CLIMPLICIT: CLIMATIC IMPLICIT EMBEDDINGS FOR GLOBAL ECOLOGICAL TASKS

Johannes Dollinger[1], **Damien Robert**[1], **Elena Plekhanova**[2], **Lukas Drees**[1], **Jan Dirk Wegner**[1]

[1]EcoVision Lab, DM3L, University of Zurich
[2]Swiss Federal Research Institute WSL

ABSTRACT

Deep learning on climatic data holds potential for macroecological applications. However, its adoption remains limited among scientists outside the deep learning community due to storage, compute, and technical expertise barriers. To address this, we introduce Climplicit, a spatio-temporal geolocation encoder pretrained to generate implicit climatic representations anywhere on Earth. By bypassing the need to download raw climatic rasters and train feature extractors, our model uses $\times 3500$ less disk space and significantly reduces computational needs for downstream tasks. We evaluate our CLIMPLICIT embeddings on biomes classification, species distribution modeling, and plant trait regression. We find that single-layer probing our CLIMPLICIT embeddings consistently performs better or on par with training a model from scratch on downstream tasks and overall better than alternative geolocation encoding models.

1 INTRODUCTION

Ecology studies how species interact with their environment, and climate is a key factor in shaping habitats and distributions (Karger et al., 2017). Recently, deep learning has been successfully applied to the modeling of species distribution from citizen science data (Cole et al., 2023; Botella et al., 2023; Brun et al., 2024; Dollinger et al., 2024). These models rely on climatic variables, which are derived from vast amounts of raw weather data collected by global stations and aggregated into climatic reanalysis datasets. Although deep learning is well suited to leverage such data, its adoption in ecology remains limited due to high computational, storage, and expertise requirements.

To this end, several works focus on pretraining lightweight, off-the-shelf geolocation encoders for downstream tasks (Rolf et al., 2021; Mai et al., 2023a; Klemmer et al., 2023; Vivanco Cepeda et al., 2024; Agarwal et al., 2024; Sastry et al., 2025). However, existing models often suffer from low resolution or produce inconsistent embeddings (see Appendix F.5, Figure 5), limiting their effectiveness. To address these shortcomings, we introduce CLIMPLICIT, a *spatio-temporal* location encoder that generates implicit climatic representations for any land location and month. By pretraining on CHELSA climatologies (Karger et al., 2017; 2018), our model captures fine-grained environmental patterns with greater consistency and resolution. Our CLIMPLICIT backbone is based on residual SIREN (RESIREN) blocks, a novel extension of SIREN (Sitzmann et al., 2020) with residual connections that improve scalability. Although previous works have explored compressing climatic data into neural networks (Huang and Hoefler, 2022; Naour et al., 2023; Liu et al., 2023; Xu et al., 2024), our approach specifically focuses on a high-resolution, lightweight location encoder. We validate CLIMPLICIT embeddings on biodiversity-related tasks and provide ablation studies for our key design choices. Code is available at `https://github.com/ecovision-uzh/climplicit`.

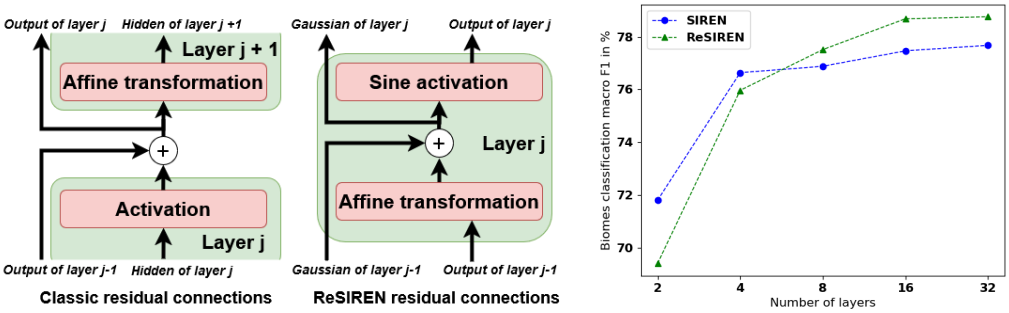


Figure 1: **Left:** Contrary to "classic" residual blocks, our ReSIREN residual connection is placed before the non-linearity rather than after. **Right:** Scaling comparison of SIREN and ReSIREN on the biome classification task (Section D). ReSIREN underperforms at small depths but scales better. See Appendix F.4 for further details.

2 Methodology

Let $i \in \mathbb{N}$ be an index over discretized month-locations on Earth. We define a location encoder as a neural network **LE** that transforms the positional encoding p_i of a month-coordinate triplet (λ_i, ϕ_i, m_i) with longitude $\lambda_i \in [-1, 1]$, latitude $\phi_i \in [-1, 1]$ and month $m_i \in \{1, ..., 12\}$ into an embedding $e_i = \mathbf{LE}(p_i)$.

2.1 Direct Positional Embedding with H-SIREN

Our positional encoding extends with a temporal dimension the Direct embedding defined by Rußwurm et al. (2024):

$$p_i = \Big[\lambda_i,\ \phi_i,\ \sin(2\pi * m_i/12),\ \cos(2\pi * m_i/12) \Big] \ \in\ [-1, 1]^4 \ .$$

We choose the Direct embedding because of its low computational cost and sufficient representational power in conjunction with the SIREN network (Rußwurm et al., 2024). It is also free of hyperparameters that predefine frequencies, unlike Fourier Features (Tancik et al., 2020), Spherical Harmonics (Rußwurm et al., 2024) and Sphere2Vec (Mai et al., 2023b). The sine and cosine transformations of the months ensure that the values are in $[-1, 1]$ and that January and December have close representations. To increase the frequencies accessible to the network, we use the H-SIREN (Gao and Jaiman, 2024) activation function in the first layer (see Appendix B.1).

2.2 Residual Connection for SIREN

Residual connections have become a staple of neural networks since they enabled the first deep architectures (He et al., 2016a). We introduce **ReSIREN**, a residual extension of SIREN (Sitzmann et al., 2020) that improves scalability and enhances information exchange between layers (He et al., 2016b). Let L_j, where $j \in \{1, ..., D\}$, be the j-th layer in a network of depth D, consisting of an affine transformation F_j and activation function A_j. Given input z_j, we define the hidden pre-activation vector as $h_j = F_j(z_j)$ and the output as $z_{j+1} = L_j(z_j)$. We introduce residual connections between the hidden vectors:

$$h'_j = \frac{h_j + h'_{j-1}}{2}, \quad z_{j+1} = A_j(h'_j).$$

This preserves the normal distribution $\mathcal{N}(0, 1)$ of h_j required for the stability of SIREN (Sitzmann et al., 2020). Figure 1 illustrates ReSIREN and its improved scaling properties compared to SIREN. See Appendix B.2 and F.4 for more details.

2.3 Climatic Pretraining

We train a ReSIREN model on the CHELSA climate dataset (Karger et al., 2017), a downscaling of ERA5 (Dee et al., 2011) data using a high-resolution digital elevation model. We use the monthly climatology from 1981-2010, containing 11 climatic variables related to humidity, wind speed, precipitation, temperature and pressure[1] that were selected in consultation with ecologists. A climatology refers to the per-month aggregate over 30 years. For each of the 12 months, we use a global raster at a horizontal resolution of 30 arc sec for each of the 11 variables. We select only land surface pixels and arrive at a total number of 224 million pixels, resulting in a file size of $12 * 11 * 0.418$ GB $= 55.2$ GB in float16. We pretrain our ReSIREN backbone followed by an affine projection to regress the Gaussian-normalized CHELSA values using the mean squared error loss. The resulting Climplicit model has a size of 16 MB: 99.97% smaller than the original CHELSA dataset is was pretrained on. Interestingly, the model could be further improved using neural compression (Neill, 2020; Gomes et al., 2025). See Appendix C for more details.

3 Experiments

We assess **Climplicit** embeddings with single-layer probing on three macroecological downstream tasks: **Biomes** classification, species distribution modeling (**SDM**) and **Plant traits** regression (see Appendix D for more details). To illustrate the storage- and compute-intensive nature of training that Climplicit aims to mitigate, we train a model from scratch (**FS**) with a location encoder (**Loc**) or a CHELSA encoder (**CH**) for each downstream task. For Loc, we use a 16-layer ReSIREN with the same hyperparameters as Climplicit, but with $[\lambda, \phi]$ as input. For CH, we incorporate CHELSA values from March, June, September, and December, processing them with a 4-layer residual network following Cole et al. (2023). For Loc + CH, we concatenate both embeddings before the probe. Additionally, we compare Climplicit embeddings with those from **SatCLIP** (Klemmer et al., 2023), **Taxabind** (Sastry et al., 2025), **SINR** (Cole et al., 2023), **CSP** (Mai et al., 2023a), and **GeoCLIP** (Vivanco Cepeda et al., 2024), which implicitly represent modalities such as satellite imagery, sound, bioclimatic variables, and species occurrence data.

4 Results

Table 1: Single-layer probing on diverse geolocation representations. Mean and standard deviation of ten random initializations of the single-layer probe. **Best**. <u>Second best</u>.

Model	Biomes (% F1 ↑)	SDM (% Acc ↑)	Plant traits (% R^2 ↑)
FS Loc	73.9 ± 2.4	2.0 ± 0.4	42.2 ± 0.0
FS CH	71.8 ± 1.9	2.5 ± 0.1	60.0 ± 0.3
FS Loc + CH	**79.6** ± 1.7	2.5 ± 0.1	<u>64.8</u> ± 0.4
SatCLIP	68.3 ± 0.4	1.3 ± 0.1	61.6 ± 0.1
Taxabind	59.3 ± 0.1	3.1 ± 0.0	56.9 ± 0.0
SINR	63.1 ± 0.3	1.7 ± 0.0	63.5 ± 0.1
CSP	58.6 ± 0.4	1.6 ± 0.1	49.7 ± 0.3
GeoCLIP	62.7 ± 0.1	**3.5** ± 0.0	57.9 ± 0.1
Climplicit (Ours)	<u>78.4</u> ± 0.3	<u>3.2</u> ± 0.0	**70.0** ± 0.1

We summarize our results in Table 1.

Climplicit representations surpass all FS ones, except for FS Loc + CH on Biomes, which slightly outperforms Climplicit but with a larger confidence interval, highlighting the high variance of representations learned from scratch on small datasets compared to pretrained representations. These results validate the potential of Climplicit for scenarios

[1]For more details, see section **7.2 Monthly** of `https://chelsa-climate.org/downloads`

with minimal storage, compute, or deep learning knowledge. Furthermore, we observe that other geolocation encoders fall short of CLIMPLICIT on all tasks, with the exception of GEOCLIP on SDM, which we attribute to the high resolution of GEOCLIP embeddings. This highlights the relevance of high-resolution climatic representations for the downstream tasks at hand. Note in Appendix F.2 Table 4 that using MLP probing lets CLIMPLICIT outperform the other methods on all three tasks. Further quantitative and qualitative results can be found in Appendix F.1, F.4, and F.5.

5 ABLATION STUDIES

We analyze the impact of key design choices in Table 2. **SIREN** removes the residual connections from our RESIREN blocks, consistently resulting in a performance drop, with -1.2 R^2 for plant traits regression. Appendix F.4 further highlights the impact of residual connections with increasing model depth. To explore the impact of learning temporal climatic data representations, **ALL- MONTHS** is pretrained to reconstruct the concatenation of all monthly CHELSA variables at each location, while **MARCH-ONLY** is pretrained exclusively on March data ($m_i = 3$). Our results support support the choice of making CLIMPLICIT a spatio-temporal location encoder rather than encoding only a single month or the entirety of yearly climatic variables. In **NO H-SIREN** we remove the H-SIREN activation function and observe a performance drop on two of our tasks. **REC-CHELSA** runs single-layer probing on the reconstructed CHELSA variables instead of the embeddings before the projection layer. As expected, the latent representation better generalizes compared to reconstructed climatic variables. In **CHELSA-CLIP**, we pretrain our model with a contrastive objective similar to Klemmer et al. (2023) which proves to produce poorer representations compared to directly regressing CHELSA. Whether this approach extends to more complicated regressions or auto-encoding settings for multi-pixel images is left to future work. Finally, **ERA5** compares representations learned from monthly **ERA5** (Dee et al., 2011) instead of CHELSA. We find that the $\times 30$ higher resolution of CHELSA (30 versus 900 arcsec) allows for learning richer representations for our marcoecological tasks.

Table 2: Ablation analysis with single-layer probing. Mean and standard deviation of ten random initializations of the single-layer probe. **Best**.

Model	Biomes (% F1 ↑)	SDM (% Acc ↑)	Plant traits (% R^2 ↑)
CLIMPLICIT	**78.4** ± 0.3	3.2 ± 0.0	**70.0** ± 0.1
SIREN	77.5 ± 0.2	3.1 ± 0.0	68.8 ± 0.2
CONCAT MONTHS	75.9 ± 0.3	2.6 ± 0.0	66.0 ± 0.1
MARCH-ONLY	78.2 ± 0.2	2.9 ± 0.0	62.8 ± 0.1
NO H-SIREN	77.9 ± 0.2	**3.6** ± 0.0	69.1 ± 0.1
REC-CHELSA	61.5 ± 0.2	1.5 ± 0.0	55.4 ± 0.1
CH-CLIP	76.5 ± 0.6	2.3 ± 0.1	66.9 ± 0.4
ERA5	63.7 ± 0.5	1.9 ± 0.1	68.6 ± 0.2

6 DISCUSSION AND CONCLUSION

We introduced CLIMPLICIT, a spatiotemporal encoder that produces implicit climatic representations anywhere on Earth, aimed at lowering the storage, compute, and machine learning expertise requirements for downstream practitioners. Our model hinges on RESIREN, a residual extension of SIREN that allows to learn high-resolution representations. Our CLIMPLICIT model occupies 99.9% less memory than the CHELSA dataset it was trained on, while producing representations that perform on par or better than from-scratch training on several macroecological tasks. Potential limitations of CLIMPLICIT embeddings are their *implicit* nature and lower resolution, compared to the original climatic raster dataset. Finally, in Appendix E, we outline how CLIMPLICIT could be extended to provide embeddings for future climatologies under different climate change scenarios, in turn allowing future predictions for the downstream tasks.

7 Acknowledgments

This project was partially funded by the Embed2Scale project, co-funded by the EU Horizon Europe Programme under Grant Agreement No. 101131841. Additional support was provided by the Swiss State Secretariat for Education, Research and Innovation (SERI) and UK Research and Innovation (UKRI).

References

M. Agarwal, M. Sun, C. Kamath, A. Muslim, P. Sarker, J. Paul, H. Yee, M. Sieniek, K. Jablonski, Y. Mayer, et al. General geospatial inference with a population dynamics foundation model. *arXiv preprint arXiv:2411.07207*, 2024.

C. Botella, B. Deneu, D. Marcos, M. Servajean, J. Estopinan, T. Larcher, C. Leblanc, P. Bonnet, and A. Joly. The geolifeclef 2023 dataset to evaluate plant species distribution models at high spatial resolution across europe. *arXiv preprint arXiv:2308.05121*, 2023.

P. Brun, D. N. Karger, D. Zurell, P. Descombes, L. C. de Witte, R. de Lutio, J. D. Wegner, and N. E. Zimmermann. Multispecies deep learning using citizen science data produces more informative plant community models. *Nature Communications*, 2024.

E. Cole, G. Van Horn, C. Lange, A. Shepard, P. Leary, P. Perona, S. Loarie, and O. Mac Aodha. Spatial implicit neural representations for global-scale species mapping. *ICML*, 2023.

D. P. Dee, S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. Balmaseda, G. Balsamo, d. P. Bauer, et al. The era-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society*, 2011.

J. Dollinger, P. Brun, V. Sainte Fare Garnot, and J. D. Wegner. Sat-sinr: High-resolution species distribution models through satellite imagery. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2024.

E. Doxsey-Whitfield, K. MacManus, S. B. Adamo, L. Pistolesi, J. Squires, O. Borkovska, and S. R. Baptista. Taking advantage of the improved availability of census data: a first look at the gridded population of the world, version 4. *Papers in Applied Geography*, 2015.

R. Gao and R. K. Jaiman. H-siren: Improving implicit neural representations with hyperbolic periodic functions. *arXiv preprint arXiv:2410.04716*, 2024.

M. J. Gidden, K. Riahi, S. J. Smith, S. Fujimori, G. Luderer, E. Kriegler, D. P. Van Vuuren, M. Van Den Berg, L. Feng, D. Klein, et al. Global emissions pathways under different socioeconomic scenarios for use in cmip6: a dataset of harmonized emissions trajectories through the end of the century. *Geoscientific Model Development*, 2019.

C. Gomes, I. Wittmann, D. Robert, J. Jakubik, T. Reichelt, S. Maurogiovanni, R. Vinge, J. Hurst, E. Scheurer, R. Sedona, et al. Lossy neural compression for geospatial analytics: A review. *IEEE Geoscience and Remote Sensing Magazine*, 2025.

K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. *CVPR*, 2016a.

K. He, X. Zhang, S. Ren, and J. Sun. Identity mappings in deep residual networks. *ECCV*, 2016b.

L. Huang and T. Hoefler. Compressing multidimensional weather and climate data into neural networks. *arXiv preprint arXiv:2210.12538*, 2022.

D. Karger, O. Conrad, J. Böhner, T. Kawohl, H. Kreft, R. Soria-Auza, N. Zimmermann, H. Linder, and M. Kessler. Climatologies at high resolution for the earth's land surface areas. envidat, 2018.

D. N. Karger, O. Conrad, J. Böhner, T. Kawohl, H. Kreft, R. W. Soria-Auza, N. E. Zimmermann, H. P. Linder, and M. Kessler. Climatologies at high resolution for the earth's land surface areas. *Scientific data*, 2017.

D. N. Karger, D. R. Schmatz, G. Dettling, and N. E. Zimmermann. High-resolution monthly precipitation and temperature time series from 2006 to 2100. *Scientific Data*, 2020.

K. Klemmer, E. Rolf, C. Robinson, L. Mackey, and M. Rußwurm. Satclip: Global, general-purpose location embeddings with satellite imagery. *arXiv preprint arXiv:2311.17179*, 2023.

J. Liu, S. Di, S. Jin, K. Zhao, X. Liang, Z. Chen, and F. Cappello. Scientific error-bounded lossy compression with super-resolution neural networks. *International Conference on Big Data*, 2023.

G. Mai, N. Lao, Y. He, J. Song, and S. Ermon. Csp: Self-supervised contrastive spatial pre-training for geospatial-visual representations. *ICML*, 2023a.

G. Mai, Y. Xuan, W. Zuo, Y. He, J. Song, S. Ermon, K. Janowicz, and N. Lao. Sphere2vec: A general-purpose location representation learning over a spherical surface for large-scale geospatial predictions. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2023b.

Á. Moreno-Martínez, G. Camps-Valls, J. Kattge, N. Robinson, M. Reichstein, P. van Bodegom, K. Kramer, J. H. C. Cornelissen, P. Reich, M. Bahn, et al. A methodology to derive global maps of leaf traits using remote sensing and climate data. *Remote Sensing of Environment*, 2018.

E. L. Naour, L. Serrano, L. Migus, Y. Yin, G. Agoua, N. Baskiotis, P. Gallinari, and V. Guigue. Time series continuous modeling for imputation and forecasting with implicit neural representations. *arXiv preprint arXiv:2306.05880*, 2023.

J. O. Neill. An overview of neural network compression. *arXiv preprint arXiv:2006.03669*, 2020.

D. M. Olson, E. Dinerstein, E. D. Wikramanayake, N. D. Burgess, G. V. Powell, E. C. Underwood, J. A. D'amico, I. Itoua, H. E. Strand, J. C. Morrison, et al. Terrestrial ecoregions of the world: A new map of life on earth: A new global map of terrestrial ecoregions provides an innovative tool for conserving biodiversity. *BioScience*, 2001.

R. K. Pace and R. Barry. Sparse spatial autoregressions. *Statistics & Probability Letters*, 1997.

E. Rolf, J. Proctor, T. Carleton, I. Bolliger, V. Shankar, M. Ishihara, B. Recht, and S. Hsiang. A generalizable and accessible approach to machine learning with global satellite imagery. *Nature Communications*, 2021.

M. Rußwurm, K. Klemmer, E. Rolf, R. Zbinden, and D. Tuia. Geographic location encoding with spherical harmonics and sinusoidal representation networks. *ICLR*, 2024.

S. Sastry, S. Khanal, A. Dhakal, A. Ahmad, and N. Jacobs. Taxabind: A unified embedding space for ecological applications. *WACV*, 2025.

V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein. Implicit neural representations with periodic activation functions. *NeurIPS*, 2020.

M. Tancik, P. Srinivasan, B. Mildenhall, S. Fridovich-Keil, N. Raghavan, U. Singhal, R. Ramamoorthi, J. Barron, and R. Ng. Fourier features let networks learn high frequency functions in low dimensional domains. *NeurIPS*, 2020.

V. Vivanco Cepeda, G. K. Nayak, and M. Shah. Geoclip: Clip-inspired alignment between locations and images for effective worldwide geo-localization. *NeurIPS*, 2024.

Z. Xu, B. Pan, H. Li, and X. Wei. Hiha: Introducing hierarchical harmonic decomposition to implicit neural compression for atmospheric data. *arXiv preprint arXiv:2411.06155*, 2024.

A Impact and Ethical Statement

This work is part of an ongoing project co-designed with ecologists to make deep learning methods more easily accessible to practitioners. Future work will apply the model to more extensive ecological problems. After finishing pretraining, this model has small requirements in regards to compute, storage capacity and know-how, which increases ease of use for non-deep learning experts and reduces the carbon footprint of downstream applications. No GenAI such as ChatGPT was used in the making of this work and writing of this publication.

B ReSIREN Details

B.1 ReSIREN Activations

Let $L_j, j \in \{1, ..., D\}$ be the j-th layer of the network of depth D, consisting of an affine transformation F_j and activation function A_j. The activation functions across the network then are:

$$A_j(x) = \begin{cases} \sin(\omega_0 \sinh(2x)) & \text{for j} = 1 \\ \sin(\omega_0 x) & \text{for } 1 < \text{j} < \text{D} \\ \text{Identity}(x), & \text{for j} = \text{D} \end{cases}$$

Following the results of Sitzmann et al. (2020), we set $\omega_0 = 30$. According to the authors, ω_0 increases the input range across multiple periods of sin, which allows the network to converge more quickly and access more frequencies. The special case in the first layer is the H-SIREN function of Gao and Jaiman (2024).

B.2 Distributional Stability of ReSIREN

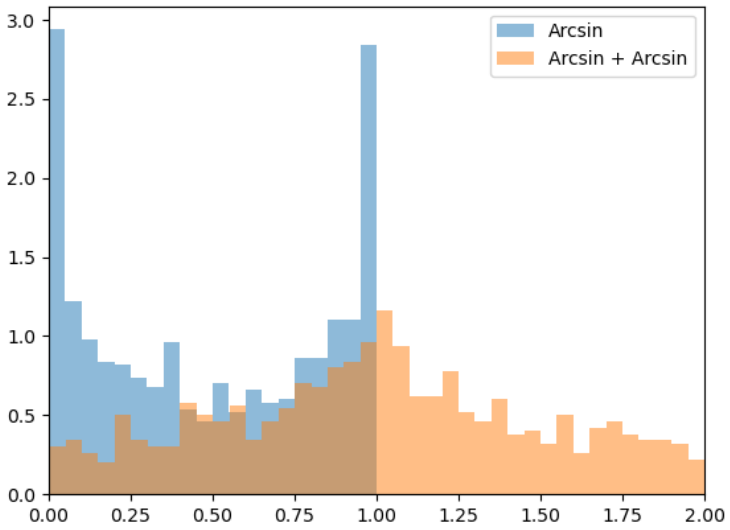


Figure 2: Experimental probability density function of adding up two independent Arcsin$(0, 1)$. The sum is a ill-behaved distribution that can not be trivially normalized back to Arcsin$(0, 1)$. Therefore, we choose to apply residual connections between to the well-behaved Gaussians in the middle of each SIREN layer.

Let z_j be the input to L_j, $h_j = F_j(z_j)$ the hidden pre-activation vector and $z_{j+1} = L_j(z_j)$ the output of the layer. According to Sitzmann et al. (2020), the distribution of z_j is Arcsin$(-1, 1)$ and $\mathcal{N}(0, 1)$ for h_j. These distributional properties are essential to the functioning of the network and are enabled by a specific initialization of the weights of F_j.

Adding residual connections the classic way by adding the output of the previously layer to the current layer's output would amount to $z'_{j+1} = z'_j + z_{j+1}$, but the sum of two Arcsin distributions turns into a badly behaved distribution (See Figure 2). Therefore, we introduce residual connections between the hidden vectors h, specifically as $h'_j = (h_j + h'_{j-1})/2$ with $z_{j+1} = A_j(h'_j)$.

Lemma B.1. *Let X,$Y \sim \mathcal{N}(0,1)$ be independent and $Z = (X+Y)/2$. Then $Z \sim \mathcal{N}(0,1)$.*

Proof. The sum of two independent normally distributed variables is equally normally distributed, with the mean and variance added up. This results in $\mathcal{N}(0+0, 1+1)$. This is regularized to $\mathcal{N}(0,1)$ by dividing through the variance 2. □

We assume independence between the two normally distributed vectors h_j and h'_{j-1}. With the residual sum being normally distributed again, we thus provide the requirements for **Lemma 1.6.** of Sitzmann et al. (2020) on keeping the activation distributions stable across the network.

C Training details

For each epoch, we iterate once over all land locations and sample a random month m_i uniformly for each location. We train the network up to 20 epochs with the Adam optimizer and early stopping, totaling up to 547 thousand training steps. We are learning a dense representation across the globe with a large number of datapoints and therefore do not require a val/test split. We use a model depth of 16, hidden size of 512, embeddings size of 256, and batch size of 8192 pixels. Although we found that a higher embedding size may slightly increase performance, we stick to 256 to align with similar geospatial encoding models. Our pretraining converges after $\sim$ 7 hours on a NVIDIA T4 with 16 GB of RAM and 16 CPUs.

D Downstream Tasks

Biomes classification. We reproduce the biomes classification task from (Klemmer et al., 2023) by sampling 100k land locations uniformly across the globe and then extracting the biome at each location from the terrestrial ecoregions of the world map (Olson et al., 2001). The train-val-test split is $[0.5, 0.1, 0.4]$. As metric we use the mean of the F1 scores for all biomes, dubbed "macro" as it gives equal weight to all classes.

Species distribution modeling. In species distribution modeling (SDM), we aim to predict if a species is likely to be observed at a location in a given month. We use the preprocessed (Dollinger et al., 2024) presence-only data from the GLC23 challenge (Botella et al., 2023), encompassing 2 million occurrences for 5.6 thousand plant species across Europe. The train-val-test split is $[0.7, 0.05, 0.25]$. We train the probe with the AN-full loss with background sampling (Cole et al., 2023), and report the Top-1 Accuracy.

Plant traits regression. We sample 100k locations as in the biomes classification and extract means and standard deviations for 4 plant functional traits from the global plant trait maps provided by Moreno-Martínez et al. (2018). The traits are: specific leaf area, leaf dry matter content, leaf nitrogen and phosphorus content per dry mass, and leaf nitrogen/phosphorus ratio. We split the data $[0.5, 0.1, 0.4]$ and report the mean $\mathbf{R^2}$ score across the eight values.

SDM has a date of observation associated with each occurrence, thus we sample Climplicit for the embedding of the respective month. In the case of biomes and plant traits, the rasters we sample from are immutable across the year. Here, we retrieve the embeddings for March, June, September and December from Climplicit and concatenate them as input to the probe for the classification and regression tasks.

E Future Work: Extension to Future Climatologies

CHELSA has been extended to include four monthly precipitation and temperature rasters for the climatologies 2011-2040, 2041-2070 and 2071-2100 (Karger et al., 2020). These are calculated from four different global circulation models for the three shared socioeconomic pathways SSP126, SSP370 and SSP585 (Gidden et al., 2019).

A future extension of Climplicit may include these future climatologies by first declaring a categorical variable $c_i \in \{-1, -0.33, 0.33, 1\}$ where each value corresponds to one of the four thirty-year intervals. This variable may be added to the positional encoding as another temporal variable:

$$p_i = \Big[\lambda_i,\ \phi_i,\ \sin(2\pi * m_i/12),\ \cos(2\pi * m_i/12)\ , c_i \Big] \ \in \ [-1, 1]^5 \ .$$

This may enable the location encoder to encode all four climatologies in a shared embedding space. It is important to note that for this method, not all eleven CHELSA variables from the contemporary climatology can be used during training, but only the four also provided for the future climatologies. This, along with the increased learning requirements from three additional climatologies might decrease performance on the downstream task setup of this work.

Having now learned a shared embedding space for the four climatologies, a downstream task may now train on the contemporary climatology $c_i = -1$, and make predictions on the future climatologies $c_i \in \{-0.33, 0.33, 1\}$. This can be seen as an downstream extrapolation to an un-sampled area, which the embeddings produced by location encoders are especially well suited to (Klemmer et al., 2023).

F Extended Results

Table 3: Single-layer probing on diverse geolocation representations for non-ecological downstream tasks. Mean and standard deviation of ten random initializations of the single-layer probe. **Best**. Best pretrained geolocation representation.

Model	Pop Density (% R^2 ↑)	Med Income (% R^2 ↑)	Cali Housing (% R^2 ↑)
FS Loc	**75.3** ± 5.1	-7.5 ± 17.2	18.9 ± 18.9
FS CH	64.9 ± 2.1	29.8 ± 2.9	**61.9** ± 2.2
FS Loc + CH	75.1 ± 3.4	23.3 ± 13.5	60.3 ± 3.5
SatCLIP	49.1 ± 0.5	30.0 ± 9.1	35.2 ± 0.4
Taxabind	41.0 ± 0.2	15.0 ± 0.1	44.2 ± 0.1
SINR	38.8 ± 0.3	38.6 ± 0.5	37.1 ± 0.3
CSP	35.8 ± 0.7	27.3 ± 1.5	50.1 ± 0.6
GeoCLIP	39.3 ± 0.1	34.7 ± 0.1	58.9 ± 0.1
Climplicit	67.0 ± 0.3	**45.0** ± 0.4	44.2 ± 0.6

F.1 Non-Ecological Downstream Tasks

We furthermore report in Table 3 the results on three non-ecological downstream tasks adapted from Klemmer et al. (2023). We sample 100k **population density** values in the United States of America from Doxsey-Whitfield et al. (2015). For **median income** regression, we use the 2021 US county-level median household income data from the US Department of Agriculture[2]. **California housing** is a long-established dataset from Pace and Barry (1997) of median house prices in California census blocks. On all tasks we report the R^2 score.

[2]https://www.ers.usda.gov/data-products/county-level-data-sets/county-level-data-sets-download-data

CLIMPLICIT is the best location encoder on two of the three tasks, also managing to outperform the fully trained models on the challenging median income task. Only on the California housing dataset does CLIMPLICIT fall short, which may be attributed to the high variation in housing prices in California independent of the local variation in climate.

F.2 MLP RESULTS

Table 4: MLP probing on diverse geolocation representations. Mean and standard deviation of ten random initializations of the linear probe. **Best**. Second best.

Model	Biomes (% F1 ↑)	SDM (% Acc ↑)	Plant traits (% R^2 ↑)
FS LOC	73.1 ± 3.2	1.64 ± 0.60	70.6 ± 0.9
FS CH	66.6 ± 1.8	3.16 ± 0.08	74.4 ± 1.2
FS LOC + CH	75.7 ± 3.1	3.17 ± 0.11	76.4 ± 0.9
SATCLIP	72.3 ± 0.7	2.80 ± 0.07	76.9 ± 0.3
TAXABIND	63.6 ± 1.1	3.04 ± 0.07	71.4 ± 0.3
SINR	64.4 ± 0.8	2.68 ± 0.07	72.8 ± 0.7
CSP	64.2 ± 1.3	3.06 ± 0.05	72.0 ± 0.4
GEOCLIP	64.9 ± 0.6	3.20 ± 0.06	71.6 ± 0.3
CLIMPLICIT (Ours)	**78.2** ± 0.7	**3.70** ± 0.04	**78.6** ± 0.3

In Table 4, instead of using a single layer as probe, we also calculate results when training a multi-layer perceptron consisting of three hidden layers with 64 neurons on the various embeddings. This is similar to the downstream setup of Klemmer et al. (2023), although the shape of this MLP has not been fine-tuned. We can see that CLIMPLICIT outperforms all comparison methods in this setup.

F.3 CHELSA RECONSTRUCTION

We samples 100k locations across earth and calculate the absolute error between the original CHELSA values and the reconstruction by CLIMPLICIT for each location and month. Note that we calculate this in the Gaussian-normalized space to keep values comparable between the different variables. We also apply log to make the visualizations less sensitive to extreme values.

In Figure 3 we observe that CLIMPLICIT reconstructs temperature and radiation the best, but struggles significantly with wind speeds. There is no discernible variation across months. Spatially, the error is concentrated in mountainous areas. As CHELSA is a down sampling of ERA5 data based on a digital elevation model, it follows that the local variation in CHELSA is strongly correlated with elevation changes. Thus the smooth reconstruction by CLIMPLICIT performs well in flat areas where CHELSA is equally smooth, but struggles in areas with significant elevation changes. Future work can tackle this issue by oversampling areas with high elevation variability to focus the models capacity there.

F.4 SCALING OF RESIREN

In Figure 4 we report both the macro and weighted F1 score, along with the European SDM accuracy. Location and global CHELSA reconstruction are regression tasks where location have been sampled across the globe similar to the biomes task. For a location, either the location itself or the respective CHELSA vector was then regressed. Lastly, we present the increase in runtime from using the residual connections.

F.5 QUALITATIVE RESULTS

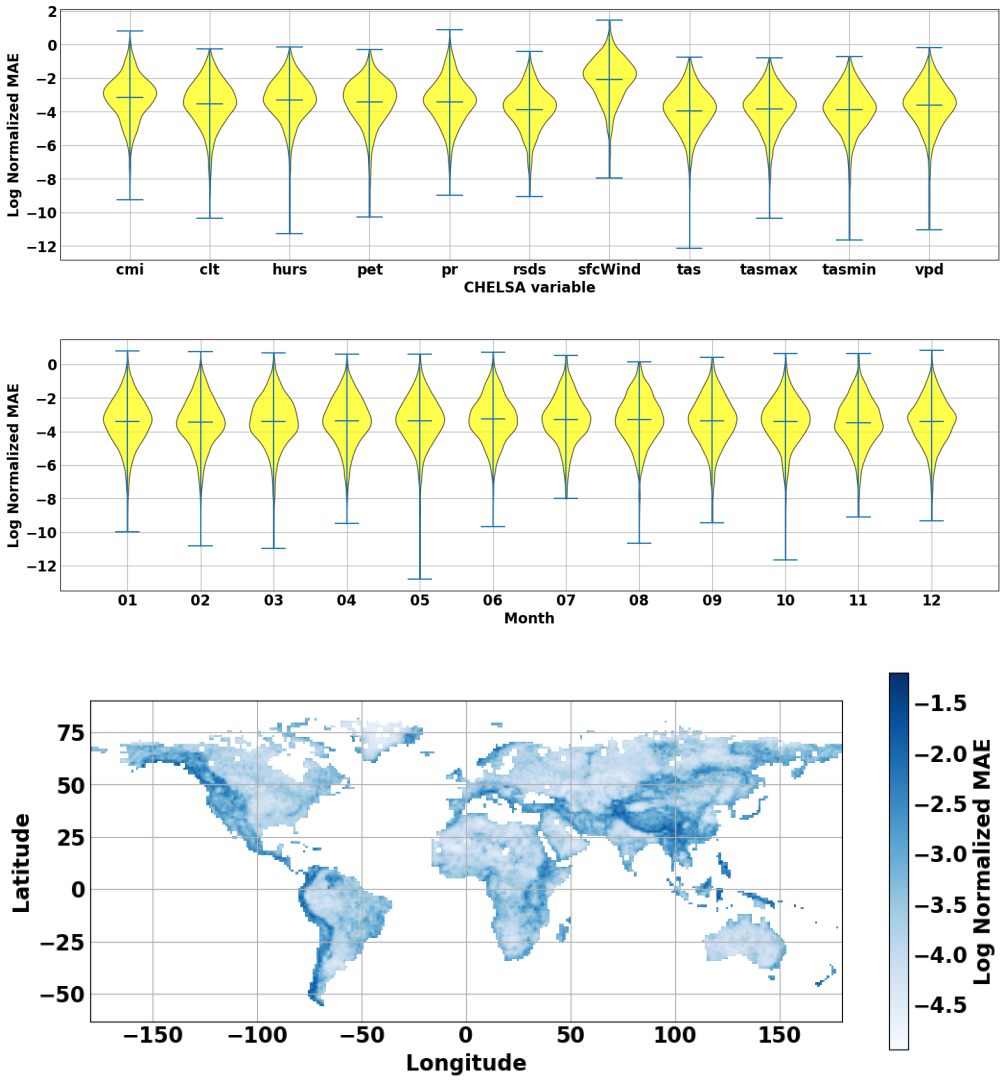


Figure 3: Analysis of the reconstruction error of CHELSA after the CLIMPLICIT pretraining. **Top**: Absolute error distribution for each climatic variable. **Middle**: Absolute error distribution for each month. **Bottom**: Mean absolute error inside 136 by 320 grid cells across the globe.

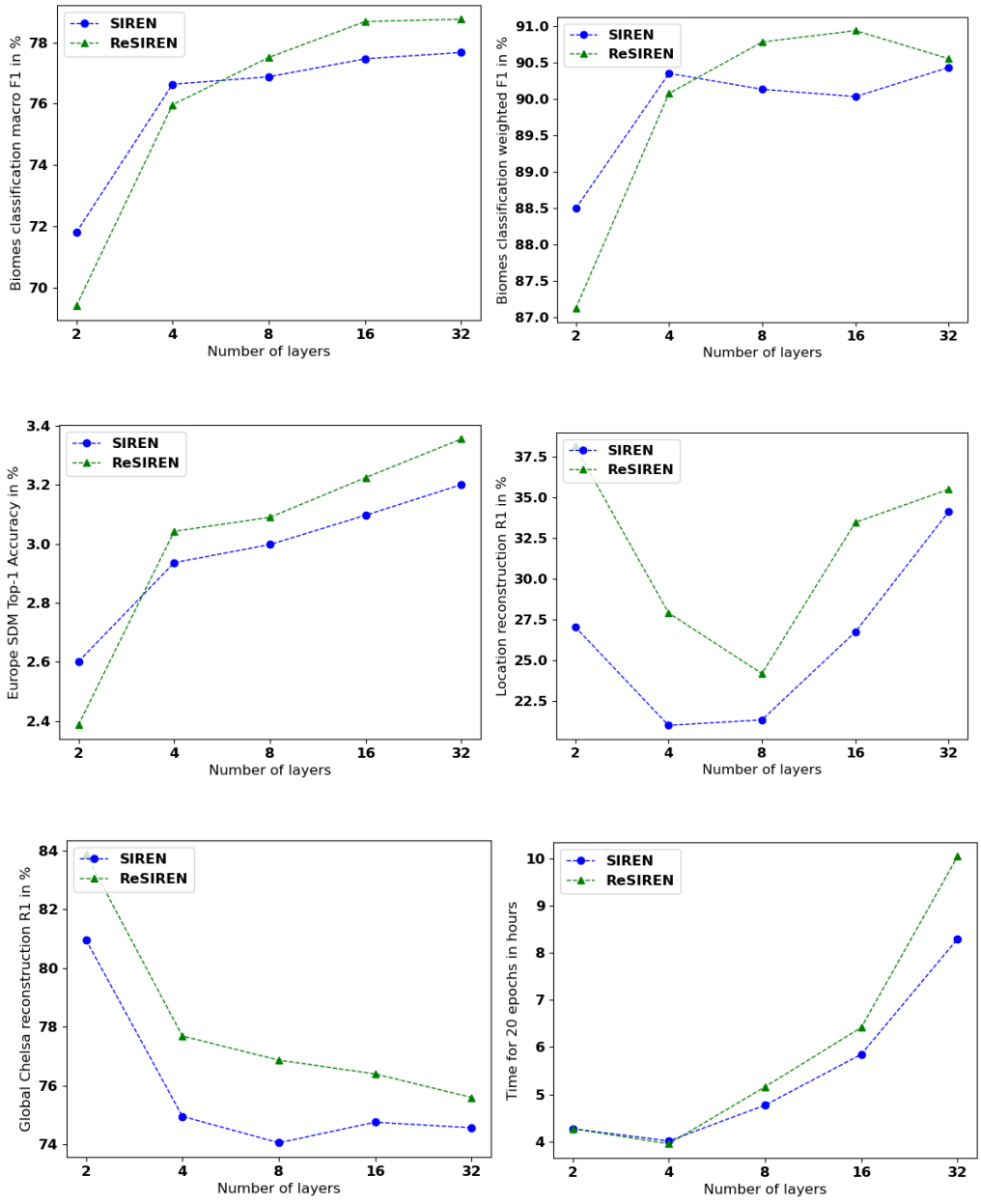
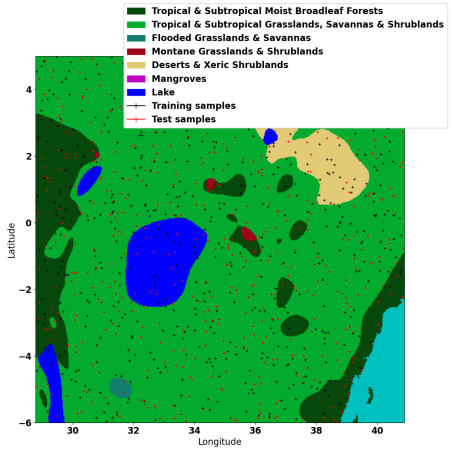
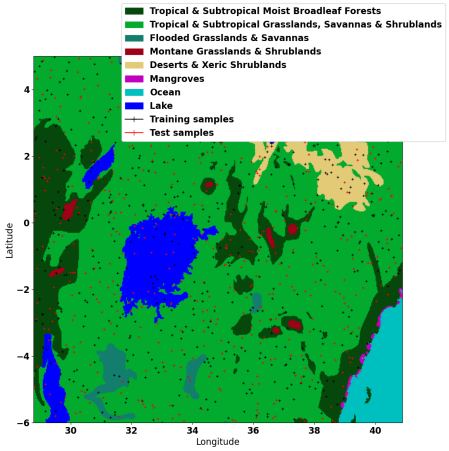


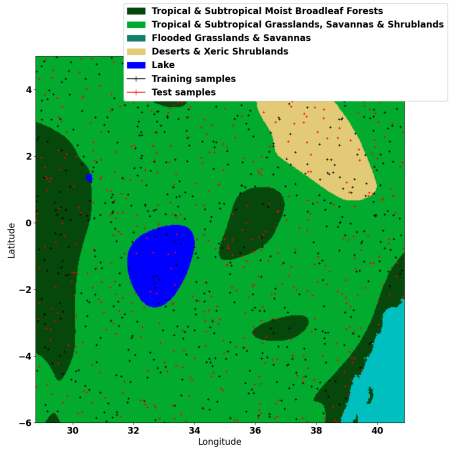
Figure 4: Comparison of the scaling behavior of ReSIREN compared to SIREN. Generally, ReSIREN performs worse at few layers, but outperforms and scales better with more layers in exchange for an increase in training time.



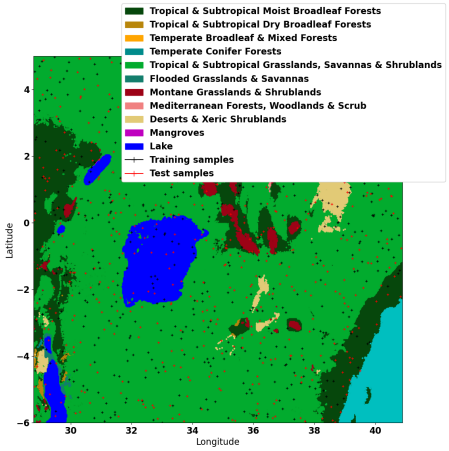
(a) Climplicit



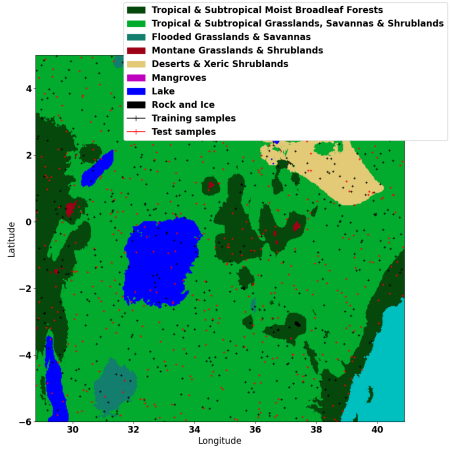
(b) Groundtruth



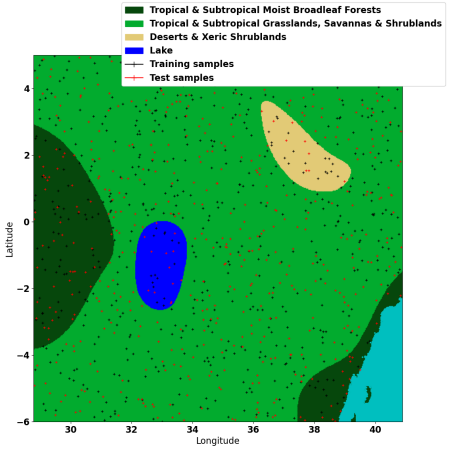
(c) FS LOC



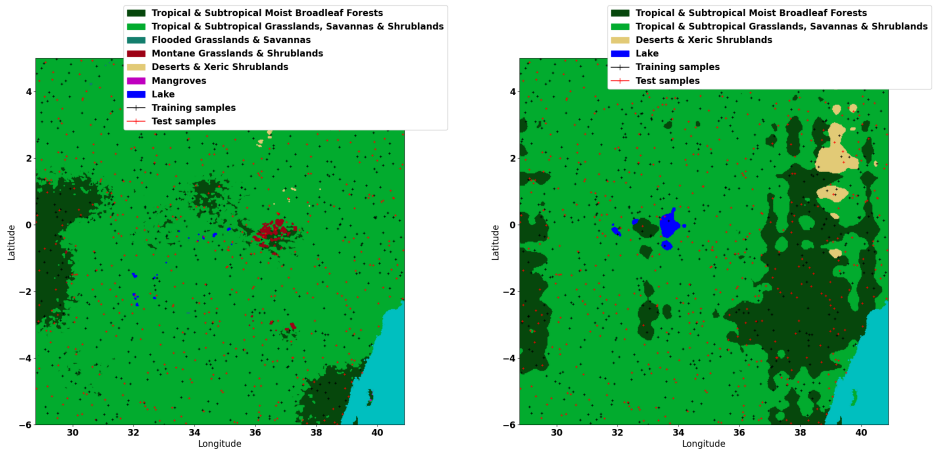
(d) FS CH



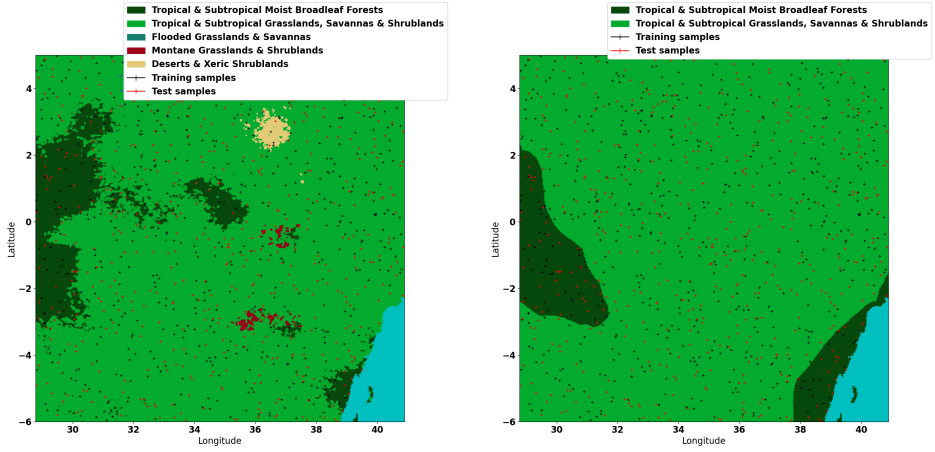
(e) FS LOC + CH



(f) SatCLIP

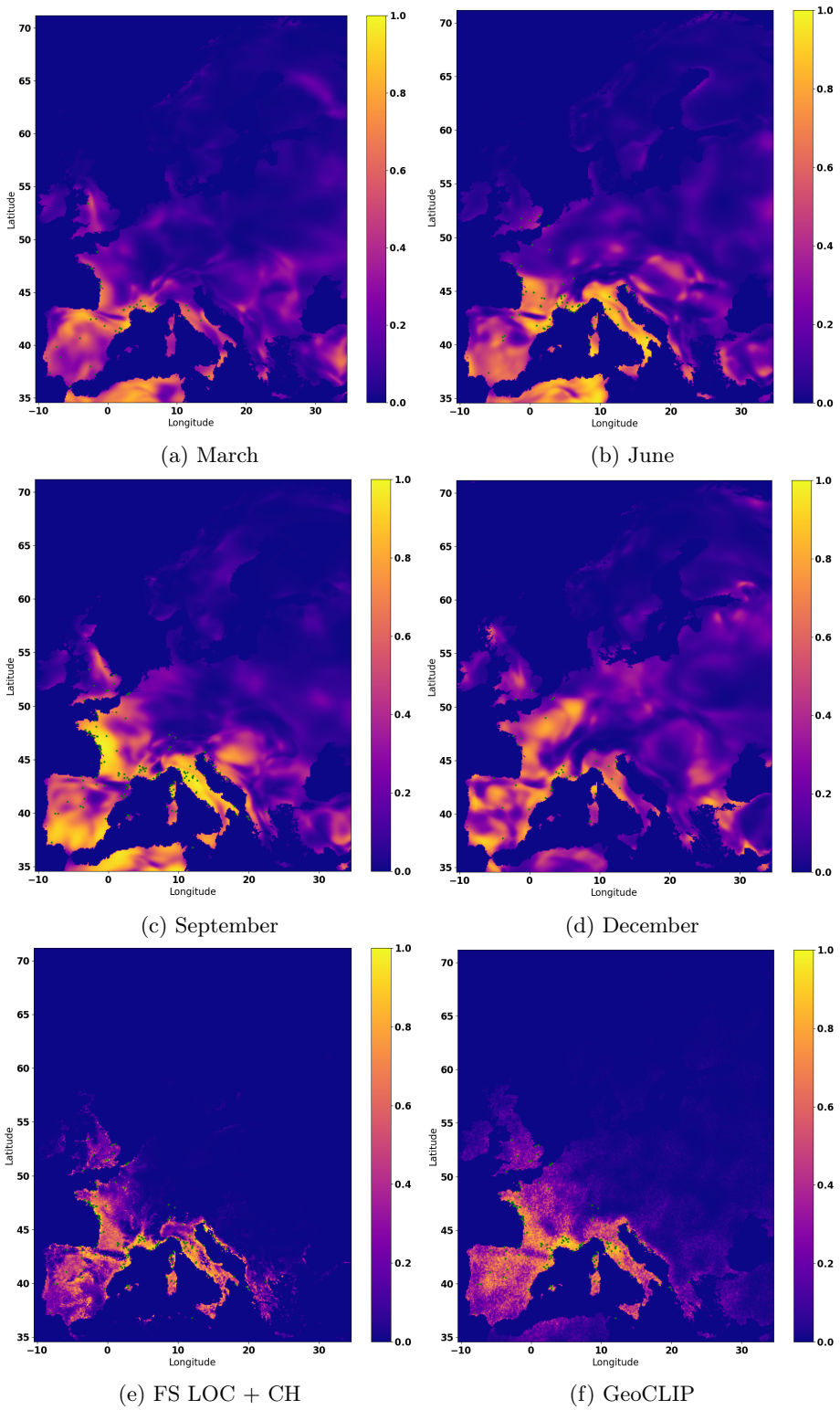


(g) GeoCLIP (h) CSP



(i) Taxabind (j) SINR

Figure 5: Comparison of the fully-trained and pretrained embeddings by mapping the predicted biomes in the region around Lake Victoria, covering parts of Kenya, Tanzania, Uganda, Rwanda, Burundi and the Democratic Republic Kongo. CLIMPLICIT achieves a good reconstruction without any hallucinations, albeit missing high-level details such as the mangroves along the coast.

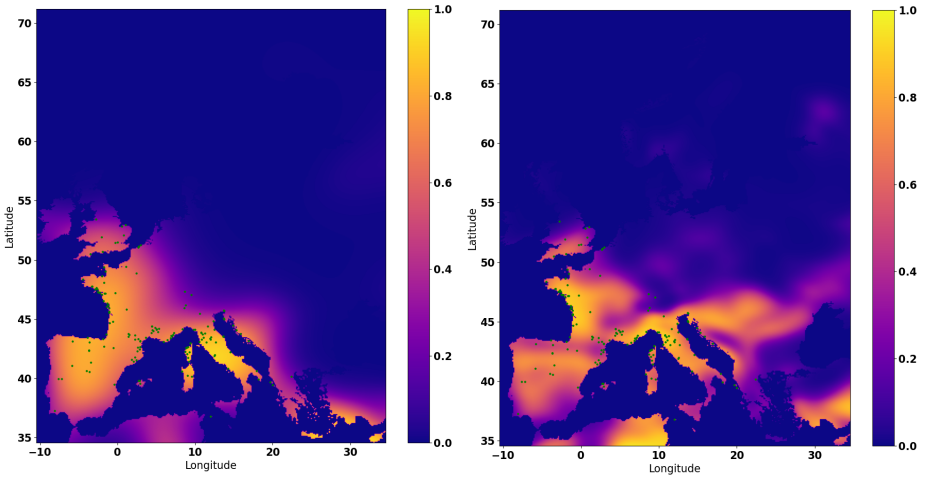


(a) March (b) June

(c) September (d) December

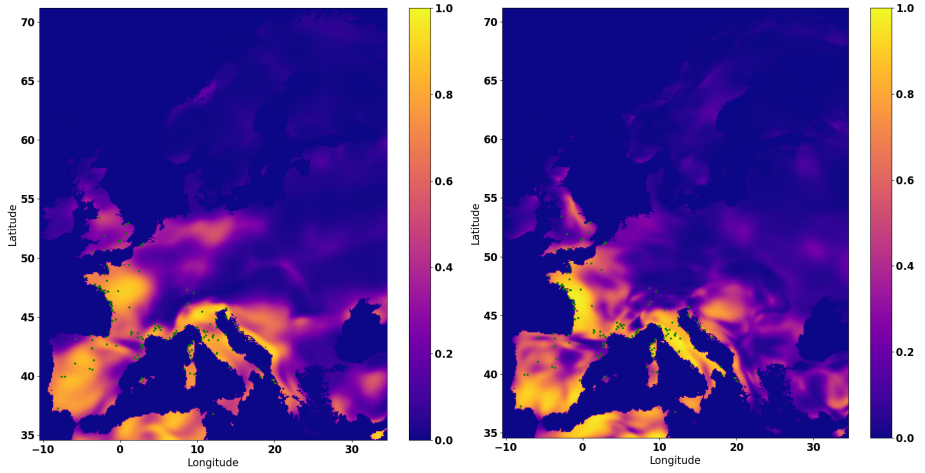
(e) FS LOC + CH (f) GeoCLIP

Figure 6: Visualization of the species distribution for *Quercus ilex* (Stone oak) for four months learned by probing the CLIMPLICIT embeddings in the SDM task. For each location, the model provides a probability of observing the oak (Blue: low, Yellow: high). The green dots represent the citizen scientist records of Quercus ilex for the respective month. Due to using months as input, the model learns different embeddings for each month, which allows for diverse distribution learning across the year. Additionally we present the distributions for the two strongest baselines.



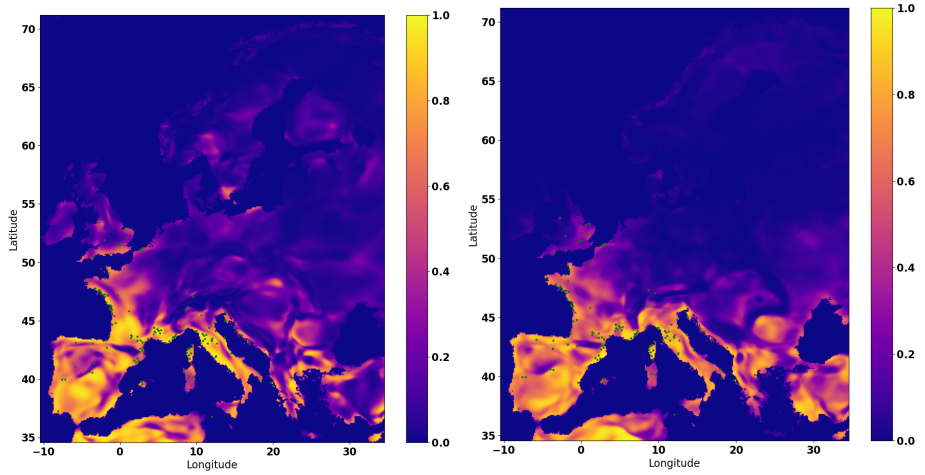
(a) 2 layers

(b) 4 layers



(c) 8 layers

(d) 16 layers



(e) 32 layers

(f) SIREN 32 layers

Figure 7: Visualization of the species distribution for *Quercus ilex* (Stone oak) after probing the ReSIREN networks pretrained with different numbers of layers. All distributions are for the month of September. Blue to yellow signifies the predicted probability and green signifies the occurrences of the oak in September. It shows that with increasing depth, the level of detail also increases, though at the cost of increasing extrapolation (e.g. Skåne county in Sweden). When using MLP probing, we do not observe such extrapolation anymore. We also present the distribution learned by a 32 layer SIREN.