NeurEO: dissecting Earth observation embeddings with computational neuroscience

Thijs L van der Plas[1*] **Jacob JW Bakermans**[2*] **Ioannis N Athanasiadis**[1]
[1]Wageningen University & Research, NL [2]University College London, UK
[*]These authors contributed equally.
{thijs.vanderplas,ioannis.athanasiadis}@wur.nl, j.bakermans@ucl.ac.uk

Abstract

Earth embeddings are compressed representations of rich Earth observation data that enable downstream models to tackle a wide range of complex geospatial tasks. Yet, this increase in performance comes at the cost of decreased interpretability, as the semantic meaning of Earth embeddings is unknown. To address this challenge, we borrow methods from the field of computational neuroscience, which aim to interpret activity in (biological) neurons. Here, we adapt and apply these methods to assign meaning to AlphaEarth embeddings. We analyse embedding features both as stand-alone data, equivalent to spontaneous neural activity, and in relation to land cover and elevation correlates, equivalent to stimulus-response neural activity. We demonstrate correlational and multi-scale structure in the embedding features. Embedding features can display highly specific and directional tuning to land cover and elevation, such as a 'North vs South woodland edge' response. When we remove the main trends of combined land cover responses (like water versus land tuning), we discover fine-grained representational details with increased effective dimensionality. Together, our analyses start to unravel what these Earth embeddings represent, which is critical for developing trustworthy downstream applications.

1 Introduction

Geospatial foundation models are self-supervised models that construct Earth embeddings and aim to capture rich, multi-modal Earth observation data in compressed embedding features [Tseng et al., 2023, Brown et al., 2025, Klemmer et al., 2025]. These embeddings introduce a paradigm shift for downstream geospatial tasks, presenting easily accessible, compressed representations of diverse Earth observation data on which a (simple) classifier or regression model can be trained. However, contrary to the raw Earth observation data with which they are trained, the semantic meaning of embedding representations is unknown *a priori*. In other words, we are presented with a vast source of data, which we know should contain rich information, yet hidden to the observer.

We argue that this scientific challenge – interpreting what geospatial embedding features represent – is commonly encountered in neuroscience, where recordings of neural activity are analysed to assign meaning to individual neurons or groups of neurons [Hubel and Wiesel, 1962, O'Keefe and Nadel, 1978, Rigotti et al., 2013, Mante et al., 2013]. In this paper we explore this connection, by analysing Earth observation embeddings (embedding features $\times$ space) in a similar fashion to neural activity (neurons $\times$ time). Hence, where biological neural activity extends in one dimension (time), embedding features extend in two (space; latitude $\times$ longitude). In this paper we explore a variety of computational neuroscience methods originally designed to interpret neural activity, either spontaneous (*i.e.*, without correlates/hypotheses) or in response to stimuli (vice versa), for both individual neurons and neuron populations, and adapt these to dissect Earth embedding features.

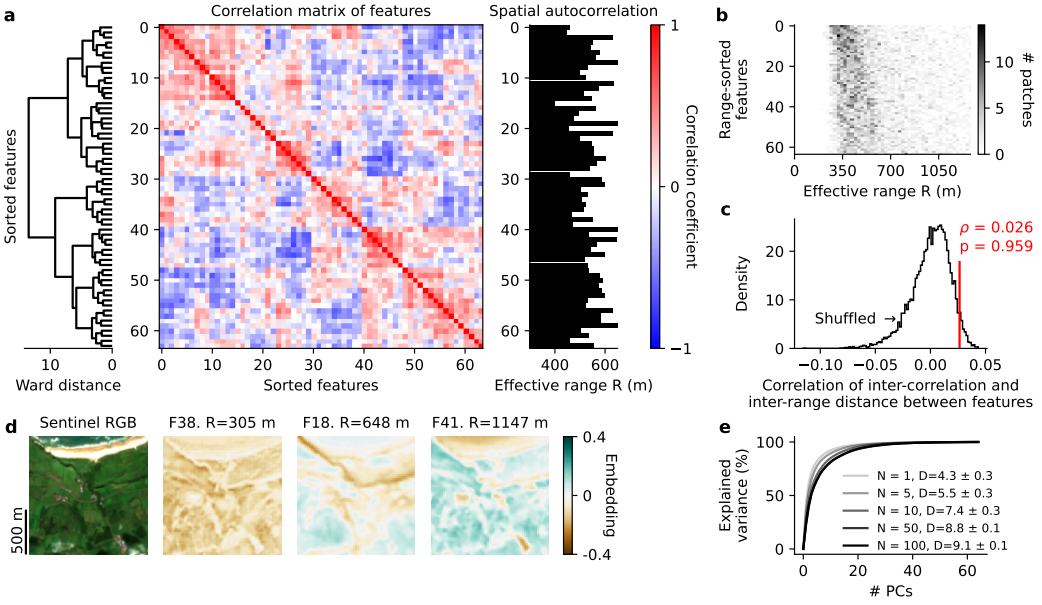

Figure 1: **Embeddings are characterised by their statistical properties. a)** Features exhibit a clustered (Pearson) correlation structure, sorted by hierarchical clustering (left). Median effective ranges (across patches) were calculated per feature (right). **b)** Effective range distributions vary slightly per feature (sorted in this panel only by median effective range). **c)** Variation in effective range is not correlated to variation in feature correlation structure (Pearson's $\rho = 0.026$, two-sided $p = 0.082$, permutation test with $N = 10,000$). **d)** Three example feature embedding maps for one example patch (F: feature, R: effective range for this patch). **e)** The effective dimensionality D of features increases with increasing number of patches N (mean $\pm$ standard error of the mean).

2 Methods

Dataset A proof-of-principle dataset was constructed by randomly subsampling 100 locations from the S2BMS dataset (that contains over 1300 locations of butterfly habitats in the UK, thus representing a wide variety of UK landscapes) [Van der Plas et al., 2025], subsampled such that each pair of locations is at least 30 km apart. For each location coordinate, we downloaded 128 x 128 pixel (= 1.28 km x 1.28 km) patches of 2024 AlphaEarth embeddings [Brown et al., 2025], 2024 DynamicWorld land cover (averaged across all 2024 data points, yielding a number between 0 and 1 for each land cover class) [Brown et al., 2022] and Copernicus 30m DSM elevation [ESA, 2019] data using Google Earth Engine [Gorelick et al., 2017]. Additionally, Sentinel-2 RGB satellite images were downloaded for visualisation purposes only. We define the set of 9 DynamicWorld land cover classes and the DSM elevation layer as the set of 10 hypotheses to correlate embedding features with.

Effective range The (variogram) effective range is defined as the inter-pixel distance where pixel values become statistically independent, and was calculated per feature per patch using a spherical fit, uniform bins and by sampling 4,000 pixels per patch with Scikit GStat [Mälicke, 2022]. Effective range values above the patch size 1,280 m were ignored to avoid patch edge issues, which presumably would require larger patches and sample sizes to resolve (at greater computational cost).

Representational similarity analysis (RSA) In neuroscience, RSA is used to test whether two separate sets of measurements (*e.g.*, neural and behavioural) co-vary together [Kriegeskorte et al., 2008]. Here, we used RSA to test whether effective range differences between features correlated significantly with pixel value correlations between features. We constructed a feature-to-feature negative distance matrix from the median effective ranges (negative distances were used such that higher values correspond to greater similarity) and a Pearson correlation matrix (of pixel values). These two matrices were correlated, and this value was compared against 10,000 permutations (where only the feature identity of the effective ranges were shuffled) to calculate a p value.

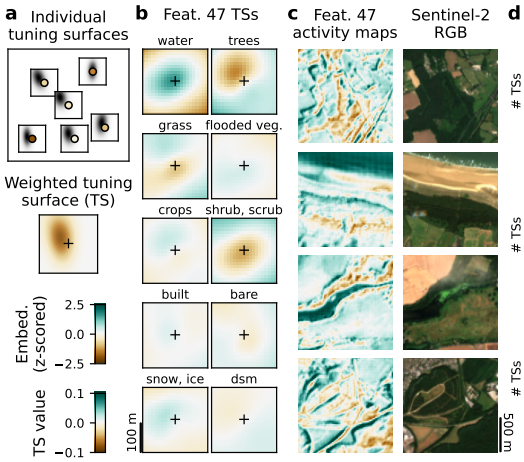

Figure 2: **Embedding tuning surfaces reveal directional coding. a)** Tuning surfaces (TSs) are computed by averaging hypothesis windows weighed by the focal embedding value. **b)** TSs of embedding feature 47, with omnidirectional coding for water and directional coding for trees. **c)** Embedding feature maps of feature 47 show positive coding for water bodies and positive/negative coding for Northern/Southern woodland edges. **d)** TS fit statistics for all feature/hypothesis combinations.

Principal component analysis (PCA) In neuroscience, PCA is used to calculate the effective dimensionality of a population of neurons [Abbott et al., 2011, Farrell et al., 2022]. Here, we used PCA to calculate the effective dimensionality of the 64 embedding features for datasets of N patches to quantify the redundancy among embeddings when used for a limited geographical scope (as defined by the N patches). Dimensionality $D = \frac{1}{M}\sum_m^M \frac{(\sum_i \lambda_i)^2}{\sum_i \lambda_i^2}$ where λ_i is the i^{th} eigenvalue and $M = 20$ subsamples of N patches were resampled for each N. For the PCA analysis of Fig 1e only, 165 instead of 100 S2BMS patches (at least 30 km apart) were used for subsampling.

Tuning surfaces In neuroscience, tuning curves/surfaces express common stimulus properties to which neurons respond, which can be calculated using the 'spike-triggered-average' by temporally aligning stimulus time-series to neural 'spike' times [Schwartz et al., 2006]. Here, we calculated a tuning surface (TS) for each feature/hypothesis combination by spatially aligning and averaging a 21×21 pixel windows of hypothesis maps, each weighted by the embedding value of the central window pixel (Fig 2a). We fit resulting TSs with 2-dimensional Gaussians to obtain TS statistics across all features/hypotheses. TS relative strength was defined as the relative peak value of each hypothesis-TS per feature. TS ellipse shape was characterised by its peak-distance from (0, 0) and eccentricity $\sqrt{1 - \frac{b^2}{a^2}}$ where $a = \max(\sigma_x, \sigma_y)$ and $b = \min(\sigma_x, \sigma_y)$ for stds. σ_x and σ_y.

Canonical correlation analysis (CCA) reconstruction In neuroscience, CCA is used find the linear projections of two sets of covariates that maximally correlate [Wang et al., 2018, Gallego et al., 2020]. We used CCA to find the $N_C = 10$ linear projections $\mathbf{C}_F$ of the embedding feature matrix $\mathbf{F}$ that maximally correlate with linear projections $\mathbf{C}_H$ of the hypotheses matrix $\mathbf{H}$. We then used ordinary least squares (OLS) regression to find the optimal fit of the N_C-dimensional $\mathbf{C}_F$ components to $\mathbf{F}$, which we call the reconstruction $Recon_C$ of $\mathbf{F}$, and equivalently $Recon_H$ using $\mathbf{H}$ to fit $\mathbf{F}$ using OLS. The residuals, $Resid_C$ and $Resid_H$, are then defined by $Resid_k = \mathbf{F} - Recon_k$.

3 Results

We use correlate-free methods to demonstrate correlational and multi-scale structure in the embeddings. We then probe the tuning of individual embedding features to land cover and elevation and find highly specific responses (*e.g.*, 'North vs South woodland edge'). We extend this tuning analysis to groups of features, which reveal fine-grained structure after the main hypothesis trends are removed.

Correlate-free Embedding features show a strong inter-feature correlation structure, leading to clusters of similarly correlated features (Fig 1a). We then used RSA to test whether clusters were predominantly caused by similarities in the spatial auto-correlation of embedding features. First, we calculated the effective range of each feature and constructed a matrix of distances between (median) effective ranges of all feature pairs. We then tested whether this matrix was significantly correlated to the inter-feature correlation structure, but found that is was very weakly correlated ($\rho = 0.026$,

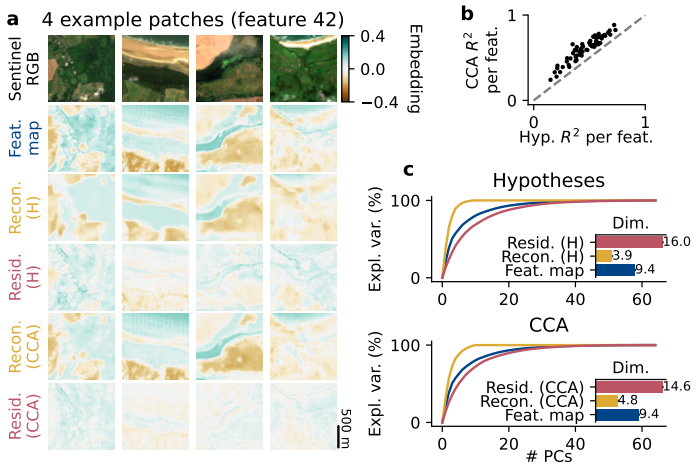

Figure 3: **Embeddings can be linearly split into low-dimensional reconstructions and high-dimensional residuals.** **a)** RGB, embedding feature maps, reconstructions and residuals for hypotheses (H) and CCA for an example embedding feature with $R^2_H = 0.72$ and $R^2_{CCA} = 0.88$. **b)** CCA reconstructions explain more embedding variance (R^2) than hypotheses reconstructions. **c)** PCA dimensionality (dim.) is greatest for residuals, followed by embeddings and reconstructions.

two-sided $p = 0.082$, Fig 1b-d). Thus, similar features may have similar scales, but there is also considerable variation of scale within clusters. Further, the PCA dimensionality increased with the number of patches (Fig 1e), demonstrating that diversifying the dataset can decorrelate the features.

Individual tuning We computed tuning surfaces for all combinations of embedding features and hypotheses to determine if any features directionally encode any hypotheses (Fig 2a). Indeed, we found a number of such examples, for example feature 47, which responds to water bodies and, specifically, distinguishes between Northern and Southern woodland edges (Fig 2b, c). The selectivity of embedding features to specific hypotheses varied, as defined by their relative tuning strength, and furthermore we observed that off-centre peaks can be used to identify directional coding (Fig 2b, d).

Population coding Finally, we asked how much variance of the embedding features is explained by the land cover and elevation hypotheses. Using OLS regression, we found that the 10 hypotheses explain on average $R^2 = 0.46$ of embedding variance. Next, because combinations of embedding features might represent combinations of hypotheses, we used canonical correlation analysis (CCA) to find the embedding reprojections that maximally correlate to the hypotheses, which explained $R^2 = 0.60$ of variance on average (Fig 3a,b). Both residuals (*i.e.*, the embeddings after subtracting the regression fits) reveal new details, as the correlations with the land cover and elevation hypotheses have effectively been removed (Fig 3a). Accordingly, the effective dimensionality of residuals is greater than the original embedding dimensionality (Fig 3c).

4 Discussion

Limitations We aimed to demonstrate methods new to Earth observation data analysis that can be helpful for interpreting Earth embeddings, using a small, proof-of-concept dataset of AlphaEarth embeddings. A larger and more diverse dataset is needed to reach more definitive conclusions about AlphaEarth embedding features, and other Earth embeddings should be considered to achieve more generalised findings. It would be particularly interesting to compare Earth embeddings trained with different model architectures and loss functions, such as Presto [Tseng et al., 2023], SatCLIP [Klemmer et al., 2025] and spherical harmonics [Rußwurm et al., 2024].

Conclusions Earth embeddings are highly compressed, yet for local applications (that do not encompass the global variety of land surfaces), these embeddings might still suffer from redundancies. We show that for a diverse dataset of locations in the United Kingdom, the effective dimensionality converges to < 10, suggesting that a more efficient embedding representation may be derived for *any* localised application. We further show that regressing out the projections maximally correlated to land cover and elevation increases the dimensionality of the residual embedding features, revealing new spatial variability that used to be obscured by large variance modes. Together, these findings suggest that Earth embeddings can be fine-tuned for local applications by reprojecting, selecting and assigning semantic meaning to embedding features, originally designed to describe the entire globe.

Acknowledgments and Disclosure of Funding

TLvdP and INA acknowledge funding from the Dutch Research Council (NWO) Large-Scale Research Infrastructures (LSRI) programme for the LTER-LIFE (`http://www.lter-life.nl`) infrastructure (grant 184.036.014). JJWB acknowledges funding from the Royal Society through a Royal Society Newton International Fellowship (NIF\R1\252778). The authors thank Marc Rußwurm for constructive feedback and suggestions.

References

Larry F Abbott, Kanaka Rajan, and Haim Sompolinsky. Interactions between intrinsic and stimulus-evoked activity in recurrent neural networks. In M Ding and Dennis L Glanzman, editors, *The dynamic brain: an exploration of neuronal variability and its functional significance*, chapter 4, pages 65–82. Oxford university press, 2011.

Christopher F Brown, Steven P Brumby, Brookie Guzder-Williams, Tanya Birch, Samantha Brooks Hyde, Joseph Mazzariello, Wanda Czerwinski, Valerie J Pasquarella, Robert Haertel, Simon Ilyushchenko, et al. Dynamic world, near real-time global 10 m land use land cover mapping. *Scientific data*, 9(1):251, 2022.

Christopher F Brown, Michal R Kazmierski, Valerie J Pasquarella, William J Rucklidge, Masha Samsikova, Chenhui Zhang, Evan Shelhamer, Estefania Lahera, Olivia Wiles, Simon Ilyushchenko, et al. Alphaearth foundations: An embedding field model for accurate and efficient global mapping from sparse label data. *arXiv preprint arXiv:2507.22291*, 2025.

ESA. Copernicus dem glo-30: Global 30m digital elevation model. `https://doi.org/10.5270/ESA-c5d3d65`, 2019.

Matthew Farrell, Stefano Recanatesi, Timothy Moore, Guillaume Lajoie, and Eric Shea-Brown. Gradient-based learning drives robust representations in recurrent neural networks by balancing compression and expansion. *Nature Machine Intelligence*, 4(6):564–573, 2022.

Juan A Gallego, Matthew G Perich, Raeed H Chowdhury, Sara A Solla, and Lee E Miller. Long-term stability of cortical population dynamics underlying consistent behavior. *Nature neuroscience*, 23 (2):260–270, 2020.

Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. Google earth engine: Planetary-scale geospatial analysis for everyone. *Remote sensing of Environment*, 202:18–27, 2017.

David H Hubel and Torsten N Wiesel. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, 160(1):106–154, 1962.

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

Nikolaus Kriegeskorte, Marieke Mur, and Peter A Bandettini. Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 2:249, 2008.

M Mälicke. Scikit-gstat 1.0: a scipy-flavored geostatistical variogram estimation toolbox written in python, 2022.

Valerio Mante, David Sussillo, Krishna V Shenoy, and William T Newsome. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, 503(7474):78–84, 2013.

John O'Keefe and Lynn Nadel. *The hippocampus as a cognitive map*. Oxford university press, 1978.

Mattia Rigotti, Omri Barak, Melissa R Warden, Xiao-Jing Wang, Nathaniel D Daw, Earl K Miller, and Stefano Fusi. The importance of mixed selectivity in complex cognitive tasks. *Nature*, 497 (7451):585–590, 2013.

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

Odelia Schwartz, Jonathan W Pillow, Nicole C Rust, and Eero P Simoncelli. Spike-triggered neural characterization. *Journal of vision*, 6(4):13–13, 2006.

Gabriel Tseng, Ruben Cartuyvels, Ivan Zvonkov, Mirali Purohit, David Rolnick, and Hannah Kerner. Lightweight, pre-trained transformers for remote sensing timeseries. *arXiv preprint arXiv:2304.14065*, 2023.

Thijs L Van der Plas, Stephen Law, and Michael JO Pocock. Predicting butterfly species presence from satellite imagery using soft contrastive regularisation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 2174–2183, June 2025. doi: https://doi.org/10.48550/arXiv.2505.09306.

Hao-Ting Wang, Jonathan Smallwood, Janaina Mourao-Miranda, Cedric Huchuan Xia, Theodore D Satterthwaite, Danielle S Bassett, and Danilo Bzdok. Finding the needle in high-dimensional haystack: A tutorial on canonical correlation analysis. *arXiv preprint arXiv:1812.02598*, 2018.