PREDICTING SPECIES OCCURRENCE PATTERNS FROM PARTIAL OBSERVATIONS

Hager Radi Abdelwahed
Mila: Quebec AI Institute
hager.radi@mila.quebec

Mélisande Teng
Mila, Université de Montréal
tengmeli@mila.quebec

David Rolnick
Mila, McGill University
drolnick@cs.mcgill.ca

ABSTRACT

To address the interlinked biodiversity and climate crises, we need an understanding of where species occur and how these patterns are changing. However, observational data on most species remains very limited, and the amount of data available varies greatly between taxonomic groups. We introduce the problem of predicting species occurrence patterns given (a) satellite imagery, and (b) known information on the occurrence of other species. To evaluate algorithms on this task, we introduce SatButterfly, a dataset of satellite images, environmental data and observational data for butterflies, which is designed to pair with the existing SatBird dataset of bird observational data. To address this task, we propose a general model, R-Tran, for predicting species occurrence patterns that enables the use of partial observational data wherever found. We find that R-Tran outperforms other methods in predicting species encounter rates with partial information both within a taxon (birds) and across taxa (birds and butterflies). Our approach opens new perspectives to leveraging insights from species with abundant data to other species with scarce data, by modelling the ecosystems in which they co-occur.

1 INTRODUCTION

The interconnectedness of the climate and biodiversity crises has been widely acknowledged (Pörtner et al., 2021). Biodiversity and associated ecosystem services play a crucial role in both mitigating and adapting to climate change, as well as being severely threatened by it. It is essential to understand species distributions to inform land use decisions and adaptation measures. Using machine learning and remote sensing data has proved promising for a variety of tasks in biodiversity monitoring (Wang et al., 2010; Reddy, 2021) including species distribution modelling (Beery et al., 2021; Estopinan et al., 2022; Joly et al., 2022), improving on traditional methods using only environmental data. Moreover, the integration of machine learning and citizen science has proved helpful to monitor biodiversity at scale, automating labelling and decreasing the need for costly field surveys (Lotfian et al., 2021; Antonelli et al., 2023).

Recently, the SatBird Teng et al. (2023) dataset was proposed for the task of predicting bird species encounter rates from remote sensing imagery, leveraging observational data from the citizen science database eBird (Kelling et al., 2013). While this framework is useful to model species' distributions in places where no data has been recorded before, birds represent an atypical case since relatively large amounts of data are available. For most taxonomic groups, records are present in significantly fewer locations. For example, the eButterfly database (Prudic et al., 2017) modelled after eBird provides high quality presence-absence data for butterfly species but has smaller geographical coverage and less observation reports than eBird. Given this challenge, a potentially promising approach is to leverage the relationships between the occurrence patterns of different species, which are used extensively by ecologists e.g. in joint species distribution models. We design a task to explore this direction, using extensive bird occurrence data to help learn from sparser butterfly occurrence data, given the correlation in abundances between these two taxonomic groups (Gilbert & Singer, 1975;

Debinski et al., 2006; Eglington et al., 2015). Butterflies are particularly affected by climate change, with many species being adapted to specific environmental conditions (Hill et al., 2002; Rödder et al., 2021; Álvarez et al., 2024). We hope furthermore that such techniques will be useful in generalizing ML predictions to other under-observed but hyper-diverse taxa threatened by climate change, such as amphibians, freshwater fish, and plants.

In this paper, we consider the practical setting in which we want to do checklist completion for species encounter rates. Our main contributions are as follows:

- We introduce SatButterfly, a dataset for predicting butterfly species encounter rates from remote sensing and environmental data. Importantly, a subset of the data are colocated with existing data from SatBird, making it possible for the first time to leverage cross-taxon relationships in predicting species occurrence patterns from satellite images.
- We propose R-Tran, a model to train and predict species encounter rates from partial information about other species. R-Tran uses a novel transformer encoder architecture to model interactions between features of satellite imagery and labels.
- We evaluate R-Tran and other methods in predicting species encounter rates from satellite images and partial information on the encounter rates of other species. We find that R-Tran surpasses the baselines while providing flexibility to use with a variable set of information.

Problem definition We consider the regression task of predicting species encounter rates from satellite imagery given partial information. Our work is in line with frameworks that leverage partial label annotations during inference such as Feedback-prop (Wang et al., 2018), a method designed to be used at inference time given partial labels and iteratively updates the scores of unknown labels using the known labels on the test set only. C-Tran (Lanchantin et al., 2021) also proposed a transformer-based architecture for data completion, using embeddings from image features, labels and states in a multi-label classification setting. We consider our task in two different settings: (1) within a single taxon (SatBird-USA-summer) where we split bird species into songbirds and non-songbirds and must predict one set of species from the other, and, (2) across taxa (SatBird-USA-summer and SatButterfly), where information for one taxon is given and the other must be inferred. Task 2 reflects the standard imbalance in data between birds and insects, and tests the ability of ML models to leverage abundant bird data to predict butterfly data. Task 1 is designed so as to further investigate the task, with songbirds (Passeriformes) representing a discrete taxon of birds with approximately half the species of the whole.
We refer to SatBird-USA-summer as SatBird throughout the rest of the paper.

2 SatButterfly dataset

We collect and publish SatButterfly, a dataset for the task of predicting butterfly species encounter rates from environmental data and satellite imagery. The dataset consists of remote sensing images and environmental data along with labels derived from eButterfly (Prudic et al., 2017) observation reports, in the continental USA. Observations in eButterfly are particularly skewed towards North America, and we work specifically with the continental USA to match SatBird. Unlike in SatBird, we consider observations over the full year as most butterflies are non-migratory. We follow the process of data collection, preparation, and splitting as in SatBird, collecting satellite images, environmental data and encounter rates from recorded observations. We propose two versions of the dataset, a) SatButterfly-v1, where hotspots do not overlap with those of SatBird and b) SatButterfly-v2, where observations are collocated with those of SatBird (Teng et al., 2023). Each version of the dataset contains $\sim$7,000 data samples.

For the targets, we use presence-absence data in the form of complete checklists from the eButterfly citizen science platform. We extract checklists recorded from 2010 to 2023 in the continental USA for a total of 601 species. We compute the encounter rates for each location h as: $\mathbf{y^h} = (y^h_{s_1}, ..., y^h_{s_n})$, where y^h_s is the number of complete checklists reporting species s at h divided by the total number of complete checklists at h. We aggregate the species observations over 13 years to construct the final targets. We follow the GeoLifeCLEF 2020 dataset (Cole et al., 2020) and SatBird in extracting satellite data and environmental variables. Further details are in Appendix A.

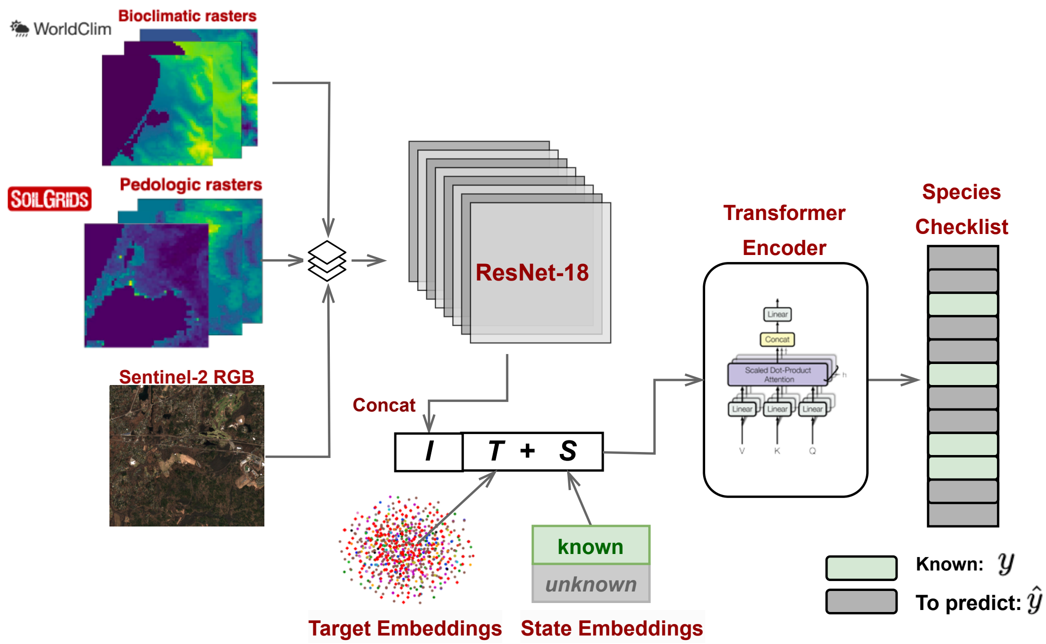

Figure 1: R-Tran architecture for predicting species encounter rates with partial information.

3 Methodology: R-Tran

Our goal is to explore methods for predicting species encounter rates given partial information. We draw inspiration from C-Tran (Lanchantin et al., 2021), an algorithm that predicts a set of target labels given an input set of known labels, and visual features from a convolutional neural network. A key ingredient of the method is a label mask training objective that encodes the three possible states of the labels (positive, negative, or unknown) during training. We propose the Regression Transformer (R-Tran) —a model adapted to our setting, which is regression not classification. Figure 1 shows the architecture for R-Tran, which acts as a general framework for predicting species encounter rates given (a) satellite images and environmental data, and (b) observations for a subset of species. Key components of R-Tran are the target embeddings T, where we train an embedding layer to represent all possible species classes, and state embeddings S, that represent state of each species whether known or unknown through training another embedding layer. Both embeddings are added, and concatenated with extracted features I from satellite and environmental data, then fed into a transformer encoder to model interactions between target classes and features. The model is flexible to accommodate any available partial species information to the full set of species. If no partial information are present, the model can still make predictions from satellite imagery and environmental variables. During training, we provide the full set of available encounter rates and randomly mask out a percentage of these as unknown. This percentage is chosen at random between $0.25n$ and n, where n is the total number of targets. During inference, we can provide information on 0 or more species to predict occurrences of remaining species.

4 Experiments

We compare R-Tran to ResNet18 (He et al., 2016) and Feedback-prop (Wang et al., 2018) models. Feedback-prop uses a trained model (in our case, ResNet18 model) and adapts the inference for unknown labels in a way that benefits from partial known labels. More experimental details are available in Appendix B. We evaluate our proposed model, R-Tran, in two settings: a) within taxon: SatBird, b) across taxa: SatBird & SatButterfly. For SatBird, we use only SatBird-USA-summer, since butterflies are rarely active in the winter.

Table 1: Evaluation within a taxon (SatBird): *A* refers to the subset of songbird classes only, *B* refers to the subset of non-songbird classes.

Model	MAE[1^{-2}]	MSE[1^{-2}]	Top-10 %	Top-30 %	Top-k %
All bird species					
ResNet18	2.19	0.65	45.28 ± 0.61	64.62 ± 0.5	66.26 ± 0.43
R-Tran	2.13	0.67	44.5 ± 0.4	64.2 ± 0.3	66.1 ± 0.26
Evaluation on classes *A*: Songbirds					
ResNet18	3.25	1.02	53.75 ± 0.66	78.94 ± 0.42	69.92 ± 0.48
ResNet18-*A*	3.23	1.02	53.91 ± 0.23	79.03 ± 0.28	70.1 ± 0.29
Feedback-prop $(A\|B)$	2.97	0.90	56.54 ± 0.6	80.71 ± 0.35	71.72 ± 0.32
R-Tran $(A\|B)$	**2.46**	**0.72**	**59.98** ± 0.16	**82.3** ± 0.1	**73.45** ± 0.2
Evaluation on classes *B*: non-songbirds					
ResNet18	1.35	0.37	59.97 ± 0.41	86.14 ± 0.32	59.72 ± 0.35
ResNet18-*B*	1.32	0.36	60.04 ± 0.29	86.09 ± 0.19	59.61 ± 0.16
Feedback-prop $(B\|A)$	1.22	0.32	63.09 ± 0.67	87.68 ± 0.4	61.94 ± 0.48
R-Tran $(B\|A)$	**0.99**	**0.26**	**66.25** ± 0.08	**89.28** ± 0.07	**64.19** ± 0.11

Within taxon (SatBird): we split SatBird's 670 species into songbirds, denoted set *A*, and non-songbirds, set *B*, composed of 298 and 372 species respectively. We train one ResNet18 model on all 670 species, as well as training individual ResNet18 models on *A* & *B* species subsets, denoted ResNet18-*A* and ResNet18-*B* respectively. R-Tran is also trained on the full set of bird species.

Table 1 shows results on SatBird: all species, subset *A* and subset *B*. R-Tran and Feedback-prop allow the use of available partial information about species, unlike ResNet18. We observe that R-Tran outperforms other baselines in predicting species *A* given information about *B*, as well as predicting *B* given information about *A*. All results reported are the average of 3 different seeds, where the standard deviation for MSE and MAE is negligible. Also, metrics reported for set *A* or *B* are masked for that set only.

Across taxa (SatButterfly & SatBird): In this setting, we use the full dataset SatBird-USA-summer, joined with SatButterfly. In SatButterfly, we select species that have at least 100 occurrences, resulting in 172 species of butterflies. When combining SatBird and SatButterfly, there are locations where only bird species are observed, only butterfly species are observed, or both species observed. For locations where not all classes of species are observed, we use a masked loss to mask out absent species; we provide more details in Appendix B.

We train ResNet18 and R-Tran on SatBird and SatButterfly jointly, where we predict $(670 + 172 = 842)$ species classes. We evaluate on the test set where both taxa (birds and butterflies) are observed, where metrics reported are masked for set *A*, or set *B*. Top-k refers to the adaptive top-k metric defined in Teng et al. (2023), where for each hotspot k is the number of species with non-zero ground truth encounter rates in the set of predicted species. In Table 2, we report results for ResNet18, Feedback-prop and RTran given partial observations of species to predict others. We find that R-Tran improves the performance on species *A* (birds) given information about butterflies, and *B* (butterflies) given information about birds. While we only evaluated the model in settings where the sets of known and unknown labels are fixed, it should be noted that R-Tran allows for more flexibility than Feedback-prop because at inference time, any arbitrary subset of species could be unknown, and R-Tran does not require multiple test examples to be effective.

5 Conclusion

We presented SatButterfly, a dataset that maps satellite images and environmental data to butterflies observations. We benchmarked different models for predicting species encounter rates considering partial information, within a taxon (SatBird), and across taxa (SatBird and SatButterfly). We find that leveraging partial information at inference time is more effective within a taxon than across taxa, and that our novel R-Tran model generally outperforms baselines. Our work offers avenues for combining data from different citizen science databases for joint prediction of species encounter

Table 2: Evaluation across taxa (SatBird and SatButterfly): *A* refers to bird species, *B* refers to butterflies. Results are evaluated on the subset of locations with both bird and butterfly data available.

Model	MAE[1^{-2}]	MSE[1^{-2}]	Top-10 %	Top-30 %	Top-k %
Evaluation on classes *A*: Birds					
ResNet18	2.07	0.62	48.02 ± 0.2	65.9 ± 0.15	73.17 ± 0.15
Feedback-prop $(A\|B)$	2.06	0.62	48.11 ± 0.16	66.18 ± 0.19	73.24 ± 0.1
R-Tran $(A\|B)$	**2.04**	**0.6**	**50.34** ± 0.3	**67.24** ± 0.4	**73.85** ± 0.18
Evaluation on classes *B*: Butterflies					
ResNet18	3.81	1.55	52.35 ± 0.43	83.74 ± 0.43	35.69 ± 0.85
Feedback-prop $(B\|A)$	3.81	**1.54**	52.51 ± 0.2	83.97 ± 0.37	35.83 ± 0.95
R-Tran $(B\|A)$	**3.6**	1.56	**52.82** ± 0.78	**84.21** ± 0.63	**36.31** ± 0.8

rates across taxa, in particular for cases where some species are less systematically surveyed than others. In future work, we hope to extend this approach to presence-only data, which represents the majority of citizen science observations, such as those from the iNaturalist platform.

REFERENCES

Census Bureau of USA. `https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html`, 2018. Accessed: 2023-06-06.

Hugo Alejandro Álvarez, Emma Walker, Mario Mingarro, Guim Ursul, Juan Pablo Cancela, Lee Bassett, and Robert J Wilson. Heterogeneity in habitat and microclimate delay butterfly community tracking of climate change over an elevation gradient. *Biological Conservation*, 289:110389, 2024.

Alexandre Antonelli, Kiran L Dhanjal-Adams, and Daniele Silvestro. Integrating machine learning, remote sensing and citizen science to create an early warning system for biodiversity. *Plants, people, planet*, 5(3):307–316, 2023.

Sara Beery, Elijah Cole, Joseph Parker, Pietro Perona, and Kevin Winner. Species distribution modeling for machine learning practitioners: A review. In *ACM SIGCAS conference on computing and sustainable societies*, pp. 329–348, 2021.

Elijah Cole, Benjamin Deneu, Titouan Lorieul, Maximilien Servajean, Christophe Botella, Dan Morris, Nebojsa Jojic, Pierre Bonnet, and Alexis Joly. The GeoLifeCLEF 2020 dataset. *Preprint arXiv:2004.04192*, 2020.

Diane M Debinski, Ron E VanNimwegen, and Mark E Jakubauskas. Quantifying relationships between bird and butterfly community shifts and environmental change. *Ecological Applications*, 16(1):380–393, 2006.

Sarah M Eglington, Tom M Brereton, Catherine M Tayleur, David Noble, Kate Risely, David B Roy, and James W Pearce-Higgins. Patterns and causes of covariation in bird and butterfly community structure. *Landscape ecology*, 30:1461–1472, 2015.

Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, pp. 226–231. AAAI Press, 1996.

Joaquim Estopinan, Maximilien Servajean, Pierre Bonnet, François Munoz, and Alexis Joly. Deep species distribution modeling from sentinel-2 image time-series: a global scale analysis on the orchid family. *Frontiers in Plant Science*, 13:839327, 2022.

Lawrence E Gilbert and Michael C Singer. Butterfly ecology. *Annual review of ecology and systematics*, 6(1):365–395, 1975.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Tomislav Hengl, Jorge Mendes de Jesus, Gerard BM Heuvelink, Maria Ruiperez Gonzalez, Milan Kilibarda, Aleksandar Blagotić, Wei Shangguan, Marvin N Wright, Xiaoyuan Geng, Bernhard Bauer-Marschallinger, et al. Soilgrids250m: Global gridded soil information based on machine learning. *PLoS one*, 12(2):e0169748, 2017.

Hijmans and al. Worldclim 1.4 (historical climate conditions). *International journal of climatology*, 25, 2005.

Jane K Hill, CD Thomas, Richard Fox, MG Telfer, SG Willis, J Asher, and B Huntley. Responses of butterflies to twentieth century climate warming: implications for future ranges. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 269(1505):2163–2171, 2002.

Alexis Joly, Hervé Goëau, Stefan Kahl, Lukáš Picek, Titouan Lorieul, Elijah Cole, Benjamin Deneu, Maximilien Servajean, Andrew Durso, Hervé Glotin, et al. Overview of lifeclef 2022: an evaluation of machine-learning based species identification and species distribution prediction. In *International Conference of the Cross-Language Evaluation Forum for European Languages*, pp. 257–285. Springer, 2022.

Steve Kelling, Jeff Gerbracht, Daniel Fink, Carl Lagoze, Weng-Keen Wong, Jun Yu, Theo Damoulas, and Carla Gomes. eBird: A human/computer learning network for biodiversity conservation and research. *AI Magazine*, 34, 03 2013.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Jack Lanchantin, Tianlu Wang, Vicente Ordonez, and Yanjun Qi. General multi-label image classification with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16478–16488, June 2021.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019. URL `https://openreview.net/forum?id=Bkg6RiCqY7`.

Maryam Lotfian, Jens Ingensand, and Maria Antonia Brovelli. The partnership of citizen science and machine learning: benefits, risks, and future challenges for engagement, data collection, and data quality. *Sustainability*, 13(14):8087, 2021.

Hans-Otto Pörtner, Robert J Scholes, John Agard, Emma Archer, A Arneth, Xuemei Bai, David Barnes, Michael Burrows, Lena Chan, WL Cheung, et al. Ipbes-ipcc co-sponsored workshop report on biodiversity and climate change. *IPBES and IPCC*, 10, 2021.

Kathleen L Prudic, Kent P McFarland, Jeffrey C Oliver, Rebecca A Hutchinson, Elizabeth C Long, Jeremy T Kerr, and Maxim Larrivée. ebutterfly: leveraging massive online citizen science for butterfly conservation. *Insects*, 8(2):53, 2017.

C Sudhakar Reddy. Remote sensing of biodiversity: what to measure and monitor from space to species? *Biodiversity and Conservation*, 30(10):2617–2631, 2021.

Dennis Rödder, Thomas Schmitt, Patrick Gros, Werner Ulrich, and Jan Christian Habel. Climate change drives mountain butterflies towards the summits. *Scientific Reports*, 11(1):14382, 2021.

Mélisande Teng, Amna Elmustafa, Benjamin Akera, Yoshua Bengio, Hager Radi, Hugo Larochelle, and David 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`.

Kai Wang, Steven E Franklin, Xulin Guo, and Marc Cattet. Remote sensing of ecology, biodiversity and conservation: a review from the perspective of remote sensing specialists. *Sensors*, 10(11): 9647–9667, 2010.

Tianlu Wang, Kota Yamaguchi, and Vicente Ordonez. Feedback-prop: Convolutional neural network inference under partial evidence. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

Table 3: Number of samples in each split for the two versions of SatButterfly.

	SatButterfly-v1	SatButterfly-v2
train	5316	4677
validation	1147	1002
test	1145	1005

A Appendix A: Dataset

In this section, we provide more details about SatButterfly dataset preparation. For the final hotspots, we excluded locations in the sea using the 5-meter resolution US nation cartographic boundaries provided by the Census Bureau's MAF/TIGER Cen (2018). We collected satellite images and environmental data for each location where butterfly species are observed. For Sentinel-2, we collected 10-meter resolution RGB and NIR reflectance data and RGB true color images, focusing on a 5 km^2 area centered on each hotspot. Our criteria for image selection included a maximum cloud cover of 10%, opting for the clearest image within specific time frames: January 1 - December 31, 2022. We assigned one image to each hotspot, based on the assumption that more recent satellite images better reflect our species data, considering the higher frequency of checklists in recent years as compared to earlier periods.

From WorldClim 1.4 (Hijmans & al., 2005), we retrieved 19 bioclimatic variables as rasters with a size of 50×50 and a spatial resolution of approximately 1 km, centered on each hotspot. These variables include annual trends in temperature, precipitation, solar radiation, wind speed, and water vapor pressure. In addition, we obtained 8 soil-related (pedologic) variables from SoilGrids (Hengl et al., 2017) at a finer resolution of 250 meters. SoilGrids offers global maps of various soil properties, such as pH, organic carbon content, and stocks, which are generated using machine learning techniques trained on soil profile data and environmental covariates from remote sensing sources.

As mentioned earlier, SatButterfly-v2 data is designed to share locations with SatBird, for both birds and butterflies data to be observed. To prepare the data, we perform BallTree-based clustering where SatBird locations are used as centriods. We then search within 1 km for neighbour butterfly observations using haversine distance. Finally, butterfly targets are aggregated and recorded for SatBird hotspots wherever available. We end up with available butterfly targets in a small subset of SatBird's train/validation/test splits.

Table 3 describes the number of hotspots present in each split for the two versions of SatButterfly dataset. Dataset is split similarly to SatBird using `scikit-learn` DBSCAN (Ester et al., 1996) clustering algorithm. Figures A and A shows distribution of data samples in train/val/test splits.

SatButterfly dataset is publicly available through this link.

B Appendix B

B.1 Experimental Details

For all models, we use the cross entropy loss: $\mathcal{L}_{CE} = \frac{1}{N_h}\sum_h \mathcal{L}_h = \frac{1}{N_h}\sum_h \sum_{s(\text{species})} -y_h^s \log(\hat{y}_h^s) - (1 - y_h^s)\log(1 - \hat{y}_h^s)$. For locations where not all classes of species are observed (when combining SatBird & SatButterfly), we define a masked loss. If we have the loss $L(y, \hat{y})$, where y is the true value and $\hat{y}$ is the predicted value, and a mask M (a binary vector or tensor of the same shape as y and $\hat{y}$), the masked loss function is defined as $L_{\text{masked}}(y, \hat{y}, M) = \frac{\sum_i M_i \cdot L(y_i, \hat{y}_i)}{\sum_i M_i}$, where M_i is the i-th element of the mask M. The loss is calculated only for the unmasked elements where $M_i = 1$, and is averaged over the number of unmasked elements.

All models use same input, RGB+ENV, where we concatenate environmental data (bioclimatic, pedologic) and RGB satellite images into different channels resulting in 30 channels. This is reported in SatBird (Teng et al., 2023) to be the best performing compared to using RGBNIR. We

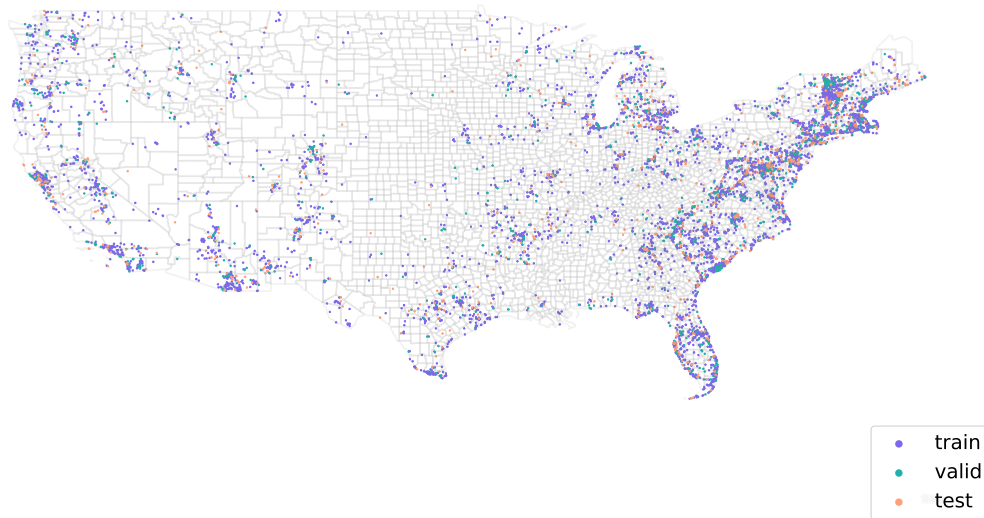

Figure 2: Distribution of hotspots across the training, validation, and test sets for SatButterfly-v1

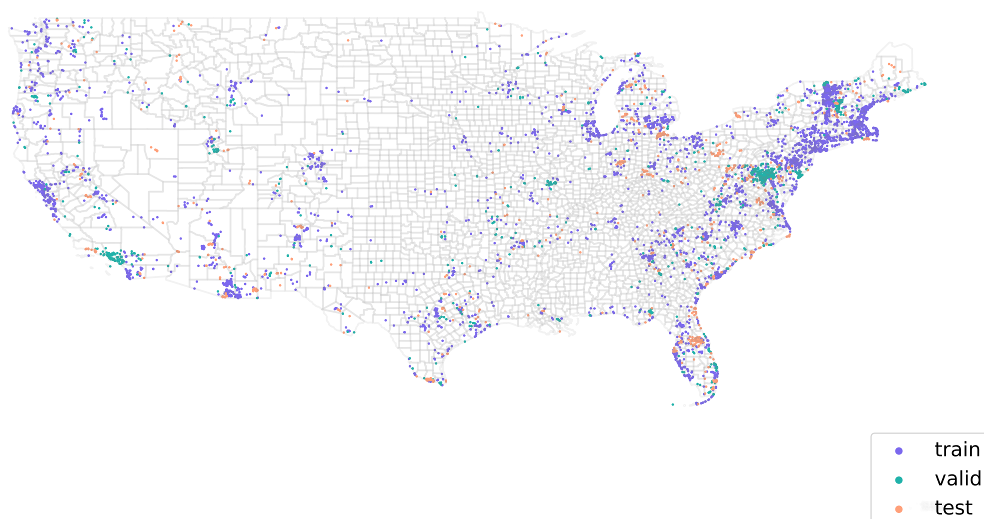

Figure 3: Distribution of hotspots across the training, validation, and test sets for SatButterfly-v2

consider a region of interest of 640 m^2, center-cropping the satellite patches to size 64×64 around the hotspot and normalizing the bands with our training set statistics. ResNet18 is adapted to take 30 channels as inputs instead of 3 channels. ResNet18 and R-Tran are trained for 50 epochs using Adam (Kingma & Ba, 2014) and AdamW (Loshchilov & Hutter, 2019) optimizers respectively.

For target embeddings T in R-Tran, we train an embedding layer of all possible classes $\{c_1, c_2, \ldots, c_m\}$ that can exist in a target y. In our experiments, we used textual species names from our datasets to train embeddings. Figure 4 shows embeddings of R-Tran after training the model on SatBird. A 2D TSNE Visualization shows clear discrimination between bird families. In state embeddings S, which are essential to mark class labels as known or unknown in a feedforward, each class is given a state s_i, where s_i takes a value of -1 if a class label is unknown, 0 if known to be absent, or a positive value $(0.25, 0.5, 0.75, 1.0)$ if the true probability y of a class > 0. This 4-bin quantization is an adaptation to the regression problem, instead of using a single value of 1 for all present classes. This is a hyper-parameter that can be further finetuned depending on the distribution of targets, but we found 4 to be performing best in our experiments. During inference, the model is still flexible to use information about presence 1 or absence 0 of species, rather than the exact encounter rate.

FeedbackProp uses the already-trained ResNet18 model to do inference with partial labels. It computes partial loss only with respect to a set of known labels *A* for input sample I. Then, it back-propagates this partially observed loss through the model, and iteratively update the input I in order to re-compute the predictions on the set of unknown classes *B*.

B.2 Evaluation

We report metrics regression metrics, MSE and MAE. Furthermore, we report the top-10, top-30 and top-k accuracies, representing the number of species present in the top k predicted species and the top k observed species. In Tables 1 and 2, all metrics reported are masked for the set of classes indicated whether it is set A or set B, using a binary mask.

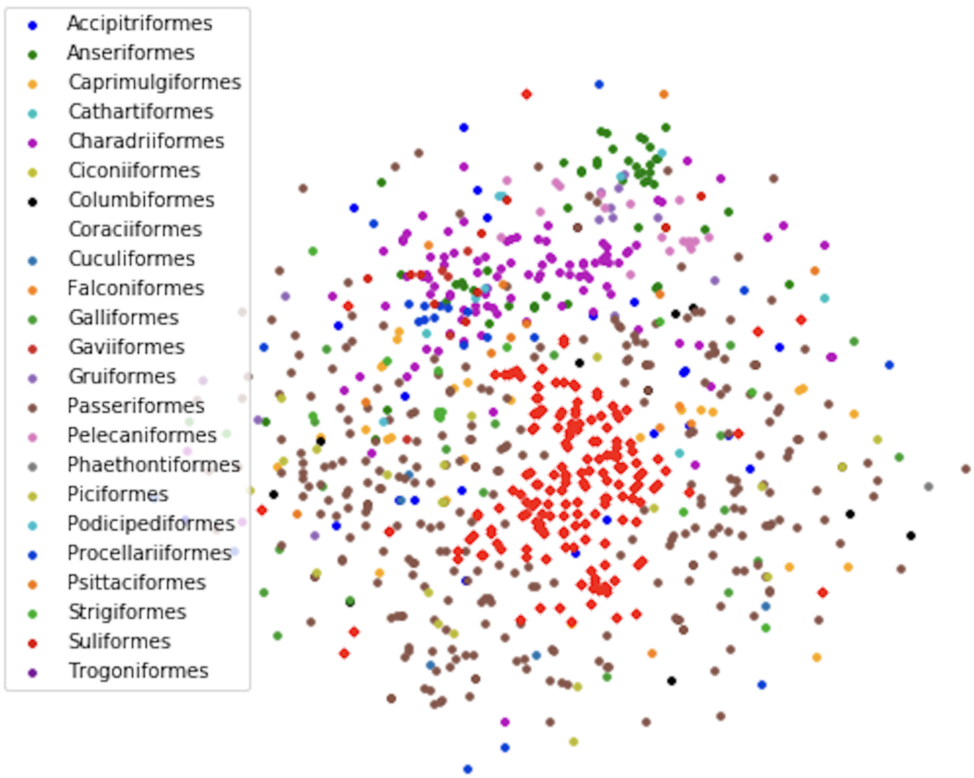

Figure 4: Visualization of trained R-Tran's target embeddings over bird species (SatBird).