

# MEASURING THE INTRINSIC DIMENSION OF EARTH REPRESENTATIONS

**Anonymous authors**

Paper under double-blind review

## ABSTRACT

Within the context of representation learning for Earth observation, geographic Implicit Neural Representations (INRs) embed low-dimensional location inputs (longitude, latitude) into high-dimensional embeddings, through models trained on geo-referenced satellite, image or text data. Despite the common aim of geographic INRs to distill Earth’s data into compact, learning-friendly representations, we lack an understanding of how much information is contained in these Earth representations, and where that information is concentrated. The intrinsic dimension of a dataset measures the number of degrees of freedom required to capture its local variability, regardless of the ambient high-dimensional space in which it is embedded. This work provides the first study of the intrinsic dimensionality of geographic INRs. Analyzing INRs with ambient dimension between 256 and 512, we find that their intrinsic dimensions fall roughly between 2 and 10 and are sensitive to changing spatial resolution and input modalities during INR pre-training. Furthermore, we show that the intrinsic dimension of a geographic INR correlates with downstream task performance and can capture spatial artifacts, facilitating model evaluation and diagnostics. More broadly, our work offers an architecture-agnostic, label-free metric of information content that can enable unsupervised evaluation, model selection, and pre-training design across INRs.

## 1 INTRODUCTION

Across vision, audio, and other modalities, seemingly high-dimensional observations often vary along far fewer degrees of freedom. This is especially true of geographic data, which is often characterized by strong spatio-temporal dependencies. For example, classical work in meteorology use dimensionality reduction techniques since large-scale oscillations in climate trends can be explained by a handful of indices (van den Dool, 2006). This phenomenon is leveraged by a class of representation learning techniques aimed at embedding signals in Earth’s data into succinct, general purpose vector representations (Rolf et al., 2025). This is done either through direct embedding of geo-referenced data with image or text encoders, or through a new class of geographic implicit neural representations (INRs) that encode geospatial signals in the weights of a location encoder network which takes geographic position (latitude and longitude) as input.

Currently, the quality of Earth representation models is evaluated largely in terms of supervised model performance for specific downstream tasks. Pre-trained geographic INRs have driven state-of-the-art performance in tasks like land cover segmentation, object detection, and image geo-localization (Cepeda et al., 2023; Klemmer et al., 2025; Mai et al., 2023a; Liu et al., 2025). In addition, geographic INRs are increasingly used to improve geospatial data interpolation (Mac Aodha et al., 2019; Chen et al., 2025; Lange et al., 2025), for instance, in global species distribution modeling (Cole et al., 2023; Dhakal et al., 2025). While task-specific metrics evaluate the “learning-friendliness” property of location encoder models (as outlined

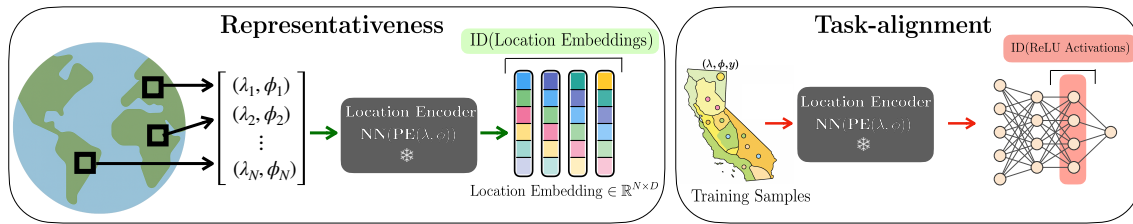


Figure 1: **Estimating the intrinsic dimension (ID) of geographic implicit neural representations (INRs).** We compute the ID of geographic INRs in two ways, to measure model representativeness and task-alignment. **Representativeness** (left): We generate location embeddings with frozen pre-trained location encoders for coordinates across Earth’s land mass. We calculate the global and local ID values on the resulting embeddings. **Task-alignment** (right): We train a downstream task-specific model using location embeddings as input. We use a TwoNN ID estimator to measure the ID of the activations of the task-specific model’s last hidden layer.

by Mai et al. (2022)), focusing only on task-specific metrics prevents us from measuring progress on a fundamental aim of location embeddings: to generate rich, general purpose representations of Earth’s data.

In this work, we study the intrinsic dimension (ID) of geographic INRs as a task-agnostic (unsupervised) metric to quantify information-richness across space and where that capacity is concentrated. Defined in Section 2.2, the intrinsic dimension measures how many independent directions a learned representation actually uses. We compute point-wise, *local* estimates of ID to capture regional effects as well as *global* ID metrics for comparison between different location encoder models. Our results show that the intrinsic dimension reveals two important, and previously unexplored properties of geographic INRs: (i) **representativeness** – the amount of independent, non-redundant variation of the INRs and (ii) **task-alignment** – indicating how well downstream predictors can compress the INRs onto a low-dimensional, target-aligned manifold. Our overall methodology is summarized in Figure 1 and our key findings can be summarized as follows:

- Global ID estimates of current geographic INRs are an order of magnitude lower than their ambient size, yet are competitive with ID estimates of Earth embeddings generated with large-scale image encoders.
- ID estimates of geographic INRs increase with additional input modalities and increased spatial resolution of the location encoder, highlighting that ID captures increases in spectral and spatial representativeness of these embeddings.
- Local ID estimates reveal spatial artifacts of pre-trained geographic INRs, which can arise from biases in the pre-training dataset coverage or properties of their model architectures.
- Global ID correlates with downstream task performance across encoders and tasks. Interestingly, correlation is positive when ID is calculated in the embedding space of frozen, pre-trained models, but is negative when ID is calculated in the activation space of supervised downstream models. This sheds light on the connections between the intrinsic dimension, representativeness, and task-alignment of geographic INRs.

While much of our analysis extends naturally to general classes of INRs, geographic INRs are a particularly interesting case to study because the underlying domain geometry is explicit. Inputs lie on the sphere ( $S^2$ ), which makes it possible to separate the known 2D manifold from learned information content above it. Moreover, the explicit goal of many geographic Earth embedding models is to compress and coalesce as much signal from Earth’s data as possible. The strategies we present for estimating different types of ID offer a new unsupervised evaluation strategy for representation learning of Earth data.

## 2 RELATED WORK

### 2.1 GEOGRAPHIC INRS AND LOCATION ENCODERS

Implicit neural representations (INRs) encode signals as functions that map coordinates to values using neural networks (Chen & Zhang, 2019; Mildenhall et al., 2020). Analogously, geographic INRs encode geographic signals on the sphere using location encoder networks. These models take geographic coordinates (longitude, latitude) as input and return a corresponding value or vector embedding (Mai et al., 2022). Location encoders can be trained directly using a supervised loss to, for instance, interpolate between species observations (Cole et al., 2023) or sea ice thickness measurements (Chen et al., 2025). Alternatively, they can be (pre)trained on unlabeled data, e.g. using a contrastive objective, to obtain an *embedding vector*.

Location encoder architectures consist of a positional encoding (PE), projecting longitude/latitude inputs into a higher-dimensional feature space, and a neural network projecting these features into the desired output space. Popular positional encodings include multi-scale sinusoidal functions (e.g. Sphere2Vec (Mai et al., 2023b) and Space2Vec (Mai et al., 2020)), multi-scale Random Fourier Features (RFFs) as used in GeoCLIP (Cepeda et al., 2023), and spherical harmonic functions that provide an orthogonal, sphere-native basis (Rußwurm et al., 2024). The “resolution” of the location embeddings is controlled by these positional encoding hyperparameters. The most common pre-training objective for location encoders is contrastive image-location matching. This way, location-specific image features are encoded in the location encoder network and—at inference time—can be accessed by providing solely coordinate inputs. Examples include SatCLIP (Klemmer et al., 2025) and GeoCLIP (Cepeda et al., 2023), which are available as pre-trained models and can be seamlessly integrated in other frameworks, for instance, in location-aware image synthesis (Sastrey et al., 2024) or super-resolution (Panangian & Bittner, 2025).

An alternative way to obtain location embedding vectors is to download raw data (e.g. a satellite image) at one location and use a pre-trained vision model as an image encoder. Single-modality pre-trained ResNets and ViTs are available in repositories like TorchGeo (Stewart et al., 2025) and SSL4EO (Wang et al., 2023). Recent work on multi-modal remote sensing foundation models like DOFA (Xiong et al., 2024), CROMA (Fuller et al., 2023), or MMEarth (Nedungadi et al., 2024) are other large-scale geospatial image encoders. In this work, we primarily study the intrinsic dimension of pre-trained location encoders as continuous geographic INRs, but also compare them quantitatively to embeddings from image encoders.

### 2.2 INTRINSIC DIMENSION (ID)

The intrinsic dimension of a dataset can be thought of as a nonlinear analogue of Principal Component Analysis (PCA). While PCA finds a single global linear rank, ID captures the minimal number of degrees of freedom needed to describe data locally. ID is uniquely suited for measuring the true dimensionality of data representations since they occupy a curved manifold (Ansuini et al., 2019). Intrinsic dimension has been used in deep learning in several ways: to define a normalized notion of task difficulty (Li et al., 2018), explain generalization ability and sample efficiency (Pope et al., 2021; Ansuini et al., 2019; Gong et al., 2019), characterize adversarial examples (Ma et al., 2018), detect AI-generated content (Lorenz et al., 2023; Tulchinskii et al., 2023), or to regularize local or joint input–feature dimensions to improve self-supervised representations (Huang et al., 2024; Zhu et al., 2018). Learning theory treats intrinsic dimensionality as a key factor influencing learnability (Narayanan & Niyogi, 2009; Narayanan & Mitter, 2010).

The intrinsic dimension can be calculated with *distance-based* or *angle-based* estimators. Distance-based estimators calculate an intrinsic dimension from how neighbor distances grow around a point. If, in a small ball, density is roughly constant, then the number of points within radius  $r$  scales like  $r^d$ . The ratios of nearest-neighbor radii encode the intrinsic dimension  $d$ . Examples include the Maximum Likelihood (MLE) (Levina & Bickel, 2005) and Two-nearest-neighbor (TwoNN) estimators. Angle-based estimators infer  $d$  from the angular spread of neighbor directions. They are robust to local spatial variabilities by recentering

the point’s neighborhood and using only the unit directions to its neighbors. Directions do not change under local rescaling, which makes angle-based estimators less sensitive to changing spatial patterns. For instance, FisherS (Albergante et al., 2019) summarizes how the neighbor directions spread around a point and converts that spread into an ID estimate. It relies on the fact that in high dimensions, points sampled uniformly on a sphere tend to be almost orthogonal, so a typical point can be linearly separated from the rest by a Fisher discriminant. Formal definitions of the different estimators are given in Appendix A.

To our knowledge, no prior work measures the intrinsic dimensionality of implicit neural representations—especially *geographic* INRs; our study is the first to provide these measurements and link ID to generalization and representativeness in geographical settings.

### 3 ESTIMATING THE ID OF GEOGRAPHIC INRS

We model a pre-trained location encoder as a map  $f : S^2 \rightarrow \mathbb{R}^D$  that returns an embedding  $z = f(x)$  for a geographic location  $x = (\lambda, \phi)$ . The intrinsic dimension at  $x$ , which we denote as  $d(x)$ , summarizes how many independent directions the embedding  $z$  varies when we perturb  $x$  slightly on Earth’s surface. Equivalently,  $d(x)$  is the smallest number of coordinates needed to describe the support of the embedding distribution in a small neighborhood of  $z$ . The ambient dimension  $D$  of the embedding is fixed by the last layer of the location encoder architecture, whereas  $d(x) \leq D$  reflects how much distinct geographic signal the encoder actually expresses at that location.

We estimate the geographic INR ID across two scales: The **local ID**  $d(x)$  reveals where the representation is complex or compressive across Earth’s surface, while the **global ID** aggregates  $d(x)$  over a specified set of locations to provide a single scalar value. We report both: local ID maps can diagnose spatial heterogeneity and the global ID allows us to compare between location encoders. High local ID values signal regions where embeddings vary along many independent directions whereas low values reveal compressive regimes where the representation is effectively one- or two-dimensional.

The choice of angle- versus distance-based ID estimators is also important in our analysis: Angle-based estimators’ robustness to spatial heterogeneity make them a natural choice to estimate a *global* intrinsic dimension over the Earth’s surface. For global analyses, distance-based estimators can be disproportionately biased by local patterns as they read changes in spacing as changes in dimension. Within the context of heterogeneous representations of the Earth, these local patterns can be induced by changing climate zones or terrain edge effects, for instance, in coastal areas. However, this sensitivity can be beneficial in *locally* analyzing where these intrinsic dimensions change spatially. Thus, our measurements of local ID use distance-based estimators.

#### 3.1 MEASURING REPRESENTATIVENESS AND TASK-ALIGNMENT WITH ID

As illustrated in Figure 1, we estimate intrinsic dimension (ID) at two different stages:

1. **Measuring representativeness in embedding space:** With a frozen location encoder  $f$  and *globally* sampled  $N$  geographic coordinates  $(\lambda, \phi)$ , we form embeddings  $Z_{\text{geo}} = f(\text{PE}(\lambda, \phi)) \in \mathbb{R}^{N \times D}$  and estimate the global and local ID of these embeddings. When correlating global ID to downstream task performance, we use the angle-based FisherS estimator.
2. **Measuring task-alignment in activation space (dataset-conditioned):** We then train a shallow classifier on task specific locations and compute ID with the distance-based TwoNN estimator on the penultimate ReLU activations evaluated only at the dataset’s coordinates for each split. Concretely, for the train/validation/test splits we pass their spatial coordinates through  $f$ , feed the resulting embeddings to the classifier, and estimate TwoNN ID jointly for the full dataset. This follows established practice of measuring the ID of neural network representations first introduced in Ansuini et al. (2019).



## 3.2 DATASETS AND EXPERIMENTS

In our experiments, we correlate the estimated ID of different geographic INRs (detailed in section 2.1) with downstream task performance on several geospatial regression and classification tasks that use either (i) only location  $(\lambda, \phi)$  context or (ii) location and additional context (e.g. an image). Location-based regression tasks include air-temperature (Hooker et al., 2018), and elevation, population density, nightlights, and tree-cover prediction from Rolf et al. (2021). Location-based classification include biome and countries classification from Klemmer et al. (2025). We also measure the ID-performance relationship on several image-location regression tasks aimed at measuring socioeconomic outcomes from the SustainBench (Yeh et al., 2021) benchmark. We use the task setup, including labels and precomputed InceptionV3 image features provided through the TorchSpatial benchmark for location encoders (Wu et al., 2024).

Our downstream classification/regression heads trained on these tasks are either a 2 or 3-layer MLP. For the SustainBench image-location regression tasks, we use one 2-layer MLP branch that receives image features as input, and a 5-layer MLP which takes the geographic coordinates as input. All fine-tuning experiments train for between 20-50 epochs with an early-stopping condition on a held-out validation loss. We use a grid search with the Optuna framework (Akiba et al., 2019) to determine the optimal learning rate, best hidden dimension size of the MLP, and best values for weight decay. Mean performance metrics across 10 random seeds are reported.

When measuring the effect additional input modalities have on geographic INR ID and representativeness (Section 4.3), we use MMEarth, a multi-modal Earth observation corpus with Sentinel-2 optical, Sentinel-1 SAR, ASTER GDEM terrain, ETH-GCHM canopy height, Dynamic World land cover, and ESA World-Cover data. Within SatCLIP we compare three variants: (i) an S2-only baseline with a MoCo-pre-trained ResNet-50 image encoder trained on MMEarth’s Sentinel-2, (ii) a SatCLIP location encoder pre-trained on Sentinel-1 and Sentinel-2 imagery, and (iii) SatCLIP pre-trained on all MMEarth pixel-level modalities. For each model we compute global FisherS ID on uniformly sampled land-only coordinates, then train a small 3-layer MLP on frozen embeddings to predict air temperature, elevation, and population-density values.

## 4 RESULTS

### 4.1 GLOBAL AND LOCAL ID OF GEOGRAPHIC INRS

**Global geographic INR ID is significantly lower than its ambient dimension but often greater than 2.** Table 1 presents global ID estimates for geographic INRs from pre-trained location encoders and commonly used geospatial image encoders. We observe that the global IDs are substantially lower than the ambient embedding dimensions ( $D$ ) for all geographic INRs. SatCLIP and CSP embeddings are 256-dimensional while GeoCLIP uses 512 dimensions yet all ID estimates for these models fall below 14.

The ID varies substantially by location encoder architecture. GeoCLIP shows IDs of 11-13 across distance-based estimators, CSP variants range from 3-6, while SatCLIP remains lowest at 2-2.5. Estimator choice also matters: Angle-based FisherS produces notably different patterns—yielding 8.08 for SatCLIP-L40 (versus 2-2.4 for distance-based methods MLE, MOM, TLE) but drops below 2 for CSP models.

**Global IDs of geographic INRs are similar to that of embeddings derived from image encoders.** The intrinsic dimension of geographic location encoders (purely with a  $(\lambda, \phi)$  context) record similar intrinsic dimension estimates compared to large-scale image encoders evaluated on S2-100K Sentinel-2 tiles (Table 1). GeoCLIP’s ID of 11-13 approaches that of foundation models like DOFA (ID: 14-16) (Xiong et al., 2024) and CROMA (ID: 17-20) (Fuller et al., 2023). This indicates that current pre-trained location encoders contain a similar amount of overall information content as embeddings of multi-spectral satellite imagery obtained through specialized image encoders, as measured through these global ID estimators.

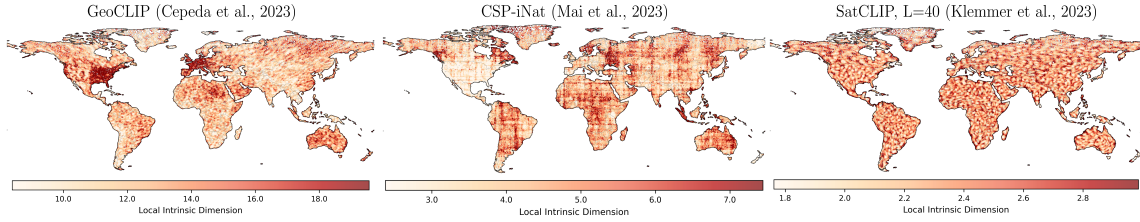


Figure 2: **Local intrinsic dimension of geographic INRs reveal spatial artifacts.** We use the MLE estimator on embeddings generated over Earth’s landmass.  $N = 100,000$  points sampled with  $k = 100$  neighbors used in the MLE ID calculation. We plot the local ID of more INRs in Appendix Figure 11.

**Local estimates of ID reveal spatial artifacts of pre-trained geographic location encoders.** Figure 2 plots local ID estimates of pre-trained geographic INRs using the distance-based MLE estimator with  $k = 100$  nearest neighbors. (Results are not very sensitive to the choice of  $k$ , which we further evaluate in Appendix B). For GeoCLIP, local ID is highest in the United States and western Europe, reflecting the spatial distribution of the social-media images on which it is pre-trained. The local IDs of CSP show a grid pattern because its positional encoding repeats at regular steps in longitude and latitude, so equally spaced locations look alike and form bands along meridians and parallels. For SatCLIP, local ID maps show no regional coverage bias (consistent with S2-100K’s global sampling), but they exhibit thin, periodic oscillations, reflecting the finite-order spherical-harmonic functions used in its location encoder.

#### 4.2 GEOGRAPHIC INR ID CORRELATES WITH TASK PERFORMANCE

##### Geographic INRs with high global ID record

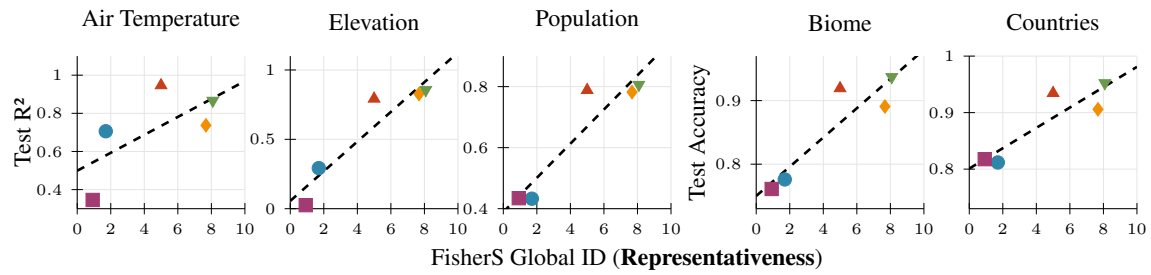
**higher task performance.** In Figure 3a, we compute the *global* FisherS ID in embedding space to measure representativeness as pictured in the left panel in Figure 1. We then correlate these ID estimates with the performance of task-specific supervised learning models (small MLPs trained on top of embeddings from frozen geographic INRs). Across datasets and tasks, the scatter plots exhibit a clear positive linear trend: geographic INRs with larger global ID lead to higher downstream performance. A higher global ID implies a richer coverage of geographic variability—i.e., more task-relevant directions are available for a shallow learner to exploit with limited supervision.

##### Lower global ID in activation space of supervised models corresponds to higher task performance.

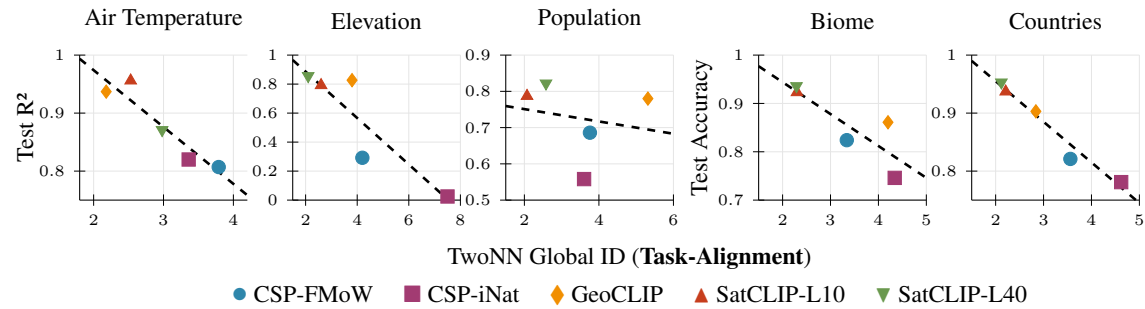
Still keeping the geographic INR frozen, we train a small task head and then measure the global TwoNN ID in activation space of this downstream task head (following the experimental settings in Ansuini et al. (2019)). This measures task-alignment as pictured in the right panel in Figure 1. In Figure 3b, we observe a strong negative correlation between activation-space ID and performance, indicating that supervised adaptation

Table 1: **Global intrinsic dimension of Earth embeddings.** Distance-based estimators use  $k = 20$  nearest neighbors. Appendix Table 2 shows results for more estimators and sampling schemes.

Model	$D$	FisherS	MLE	MOM	TLE
<i>Location encoders, Land sampling (100k points)</i>					
SatCLIP-L10	256	5.00	1.96	2.02	2.16
SatCLIP-L40	256	<b>8.08</b>	2.03	2.39	2.32
GeoCLIP	512	7.68	<b>11.21</b>	<b>13.02</b>	<b>11.53</b>
CSP-fMoW	256	1.70	5.18	5.23	6.25
CSP-iNat	256	0.92	3.37	4.64	4.14
<i>Image encoders on S2-100K (Klemmer et al., 2025)</i>					
RCF	512	1.64	6.32	5.23	7.10
CROMA	768	<b>9.79</b>	19.57	17.00	20.30
DOFA	768	3.32	15.58	13.78	16.20
ResNet18	512	6.32	16.14	12.27	16.80
ResNet50	2048	6.42	16.27	13.18	17.00
ResNet152	2048	7.60	<b>20.72</b>	<b>17.50</b>	<b>21.50</b>
ViT-Small	384	3.33	18.53	15.80	19.20
ScaleMAE (RGB)	1024	2.96	10.16	8.90	11.00
ResNet18 (RGB)	512	0.92	10.85	8.70	11.70
ResNet50 (RGB)	2048	0.92	9.92	8.10	10.80



(a) FisherS global ID of geographic INRs vs test  $R^2$  and top-1 accuracy using land-based coordinate sampling.



(b) TwoNN global ID of ReLU activations of a 3-hidden-layer MLP's penultimate layer vs test  $R^2$  and top-1 accuracy.

Figure 3: **Relationship between global ID of geographic INRs and downstream task performance** measured across five regression and classification tasks. In both rows, the location embeddings are frozen while task-specific predictions heads (3 layer MLPs) are learned. In (a), ID (horizontal axis) is calculated on the frozen pre-trained embeddings as in Table 1. In (b), ID is measured in *activation space* using the TwoNN estimator on a learned classifier's penultimate layer.

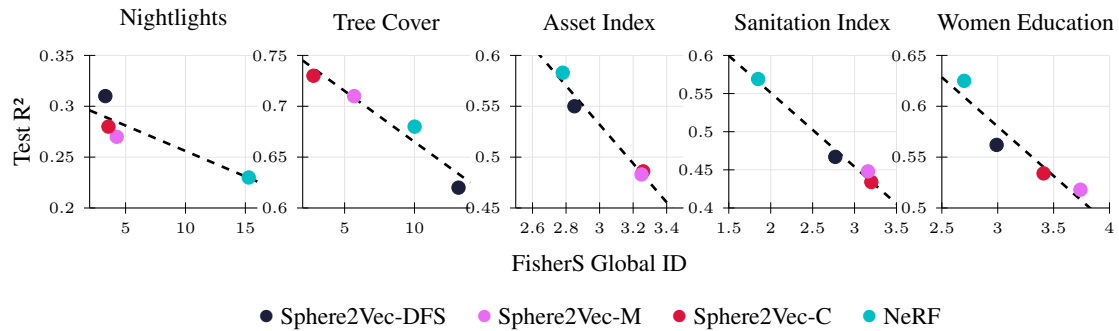


Figure 4: **FisherS global ID of task-specific location embeddings learned via supervised learning vs test  $R^2$  of four continuous location encoders on five tasks from TorchSpatial (Wu et al., 2024)** FisherS ID is calculated on the intermediate location embeddings from the location encoder, similar to Figure 3a. The asset index, sanitation index, and women education tasks are image-location regression tasks, as detailed in Section 3.2.

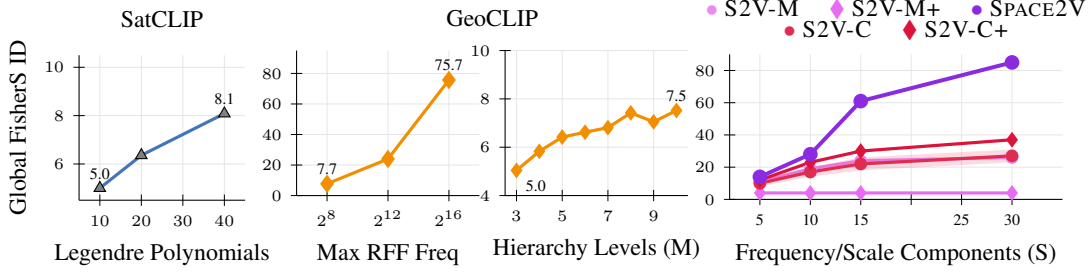


Figure 5: **Effect of location encoder spatial resolution on global ID.** (Left) for SatCLIP, we pre-train the location encoder with  $L = 10, 20$ , and 40 Legendre Polynomials. (Middle) For GeoCLIP, we increase both the maximum RFF frequency and the number of hierarchical levels ( $M$ ) used by the location encoder by fine-tuning the new higher-frequency branches on a YFCC (Thomee et al., 2016) image geo-localization task. (Right) For Sphere2Vec (S2V) and Space2Vec (SPACE2V) encoders, we increase the number of frequency components ( $S$ ) and train the location encoder with supervised learning on the MOSAICS nightlights regression task.

compresses the INR features onto a task-aligned manifold with fewer effective degrees of freedom. This is consistent with past work, that found lower ID indicates more concentrated, linearly separable structure and thus better generalization (Ansuini et al., 2019; Pope et al., 2021; Zhu et al., 2018).

**Within a supervised learning setting, lower ID of task-specific location encoders correspond to higher task performance.** To further investigate task-alignment, we train continuous location encoders Sphere2Vec and Space2Vec end-to-end with supervised learning on the image-location regression tasks outlined in Section 3.2. We then compute the global FisherS ID on the task-specific learned embeddings, across training sample locations. In Figure 4, interestingly, we again find a consistent negative relationship between ID and  $R^2$ , suggesting that direct supervision drives the representations toward a lower-ID, task-specific manifold that is easier to separate or regress on. This mirrors Figure 3b and reinforces the picture that the benefits of self-supervised pre-training stem from *high* global ID (expressivity/coverage), while subsequent supervised models benefit from *low* global ID (compression/task-alignment).

#### 4.3 EFFECT OF RESOLUTION AND INPUT MODALITIES ON THE ID OF GEOGRAPHIC INRS

Having established that ID measures can capture the relative information content of geographic INRs across space and relate to downstream task-specific performance, we now examine how ID values change when the properties of geographic location encoders change. We focus on two properties along which geographic location encoders are routinely modified that should result in different information content in their embeddings: the spatial resolution of the location encoder architecture, and the input modalities used during pre-training.

**Increased spatial resolution of geographic INRs increases global ID and representativeness.** The spatial resolution of several location encoder architectures is controlled by specific model hyperparameters: spherical harmonics by the number of Legendre polynomials used  $L$ , Random Fourier Features (RFF) by their maximum frequency  $\sigma_{\max}$  and hierarchy depth  $M$ , and Space2Vec/Sphere2Vec via the number of multi-scale components  $S$ . In Figure 5, we plot how ID changes as we vary these hyperparameters. Intuitively, increasing resolution should allow the encoder to resolve finer geospatial phenomena and thus use more independent directions in representation space. For SatCLIP (equipped with a spherical harmonic and sinusoidal representation network positional encoder), the global FisherS ID rises with  $L$ . Similarly, for GeoCLIP, increasing  $\sigma_{\max}$  by appending high-frequency branches to the original pre-trained set and increasing the hi-

erarchy density  $M$  both increase global ID. Increasing  $\sigma_{\max}$  produces a sharp increase in global ID, whereas increasing  $M$  produces more gradual increases. Across Space2Vec variants, as we vary the  $S$  parameter, the rise in ID is steepest for the theory/grid formulation of Space2Vec, moderate for the compositional variants, and smallest for the multiplicative variants. This supports the view that adding multi-scale components enlarges the set of independent directions the encoder can use, thereby increasing the representational capacity of the location embedding.

#### Additional pre-training data modalities increases global ID and representativeness.

In Figure 6, we train SatCLIP models with different subsets of pre-training data modalities from the MMEarth dataset. We find that increasing the number of geographic layers seen during training increases both the global ID and its downstream task performance. SatCLIP location encoders with the most number of input modalities (optical Sentinel-1, multispectral Sentinel-2 and all pixel-level modalities on MMEarth) records both the highest global ID and performance across an air-temperature, elevation, and population regression task. This confirms that ID can capture differences in information content due to increasing the number modalities beyond multi-spectral Sentinel-2 imagery. The clear gains to downstream performance in Figure 6 echoes our results linking increased representativeness to increased downstream task performance in Figure 3.

## 5 DISCUSSION AND CONCLUSION

Intrinsic dimension (ID) offers an architecture- and task-agnostic metric to measure the information content encoded in geographic INRs. Despite being a task-agnostic metric, global ID still informs the “learning-friendliness” (Mai et al., 2022) property of a location encoder. When measured on frozen embeddings, higher global ID aligns with stronger downstream performance after fine-tuning, indicating greater representativeness. When measured on the activations of a supervised prediction head, lower global ID accompanies better generalization, consistent with task-aligned compression. Local ID maps expose spatial artifacts that could influence the performance of geographic INRs on downstream applications. Thus, both global and local ID analysis can support model selection (e.g., choosing positional encoders or resolution) via label-free model evaluation at the *pre-training* stage.

We view ID as one component of a broader evaluation toolkit for geographic representation learning. Complementary measures could quantify data provenance by attributing information content to specific pre-training corpora or regions, and can improve interpretability by localizing that content to subspaces of the embedding. The rich connection between ID, model representativeness, and task-alignment that we evidence here has several implications for future work. While our analysis linking ID to downstream performance is descriptive and not causal, future work could examine whether using ID as an optimization target during pre-training leads to information-rich geographic location encoders. Future work could also explore using patterns in local ID to guide the design of task-specific fine-tuning strategies that could differ in regions of high and low information content.

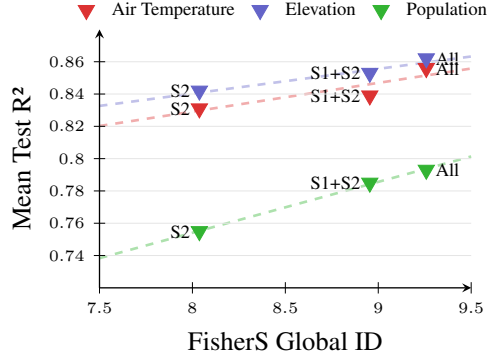


Figure 6: Using additional input modalities during pre-training increases ID and downstream task performance of geographic INRs. Colors represent different tasks. Models are pre-trained on subsets of the MMEarth dataset: Sentinel-2 (S2), Sentinel-1 and 2 (S1+S2), and all available rasters (All).

## REFERENCES

- Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pp. 2623–2631, 2019.
- Luca Albergante, Jonathan Bac, and Andrei Zinovyev. Estimating the effective dimension of large biological datasets using Fisher separability analysis. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. IEEE, 2019.
- Laurent Amsaleg, Oussama Chelly, Teddy Furon, Stéphane Girard, Michael E Houle, Ken-ichi Kawarabayashi, and Michael Nett. Extreme-value-theoretic estimation of local intrinsic dimensionality. *Data Mining and Knowledge Discovery*, 32(6):1768–1805, 2018.
- Laurent Amsaleg, Oussama Chelly, Michael E Houle, Ken-ichi Kawarabayashi, Miloš Radovanović, and Weeris Treeratanajaru. Intrinsic dimensionality estimation within tight localities. In *Proceedings of the 2019 SIAM International Conference on Data Mining*, pp. 181–189. SIAM, 2019.
- Alessio Ansuini, Alessandro Laio, Jakob H. Macke, and Davide Zoccolan. Intrinsic dimension of data representations in deep neural networks. In *Advances in Neural Information Processing Systems*, volume 32, pp. 6109–6119, 2019.
- Vicente Vivanco Cepeda, Gaurav Kumar Nayak, and Mubarak Shah. GeoCLIP: Clip-inspired alignment between locations and images for effective worldwide geo-localization. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. URL <https://openreview.net/forum?id=I18BXotQ7j>.
- Weibin Chen, Azhir Mahmood, Michel Tsamados, and So Takao. Deep random features for scalable interpolation of spatiotemporal data. In *The Thirteenth International Conference on Learning Representations*, 2025. URL <https://openreview.net/forum?id=OD1MV7vf41>.
- Zhiqin Chen and Hao Zhang. Learning implicit fields for generative shape modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5932–5941, 2019.
- Gordon Christie, Neil Fendley, James Wilson, and Ryan Mukherjee. Functional Map of the World. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6172–6180, 2018.
- Elijah Cole, Grant Van Horn, Christian Lange, Alexander Shepard, Patrick Leary, Pietro Perona, Scott Loarie, and Oisín Mac Aodha. Spatial implicit neural representations for global-scale species mapping. In *International conference on machine learning*, pp. 6320–6342. PMLR, 2023.
- 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 Computer Vision and Pattern Recognition Conference*, pp. 24680–24689, 2025.
- Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. URL <https://openreview.net/forum?id=YicbFdNTTy>.
- Anthony Fuller, Koreen Millard, and James Green. CROMA: Remote sensing representations with contrastive radar-optical masked autoencoders. *Advances in Neural Information Processing Systems*, 36: 5506–5538, 2023.

- Sixue Gong, Vishnu Naresh Boddeti, and Anil K Jain. On the intrinsic dimensionality of image representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3987–3996, 2019.
- Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016. doi: 10.1109/CVPR.2016.90.
- Josh Hooker, Gregory Duveiller, and Alessandro Cescatti. A global dataset of air temperature derived from satellite remote sensing and weather stations. *Scientific data*, 5(1):1–11, 2018.
- Hanxun Huang, Ricardo J. G. B. Campello, Sarah Monazam Erfani, Xingjun Ma, Michael E. Houle, and James Bailey. LDReg: Local dimensionality regularized self-supervised learning. In *The Twelfth International Conference on Learning Representations*, 2024. URL <https://openreview.net/forum?id=oZyAqjAjJW>.
- Kerstin Johnsson, Charlotte Soneson, and Magnus Fontes. Low bias local intrinsic dimension estimation from expected simplex skewness. *IEEE transactions on pattern analysis and machine intelligence*, 37(1): 196–202, 2014.
- 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, pp. 4347–4355, 2025.
- Christian Lange, Max Hamilton, Elijah Cole, Alexander Shepard, Samuel Heinrich, Angela Zhu, Subhansu Maji, Grant Van Horn, and Oisín Mac Aodha. Few-shot species range estimation. *arXiv preprint arXiv:2502.14977*, 2025.
- Elizaveta Levina and Peter J. Bickel. Maximum likelihood estimation of intrinsic dimension. In *Advances in Neural Information Processing Systems*, volume 17, pp. 777–784. MIT Press, 2005.
- Chunyu Li, Heerad Farkhor, Rosanne Liu, and Jason Yosinski. Measuring the intrinsic dimension of objective landscapes. In *International Conference on Learning Representations*, 2018.
- Zeping Liu, Fan Zhang, Junfeng Jiao, Ni Lao, and Gengchen Mai. GAIR: Improving multimodal geo-foundation model with geo-aligned implicit representations. *arXiv preprint arXiv:2503.16683*, 2025. URL <https://arxiv.org/abs/2503.16683>.
- Peter Lorenz, Ricard L Durall, and Janis Keuper. Detecting images generated by deep diffusion models using their local intrinsic dimensionality. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 448–459, 2023.
- Xingjun Ma, Bo Li, Yisen Wang, Sarah M. Erfani, Sudanthi N. R. Wijewickrema, Grant Schoenebeck, Dawn Song, Michael E. Houle, and James Bailey. Characterizing adversarial subspaces using local intrinsic dimensionality. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2018. URL <https://openreview.net/forum?id=B1gJ1L2aW4>.
- Oisín Mac Aodha, Elijah Cole, and Pietro Perona. Presence-only geographical priors for fine-grained image classification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9596–9606, 2019.
- 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. OpenReview.

- 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.
- 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)*. PMLR, 2023a.
- 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.
- Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. In *European Conference on Computer Vision (ECCV)*, pp. 405–421, 2020.
- Hariharan Narayanan and Sanjoy Mitter. Sample complexity of testing the manifold hypothesis. In *Advances in Neural Information Processing Systems*, volume 23, 2010.
- Hariharan Narayanan and Partha Niyogi. On the sample complexity of learning smooth cuts on a manifold. In *The 22nd Conference on Learning Theory*, 01 2009.
- Vishal Nedungadi, Ankit Kariryaa, Stefan Oehmcke, Serge Belongie, Christian Igel, and Nico Lang. MMEarth: Exploring multi-modal pretext tasks for geospatial representation learning. In *European Conference on Computer Vision*, pp. 164–182. Springer, 2024.
- Daniel Panangian and Ksenia Bittner. Can location embeddings enhance super-resolution of satellite imagery? In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 6136–6145. IEEE, 2025.
- Phillip Pope, Chen Zhu, Ahmed Abdelkader, Micah Goldblum, and Tom Goldstein. The intrinsic dimension of images and its impact on learning. In *9th International Conference on Learning Representations (ICLR)*, 2021. URL <https://openreview.net/forum?id=XJk19XzGq2J>.
- Catherine Reed, Ritwik Gupta, Shufan Li, Sarah Brockman, Christopher Funk, Brian Clipp, Salvatore Candido, Matt Uyttendaele, and Trevor Darrell. Scale-MAE: A scale-aware masked autoencoder for multi-scale geospatial representation learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4065–4076, 2023.
- Esther Rolf, Jonathan Proctor, Tamma Carleton, Ian Bolliger, Vaishal 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.
- Esther Rolf, Lucia Gordon, Milind Tambe, and Andrew Davies. Contrasting local and global modeling with machine learning and satellite data: A case study estimating tree canopy height in African savannas. *arXiv preprint arXiv:2411.14354*, 2024.
- Esther Rolf, Konstantin Klemmer, and Marc Rußwurm. Earth Embeddings: Harnessing the information in Earth observation data with machine learning. In *Proceedings of the Special Interest Group on Computer Graphics and Interactive Techniques Conference Frontiers, SIGGRAPH Frontiers '25*, New York, NY, USA, 2025. Association for Computing Machinery. doi: 10.1145/3736539.3754446.



- Marc Rußwurm, Konstantin Klemmer, Esther Rolf, Robin Zbinden, and Devis Tuia. Geographic location encoding with spherical harmonics and sinusoidal representation networks. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2024. URL <https://iclr.cc/virtual/2024/poster/18690>.
- Srikumar Sastry, Subash Khanal, Aayush Dhakal, and Nathan Jacobs. Geosynth: Contextually-aware high-resolution satellite image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 460–470, 2024.
- Vincent Sitzmann, Julien Martel, Alexander Bergman, David Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. *Advances in Neural Information Processing Systems*, 33:7462–7473, 2020.
- Adam J Stewart, Caleb Robinson, Isaac A Corley, Anthony Ortiz, Juan M Lavista Ferres, and Arindam Banerjee. Torchgeo: deep learning with geospatial data. *ACM Transactions on Spatial Algorithms and Systems*, 11(4):1–28, 2025.
- Bart Thomee, David A. Shamma, Gerald Friedland, Benjamin Elizalde, Karl Ni, Douglas Poland, Damian Borth, and Li-Jia Li. Yfcc100m: the new data in multimedia research. *Commun. ACM*, 59(2):64–73, January 2016. ISSN 0001-0782. doi: 10.1145/2812802. URL <https://doi.org/10.1145/2812802>.
- Eduard Tulchinskii, Kristian Kuznetsov, Laida Kushnareva, Daniil Cherniavskii, Sergey Nikolenko, Evgeny Burnaev, Serguei Barannikov, and Irina Piontkovskaya. Intrinsic dimension estimation for robust detection of AI-generated texts. *Advances in Neural Information Processing Systems*, 36:39257–39276, 2023.
- Huug van den Dool. *Empirical Methods in Short-Term Climate Prediction*. Oxford University Press, 12 2006. ISBN 9780199202782. doi: 10.1093/oso/9780199202782.001.0001. URL <https://doi.org/10.1093/oso/9780199202782.001.0001>.
- Yi Wang, Nassim Ait Ali Braham, Zhitong Xiong, Chenying Liu, Conrad M Albrecht, and Xiao Xiang Zhu. SSL4EO-S12: A large-scale multimodal, multitemporal dataset for self-supervised learning in Earth observation [software and data sets]. *IEEE Geoscience and Remote Sensing Magazine*, 11(3):98–106, 2023.
- Nemin Wu, Qian Cao, Zhangyu Wang, Zeping Liu, Yanlin Qi, Jielu Zhang, Joshua Ni, X. Angela 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 *The Thirty-eighth Conference on Neural Information Processing Systems (NeurIPS 2024) – Datasets and Benchmarks Track*, 2024. URL <https://openreview.net/forum?id=DErtzUdhkk>.
- Zhitong Xiong, Yi Wang, Fahong Zhang, Adam J Stewart, Joëlle Hanna, Damian Borth, Ioannis Papoutsis, Bertrand Le Saux, Gustau Camps-Valls, and Xiao Xiang Zhu. Neural plasticity-inspired foundation model for observing the Earth crossing modalities. *arXiv preprint arXiv:2403.15356*, 2024.
- Christopher Yeh, Chenlin Meng, Sherrie Wang, Anne Driscoll, Erik Rozi, Patrick Liu, Jihyeon Lee, Marshall Burke, David Lobell, and Stefano Ermon. SustainBench: Benchmarks for monitoring the sustainable development goals with machine learning. In *Thirty-fifth Conference on Neural Information Processing Systems, Datasets and Benchmarks Track (Round 2)*, 12 2021. URL <https://openreview.net/forum?id=5HR3vCylqD>.
- Wei Zhu, Qiang Qiu, Jiaji Huang, Robert Calderbank, Guillermo Sapiro, and Ingrid Daubechies. LDM-Net: Low dimensional manifold regularized neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2743–2751, 2018.