Deep Multi-modal Species Occupancy Modeling

Timm Haucke[1,*] Lauren Harrell[2] Yunyi Shen[1] Levente Klein[3]
David Rolnick[4] Lauren Gillespie[1] Sara Beery[1]

[1]MIT
[2]Google
[3]IBM
[4]McGill University

[*]Corresponding author: haucke@mit.edu

Abstract

Occupancy models are tools for modeling the relationship between habitat and species occurrence while accounting for the fact that species may still be present even if not detected. The types of environmental variables typically used for characterizing habitats in such ecological models, such as precipitation or tree cover, are frequently of low spatial resolution, with a single value for a spatial pixel size of, e.g., $1\,\mathrm{km}^2$. This spatial scale fails to capture the nuances of micro-habitat conditions that can strongly influence species presence, and additionally, as many of these are derived from satellite data, there are aspects of the environment they cannot capture, such as the structure of vegetation below the forest canopy. We propose to combine high-resolution satellite *and* ground-level imagery to produce multi-modal environmental features that better capture micro-habitat conditions, and incorporate these multi-modal features into hierarchical Bayesian species occupancy models. We leverage pre-trained deep learning models to flexibly capture relevant information directly from raw imagery, in contrast to traditional approaches which rely on derived and/or hand-crafted sets of ecosystem covariates. We implement deep multi-modal species occupancy modeling using a new open-source Python package for ecological modeling, designed for bridging machine learning and statistical ecology. We test our method under a strict evaluation protocol on 16 mammal species across thousands of camera traps in Snapshot USA surveys, and find that multi-modal features substantially enhance predictive power compared to traditional environmental variables alone. Our results not only highlight the predictive value and complementarity of in-situ samples, but also make the case for more closely integrating deep learning models and traditional statistical ecological models while maintaining their interpretability.

1 Introduction

Understanding where and under what conditions species occur is of critical importance, since it enables understanding species niches [25], planning effective conservation measures [40], and preventing human-wildlife conflict [27]. Species distribution models formalize this idea by using statistical models to link environmental factors such as local climate with observations of species presence. Unlike classic species distribution models that rely primarily on presence-only observations, *occupancy* models are able to explicitly account for imperfect detection by using both observed detections and non-detections [62]. The ability to disentangle the probability of occupancy from the probability of detection allows occupancy models to produce less biased estimates of species occupancy [38, 39]. Occupancy models have successfully been

deployed to predict the risk of wildlife road crossings [84], support early detection of invasive species [6], identify high-priority conservation areas [19] and wildlife corridors [103], prioritize habitat restoration efforts [86], and support harvesting decision making [31].

However, classical approaches predominantly predict species occurrence based on environmental variables inferred by global models (e.g., climate or soil models), which often have limited spatial and temporal resolution [3, 41, 33, 30, 28]. Recent work has started to address these issues by leveraging higher-resolution satellite imagery as input to deep learning models that directly predict species distributions and encounter rates [93, 74, 35]. However, satellite imagery still frequently fails to capture micro-habitat conditions, with intuitive examples being the inability to "see" under the canopy or in areas with a lot of cloud cover [66, 104].

The in-situ imagery of large-scale camera trapping efforts gives us a complementary and direct picture of micro-habitats on the ground, and additionally captures detection *and* non-detection data for a wide variety of species over relatively long time-spans (weeks to months). We therefore propose to combine satellite-derived macro-environmental context with the aforementioned camera-driven micro-habitat clues. Deep learning models provide a way to extract these micro-habitat clues via their high-level internal representations of the raw image data. By treating the resulting multi-modal deep features as covariates within a Bayesian occupancy modeling framework, we obtain models that exploit the full depth of environmental information while maintaining the statistical rigor of detection/non-detection inference.

We demonstrate the value of multi-modal deep features by fitting single-species single-season occupancy models on large-scale data from the Wildlife Insights platform [1], which allows camera trap users to upload and automatically classify their imagery. We then evaluate the fitted models on publicly available camera trap data from the nationwide Snapshot USA camera trapping surveys [80], showing that 12 out of our 16 focal species are modeled most accurately when using multi-modal deep features as predictors of species occupancy.

To summarize, our contributions are as follows:

- We propose a novel method that **combines modern AI with statistical ecology**, enabling Bayesian hierarchical ecological models to benefit from environmental features extracted by deep learning models.
- Our method improves the predictive power of occupancy models by incorporating **multi-modal** deep features extracted from satellite and in-situ camera trap imagery, and propose methods for **interpreting** them qualitatively.
- We propose a strict **evaluation protocol** for occupancy models that enforces separation across surveys and geography. This enables realistic evaluation and fair comparisons of models that leverage high-dimensional deep features.
- We introduce a new open-source ecological modeling software package written in the Python programming language, to help **bridge gaps** between machine learning and statistical ecology.

2 Related work

2.1 Species distribution modeling

In 1904, Grinnell first established qualitatively that patterns of species occurrence can be inferred from environmental factors such as humidity [36]. In the century since, a rich body of quantitative research has emerged around associating environmental factors with the occurrence of specific species. The most widespread approaches are called species distribution models (SDMs), which generally aim to model the relationship between species and habitats through environmental variables such as climate, elevation, or human presence. These models are used by land managers and policy makers to, e.g., understand risks to species and prioritize conservation action, critical for both ecological research and practical on-the-ground species and habitat conservation. The model methodologies used for SDMs and the interpretation of the model results are primarily driven by what kind of

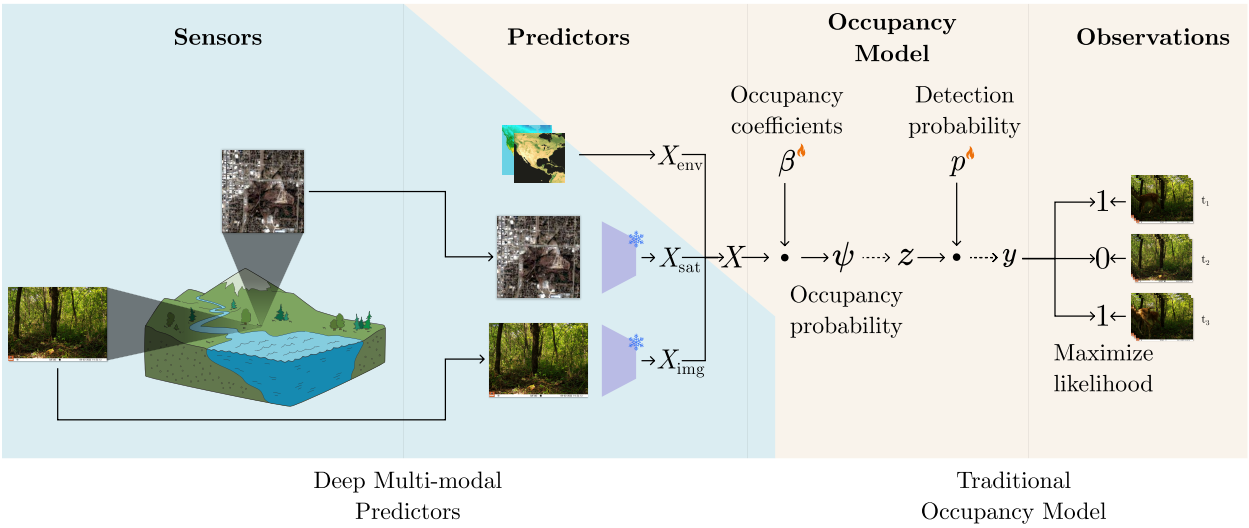

Figure 1: **We propose the first method that enables deep features extracted from multiple ecological data modalities to be directly incorporated as additional environmental context in ecological occupancy models.** Our method combines environmental variables X_{env} with deep feature vectors extracted from a single blank image per site X_{img} and from satellite imagery X_{sat}. These features are concatenated into a unified feature vector X, which is then used by the occupancy model to regress the occupancy probability ψ. The occupancy model is otherwise fit as usual: ψ is used to draw discrete occupancy states $z \in 0, 1$, which are multiplied by the detection probability p derived from detection coefficients α to draw binary detection/non-detection observations y, the likelihoods of which are then maximized under true observations derived from image labels. As used in the machine learning community, the snowflake icons represent parts of the model that are "frozen", i.e. fixed and not updated during the fitting process. In contrast, the flame icons emphasize the parameters in the model that are fit to data. Example camera trap images are from Snapshot USA 2022 [80].

species occurrence records are used as inputs. There are two primary types of data commonly used to fit SDMs: (1) presence-only data and (2) detection/non-detection data (also called presence-absence[1] data). The majority of SDM methods are designed to address the specific nuances of either presence-only or detection/non-detection data, however there are some models that are designed to integrate both types [24, 23] or other types of occurrence records such as abundance data [53].

Presence-only data, in which only records that contain a positive detection or incidence of a species occurrence, is the most abundant data available for modeling. Due to a lack of "negative" detections in the data, methods developed for presence-only data typically derive either "pseudo-absence" or "background" locations [7]. For single-species distribution models, common model methodologies include generalized linear models, generalized additive models, MaxEnt [73], and random forest/tree-based models [99] and leverage environmental features to characterize the species distribution. The typical interpretation of the output

[1]Although detection/non-detection and presence/absence are often used interchangeably in the literature to express observations where there can be either positives or negatives for a species occurrence at a location, we use the language of detection/non-detection to describe the data because the treatment of non-detections as true absences can bias or confound SDM estimation [37, 60, 61]. Using the terms presence-absence implies that the absence data is reflective of true "non-occupancy" and not a lack of detection.

of these models is not occupancy but rather "habitat suitability" – where locations more similar to sites where species have been detected in the environmental feature space are predicted to have higher values.

Over the last decade, neural-network-based SDMs have emerged that can scalably model presence-only species detection data through species co-occurrence [15], where many recent innovations focus on deep-learning primarily through either novel loss functions that have different mechanisms for defining and weighting pseudo-absences and/or the inclusion of remote sensing features [101, 35, 18]. For example, Cole et al. have shown that spatial implicit neural representations (SINR) are effective at learning from large-scale citizen science data across hundreds of thousands of species and perform well for tasks like range estimation from jointly modeling presence-only species occurrence records with few or no environmental characteristics [18]. Heterogeneous graph neural networks have also emerged as deep-learning approaches that can add additional structure in learning species distributions and interactions from presence-only data [43].

Detection/non-detection data contains records of both positive species detections, where target species were found at surveyed locations, as well as non-detections across different places or areas for a given survey – where non-detections indicate a species that was included in the target species set of a survey, but not found at the location during the survey time. Detection/non-detection data can be used as positives/negatives for training binary classification models, but these results can be highly misleading similar to results from presence-only data [37]. Because detection is frequently imperfect, a species can still be present or occupy a given habitat without being detected, which means that *observed* non-detection (or absence) does not equal *true* absence [60].

The probability of detecting a species, if the species is truly present, may be highly sensitive to the observation method (e.g., human experts or automated sensors such as camera traps) and survey effort (e.g., survey duration, time of day, area surveyed, and number of repeated visits, if any). Accounting for these effects is difficult, especially if observer effects and effort are unknown or not characterized in a standardized way across surveys, as is the case with large-scale citizen science data repositories. Even if there is a combination of presence-only data and detection/non-detection data, it is common for many SDMs to reduce their training data to presence-only records because dealing with the nuances of survey details across many study design types increases the complexity. The focus of this paper is leveraging large-scale detection/non-detection records from repeated surveys using wildlife camera trap image data, but could be extended for integrated models that include both presence-only and detection/non-detection surveys.

2.2 Occupancy modeling

In contrast to presence-only SDMs, occupancy models disentangle the probability of species presence (or *occupancy*) from the probability of detecting said species [62]. As opposed to treating a species' use of a habitat as a binary classification model on detection/non-detection status, occupancy models treat the habitat usage or site occupancy as a latent, or unobserved variable, where there is no observable "ground truth" absence (or presence if false-positives are assumed). Occupancy models are able to estimate separate processes for the probability of detection and probability of occupancy by incorporating repeated survey data of species detection and non-detection - where the rate of positive detections versus non-detections allows for estimation of the probability of detecting a species given a site is occupied. Site occupancy is then estimated through a posterior distribution given the observed data, which could also include covariates for both the occupancy and the detection processes.

The ability to account for separate detection versus occupancy processes is particularly crucial for animals that are difficult to detect, and for which naively assuming absence based on few negative observations would significantly bias the occupancy estimate to a lower than actual level. For example, consider a nocturnal and elusive species like the American marten (Martes americana). An expert surveying this species during the day could be standing right on top of a marten den while at the same time being unaware of their presence, even over the course of multiple visits. In practice, not accounting for imperfect detection

of species like the American marten can lead to an underestimation in presence at 15–52% of surveyed sites [88].

Occupancy models have been widely applied to infer species distributions, population trends, and responses to environmental factors. Early applications often relied on manual surveys (e.g., visual observations by researchers in the field) and were thus limited to relatively small areas. In recent years, improved autonomous sensor technology such as camera traps and autonomous acoustic recording units has greatly expanded the scale and efficiency of occupancy studies.

Just like other SDMs, occupancy models typically use a small set of environmental variables as predictors of occupancy. Common environmental variables range from habitat structure (e.g., land-cover heterogeneity [67] and vegetation structure [57]), topography (e.g., elevation [5]), climate (e.g., mean annual temperature [5]), and human influence (e.g., presence of non-native fish species [42]). Other works are simply concerned with estimating average occupancy and thus forgo environmental variables as predictors and estimating a constant occupancy probability across sites [54].

2.3 Deep representations

Deep learning models are known to learn powerful, high-level representations of their input data. Representations can be learned using manually labeled data. For example, an image classification model such as SpeciesNet [32], which is trained to recognize various animal species, will likely learn representations associated with the species' specific shapes and textures. For intuition about what type of information these types of representations capture, we refer the reader to the work of Zellweger et al. [102].

Alternatively, representations can be learned without manually labeled data using a process called self-supervised learning (SSL), in which the model uses structure within the training data to obtain useful learning signals. For example, approaches such as MoCo [46] and SimCLR [16] process two slightly altered versions of a single image (e.g., different cropped regions) and try to make their representations as similar as possible, thereby capturing the high-level information common across both altered versions while minimizing the impact of "nuisance" information that is unique to each individual version and therefore the result of the alterations. These types of approaches have since been improved upon primarily by scaling the volume of training data, such as by DINOv2, which is trained on 1.2B images [69]. Although popularized by machine learning models trained on regular imagery, SSL methods have furthermore successfully been applied to learn generally useful representations of single-time-step satellite imagery [52, 63, 4]. More recent methods incorporate temporal information by focusing on series of satellite images over time, thereby capturing effects such as seasonality [97, 98, 51, 91, 10].

2.4 Deep features in ecological models

There has been growing interest in the research community in combining raw satellite imagery with features extracted from deep learning models as predictors for species distribution models [17, 21, 92]. [93] finds that leveraging satellite imagery in the RGB and NIR bands alongside environmental variables improves prediction of bird species encounter rates. Similarly, Sat-SINR [22] extends SINR [18] by adding satellite imagery as additional predictor, alongside environmental variables and geographic location, to map the distribution of plant species at high resolution. Compared to the aforementioned works, we go beyond incorporating satellite imagery and additionally consider in-situ imagery. TaxaBind [85] learns commonalities between satellite and in-situ citizen science imagery using contrastive self-supervised learning and explores species distribution modeling qualitatively by visually comparing generated maps with point-wise citizen science observations. In contrast, here we are focusing on Bayesian hierarchical occupancy models that explicitly account for imperfect detection, and we evaluate those models quantitatively.

3 Material and methods

3.1 Occupancy model

To take advantage of the rich detection/non-detection data provided by camera trapping surveys, we lever-

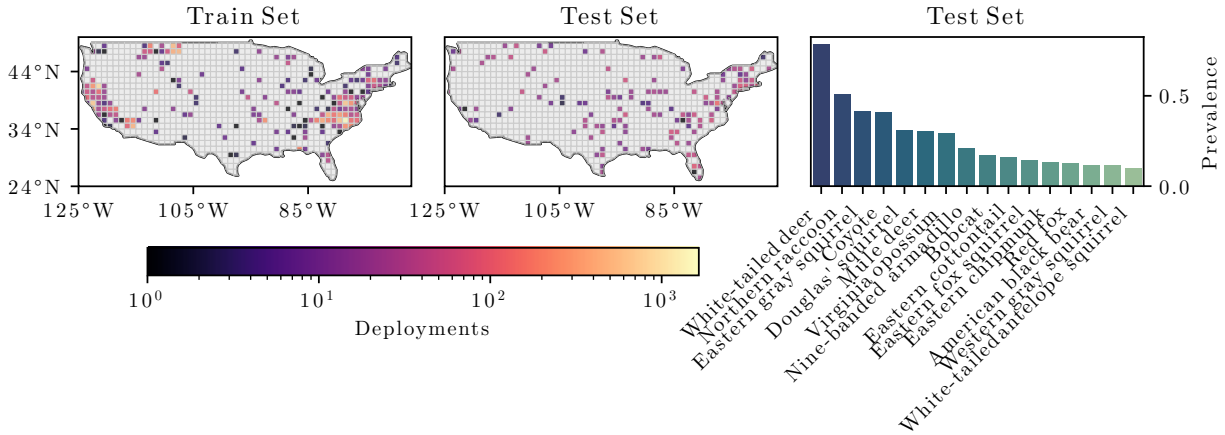

Figure 2: **We train and evaluate occupancy models for 16 species of varying prevalence across the contiguous United States.** The left and center panels show the density of train and test sites, respectively. The right panel shows prevalence across our focal species, here referring to the fraction of test-set sites within each species' range (see section 3.4) where the species was detected at least once. To protect the integrity of non-public train-set sites, we obfuscated their geographic location and only show their density at low resolution.

age the occupancy model proposed by MacKenzie et al. [62]. Instead of assuming perfect observations of species presence or absence, the occupancy status of a species is modeled as a latent variable. Specifically, for site i, the species' occupancy status is modeled as a Bernoulli latent variable $z_i \in \{0, 1\}$ with probability of occupancy ψ_i, i.e.:

$$z_i \sim \text{Bernoulli}(\psi_i) \tag{1}$$

For each site i and discrete time steps $j \in 1, ..., J_i$, detection/non-detections $y_{i,j} \in \{0, 1\}$ are then modeled conditional on the occupancy status with probability of detection p. Specifically, we have:

$$y_{i,j}|z_i \sim \text{Bernoulli}(p \times z_i) \tag{2}$$

We model the probability of occupancy ψ_i using a linear regression with the logit-link function, parameterized by coefficients β and predictors X_i at site i:

$$\log\left(\frac{\psi_i}{1-\psi_i}\right) = \beta^\top X_i \tag{3}$$

Although our model could also incorporate detection-level covariates, in this particular study, we forgo observation-level detection probability covariates and instead estimate a fixed detection probability p for each species across all sites to focus on how deep-learned features can improve occupancy-specific estimation. Incorporating detection covariates would add additional complexity that could make interpretation difficult in comparing candidate models.

3.2 Implementation

For simplicity, we do not consider species interactions and fit separate models for each species. Since our occupancy model expects observations at discrete time-steps, but our data consists of detections / non-detections at continuous times, we discretize observations $y_{i,j}$ at site i over one week long time steps j. We discretize such that any week with at least one observation counts as observed detection ($y_{i,j} = 1$) and any week without at least one observation counts as a non-detection ($y_{i,j} = 0$). This type of discretization into week-long time-steps is in line with prior work [55, 71] and provides a sufficient number of repeated samples while keeping computational cost reasonable. We ex-

clusively rely on human-labeled camera trap images to minimize the chance for false positive observations.

To estimate the parameters of this hierarchical Bayesian model, we have to integrate or sample over the posterior distribution of the latent space. To achieve this, we introduce a new open-source implementation of occupancy models we call *Biolith* [45]. Biolith is written in the Python programming language and uses the NumPyro probabilistic programming library [72] to perform Hamiltonian Monte Carlo inference using the No-U-Turn Sampler [49]. Being written in Python, our implementation enables machine learning to easily be incorporated into the ecological modeling workflow. More details on Biolith can be found in appendix E.

Reducing model complexity is crucial for reducing the risk of overfitting and for making models easier to interpret. Since our predictors X have on the order of 100 dimensions, fitting and performing model selection across all possible subsets of predictors is infeasible. Instead, we exclusively rely on regularization to reduce model complexity, a common approach, for example, in MaxEnt species distribution models [100]. We impose L2 or L1 regularization by placing Gaussian or Laplace priors on the model coefficients, respectively [94]. To find the optimal configuration of prior type and scale for each species and combination of predictors (see section 3.3), we perform 3-fold cross validation on the train set, stratified by naive occupancy (i.e., whether the site has at least one observation). This stratification ensures that there are sites with and without detections in the splits used to fit and validate the model. We then select the best configuration according to LPPD on the held-out portion, and re-fit on the entire train set.

3.3 Predictors

As baseline predictors, we select a similar set of environmental variables X_{env} as used by [93] for large-scale species distribution modeling. These covariates consist of bioclimatic variables [48] and soil properties [75]. The full set of variables can be found in appendix B.

Additionally, we make use of satellite and in-situ imagery to capture environmental information about the location that is higher resolution and perhaps more nuanced than what is captured by the environmental variables described above. In order to do so, we make use of various machine learning models that were trained using representation learning approaches (see section 2.3).

Specifically, we sample a single blank (i.e. containing no animal) daylight image per location and extract a high-dimensional representation X_{img} using a machine learning model. Depending on the exact model used, these representations vary between hundreds to thousands of dimensions and thus leads to significant memory requirements. To alleviate this and keep all data in memory, we apply principal component analysis and keep only the 128 dimensions with the most explained variance. We follow the same process to extract representations of satellite imagery X_{sat}. Finally, we simply concatenate all site-specific predictors into a single multi-modal habitat representation X_i for site i.

To choose which machine learning models to use for extracting representations from satellite and in-situ imagery, we fit occupancy models using combinations of environmental variables and the respective imagery ($X_{\text{env}}, X_{\text{sat}}$ and $X_{\text{env}}, X_{\text{img}}$, respectively). We then select the model that achieves the optimal normalized LPPD averaged over species and cross-validation splits (see section 3.2). In our experiments, we find that DINOv2 ViT-B/14 [69] and AlphaEarth Foundations [10] perform best, although performance differences between models are not always significant. We show results of additional models in appendix D.

3.4 Data

As a test bed, we focus our experiments on data from Wildlife Insights [1], an online platform that users upload, label and automatically classify species in large volumes of camera trap images. On Wildlife Insights, each camera trap image is associated with a single camera trap site (which might have an associated geographic location), and each site is in turn associated with a single project. Projects might represent individual camera trapping surveys or be part of a larger collection of surveys.

To learn the relationships between environmental context, provided by both environmental variables

and deep features from in-situ and satellite imagery, and species occupancy, we "fit" or "train" our deep multi-modal occupancy models across 16 species on a combination of public and non-public data from the Wildlife Insights platform (our "training set").

Since ecological data such as species detections/non-detections frequently has spatial and temporal structure, it is critical to account for such structure when evaluating models in order not to violate independence assumptions and report overly optimistic results [79]. This issue is made even more critical when using deep features within a model, since their high-dimensionality can more easily lead models to rely on nuisance variables that have no relation to the ecological processes driving species occupancy. As an example, camera trap hardware within a single survey is frequently identical, and model-specific metadata embedded within each image could theoretically be used by models to learn "shortcuts" between camera trap model and species occurrence within the survey's geographic region.

To make sure our evaluation is not overly optimistic and enable fair comparisons between models using different modalities, we extrapolate our models to a held-out "test set" of data from the public Snapshot USA camera trapping surveys from 2020 to 2023, also accessed through the Wildlife Insights platform. To prevent data leakage due to overlapping or visually similar satellite imagery, we furthermore remove any sites from the train set that are closer than 10 km to any test site.

We filter the set of sites used for fitting and evaluating models for each individual species based on range bounding boxes obtained from the Map of Life platform[2]. Restricting the area for each species has two primary motivations in this particular work: (1) limiting to known ranges helps focus learning fine-grained occurrence patterns rather than broad geographical patterns and (2) improves likely convergence of the models given that we are not estimating occupancy for highly improbable locations (which could result in rank deficiencies in the observed data matrices). If the goal is to provide larger-scale range maps from occupancy output or to estimate the effect of large-scale environmental patterns on the species occupancy probability, then filtering the data to the existing coarse range is not recommended.

The Eastern gray squirrel (Sciurus carolinensis), for example, is prevalent in the Eastern and Midwestern United States, but absent from most of the western United States despite suitable habitat. Filtering the set of considered sites in this way is in accordance with existing approaches that specifically focus on geographic areas where the species of interest is thought to occur a priori [44, 68]. The interpretation/inference of the resulting occupancy model is then limited to this range and similar environments, but this allows us to address nuances in environments where the species could occur to identify features that are more specific to the species occupancy within this range rather than across broader ranges. Since our datasets include a long tail of rarely observed species that are difficult to model, we focus on species that are observed at least once at 10% of sites or more across their range. This results in 16 species that are reasonably prevalent within their range, some of which encompass close to the entire contiguous United States and some of which are highly localized. Figure 2 shows the spatial density of camera trap sites and summarizes species prevalence across our focal species.

3.5 Evaluation protocol

Since occupancy models assume there are no single ground truth species presence/absence labels (see section 2.2), only detections / non-detections, we evaluate using the likelihood of test-set detection / non-detection data under the fitted model through the posterior predictive distribution. To construct the likelihood measure for each candidate occupancy model, we draw Q simulated datasets from the posterior distribution from the fitted model and compute the likelihood for each observation of the generated datasets to represent the posterior predictive distribution.

We then use log point-wise predictive density (LPPD) [34], computed on the held-out test data, to summarize the likelihoods across samples and observations:

$$\text{LPPD} = \sum_{i=1}^{n} \log \left(\frac{1}{Q} \sum_{q=1}^{Q} p(y_i|\theta^{(q)}) \right) \tag{4}$$

[2] https://mol.org

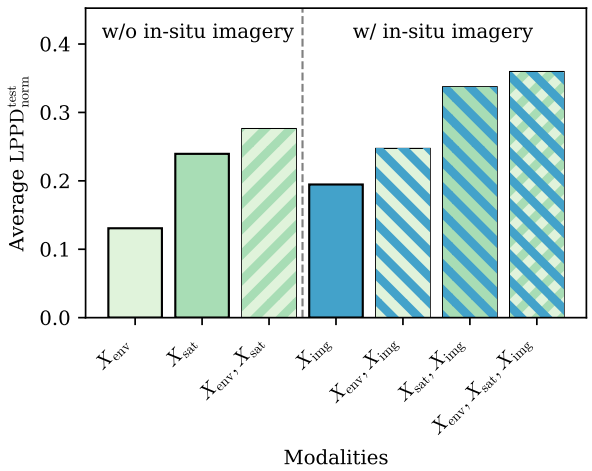

Figure 3: Predictive performance of different combinations of modalities, averaged across species, weighting each species equally. **Satellite imagery embeddings are the strongest single modality, but adding in-situ imagery yields a substantial improvement in predictive performance.** Environmental variables do not significantly improve models already incorporating satellite features, suggesting that deep features obtained from satellite imagery already capture most of the relevant information contained in environmental variables.

where Q is the number of posterior samples, n is the number of sites, J_i is the number of visits at site i, y_i is the observed detection history for site i across J_i revisits, y_{ij} is the observed detection at site i and visit j, $z_i^{(q)}$ is the latent occupancy state for site i in sample q, and $p_{ij}^{(q)}$ is the detection probability for site i and visit j in sample q.

The LPPD provides a measure of how divergent a set of observed data is from the posterior distribution of a fitted model. It is helpful because it allows us to evaluate model fit using only repeated observations and without having explicit ground truth labels for species presence/absence. However, similar to other likelihood-based measures, the magnitude of LPPD can be difficult to interpret and compare across species. To create a metric which provides more insight into the absolute performance of a given model and make it easier to compare across species, we compute a linear transformation of LPPD that places each model on a 0-1 scale defined by the LPPD of a "null" and an "oracle" model within a given species. The null model is a model that incorporates nothing but an intercept term for occupancy and detection, and is therefore limited to estimating a single mean occupancy probability over all sites. In contrast, the oracle model is fit to the otherwise held-out *test* data using site-level random effects. The oracle model therefore acts as an upper bound and intuitively represents a model whose predictors and parameters perfectly mirror the underlying ecological occurrence process on the test data. Specifically, given the LPPD of null and oracle models $\text{LPPD}_{\text{null}}, \text{LPPD}_{\text{oracle}}$, we normalize a given LPPD as follows:

$$\text{LPPD}_{\text{norm}} = \frac{\text{LPPD} - \text{LPPD}_{\text{null}}}{\text{LPPD}_{\text{oracle}} - \text{LPPD}_{\text{null}}} \tag{5}$$

We note that this normalization can result in negative values if the model of interest fails to outperform the null model.

4 Experiments

4.1 Predictive value

To compare the power of different predictors, we use the normalized LPPD and its macro-average over species (i.e., averaging across species while weighting each species equally). Figure 3 summarizes those results over different species and combinations of environmental variables, satellite imagery embeddings, and in-situ imagery embeddings. Satellite imagery embeddings are by far the strongest single predictor; however, combining satellite imagery with embeddings from in-situ imagery results in the strongest models overall. Whether those models are best with or without environmental variables varies between species. Models using satellite imagery embeddings seem to be largely invariant to whether or not environmental variables are also included as predictors, which suggests that the features extracted by the deep model capture most of the information contained within these variables.

Table 1 shows the normalized LPPD of models fit using different combinations of modalities as predictors for occupancy. Figure 4 highlights which species benefit the most and least from in-situ image features by comparing the best-performing model with in-situ image features to the best-performing model without them. We find that the value of in-situ imagery varies between species, with most species seeing clear benefits from its inclusion, the most significant being the Virginia opossum.

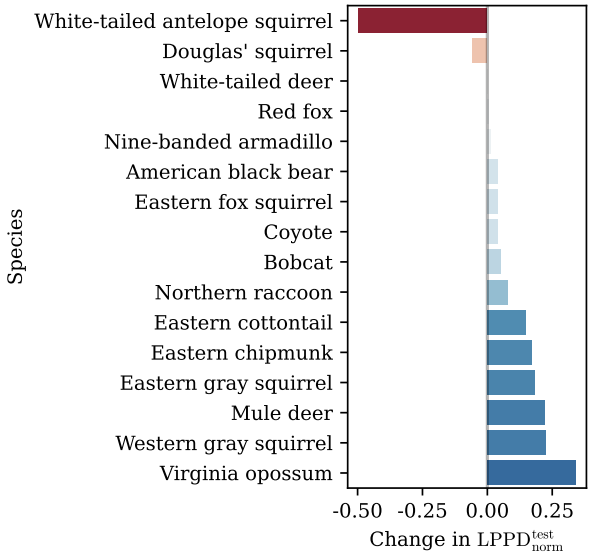

Figure 4: Improvement in test-set normalized LPPD of the best model with in-situ image features over the best model without. **Incorporating in-situ image features improves predictive power for the majority of species.** One species, the white-tailed antelope squirrel, is modeled perfectly using environmental variables only (see table 1) and including additional predictors seems to lead to overfitting (see section 4.2).

4.2 Overfitting and domain shift

Although most species are modeled more accurately by incorporating in-situ image embeddings, two outliers instead see a performance drop when these embeddings are added to environmental-variable-only models. This performance drop is significant for the white-tailed antelope squirrel, which is perfectly modeled using environmental variables alone. When additionally using in-situ image embeddings, performance degrades by over 60% (see table 1).

We observe the same trends for the simplified class of "naive" occupancy models that do not account for imperfect detection (see appendix A for detailed results). We hypothesize that the primary reason for this drop in performance and the variability of the predictive value of in-situ imagery in general is overfitting to the train set, amplified by domain shift between images of the train and the test sets. One way to quantify this shift is by computing the Fréchet distance between the distributions of features extracted from the train and test sets [47]. For the Virginia opossum, this distance between in-situ train and test images is equal to 21.2, whereas it is much higher at 121.7 for the white-tailed antelope squirrel. This indicates significant domain shift for train and test sites of the white-tailed antelope squirrel and explains why this species is modeled worse when incorporating in-situ imagery.

5 Discussion

Our multi-modal deep occupancy modeling approach offers several key advantages. First, we are able to directly represent and benefit from complementary fine-grained habitat details from in-situ imagery. Second, the use of deep features extracted from satellite imagery enables us to learn complex, and potentially previously unknown, relationships between environmental factors beyond traditional covariates and species occupancy directly from the data. Third, we achieve a significant improvement in predictive performance for 12 out of 16 focal species. This success is just the beginning, and we see many opportunities to expand upon and improve our approach in future work.

	X_{env}	X_{sat}	X_{img}	$X_{\text{env}}, X_{\text{sat}}$	$X_{\text{env}}, X_{\text{img}}$	$X_{\text{sat}}, X_{\text{img}}$	$X_{\text{env}}, X_{\text{sat}}, X_{\text{img}}$
American black bear	-0.06	0.33	0.02	0.28	0.01	**0.37**	0.29
Bobcat	0.21	0.13	0.17	0.20	0.25	0.19	**0.26**
Coyote	-0.04	-0.01	-0.02	0.02	**0.06**	0.01	0.05
Douglas' squirrel	-0.04	0.39	0.29	**0.51**	0.33	0.43	0.45
Eastern chipmunk	0.06	0.32	0.44	0.30	0.48	0.47	**0.49**
Eastern cottontail	0.04	0.12	0.19	0.08	0.08	**0.27**	0.21
Eastern fox squirrel	0.52	0.63	0.25	0.60	0.46	**0.67**	0.66
Eastern gray squirrel	0.22	0.46	0.53	0.38	0.55	**0.64**	0.61
Mule deer	-0.52	-0.49	-0.06	-0.23	-0.13	-0.12	**-0.01**
Nine-banded armadillo	0.47	0.71	0.30	0.66	0.55	0.66	**0.72**
Northern raccoon	0.08	0.34	0.25	0.35	0.37	0.40	**0.43**
Red fox	0.37	0.27	-0.01	0.32	0.28	0.24	**0.38**
Virginia opossum	-0.04	0.07	0.39	0.03	0.39	**0.41**	0.41
Western gray squirrel	-0.10	-0.22	0.24	0.03	-0.18	0.24	**0.25**
White-tailed antelope squirrel	**1.01**	0.66	0.03	0.78	0.39	0.41	0.51
White-tailed deer	0.22	0.36	0.23	0.44	0.43	0.38	**0.45**

Table 1: Test-set predictive performance as measured by normalized LPPD for different combinations of modalities across species (higher is better). **Overall, incorporating additional modalities improves predictive performance, with 12 out of 16 species being modeled best using satellite and in-situ imagery.** Entries outperforming the null model, which estimates a constant occupancy probability across all sites, are colored in green, others red. The best-performing model per species is bolded. The white-tailed antelope squirrel is an apparent outlier that is modeled perfectly using environmental variables alone. None of our models outperforms the null model for the mule deer.

5.1 Environmental variable selection

Covariate selection in ecological models is notoriously difficult and oftentimes driven by prior species-specific expertise about which environmental factors are ecologically relevant. Due to the large number of species considered in this work, and in order to make results comparable across species, we relied on general sets of environmental variables that are widespread and well-proven in the species distribution modeling literature [93]. In our experiments, we see significant evidence that a majority of our environmental variables do not add significant predictive power to models already incorporating satellite-derived features (see fig. 3). Therefore, identifying which environmental variables are potentially superfluous and could therefore be removed, and which additional environmental variables provide additive value and should therefore be added, is of particular interest.

5.2 Tradeoff between interpretability and predictive accuracy

Our results underscore the well-known trade-off between interpretability and predictive accuracy. Simple models are easy to interpret but may lack in predictive performance. Complex models tend to be more accurate but are more difficult to interpret.

When the goal is to make claims about causal effects, explainability is invaluable. For example, in a logistic regression, where species occupancy ψ_i at site i follows

$$\log\left(\frac{\psi_i}{1-\psi_i}\right) = \beta_0 + \beta_1 T_i,$$

the coefficient β_1 is commonly read as the direction and strength of the influence of temperature T_i on habitat suitability – an interpretation that can support causal hypotheses when paired with appropriate design and assumptions.

The challenge is that ecological systems rarely obey such simple relationships and narrow sets of environmental variables as predictors. Constraining ourselves to overly simple models to preserve interpretability thus risks misrepresentation of the system. And is it worth understanding a model that is fundamentally wrong?

By contrast, more complex models that rely on stronger non-linearities and broader sets of predictors are capable of more faithfully representing such complex systems, but require substantially more data to fit robustly and avoid overfitting. A clear illustration comes from the field of machine learning. As data volume and computational resources increased, the field pivoted from simple and interpretable models towards deep learning: heavily over-parameterized models that were difficult to interpret but had far improved predictive power in practice [58]. The community was unwilling to give up these gains in predictive power, and thus developed the new field of "Interpretable AI" to help explain these black-box models.

As data collected by autonomous sensors, remote sensing, and citizen science becomes more readily available, the ecological community is experiencing a similar paradigm shift. We argue that complex systems necessitate complex models, but the need to understand and translate these models into ecological insights remains essential. We lay out an initial step in this direction for our occupancy models, by visualizing individual data points, such as satellite and in-situ images, and how they influence occupancy estimates, in appendix C. In this way, complex models might even enable the discovery of candidate causal effects which could then be more rigorously evaluated through additional experimentation and analysis [59].

5.3 Representing environmental features in imagery with deep feature extractors

The presented results are obtained using DINOv2 [69] for extracting features from in-situ imagery and the AlphaEarth Foundations model for extracting features from satellite imagery. We additionally evaluate alternative choices in appendix D. We note that while the features extracted by the majority of these models improve predictive performance by capturing *some* evidently meaningful habitat dimensions, they are not explicitly and solely optimized towards this task. This fact highlights significant potential for future work to adapt these models towards capturing more meaningful fine-grained habitat dimensions by adapting training objectives, training datasets, and potentially fine-tuning model weights explicitly for ecological modeling tasks.

5.4 Alternative modalities and ecological models

In this study, we focus exclusively on occupancy modeling camera trap data. However, there are other ecological models where one could investigate incorporating deep features in addition to or in lieu of traditional covariates. Examples include presence-only SDMs (see section 2.1), N-mixture models (e.g., [81]), distance sampling models (e.g., [50]), and resource selection functions (e.g., [64]).

There are also other ecological data modalities that capture useful fine-grained ecological context alongside species occurrence, such as passive acoustic monitoring [13] or environmental DNA [12]. One can easily imagine taking a similar approach and using deep learning to represent and enable access to the ecological context of these samples. Blank images, the choice we made in this initial work, are an intuitive way to define those features for camera traps, but the use of a single blank image per location is just a starting point. There is additional, valuable environmental context in the temporal changes to that environment over the time period of deployment. The key challenge when expanding our approach, to handle temporal variation or to include additional modalities, is to define a meaningful and informative mechanism to generate a feature embedding that is representative of the micro-habitat, and useful for the modeling task of interest (see section 5.3). It is an interesting avenue of future work to explore how to *optimally* generate these representative features for each modality, and for each of the many possible ecological models beyond occupancy.

5.5 Practical implications

Given the tradeoffs around model complexity and interpretability, should camera trappers, conservation managers, and ecological modelers who use occupancy models leverage deep features as predictors? If the singular goal is to obtain a simple and interpretable model, the answer is probably no. However, if predictive performance or discovering previously unknown relationships between habitat and occupancy is of greater interest, incorporating deep features derived from satellite and in-situ imagery is well supported by our experiments. In fact, most of our focal species are modeled significantly better using such features, and only one (the white-tailed antelope squirrel) is modeled significantly worse. While in-situ data, such as images and other modalities (see section 5.4) can provide valuable micro-habitat context, such data is also expensive to collect from scratch and at scale. Luckily, sensors are decreasing in cost [96, 11, 2], cultural norms around data sharing in ecology are shifting [76, 77], and citizen science platforms such as iNaturalist[3] or Xeno-canto[4] are decentralizing the burden of in-situ data collection [20, 14]. As a result, growing volumes of such data are openly available. Our work suggests an additional, and valuable, opportunity for citizen scientists–ecologists seeking to improve access to fine-grained ecological context in a given area could prioritize "useful" habitat data collection, i.e. via bioblitzes [70, 65] targeting ecological context as opposed to species occurrence. What remains to be seen is *which* data is most useful when seeking to fill gaps and improve our understanding? Clearly, citizen scientists cannot be deployed everywhere on earth, but what we learn through this process about which data is "most valuable" to collect on the ground can inform data prioritization strategies in remote areas, such as the Amazon or remote Arctic.

[3]https://www.inaturalist.org
[4]https://xeno-canto.org

6 Conclusions

We show that deep features derived from satellite and in-situ imagery can be used to augment traditionally used environmental variables in occupancy models and achieve significantly improved predictive performance for the majority of focal species. In general, we find that deep features extracted from satellite imagery are the best single predictor of species occupancy; however, incorporating in-situ imagery improves models even further. To fairly evaluate and prevent data leakage, we introduce a strict evaluation protocol that relies on spatially separate data splitting. We fit models using a new implementation of ecological models in Python, which facilitates integration of modern machine learning approaches into statistical ecological models. Our approach paves the way for deeper integration of deep learning and quantitative ecological modeling, demonstrating that machine learning can enhance rather than replace existing statistical ecology frameworks. By maintaining the interpretability and uncertainty quantification of such frameworks while leveraging the feature extraction capabilities of deep models, we provide a template for future work combining these complementary approaches.

Acknowledgments

This work was supported in part by the MIT-IBM Watson AI Lab, the AI and Biodiversity Change (ABC) Global Center, which is funded by the US National Science Foundation under Award No. 2330423 and Natural Sciences and Engineering Research Council of Canada under Award No. 585136, NSF CAREER Award No. 2441060, and the MIT-Google program for Computing Innovation and Planetary Health. This work draws on research supported in part by the Social Sciences and Humanities Research Council. We thank Bill McShea, Roland Kays, and all other collaborators of Snapshot USA.

References

[1] J. A. Ahumada, E. Fegraus, T. Birch, N. Flores, R. Kays, T. G. O'Brien, J. Palmer, S. Schuttler, J. Y. Zhao, W. Jetz, et al. Wildlife insights: A platform to maximize the potential of camera trap and other passive sensor wildlife data for the planet. *Environmental Conservation*, 47(1): 1–6, 2020.

[2] M. C. Allen, J. L. Lockwood, R. Ibanez, J. D. Butler, J. C. Angle, and B. D. Jaffe. edna offers opportunities for improved biodiversity monitoring within forest carbon markets. *Communications Earth & Environment*, 5(1):801, 2024.

[3] M. P. Austin and K. P. Van Niel. Improving species distribution models for climate change studies: variable selection and scale, 2011.

[4] K. Ayush, B. Uzkent, C. Meng, K. Tanmay, M. Burke, D. Lobell, and S. Ermon. Geography-aware self-supervised learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10181–10190, 2021.

[5] S. Bajaru, S. Pal, M. Prabhu, P. Patel, R. Khot, and D. Apte. A multi-species occupancy modeling approach to access the impacts of land use and land cover on terrestrial vertebrates in the mumbai metropolitan region (mmr), western ghats, india. *PLoS One*, 15(10):e0240989, 2020.

[6] A. W. Bajcz, W. J. Glisson, J. W. Doser, D. J. Larkin, and J. R. Fieberg. A within-lake occupancy model for starry stonewort, nitellopsis obtusa, to support early detection and monitoring. *Scientific reports*, 14(1):2644, 2024.

[7] M. Barbet-Massin, F. Jiguet, C. H. Albert, and W. Thuiller. Selecting pseudo-absences for species distribution models: how, where and how many? *Methods Ecol. Evol.*, 3(2):327–338, Apr. 2012. doi: 10.1111/j.2041-210x.2011.00172.x.

[8] S. Beery, D. Morris, and S. Yang. Efficient pipeline for camera trap image review, 2019.

[9] J. Bradbury, R. Frostig, P. Hawkins, M. J. Johnson, C. Leary, D. Maclaurin, G. Necula, A. Paszke, J. VanderPlas, S. Wanderman-Milne, and Q. Zhang. JAX: composable transformations of Python+NumPy programs, 2018. URL `http://github.com/jax-ml/jax`.

[10] C. F. Brown, M. R. Kazmierski, V. J. Pasquarella, W. J. Rucklidge, M. Samsikova, C. Zhang, E. Shelhamer, E. Lahera, O. Wiles, S. 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.

[11] E. Browning, R. Gibb, P. Glover-Kapfer, and K. E. Jones. Passive acoustic monitoring in ecology and conservation. 2017.

[12] A. Burian, Q. Mauvisseau, M. Bulling, S. Domisch, S. Qian, and M. Sweet. Improving the reliability of edna data interpretation. *Molecular Ecology Resources*, 21(5):1422–1433, 2021.

[13] M. Campos-Cerqueira and T. M. Aide. Improving distribution data of threatened species by combining acoustic monitoring and occupancy modelling. *Methods in Ecology and Evolution*, 7(11):1340–1348, 2016.

[14] C. Carroll, R. E. Furrow, and L. M. Gerhart. Consistency and validity of participatory science data: A comparison of seasonality patterns of northern california and nevada birds across ebird and inaturalist. *Citizen Science: Theory and Practice*, 10(1), 2025.

[15] D. Chen, Y. Xue, S. Chen, D. Fink, and C. Gomes. Deep multi-species embedding. *arXiv preprint arXiv:1609.09353*, 2016.

[16] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PmLR, 2020.

[17] E. Cole, B. Deneu, T. Lorieul, M. Servajean, C. Botella, D. Morris, N. Jojic, P. Bonnet, and A. Joly. The geolifeclef 2020 dataset. *arXiv preprint arXiv:2004.04192*, 2020.

[18] 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. In *ICML*, 2023.

[19] A. A. De Wan, P. J. Sullivan, A. J. Lembo, C. R. Smith, J. C. Maerz, J. P. Lassoie, and M. E. Richmond. Using occupancy models of forest breeding birds to prioritize conservation planning. *Biological Conservation*, 142(5):982–991, 2009.

[20] G. J. Di Cecco, V. Barve, M. W. Belitz, B. J. Stucky, R. P. Guralnick, and A. H. Hurlbert. Observing the observers: How participants contribute data to inaturalist and implications for biodiversity science. *BioScience*, 71(11):1179–1188, 2021.

[21] S. Z. Dobrowski, H. D. Safford, Y. B. Cheng, and S. L. Ustin. Mapping mountain vegetation using species distribution modeling, image-based texture analysis, and object-based classification. *Applied Vegetation Science*, 11(4): 499–508, 2008.

[22] 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*, 10:41–48, 2024.

[23] J. W. Doser, A. O. Finley, M. Kéry, and E. F. Zipkin. spoccupancy: An r package for single-species, multi-species, and integrated spatial occupancy models. *Methods in Ecology and Evolution*, 13(8):1670–1678, 2022.

[24] J. W. Doser, W. Leuenberger, T. S. Sillett, M. T. Hallworth, and E. F. Zipkin. Integrated community occupancy models: A framework to assess occurrence and biodiversity dynamics using multiple data sources. *Methods in Ecology and Evolution*, 13(4):919–932, 2022.

[25] J. Elith and J. R. Leathwick. Species distribution models: ecological explanation and prediction across space and time. *Annual review of ecology, evolution, and systematics*, 40(1): 677–697, 2009.

[26] Z. Feng, C. Atzberger, S. Jaffer, J. Knezevic, S. Sormunen, R. Young, M. C. Lisaius, M. Immitzer, D. A. Coomes, A. Madhavapeddy, et al. Tessera: Temporal embeddings of surface spectra for earth representation and analysis. *arXiv preprint arXiv:2506.20380*, 2025.

[27] M. Fidino, E. W. Lehrer, C. A. Kay, N. T. Yarmey, M. H. Murray, K. Fake, H. C. Adams, and S. B. Magle. Integrated species distribution models reveal spatiotemporal patterns of human–wildlife conflict. *Ecological Applications*, 32(7): e2647, 2022.

[28] K. J. Filz, T. Schmitt, and J. O. Engler. How fine is fine-scale? questioning the use of fine-scale bioclimatic data in species distribution models used for forecasting abundance patterns in butterflies. *European Journal of Entomology*, 110(2), 2013.

[29] I. Fiske and R. Chandler. Unmarked: an r package for fitting hierarchical models of wildlife

occurrence and abundance. *Journal of statistical software*, 43:1–23, 2011.

[30] J. Franklin, F. W. Davis, M. Ikegami, A. D. Syphard, L. E. Flint, A. L. Flint, and L. Hannah. Modeling plant species distributions under future climates: how fine scale do climate projections need to be? *Global change biology*, 19 (2):473–483, 2013.

[31] A. K. Fuller, D. W. Linden, and J. A. Royle. Management decision making for fisher populations informed by occupancy modeling. *The Journal of Wildlife Management*, 80(5):794–802, 2016.

[32] T. Gadot, S. Istrate, H. Kim, D. Morris, S. Beery, T. Birch, and J. Ahumada. To crop or not to crop: Comparing whole-image and cropped classification on a large dataset of camera trap images. *IET Computer Vision*, 18(8): 1193–1208, 2024.

[33] J. A. Gamon, R. Wang, H. Gholizadeh, B. Zutta, P. A. Townsend, and J. Cavender-Bares. Consideration of scale in remote sensing of biodiversity. In *Remote sensing of plant biodiversity*, pages 425–447. Springer International Publishing Cham, 2020.

[34] A. Gelman, J. Hwang, and A. Vehtari. Understanding predictive information criteria for bayesian models. *Statistics and computing*, 24: 997–1016, 2014.

[35] L. E. Gillespie, M. Ruffley, and M. Exposito-Alonso. Deep learning models map rapid plant species changes from citizen science and remote sensing data. *Proceedings of the National Academy of Sciences*, 121(37):e2318296121, 2024.

[36] J. Grinnell. The origin and distribution of the chest-nut-backed chickadee. *The Auk*, 21(3): 364–382, 1904.

[37] W. Gu and R. K. Swihart. Absent or undetected? effects of non-detection of species occurrence on wildlife–habitat models. *Biological conservation*, 116(2):195–203, 2004.

[38] G. Guillera-Arroita, J. J. Lahoz-Monfort, D. I. MacKenzie, B. A. Wintle, and M. A. McCarthy. Ignoring imperfect detection in biological surveys is dangerous: A response to 'fitting and interpreting occupancy models'. *PloS one*, 9(7): e99571, 2014.

[39] G. Guillera-Arroita, J. J. Lahoz-Monfort, J. Elith, A. Gordon, H. Kujala, P. E. Lentini, M. A. McCarthy, R. Tingley, and B. A. Wintle. Is my species distribution model fit for purpose? matching data and models to applications. *Global ecology and biogeography*, 24(3): 276–292, 2015.

[40] A. Guisan, R. Tingley, J. B. Baumgartner, I. Naujokaitis-Lewis, P. R. Sutcliffe, A. I. Tulloch, T. J. Regan, L. Brotons, E. McDonald-Madden, C. Mantyka-Pringle, et al. Predicting species distributions for conservation decisions. *Ecology letters*, 16(12):1424–1435, 2013.

[41] S. Haesen, J. Lenoir, E. Gril, P. De Frenne, J. J. Lembrechts, M. Kopecký, M. Macek, M. Man, J. Wild, and K. Van Meerbeek. Microclimate reveals the true thermal niche of forest plant species. *Ecology Letters*, 26(12):2043–2055, 2023.

[42] B. J. Halstead, P. M. Kleeman, J. P. Rose, and G. M. Fellers. Sierra nevada amphibians demonstrate stable occupancy despite precipitation volatility in the early 21st century. *Frontiers in Ecology and Evolution*, 10:1040114, 2023.

[43] L. Harrell, C. Kaeser-Chen, B. K. Ayan, K. Anderson, M. Conserva, E. Kleeman, M. Neumann, M. Overlan, M. Chapman, and D. Purves. Heterogeneous graph neural networks for species distribution modeling. *arXiv preprint arXiv:2503.11900*, 2025.

[44] G. Harris and S. L. Pimm. Range size and extinction risk in forest birds. *Conservation Biology*, 22(1):163–171, 2008.

[45] T. Haucke, L. Harrell, and S. Beery. Biolith: Bayesian ecological modeling in python, 2025. URL `https://github.com/timmh/biolith`.

[46] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9729–9738, 2020.

[47] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems*, 30, 2017.

[48] R. J. Hijmans, S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 25(15): 1965–1978, 2005.

[49] M. D. Hoffman, A. Gelman, et al. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623, 2014.

[50] E. J. Howe, S. T. Buckland, M.-L. Després-Einspenner, and H. S. Kühl. Distance sampling with camera traps. *Methods in Ecology and Evolution*, 8(11):1558–1565, 2017.

[51] J. Jakubik, S. Roy, C. E. Phillips, P. Fraccaro, D. Godwin, B. Zadrozny, D. Szwarcman, C. Gomes, G. Nyirjesy, B. Edwards, D. Kimura, N. Simumba, L. Chu, S. K. Mukkavilli, D. Lambhate, K. Das, R. Bangalore, D. Oliveira, M. Muszynski, K. Ankur, M. Ramasubramanian, I. Gurung, S. Khallaghi, H. S. Li, M. Cecil, M. Ahmadi, F. Kordi, H. Alemohammad, M. Maskey, R. Ganti, K. Weldemariam, and R. Ramachandran. Foundation Models for Generalist Geospatial Artificial Intelligence. *Preprint Available on arxiv:2310.18660*, Oct. 2023.

[52] N. Jean, S. Wang, A. Samar, G. Azzari, D. Lobell, and S. Ermon. Tile2vec: Unsupervised representation learning for spatially distributed data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 3967–3974, 2019.

[53] L. N. Joseph, S. A. Field, C. Wilcox, and H. P. Possingham. Presence–absence versus abundance data for monitoring threatened species. *Conservation biology*, 20(6):1679–1687, 2006.

[54] L. K. Katsis, T. A. Rhinehart, E. Dorgay, E. E. Sanchez, J. L. Snaddon, C. P. Doncaster, and J. Kitzes. A comparison of statistical methods for deriving occupancy estimates from machine learning outputs. *Scientific Reports*, 15(1):1–13, 2025.

[55] K. F. Kellner, A. W. Parsons, R. Kays, J. J. Millspaugh, and C. T. Rota. A two-species occupancy model with a continuous-time detection process reveals spatial and temporal interactions. *Journal of Agricultural, Biological and Environmental Statistics*, 27(2):321–338, 2022.

[56] K. F. Kellner, A. D. Smith, J. A. Royle, M. Kéry, J. L. Belant, and R. B. Chandler. The unmarked r package: Twelve years of advances in occurrence and abundance modelling in ecology. *Methods in Ecology and Evolution*, 14(6):1408–1415, 2023. doi: https://doi.org/10.1111/2041-210X.14123. URL `https://besjournals.onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.14123`.

[57] A. K. Killion, A. Honda, E. Trout, and N. H. Carter. Integrating spaceborne estimates of structural diversity of habitat into wildlife occupancy models. *Environmental Research Letters*, 18(6):065002, 2023.

[58] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[59] Y. Liu, K. Duffy, J. G. Dy, and A. R. Ganguly. Explainable deep learning for insights in el niño and river flows. *Nature Communications*, 14(1): 339, 2023.

[60] D. I. MacKENZIE. What are the issues with presence-absence data for wildlife managers?

The Journal of Wildlife Management, 69(3):849–860, 2005.

[61] D. I. Mackenzie. Was it there? dealing with imperfect detection for species presence/absence data. *Australian & New Zealand Journal of Statistics*, 47(1):65–74, 2005.

[62] D. I. MacKenzie, J. D. Nichols, G. B. Lachman, S. Droege, J. Andrew Royle, and C. A. Langtimm. Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, 83(8):2248–2255, 2002.

[63] O. Manas, A. Lacoste, X. Giró-i Nieto, D. Vazquez, and P. Rodriguez. Seasonal contrast: Unsupervised pre-training from uncurated remote sensing data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9414–9423, 2021.

[64] B. F. Manly, L. L. McDonald, D. L. Thomas, T. L. McDonald, and W. P. Erickson. *Resource selection by animals: statistical design and analysis for field studies*. Springer, 2002.

[65] S. Meeus, I. Silva-Rocha, T. Adriaens, P. M. Brown, N. Chartosia, B. Claramunt-López, A. F. Martinou, M. J. Pocock, C. Preda, H. E. Roy, et al. More than a bit of fun: the multiple outcomes of a bioblitz. *BioScience*, 73(3):168–181, 2023.

[66] D. J. Mildrexler, M. Zhao, and S. W. Running. A global comparison between station air temperatures and modis land surface temperatures reveals the cooling role of forests. *Journal of Geophysical Research: Biogeosciences*, 116(G3), 2011.

[67] J. Niedballa, R. Sollmann, A. b. Mohamed, J. Bender, and A. Wilting. Defining habitat covariates in camera-trap based occupancy studies. *Scientific reports*, 5(1):17041, 2015.

[68] N. Ocampo-Peñuela, C. N. Jenkins, V. Vijay, B. V. Li, and S. L. Pimm. Incorporating explicit geospatial data shows more species at risk of extinction than the current red list. *Science Advances*, 2(11):e1601367, 2016.

[69] M. Oquab, T. Darcet, T. Moutakanni, H. V. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby, R. Howes, P.-Y. Huang, H. Xu, V. Sharma, S.-W. Li, W. Galuba, M. Rabbat, M. Assran, N. Ballas, G. Synnaeve, I. Misra, H. Jegou, J. Mairal, P. Labatut, A. Joulin, and P. Bojanowski. Dinov2: Learning robust visual features without supervision, 2023.

[70] S. S. Parker, G. B. Pauly, J. Moore, N. S. Fraga, J. J. Knapp, Z. Principe, B. V. Brown, J. M. Randall, B. S. Cohen, and T. A. Wake. Adapting the bioblitz to meet conservation needs. *Conservation biology*, 32(5):1007–1019, 2018.

[71] L. Pautrel, S. Moulherat, O. Gimenez, and M.-P. Etienne. Analysing biodiversity observation data collected in continuous time: Should we use discrete-or continuous-time occupancy models? *Methods in Ecology and Evolution*, 15(5):935–950, 2024.

[72] D. Phan, N. Pradhan, and M. Jankowiak. Composable effects for flexible and accelerated probabilistic programming in numpyro. *arXiv preprint arXiv:1912.11554*, 2019.

[73] S. J. Phillips, R. P. Anderson, and R. E. Schapire. Maximum entropy modeling of species geographic distributions. *Ecological modelling*, 190(3-4):231–259, 2006.

[74] L. Picek, C. Botella, M. Servajean, C. Leblanc, R. Palard, T. Larcher, B. Deneu, D. Marcos, P. Bonnet, and A. Joly. Geoplant: Spatial plant species prediction dataset. *Advances in Neural Information Processing Systems*, 37:126653–126676, 2024.

[75] L. Poggio, L. M. De Sousa, N. H. Batjes, G. B. Heuvelink, B. Kempen, E. Ribeiro, and D. Rossiter. Soilgrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *Soil*, 7(1):217–240, 2021.

[76] S. M. Powers and S. E. Hampton. Open science, reproducibility, and transparency in ecology. *Ecological applications*, 29(1):e01822, 2019.

[77] O. J. Reichman, M. B. Jones, and M. P. Schildhauer. Challenges and opportunities of open data in ecology. *Science*, 331(6018):703–705, 2011.

[78] T. A. Rhinehart, D. Turek, and J. Kitzes. A continuous-score occupancy model that incorporates uncertain machine learning output from autonomous biodiversity surveys. *Methods in Ecology and Evolution*, 13(8):1778–1789, 2022.

[79] D. R. Roberts, V. Bahn, S. Ciuti, M. S. Boyce, J. Elith, G. Guillera-Arroita, S. Hauenstein, J. J. Lahoz-Monfort, B. Schröder, W. Thuiller, et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography*, 40(8):913–929, 2017.

[80] B. Rooney, R. Kays, M. V. Cove, A. Jensen, B. R. Goldstein, C. Pate, P. Castiblanco, M. E. Abell, J. Adley, B. Agenbroad, A. A. Ahlers, P. D. Alexander, D. Allen, M. L. Allen, J. M. Alston, M. Alyetama, T. L. Anderson, R. Andrade, C. Anhalt-Depies, C. L. Appel, L. Armendariz, C. R. Ayers, A. B. Baird, C. Bak, G. Bandler, E. E. Barding, E. G. Barr, C. Baruzzi, K. Bashaw, S. C. Beers, J. L. Belant, E. Bell, J. F. Benson, A. Berg, D. L. Bergman, B. M. Bernhardt, M. A. Bethel, T. Bird, A. B. Bishop, D. A. Bogan, L. Brandt, L. C. Brandt, A. B. Branney, C. Bratton, C. E. Bresnan, J. M. Brooke, E. K. Buchholtz, F. Buderman, A. D. Burnett, E. E. Burns, D. A. Byrd, S. A. Cannella, K. A. Carey, W. A. Carlile, K. L. Carter, B. J. Cassidy, I. Castro-Arellano, S. Cendejas-Zarelli, N. Chatterjee, A. E. Cheeseman, C. Chevalier, M. C. Chitwood, P. Chrysafis, B. A. Collier, D. P. Collins, J. A. Compton, R. Cone, L. M. Conner, B. L. Cook, O. G. Cosby, S. S. Coster, A. P. Crupi, A. K. Darracq, J. M. Davenport, D. Davis, D. R. Davis, M. L. Davis, R. J. Davis, B. A. DeGregorio, A. Deshwal, K. D. Dougherty, A. Drauglis, C. J. Durbin, A. J. Edelman, V. Elder, B. Eller, E. H. Ellington, S. N. Ellis-Felege, C. N. Ellison, J. E. Fantle-Lepczyk, J. J. Farr, Z. J. Farris, S. P. Finnegan, M. C. Fisher-Reid, E. A. Flaherty, G. Franzoi Dri, S. Fritts, J. Fuller, T. Gallo, L. S. Ganoe, C. N. Ganong, R. Garibay, B. D. Gerber, F. D. Gerraty, S. T. Giery, S. M. Gilyot, J. L. Glasscock, B. Goldfarb, L. E. Good, G. Granados, A. M. Green, J. K. Grewal, A. Grusenmeyer, J. M. Guthrie, M. T. Hallett, C. Hansen, L. P. Hansen, C. Hanson, E. J. Harrity, S. C. M. Hasstedt, M. Hebblewhite, D. J. Herrera, A. Holland, B. R. Humphreys, H. D. Island, A. R. Jack, E. P. Johansson, A. M. Johnson, L. Johnson, T. L. Johnstone-Yellin, M. L. S. P. Jorge, W. Kahano, M. A. Kinsey, B. E. Klossing, T. W. Knowles, M. M. Koeck, J. L. Koprowski, K. M. Kuhn, E. K. Kuprewicz, D. J. R. Lafferty, J. A. Lamberton-Moreno, T. J. Land, A. M. Langston, S. LaPoint, E. N. Largent, M. A. Lashley, R. G. Lathrop, T. E. Lee Jr, C. A. Lepczyk, D. B. Lesmeister, C. Leung, J. V. Lombardi, R. Long, R. C. Lonsinger, I. Lord, S. S. Madere, S. P. Maher, J. A. Mallinoff, A. Martinez, D. S. Mason, H. A. Mathewson, A. E. Mayer, K. P. McCarthy, S. F. McCracken, B. McDonald, B. McGarry, S. T. McMurry, L. E. McTigue, B. M. Mena, M. Mercer, M. R. Merz, S. Millar, G. D. Miller, J. J. Millspaugh, R. J. Moll, T. W. Mong, J. D. Monzón, J. C. Moore, A. Mortelliti, K. W. Mote, K. Mullen, A. Mychajliw, C. Nagy, S. A. Neiswenter, D. R. Neyland, L. P. Nicholson, M. T. O'Mara, B. J. O'Neill, E. A. Olson, M. J. Orgill, G. Palomo-Munoz, S. M. Parsons, L. E. Patrick, J. R. Patterson, D. L. Pearce, M. E. Pendergast, B. S. Perla, T. R. Petroelje, H. Pliske, M. K. P. Poisson, M. R. Price, M. D. Proctor, N. J. Proudman, J. L. Rachlow, R. E. Ramos, M. Reabold, J. Redinger, A. E. Reed, C. C. Rega-Brodsky, E. Rehm, K. R. Remine, M. S. Rentz, E. Ridder, D. R. Risch, L. L. Robbins, J. P. Roemer, A. Romero, C. Rota, C. M. Schalk, B. D. Scholten, C. L. Scott, B. M. Scurlock, M. Sergeyev, W. J. Severud, J. Sevin, H. Shamon, C. Sharp, M. Shaw, V. Siverls-Dunham, A. B. Smith, D. S. Smith, M. H.

Snider, D. A. Sossover, A. R. Sovie, J. A. Sparks, J. Speiser, M. T. Springer, J. L. Spurlin, E. A. Steinkamp, J. L. Stenglein, J. Stewart Kloker, C. M. Stitzman, M. Stokes, K. Stringer, J. Stutzman, D. S. Sullins, C. Sullivan, N. B. Sullivan, E. P. Tanner, A. M. Tanner, E. B. Thornock, J. Titus, J. M. Tleimat, K. Toomey, L. T. Toussaint, M. Uribe, M. Van der Merwe, D. J. Van Parys, J. P. Vanek, J. Varner, B. V. Walker, C. Wallace, D. Ward, B. H. Warner, D. T. Warren, J. R. Wasdin, S. L. Webb, K. L. Wehr, N. H. Wehr, E. G. Weigel, T. J. Werdel, L. S. Whipple, C. A. Whittier, C. Wiersema, A. M. Wilson, M. F. H. Wilson, A. J. Wolf, J. P. Wolford, D. W. Wolfson, D. J. Woolsey, M. A. Wuensch, G. Xu, K. L. Yurewicz, V. Zancho, M. Zimova, A. Zorn, and W. J. McShea. Snapshot usa 2019–2023: The first five years of data from a coordinated camera trap survey of the united states. *Global Ecology and Biogeography*, 34(1): e13941, 2025. doi: https://doi.org/10.1111/geb.13941. URL `https://onlinelibrary.wiley.com/doi/abs/10.1111/geb.13941`. e13941 GEB-2024-0441.R1.

[81] J. A. Royle. N-mixture models for estimating population size from spatially replicated counts. *Biometrics*, 60(1):108–115, 2004.

[82] J. A. Royle and W. A. Link. Generalized site occupancy models allowing for false positive and false negative errors. *Ecology*, 87(4):835–841, 2006.

[83] J. A. Royle and J. D. Nichols. Estimating abundance from repeated presence–absence data or point counts. *Ecology*, 84(3):777–790, 2003.

[84] R. A. Santos, M. Mota-Ferreira, L. M. Aguiar, and F. Ascensão. Predicting wildlife road-crossing probability from roadkill data using occupancy-detection models. *Science of the Total Environment*, 642:629–637, 2018.

[85] S. Sastry, S. Khanal, A. Dhakal, A. Ahmad, and N. Jacobs. Taxabind: A unified embedding space for ecological applications. In *Winter Conference on Applications of Computer Vision*. IEEE/CVF, 2025.

[86] M. P. Scroggie, K. Preece, E. Nicholson, M. A. McCarthy, K. M. Parris, and G. W. Heard. Optimizing habitat management for amphibians: From simple models to complex decisions. *Biological Conservation*, 236:60–69, 2019.

[87] O. Siméoni, H. V. Vo, M. Seitzer, F. Baldassarre, M. Oquab, C. Jose, V. Khalidov, M. Szafraniec, S. Yi, M. Ramamonjisoa, F. Massa, D. Haziza, L. Wehrstedt, J. Wang, T. Darcet, T. Moutakanni, L. Sentana, C. Roberts, A. Vedaldi, J. Tolan, J. Brandt, C. Couprie, J. Mairal, H. Jégou, P. Labatut, and P. Bojanowski. DINOv3, 2025. URL `https://arxiv.org/abs/2508.10104`.

[88] J. B. Smith, J. A. Jenks, and R. W. Klaver. Evaluating detection probabilities for american marten in the black hills, south dakota. *The Journal of Wildlife Management*, 71(7):2412–2416, 2007.

[89] H. R. Sofaer, J. A. Hoeting, and C. S. Jarnevich. The area under the precision-recall curve as a performance metric for rare binary events. *Methods in Ecology and Evolution*, 10(4):565–577, 2019.

[90] S. Stevens, J. Wu, M. J. Thompson, E. G. Campolongo, C. H. Song, D. E. Carlyn, L. Dong, W. M. Dahdul, C. Stewart, T. Berger-Wolf, W.-L. Chao, and Y. Su. BioCLIP: A vision foundation model for the tree of life. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 19412–19424, 2024.

[91] D. Szwarcman, S. Roy, P. Fraccaro, Þ. E. Gíslason, B. Blumenstiel, R. Ghosal, P. H. de Oliveira, J. L. d. S. Almeida, R. Sedona, Y. Kang, et al. Prithvi-eo-2.0: A versatile multi-temporal foundation model for earth observation applications. *arXiv preprint arXiv:2412.02732*, 2024.

[92] L. Tang, Y. Xue, D. Chen, and C. Gomes. Multi-entity dependence learning with rich context via conditional variational auto-encoder. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

[93] M. Teng, A. Elmustafa, B. Akera, Y. Bengio, H. Radi, H. Larochelle, and D. Rolnick. Satbird: a dataset for bird species distribution modeling using remote sensing and citizen science data. In *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2023. URL `https://openreview.net/forum?id=Vn5qZGxGj3`.

[94] R. Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 58 (1):267–288, 1996.

[95] L. A. Tigas, D. H. Van Vuren, and R. M. Sauvajot. Behavioral responses of bobcats and coyotes to habitat fragmentation and corridors in an urban environment. *Biological Conservation*, 108 (3):299–306, 2002.

[96] M. W. Tobler, S. E. Carrillo-Percastegui, R. Leite Pitman, R. Mares, and G. Powell. An evaluation of camera traps for inventorying large-and medium-sized terrestrial rainforest mammals. *Animal conservation*, 11(3):169–178, 2008.

[97] G. Tseng, R. Cartuyvels, I. Zvonkov, M. Purohit, D. Rolnick, and H. Kerner. Lightweight, pre-trained transformers for remote sensing timeseries. *arXiv preprint arXiv:2304.14065*, 2023.

[98] G. Tseng, A. Fuller, M. Reil, H. Herzog, P. Beukema, F. Bastani, J. R. Green, E. Shelhamer, H. Kerner, and D. Rolnick. Galileo: Learning global and local features in pretrained remote sensing models, 2025. URL `https://arxiv.org/abs/2502.09356`.

[99] R. Valavi, G. Guillera-Arroita, J. J. Lahoz-Monfort, and J. Elith. Predictive performance of presence-only species distribution models: a benchmark study with reproducible code. *Ecological monographs*, 92(1):e01486, 2022.

[100] D. L. Warren and S. N. Seifert. Ecological niche modeling in maxent: the importance of model complexity and the performance of model selection criteria. *Ecological applications*, 21(2): 335–342, 2011.

[101] R. Zbinden, N. van Tiel, B. Kellenberger, L. Hughes, and D. Tuia. On the selection and effectiveness of pseudo-absences for species distribution modeling with deep learning. *Ecological Informatics*, 81:102623, July 2024. ISSN 1574-9541. doi: 10.1016/j.ecoinf.2024.102623. URL `http://dx.doi.org/10.1016/j.ecoinf.2024.102623`.

[102] M. D. Zeiler and R. Fergus. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pages 818–833. Springer, 2014.

[103] K. A. Zeller, S. Nijhawan, R. Salom-Pérez, S. H. Potosme, and J. E. Hines. Integrating occupancy modeling and interview data for corridor identification: a case study for jaguars in nicaragua. *Biological Conservation*, 144(2):892–901, 2011.

[104] F. Zellweger, E. Sulmoni, J. T. Malle, A. Baltensweiler, T. Jonas, N. E. Zimmermann, C. Ginzler, D. N. Karger, P. De Frenne, D. Frey, et al. Microclimate mapping using novel radiative transfer modelling. *Biogeosciences*, 21(2): 605–623, 2024.

Wildlife Insights data references

[1] A One Tam project managed collaboratively by the National Park Service, California State Parks, Marin Water, Marin County Parks and Golden Gate National Parks Conservancy. Last updated august 2025. `http://n2t.net/ark:/63614/w12001631`, 2017. Accessed via wildlifeinsights.org on September 1, 2025.

[2] A One Tam project managed collaboratively by the National Park Service, California State Parks, Marin Water, Marin County Parks and Golden Gate National Parks Conservancy. Last updated july 2025. http://n2t.net/ark:/63614/w12001634, 2017. Accessed via wildlifeinsights.org on September 1, 2025.

[3] A One Tam project managed collaboratively by the National Park Service, California State Parks, Marin Water, Marin County Parks and Golden Gate National Parks Conservancy. Last updated august 2025. http://n2t.net/ark:/63614/w12001639, 2017. Accessed via wildlifeinsights.org on September 1, 2025.

[4] A One Tam project managed collaboratively by the National Park Service, California State Parks, Marin Water, Marin County Parks and Golden Gate National Parks Conservancy. Last updated august 2025. http://n2t.net/ark:/63614/w12001648, 2017. Accessed via wildlifeinsights.org on September 1, 2025.

[5] Aiello, C. M., Galloway, N. L., Fratella, K., Prentice, P. R., Darby, N. W., Hughson D., Epps, C. W. Last updated november 2024. http://n2t.net/ark:/63614/w12003007, 2018. Accessed via wildlifeinsights.org on September 1, 2025.

[6] Bresnan, C. Last updated november 2024. http://n2t.net/ark:/63614/w12006261, 2023. Accessed via wildlifeinsights.org on September 1, 2025.

[7] Bresnan, C., Shamon, H. Last updated september 2024. http://n2t.net/ark:/63614/w12004234, 2022. Accessed via wildlifeinsights.org on September 1, 2025.

[8] Casady, D. Last updated march 2025. http://n2t.net/ark:/63614/w12002889, 2016. Accessed via wildlifeinsights.org on September 1, 2025.

[9] Casady, D. Last updated may 2025. http://n2t.net/ark:/63614/w12002877, 2017. Accessed via wildlifeinsights.org on September 1, 2025.

[10] Casady, D. Last updated march 2025. http://n2t.net/ark:/63614/w12002923, 2017. Accessed via wildlifeinsights.org on September 1, 2025.

[11] Casady, D. Last updated january 2025. http://n2t.net/ark:/63614/w12002901, 2019. Accessed via wildlifeinsights.org on September 1, 2025.

[12] Critter Camera Data. http://n2t.net/ark:/63614/w12001313. Accessed via wildlifeinsights.org on September 1, 2025.

[13] Davis, J. Last updated august 2025. http://n2t.net/ark:/63614/w12004572, 2016. Accessed via wildlifeinsights.org on September 1, 2025.

[14] Dorcy, J., Rich, L., Furnas, B., Dorcy-Ponce, J. Last updated march 2025. http://n2t.net/ark:/63614/w12002828, 2017. Accessed via wildlifeinsights.org on September 1, 2025.

[15] Flores, N., singh, a., Rich, L., Cook, V., Vollmer, J., Whetzel, M. Last updated july 2025. http://n2t.net/ark:/63614/w12004221, 2019. Accessed via wildlifeinsights.org on September 1, 2025.

[16] Flores, N., Smithsonian, e., Kays, R. Last updated march 2024. http://n2t.net/ark:/63614/w12003860, 2008. Accessed via wildlifeinsights.org on September 1, 2025.

[17] Flores, N., Smithsonian, e., Kays, R., Chrysafis, P. Last updated april 2023. http://n2t.net/ark:/63614/w12003854, 2011. Accessed via wildlifeinsights.org on September 1, 2025.

[18] Flores, N., Smithsonian, e., McShea, W., Shamon, H. Last updated august 2025. http://n2t.net/ark:/63614/w12006132, 2013. Accessed via wildlifeinsights.org on September 1, 2025.

[19] Forrester, T. . Last updated january 2023. http://n2t.net/ark:/63614/w12004287, 2011. Accessed via wildlifeinsights.org on September 1, 2025.

[20] Forrester, T. . Last updated april 2023. http://n2t.net/ark:/63614/w12004316, 2011. Accessed via wildlifeinsights.org on September 1, 2025.

[21] Forrester, T. Last updated august 2025. http://n2t.net/ark:/63614/w12004302, 2000. Accessed via wildlifeinsights.org on September 1, 2025.

[22] Green, A. Last updated june 2025. http://n2t.net/ark:/63614/w12003225, 2020. Accessed via wildlifeinsights.org on September 1, 2025.

[23] C. P. Hansen, A. W. Parsons, R. Kays, and J. J. Millspaugh. Does use of backyard resources explain the abundance of urban wildlife? *Frontiers in Ecology and Evolution*, 8:570771, 2020. URL http://n2t.net/ark:/63614/w12004526. Accessed via wildlifeinsights.org on September 1, 2025.

[24] Hatfield, B., Rich, L., Malleshappa, V. Last updated august 2025. http://n2t.net/ark:/63614/w12002818, 2020. Accessed via wildlifeinsights.org on September 1, 2025.

[25] Higdon, S. Last updated april 2023. http://n2t.net/ark:/63614/w12004592, 2000. Accessed via wildlifeinsights.org on September 1, 2025.

[26] Jachowski, D.S.; Katzner, T.E.; Harris, S.N.; Marneweck, C.J. Last updated august 2023. http://n2t.net/ark:/63614/w12001570, 2007. Accessed via wildlifeinsights.org on September 1, 2025.

[27] Jachowski, D.S.; Katzner, T.E.; Harris, S.N.; Marneweck, C.J. Last updated july 2024. http://n2t.net/ark:/63614/w12003589, 2010. Accessed via wildlifeinsights.org on September 1, 2025.

[28] Jansen, P. Last updated february 2023. http://n2t.net/ark:/63614/w12004587, 2015. Accessed via wildlifeinsights.org on September 1, 2025.

[29] Johnson, Aaron P., Davis, Allegra G., Garcia, Diego G., Tobio, Ana, Wesling, Lauren A., Dibble, Josh. . Last updated july 2025. http://n2t.net/ark:/63614/w12004413, 2015. Accessed via wildlifeinsights.org on September 1, 2025.

[30] Jones, A. Last updated september 2024. http://n2t.net/ark:/63614/w12003849, 2016. Accessed via wildlifeinsights.org on September 1, 2025.

[31] Kalies, E.L. Last updated november 2022. http://n2t.net/ark:/63614/w12004569, 2016. Accessed via wildlifeinsights.org on September 1, 2025.

[32] Kays, R. . Last updated june 2025. http://n2t.net/ark:/63614/w12004713, 2000. Accessed via wildlifeinsights.org on September 1, 2025.

[33] Kays, R. . Last updated january 2023. http://n2t.net/ark:/63614/w12004780, 2011. Accessed via wildlifeinsights.org on September 1, 2025.

[34] Kays, R. . Last updated september 2024. http://n2t.net/ark:/63614/w12004748, 2012. Accessed via wildlifeinsights.org on September 1, 2025.

[35] Kays, R. . Last updated june 2025. http://n2t.net/ark:/63614/w12004318, 2015. Accessed via wildlifeinsights.org on September 1, 2025.

[36] Kays, R. . Last updated march 2025. http://n2t.net/ark:/63614/w12004782, 2017. Accessed via wildlifeinsights.org on September 1, 2025.

[37] Kays, R., Levin, M., Kalies, L. Last updated february 2023. http://n2t.net/ark:/63614/w12004578, 2022. Accessed via wildlifeinsights.org on September 1, 2025.

[38] Kays, R., Support, W., Ahumada, J., Alyetama, M. Last updated august 2025. http://n2t.net/ark:/63614/w12003602, 2021. Accessed via wildlifeinsights.org on September 1, 2025.

[39] Kays, Roland; Hess, George . Last updated august 2025. http://n2t.net/ark:/63614/w12004290, 2000. Accessed via wildlifeinsights.org on September 1, 2025.

[40] kumar, s., Stutzman, J. Last updated may 2024. http://n2t.net/ark:/63614/w12003608, 2021. Accessed via wildlifeinsights.org on September 1, 2025.

[41] LaPoint, S. . Last updated november 2022. http://n2t.net/ark:/63614/w12004076, 2019. Accessed via wildlifeinsights.org on September 1, 2025.

[42] Long, R. . Last updated january 2024. http://n2t.net/ark:/63614/w12004625, 2000. Accessed via wildlifeinsights.org on September 1, 2025.

[43] M. Lasky, et al. CAROLINA CRITTERS: a collection of camera trap data from wildlife surveys across North Carolina. Ecology. e03372 (2021). Last updated december 2023. http://n2t.net/ark:/63614/w12004307, 2000. Accessed via wildlifeinsights.org on September 1, 2025.

[44] Malleshappa, V., Flores, N., Rich, L., Found-Jackson, C., Carrothers, D., Anderson, S. Last updated march 2025. http://n2t.net/ark:/63614/w12003821, 2019. Accessed via wildlifeinsights.org on September 1, 2025.

[45] Malleshappa, V., Smithsonian, e., Kays, R. Last updated july 2023. http://n2t.net/ark:/63614/w12004600, 2000. Accessed via wildlifeinsights.org on September 1, 2025.

[46] Malleshappa, V., Smithsonian, e., McShea, W., Kays, R. Last updated june 2023. http://n2t.net/ark:/63614/w12004286, 2013. Accessed via wildlifeinsights.org on September 1, 2025.

[47] Malleshappa, V., Smithsonian, e., McShea, W., Schuttler, S., Kays, R. Last updated april 2023. http://n2t.net/ark:/63614/w12004293, 2014. Accessed via wildlifeinsights.org on September 1, 2025.

[48] Malleshappa, V., Smithsonian, e., Rooney, B., Kays, R. Last updated july 2025. http://n2t.net/ark:/63614/w12004090, 2012. Accessed via wildlifeinsights.org on September 1, 2025.

[49] Malleshappa, V., Smithsonian, e., Schuttler, S., Kays, R. Last updated december 2023. http://n2t.net/ark:/63614/w12004311, 2014. Accessed via wildlifeinsights.org on September 1, 2025.

[50] Martin, Marie E., Green, David S., and Matthews, Sean M. http://n2t.net/ark:/63614/w12001294. Accessed via wildlifeinsights.org on September 1, 2025.

[51] S. McMurry, A. W. Parsons, M. Lasky, H. Boone, R. Kays, M. Snider, E. Kays, J. Luongo, J. Clark, C. L. Scher, and W. J. McShea. Standardized camera trapping data from National Ecological Observatory Network (NEON) sites and other forest plots. Ecology, in review. Accessed via wildlifeinsights.org on September 1, 2025. URL http://n2t.net/ark:/63614/w12004578.

[52] McMurry, S., Kays, R., Alyetama, M., Patel, L. Last updated june 2025. http://n2t.net/ark:/63614/w12006449, 2023. Accessed via wildlifeinsights.org on September 1, 2025.

[53] McMurry, S., Kays, R., Spurlin, J., Martin, G., Frech, G., Barajas-Salazar, K., Snider, M. Last updated october 2023. http://n2t.net/ark:/63614/w12004251, 2022. Accessed via wildlifeinsights.org on September 1, 2025.

[54] McShea, W. . Last updated june 2025. http://n2t.net/ark:/63614/w12004765, 1999. Accessed via wildlifeinsights.org on September 1, 2025.

[55] McShea, W. . Last updated august 2025. http://n2t.net/ark:/63614/w12004781, 2000. Accessed via wildlifeinsights.org on September 1, 2025.

[56] Mueldener, M. Last updated april 2023. http://n2t.net/ark:/63614/w12004588, 2017. Accessed via wildlifeinsights.org on September 1, 2025.

[57] Myers, J. . Last updated april 2023. http://n2t.net/ark:/63614/w12004295, 2014. Accessed via wildlifeinsights.org on September 1, 2025.

[58] Nelson, D. L. Shamon, H. Jachowski, D. S. Last updated april 2025. http://n2t.net/ark:/63614/w12003597, 2021. Accessed via wildlifeinsights.org on September 1, 2025.

[59] Nelson, D. L. Shamon, H. Jachowski, D. S. Last updated november 2024. http://n2t.net/ark:/63614/w12004526, 2022. Accessed via wildlifeinsights.org on September 1, 2025.

[60] Nelson, Dana L. . Last updated july 2025. http://n2t.net/ark:/63614/w12002979, 2021. Accessed via wildlifeinsights.org on September 1, 2025.

[61] Rich, L. Last updated april 2025. http://n2t.net/ark:/63614/w12003401, 2012. Accessed via wildlifeinsights.org on September 1, 2025.

[62] Rich, L. N., Toenies, M., Cornelius, N., Boynton, M., Chappell, E. Last updated august 2025. http://n2t.net/ark:/63614/w12001650, 2022. Accessed via wildlifeinsights.org on September 1, 2025.

[63] Rooney, Brigit; McShea, William. Last updated august 2025. http://n2t.net/ark:/63614/w12006220, 2023. Accessed via wildlifeinsights.org on September 1, 2025.

[64] Rota, C. . Last updated november 2023. http://n2t.net/ark:/63614/w12003869, 2016. Accessed via wildlifeinsights.org on September 1, 2025.

[65] Rota, C. Last updated december 2023. http://n2t.net/ark:/63614/w12003912, 2016. Accessed via wildlifeinsights.org on September 1, 2025.

[66] Samsoe, E. Last updated october 2023. http://n2t.net/ark:/63614/w12003109, 2007. Accessed via wildlifeinsights.org on September 1, 2025.

[67] Sargis, E. , Widness, J. Last updated september 2024. http://n2t.net/ark:/63614/w12003858, 2017. Accessed via wildlifeinsights.org on September 1, 2025.

[68] Sevin, J. Last updated august 2025. http://n2t.net/ark:/63614/w12004615, 2020. Accessed via wildlifeinsights.org on September 1, 2025.

[69] Shamon, H. Last updated may 2023. http://n2t.net/ark:/63614/w12003074, 2021. Accessed via wildlifeinsights.org on September 1, 2025.

[70] Shamon, H. Last updated july 2023. http://n2t.net/ark:/63614/w12006229, 2023. Accessed via wildlifeinsights.org on September 1, 2025.

[71] singh, a., Malleshappa, V., Rich, L., Cook, V., Vollmer, J., Whetzel, M. Last updated july 2025. http://n2t.net/ark:/63614/w12003570, 2020. Accessed via wildlifeinsights.org on September 1, 2025.

[72] Sink, C. Last updated june 2023. http://n2t.net/ark:/63614/w12002779, 2020. Accessed via wildlifeinsights.org on September 1, 2025.

[73] Smithsonian, e. Last updated april 2023. http://n2t.net/ark:/63614/w12002684, 2020. Accessed via wildlifeinsights.org on September 1, 2025.

[74] Smithsonian, e. Last updated april 2023. http://n2t.net/ark:/63614/w12002685, 2020. Accessed via wildlifeinsights.org on September 1, 2025.

[75] Smithsonian, e., Farkas, D. Last updated december 2023. http://n2t.net/ark:/63614/w12004582, 2011. Accessed via wildlifeinsights.org on September 1, 2025.

[76] Smithsonian, e., McShea, W. Last updated march 2025. http://n2t.net/ark:/63614/w12003833, 2007. Accessed via wildlifeinsights.org on September 1, 2025.

[77] Smithsonian, e., McShea, W. Last updated april 2023. http://n2t.net/ark:/63614/w12004583, 2017. Accessed via wildlifeinsights.org on September 1, 2025.

[78] Smithsonian, e., McShea, W., Shamon, H. Last updated june 2025. http://n2t.net/ark:/63614/w12006140, 2012. Accessed via wildlifeinsights.org on September 1, 2025.

[79] Tucker, J., Support, B., Heiman, J., Heiman, J., Eyes, S. Last updated july 2025. http://n2t.net/ark:/63614/w12004223, 2021. Accessed via wildlifeinsights.org on September 1, 2025.

[80] USDA Forest Service, Pacific Southwest Region, Carnivore Monitoring Program. http://n2t.net/ark:/63614/w12000517. Accessed via wildlifeinsights.org on September 1, 2025.

[81] VanNiel, J. Last updated april 2023. http://n2t.net/ark:/63614/w12003830, 2017. Accessed via wildlifeinsights.org on September 1, 2025.

[82] Wainstein, M. Last updated april 2023. http://n2t.net/ark:/63614/w12003862, 2014. Accessed via wildlifeinsights.org on September 1, 2025.

[83] Woodrow, A. Last updated october 2022. http://n2t.net/ark:/63614/w12003864, 2011. Accessed via wildlifeinsights.org on September 1, 2025.

A Naive occupancy experiments

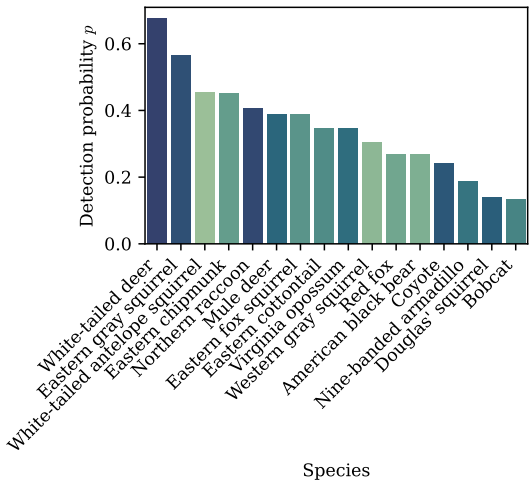

Figure 5: Posterior mean detection probability across species. **Detection probabilities are relatively high for many species, indicating that non-hierarchical logistic regression might be sufficient in some cases.** Blue colors represent more prevalent, green colors less prevalent species (see section 3.4).

In addition to fitting full hierarchical occupancy models, we can fit a simpler class of models we call *naive* occupancy models. Instead of presence and non-detection observations, these models predict a single binary *occupied* or *unoccupied* class for each site. We say a site i is naively occupied ($y_i^{\text{NO}} = 1$) if there is at least one observation of the focal species at the site, otherwise we say the site is naively unoccupied ($y_i^{\text{NO}} = 0$).

While this simplified formulation does not properly account for imperfect detection, this might be of less concern for highly detectable species. Figure 5 shows that detection probability estimates for some of our focal species are relatively high. This simplified formulation then allows us to use any binary classification model such as logistic regression that can be optimized without having to rely on computationally expensive Markov-chain Monte Carlo. Additionally, naive occupancy can be evaluated using intuitive binary classification metrics such as average precision, also called PR-AUC. PR-AUC is the area under the precision-recall curve plotted over different confidence thresholds. It is popular for evaluating species distributions models, as it is robust to class imbalance, as, for example, induced by rare species [89]. Given presence-absence confidence scores s_i produced by the naive occupancy classifier for site i and a set of confidence thresholds t_n, recall and precision (sometimes *specificity*) are computed for each of those thresholds as:

$$P_n = \frac{\text{TP}_n}{\text{TP}_n + \text{FP}_n} = \frac{|\{i \mid y_i^{NO} = 1 \wedge s_i \geq t_n\}|}{|\{i \mid s_i \geq t_n\}|} \quad (6)$$

$$R_n = \frac{\text{TP}_n}{\text{TP}_n + \text{FN}_n} = \frac{|\{i \mid y_i^{NO} = 1 \wedge s_i \geq t_n\}|}{|\{i \mid y_i^{NO} = 1\}|} \quad (7)$$

We then compute the approximate area under the precision recall curve as:

$$\text{AP} = \sum_n (R_n - R_{n-1}) P_n \quad (8)$$

Figure 6 shows the improvement in PR-AUC of the best-performing in-situ image model over the best-performing non in-situ image model and fig. 7 shows the same for satellite image features. Overall, we observe the same trends between naive and full occupancy models. Species benefiting more strongly from in-situ and satellite imagery in full occupancy models benefit similarly in naive occupancy models, and vice-versa.

This shows that, despite their inherent limitations, naive occupancy models might be beneficial for selecting predictor variables

B Environmental variables

Table 2 shows the full set of environmental variables we use as occupancy covariates.

C Qualitative interpretability

Although the features themselves generated by deep models can be difficult to interpret, we can still infer

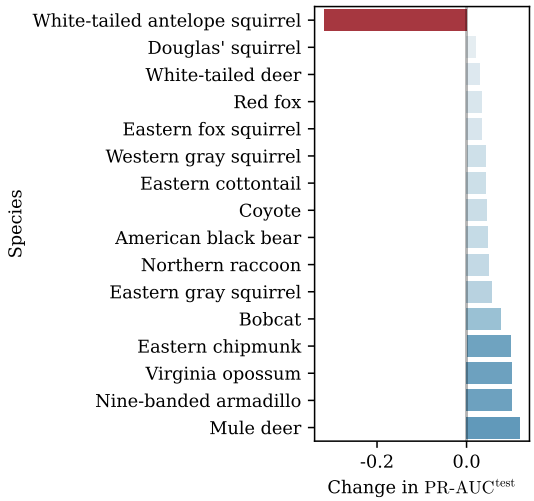

Figure 6: Test-set PR-AUC improvement of the best-performing model including in-situ image features over the best-performing model not including in-situ image features.

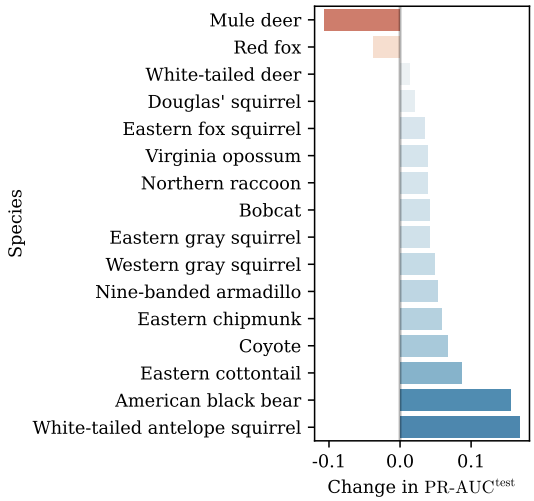

Figure 7: Test-set PR-AUC improvement of the best-performing model including satellite image features over the best-performing model not including satellite image features.

Name	Description	Unit
BIO01	Annual mean temperature	°C
BIO02	Mean diurnal range (mean monthly $\langle T_{\max} - T_{\min} \rangle$)	°C
BIO03	Isothermality (BIO02 / BIO07 ×100)	%
BIO04	Temperature seasonality (SD ×100)	°C
BIO05	Max. temperature of warmest month	°C
BIO06	Min. temperature of coldest month	°C
BIO07	Temperature annual range (BIO05 – BIO06)	°C
BIO08	Mean temperature of wettest quarter	°C
BIO09	Mean temperature of driest quarter	°C
BIO10	Mean temperature of warmest quarter	°C
BIO11	Mean temperature of coldest quarter	°C
BIO12	Annual precipitation	mm
BIO13	Precipitation of wettest month	mm
BIO14	Precipitation of driest month	mm
BIO15	Precipitation seasonality	Coefficient of Variation
BIO16	Precipitation of wettest quarter	mm
BIO17	Precipitation of driest quarter	mm
BIO18	Precipitation of warmest quarter	mm
BIO19	Precipitation of coldest quarter	mm
orcdrc	Soil organic carbon density (15–30 cm)	$\mathrm{dg\,dm^{-3}}$
phihox	Soil pH in water (pH×10) at 15–30 cm	—
cecsol	Cation exchange capacity at pH 7 (15–30 cm)	$\mathrm{mmol\,kg^{-1}}$
clyppt	Clay content (15–30 cm)	$\mathrm{g\,kg^{-1}}$
sltppt	Silt content (15–30 cm)	$\mathrm{g\,kg^{-1}}$
sndppt	Sand content (15–30 cm)	$\mathrm{g\,kg^{-1}}$
bdod	Bulk density (15–30 cm)	—

Table 2: The full set of environmental variables used for our experiments, consisting of WorldClim BIO [48] and SoilGrids [75] variables sampled via Google Earth Engine. We adapted this set of environmental variables from SatBird [93] but excluded the SoilGrids "absolute depth to bedrock in cm" since it is not readily available on Google Earth Engine.

how and to what magnitude the underlying individual samples such as in-situ images influence the occupancy estimate. Given linear model occupancy probability coefficients β, we isolate those concerned with in-situ images, β_{img}. We then find in-situ image features X_{img} that maximize or minimize the term $\beta_{\text{img}} X_{\text{img}}$, which finds those features that maximize or minimize the model's estimate of occupancy probability ψ. We show the top 3 (maximizing) and bottom 3 (minimizing) examples for different species in fig. 8.

D Alternative deep feature extractors

In addition to DINOv2 [69] for in-situ-images and the AlphaEarth Foundations model [10], we evaluate the effectiveness of alternative deep feature extractors. For in-situ imagery, we additionally evaluate DINOv3

Figure 8: **To help interpret the types of relationships our deep multi-modal occupancy models learn between in-situ imagery and species occupancy, we visualize in-situ images that contribute maximally to an occupancy (maximizing ψ, top) or non-occupancy (minimizing ψ, bottom) estimate.** These examples qualitatively demonstrate that our models learn ecologically relevant patterns. For example, the model seems to associate unpaved roads with Coyote occupancy. Indeed, Coyotes are known to specifically use roads for movement [95]. Conversely, the model has learned that Coyotes are unlikely to occupy habitats close to human infrastructure.

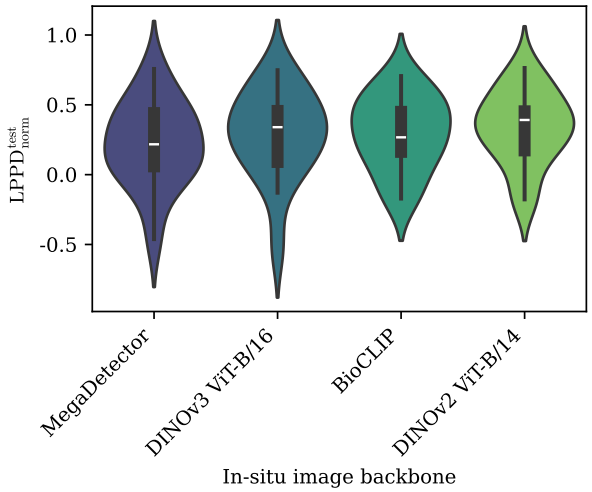

Figure 9: Predictive performance of various in-situ image backbones across focal species. We evaluate backbones using the full set of environmental, satellite (using AlphaEarth Foundations) and in-situ imagery. Although none of these models are explicitly trained to perform fine-grained habitat characterization, they evidently capture varying degrees of useful information.

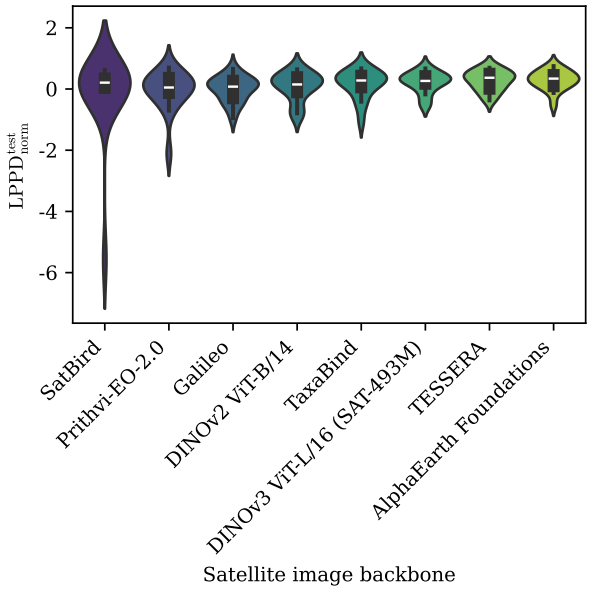

Figure 10: Predictive performance of various satellite image backbones across focal species. We evaluate backbones using the full set of environmental, satellite and in-situ imagery (using DINOv2).

[87], BioCLIP [90] (a model trained to recognize over 450k diverse taxonomic labels), and embeddings extracted from MegaDetector [8]. We note that despite using blank camera trap images as input, these models might capture useful prior information based on the scene's background. Figure 9 shows macro-averaged normalized LPPD for those different models.

For satellite imagery, in addition to AlphaEarth Foundations [10], we evaluate Prithvi-EO-2.0 [91], SatBird [93], TaxaBind [85], Galileo [98], and TESSERA [26]. We furthermore evaluate DINOv3 pre-trained on satellite imagery [87], and the DINOv2 [69] general image model. We obtain embeddings from AlphaEarth Foundations [10] by sampling them from the Google Earth Engine Satellite Embedding V1 dataset[5]. We use the annualized embedding from the year the sensor was deployed, at a spatial scale of 250 m, resulting in a single 64-dimensional feature vector per location. Figure 10 shows predictive performance across those models.

[5] `https://developers.google.com/earth-engine/datasets/catalog/GOOGLE_SATELLITE_EMBEDDING_V1_ANNUAL`

E Biolith implementation details

To support the experiments presented in this manuscript, we developed *Biolith*, a new implementation of occupancy and related statistical ecological models in the Python programming language. Since Python is the lingua franca of machine learning research, we argue that this reduces the barrier of applying new machine learning algorithms to ecological modeling problems. Biolith is highly influenced by popular libraries in the statistical ecology ecosystem, including Unmarked [29, 56] and spOccupancy [23] in particular.

Beyond the classic Bernoulli occupancy model by

[62] used in this manuscript, Biolith implements the abundance model by [83], the continuous-score occupancy model by [78] for incorporating uncertain machine learning classifications, and the discrete-time counting occurrences model by [71]. In contrast to existing implementations, Biolith places great emphasis on modularity; all of the models named above can optionally be extended with modules such as false positives (thereby accounting for faulty observations of *presence*, which is useful when incorporating classifications produced by machine learning models) [82], site- and observation-level random effects, and spatial explicitness to model spatial autocorrelation.

Biolith is implemented using the NumPyro [72] probabilistic programming library, which enables defining models declaratively in a fashion similar to languages such as BUGS / JAGS. To fit models, Biolith uses the NumPyro implementation of Hamiltonian Monte Carlo inference and the No-U-Turn Sampler [49].

Since NumPyro is implemented using JAX [9], inference can be accelerated by graphics processing units (GPUs). The massive parallelism of GPUs unlocks potential reductions in inference runtime; however, currently this manifests itself exclusively on large datasets (see fig. 11).

We perform all of our experiments on 16 cores of an AMD EPYC 9554 with 64 GB of system memory and a single NVIDIA L40S GPU with 48 GB of memory. We use JAX 0.5.2, NumPyro 0.18.0, Biolith 0.0.9, and CUDA 12.9.

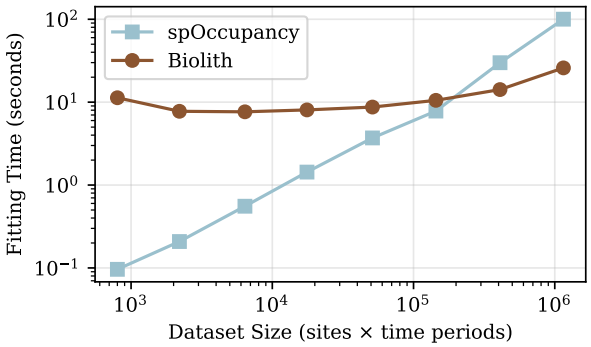

Figure 11: Benchmark of Biolith (GPU-accelerated) and spOccupancy [23] (CPU-only, single-threaded) runtimes over different dataset sizes. spOccupancy shows an almost perfect linear relationship and performs best at small to moderate dataset sizes, whereas the runtime of Biolith is initially strongly dominated by constant overhead and only becomes competitive on larger datasets. Details about the system used to run this benchmark can be found in appendix E.