MIND THE GAP: A GEOGRAPHIC IMPLICIT NEURAL REPRESENTATION WITH ADJUSTABLE SPATIAL SCALE

Anonymous authors
Paper under double-blind review

ABSTRACT

Geographic measurements are often sparse, leaving large areas without labels for the quantities we want to map. Geographic implicit neural representations (INRs) provide coordinate-based embeddings that can be combined with sparse labels to predict at unsampled locations without satellite imagery at inference. Yet existing INRs are largely evaluated with random holdouts, leaving their ability to generalize across larger geographic gaps unclear. We introduce Matryoshka Implicit Neural Distillation (MIND), a geographic INR whose spatial granularity can be adjusted after training. MIND distills several pretrained geospatial models using nested supervision at increasing embedding dimensions, dividing the representation into contiguous chunks. Early chunks capture broad spatial patterns, while later chunks add increasingly local variation. Downstream models can retain only the leading chunks or use our Chunked Penalty to reduce reliance on later chunks without retraining the INR. We also introduce CoordBench, comprising 52 datasets and 78 targets with both random and regional holdouts at multiple spatial scales. Across CoordBench, fine-scale features help most when labels are nearby, while smoother representations generalize better across larger geographic gaps. MIND with the Chunked Penalty achieves the highest aggregate regression and classification performance among tested INRs and the highest overall performance under regional holdout. These results show that geographic representations should be evaluated and adapted according to the spatial separation between labeled and prediction locations.

1 INTRODUCTION

Measurements of income, health, and other socioeconomic and environmental quantities are unevenly distributed across the world, while maps require dense predictions between and beyond these observations for visualization and decision-making (Rolf et al., 2021; Agarwal et al., 2024; CDC, 2023; Yeh et al., 2021; Fick & Hijmans, 2017). Some unobserved locations lie near labeled samples, while others lie in regions with few or no labels for the target of interest (Mai et al., 2020). Geographic implicit neural representations (INRs) encode satellite observations and other geospatial data as a continuous function over the spherical surface of planet Earth (Mai et al., 2022; 2020; Rußwurm et al., 2024). Queried via location (latitude/longitude) inputs, these implicit data representations can be combined with available labels to interpolate or extrapolate at new unsampled locations, without requiring satellite imagery at inference time (Vivanco Cepeda et al., 2023; Klemmer et al., 2025).

Standard evaluations of INRs fit linear predictors to frozen features and evaluate on randomly assigned folds of the labeled dataset (Dollinger et al., 2025; Astruc et al., 2026; Liu et al., 2026; Mai et al., 2023a). Because nearby locations often have similar target values due to *spatial autocorrelation*, these splits primarily test local interpolation near labeled locations (Ploton et al., 2020; Shen et al., 2026). In our benchmark, 93% of random-fold test observations lie within 25 km of a training point, and inverse-distance weighting (IDW) on coordinates achieves higher aggregate regression R^2 than every tested INR under random folds (Table 3). Under regional holdout, features that capture local variation can lose predictive accuracy as held-out regions grow, and the ranking of encoders changes. Evaluating across these scales is necessary when maps extend beyond surveyed areas.

A general-purpose coordinate representation should combine diverse geographic signals while supporting the different spatial scales needed for nearby and regional prediction. Existing INRs are

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

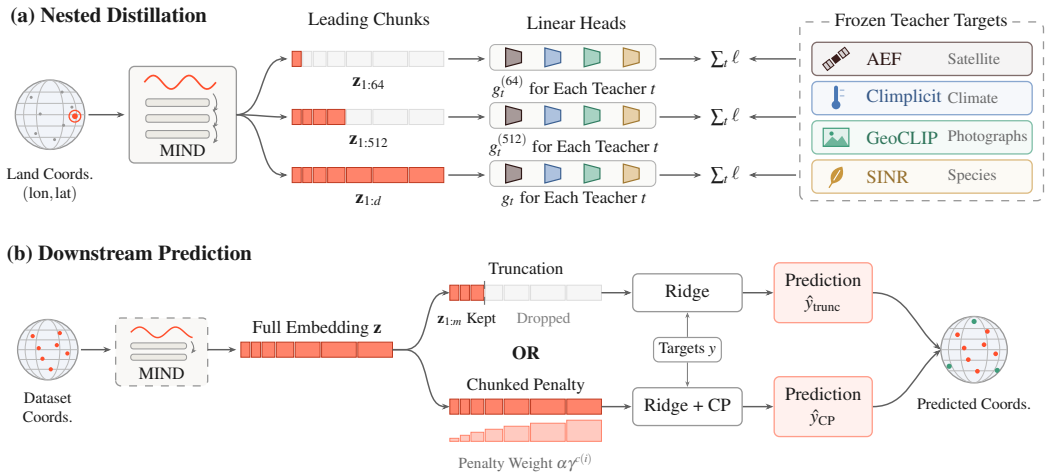

Figure 1: **MIND training and downstream prediction.** (a) Training learns a shared embedding of multiple teachers. At each supervised dimension, linear heads reconstruct all teacher embeddings from the leading chunks. The heads are discarded after training. (b) Prediction takes the full embedding, truncates (keeping leading chunks), or regularizes (penalizing later chunks) to suit the needs of the task.

often trained and evaluated within one modality, such as climate, photographs, or species observations (Dollinger et al., 2025; Vivanco Cepeda et al., 2023; Cole et al., 2023). Distilling specialist embeddings at shared coordinates offers a way to combine these signals without processing the original datasets.

We introduce MIND (Matryoshka Implicit Neural Distillation), a geographic INR with adjustable spatial scale (Fig. 1). MIND learns to reconstruct embeddings from pretrained models, called *teachers*, using MINDSET, our dataset of paired coordinates and teacher embeddings. Following Matryoshka Representation Learning (Kusupati et al., 2022), we supervise cumulative sets of dimensions at 64, 128, and successively larger sizes up to the full embedding. These boundaries define contiguous *chunks*, with each reconstruction using all preceding chunks. In our experiments, early chunks capture smoother, large-scale geographic patterns, while later chunks add more localized variation. The encoder produces a fixed embedding at each coordinate, with spatial scale selected through truncation or coefficient regularization when fitting the downstream predictor. Our Chunked Penalty (CP) retains all chunks but penalizes later chunks more strongly, balancing performance under random folds and regional holdout without retraining the encoder.

Prior geographic INR studies use different datasets and evaluation procedures, making encoder comparisons difficult. To enable consistent comparisons, we introduce CoordBench, with 52 datasets and 78 targets evaluated using common predictors, metrics, random folds, and regional holdout. MIND+CP achieves the highest aggregate scores among the tested INRs (Fig. 2, right) and improves regional regression over selecting a fixed number of leading chunks to retain. The same MIND encoder trained on the same data without nested supervision does not show the coarse-to-fine ordering across chunks, and the benefit of the penalty depends on the trained dimension order. We release MIND, CoordBench, MINDSET, and a global 1 km grid of MIND embeddings, so users can select chunks without recomputing embeddings.

2 Related Work

Geographic representations. Geographic INRs encode coordinates using multiscale features, spherical harmonics, or learned networks (Mai et al., 2020; 2023b; Rußwurm et al., 2024; Astruc et al., 2026). Sphere2Vec explicitly encodes spherical geometry to preserve geodesic distance (Mai et al., 2023b). Their supervision includes imagery (Klemmer et al., 2025; Vivanco Cepeda et al., 2023; Mai et al., 2023a; Liu et al., 2026), species observations (Cole et al., 2023; Sastry et al., 2025), climate data (Dollinger et al., 2025), and geographic partitions (Cher et al., 2026). Other Earth em-

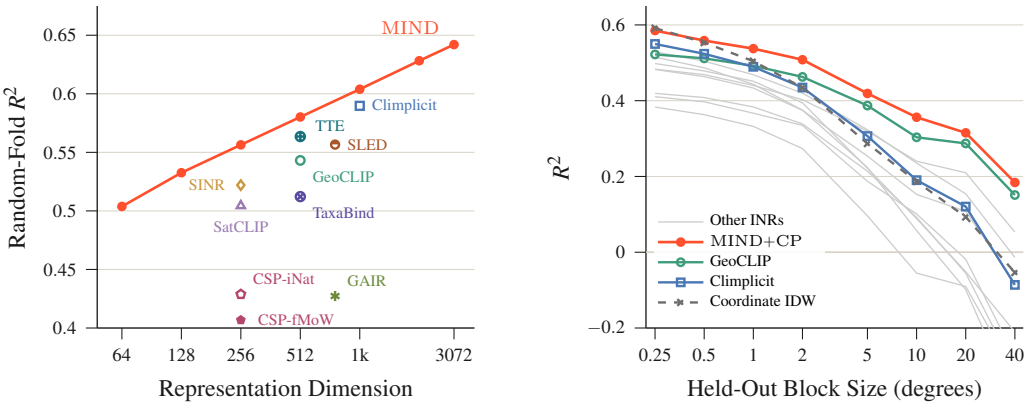

Figure 2: **Representation dimension and geographic separation.** Without separation, MIND has higher random-fold R^2 than the other tested INRs (left). With enforced geographic separation of the train and test sets, MIND with the Chunked Penalty (MIND+CP) has the highest R^2 among tested INRs at each spatial scale (right), exceeding IDW at block sizes of 0.5° and larger.

beddings include AlphaEarth Foundations (AEF) (Brown et al., 2025) and models derived from multiple geospatial data sources (Agarwal et al., 2024; Klemmer et al., 2026). Scalable Location Encoding via Distillation (SLED) (Lane et al., 2026), the closest distillation approach to MIND, distills frozen satellite image encoders into a coordinate representation. MIND distills teachers trained on satellite imagery, climate data, photographs, and species observations into a single coordinate embedding.

Nested supervision and spatial scale. Matryoshka Representation Learning supervises cumulative sets of dimensions so one embedding can serve different dimension budgets (Kusupati et al., 2022; Zhang et al., 2025). TESSERA v2 uses Matryoshka distillation to produce compact image-derived embeddings at several dimensions (Feng et al., 2026). Neural fields can explicitly constrain spatial frequencies (Lindell et al., 2022; Lin et al., 2021; Yang et al., 2023). RANGE (Dhakal et al., 2025) augments frozen SatCLIP embeddings through an external image-feature bank and controls spatial smoothness through retrieval weights. MIND learns a coordinate embedding with nested supervision and adjusts its spatial scale through downstream truncation or the Chunked Penalty.

Geographic INR evaluation. Regional holdout is an established practice to address spatial autocorrelation during evaluation in ecology and remote sensing (Roberts et al., 2017; Valavi et al., 2019; Karasiak et al., 2022; Rolf, 2023). Spatial baselines and analyses of train–test distances show that geospatial evaluations need to account for proximity to training data (Hijmans, 2012; Bahn & McGill, 2007). Buffered exclusion removes nearby training points directly (Shen et al., 2024; 2026), and distance-matched folds aim to reproduce the expected distances between train and test locations (Milà et al., 2022). Cai & Balestriero (2025) introduce spherical wavelet encodings and the FAIR-Earth dataset for evaluating geographic subgroup disparities. CoordBench complements these and other INR evaluations (Dollinger et al., 2025; Astruc et al., 2026; Liu et al., 2026; Klemmer et al., 2025; Wu et al., 2024) by testing both nearby prediction and generalization across larger geographic gaps.

3 MIND FOR GEOGRAPHIC REPRESENTATIONS

Matryoshka Implicit Neural Distillation (MIND) maps a geographic coordinate (latitude/longitude) to an embedding divided into contiguous chunks. In our experiments, earlier chunks capture broader spatial variation than later chunks (Sec. 5.2). During training, linear layers called reconstruction heads map each cumulative set of chunks to the teacher embeddings (Fig. 1). After training, we discard these heads and fit a downstream predictor to either the leading chunks or the full embedding.

Table 1: **CoordBench sources and coverage.** CoordBench is composed of 78 target variables across 52 datasets, combining the disjoint benchmark datasets used in many Geographic INR evaluations. Spatial unit gives the source label unit or raster resolution where available.

Source	Targets	Spatial unit	Coverage
PDFM (Agarwal et al., 2024)	27 socioeconomic measures	US place-level	Contiguous US
PLACES (CDC, 2023)	12 health measures	ZCTA centroids	US postal areas
SustainBench (Yeh et al., 2021)	Assets, water, sanitation	DHS cluster centroids	Global
MOSAIKS/USAVars (Rolf et al., 2021)	Population, nightlights, income	1 km points	US
SatCLIP (Klemmer et al., 2025)	Population, elevation, biome	Global point samples	Global
AEF evaluation (Brown et al., 2025)	Land cover, crops, emissivity	Task-dependent	Global
WorldClim (Fick & Hijmans, 2017)	Temperature, precipitation	10-arc-minute grid	Land
SoilGrids (Poggio et al., 2021)	Organic carbon, pH	Global raster points	Land
California Housing (Pace & Barry, 1997)	Median house value	Census block groups	California

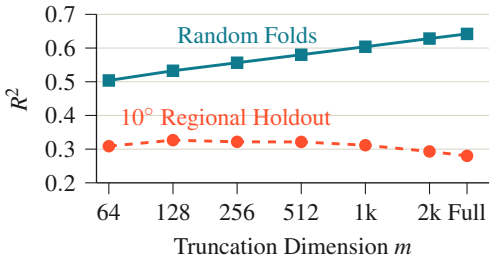

Figure 3: **Chunk performance across different holdouts.** Adding chunks improves random-fold R^2 but lowers 10° blocked regional-holdout R^2 beyond 128 dimensions.

Table 2: **The effect of nested supervision.** At the same dimension, MIND's leading chunks have higher R^2 on CoordBench than the identical model trained without nested supervision, with larger differences under regional holdout.

Dim.	Model	Random	10°	40°
64	Non-nested	0.452	0.225	−0.019
	MIND	**0.504**	**0.309**	**0.128**
256	Non-nested	0.550	0.312	0.059
	MIND	**0.556**	**0.322**	**0.116**

3.1 Nested distillation

Following the notation of Matryoshka Representation Learning (Kusupati et al., 2022), let $\mathbf{z} = f_\theta(\mathbf{x}) \in \mathbb{R}^d$ be the full embedding at coordinate $\mathbf{x}$. Truncation at m keeps the first m dimensions, which form $\mathbf{z}_{1:m} \in \mathbb{R}^m$, where $m \in [d]$ and $[d] = \{1, \ldots, d\}$. We denote this representation by MIND_m. The supervised dimensions form a set $\mathscr{M} \subset [d]$ that includes d, with consecutive values defining the chunk boundaries. Each teacher provides a target vector $\mathbf{e}_t \in \mathbb{R}^{d_t}$ at the same coordinate. We normalize the teacher targets by the PHI-S transformation (Ranzinger et al., 2024) and minimize the following distillation loss:

$$\mathscr{L} = \sum_t \ell\big(g_t(\mathbf{z}), \mathbf{e}_t\big) + \frac{1}{|\mathscr{M}| - 1} \sum_{m \in \mathscr{M} \setminus \{d\}} \sum_t \ell\big(g_t^{(m)}(\mathbf{z}_{1:m}), \mathbf{e}_t\big). \tag{1}$$

Each teacher has a separate linear reconstruction head at every supervised dimension. The head $g_t^{(m)}$ maps m student dimensions to all d_t teacher dimensions, while g_t uses the full d-dimensional student embedding. For teacher t, let $\mathbf{u}$ and $\mathbf{v}$ be the d_t-dimensional prediction and target. The reconstruction loss combines cosine distance and mean squared error:

$$\ell(\mathbf{u}, \mathbf{v}) = 1 - \bar{\mathbf{u}}^\top \bar{\mathbf{v}} + \frac{1}{d_t} \|\bar{\mathbf{u}} - \bar{\mathbf{v}}\|_2^2, \qquad \bar{\mathbf{u}} = \frac{\mathbf{u}}{\|\mathbf{u}\|_2}. \tag{2}$$

Here $\bar{\mathbf{v}} = \mathbf{v}/\|\mathbf{v}\|_2$. App. A gives the PHI-S normalization details. We weight the full-dimensional loss separately and average the losses for $m < d$ (Zhang et al., 2025). We measure how spatial scale changes as more chunks are included in Sec. 5.2.

Our encoder is a residual sinusoidal representation network (ReSIREN) (Dollinger et al., 2025; Sitzmann et al., 2020) that maps coordinates from the Equal Earth map projection to an embedding of full dimension $d = 3{,}072$, denoted $\text{MIND}_{\text{Full}}$. The teachers are AEF, Climplicit, GeoCLIP, and

Table 3: **Performance of Different Methods on CoordBench.** MIND+CP has the highest R^2 and accuracy among tested INRs while Coordinate IDW has the highest random-fold R^2. Degrees indicate the block size for regional holdout, and we report means across five fold-assignment seeds. Appendix Table 8 reports standard deviations. Best result in **bold**, second-best in *italics*.

	R^2					Accuracy (%)				
Method	Random	2°	10°	20°	40°	Random	2°	10°	20°	40°
Coord. IDW	**0.675**	0.435	0.187	0.093	-0.053	67.7	62.8	55.7	53.0	51.2
Cartesian 3D	0.137	0.089	-0.088	-0.173	-0.423	51.8	50.6	46.4	45.8	44.4
Wrap (Sin/Cos)	0.204	0.155	-0.025	-0.116	-0.344	53.1	52.2	48.3	46.8	45.2
SINR *(ICML '23)*	0.522	0.394	0.089	-0.051	-0.211	63.8	61.4	55.2	51.4	49.5
SatCLIP *(AAAI '25)*	0.504	0.375	0.153	0.108	-0.082	63.8	60.8	55.0	52.4	51.9
Climplicit *(ICLRW '25)*	0.590	0.435	0.190	0.120	-0.086	66.7	63.2	57.1	54.6	52.8
GeoCLIP *(NeurIPS '23)*	0.543	0.463	0.303	0.287	*0.151*	63.5	60.6	56.0	54.6	54.2
CSP-iNat *(ICML '23)*	0.429	0.339	0.092	-0.055	-0.310	61.4	59.4	54.0	51.3	48.2
CSP-fMoW *(ICML '23)*	0.407	0.335	0.101	-0.018	-0.276	61.7	60.0	54.8	51.9	47.7
GAIR *(ISPRS '26)*	0.427	0.274	-0.055	-0.090	-0.322	60.3	57.1	51.7	49.6	48.5
TTE *(ECCV '26)*	0.563	0.420	0.235	0.154	-0.014	66.2	62.5	58.2	55.8	54.7
TaxaBind *(WACV '25)*	0.512	0.402	0.240	0.210	0.053	63.9	61.0	55.9	54.4	53.7
SLED *(arXiv '26)*	0.557	0.375	0.057	-0.099	-0.394	66.0	62.0	53.9	49.9	46.0
MIND_{64}	0.504	0.447	0.309	0.281	0.128	62.5	60.5	56.9	55.3	54.2
MIND_{128}	0.533	0.465	*0.327*	*0.303*	0.136	64.0	62.0	57.7	56.3	54.8
MIND_{256}	0.556	0.476	0.322	0.279	0.116	65.2	63.0	58.1	*56.8*	*55.3*
$\text{MIND}_{\text{Full}}$	0.642	*0.492*	0.280	0.211	0.019	*67.8*	*64.4*	*59.1*	56.2	54.6
MIND+CP	*0.643*	**0.508**	**0.356**	**0.315**	**0.184**	**67.8**	**64.6**	**59.9**	**57.8**	**56.3**

Spatial Implicit Neural Representations (SINR), which are supervised by satellite imagery, gridded climate data, photographs, and species observations, respectively (Brown et al., 2025; Dollinger et al., 2025; Vivanco Cepeda et al., 2023; Cole et al., 2023). We train on MINDSET, a pretraining dataset we compiled that pairs teacher embeddings with 12M land coordinates sampled densely in and around cities (App. C). Architecture, teacher dimensions, and optimization settings are in App. A.

3.2 Chunked Penalty for Scale-aware Regularization

Truncation restricts the predictor to the leading chunks, discarding later dimensions and any useful information they contain. As a soft alternative, we propose the Chunked Penalty to regularize the full embedding by penalizing coefficients for later chunks more strongly. For features X and regression targets y, the predictor coefficients solve

$$(\hat{\beta}, \hat{b}) = \arg\min_{\beta, b} \|y - X\beta - b\mathbf{1}\|_2^2 + \alpha \sum_i \gamma^{c(i)} \beta_i^2, \tag{3}$$

Here β contains the feature coefficients and b is the intercept. The chunk index $c(i)$ is fixed by the training boundaries, with $c(i) = 0$ for dimensions 1–64, $c(i) = 1$ for 65–128, and so on. The parameter α sets the overall penalty and γ determines how much it grows between chunks. Standard ridge regression is the special case $\gamma = 1$. MIND+CP applies the Chunked Penalty to the same full embedding used by $\text{MIND}_{\text{Full}}$ with standard ridge regression. The penalty downweights later chunks, where we observe more local spatial variation. The penalty can also be applied to other INR embeddings, which we evaluate in Sec. 5.3. For classification, we use a one-hot ridge predictor, fitting Eq. 3 for each class with shared α and γ. We predict the class with the largest linear output.

4 CoordBench: Measuring Spatial Generalization

CoordBench contains 52 datasets from nine sources (Agarwal et al., 2024; CDC, 2023; Yeh et al., 2021; Rolf et al., 2021; Klemmer et al., 2025; Brown et al., 2025; Fick & Hijmans, 2017; Poggio

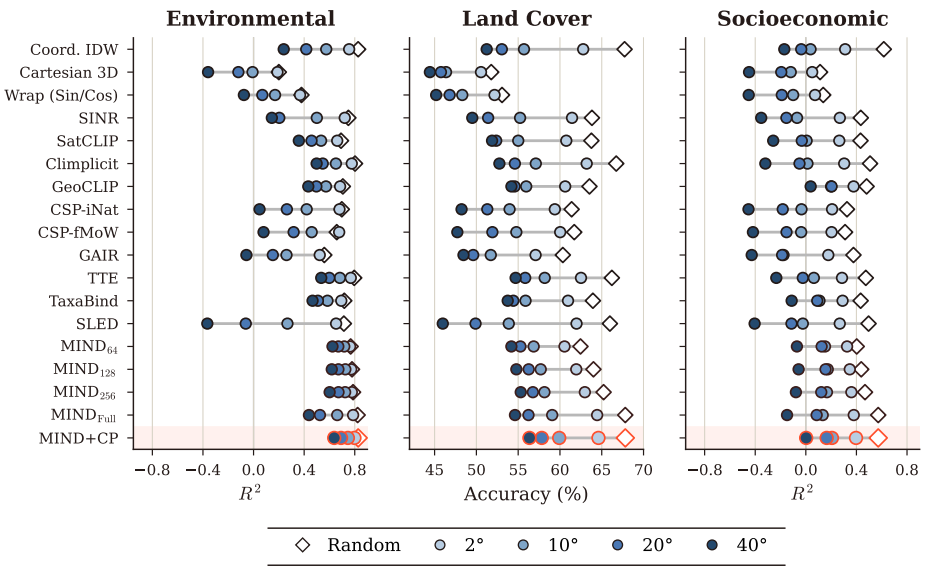

Figure 4: **CoordBench results by target category.** Each row shows one method's score under random folds (diamond) and regional holdouts with varying block sizes in degrees (darker for larger blocks). MIND+CP has the highest score among INRs at every block size for environmental and land-cover targets. Metrics are the mean scores across five seeds. See Table 7 for the full results.

et al., 2021; Pace & Barry, 1997), with 78 prediction targets (Table 1). The benchmark suite covers socioeconomic targets such as income and housing (26 datasets), environmental quantities such as temperature and soil pH (10), and land-cover classes (16). Each dataset contains observations at a shared set of coordinates and one or more target columns. We normalize them to a common data format and fold assignments so that new encoders can be compared without rebuilding each source's preprocessing.

Spatial coverage and resolution. The CoordBench datasets differ in label resolution, sampling density, and geographic coverage. WorldClim provides annual temperature and precipitation on a 10-arc-minute grid (Fick & Hijmans, 2017) and MOSAIKS densely sampled 1 km population labels (Rolf et al., 2021), while the Africa crop mask (Brown et al., 2025) has 2,556 clustered locations. For a coordinate-only encoder, these tasks test both broad patterns in climate and population and local variation in land cover and soil carbon.

Splits. In all splits, we assign observations to five random folds. For *random folds*, we allocate points to folds uniformly at random. To evaluate prediction across geographic gaps, CoordBench adds common *regional folds* by spatial blocking, assigning latitude–longitude cells to five folds using a seeded permutation with balanced cell counts. Increasing block cell size from 0.25° to 40° generally increases train–test distance, but east–west extent shrinks toward the poles, so degrees do not specify a uniform physical holdout size. Appendix Fig. 8 illustrates these folds. Both splits withhold downstream labels from predictor training, while the encoders remain pretrained on globally distributed data. We also perform a buffered holdout experiment and find that performance declines as test points near training points are removed (App. E). For comparison with existing INR benchmarks, LocBench and TorchSpatial neglect regional holdouts and instead use random 80/20 splits, even ignoring SustainBench's original country-level splits (Wu et al., 2024).

Evaluation. For regression, we floor each dataset's R^2 score at -1 before averaging across datasets so a few divergent targets do not dominate the aggregate (App. F). All methods use the same datasets at each block size (App. A). We repeat five-fold evaluation with five random seeds to shuffle the fold assignments and report the mean.

Table 4: **Chunk selection and regularization.** The Chunked Penalty improves regional R^2 for MIND, but not for the non-nested model, GeoCLIP, or permuted MIND.

Method	CP	10°	40°
Non-nested	×	**0.338**	**0.127**
	✓	0.330	0.114
GeoCLIP	×	**0.303**	**0.151**
	✓	0.303	0.148
MIND (chunk sel.)	–	0.321	0.134
MIND	✓	**0.356**	**0.184**
$\text{MIND}_{\text{Full}}$ (Perm.)	×	**0.289**	**0.115**
	✓	0.286	0.099

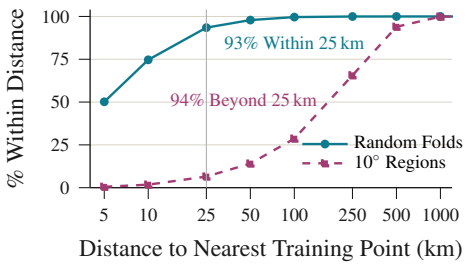

Figure 5: **Held-out distances.** Cumulative distributions of distance to the nearest training point, pooled across CoordBench. Under random folds, 93% of observations lie within 25 km of training data; under 10° regional holdout, 94% lie farther than 25 km.

5 Experiments

We compare MIND with nine pretrained INR families: SINR (Cole et al., 2023), SatCLIP (Klemmer et al., 2025), Climplicit (Dollinger et al., 2025), GeoCLIP (Vivanco Cepeda et al., 2023), the two Contrastive Spatial Pretraining (CSP) checkpoints (Mai et al., 2023a), trained on iNaturalist (Van Horn et al., 2018) and Functional Map of the World (fMoW) (Christie et al., 2018), Geo-aligned Implicit Representations (GAIR) (Liu et al., 2026), Tessellating The Earth (TTE) (Cher et al., 2026), TaxaBind (Sastry et al., 2025), and SLED (Lane et al., 2026). We exclude UniGeoCLIP (Astruc et al., 2026) and FAIR-Earth's spherical wavelet models (Cai & Balestriero, 2025) because neither releases trained checkpoints or training data. We restrict encoder comparisons to coordinate-only models, excluding the imagery-based AEF and TESSERA v2 (Brown et al., 2025; Feng et al., 2026). We baseline against the Wrap (Sin/Cos) and Cartesian 3D (XYZ) nonparametric coordinate encoders, as well as Coordinate IDW to measure local interpolation. Coordinate IDW uses distance-weighted k-nearest neighbors, taking a weighted average for regression and a weighted class vote for classification. We fit linear predictors to all learned representations using frozen features and standard ridge regression, except that MIND+CP uses the Chunked Penalty described in Eq. 3. Further details are described in App. A.

Experimental setup. We use nested cross-validation to tune the truncation dimension and regularization parameters separately for each target (Cawley & Talbot, 2010). For each outer test fold, we choose these settings using inner validation folds within the remaining data. Both levels use regional folds for regional holdout and random folds otherwise.

5.1 Prediction in held-out regions

Under random folds, 93% of held-out observations are within 25 km of a training point (Fig. 5), and Coordinate IDW outperforms every INR in aggregate regression R^2 (Table 3). The same effect appears within MIND: the full embedding performs better than the first 64 dimensions for nearby test points, but the difference is much smaller beyond 50 km. Removing nearby training labels similarly reduces performance more for the full embedding than for the first chunk (App. E). This suggests that later MIND dimensions are most useful when nearby labeled data are available, while earlier dimensions are less dependent on local supervision.

Among the pretrained INR baselines, Climplicit performs best under random folds, while GeoCLIP performs best under regional holdouts (Table 3). MIND+CP exceeds both, with the highest overall R^2 and accuracy among the tested INRs in every split. This result is consistent across fold assignments: MIND+CP outperforms GeoCLIP in 24 of 25 split-by-seed comparisons and Climplicit in all 25 (Table 8). It also exceeds Coordinate IDW at every regional block size, while simply concatenating the teachers performs worse than MIND+CP in every evaluated split (App. A). Overall, MIND+CP remains competitive when nearby training data are available while improving performance as training and test locations become more geographically separated.

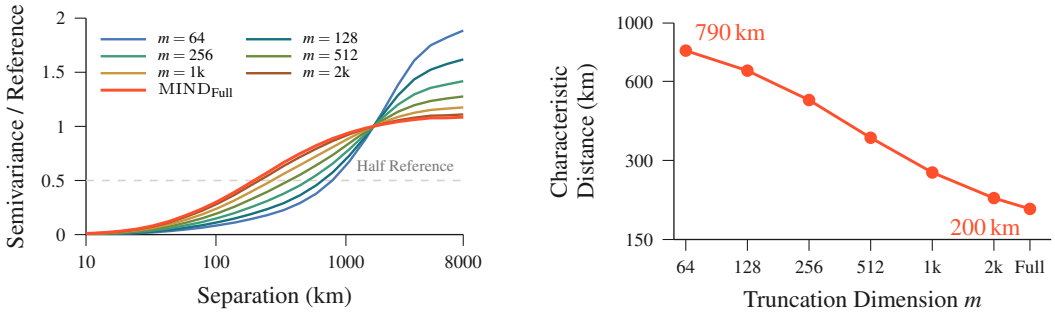

Figure 6: **Spatial scale across truncations.** Left: semivariance against geographic separation, normalized by the median semivariance over the largest quartile of sampled separations. Curves for more chunks rise at shorter distances. Right: the separation where each curve reaches half this reference value, decreasing from 790 km for the first chunk to 200 km for the full embedding.

5.2 Spatial scale across chunks

MIND's early chunks vary smoothly across broad geographic areas, while later chunks add increasingly local variation (Figs. 6 and 7). We hypothesize that the limited capacity of early chunks, together with the spectral bias of neural networks (Tancik et al., 2020; Rahaman et al., 2019), favors broad spatial patterns shared across teachers, such as latitudinal climate gradients, while finer local variation is pushed into later chunks. Notably, nested supervision does not explicitly assign a spatial scale to any chunk but it emerges implicitly through the nested distillation training from specialist teachers.

We quantify this pattern with an empirical *semivariogram*, which measures how quickly embeddings change with geographic distance (Fig. 6). We summarize it using a *characteristic distance*: the separation at which semivariance reaches half its large-distance reference value. Shorter distances indicate finer spatial variation. As more chunks are included, the characteristic distance falls from about 790 km for the first 64 dimensions to 200 km for the full embedding. App. D describes the estimator.

Adding chunks improves random-fold regression, while truncating MIND gives higher R^2 for regional holdouts (Table 3, Fig. 3). Performance also drops below 64 dimensions, indicating that shortening the first chunk further loses useful information (App. D). *This coarse-to-fine ordering appears to come from nested supervision rather than representation size alone.* The characteristic distance decreases as chunks are added to MIND, but remains nearly constant across dimensions in models trained at full width, with or without the addition of feature decorrelation methods like VICReg (Bardes et al., 2022) (App. D). Likewise, separately trained 64- and 256-dimensional models underperform MIND truncated to the same dimensions across all splits (Table 2).

5.3 Chunk selection, regularization, and spatial ordering

The Chunked Penalty improves regional R^2 over per-target chunk selection at 10° and 40° without reducing $\text{MIND}_{\text{Full}}$'s random-fold performance (Table 4); at 10°, chunk selection also performs worse than simply using MIND_{128} (Table 3). Its benefit depends on MIND's nested representation. An identical architecture trained only at full dimension performs better than $\text{MIND}_{\text{Full}}$ without the penalty (0.338 vs. 0.280 at 10°), but does not improve with the Chunked Penalty, and MIND+CP outperforms this non-nested variant. GeoCLIP likewise does not benefit from the penalty because its dimensions are not nested. Randomly permuting MIND's dimensions leaves performance unchanged but removes the gain from the Chunked Penalty (Table 4), suggesting that the improvement comes from the ordering induced by nested supervision rather than regularization alone.

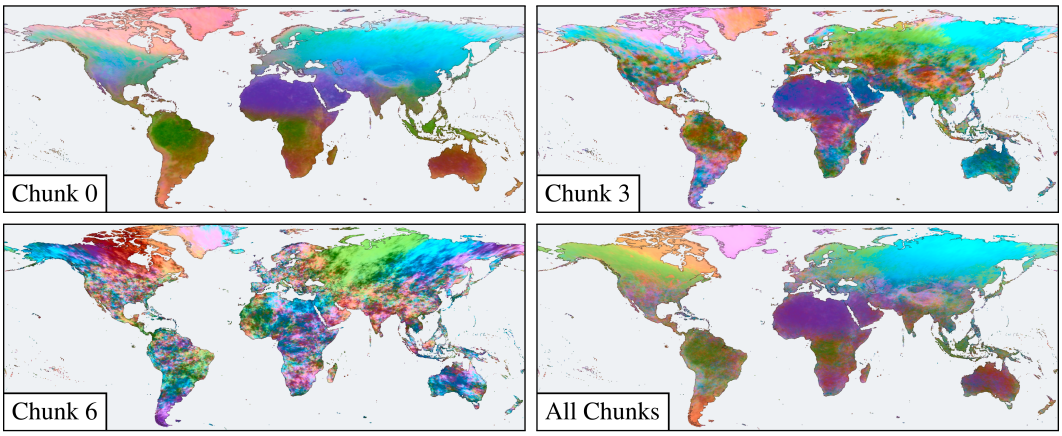

Figure 7: **Spatial scale across chunks.** Chunk 0 varies gradually across continents (top left), Chunk 3 adds finer regional structure (top right), and Chunk 6 varies at finer scales within regions (bottom left). The full embedding combines variation across spatial scales (bottom right). Each panel maps the top three principal components of the dimensions.

6 Discussion and Conclusion

Geographic evaluation. Coordinate IDW is strongest in regression under random folds, where test points are often close to training observations, but its advantage reverses under regional holdout (Table 3). The buffered-exclusion experiment shows that this effect is driven in part by proximity itself (App. E). Random folds are therefore useful for measuring interpolation near existing observations, but not regional generalization. We recommend reporting IDW, train–test distances, and regional holdouts at scales relevant to deployment. A natural extension is to krige the residuals of an embedding-based predictor (Hengl et al., 2007), using nearby labels for local correction and the learned representation farther away.

Spatial scale. MIND adapts the spatial scale of a fixed representation without retraining. Early chunks capture broader patterns, while later chunks add finer variation that can help nearby prediction but transfer less well to unseen regions. Truncation removes later chunks entirely, while the Chunked Penalty downweights them, which may explain its higher regional R^2 than chunk selection (Table 4). The non-nested and permutation experiments suggest that this benefit depends primarily on the learned ordering rather than regularization alone. When spatial validation is impractical, characteristic distance could provide a label-free way to choose an appropriate representation scale.

Limitations and future work. CoordBench is geographically imbalanced, with 43 of 78 targets from the United States, so evidence is weaker in data-poor regions. Its common format makes it straightforward to add more spatial and temporal datasets. MIND's learned ordering also depends on its teacher representations and training samples. Broader sampling beyond MINDSET's current emphasis on cities may improve reconstruction in rural and sparsely sampled areas. MINDSET's annual AEF embeddings could also be used to study whether a similar ordering emerges over time.

Conclusion. Fine-scale features are most useful near labeled data, while smoother features transfer better across larger gaps. MIND captures both in a single nested representation, allowing downstream models to adjust spatial scale without retraining the encoder. Truncation is useful when the desired scale is known and also reduces storage and probing cost. When it is not, the Chunked Penalty provides a softer way to reduce reliance on finer-scale features. CoordBench makes these differences measurable across random and regional holdouts, showing that the best geographic representation depends not only on the target, but also on how far predictions are made from labeled data.

Reproducibility Statement

The supplementary material contains training and evaluation scripts, model configuration, and per-dataset results for reproducing the experiments. Appendix A specifies the objective, optimization, predictors, folds, and aggregation. Apps. B and C describe data provenance. The table and figure scripts derive reported aggregates from these results. CoordBench is at anonymous-hf.com/a/f9m59io7ve68, MIND model checkpoints at anonymous-hf.com/a/yt7qkf7ky3vv, and the MINDSET pretraining dataset at anonymous-hf.com/a/yect8xpopd3m.

AI Use Statement

We used language models for manuscript feedback and revision, consistency checks, and assistance with code development and review.

Ethics Statement

MIND can be used to predict socioeconomic, health, and environmental variables in places with few measurements. If those predictions are wrong, they could lead to poor decisions, especially where there is little local data to check them against. Performance may also be worse in rural or under-sampled regions because the training data are concentrated around cities, and biases in the teacher models may carry into MIND. We therefore do not treat these predictions as a replacement for local measurements in high-stakes settings. All datasets used in this work are public, and their provenance and license status are documented in the dataset card.

References

Mohit Agarwal, Mimi Sun, Chaitanya Kamath, Arbaaz Muslim, Prithul Sarker, Joydeep Paul, Hector Yee, Marcin Sieniek, Kim Jablonski, Swapnil Vispute, Atul Kumar, Yael Mayer, David Fork, Sheila de Guia, Jamie McPike, Adam Boulanger, Tomer Shekel, David Schottlander, Yao Xiao, Manjit Chakravarthy Manukonda, Yun Liu, Neslihan Bulut, Sami Abu-El-Haija, Bryan Perozzi, Monica Bharel, Von Nguyen, Luke Barrington, Niv Efron, Yossi Matias, Greg Corrado, Krish Eswaran, Shruthi Prabhakara, Shravya Shetty, and Gautam Prasad. General geospatial inference with a population dynamics foundation model. arXiv preprint arXiv:2411.07207, 2024. URL https://arxiv.org/abs/2411.07207.

Guillaume Astruc, Eduard Trulls, Jan Hosang, Loic Landrieu, and Paul-Edouard Sarlin. UniGeoCLIP: Unified geospatial contrastive learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 7847–7856, 2026. URL https://arxiv.org/abs/2604.11668.

Volker Bahn and Brian J. McGill. Can niche-based distribution models outperform spatial interpolation? *Global Ecology and Biogeography*, 16(6):733–742, 2007. doi: 10.1111/j.1466-8238.2007.00331.x.

Adrien Bardes, Jean Ponce, and Yann LeCun. VICReg: Variance-invariance-covariance regularization for self-supervised learning. In *International Conference on Learning Representations (ICLR)*, 2022. URL https://arxiv.org/abs/2105.04906.

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.

Clara R. Burgert, Josh Colston, Thea Roy, and Blake Zachary. Geographic displacement procedure and georeferenced data release policy for the Demographic and Health Surveys. DHS Spatial Analysis Report 7, ICF International, 2013. URL https://dhsprogram.com/pubs/pdf/SAR7/SAR7.pdf.

Daniel Cai and Randall Balestriero. No location left behind: Measuring and improving the fairness of implicit representations for Earth data. In *International Conference on Learning Representations (ICLR)*, pp. 59288–59314, 2025. URL https://openreview.net/forum?id=hSZaCIznB2.

Gavin C. Cawley and Nicola L. C. Talbot. On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, 11:2079–2107, 2010. URL https://www.jmlr.org/papers/v11/cawley10a.html.

CDC. PLACES: Local Data for Better Health, ZCTA data (2023 release). https://data.cdc.gov/d/9umn-c3jf, 2023.

Daniel Cher, Hamza Iqbal, Eric Xing, Brian Wei, and Nathan Jacobs. Tessellating the Earth. In *European Conference on Computer Vision (ECCV)*, 2026. URL https://arxiv.org/abs/2606.27514.

Gordon Christie, Neil Fendley, James Wilson, and Ryan Mukherjee. Functional map of the world. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6172–6180, 2018. doi: 10.1109/CVPR.2018.00646.

Elijah Cole, Grant Van Horn, Christian Lange, Alexander Shepard, Patrick Leary, Pietro Perona, Scott Loarie, and Oisin Mac Aodha. Spatial implicit neural representations for global-scale species mapping. In *International Conference on Machine Learning (ICML)*, pp. 6320–6342. PMLR, 2023. URL https://arxiv.org/abs/2306.02564.

Aayush Dhakal, Srikumar Sastry, Subash Khanal, Adeel Ahmad, Eric Xing, and Nathan Jacobs. RANGE: Retrieval augmented neural fields for multi-resolution geo-embeddings. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24680–24689. IEEE, 2025. URL https://arxiv.org/abs/2502.19781.

Johannes Dollinger, Damien Robert, Elena Plekhanova, Lukas Drees, and Jan Dirk Wegner. Climplicit: Climatic implicit embeddings for global ecological tasks. In *International Conference on Learning Representations (ICLR) Workshop on Tackling Climate Change with Machine Learning*, 2025. URL https://arxiv.org/abs/2504.05089.

Zhengpeng Feng, Sadiq Jaffer, Ira Shokar, Jovana Knezevic, James Ball, Pedro Sousa, Mark Elvers, Madeline Lisaius, Clement Atzberger, Robin Young, Aneesh Naik, Niall Robinson, David Coomes, Anil Madhavapeddy, and Srinivasan Keshav. TESSERA v2: Scaling pixel-wise Earth foundation models. arXiv preprint arXiv:2607.03949, 2026. URL https://arxiv.org/abs/2607.03949.

Stephen E. Fick and Robert J. Hijmans. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37(12):4302–4315, 2017. doi: 10.1002/joc.5086.

GeoNames. GeoNames geographical database. https://www.geonames.org/, 2026. Accessed: 2026-09-19.

Tomislav Hengl, Gerard B. M. Heuvelink, and David G. Rossiter. About regression-kriging: From equations to case studies. *Computers & Geosciences*, 33(10):1301–1315, 2007. doi: 10.1016/j.cageo.2007.05.001.

Robert J. Hijmans. Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology*, 93(3):679–688, 2012. doi: 10.1890/11-0826.1.

Nicolas Karasiak, Jean-François Dejoux, Claude Monteil, and David Sheeren. Spatial dependence between training and test sets: another pitfall of classification accuracy assessment in remote sensing. *Machine Learning*, 111(7):2715–2740, 2022. doi: 10.1007/s10994-021-05972-1.

Konstantin Klemmer, Esther Rolf, Caleb Robinson, Lester Mackey, and Marc Rußwurm. SatCLIP: Global, general-purpose location embeddings with satellite imagery. *Proceedings of the AAAI Conference on Artificial Intelligence*, 39(4):4347–4355, 2025. doi: 10.1609/aaai.v39i4.32457. URL https://arxiv.org/abs/2311.17179.

Konstantin Klemmer, Esther Rolf, Marc Rußwurm, Gustau Camps-Valls, Mikolaj Czerkawski, Stefano Ermon, Alistair Francis, Nathan Jacobs, Hannah 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: Toward artificial intelligence-centric representations of our planet. *IEEE Geoscience and Remote Sensing Magazine*, pp. 2–15, 2026. doi: 10.1109/MGRS.2026.3710416.

Aditya Kusupati, Gantavya Bhatt, Aniket Rege, Matthew Wallingford, Aditya Sinha, Vivek Ramanujan, William Howard-Snyder, Kaifeng Chen, Sham Kakade, Prateek Jain, and Ali Farhadi. Matryoshka representation learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. URL https://arxiv.org/abs/2205.13147.

Kevin Lane, Zhongying Wang, Esther Rolf, and Morteza Karimzadeh. SLED: Scalable location encoding via distillation. arXiv preprint arXiv:2608.06612, 2026. URL https://arxiv.org/abs/2608.06612.

Chen-Hsuan Lin, Wei-Chiu Ma, Antonio Torralba, and Simon Lucey. BARF: Bundle-adjusting neural radiance fields. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5721–5731, 2021. doi: 10.1109/ICCV48922.2021.00569. URL https://arxiv.org/abs/2104.06405.

David B. Lindell, Dave Van Veen, Jeong Joon Park, and Gordon Wetzstein. BACON: Band-limited coordinate networks for multiscale scene representation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16231–16241, 2022. doi: 10.1109/CVPR52688.2022.01577. URL https://arxiv.org/abs/2112.04645.

Zeping Liu, Ni Lao, Zhangyu Wang, Junfeng Jiao, and Gengchen Mai. GAIR: Location-aware self-supervised contrastive pre-training with geo-aligned implicit representations. *ISPRS Journal of Photogrammetry and Remote Sensing*, 237:166–182, 2026. doi: 10.1016/j.isprsjprs.2026.04.035. URL https://arxiv.org/abs/2503.16683.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations (ICLR)*, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7. arXiv:1711.05101.

Gengchen Mai, Krzysztof Janowicz, Bo Yan, Rui Zhu, Ling Cai, and Ni Lao. Multi-scale representation learning for spatial feature distributions using grid cells. In *International Conference on Learning Representations (ICLR)*, 2020. URL https://arxiv.org/abs/2003.00824.

Gengchen Mai, Krzysztof Janowicz, Yingjie Hu, Song Gao, Bo Yan, Rui Zhu, Ling Cai, and Ni Lao. A review of location encoding for GeoAI: methods and applications. *International Journal of Geographical Information Science*, 36(4):639–673, 2022. doi: 10.1080/13658816.2021.2004602.

Gengchen Mai, Ni Lao, Yutong He, Jiaming Song, and Stefano Ermon. CSP: Self-supervised contrastive spatial pre-training for geospatial-visual representations. In *International Conference on Machine Learning (ICML)*, pp. 23498–23515. PMLR, 2023a. URL https://arxiv.org/abs/2305.01118.

Gengchen Mai, Yao Xuan, Wenyun Zuo, Yutong He, Jiaming Song, Stefano Ermon, Krzysztof Janowicz, and Ni Lao. Sphere2Vec: A general-purpose location representation learning over a spherical surface for large-scale geospatial predictions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 202:439–462, 2023b. doi: 10.1016/j.isprsjprs.2023.06.016.

Carles Milà, Jorge Mateu, Edzer Pebesma, and Hanna Meyer. Nearest neighbour distance matching leave-one-out cross-validation for map validation. *Methods in Ecology and Evolution*, 13(6):1304–1316, 2022. doi: 10.1111/2041-210X.13851.

R. Kelley Pace and Ronald Barry. Sparse spatial autoregressions. *Statistics & Probability Letters*, 33(3):291–297, 1997. doi: 10.1016/S0167-7152(96)00140-X.

Pierre Ploton, Frédéric Mortier, Maxime Réjou-Méchain, Nicolas Barbier, Nicolas Picard, Vivien Rossi, Carsten Dormann, Guillaume Cornu, Gaëlle Viennois, Nicolas Bayol, Alexei Lyapustin, Sylvie Gourlet-Fleury, and Raphaël Pélissier. Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature Communications*, 11:4540, 2020. doi: 10.1038/s41467-020-18321-y.

Laura Poggio, Luis M. de Sousa, Niels H. Batjes, Gerard B. M. Heuvelink, Bas Kempen, Eloi Ribeiro, and David Rossiter. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *SOIL*, 7(1):217–240, 2021. doi: 10.5194/soil-7-217-2021.

Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred A. Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *International Conference on Machine Learning (ICML)*, pp. 5301–5310. PMLR, 2019. URL https://arxiv.org/abs/1806.08734.

Mike Ranzinger, Jon Barker, Greg Heinrich, Pavlo Molchanov, Bryan Catanzaro, and Andrew Tao. PHI-S: Distribution balancing for label-free multi-teacher distillation. arXiv preprint arXiv:2410.01680, 2024. URL https://arxiv.org/abs/2410.01680.

David R. Roberts, Volker Bahn, Simone Ciuti, Mark S. Boyce, Jane Elith, Gurutzeta Guillera-Arroita, Severin Hauenstein, José J. Lahoz-Monfort, Boris Schröder, Wilfried Thuiller, David I. Warton, Brendan A. Wintle, Florian Hartig, and Carsten F. Dormann. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8): 913–929, 2017. doi: 10.1111/ecog.02881.

Esther Rolf. Evaluation challenges for geospatial ML. In *International Conference on Learning Representations (ICLR) Workshop on Machine Learning for Remote Sensing*, 2023. URL https://arxiv.org/abs/2303.18087.

Esther Rolf, Jonathan Proctor, Tamma Carleton, Ian Bolliger, Vaishaal Shankar, Miyabi Ishihara, Benjamin Recht, and Solomon Hsiang. A generalizable and accessible approach to machine learning with global satellite imagery. *Nature Communications*, 12(1):4392, 2021. doi: 10.1038/s41467-021-24638-z.

Marc Rußwurm, Konstantin Klemmer, Esther Rolf, Robin Zbinden, and Devis Tuia. Geographic location encoding with spherical harmonics and sinusoidal representation networks. In *International Conference on Learning Representations (ICLR)*, pp. 1746–1759, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/073c8584ef86bee26fe9d639ec648e28-Paper-Conference.pdf.

Srikumar Sastry, Subash Khanal, Aayush Dhakal, Adeel Ahmad, and Nathan Jacobs. TaxaBind: A unified embedding space for ecological applications. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 1765–1774, 2025. URL https://arxiv.org/abs/2411.00683.

Siyuan Shen, Chi Li, Aaron van Donkelaar, Nathan Jacobs, Chenguang Wang, and Randall V. Martin. Enhancing global estimation of fine particulate matter concentrations by including geophysical a priori information in deep learning. *ACS ES&T Air*, 1(5):332–345, 2024. doi: 10.1021/acsestair.3c00054.

Siyuan Shen, Aaron van Donkelaar, Nathan Jacobs, Chi Li, and Randall V. Martin. Enhancing estimation of fine particulate matter chemical composition across North America by including geophysical a priori information in deep learning with uncertainty quantification. *ACS ES&T Air*, 3(2):336–350, 2026. doi: 10.1021/acsestair.5c00251.

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)*, 2020. URL https://arxiv.org/abs/2006.09661.

Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. URL https://arxiv.org/abs/2006.10739.

Roozbeh Valavi, Jane Elith, José J. Lahoz-Monfort, and Gurutzeta Guillera-Arroita. blockCV: An R package for generating spatially or environmentally separated folds for k-fold cross-validation of species distribution models. *Methods in Ecology and Evolution*, 10(2):225–232, 2019. doi: 10.1111/2041-210X.13107.

Grant Van Horn, Oisin Mac Aodha, Yang Song, Yin Cui, Chen Sun, Alex Shepard, Hartwig Adam, Pietro Perona, and Serge Belongie. The iNaturalist species classification and detection dataset. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8769–8778, 2018. doi: 10.1109/CVPR.2018.00914.

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, pp. 8690–8701, 2023. doi: 10.52202/075280-0379.

Nemin Wu, Qian Cao, Zhangyu Wang, Zeping Liu, Yanlin Qi, Jielu Zhang, Joshua Ni, Xiaobai Yao, Hongxu Ma, Lan Mu, Stefano Ermon, Tanuja Ganu, Akshay Nambi, Ni Lao, and Gengchen Mai. TorchSpatial: A location encoding framework and benchmark for spatial representation learning. In *Advances in Neural Information Processing Systems (NeurIPS) Datasets and Benchmarks Track*, 2024. URL https://arxiv.org/abs/2406.15658.

Jiawei Yang, Marco Pavone, and Yue Wang. FreeNeRF: Improving few-shot neural rendering with free frequency regularization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8254–8263, 2023. doi: 10.1109/CVPR52729.2023.00798. URL https://arxiv.org/abs/2303.07418.

Christopher Yeh, Chenlin Meng, Sherrie Wang, Anne Driscoll, Erik Rozi, Patrick Liu, Jihyeon Lee, Marshall Burke, David B. Lobell, and Stefano Ermon. SustainBench: Benchmarks for monitoring the Sustainable Development Goals with machine learning. In *Neural Information Processing Systems Track on Datasets and Benchmarks (NeurIPS)*, 2021. URL https://arxiv.org/abs/2111.04724.

Biao Zhang, Lixin Chen, Tong Liu, and Bo Zheng. SMEC: Rethinking Matryoshka representation learning for retrieval embedding compression. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 26209–26222, 2025. doi: 10.18653/v1/2025.emnlp-main.1332. URL https://arxiv.org/abs/2510.12474.