DeepEarth: Geospatial Deep Simulator of Earth's Ecosystems at Landscape Scale

Lance Legel[1]* Daniel Neamati[2] Mira Partha[2] Barnabas Daru[3] Zachary Brym[4] Teagan Holly Young[4] Ryan Klein[4] Jorg Peters[5] Mayra Vidal[6] David Althoff[7] Kari Segraves[7] Makenzie Mabry[8] Robert Guralnick[8] Douglas Soltis[8] Pamela Soltis[8]

[1] *Ecological Intelligence Lab, Ecodash.ai*
[2] *Navigation & Autonomous Vehicles Lab, Stanford University*
[3] *Biogeography Lab, Stanford University*
[4] *Institute of Food and Agricultural Sciences, University of Florida*
[5] *Geometric Algorithms, 3D Graphics, & Simulation Lab, University of Florida*
[6] *Species Interactions Lab, University of Massachusetts Boston*
[7] *Evolutionary Ecology Lab, Syracuse University*
[8] *Florida Museum of Natural History, University of Florida*

Abstract

We propose *DeepEarth*, a deep neural network for modeling Earth's ecosystems, especially for prediction of *plant-pollinator interactions*. DeepEarth learns to simulate biological, geological, and ecological data across space and time. Training data includes biodiversity records, remote imagery, soil surveys, and climate models. DeepEarth learns by reconstructing partially or fully masked data, sampled around (x, y, z, t) coordinates of ecological observations. Prior knowledge from validated scientific models guides optimization as Bayesian constraints. We propose to protoype DeepEarth for simulation of flowering and pollination in native ecosystems of California and Florida between 2010 to 2025. We will evaluate our model through (*i*) *in silico* accuracy of predicted plant-pollinator distributions vs. real observations unseen during model training, and (*ii*) abundance and biodiversity of pollinators attracted through native pollinator garden field trials advised by our model.

1 Introduction

Pollinators (*e.g.* bees, butterflies, hummingbirds) provide vital ecosystem services for life on Earth. They cultivate $500 billion per year in global food crops and 88% of all flowering plants [1–4]. Yet pollinators are declining worldwide with dire consequences [5–10]. Scientific fields such as biodiversity, biogeography, conservation biology, restoration ecology, and natural capital have emerged to support global sustainability of such ecosystems, inspiring generations of ecologists and conservationists to help preserve over 16% of Earth's land as of 2025 [11–22]. Yet models of Earth's ecosystems – *e.g.* habitats, phenology, mutualisms – often fail to accurately and preciely predict real distributions at landscape scale ($< 1 \times 1$ m), limiting their utility [23–30].

Here we introduce *DeepEarth* for scientific modeling of Earth's ecosystems at landscape scale. DeepEarth is a Bayesian deep neural network designed to scale geographically across the planet, and biologically across the tree of life. DeepEarth represents an integrative deep encoding of diverse biological, geological, and ecological data, including herbaria and iNaturalist records of species presence, visible and hyperspectral aerial and remote sensing, 3D topographic maps from LiDAR and photogrammetry, laboratory-grade soil surveys including metrics of soil chemistry and texture,

*Correspondence: `lance@ecodash.ai`

Preprint.

as well as dynamic weather and climate models at hourly resolution over large periods of the study. This data (and additional candidate data sources for unsupervised machine learning) are detailed in Table 1 below, while variables to be jointly encoded are detailed in Table 2.

Table 1: **Datasets for DeepEarth.** Multi-modal data sources for training and validation of DeepEarth, selected especially for prediction of plant–pollinator interactions.

Reference	Details & Value for Ecosystem Modeling	Source
iNaturalist	Millions of *in situ* timestamped geotagged photos of species for modeling distributions, phenology, and interactions.	[31]
iDigBio	*iDigBio* features millions of historical biodiversity records prior to 1900 for modeling of native ecosystems.	[32]
PhyloRegion	*Phyloregion* is a geographic species distribution model with biological constraints from evolutionary genetics.	[33]
PhenoVision	*PhenoVision* deep learned a model for 50 million iNaturalist images, trained to classify flowering with 98.5% accuracy.	[34]
GloBI	330,000 pollination events between plants & pollinators have been indexed by *Global Biotic Interactions* since 2014.	[35]
InVEST	*InVEST Pollination* models pollination using LULC maps, land cover attributes, pollinator guild presence and flight ranges.	[36]
USDA-Aerial	HD aerial RGBI imagery (0.33 meters / pixel) from USDA (*NAIP*) for vivid agricultural intelligence across the US.	[37]
NASA-Thermal	HD infrared satellite from NASA (*GOES-R*) for hourly tracking of weather, cloud, and temperature fields.	[38]
ESA-Hyperspectral	Hyperspectral satellite (*DESIS*) with 235 frequencies for spectrographic deep learning of physics & chemistry.	[39]
WorldClim	19 bioclimatic metrics of temperature and precipitation between 1970-2000 at 1km resolution from *WorldClim*.	[40]
NOAA-Weather	Hourly simulation of land hydrology and energy (*e.g.* evapotranspiration, heat flux) from NASA and NOAA.	[41]
SoilSurvey	Precision USDA soil surveys (*e.g.* pH, organic matter, texture) for constraining underlying properties of ecosystems.	[42]
HydroSHEDS	HydroSHEDS v2 global watersheds and river flow networks from elevation data at 12m scale.	[43]
USGS-3DEP	USGS 3D topography data shaping local microclimates and water flow, structuring species–landscape interactions.	[44]
GeoFusion	Fuses 3D data from LiDAR, iOS ARKit, GCPs, and GNSS RTK into global geodetic coordinates for photorealistic 3D mapping.	[45]

2 DeepEarth: A Deep Generative AI Foundation Model for Ecology

DeepEarth is a deep generative AI foundation model for geology, biology, and ecology [46]. It is an autoregressive self-supervised machine learning model. We previously proposed a self-supervised neural network architecture for training DeepEarth, described as an *inductive* neural network [47]. The essence of an inductive neural network is to learn generative relationships across multi-modal data in a latent space, such that the model learns to not only reconstruct hidden ground truth data, but also to interpolate dense field models from sparse data. In this way, we expect to reconstruct in the latent space of DeepEarth the equivalent of a dense matrix of plant-pollinator species interactions, conditioned on time of year and location on the planet.

Table 2: **Modeling Variables for DeepEarth.** Below are data for ecological deep learning, including geospatial and temporal resolutions, metric units, and sources. Sampling of data for machine learning will be constrained to native plants & pollinators in California and Florida, between 2010 and 2025. Following successful experiments, deep learning can scale globally across the tree of life.

REMOTE IMAGERY	***Spatial***	***Temporal***	***Shape × Units***	***Data Source***
Visible (aerial)	0.3 m	2yr	$H \times W \times$ RGBI	**USDA-Aerial**
Infrared (satellite)	2 km	1 hr	$H \times W \times 16 \times$ m	**NASA-Thermal**
Hyperspectral (satellite)	30 m	5 yr	$H \times W \times 235 \times$ nm	**ESA-Hyperspectral**
Geotagged 3D Gaussian Splats	0.01 m	1 sec	N × (x, y, z, r, g, b,)	**GeoFusion**
BIOLOGY & ECOLOGY				
In Situ Nature Imagery	0.01 m	1 sec	$3 \times H \times W \times$ RGB	**iNaturalist**
Species Habitat Range Model	0.1 km	1 day	12k × taxa	**PhyloRegion**
Species Distribution Records	10 m	1 sec	$N \times$ taxa	**iDigBio**
Plant Flowering Phenology	10 m	1 sec	12k × taxa	**PhenoVision**
Plant-Pollinator Flower Visits	10 m	1 sec	2 × taxa	**GloBI**
Floral Resources Index	30 m	3 mo	blooms/km^2/season	**InVEST**
Habitat Nesting Suitability	30 m	3 mo	nests/km^2	**InVEST**
Relative Species Abundance	30 m	3 mo	species/km^2/season	**InVEST**
Pollinator Abundance	30 m	3 mo	visits/flower/season	**InVEST**
CLIMATE				
Precipitation	12 km	1 hr	kg/m^2	**NOAA-Weather**
Air Temperature (2m)	12 km	1 hr	K	**NOAA-Weather**
Specific Humidity (2m)	12 km	1 hr	%	**NOAA-Weather**
Convective Available Potential Energy	12 km	1 hr	J/kg	**NOAA-Weather**
Surface Albedo	12 km	1 hr	%	**NOAA-Weather**
Sensible / Ground / Latent Heat Flux	12 km	1 hr	W/m^2	**NOAA-Weather**
Wind (speed, direction)	2 km	1 hr	m/s, $^\circ$	**NASA-Thermal**
Downward Shortwave Radiation	2 km	1 hr	W/m^2	**NASA-Thermal**
Bioclimatic Metrics	1 km	1 yr	19 × kg/m^2, K	**WorldClim**
SOIL				
pH	1 m	10 yr	$0 - 14$	**SoilSurvey**
Sand, Silt, Clay, Organic Matter	1 m	10 yr	4 × %	**SoilSurvey**
Bulk Density	1 m	10 yr	g/cm^3	**SoilSurvey**
Depth to Water Table	1 m	10 yr	cm	**SoilSurvey**
Saturated Hydraulic Conductivity	1 m	10 yr	μm/s	**SoilSurvey**
Available Water Capacity	1 m	10 yr	%	**SoilSurvey**
HYDROLOGY & TOPOGRAPHY				
Surface / Subsurface Runoff	12 km	1 hr	kg/m^2	**NOAA-Weather**
Streamflow	12 km	1 hr	m^3/s	**NOAA-Weather**
Snow Water Equivalent	12 km	1 hr	kg m^{-2}	**NOAA-Weather**
Evapotranspiration (mass/energy flux)	12 km	1 hr	kg, W/m^2	**NOAA-Weather**
Mean Annual Discharge	12 m	10 yr	m^3 s^{-1}	**HydroSHEDS**
Digital Elevation Model	10 m	5yr	m	**USGS-3DEP**
3D Topographic LiDAR	10 m	5yr	$N \times (x, y, z)$ m	**USGS-3DEP**
Slope / Aspect	10 m	5yr	$^\circ$	**USGS-3DEP**

Through loss regularlization, we will integrate validated scientific models (Bayesian constraints) into DeepEarth machine learning. In this way, our model can naturally discover new patterns given sufficient evidence, but rely on existing scientific models otherwise.

We propose to first protoype DeepEarth for simulating pollination ecosystems. We will simulate:

- Plant and pollinator habitat distributions at landscape scale, including migration patterns
- Plant flowering phenology with geospatially distributed bloom schedules
- Mutualistic plant-pollinator relationships through species visitation networks

We will focus on simulating ecological observations around native plant and pollinator species in California and Florida between 2010 to 2025. Our initial prototyping will focus on commercially–cultivated native plant species, distributed across ecologically diverse and economically important regions of the states.

We will open-source the DeepEarth model at `https://github.com/legel/deepearth`. Through integrative ecological modeling ranging from landscape to regional climate scales, DeepEarth could fundamentally transform our ability to simulate and understand global ecosystem dynamics. Such an AI model could dramatically help to automate, scale, and optimize our engagement with complex ecosystems, which globally yield over $175 trillion per year in services to society [48].

References

[1] Jeff Ollerton, Rachael Winfree, and Sam Tarrant. **How many flowering plants are pollinated by animals?** *Oikos* 120.3 (2011), pp. 321–326.

[2] IPBES. *The Assessment Report on Pollinators, Pollination and Food Production*. IPBES secretariat, Bonn, Germany. 2016.

[3] Sven Lautenbach et al. **Spatial and temporal trends of global pollination benefit**. *PloS one* 7.4 (2012), e35954.

[4] Taylor H Ricketts et al. **Economic value of tropical forest to coffee production**. *Proceedings of the National Academy of Sciences* 101.34 (2004), pp. 12579–12582.

[5] Simon G. Potts et al. **Global pollinator declines: trends, impacts and drivers**. *Trends in Ecology & Evolution* 25.6 (2010), pp. 345–353. doi: `10.1016/j.tree.2010.01.007`.

[6] Insu Koh et al. **Modeling the status, trends, and impacts of wild bee abundance in the United States**. *Proceedings of the National Academy of Sciences* 113.1 (2016), pp. 140–145.

[7] Robert Watson et al. **Summary for policymakers of the global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services**. *IPBES Secretariat: Bonn, Germany* (2019), pp. 22–47.

[8] Joseph Millard et al. **Global effects of land-use intensity on local pollinator biodiversity**. *Nature Communications* 12.1 (2021), p. 2902.

[9] Lynn V Dicks et al. **A global-scale expert assessment of drivers and risks associated with pollinator decline**. *Nature ecology & evolution* 5.10 (2021), pp. 1453–1461.

[10] Christian Lippert, Arndt Feuerbacher, and Manuel Narjes. **Revisiting the economic valuation of agricultural losses due to large-scale changes in pollinator populations**. *Ecological Economics* 180 (2021), p. 106860.

[11] Michael E Soulé. **What is conservation biology?** *BioScience* 35.11 (1985), pp. 727–734.

[12] Navjot S Sodhi and Paul R Ehrlich. *Conservation biology for all*. Oxford University Press, 2010.

[13] Edward O Wilson et al. **Biodiversity** (1988).

[14] Julie M Allen et al. **Biodiversity synthesis across the green branches of the tree of life**. *Nature Plants* 5.1 (2019), pp. 11–13.

[15] Mark V Lomolino, Brett R Riddle, and Robert J Whittaker. **Biogeography** (2017).

[16] Barnabas H Daru, Michelle van der Bank, and T Jonathan Davies. **Unravelling the evolutionary origins of biogeographic assemblages**. *Diversity and Distributions* 24.3 (2018), pp. 313–324.

[17] Michael P Perring et al. **Advances in restoration ecology: rising to the challenges of the coming decades**. *Ecosphere* 6.8 (2015), pp. 1–25.

[18] TPl Young, DA Petersen, and JJ Clary. **The ecology of restoration: historical links, emerging issues and unexplored realms**. *Ecology letters* 8.6 (2005), pp. 662–673.

[19] Peter Kareiva et al. *Natural Capital: Theory and Practice of Mapping Ecosystem Services*. Oxford University Press, Apr. 2011. isbn: 9780199588992. doi: `10.1093/acprof:oso/9780199588992.001.0001`.

[20] James EM Watson et al. **The performance and potential of protected areas**. *Nature* 515.7525 (2014), pp. 67–73.

[21] UNEP World Conservation Monitoring Centre and IUCN. *World Database on Protected Areas (WDPA)*. `https://www.protectedplanet.net`. Accessed: March 22, 2025. 2025.

[22] Natural Capital Project. *Database of publications using InVEST and other Natural Capital Project software*. 2024. doi: `10.25740/bb284rg5424`.

[23] Christian König et al. **Scale dependency of joint species distribution models challenges interpretation of biotic interactions**. *Journal of Biogeography* 48.7 (2021), pp. 1541–1551.

[24] Julie A. Lee-Yaw et al. **Species distribution models rarely predict the biology of real populations**. *Ecography* 2022.6 (2022), e05877.

[25] Guadalupe Peralta et al. **Predicting plant–pollinator interactions: concepts, methods, and challenges**. *Trends in Ecology & Evolution* 39.5 (2024), pp. 494–505.

[26] Gretchen C Daily et al. **Ecosystem services in decision making: time to deliver**. *Frontiers in Ecology and the Environment* 7.1 (2009), pp. 21–28.

[27] Rebecca Chaplin-Kramer et al. **Global modeling of nature's contributions to people**. *Science* 366.6462 (2019), pp. 255–258.

[28] Barnabas H Daru. **Predicting undetected native vascular plant diversity at a global scale**. *Proceedings of the National Academy of Sciences* 121.34 (2024), e2319989121.

[29] Yujing Yan, Bin Shao, and Charles C Davis. **A generative deep learning approach for global species distribution prediction**. *bioRxiv* (2024), pp. 2024–12.

[30] Marie C Dade et al. **Landscapes—a lens for assessing sustainability**. *Landscape Ecology* 40.2 (2025), p. 28.

[31] iNaturalist. *iNaturalist: Exploring and Connecting with Nature*. Accessed: 2025-03-11. 2021.

[32] iDigBio. *Integrated Digitized Biocollections*. `https://www.idigbio.org`. Accessed: March 12, 2025. 2025.

[33] Barnabas H Daru et al. **Understanding the processes underpinning patterns of phylogenetic regionalization**. *Trends in ecology & evolution* 32.11 (2017), pp. 845–860.

[34] Russell Dinnage et al. **PhenoVision: A framework for automating and delivering research-ready plant phenology data from field images**. *bioRxiv* (2024). doi: `10.1101/2024.10.10.617505`. eprint: `https://www.biorxiv.org/content/early/2024/10/14/2024.10.10.617505.full.pdf`.

[35] Jorrit H. Poelen, James D. Simons, and Chris J. Mungall. **Global Biotic Interactions: An open infrastructure to share and analyze species-interaction datasets**. *Ecological Informatics* 24 (2014), pp. 148–159. doi: `10.1016/j.ecoinf.2014.08.005`.

[36] Natural Capital Project. *InVEST 3.15.0*. 2025.

[37] United States Department of Agriculture. *USDA NAIP: National Agriculture Imagery Program*. Aerial Imagery Dataset. 2023.

[38] National Aeronautics and Space Administration. *NASA GOES-R: Geostationary Operational Environmental Satellite*. Infrared and Radiance Data. 2022.

[39] European Space Agency. *ESA DESIS: DLR Earth Sensing Imaging Spectrometer*. Satellite Hyperspectral Dataset. 2023.

[40] Stephen E. Fick and Robert J. Hijmans. *WorldClim: Global Climate Data*. Bioclimatic Data Layers. 2022.

[41] National Oceanic and Atmospheric Administration. *NOAA NLDAS: North American Land Data Assimilation System*. 2023.

[42] United States Department of Agriculture. *USDA NCSS: National Cooperative Soil Survey*. Soil Properties Database. 2023.

[43] Bernhard Lehner et al. "HydroSHEDS v2. 0-Refined global river network and catchment delineations from TanDEM-X elevation data". *EGU General Assembly Conference Abstracts*. 2021, EGU21–9277.

[44] United States Geological Survey. *USGS 3DEP: 3D Elevation Program*. Topography and LIDAR Data. 2024.

[45] Christoph Strecha, Martin Rehak, and Davide Cucci. **The Accuracy of RTK enabled Mobile Phones in RTK denied Areas**. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 10 (2024), pp. 327–332.

[46] Lance Legel. *Artificial Intelligence Foundation Models for Biogeography and Ecophysiology*. Pre-print, Ecodash.ai. Accessed: 2025-03-11. Dec. 2024.

[47] Lance Legel. *Multimodal Autoregressive Neural Networks for Ecology*. Pre-print, Ecodash.ai. Accessed: 2025-01-25. 2025.

[48] Robert Costanza et al. **Changes in the global value of ecosystem services**. *Global environmental change* 26 (2014), pp. 152–158.