prior is what makes the latent space readable rather than a lookup table: geography, biome structure and coherent ocean basins are legible in the location embedding, and a development gradient emerges from what the model predicts about each country once the development indicators themselves are removed from the projection. Fields, records and image-derived embeddings admit a compatible representation, although here that compatibility is built by explicit alignment rather than emerging from data alone. The coupling buys capabilities that neither geometry expresses alone, and cross-geometry retrieval and national-to-gridded downscaling are one coupling read in two directions: retrieval names the country a territory belongs to, downscaling writes a national quantity back onto the grid, both supported by the same population-weighted alignment. The ablation shows that the two alignment objectives do separable work, one building the common space and the other enriching it.

Uncertainty is a first-class output. The width responds to evidence, widening as a measurement lattice thins and as a query moves beyond the years a record covers, so a model that reports where it is uncertain flags its own extrapolation instead of failing silently. Cheap adaptation follows from the same architectural choice that produces the benchmark margin: because the decoder is generated per query from the task and the spatiotemporal state, a new variable is reached inside that fusion without touching any pretrained task.

Limitations. The comparison against specialised geospatial encoders is not capacity-matched (§6.1). The lead is delivered by the task-conditioned decoder and the fusion it adapts rather than by the frozen embedding, and at the smallest label budgets the strongest baseline is still ahead. Downscaling is a capability of the coupled representation rather than a validated product, and it recovers nothing where a quantity barely varies within countries (§6.5). The socioeconomic and institutional inputs are observational and carry historical and measurement bias that a learned representation can propagate; predictions for under-observed countries carry the widest intervals and should be read with them attached. TerraNova is a representation layer, not a causal model of climate damages, an integrated-assessment model, or a replacement for process-based simulation. It does not identify causal effects, simulate climate-society feedbacks or project policy scenarios, and reading its predictions as if it did would be a misuse. What it provides is a joint observational feature layer in which transfer, spatial inference with uncertainty, cross-domain prediction and hypothesis generation can be studied in one latent space, alongside the causal and mechanistic tools that ultimately answer climate-society questions.

Outlook. The model treats time as an additional input variable, which limits its ability to learn conditioned dynamics, such as the evolution of an economy under different radiative-forcing assumptions. Lifting TerraNova into a representation better suited to those dynamics could enable high-resolution, uncertainty-aware scenario generation. Beyond the three modalities used here, LLM-derived embeddings of locations, cultures or events could carry information that is not directly quantifiable, although this should be done with the utmost care over bias, transparency, and representation of the Global South. Finally, TerraNova implicitly assumes that countries are somewhat independent, since we do not explicitly

model bilateral trade or migration networks; graph neural networks over migration and trade networks, which predict regional economic outcomes at the sub-national scale (Xu et al., 2020), could extend it to cascading climate and economic impacts between countries. These extensions sit on the same foundation this paper sets out: making the coupled data streams of the Anthropocene learnable, transferable and queryable within one model, cheaply enough to put it within reach of groups far from the one that trained it. That is the integrated, cross-disciplinary modelling named a central goal for artificial intelligence in climate research (Ou et al., 2026): not a coupling of separate models, but a shared representation in which local environmental structure and national outcomes inform one another directly.

Broader Impact Statement

TerraNova lowers the cost of using a planetary-scale model. A new variable is fitted with a compact adapter on a frozen backbone, so a group with a regional survey or a bespoke indicator can specialise the representation on ordinary hardware instead of depending on a supercomputer. This matters most in the data-poor settings where such models are hardest to train from scratch. Calibrated uncertainties are the other half of that story: a model that reports where it is uncertain is harder to trust blindly than a bare point predictor. The limit of this argument is that cheap adaptation distributes the product of a planetary representation while leaving the means of producing one, the corpus, the compute and the labour that assembles both, concentrated where they already were Mohamed et al. (2020); Crawford (2021).

The risks specific to this model follow from its breadth. Because TerraNova learns across such a wide variety of tasks, its representation reaches well beyond physical fields into human development, governance and institutions, and it may learn associations between environmental conditions and national outcomes in a world whose present distribution of those outcomes was produced by history: colonisation, imperialism, extractive economic practices, wars, occupations, trade structure and policy decisions. The model observes none of that history explicitly and does not represent it directly. Its representation therefore recovers real regularities whose causes lie outside its inputs, and the development axis is a statistical description of that confounded world rather than an explanation of it. Reading a learned environment-development association as though geography produced development would recast institutional and historical processes Acemoglu et al. (2001) as natural determinism: a history of relations between places would appear instead as a property of the places themselves, with the relations that produced it no longer visible in the result Winner (1980). We report the axis as an insight about the representation, not as a finding about why nations differ.

The corpus carries a viewpoint of its own. National indicators are constructed rather than observed, the product of statistical capacity that is itself unequally distributed, and governance and institutional quality are the hardest of them to pin down. The construct has to be defined before it can be scored, that definition carries a normative model of what institutions ought to look like, and the resulting indices are assembled largely by institutions of the Global North. Rendering such heterogeneous records commensurable, as our per-task standardisation does, is a convenience for training and also a substantive operation: quantities produced under different conditions, by different institutions, for different purposes, are placed

on one scale and made available to a single objective Espeland and Stevens (1998). Coverage is unequal for historical reasons, so the model's ignorance is not randomly distributed: it is greatest where its predictions are most likely to be sought. This caution extends across every societal variable the model spans.

The same caution should apply to the model's outputs. A downscaled field is a prediction carrying the visual authority of a measurement, and a caveat in a manuscript does not travel with a raster. Downscaling recovers structure only to the extent that a quantity varies within countries at all, and it fails quietly otherwise, painting a plausible map for a target it cannot resolve. Predictions are least constrained where observations are sparsest, which is also where they are most likely to be taken as the only estimate available. Such surfaces are attractive precisely because they render populations legible to administration at a resolution at which they were never measured Scott (1998), and that legibility is not an end in itself but an input to distribution: climate finance and loss-and-damage eligibility, humanitarian targeting, subnational resource allocation. That asymmetry is why per-cell uncertainty is part of the prediction here rather than a diagnostic beside it, and why gridded socioeconomic outputs from this model should not be used directly for allocation, targeting or eligibility decisions about the places they describe. Where such uses are contemplated we advise domain expertise and co-design with policy-makers and affected communities, while noting that participation is not a design fix and becomes participation-washing where it is short-term or extractive Sloane et al. (2022). Cheap adaptation can also be problematic. The backbone accepts any new task, so the mechanism that lets, for instance, a public-health group attach a regional indicator also lets anyone attach a target we neither anticipated nor evaluated, inheriting the representation's biases without inheriting its evaluation. Upon acceptance, we will release the model weights and configuration, evaluation code and tutorials under permissive licenses.

Acknowledgments and Disclosure of Funding

Carlos Rodriguez-Pardo and Massimo Tavoni acknowledge support from the European Research Council, ERC grant agreement number 101044703 (EUNICE), CUP D87G22000340006.

References

Daron Acemoglu, Simon Johnson, and James A. Robinson. The colonial origins of comparative development: An empirical investigation. *American Economic Review*, 91(5):1369–1401, 2001. doi: 10.1257/aer.91.5.1369.

Mohit Agarwal, Mimi Sun, Chaitanya Kamath, Arbaaz Muslim, Prithul Sarker, Joydeep Paul, Hector Yee, Marcin Sieniek, Kim Jablonski, et al. General geospatial inference with a population dynamics foundation model. *arXiv preprint arXiv:2411.07207*, 2024.

Anna Allen, Stratis Markou, Will Tebbutt, James Requeima, Wessel P. Bruinsma, Tom R. Andersson, Michael Herzog, Nicholas D. Lane, Matthew Chantry, J. Scott Hosking, and Richard E. Turner. End-to-end data-driven weather prediction. *Nature*, 641(8065):1172–1179, 2025. doi: 10.1038/s41586-025-08897-0.

Felipe Almeida and Geraldo Xexéo. Word embeddings: A survey. *arXiv preprint arXiv:1901.09069*, 2019.

Alexander Amini, Wilko Schwarting, Ava Soleimany, and Daniela Rus. Deep evidential regression. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 14927–14937, 2020.

Lint Barrage and William D Nordhaus. Policies, projections, and the social cost of carbon: Results from the DICE-2023 model. *Proceedings of the National Academy of Sciences*, 121 (13):e2312030121, 2024. doi: 10.1073/pnas.2312030121.

Kaifeng Bi, Lingxi Xie, Hengheng Zhang, Xin Chen, Xiaotao Gu, and Qi Tian. Accurate medium-range global weather forecasting with 3D neural networks. *Nature*, 619(7970): 533–538, 2023. doi: 10.1038/s41586-023-06185-3.

Cristian Bodnar, Wessel P. Bruinsma, Ana Lucic, Megan Stanley, Anna Allen, Johannes Brandstetter, Patrick Garvan, Maik Riechert, Jonathan A. Weyn, Haiyu Dong, et al. A foundation model for the Earth system. *Nature*, 641(8065):1180–1187, 2025. doi: 10.1038/s41586-025-09005-y.

Christopher F. Brown, Michal R. Kazmierski, Valerie J. Pasquarella, William J. Rucklidge, Masha Samsikova, Chenhui Zhang, Evan Shelhamer, Estefania Lahera, Olivia Wiles, Simon Ilyushchenko, Noel Gorelick, Lihui Lydia Zhang, Sophia Alj, Emily Schechter, Sean Askay, Oliver Guinan, Rebecca Moore, Alexis Boukouvalas, and Pushmeet Kohli. AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data. *arXiv preprint arXiv:2507.22291*, 2025. URL https://arxiv.org/abs/2507.22291.

Marshall Burke, Solomon M Hsiang, and Edward Miguel. Global non-linear effect of temperature on economic production. *Nature*, 527(7577):235–239, 2015. doi: 10.1038/nature15725.

Leonardo Chiani, Emanuele Borgonovo, Elmar Plischke, and Massimo Tavoni. Global sensitivity analysis of integrated assessment models with multivariate outputs. *Risk Analysis*, 45(8):2132–2156, 2025. doi: 10.1111/risa.70002.

Clay Foundation. Clay Foundation Model: An open source AI model for Earth, 2024. URL https://github.com/Clay-foundation/model. Version 1.5.

Benoît Courty, Victor Schmidt, Sasha Luccioni, Goyal-Kamal, Marion Coord, Boris Feld, et al. mlco2/codecarbon: v2.4.1, 2024. URL https://doi.org/10.5281/zenodo.11171501.

Kate Crawford. *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. Yale University Press, New Haven, CT, 2021. ISBN 978-0-300-20957-0. doi: 10.12987/9780300252392.

Paul J Crutzen. Geology of mankind. *Nature*, 415(6867):23, 2002. doi: 10.1038/415023a.

Aayush Dhakal, Srikumar Sastry, Subash Khanal, Adeel Ahmad, Eric Xing, and Nathan Jacobs. RANGE: Retrieval augmented neural fields for multi-resolution geo-embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 24680–24689, 2025. doi: 10.1109/CVPR52734.2025.02298.

Noah S Diffenbaugh and Marshall Burke. Global warming has increased global economic inequality. *Proceedings of the National Academy of Sciences*, 116(20):9808–9813, 2019. doi: 10.1073/pnas.1816020116.

Wendy Nelson Espeland and Mitchell L. Stevens. Commensuration as a social process. *Annual Review of Sociology*, 24:313–343, 1998. doi: 10.1146/annurev.soc.24.1.313.

Veronika Eyring, Sandrine Bony, Gerald A. Meehl, Catherine A. Senior, Bjorn Stevens, Ronald J. Stouffer, and Karl E. Taylor. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, 9(5):1937–1958, 2016. doi: 10.5194/gmd-9-1937-2016.

Ajay Gambhir, Isabela Butnar, Pei-Hao Li, Pete Smith, and Neil Strachan. A review of criticisms of integrated assessment models and proposed approaches to address these, through the lens of BECCS. *Energies*, 12(9):1747, 2019. doi: 10.3390/en12091747.

Rohit Girdhar, Alaaeldin El-Nouby, Zhuang Liu, Mannat Singh, Kalyan Vasudev Alwala, Armand Joulin, and Ishan Misra. ImageBind: One embedding space to bind them all. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15180–15190, 2023.

David Ha, Andrew M. Dai, and Quoc V. Le. Hypernetworks. In *International Conference on Learning Representations (ICLR)*, 2017.

Ed Hawkins and Rowan Sutton. The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, 90(8):1095–1108, 2009. doi: 10.1175/2009BAMS2607.1.

Ed Hawkins and Rowan Sutton. The potential to narrow uncertainty in projections of regional precipitation change. *Climate Dynamics*, 37(1-2):407–418, 2011. doi: 10.1007/s00382-010-0810-6.

Solomon Hsiang, Robert Kopp, Amir Jina, James Rising, Michael Delgado, Shashank Mohan, D. J. Rasmussen, Robert Muir-Wood, Paul Wilson, Michael Oppenheimer, Kate Larsen, and Trevor Houser. Estimating economic damage from climate change in the United States. *Science*, 356(6345):1362–1369, 2017. doi: 10.1126/science.aal4369.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations (ICLR)*, 2022. arXiv:2106.09685.

Minyoung Huh, Brian Cheung, Tongzhou Wang, and Phillip Isola. Position: The Platonic Representation Hypothesis. In *Proceedings of the 41st International Conference on Machine Learning (ICML)*, volume 235 of *Proceedings of Machine Learning Research*, pages 20617–20642. PMLR, 2024.

Andrew Jaegle, Sebastian Borgeaud, Jean-Baptiste Alayrac, Carl Doersch, Catalin Ionescu, David Ding, Skanda Koppula, Daniel Zoran, Andrew Brock, Evan Shelhamer, Olivier J. Hénaff, Matthew M. Botvinick, Andrew Zisserman, Oriol Vinyals, and João Carreira. Perceiver IO: A general architecture for structured inputs and outputs. In *International Conference on Learning Representations (ICLR)*, 2022. URL https://openreview.net/forum?id=fILj7WpI-g.

Johannes Jakubik, Felix Yang, Benedikt Blumenstiel, Erik Scheurer, Rocco Sedona, Stefano Maurogiovanni, Jente Bosmans, Nikolaos Dionelis, Valerio Marsocci, Niklas Kopp, et al. TerraMind: Large-scale generative multimodality for Earth observation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 7383–7394, 2025. doi: 10.1109/ICCV51701.2025.00693.

Jiale Kang and Qingyu Yin. MiSS: Revisiting the trade-off in LoRA with an efficient shard-sharing structure. In *International Conference on Learning Representations (ICLR)*, 2026. URL https://openreview.net/forum?id=gohmWoUSoS. Formerly titled "Bone: Block Affine Transformation as Parameter Efficient Fine-tuning Methods for Large Language Models".

Konstantin Klemmer, Esther Rolf, Caleb Robinson, Lester Mackey, and Marc Rußwurm. SatCLIP: Global, general-purpose location embeddings with satellite imagery. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 4347–4355, 2025a. doi: 10.1609/aaai.v39i4.32457.

Konstantin Klemmer, Esther Rolf, Marc Rußwurm, Gustau Camps-Valls, Mikolaj Czerkawski, Stefano Ermon, Alistair Francis, Nathan Jacobs, Hannah Rae Kerner, Lester Mackey, Gengchen Mai, Oisin Mac Aodha, Markus Reichstein, Caleb Robinson, David Rolnick, Evan Shelhamer, Vincent Sitzmann, Devis Tuia, and Xiaoxiang Zhu. Earth embeddings: Towards AI-centric representations of our planet. *EarthArXiv preprint*, 2025b. doi: 10.31223/X5HX9S. Preprint, not peer reviewed.

Dawid J. Kopiczko, Tijmen Blankevoort, and Yuki M. Asano. VeRA: Vector-based random matrix adaptation. In *International Conference on Learning Representations (ICLR)*, 2024. arXiv:2310.11454.

Matti Kummu, Maija Taka, and Joseph H. A. Guillaume. Gridded global datasets for Gross Domestic Product and Human Development Index over 1990–2015. *Scientific Data*, 5(1): 180004, 2018. doi: 10.1038/sdata.2018.4.

Remi Lam, Alvaro Sanchez-Gonzalez, Matthew Willson, Peter Wirnsberger, Meire Fortunato, Ferran Alet, Suman Ravuri, Timo Ewalds, Zach Eaton-Rosen, Weihua Hu, Alexander Merose, Stephan Hoyer, George Holland, Oriol Vinyals, Jacklynn Stott, Alexander Pritzel, Shakir Mohamed, and Peter Battaglia. Learning skillful medium-range global weather forecasting. *Science*, 382(6677):1416–1421, 2023. doi: 10.1126/science.adi2336.

Ilya Loshchilov and Frank Hutter. SGDR: Stochastic gradient descent with warm restarts. In *International Conference on Learning Representations (ICLR)*, 2017.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations (ICLR)*, 2019. arXiv:1711.05101.

Ming Ronnier Luo. CIELAB. In *Encyclopedia of Color Science and Technology*, pages 251–257. Springer International Publishing, 2023. doi: 10.1007/978-3-030-89862-5_11.

Michael D. Mastrandrea, Katharine J. Mach, Gian-Kasper Plattner, Ottmar Edenhofer, Thomas F. Stocker, Christopher B. Field, Kristie L. Ebi, and Patrick R. Matschoss. The IPCC AR5 guidance note on consistent treatment of uncertainties: A common approach across the working groups. *Climatic Change*, 108(4):675–691, 2011. doi: 10.1007/s10584-011-0178-6.

Leland McInnes, John Healy, and James Melville. UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018. doi: 10.48550/arXiv.1802.03426.

Nis Meinert, Jakob Gawlikowski, and Alexander Lavin. The unreasonable effectiveness of deep evidential regression. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 9134–9142, 2023. doi: 10.1609/aaai.v37i8.26096.

Tomas Mikolov, Wen-tau Yih, and Geoffrey Zweig. Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*, pages 746–751. Association for Computational Linguistics, 2013.

Shakir Mohamed, Marie-Therese Png, and William Isaac. Decolonial AI: Decolonial theory as sociotechnical foresight in artificial intelligence. *Philosophy & Technology*, 33(4):659–684, 2020. doi: 10.1007/s13347-020-00405-8.

Frances C. Moore, Moritz A. Drupp, James Rising, Simon Dietz, Ivan Rudik, and Gernot Wagner. Synthesis of evidence yields high social cost of carbon due to structural model variation and uncertainties. *Proceedings of the National Academy of Sciences*, 121(52): e2410733121, December 2024. doi: 10.1073/pnas.2410733121. URL https://www.pnas.org/doi/10.1073/pnas.2410733121.

Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics (TOG)*, 41(4):102:1–102:15, 2022. doi: 10.1145/3528223.3530127.

Tung Nguyen, Johannes Brandstetter, Ashish Kapoor, Jayesh K Gupta, and Aditya Grover. ClimaX: A foundation model for weather and climate. In *International Conference on Machine Learning*, pages 25904–25938. PMLR, 2023.

William D Nordhaus. Revisiting the social cost of carbon. *Proceedings of the National Academy of Sciences*, 114(7):1518–1523, 2017. doi: 10.1073/pnas.1609244114.

Brian C O'Neill, Elmar Kriegler, Keywan Riahi, Kristie L Ebi, Stéphane Hallegatte, Timothy R Carter, Ritu Mathur, and Detlef P van Vuuren. A new scenario framework for climate change research: the concept of shared socioeconomic pathways. *Climatic Change*, 122(3):387–400, 2014. doi: 10.1007/s10584-013-0905-2.

Yang Ou et al. Artificial intelligence to support cross-disciplinary climate change research. *Nature Climate Change*, 16(5):505–507, 2026. doi: 10.1038/s41558-026-02624-x.

Jaideep Pathak, Shashank Subramanian, Peter Harrington, Sanjeev Raja, Ashesh Chattopadhyay, Morteza Mardani, Thorsten Kurth, David Hall, Zongyi Li, Kamyar Azizzadenesheli, Pedram Hassanzadeh, Karthik Kashinath, and Animashree Anandkumar. FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. *arXiv preprint arXiv:2202.11214*, 2022.

Ilan Price, Alvaro Sanchez-Gonzalez, Ferran Alet, Tom R. Andersson, Andrew El-Kadi, Dominic Masters, Timo Ewalds, Jacklynn Stott, Shakir Mohamed, Peter Battaglia, Remi Lam, and Matthew Willson. Probabilistic weather forecasting with machine learning. *Nature*, 637(8044):84–90, 2025. doi: 10.1038/s41586-024-08252-9.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning (ICML)*, volume 139 of *Proceedings of Machine Learning Research*, pages 8748–8763. PMLR, 2021.

Arjun Rao, Marc Rußwurm, Konstantin Klemmer, and Esther Rolf. Measuring the intrinsic dimension of Earth representations. In *International Conference on Learning Representations (ICLR)*, 2026. URL https://openreview.net/forum?id=gQPD83DrGp.

Keywan Riahi, Detlef P van Vuuren, Elmar Kriegler, Jae Edmonds, Brian C O'Neill, Shinichiro Fujimori, Nico Bauer, Katherine Calvin, Rob Dellink, Oliver Fricko, et al. The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, 42:153–168, 2017. doi: 10.1016/j.gloenvcha.2016.05.009.

Katherine Richardson, Will Steffen, Wolfgang Lucht, Jørgen Bendtsen, Sarah E Cornell, Jonathan F Donges, Markus Drüke, Ingo Fetzer, Govindasamy Bala, Werner von Bloh, et al. Earth beyond six of nine planetary boundaries. *Science Advances*, 9(37):eadh2458, 2023. doi: 10.1126/sciadv.adh2458.

Carlos Rodriguez-Pardo and Massimo Tavoni. A harmonised dataset for Earth system foundation models. *Scientific Data*, 2026. ISSN 2052-4463. doi: 10.1038/s41597-026-07913-w. URL https://doi.org/10.1038/s41597-026-07913-w.

Carlos Rodriguez-Pardo, Konstantinos Kazatzis, Jorge Lopez-Moreno, and Elena Garces. NeuBTF: Neural fields for BTF encoding and transfer. *Computers & Graphics*, 114: 239–246, 2023. doi: 10.1016/j.cag.2023.06.018.

Carlos Rodriguez-Pardo, Leonardo Chiani, Emanuele Borgonovo, and Massimo Tavoni. Neural conditional transport maps. *Transactions on Machine Learning Research*, 2026. arXiv:2505.15808.

Marc Rußwurm, Konstantin Klemmer, Esther Rolf, Robin Zbinden, and Devis Tuia. Geographic location encoding with spherical harmonics and sinusoidal representation networks. In *The Twelfth International Conference on Learning Representations*, 2024.

Vishwanath Saragadam, Daniel LeJeune, Jasper Tan, Guha Balakrishnan, Ashok Veeraraghavan, and Richard G. Baraniuk. WIRE: Wavelet implicit neural representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 18507–18516, 2023.

James C. Scott. *Seeing Like a State: How Certain Schemes to Improve the Human Condition Have Failed*. Yale University Press, New Haven, CT, 1998. ISBN 978-0-300-07016-3.

Vincent Sitzmann, Julien N. P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 7462–7473, 2020.

Mona Sloane, Emanuel Moss, Olaitan Awomolo, and Laura Forlano. Participation is not a design fix for machine learning. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '22)*, pages 1–6. ACM, 2022. doi: 10.1145/3551624.3555285.

Will Steffen, Katherine Richardson, Johan Rockström, Sarah E. Cornell, Ingo Fetzer, Elena M. Bennett, Reinette Biggs, Stephen R. Carpenter, Wim de Vries, Cynthia A. de Wit, et al. Planetary boundaries: Guiding human development on a changing planet. *Science*, 347 (6223):1259855, 2015. doi: 10.1126/science.1259855.

Elke Stehfest, Detlef van Vuuren, Tom Kram, Lex Bouwman, Rob Alkemade, Michel Bakkenes, Hester Biemans, Arno Bouwman, Michel den Elzen, Jan Janse, et al. *Integrated Assessment of Global Environmental Change with IMAGE 3.0: Model description and policy applications*. Netherlands Environmental Assessment Agency, 2014.

Richard Swinbank and R. James Purser. Fibonacci grids: A novel approach to global modelling. *Quarterly Journal of the Royal Meteorological Society*, 132(619):1769–1793, 2006. doi: 10.1256/qj.05.227.

Daniela Szwarcman, Sujit Roy, Paolo Fraccaro, Þorsteinn Elí Gíslason, Benedikt Blumenstiel, Rinki Ghosal, Pedro Henrique de Oliveira, Joao Lucas de Sousa Almeida, et al. Prithvi-EO-2.0: A versatile multitemporal foundation model for Earth observation applications. *IEEE Transactions on Geoscience and Remote Sensing*, 64:1–20, 2026. doi: 10.1109/TGRS.2025.3642610.

Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in neural information processing systems*, 33:7537–7547, 2020.

Massimo Tavoni and Giovanni Valente. Uncertainty in Integrated Assessment Modeling of Climate Change. *Perspectives on Science*, 30(2):321–351, 2022. doi: 10.1162/posc_a_00417. URL https://direct.mit.edu/posc/article-abstract/30/2/321/109169/Uncertainty-in-Integrated-Assessment-Modeling-of?redirectedFrom=fulltext.

Massimo Tavoni, Marta Mastropietro, and Jonathan Spinoni. Past and projected climate impacts on human development. Research Square preprint, July 2025. Preprint, not peer reviewed. doi: 10.21203/rs.3.rs-6923904/v1.

Jan Teorell, Aksel Sundström, Sören Holmberg, Bo Rothstein, Natalia Alvarado Pachon, and Cem Mert Dalli. The Quality of Government Standard Dataset, version Jan21. University of Gothenburg: The Quality of Government Institute, 2021. doi: 10.18157/qogstdjan21.

Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30, pages 5998–6008, 2017.

Vicente Vivanco Cepeda, Gaurav Kumar Nayak, and Mubarak Shah. GeoCLIP: CLIP-inspired alignment between locations and images for effective worldwide geo-localization. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, pages 8690–8701, 2023. doi: 10.52202/075280-0379.

Yi Wang, Zhitong Xiong, Chenying Liu, Adam J. Stewart, Thomas Dujardin, Nikolaos Ioannis Bountos, Angelos Zavras, Franziska Gerken, Ioannis Papoutsis, Laura Leal-Taixé, and Xiao Xiang Zhu. Towards a unified Copernicus foundation model for Earth vision. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9888–9899, 2025. Copernicus-FM.

John Weyant. Some Contributions of Integrated Assessment Models of Global Climate Change. *Review of Environmental Economics and Policy*, 11(1):115–137, January 2017. doi: 10.1093/reep/rew018.

Langdon Winner. Do artifacts have politics? *Daedalus*, 109(1):121–136, 1980. Modern Technology: Problem or Opportunity?

Fengli Xu, Yong Li, and Shusheng Xu. Attentional multi-graph convolutional network for regional economy prediction with open migration data. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2225–2233. ACM, 2020. doi: 10.1145/3394486.3403273.