

THESIS

FROM OCCURRENCES TO PREDICTORS: MODELING SPECIES DISTRIBUTION  
AND ENVIRONMENTAL DRIVERS OF NYMPHALIDAE

Submitted by

Melissa Morales

Graduate Degree Program in Ecology

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2025

Master's Committee:

Advisor: Greg Newman

Ruth Hufbauer

Seth Davis

Copyright by Melissa Morales 2025

All Rights Reserved

## ABSTRACT

### FROM OCCURRENCES TO PREDICTORS: MODELING SPECIES DISTRIBUTION AND ENVIRONMENTAL DRIVERS OF NYMPHALIDAE

Anthropogenic climate change poses growing threats to global biodiversity. These threats put pollinators at risk, which affects their ecological and agricultural roles. Butterflies in the Nymphalidae family act as both pollinators and indicators of environmental change, making them valuable models for ecological research. The reliability of occurrence data--from professional or participatory--significantly shapes species distribution modeling outcomes.

Chapter 1 evaluates the comparative performance of professional, participatory, and combined occurrence datasets in modeling Nymphalidae distributions across two climatically sensitive North American ecoregions, the Western Cordillera and South Central Semi-Arid Prairies. Using MaxEnt and Random Forest algorithms across four temporal bins (2008-2022), we demonstrate that Random Forest consistently outperforms MaxEnt in predictive performance. Spatial autocorrelation analyses revealed fundamental differences between data sources: participatory records exhibited strong spatial clustering (Moran's  $I = 0.36-0.65$ ), while professional records showed weak to moderate clustering (Moran's  $I = 0.13-0.32$ ), reflecting systematic differences in sampling efforts. Random Forest models built with professional datasets achieved the highest performance (AUC = 0.984; TSS = 0.944), while combined datasets offered the best trade-off between spatial coverage and predictive strength.

Chapter 2 applies these methodological insights to *Vanessa cardui* (Nymphalidae), a migratory butterfly and ecological generalist, examining species distribution across the same ecoregions. Random Forest models again achieved strong predictive performance (AUC = 0.968; TSS = 0.895). Variable importance analyses identified precipitation seasonality, maximum temperature of the warmest month, and mean diurnal range as the strongest predictors of species distribution. Residual diagnostics revealed systematic deviations at low and high suitability values, highlighting challenges in predicting rare outcomes. However, removing abundance-related bias shifted environmental relationships: population change was more strongly associated with temperature seasonality and diurnal variation, while consistently warm conditions showed negative correlations. Spatial predictions revealed persistent but patchy areas of suitability from 2008-2022, concentrated in both urban-adjacent and remote landscapes, with overall suitable habitat comprising a small proportion of the total study area.

Together, this research demonstrates both the potential of integrating diverse datasets with machine learning to assess species distributions and the importance of selecting appropriate datasets, as outcomes and interpretations highly depend on data source and quality. Simultaneously, it emphasizes the methodological and interpretive challenges posed by sampling bias and residual patterns. By identifying key environmental drivers and spatial hotspots, these models inform future monitoring and conservation strategies for butterflies in the face of accelerating environmental change.

## ACKNOWLEDGEMENTS

I want to express my gratitude to my advisor, Dr. Greg Newman, for his guidance and support. I appreciate the incredible opportunities he has provided me with. You are truly exceptional at what you do, and I thank you for inspiring me to become a better scientist and person. I would like to sincerely thank my committee members, Dr. Ruth Hufbauer and Dr. Seth Davis, for their expertise and guidance throughout my research, it has helped me grow in ways that I will carry with me throughout my career. Additional thanks to those at CSU's Geospatial Centroid and the C.P. Gillette Museum for their resources, space, and time.

To Dr. John Leiser, thank you for being an incredible first mentor in this field and for being the first educator to ever offer me genuine support; your example continues to inspire me today. To Dr. Joshua Lord, thank you for your mentorship and for the opportunities that have shaped my path and the scientist I am today.

To the National Parks, whose beauty and presence continues to guide my journey, it has been one of my greatest privileges to work alongside and for you. These lands kindled my love for the natural world, and in return, I feel a deep responsibility to protect public lands and honor the irreplaceable wonders of Mother Nature and all the species who call her home.

Thank you to my family for your endless love and unwavering support. To my Mom and Dad, thank you for your embrace and for always being my guiding light. From the bottom of my heart, thank you for letting me discover my path and for supporting

and cheering me on every step of the way. My path would have been much more difficult without your sacrifice in moving our family from New York City to the Pocono Mountains. I'm lovingly indebted to you both--thank you for being the most amazing parents that I could ask for.

To my younger siblings, thank you for the countless laughs and the much-needed tears, no matter the hour. To my brother, I'm so excited to see you stepping into the environmental science field and can't wait to see all the incredible opportunities that come your way. This field urgently needs passionate, capable minds like yours, and I know you'll give your absolute best to make a real difference. To my sister, if I could struggle through high school and still end up where I am today, then you--of all people--can achieve anything you set your heart and mind to. Be patient and kind with yourself; I'm excited to see where your wildest dreams take you. You both have my eternal love, support, and friendship.

Last but certainly not least, I owe endless thanks to my best friend and the love of my life, Tommy, for your unwavering love, support, and companionship. I could not have completed this marathon without you. Thank you for believing in me and guiding me forward with patience, care, and encouragement at every turn. I'm grateful for every breather we took during each paddle, in all the cabins, at every picnic, and on each trail from the mountains to the deserts and everywhere in between. We've shared so many incredible chapters since high school, and I'm so excited for all the ones still to come.

## TABLE OF CONTENTS

ABSTRACT .....	ii
ACKNOWLEDGEMENTS.....	iv
CHAPTER 1: BUTTERFLIES AND BIAS: UNDERSTANDING NYMPHALIDAE DISTRIBUTIONS THROUGH DIVERSE DATA LENSES .....	1
Introduction.....	1
Methods.....	4
Ecoregions and Occurrence Data .....	4
Population Density.....	6
Environmental Variables .....	7
Fishnet Grid with Zonal Statistics .....	10
Background points.....	10
Spatial Autocorrelation Analysis .....	11
SDM.....	12
MaxEnt .....	12
Random Forest.....	12
Model Performance Metrics .....	13
Results .....	14
Trend Analysis for Occurrences.....	14
Spatial Autocorrelation Analysis for Occurrence Datasets .....	14
SDM: MaxEnt.....	15
Participatory Dataset Performance .....	16
Professional Dataset Performance .....	16
Combined Dataset Performance.....	16
MaxEnt Heatmaps.....	17
SDM: Random Forest.....	19
Participatory Dataset Performance .....	19
Professional Dataset Performance .....	19

Combined Dataset Performance.....	20
Random Forest Heatmaps .....	20
Discussion .....	23
REFERENCES .....	27
CHAPTER 2: PATTERNS IN PAINTED LADIES: RANDOM FOREST INSIGHTS INTO DISTRIBUTION AND ENVIRONMENTAL DRIVERS .....	34
Introduction.....	34
Methods.....	36
Ecoregions and Occurrence Data .....	36
Environmental Variables .....	36
Background Points and Data Preparation.....	37
Random Forest Modeling.....	37
Model Evaluation .....	37
Species Distribution .....	38
Visualization and Output .....	38
Results .....	38
Discussion .....	46
REFERENCES .....	50

# CHAPTER 1: BUTTERFLIES AND BIAS: UNDERSTANDING NYMPHALIDAE DISTRIBUTIONS THROUGH DIVERSE DATA LENSES

## **Introduction**

Species occurrence data, which record the presence or absence of a species at a specific location and time, are fundamental to ecological research, management, and policy. These georeferenced datasets allow researchers to study biodiversity patterns, community assembly, ecosystem functioning, and species-environment relationships, and they are essential for conservation planning, including identifying biodiversity hotspots, evaluating vulnerability to climate change, and designing protected areas (Franklin, 2023; Jetz et al., 2019; Petersen et al., 2021; Zhang et al., 2022). Occurrence data are widely accessible through networks such as the Global Biodiversity Information Facility (GBIF), which compiles contributions from natural history museums, research institutions, and participatory science platforms like iNaturalist. While museums remain important, participatory contributions are rapidly increasing and, in some cases, surpassing professional collections, particularly for birds (Wagner et al., 2021; Galván, Barrientos et al., 2022). Recent advances in machine-based pattern recognition have enabled the creation of occurrence maps for over 600,000 species using GBIF data, with strong alignment to expert-generated maps across taxa (Dasgupta et al., 2024), reinforcing the value of occurrence data in biodiversity research and conservation.

As global climates shift, species distribution models (SDM), also known as habitat suitability or climate envelope models, have become essential for forecasting where species are likely to persist under environmental change (Guisan & Thuiller,

2005; Elith & Leathwick, 2009). Using occurrence data and abiotic variables to define ecological niches, SDM estimate past, present, and future ranges. When aggregated across species, they inform broader richness and diversity analyses, making them crucial to segments of conservation planning (Franklin, 2013).

Despite their usefulness, SDM carry uncertainty. Biases in species selection, study region, and uneven sampling can influence results (Thomas et al., 2004; Loiselle et al., 2003). Data sources also shape outcomes: participatory science platforms such as iNaturalist, eBird, and Zooniverse provide vast coverage but can overrepresent accessible areas, whereas museum collections mainly reflect historical sampling patterns (Beck et al., 2014). Most SDM emphasize abiotic variables while neglecting biotic interactions (Araújo & Luoto, 2007) or microhabitat features (Dormann et al., 2018; Guisan et al., 2017), resulting in less reliable predictions in some situations. Advances in machine learning have expanded SDM accessibility, but these methods highlight correlations without demonstrating causality, unlike mechanistic approaches that require more data and expertise (Arif & MacNeil, 2022; Baker et al., 2017).

To strengthen predictive capacity, this study applies two widely used algorithms: MaxEnt and Random Forest. MaxEnt is a presence-only method that contrasts environmental conditions at observed occurrence points with background locations, making it effective when only opportunistic or spatially biased presence data are available (Syfert et al., 2013). Random Forest is a presence-absence classifier that integrates pseudo-absence points into an ensemble of decision trees, capturing nonlinear relationships and remaining robust to noise and correlation among predictors

(Zhao et al., 2022). Their complementary strengths make them valuable for evaluating the reliability of participatory versus professional datasets in SDM.

Professional data sources--such as museum collections, research sites, and public lands--provide decades of spatially and temporally rich information with high taxonomic accuracy, which is especially important for groups that are difficult to identify, such as invertebrates (Suarez & Tsutsui, 2004; Spear et al., 2017; Shultz et al., 2021; Konowalik & Nosol, 2021; Turley et al., 2024). However, they are often biased toward areas of scientific interest and remain only partially digitized (Lazagabaster et al., 2024; Fisher-Phelps et al., 2017; Barends et al., 2020). Participatory science platforms such as iNaturalist and CitSci.org broaden datasets through casual to structured observations, often with photo documentation and ID support (Pocock et al., 2014). They provide broad coverage and long-term monitoring (e.g., Bald Eagle Watch; Bove et al., 2024), while also engaging the public in biodiversity stewardship. Yet challenges remain, including observer bias, inconsistent sampling, and retention of participants (Kosmala et al., 2016; Johnston et al., 2023). Professional science offers taxonomic rigor and historical depth, while participatory science expands spatial and temporal coverage; integrating both with quality assurance can maximize their complementary strengths (Araújo et al., 2019; Rocchini et al., 2023).

Given the growing use of participatory science in ecological research, it is essential to evaluate how different data sources influence SDM performance, particularly for Lepidoptera, which are sensitive to environmental change, well-studied, and widely observed. Few studies have directly compared SDM built from participatory versus professional sources for butterflies in the Western United States, leaving an

important gap. To address this, this study asks: Do SDM built with participatory data differ in accuracy from those built with professional data? Using Nymphalidae butterflies as a focal group, I compare models based on participatory and professionally collected records, applying both MaxEnt and Random Forest, to evaluate whether participatory data can achieve predictive accuracy and ecological reliability comparable to professional data.

## **Methods**

### Ecoregions and Occurrence Data

Understanding ecological changes benefits from analyzing species within ecoregions rather than political boundaries such as counties or states (Omernik, 2004). This study focused on the Western Cordillera and South Central Semi-Arid Prairies Level II ecoregions in the United States, extracted from the U.S. EPA Ecoregion shapefile (using coordinate system WGS 84) employing ArcGIS Pro (Figure 1). I excluded the Canadian portion of the Western Cordillera because the available PRISM climate data (1961-1990) did not align with this study's temporal scope (2008-2022).



**Figure 1:** Map of the study area showing the South Central Semi-Arid Prairies and Western Cordillera ecoregions, delineated using the U.S. EPA Level II Ecoregions shapefile. These two distinct ecological regions form the spatial extent for the SDM in this study.

I downloaded occurrence data from GBIF (GBIF.org, 2023), querying Nymphalidae records within a polygon covering both ecoregions from 2008 through 2022. This query returned records from 70 datasets, however, datasets outside of the United States were removed. Using GBIF metadata, I classified datasets as either participatory initiatives (no formal training for collectors) or professional initiatives (involving some form of training) (Table 1).

I created three datasets: professional, participatory, and combined (Table 1), each subdivided into four fixed time bins. Using identical temporal bins allowed direct comparison between all datasets, since each model represented the same time frame.

**Table 1:** Number of Nymphalidae occurrence records from professional and participatory datasets, as well as a separate combined dataset integrating both sources, summarized across four temporal bins (2008-2011, 2012-2015, 2016-2019, and 2020-2022).

<b>Fixed year bins</b>	<b>Professional Records</b>	<b>Participatory Records</b>	<b>Combined Records</b>
<b>2008-2011</b>	1,226	1,343	2,569
<b>2012-2015</b>	3,961	5,852	9,813
<b>2016-2019</b>	2,505	46,547	49,052
<b>2020-2022</b>	2,810	79,145	81,955
<b>Sum of all bins</b>	10,502	132,887	143,389

### Population Density

I downloaded the 2020 human population density data for North America from ArcGIS Online. The Center for International Earth Science Information Network at Columbia University developed the source dataset, the Gridded Population of the World, Version 4 (GPWv4), in 2018, with a spatial resolution of 30 arc-seconds per

pixel. I then imported the population density layer into ArcGIS Pro and overlaid it with ecoregions to highlight areas of highest population density.

### Environmental Variables

Land cover data provides critical information on habitat availability, quality, and connectivity for Nymphalidae butterflies (Tzortzakaki et al., 2019; Diengdoh et al., 2023). As a categorical variable, land cover represents vegetation type, urban areas, and water bodies, all of which influence butterfly population dynamics and distribution (Tong et al., 2025; Shrestha et al., 2024; Senapathi et al., 2015; Huang et al., 2024; Munisi et al., 2024). I obtained land cover data from the National Land Cover Dataset (NLCD) via the U.S. Geological Survey (USGS) through the Living Atlas Portal in ArcGIS Pro, and I processed them to restrict coverage to the study regions. The dataset spans a time series of land cover categories from 2001 to 2021.

I included topographic variables (i.e., elevation, slope, and aspect) because they shape microclimatic conditions, habitat structure, and resource availability (Popović et al., 2021; Shrestha et al., 2024; Svancara et al., 2019; Rodríguez-Castañeda et al., 2019). Elevation strongly influences temperature, moisture, and sunlight exposure, which in turn affect butterfly survival and reproduction (Mahata et al., 2023). I obtained elevation data from USGS Earth Explorer, mosaicked the raster's in ArcGIS Pro, and clipped them to the study ecoregions. From this layer, I derived slope and aspect using ArcGIS spatial analyst tools.

I also included bioclimatic variables, which describe patterns of temperature and precipitation that directly influence butterfly physiology, life cycles, and host plant

availability (Hill et al., 2021; Wells & Tonkyn, 2018). These variables are essential for modeling species distribution under climate change scenarios. To generate them, I downloaded precipitation (PPT), maximum temperature (TMAX), and minimum temperature (TMIN) data (2008-2022) from the PRISM database at Oregon State University. After importing and clipping the datasets to the ecoregions in ArcGIS Pro, I used the R package *bioclim* (Booth et al., 2014) to calculate 19 bioclimatic variables (Table 2).

**Table 2:** Bioclimatic variables using PPT, TMIN and TMAX from PRISM.

<b>Variable</b>	<b>Name</b>	<b>Description</b>
<b>BIO1</b>	Annual Mean Temperature	Mean of all monthly temperatures
<b>BIO2</b>	Mean Diurnal Range	Mean of monthly (max temp - min temp)
<b>BIO3</b>	Isothermality	$(\text{BIO2} / \text{BIO7}) \times 100$ ; day-night temp oscillation vs annual range
<b>BIO4</b>	Temperature Seasonality	Standard deviation of monthly temps $\times 100$ ; measures temperature variability
<b>BIO5</b>	Max Temperature of Warmest Month	Highest average maximum temperature in the warmest month
<b>BIO6</b>	Min Temperature of Coldest Month	Lowest average minimum temperature in the coldest month
<b>BIO7</b>	Temperature Annual Range	$\text{BIO5} - \text{BIO6}$ ; overall temperature range
<b>BIO8</b>	Mean Temperature of Wettest Quarter	Average temp of the 3-month period with highest precipitation
<b>BIO9</b>	Mean Temperature of Driest Quarter	Average temp of the 3-month period with lowest precipitation
<b>BIO10</b>	Mean Temperature of Warmest Quarter	Average temperature during the warmest quarter
<b>BIO11</b>	Mean Temperature of Coldest Quarter	Average temperature during the coldest quarter
<b>BIO12</b>	Annual Precipitation	Total yearly precipitation
<b>BIO13</b>	Precipitation of Wettest Month	Total precipitation in the wettest month
<b>BIO14</b>	Precipitation of Driest Month	Total precipitation in the driest month
<b>BIO15</b>	Precipitation Seasonality	Coefficient of variation (CV) of monthly precipitation
<b>BIO16</b>	Precipitation of Wettest Quarter	Total precipitation during the wettest quarter
<b>BIO17</b>	Precipitation of Driest Quarter	Total precipitation during the driest quarter
<b>BIO18</b>	Precipitation of Warmest Quarter	Total precipitation during the warmest quarter
<b>BIO19</b>	Precipitation of Coldest Quarter	Total precipitation during the coldest quarter

## Fishnet Grid with Zonal Statistics

I created a uniform fishnet grid using the WGS 84 coordinate system, with cells measuring  $0.054^\circ \times 0.054^\circ$  ( $\approx 6 \text{ km} \times 6 \text{ km}$  at the equator), an appropriate resolution for regional-scale ecological modeling. To align the grid with ecologically relevant boundaries, I clipped it to an ecoregion shapefile covering the Western Cordillera and South Central Semi-Arid Prairies. I then exported the clipped grid as a shapefile to ensure consistent spatial referencing and integration with other datasets.

For each grid cell, I extracted zonal statistics of all environmental variables using the `exactextractr` package in R (Baston, 2020). This method computed minimum, mean, and maximum values of raster-based covariates overlapping each cell, allowing precise aggregation of spatial data while preserving environmental variation across heterogeneous landscapes.

## Background points

I generated background points for MaxEnt through simple random sampling across the entire study area using a predefined fishnet grid. This approach ensured an unbiased representation of the available environmental space and avoided spatial or environmental bias that can arise from non-random sampling or observer behavior. By distributing background points uniformly and independently of presence data, the model more accurately distinguished suitable habitat conditions from the broader landscape (Merow et al., 2013). This method established a neutral baseline for comparison, allowing the model to contrast presence locations against a consistent environmental background without overrepresenting heavily sampled or easily accessible regions. For

Random Forest, I generated pseudo-absence points from the fishnet grid at a ratio of five background points for every presence point.

### Spatial Autocorrelation Analysis

I evaluated the spatial independence of model predictions and residual errors for both MaxEnt and Random Forest models using Moran's I statistics, which quantify the degree to which residuals are spatially clustered or dispersed. In SDM, significant residual autocorrelation can signal model misspecification, sampling bias, or unaccounted environmental structures. I applied Moran's I to residuals from each fixed year bin to assess whether participatory or professional collections produced spatially biased errors. I hypothesized that models developed from participatory collections would differ significantly from those based on professional collections, while the alternative hypothesis predicted no significant difference between data sources.

### Variable Selection and Multicollinearity Assessment

I initially ran MaxEnt with all 24 environmental variables to assess their contributions, then applied variable selection to reduce multicollinearity and improve interpretability. For MaxEnt, I used Variance Inflation Factor (VIF) analysis and retained only predictors with  $VIF < 10$ ; for Random Forest, I applied correlation-based filtering with the findCorrelation function ( $r > 0.85$ ). From this process, I selected ten biologically relevant and statistically independent predictors, all with exceptionally low VIF values ( $< 1$ ), ensuring independent contributions to the models. The final predictors included climatic variables (BIO3, BIO4, BIO7, BIO14, BIO15, BIO18, BIO19), topography (elevation, slope), and land cover. This approach produced ecologically robust,

interpretable models for both MaxEnt and Random Forest with minimal redundancy among predictors.

## SDM

### *MaxEnt*

I implemented MaxEnt models using linear, quadratic, and hinge features (lqh) with a regularization multiplier ( $\beta$ ) of 1.0, randomly sampling 10,000-20,000 background points depending on dataset size (combined datasets used 20,000). I excluded background points from fishnet grids missing required variables (e.g., bioclimatic, elevation, land cover). I ran separate models for each dataset (participatory, professional, combined) and fixed year bin (2008-2011, 2012-2015, 2016-2019, 2020-2022) to evaluate temporal changes in performance, variable importance, and predicted distributions. I assessed model performance using 5-fold spatial block cross-validation with  $6 \times 6$  km stratified blocks and repeated fold assignments over 100 iterations. Performance metrics included Area Under the Curve, True Skill Statistic and Akaike Information Criterion.

### *Random Forest*

I implemented Random Forest (RF) models to predict Nymphalidae distributions and compare results with MaxEnt. For each dataset (participatory, professional, combined) and fixed year bin, I ran RF using the same predictors selected for MaxEnt, with 500 trees, default node size, and cross-validated *mtry* optimization. I capped sampling at 50,000 points and used a fixed seed to ensure reproducibility of pseudo-absence selection. I evaluated model performance with Area Under the Curve, True

Skill Statistic, and Cohen's Kappa, and calculated Moran's I on residuals using k-nearest neighbors ( $k = 10$ ). I generated predictions, performance metrics, and variable rankings for each fixed year bin to allow direct comparisons with MaxEnt results.

### Model Performance Metrics

I assessed the predictive performance and reliability of MaxEnt and Random Forest models using multiple evaluation metrics. Both models were evaluated with Area Under the Curve (AUC) and True Skill Statistic (TSS) to provide standardized measures for cross-model comparison. AUC quantifies a model's ability to distinguish between presence and absence (or pseudo-absence), ranging from 0.5 (random) to 1.0 (perfect discrimination). TSS accounts for both sensitivity and specificity, ranging from -1 to 1, where 1 indicates flawless classification and values below 0 indicate no better than random performance. For each model, I included a third, model-specific metric to complement AUC and TSS: for MaxEnt, I calculated the corrected Akaike Information Criterion (AICc), which balances goodness-of-fit against model complexity and small sample sizes, with lower values indicating more parsimonious models and reduced overfitting risk. For Random Forest, I calculated Cohen's Kappa, which measures agreement between predicted and observed presence/absence values beyond chance, ranging from -1 (complete disagreement) to 1 (perfect agreement).

AICc does not apply to Random Forest because it is a non-parametric ensemble method, while Cohen's Kappa could be applied to MaxEnt only by imposing an arbitrary threshold on continuous suitability scores, introducing subjectivity. In addition to Cohen's Kappa, Random Forest performance was also evaluated using out-of-bag (OOB) error. This measure is calculated during model training by predicting data excluded from each

bootstrap sample, thereby providing an unbiased estimate of prediction error without requiring a separate test dataset, although such datasets were also employed in this study. AUC remains the preferred metric for presence-only models because it evaluates predictive performance without threshold dependence. Taken together, these metrics provide a robust evaluation of model performance and allow me to test the hypothesis of whether SDM based on participatory data differ in accuracy from those based on professional collection data.

## **Results**

### Trend Analysis for Occurrences

Professional records declined modestly across time periods, decreasing by 17% from 2,423 records in 2008-2011 to 2,014 in 2012-2015, then by 11% to 1,782 records in 2016-2019, before slightly recovering with a 9% increase to 1,948 records in 2020-2022. In contrast, participatory records expanded rapidly, rising by 229% from 2,215 (2008-2011) to 7,276 (2012-2015), then by 168% to 19,470 (2016-2019), and finally by 236% to 65,402 records in 2020-2022. The combined dataset mirrors the participatory trajectory after 2011, reaching 67,350 records in the most recent period. Participatory became the dominant driver of data availability for Nymphalidae, contributing more than 97% of records by 2020-2022.

### Spatial Autocorrelation Analysis for Occurrence Datasets

Moran's I analysis revealed significant spatial clustering ( $p < 0.001$ ) across all datasets and time periods, with distinct patterns between data sources (Table 3). Professional data exhibited weak to moderate spatial autocorrelation (Moran's I = 0.127-0.319), while participatory data showed moderate to strong clustering (Moran's I =

0.362-0.655). Professional data maintained weak autocorrelation during 2008-2019, shifting to moderate clustering in 2020-2022 (Moran's I = 0.319). The combined dataset closely mirrored participatory patterns as its dominance increased, ranging from Moran's I = 0.298 in 2008-2011 to a maximum of 0.711 in 2016-2019. Despite large differences in sampling intensity (1,226-79,145 presences), background points remained relatively constant across datasets and periods, indicating stable geographic extent. Standard errors of Moran's I estimates were inversely related to sample size across the range of 1,226 to 79,145 observations. These results indicate that the spatial distribution of errors differs between data sources, with participatory data exhibiting systematically different spatial patterns compared to professional data.

**Table 3:** Moran's I results for participatory, professional, and combined datasets.

<b>Dataset</b>	<b>Time period</b>	<b>Moran's I</b>	<b>Interpretation</b>
<b>Participatory</b>	2008-2011	0.3615	Moderate spatial clustering
	2012-2015	0.6546	Strong spatial clustering
	2016-2019	0.6507	Strong spatial clustering
	2020-2022	0.6142	Strong spatial clustering
<b>Professional</b>	2008-2011	0.1266	Weak spatial clustering
	2012-2015	0.1902	Weak spatial clustering
	2016-2019	0.1851	Weak spatial clustering
	2020-2022	0.3188	Moderate spatial clustering
<b>Combined</b>	2008-2011	0.2981	Moderate spatial clustering
	2012-2015	0.6192	Strong spatial clustering
	2016-2019	0.7106	Strong spatial clustering
	2020-2022	0.6797	Strong spatial clustering

SDM: MaxEnt

MaxEnt models exhibited distinct performance trajectories across data sources.

Participatory models achieved the highest overall performance levels and demonstrated

the most consistent temporal improvement, while professional models showed greater variability with notable performance fluctuations mid-series. Combined dataset models displayed intermediate characteristics, benefiting from increased sample sizes while maintaining stable performance metrics. All models achieved AUC values above 0.7, indicating acceptable to excellent discriminatory ability, with TSS values approaching or exceeding the 0.6 threshold for good model performance in later periods across most datasets.

#### *Participatory Dataset Performance*

These MaxEnt models showed progressively improved performance across time periods (Table 4). AUC values increased from 0.804 in 2008-2011 to 0.873 in 2020-2022. TSS values improved from 0.494 to 0.625, with the final two periods meeting or exceeding the 0.6 threshold for good model performance. AICc values declined from 107.1 to 72.1, reflecting more parsimonious models. A temporary AICc increase during 2012-2015 was followed by substantial declines.

#### *Professional Dataset Performance*

These MaxEnt models showed variable performance across periods (Table 4). AUC values fluctuated from 0.719 in 2008-2011 to a peak of 0.863 in 2020-2022, with a dip to 0.752 in 2016-2019. TSS values followed a similar pattern, reaching 0.668 in the final period. AICc values peaked at 117.5 in 2016-2019.

#### *Combined Dataset Performance*

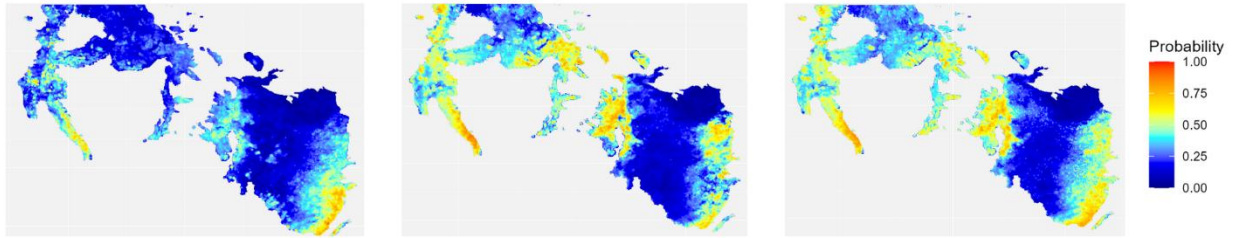
These MaxEnt models demonstrated progressively improving performance (Table 4). AUC values increased from 0.773 in 2008-2011 to a peak of 0.859 in 2016-

2019, then slightly declined to 0.851 in 2020-2022. TSS values consistently improved from 0.447 to 0.585, approaching the 0.6 threshold in later periods. AICc values decreased steadily from 106.4 to 73.4.

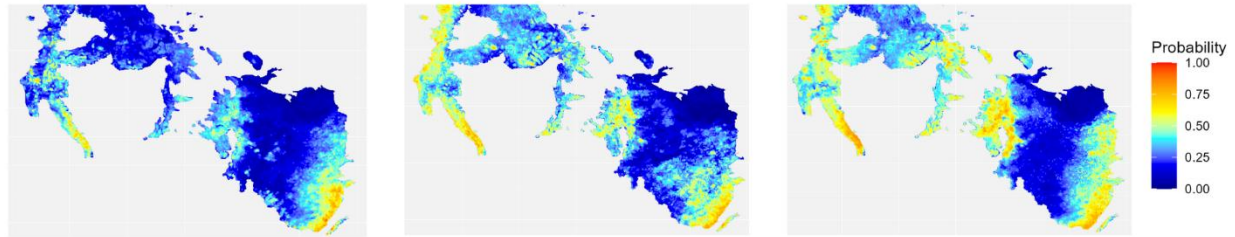
### *MaxEnt Heatmaps*

MaxEnt species distribution predictions showed substantial temporal variation across datasets and periods (Figure 2). Participatory models exhibited progressive spatial expansion, with high-probability areas in 2008-2011 spreading broadly across the western regions by 2020-2022. In contrast, professional models displayed range contraction, shifting from extensive Pacific and Rocky Mountain distributions early on to spatially restricted Western Cordillera predictions later, specifically in Idaho. Combined dataset predictions were intermediate, gradually expanding over time with the most extensive suitable habitat in 2020-2022. Unlike Random Forest models, MaxEnt predictions reflected considerable temporal shifts in suitable habitat, with changing probability distributions over the 15-year study period.

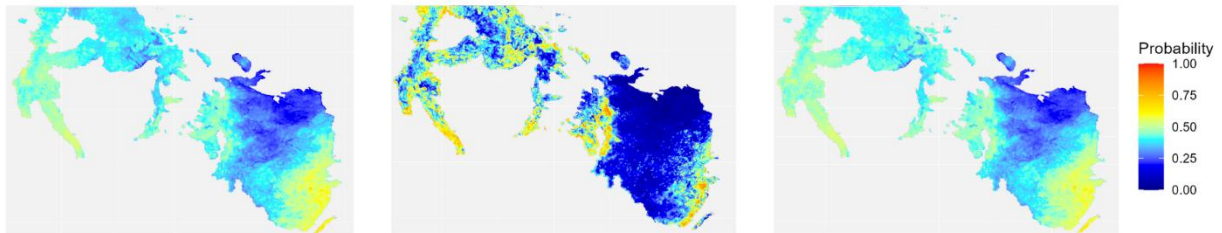
MaxEnt results for 2008-2011 in the following order:  
participatory, professional and combined.



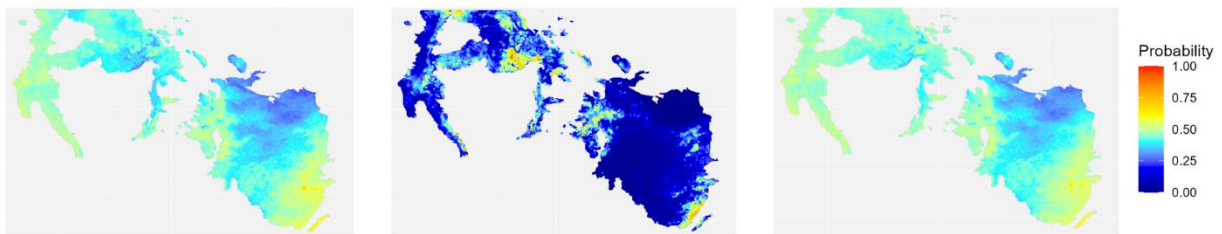
MaxEnt results for 2012-2015 in the following order:  
participatory, professional and combined.



MaxEnt results for 2016-2019 in the following order:  
participatory, professional and combined.



MaxEnt results for 2020-2022 in the following order:  
participatory, professional and combined.



**Figure 2:** MaxEnt species distribution predictions for Nymphalidae across four time periods (2008-2011, 2012-2015, 2016-2019, 2020-2022). Each row demonstrates results from participatory (left), professional (center), and combined datasets (right). Colors represent species distribution probabilities from 0.00 (dark blue, unsuitable) to 1.00 (red, highly suitable) across the Western Cordillera and South Central Semi-Arid Prairies ecoregions.

### SDM: Random Forest

All Random Forest models demonstrated strong performance with Moran's I values remaining low across datasets and time periods (-0.003 to 0.055), indicating minimal spatial autocorrelation in residuals and effective capture of spatial patterns. Professional models exhibited slightly higher Moran's I values (0.021-0.037) than participatory models, though all remained well below the threshold of concern (>0.1).

#### *Participatory Dataset Performance*

These Random Forest models performed strongly and improved over time (Table 4). AUC values rose from 0.811 in 2008-2011 to a peak of 0.953 in 2016-2019, declining slightly to 0.946 in 2020-2022. TSS values increased from 0.527 to 0.804-0.812 in later periods. Cohen's Kappa rose from 0.593 to 0.799-0.813. OOB error rates remained below 10%, with the lowest in 2012-2015 (6.0%).

#### *Professional Dataset Performance*

These Random Forest models consistently produced the highest model performance across all periods (Table 4). AUC values rose from 0.895 to a peak of 0.984 in 2020-2022. TSS values improved from 0.745 to 0.944, reaching exceptional performance above 0.8 in the latter three periods. Cohen's Kappa increased from 0.808

to 0.943. OOB error rates declined from 4.7% to 1.7%, confirming superior performance.

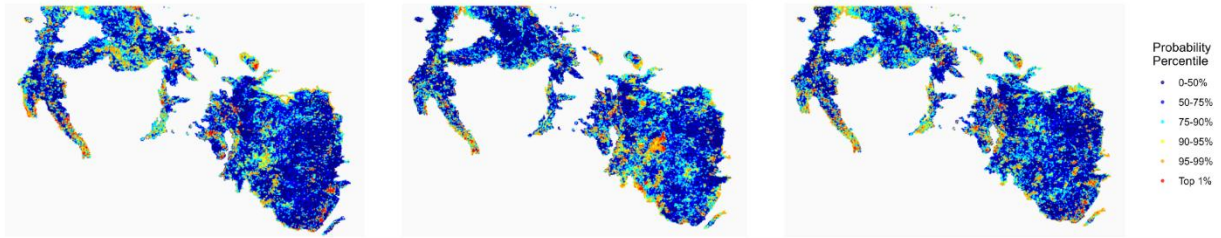
### *Combined Dataset Performance*

These Random Forest models displayed strong performance with distinct temporal patterns (Table 4). AUC values increased from 0.885 to 0.953 in 2016-2019, with slight decline to 0.948 in 2020-2022. TSS values rose to 0.855 in 2012-2015, then declined but remained above the 0.6 threshold in all periods. Cohen's Kappa peaked at 0.839 in 2012-2015. OOB error rates showed a bimodal pattern across time periods, with the lowest error rate in 2012-2015 (4.8%) and higher rates in 2008-2011 (7.2%) and the two later periods (9.4% and 9.7%). Moran's I declined over time from modest early values to near zero in recent years, reflecting improved spatial structure capture.

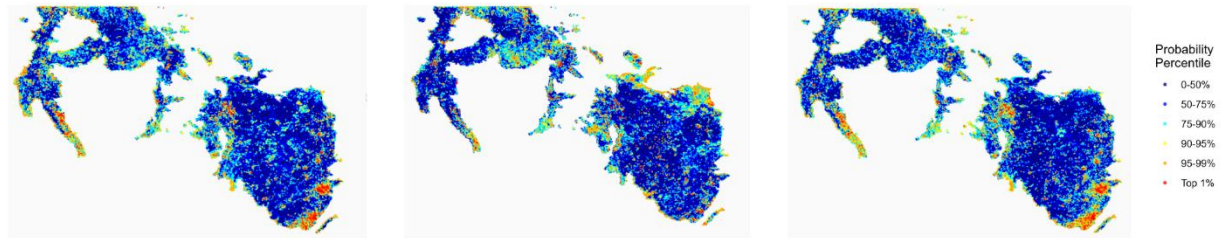
### *Random Forest Heatmaps*

Random Forest species distribution predictions showed remarkable temporal stability across all datasets and time periods (Figure 3). Participatory models maintained consistent spatial patterns throughout all periods. Professional models exhibited similar stability, displaying nearly identical spatial distributions across time. Combined dataset predictions mirrored this pattern of temporal consistency, showing stable species distributions with minimal variation between periods. Unlike MaxEnt models, Random Forest predictions demonstrated little temporal expansion or contraction of suitable habitat, maintaining consistent geographic boundaries and probability distributions across the 15-year study period regardless of increasing sample sizes or changing data composition.

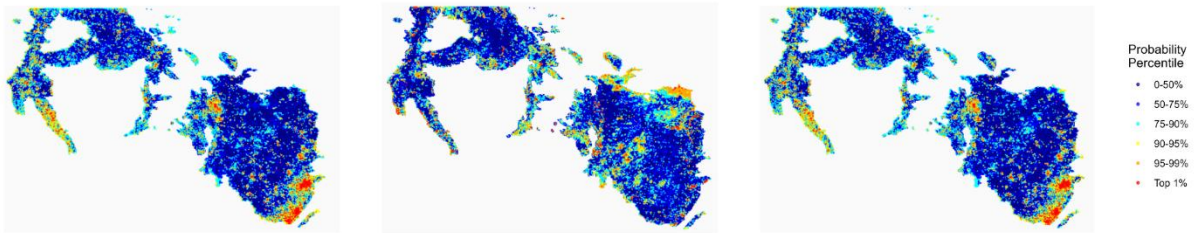
Random Forest results for 2008-2011 in the following order:  
participatory, professional and combined.



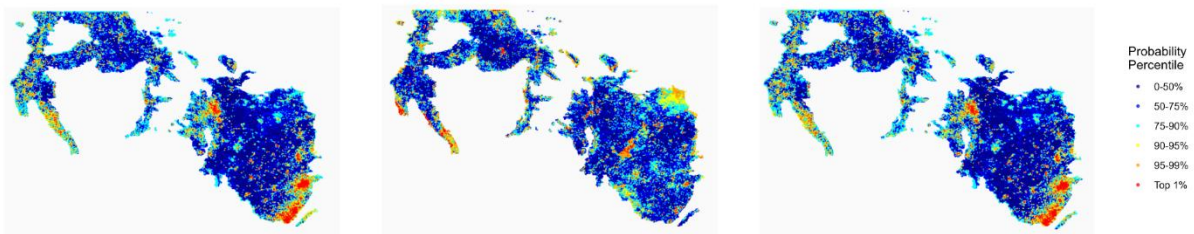
Random Forest results for 2012-2015 in the following order:  
participatory, professional and combined.



Random Forest results for 2016-2019 in the following order:  
participatory, professional and combined.



Random Forest results for 2020-2022 in the following order:  
participatory, professional and combined.



**Figure 3:** Random Forest species distribution predictions for Nymphalidae across four time periods (2008-2011, 2012-2015, 2016-2019, 2020-2022). Each row demonstrates results from participatory (left), professional (center), and combined datasets (right). Colors represent probability percentiles from lowest suitability (dark blue) to highest suitability (red, top 1%) across the Western Cordillera and South Central Semi-Arid Prairies ecoregions.

**Table 4:** Comparative performance metrics for MaxEnt and Random Forest (RF) SDM evaluating Nymphalidae species distribution across four time periods (2008-2011, 2012-2015, 2016-2019, 2020-2022) using participatory, professional, and combined datasets. Metrics include AUC, TSS, and AICc for MaxEnt, and AUC, TSS, Kappa, Moran's I, and OOB Error for Random Forest.

Dataset	Period	MaxEnt AUC	MaxEnt TSS	MaxEnt AICc	RF AUC	RF TSS	RF Kappa	RF Moran's I	RF OOB Error
Participatory	2008-2011	0.804	0.494	107.1	0.8115	0.5278	0.5928	0.054537	0.08835
	2012-2015	0.859	0.594	111.3	0.9292	0.8047	0.8	0.032446	0.06042
	2016-2019	0.867	0.6	92.3	0.9535	0.8129	0.8128	-0.00383	0.09504
	2020-2022	0.873	0.625	72.1	0.9463	0.7778	0.7806	0.006225	0.0978
Professional	2008-2011	0.719	0.4	107	0.8959	0.7456	0.8081	0.021986	0.04737
	2012-2015	0.759	0.463	110.3	0.969	0.8767	0.8839	0.028512	0.02938
	2016-2019	0.752	0.421	117.5	0.9569	0.8844	0.8934	0.033361	0.03068
	2020-2022	0.863	0.668	111.6	0.9849	0.9444	0.9439	0.037674	0.01753
Combined	2008-2011	0.773	0.447	106.4	0.8853	0.735	0.7619	0.035022	0.07227
	2012-2015	0.814	0.517	101.5	0.9482	0.8555	0.8393	0.043175	0.04818
	2016-2019	0.859	0.579	84.2	0.9532	0.8108	0.8108	0.008574	0.09441
	2020-2022	0.851	0.585	73.4	0.948	0.7846	0.7868	0.007289	0.09702

## Discussion

This study demonstrates that participatory and professional data contribute differently to SDM for Nymphalidae, and that combining them can provide complementary benefits. Participatory data increased nearly 80-fold from 1,343 records in 2008-2011 to over 79,000 in 2020-2022, reflecting heightened public engagement. This growth largely reflects the scale and distributed participation of citizen science programs, which often generate more occurrence records than professional surveys (Goldstein et al., 2024; Theobald et al., 2015; Di Cecco et al., 2021), as seen in eBird (Sullivan et al., 2014). Beyond size, participatory science enhances monitoring capacity and public engagement (Cooper et al., 2014; Kays et al., 2020; Bonney, 2021; Callaghan et al., 2021). Professional datasets remained stable (1,226-3,961 records), potentially constrained by funding and staffing, and by 2020-2022 participatory records comprised 96.5% of the combined dataset, highlighting spatial bias and the need for structured participatory approaches (Johnston et al., 2023).

Spatial analyses revealed that participatory and combined datasets were moderately to strongly clustered (Moran's  $I = 0.36-0.65$ ), reflecting possible bias toward accessible areas, whereas professional data were more evenly distributed (Moran's  $I = 0.12-0.31$ ). Strong spatial autocorrelation can violate independence assumptions and inflate apparent model performance (Santos-Fernandez & Mengersen, 2021; Aiello-Lammens et al., 2015; Steen et al., 2021; Boria et al., 2013), and participatory data tend to cluster in urban or accessible regions (Leong & Trautwein, 2019).

MaxEnt models mirrored these differences. Participatory models improved over time (AUC:  $0.804 \rightarrow 0.873$ ; TSS:  $0.494 \rightarrow 0.625$ ) but became diffuse, reflecting

overgeneralization from spatial bias and rapid sample growth. Professional models produced restricted, ecologically targeted predictions with high performance (AUC = 0.863, TSS = 0.668). Combined datasets yielded geographically stable predictions capturing both fine-scale habitat associations and broader coverage (AUC = 0.773-0.859, TSS = 0.447-0.585), though heatmaps still showed overgeneralization from participatory bias. These results indicate that participatory and professional datasets complement rather than replicate each other, with participatory data influenced by accessibility rather than systematic design (Mandeville et al., 2022).

Random Forest outperformed MaxEnt. Professional Random Forest models achieved the highest performance (AUC = 0.984, TSS = 0.944, Cohen's Kappa = 0.944), underscoring the value of structured, high-quality data. Expertise remains important: trained observers often detect more species per survey than volunteers, particularly those difficult to identify, due to skills untrained observers may lack (Farr et al., 2023). However, validated participatory data can produce Random Forest models comparable to professional data (Matutini et al., 2021). Participatory Random Forest models improved over time, peaking in 2016-2019 (AUC = 0.953) before minor declines from clustered, noisy observations. Combined datasets leveraged complementary strengths, providing broader coverage while mitigating some spatial bias, though Random Forest's correlation-based approach may reduce generalizability with heterogeneous sources.

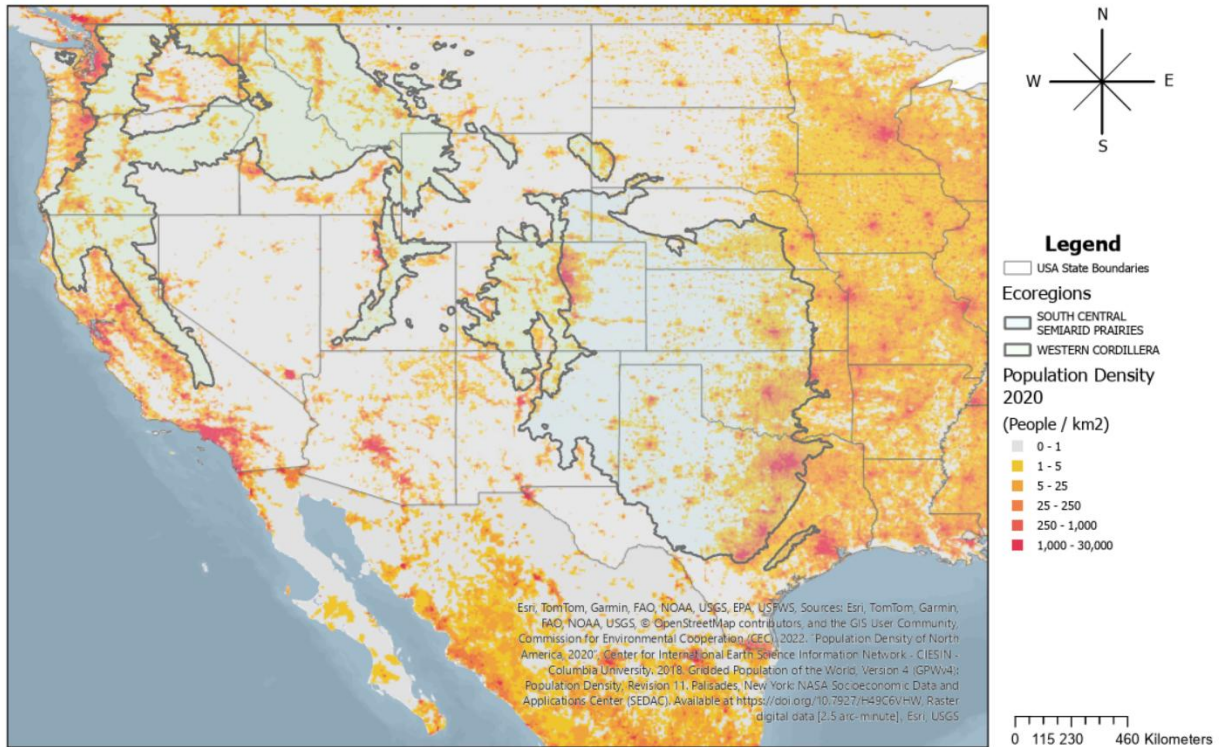
Visual inspection of the heatmaps highlight that MaxEnt predictions from participatory data expanded temporally, likely reflecting sampling artifacts, while professional models remained restricted and high confidence. Combined models

produced intermediate patterns that integrated both sampling approaches, though they maintained the excessive generalizability that obscured fine-scale habitat distinctions. Random Forest predictions showed temporal stability in core habitat identification (e.g., Pacific coast, Rockies, Texas Prairies), indicating reduced susceptibility to sampling bias and better spatial specificity than the overgeneralized MaxEnt distributions. In this study, high-probability areas in participatory models often coincide with dense human populations (Figure 4), indicating observation density may reflect accessibility rather than species distribution.

Debates about participatory data quality persist. Aceves-Bueno et al. (2017) argued that fewer than two-thirds of participatory datasets meet minimum accuracy standards, though critiques note substantial variation across projects and taxa (Specht & Lewandowski, 2018). Integrating participatory and professional data provides broader ecological coverage than either source alone, as highlighted in comparative analyses of insect datasets (Díaz-Calafat et al., 2024).

Overall, SDM differs by data source. Professional data yield high-quality, ecologically realistic models even with limited samples, while participatory data expand coverage, improving performance but potentially introducing clustered and spatial bias (Pocock et al., 2014; Pocock et al., 2023). Integrating datasets with bias-correction techniques (e.g., spatial thinning or block cross-validation) merges broad coverage with high-confidence observations. Future research should refine integration strategies, address spatial autocorrelation, and evaluate ecological realism using independent validation and spatially explicit frameworks. Model choice should align with data: MaxEnt suits presence-only or smaller datasets, capturing broad patterns, whereas

Random Forest handles larger, heterogeneous datasets, producing stable predictions robust to spatial bias.



**Figure 4:** Human population density map for 2020 across the Western Cordillera and South Central Semi-Arid Prairies ecoregions. Colors indicate population density from low (white, 0-1 people/km<sup>2</sup>) to high (dark red, 1,000-30,000 people/km<sup>2</sup>).

## REFERENCES

- Aceves-Bueno, E., Adeleye, A. S., Feraud, M., Huang, Y., Tao, M., Yang, Y., & Anderson, S. E. (2017). The accuracy of citizen science data: a quantitative review. *Bulletin of the Ecological Society of America*, 98(4), 278-290.
- Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., & Anderson, R. P. (2015). spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, 38(5), 541-545.
- Allouche, O., Tsoar, A., & Kadmon, R. (2006). Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, 43(6), 1223-1232.
- Araújo, M. B., & Luoto, M. (2007). The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography*, 16(6), 743-753.
- Araújo, M. B., Anderson, R. P., Márcia Barbosa, A., Beale, C. M., Dormann, C. F., Early, R., ... & Rahbek, C. (2019). Standards for distribution models in biodiversity assessments. *Science Advances*, 5(1), eaat4858.
- Arenas-Castro, S., Goncalves, J., Alves, P., Alcaraz-Segura, D., & Honrado, J. P. (2018). Assessing the multi-scale predictive ability of ecosystem functional attributes for species distribution modelling. *PLoS ONE*, 13(6), e0199292.
- Arif, S., & MacNeil, M. A. (2022). Predictive models aren't for causal inference. *Ecology Letters*, 25(8), 1741-1745.
- Baker, B., Gupta, O., Raskar, R., & Naik, N. (2017). Accelerating neural architecture search using performance prediction. *arXiv preprint arXiv:1705.10823*.
- Baker, E., Drury, J. P., Judge, J., Roy, D. B., Smith, G. C., & Stephens, P. A. (2021). The verification of ecological citizen science data: current approaches and future possibilities. *Citizen Science: Theory and Practice*, 6(1).
- Barends, J. M., Pietersen, D. W., Zambatis, G., Tye, D. R., & Maritz, B. (2020). Sampling bias in reptile occurrence data for the Kruger National Park. *Koedoe: African Protected Area Conservation and Science*, 62(1), 1-9.
- Baston, D. (2020). exactextractr: Fast Extraction from Raster Datasets using Polygons. R package version 0.5.0. <https://CRAN.R-project.org/package=exactextractr>
- Beck, J., Böller, M., Erhardt, A., & Schwanghart, W. (2014). Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics*, 19, 10-15.
- Booth, T. H., Nix, H. A., Busby, J. R., & Hutchinson, M. F. (2014). BIOCLIM: the first species distribution modelling package, its early applications and relevance to most current MAXENT studies. *Diversity and Distributions*, 20(1), 1-9.
- Bonney, R. (2021). Expanding the impact of citizen science. *BioScience*, 71(5), 448-451.

- Boria, R. A., Olson, L. E., Goodman, S. M., & Anderson, R. P. (2014). Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling*, 275, 73-77.
- Bove, D. J., Anderson, H. A., Smith, M. A., & Kuhn, T. A. (2024). Nesting bald eagle population numbers, density, territorial resources, and relationship to human development in northern Colorado's Front Range. *Western Birds*, 55(1).
- Brown, E. D., & Williams, B. K. (2019). The potential for citizen science to produce reliable and useful information in ecology. *Conservation Biology*, 33(3), 561-569.
- Callaghan, C. T., Poore, A. G., Mesaglio, T., Moles, A. T., Nakagawa, S., Roberts, C., ... & Cornwell, W. K. (2021). Three frontiers for the future of biodiversity research using citizen science data. *BioScience*, 71(1), 55-63.
- Castañeda, S., Botello, F., Sánchez-Cordero, V., & Sarkar, S. (2019). Spatio-temporal distribution of monarch butterflies along their migratory route. *Frontiers in Ecology and Evolution*, 7, 400.
- Chandler, M., See, L., Copas, K., Bonde, A. M., López, B. C., Danielsen, F., ... & Turak, E. (2017). Contribution of citizen science towards international biodiversity monitoring. *Biological Conservation*, 213, 280-294.
- Cooper, C. B., Shirk, J., & Zuckerberg, B. (2014). The invisible prevalence of citizen science in global research: migratory birds and climate change. *PLoS ONE*, 9(9), e106508.
- Couvet, D., & Prevot, A. C. (2015). Citizen-science programs: Towards transformative biodiversity governance. *Environmental Development*, 13, 39-45.
- Dasgupta, S., Blankespoor, B., & Wheeler, D. (2024). Toward Better Conservation: A Spatial Analysis of Species Occurrence Data from the Global Biodiversity Information Facility. *Earth System Science Data Discussions*, 2024, 1-27.
- Díaz-Calafat, J., Jaume-Ramis, S., Soacha, K., Álvarez, A., & Piera, J. (2024). Revealing biases in insect observations: A comparative analysis between academic and citizen science data. *PLoS ONE*, 19(7), e0305757.
- Di Cecco, G. J., Barve, V., Belitz, M. W., Stucky, B. J., Guralnick, R. P., & Hurlbert, A. H. (2021). Observing the observers: How participants contribute data to iNaturalist and implications for biodiversity science. *BioScience*, 71(11), 1179-1188.
- Dickinson, J. L., Shirk, J., Bonter, D., Bonney, R., Crain, R. L., Martin, J., ... & Purcell, K. (2012). The current state of citizen science as a tool for ecological research and public engagement. *Frontiers in Ecology and the Environment*, 10(6), 291-297.
- Diengdoh, V. L., Ondei, S., Amin, R. J., Hunt, M., & Brook, B. W. (2023). Landscape functional connectivity for butterflies under different scenarios of land-use, land-cover, and climate change in Australia. *Biological Conservation*, 277, 109825.
- Dormann, C. F., Bobrowski, M., Dehling, D. M., Harris, D. J., Hartig, F., Lischke, H., ... & Kraan, C. (2018). Biotic interactions in species distribution modelling: 10 questions to guide

- interpretation and avoid false conclusions. *Global Ecology and Biogeography*, 27(9), 1004-1016.
- Elith, J., & Leathwick, J. R. (2009). Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology, Evolution, and Systematics*, 40(1), 677-697.
- Farr, C. M., Ngo, F., & Olsen, B. (2023). Evaluating data quality and changes in species identification in a citizen science bird monitoring project. *Citizen Science: Theory and Practice*, 8(1).
- Feldman, M. J., Imbeau, L., Marchand, P., Mazerolle, M. J., Darveau, M., & Fenton, N. J. (2021). Trends and gaps in the use of citizen science derived data as input for species distribution models: A quantitative review. *PLoS ONE*, 16(3), e0234587.
- Ferro, M. L., & Flick, A. J. (2015). "Collection Bias" and the importance of natural history collections in species habitat modeling: A case study using *Thoracophorus costalis* Erichson (Coleoptera: Staphylinidae: Osoriinae), with a critique of GBIF.org. *The Coleopterists Bulletin*, 69(3), 415-425.
- Fisher-Phelps, M., Cao, G., Wilson, R. M., & Kingston, T. (2017). Protecting bias: across time and ecology, open-source bat locality data are heavily biased by distance to protected area. *Ecological Informatics*, 40, 22-34.
- Franklin, J. (2013). Species distribution models in conservation biogeography: developments and challenges. *Diversity and Distributions*, 19(10), 1217-1223.
- Franklin, J. (2023). Species distribution modelling supports the study of past, present and future biogeographies. *Journal of Biogeography*, 50(9), 1533-1545.
- Fritz, S., See, L., & Grey, F. (2022). The grand challenges facing environmental citizen science. *Frontiers in Environmental Science*, 10, 1019628.
- Galván, S., Barrientos, R., & Varela, S. (2022). No bird database is perfect: citizen science and professional datasets contain different and complementary biodiversity information. *Ardeola*, 69(1), 97-114.
- GBIF.org (11 December 2023) GBIF Occurrence Download <https://doi.org/10.15468/dl.nrgwxj>
- Geldmann, J., Heilmann-Clausen, J., Holm, T. E., Levinsky, I., Markussen, B. O., Olsen, K., ... & Tøttrup, A. P. (2016). What determines spatial bias in citizen science? Exploring four recording schemes with different proficiency requirements. *Diversity and Distributions*, 22(11), 1139-1149.
- Goldstein, B. R., Stoudt, S., Lewthwaite, J. M., Shirey, V., Mendoza, E., & Guzman, L. M. (2024). Logistical and preference bias in participatory science butterfly data. *Frontiers in Ecology and the Environment*, 22(8), e2783.
- Guisan, A., & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8(9), 993-1009.

- Guisan, A., Thuiller, W., & Zimmermann, N. E. (2017). *Habitat suitability and distribution models: with applications in R*. Cambridge University Press.
- Hansen, S. E., Monfils, M. J., Hackett, R. A., Goebel, R. T., & Monfils, A. K. (2024). Data-centric species distribution modeling: Impacts of modeler decisions in a case study of invasive European frog-bit. *Applications in Plant Sciences*, *12*(3), e11573.
- Hill, G. M., Kawahara, A. Y., Daniels, J. C., Bateman, C. C., & Scheffers, B. R. (2021). Climate change effects on animal ecology: butterflies and moths as a case study. *Biological Reviews*, *96*(5), 2113-2126.
- Huang, S., Lin, Y., Dong, J., Lin, Y., Su, Z., Li, J., ... & Fu, W. (2024). Relationship between Plant Habitat Types and Butterfly Diversity in Urban Mountain Parks. *Forests*, *15*(8), 1390.
- Jetz, W., McGeoch, M. A., Guralnick, R., Ferrier, S., Beck, J., Costello, M. J., ... & Turak, E. (2019). Essential biodiversity variables for mapping and monitoring species populations. *Nature Ecology & Evolution*, *3*(4), 539-551.
- Johnston, A., Matechou, E., & Dennis, E. B. (2023). Outstanding challenges and future directions for biodiversity monitoring using citizen science data. *Methods in Ecology and Evolution*, *14*(1), 103-116.
- Kays, R., Lasky, M., Parsons, A. W., Pease, B., & Pacifici, K. (2021). Evaluation of the spatial biases and sample size of a statewide citizen science project. *Citizen Science: Theory and Practice*, *6*(1).
- Konowalik, K., & Nosol, A. (2021). Evaluation metrics and validation of presence-only species distribution models based on distributional maps with varying coverage. *Scientific Reports*, *11*(1), 1482.
- Kosmala, M., Wiggins, A., Swanson, A., & Simmons, B. (2016). Assessing data quality in citizen science. *Frontiers in Ecology and the Environment*, *14*(10), 551-560.
- Lazagabaster, I. A., Thomas, C. D., Spedding, J. V., Ikram, S., Solano-Regadera, I., Snape, S., & Bro-Jørgensen, J. (2024). Evaluating species distribution model predictions through time against paleozoological records. *Ecology and Evolution*, *14*(10), e70288.
- Leong, M., & Trautwein, M. (2019). A citizen science approach to evaluating US cities for biotic homogenization. *PeerJ*, *7*, e6879.
- Loiselle, B. A., Howell, C. A., Graham, C. H., Goerck, J. M., Brooks, T., Smith, K. G., & Williams, P. H. (2003). Avoiding pitfalls of using species distribution models in conservation planning. *Conservation Biology*, *17*(6), 1591-1600.
- Mahata, A., Panda, R. M., Dash, P., Naik, A., Naik, A. K., & Palita, S. K. (2023). Microclimate and vegetation structure significantly affect butterfly assemblages in a tropical dry forest. *Climate*, *11*(11), 220.
- Mandeville, C. P., Nilsen, E. B., & Finstad, A. G. (2022). Spatial distribution of biodiversity citizen science in a natural area depends on area accessibility and differs from other recreational area use. *Ecological Solutions and Evidence*, *3*(4), e12185.

- Matutini, F., Baudry, J., Pain, G., Sineau, M., & Pithon, J. (2021). How citizen science could improve species distribution models and their independent assessment. *Ecology and Evolution*, *11*(7), 3028-3039.
- Merow, C., Smith, M. J., & Silander Jr, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, *36*(10), 1058-1069.
- Munisi, E. J., Masenga, E. H., Nkwabi, A. K., Kiwango, H. R., & Mjingo, E. E. (2024). Butterfly Abundance and Diversity in Different Habitat Types in the Usangu Area, Ruaha National Park. *Psyche: A Journal of Entomology*, *2024*(1), 8833655.
- Omernik, J. M. (2004). Perspectives on the nature and definition of ecological regions. *Environmental Management*, *34*(Suppl 1), S27-S38.
- Paseka, R. E., White, L. A., Van de Waal, D. B., Strauss, A. T., González, A. L., Everett, R. A., ... & Borer, E. T. (2020). Disease-mediated ecosystem services: pathogens, plants, and people. *Trends in Ecology & Evolution*, *35*(8), 731-743.
- Petersen, T. K., Speed, J. D., Grøtan, V., & Austrheim, G. (2021). Species data for understanding biodiversity dynamics: The what, where and when of species occurrence data collection. *Ecological Solutions and Evidence*, *2*(1), e12048.
- Pinto-Ledezma, J. N., & Cavender-Bares, J. (2021). Predicting species distributions and community composition using satellite remote sensing predictors. *Scientific Reports*, *11*(1), 16448.
- Pocock, M. J., Chapman, D. S., Sheppard, L. J., & Roy, H. E. (2014). Choosing and using citizen science: A guide to when and how to use citizen science to monitor biodiversity and the environment. NERC/Centre for Ecology & Hydrology. <https://nora.nerc.ac.uk/id/eprint/510644/>
- Pocock, M. J., Hamlin, I., Christelow, J., Passmore, H. A., & Richardson, M. (2023). The benefits of citizen science and nature-noticing activities for well-being, nature connectedness and pro-nature conservation behaviours. *People and Nature*, *5*(2), 591-606.
- Popović, M., Micevski, B., & Verovnik, R. (2021). Effects of elevation gradient and aspect on butterfly diversity on Galičica Mountain in the Republic of Macedonia (south-eastern Europe). *Animal Biodiversity and Conservation*, *44*(1), 67-78.
- Rocchini, D., Tordoni, E., Marchetto, E., Marcantonio, M., Barbosa, A. M., Bazzichetto, M., ... & Malavasi, M. (2023). A quixotic view of spatial bias in modelling the distribution of species and their diversity. *npj Biodiversity*, *2*(1), 10.
- Rodríguez-Castañeda, G., Hof, A. R., Jansson, R., & Harding, L. E. (2012). Predicting the fate of biodiversity using species' distribution models: Enhancing model comparability and repeatability. *PLoS ONE*, *7*(9), e44402. <https://doi.org/10.1371/journal.pone.0044402>

- Santos-Fernandez, E., & Mengersen, K. (2021). Understanding the reliability of citizen science observational data using item response models. *Methods in Ecology and Evolution*, 12(8), 1533-1548.
- Senapathi, D., Carvalheiro, L. G., Biesmeijer, J. C., Dodson, C. A., Evans, R. L., McKerchar, M., ... & Potts, S. G. (2015). The impact of over 80 years of land cover changes on bee and wasp pollinator communities in England. *Proceedings of the Royal Society B: Biological Sciences*, 282(1806), 20150294.
- Shrestha, B. R., Baral, S., Budha-Magar, S., Thapa Magar, K., Gaudel, P., Suwal, S. P., ... & Shrestha, P. (2024). Vegetation productivity determines the response of butterflies along elevation gradients in the trans-Himalayas, Nepal. *Ecosphere*, 15(10), e70019.
- Shirey, V., Larsen, E., Doherty, A., Kim, C. A., Al-Sulaiman, F. T., Hinolan, J. D., ... & Ries, L. (2022). LepTraits 1.0 A globally comprehensive dataset of butterfly traits. *Scientific Data*, 9(1), 382.
- Shultz, A. J., Adams, B. J., Bell, K. C., Ludt, W. B., Pauly, G. B., & Vendetti, J. E. (2021). Natural history collections are critical resources for contemporary and future studies of urban evolution. *Evolutionary Applications*, 14(1), 233-247.
- Spear, D. M., Pauly, G. B., & Kaiser, K. (2017). Citizen science as a tool for augmenting museum collection data from urban areas. *Frontiers in Ecology and Evolution*, 5, 86.
- Specht, H., & Lewandowski, E. (2018). Biased assumptions and oversimplifications in evaluations of citizen science data quality. *Bulletin of the Ecological Society of America*, 99(2), 251-256.
- Steen, V. A., Tingley, M. W., Paton, P. W., & Elphick, C. S. (2021). Spatial thinning and class balancing: Key choices lead to variation in the performance of species distribution models with citizen science data. *Methods in Ecology and Evolution*, 12(2), 216-226.
- Suarez, A. V., & Tsutsui, N. D. (2004). The value of museum collections for research and society. *BioScience*, 54(1), 66-74.
- Sullivan, B. L., Aycrigg, J. L., Barry, J. H., Bonney, R. E., Bruns, N., Cooper, C. B., ... & Kelling, S. (2014). The eBird enterprise: An integrated approach to development and application of citizen science. *Biological Conservation*, 169, 31-40.
- Svancara, L. K., Abatzoglou, J. T., & Waterbury, B. (2019). Modeling current and future potential distributions of milkweeds and the monarch butterfly in Idaho. *Frontiers in Ecology and Evolution*, 7, 168.
- Syfert, M. M., Smith, M. J., & Coomes, D. A. (2013). The effects of sampling bias and model complexity on the predictive performance of MaxEnt species distribution models. *PLoS ONE*, 8(2), e55158.
- Theobald, E. J., Ettinger, A. K., Burgess, H. K., DeBey, L. B., Schmidt, N. R., Froehlich, H. E., ... & Parrish, J. K. (2015). Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. *Biological Conservation*, 181, 236-244.

- Thomas, C. D., Cameron, A., Green, R. E., Bakkenes, M., Beaumont, L. J., Collingham, Y. C., & Williams, S. E. (2004). Extinction risk from climate change. *Nature*, *427*(6970), 145-148.
- Thuiller, W., Pollock, L. J., Gueguen, M., & Münkemüller, T. (2015). From species distributions to meta-communities. *Ecology Letters*, *18*(12), 1321-1328.
- Tong, X. Y., Dong, R., & Zhu, X. X. (2025). Global high categorical resolution land cover mapping via weak supervision. *ISPRS Journal of Photogrammetry and Remote Sensing*, *220*, 535-549.
- Turley, N. E., Kania, S. E., Petitta, I. R., Otruba, E. A., Biddinger, D. J., Butzler, T. M., ... & López-Urbe, M. M. (2024). Bee monitoring by community scientists: comparing a collections-based program with iNaturalist. *Annals of the Entomological Society of America*, *117*(4), 220-233.
- Tzortzakaki, O., Kati, V., Panitsa, M., Tzanatos, E., & Giokas, S. (2019). Butterfly diversity along the urbanization gradient in a densely-built Mediterranean city: Land cover is more decisive than resources in structuring communities. *Landscape and Urban Planning*, *183*, 79-87.
- Wagner, D. L. (2020). Insect declines in the Anthropocene. *Annual Review of Entomology*, *65*(1), 457-480.
- WallisDeVries, M. F., Baxter, W., & Van Vliet, A. J. (2011). Beyond climate envelopes: effects of weather on regional population trends in butterflies. *Oecologia*, *167*(2), 559-571.
- Wells, C. N., & Tonkyn, D. (2018). Changes in the geographic distribution of the Diana fritillary (*Speyeria diana*: Nymphalidae) under forecasted predictions of climate change. *Insects*, *9*(3), 94.
- Zhang, W., Bussmann, R. W., Li, J., Liu, B., Xue, T., Yang, X., ... & Yu, S. (2022). Biodiversity hotspots and conservation efficiency of a large drainage basin: Distribution patterns of species richness and conservation gaps analysis in the Yangtze River Basin, China. *Conservation Science and Practice*, *4*(4), e12653.
- Zhao, Z., Xiao, N., Shen, M., & Li, J. (2022). Comparison between optimized MaxEnt and Random Forest modeling in predicting potential distribution: A case study with *Quasipaa boulengeri* in China. *Science of the Total Environment*, *842*, 156867.

## CHAPTER 2: PATTERNS IN PAINTED LADIES: RANDOM FOREST INSIGHTS INTO DISTRIBUTION AND ENVIRONMENTAL DRIVERS

### Introduction

*Vanessa cardui*, commonly known as the Painted Lady, belongs to the Nymphalidae family and is one of the world's most widespread butterflies, found on every continent except Antarctica, while mainly absent from Australia and South America (Talavera & Vila, 2017). As a generalist pollinator, *Vanessa cardui* contributes to biodiversity by visiting a wide variety of flowers, particularly species in the Asteraceae family (e.g., rabbitbrush, purple coneflower, dandelion, and Canada thistle), along with species of Fabaceae and Scrophulariaceae (Scott, 1992). The caterpillars and adults serve as food for numerous animals and host parasites and parasitoids, while larval host plants, such as thistles, hold additional ecological significance (Stefanescu, 2023; Lampert et al., 2014; Muchoney et al., 2022; Hernandez & Bowers, 2024). The species' mobility, ecological breadth, and availability of occurrence data make it an ideal focal taxon for modeling species distribution across diverse ecoregions in the United States. Importantly, *Vanessa cardui* also functions as a sensitive indicator of ecosystem change, reflecting the effects of habitat fragmentation, drought, agricultural intensification, and climate-driven range shifts (Menchetti et al., 2019; Stefanescu et al., 2007; Peterson et al., 2019; Hu et al., 2021; Mesler & Mabry, 2024; Granato et al., 2024).

Like many insects (Sánchez-Bayo & Wyckhuys, 2019), *Vanessa cardui* is vulnerable to intensifying effects of anthropogenic change. Urban development and

agricultural intensification eliminate essential habitat patches, concentrating populations into fewer locations where disease transmission and demographic stress increase (Chowdhury et al., 2021; Di Mauro, 2004). Climate change exacerbates these pressures by altering temperature and precipitation cycles, with extremes impairing butterfly development and movement (Tüzün et al., 2018; Bristow et al., 2024). Winter periods (December-March) already impose demographic constraints that climate shifts may further intensify (Menchetti et al., 2019). Meanwhile, herbicide-driven thistle declines threaten larval food plants and long-term stability (Chowdhury et al., 2021). Studying a mobile pollinator like *Vanessa cardui* is therefore crucial for understanding landscape connectivity, ecosystem resilience, and how multiple stressors interact across prairie, alpine, and subalpine ecosystems.

Despite its ecological importance, significant gaps remain in understanding *Vanessa cardui* habitat distribution patterns within the United States. While migration behaviors and general distributions are documented (Abbott, 1951; Stefanescu et al., 2007; Talavera & Vila, 2017; Stefanescu et al., 2017), no studies to our current knowledge have applied machine learning approaches such as Random Forest to predict *Vanessa cardui* species distribution across anywhere in Western North America. In contrast, the monarch butterfly (*Danaus plexippus*) has been the focus of numerous modeling efforts that leverage advanced statistical methods (Jepsen et al., 2015; Dingle et al., 2005; Belsky & Joshi, 2018; Svancara et al., 2019; Reppert & de Roode, 2018; Brower et al., 2006; Bartel et al., 2011). This imbalance represents a critical knowledge gap, limiting our ability to assess climate change impacts on a globally important migratory species.

Building on the established effectiveness of Random Forest modeling for species distribution, this study asks: How can Random Forests best predict *Vanessa cardui* species distribution across two diverse landscapes? To answer this, we modeled species distribution in two ecologically distinct regions--the Western Cordillera and South Central Semi-Arid Prairies--while identifying key environmental drivers.

## **Methods**

### Ecoregions and Occurrence Data

I prepared ecoregion data following the same procedures described in Chapter 1 (see Methods, Chapter 1). I restricted occurrence records to professionally verified *Vanessa cardui* observations spanning 2008-2022 (GBIF.org, 2023) and excluded all non-focal species. To assess temporal change, I divided records into two periods: Early (2008-2014) and Late (2016-2022), excluding intermediate years to maximize temporal contrast.

### Environmental Variables

I assembled environmental covariates that largely mirrored those described in Chapter 1, including land cover, bioclimatic variables, elevation, slope, and aspect. To capture additional ecological drivers, I incorporated soil and anthropogenic variables. I obtained soil characteristics from the Soil Survey Geographic Database (SSURGO) via the United States Department of Agriculture Geospatial Data Gateway, derived road density from the 2014 U.S. Census TIGER dataset (accessed through ArcGIS Pro's Living Atlas) and obtained impervious surface data from the National Land Cover Database (NLCD) 2019 Impervious Product. I imported all layers into ArcGIS Pro and

clipped them to the Western Cordillera and South Central Semi-Arid Prairies ecoregions using the "Extract by Mask" tool.

### Background Points and Data Preparation

I generated background points using stratified random sampling with a 5:1 background-to-presence ratio to balance model efficiency and reliability. To reduce temporal sampling bias, I assigned random years within the study range to background points. I screened environmental predictors for multicollinearity and removed variables with Pearson's  $r > 0.85$  using the caret package (Kuhn, 2008). I retained only complete cases for modeling. I stratified occurrence and background data by time period to preserve temporal representation during training and evaluation.

### Random Forest Modeling

I implemented Random Forest modeling in R using the randomForest package (Liaw, 2002). I trained models with 500 decision trees, setting the number of variables considered at each split (mtry) to the square root of the total number of predictors. I applied bootstrap sampling and used out-of-bag (OOB) samples for internal validation. I also partitioned data into training (80%) and testing (20%) subsets using stratified sampling to balance presence and background points across time periods.

### Model Evaluation

I assessed model performance using both OOB estimates and independent test sets. I calculated accuracy, sensitivity, specificity, precision, F1-score, the Area Under the Curve (AUC), and the True Skill Statistic (TSS) and Cohen's Kappa. I quantified

variable importance using mean decrease in accuracy and mean decrease in Gini impurity and ranked predictors accordingly.

### Species Distribution

I generated predicted species distribution across the study area on a continuous scale from 0 to 1 and categorized values into five classes: Very Low (0.0-0.2), Low (0.2-0.4), Moderate (0.4-0.6), High (0.6-0.8), and Very High (0.8-1.0). I calculated distributional change between Early (2008-2014) and Late (2016-2022) periods as absolute and proportional differences in predicted abundance per grid cell. To control for sampling bias, I extracted residuals from linear regression models and used them as bias-corrected measures of distributional change. I then calculated correlations between environmental predictors and both raw and residual change metrics to identify potential ecological drivers.

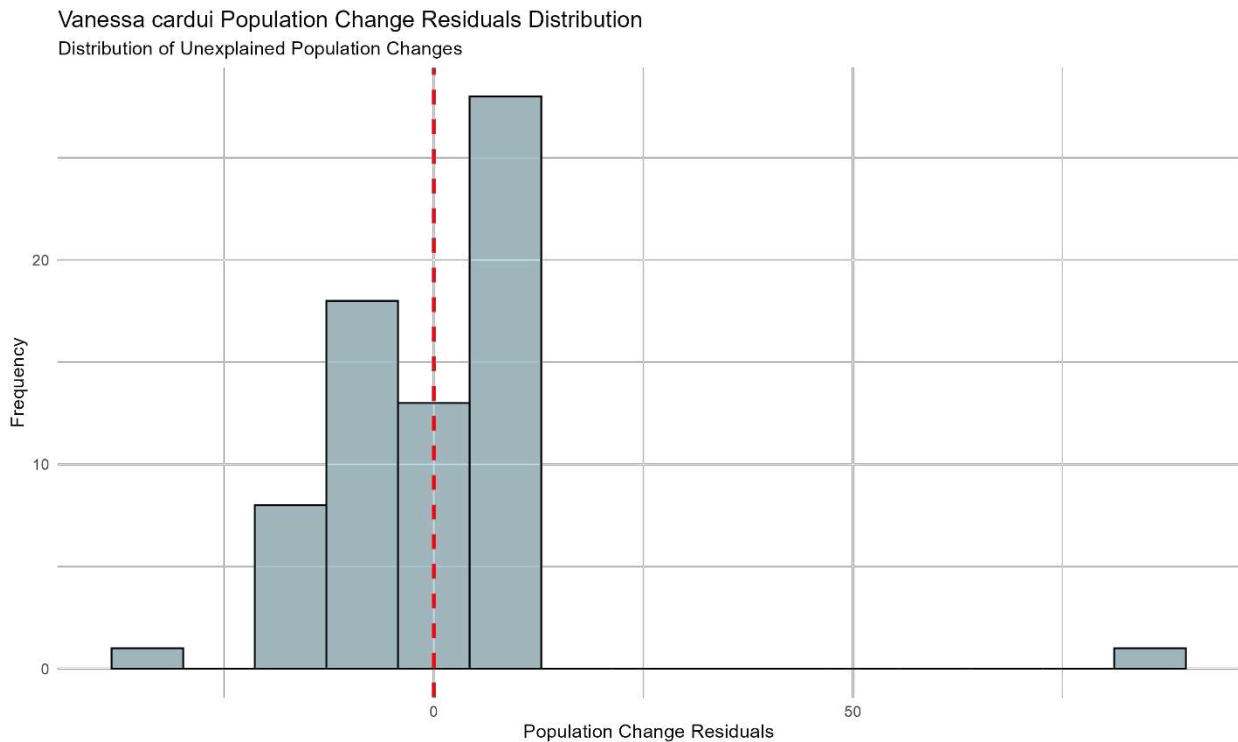
### Visualization and Output

I visualized results using heatmaps of model performance, predictor correlations, and variable importance. I applied hexagonal aggregation to highlight spatial patterns in predicted suitability and reduce ambiguity (Birch et al., 2007). I conducted analyses in R using the randomForest, caret, pROC, dplyr, and ggplot2 packages (Liaw, 2002; Kuhn, 2008; Robin et al., 2011; Wickham et al., 2025; Wickham, 2016).

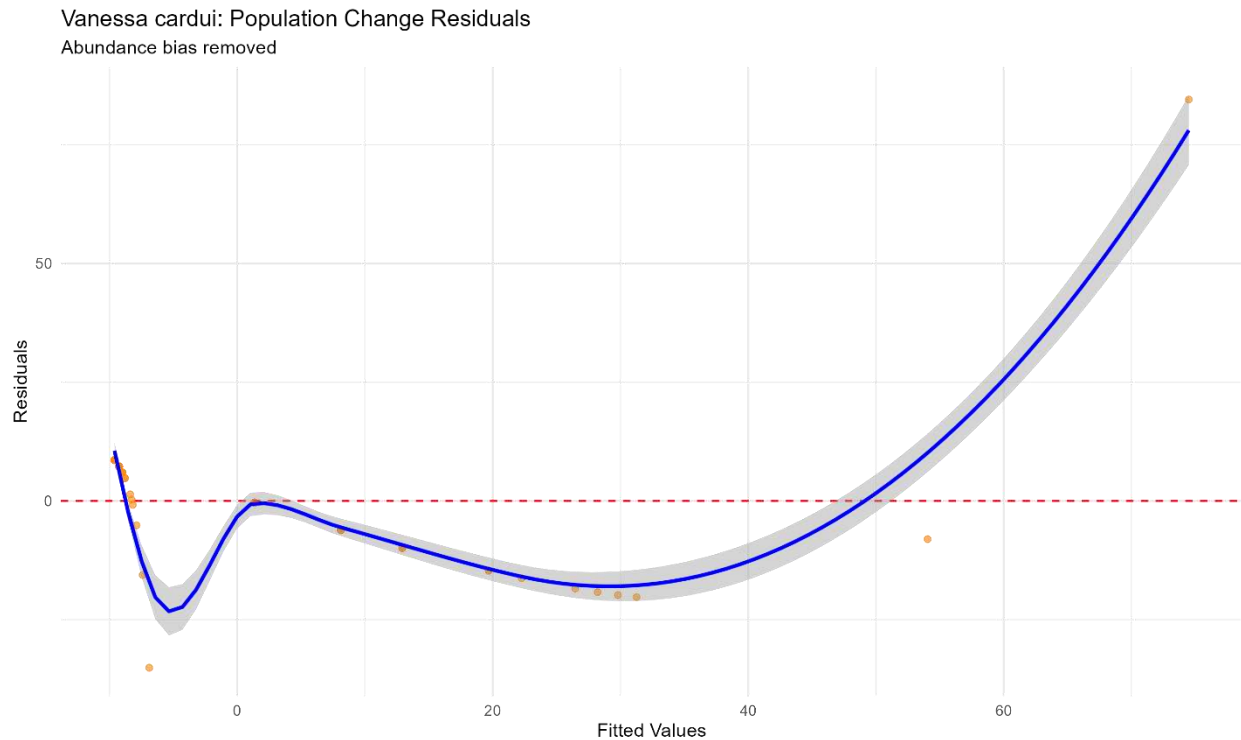
## **Results**

Residual diagnostics revealed systematic patterns in model performance (Figures 5 and 6). The distribution of residuals was approximately normal and centered near zero (Figure 5), indicating that the model captured central tendencies effectively for

most predictions. However, in the residuals versus fitted values plot (Figure 6), systematic deviations were evident at the extremes, with underestimation at low suitability values and overestimation at high values. The few extreme residual values visible in the distribution (Figure 5) correspond to these systematic deviations, indicating limitations in predicting rare distributional outcomes.



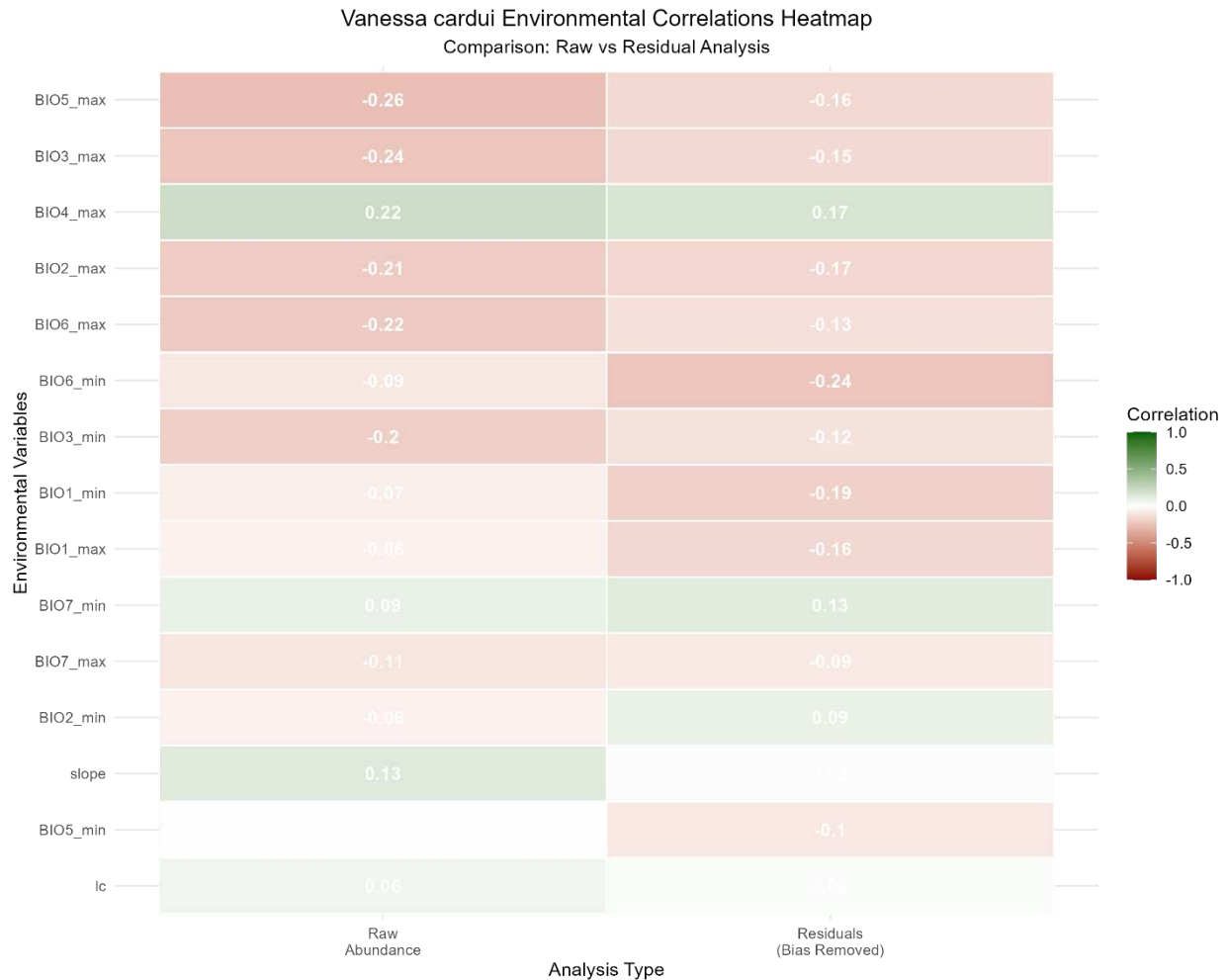
**Figure 5:** Distribution of population change residuals for *Vanessa cardui*. Histogram shows the frequency distribution of unexplained population changes (residuals) from the SDM. The red dashed line indicates zero change, with most residuals clustered near zero and few extreme values at the distribution tails.



**Figure 6:** Residuals vs. fitted values plot for *Vanessa cardui* population change model. Scatter plot showing model residuals (with abundance bias removed) against fitted values, with locally weighted smoothing line (blue) and 95% confidence interval (gray shading). The red dashed line indicates zero residuals.

Removing abundance-related bias through residual extraction revealed altered environmental relationships with *Vanessa cardui* population change compared to raw abundance data (Figure 7). Several bioclimatic variables emerged as positively associated with residuals, with temperature seasonality (BIO4\_max,  $r = 0.17$ ) showing the strongest positive correlation, followed by diurnal temperature range (BIO7\_min,  $r = 0.14$ ) and mean diurnal range (BIO2\_min,  $r = 0.09$ ). Minimum temperature of the coldest month (BIO6\_min) shifted from near-zero correlation ( $r = 0.09$ ) in raw data to a negative

correlation ( $r = -0.24$ ) in residuals, while annual mean temperature (BIO1\_min and BIO1\_max) also showed stronger negative correlations ( $r = -0.19$  and  $r = -0.18$ , respectively). Conversely, maximum temperature of the warmest month (BIO5\_max) decreased in correlation strength from  $r = -0.26$  to  $r = -0.16$ , and isothermality (BIO3\_max) weakened from  $r = -0.24$  to  $r = -0.15$ .

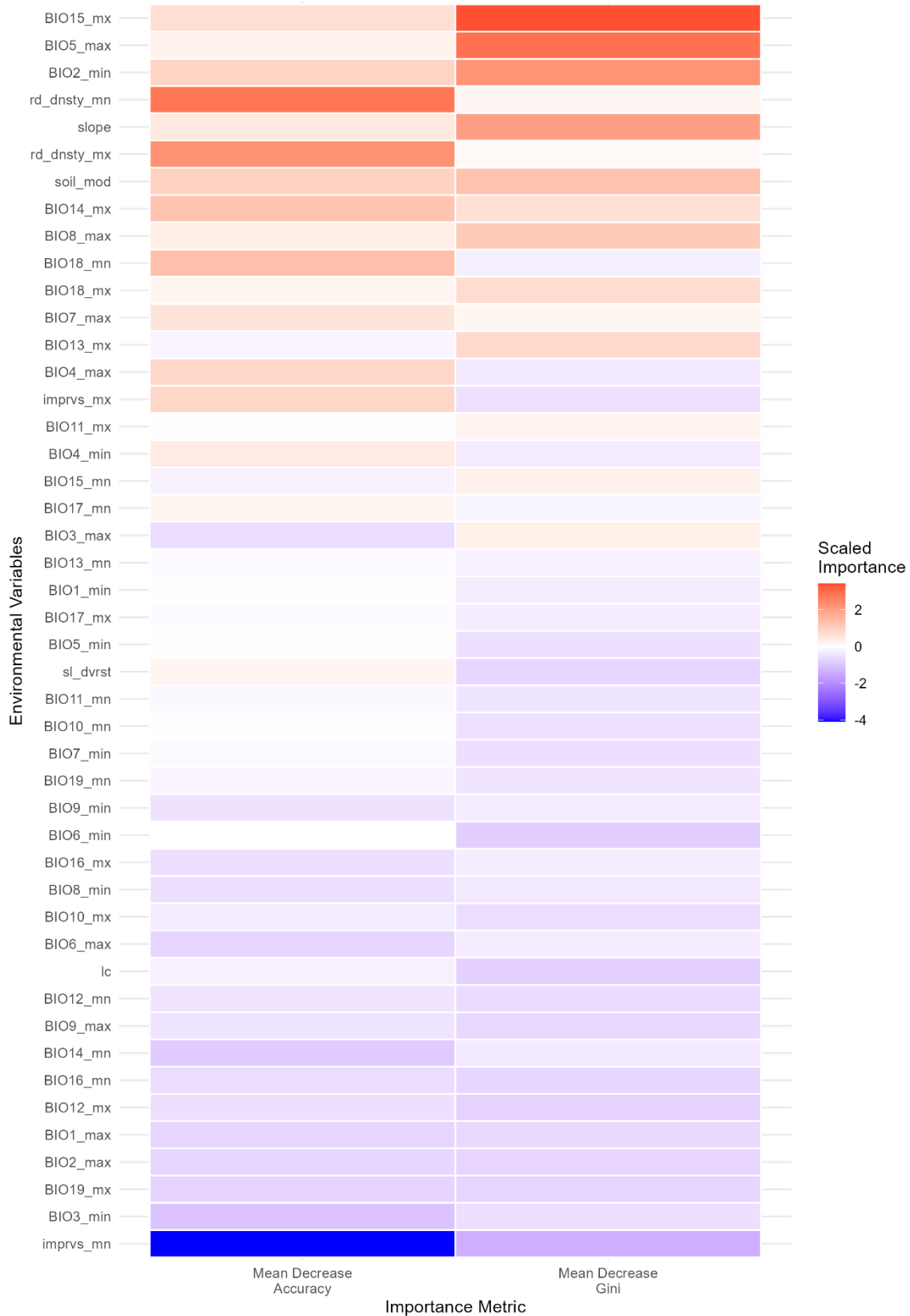


**Figure 7:** Environmental correlations with *Vanessa cardui* population change, comparing raw abundance data versus residuals with abundance bias removed. Correlation coefficients are shown for the top 15 environmental variables, with green indicating positive correlations and red indicating negative correlations. The residual

analysis (right column) reveals environmental relationships after controlling for abundance-related biases.

RF variable importance analysis identified key environmental predictors driving species distribution patterns (Figure 8). Precipitation seasonality (BIO15\_max) and maximum temperature of the warmest month (BIO5\_max) emerged as the most important variables based on mean decrease in accuracy, followed by minimum mean diurnal range (BIO2\_min). Road density variables (both minimum and maximum) also ranked highly, indicating the importance of anthropogenic landscape features. The standardized importance heatmap revealed that Gini-based importance measures generally aligned with accuracy-based measures, though with some notable differences in ranking, particularly for slope and soil characteristics which showed higher Gini importance relative to their accuracy contributions.

### Vanessa cardui Variable Importance Heatmap Standardized Importance Metrics



**Figure 8:** Variable importance heatmap for *Vanessa cardui* species distribution model showing standardized importance metrics. Variables are ranked by mean decrease in accuracy (left column) and mean decrease in Gini (right column). Red colors indicate high importance while blue indicates lower relative importance. The top variables include precipitation seasonality (BIO15\_max), temperature variables (BIO5\_max, BIO2\_min), and anthropogenic features (road density).

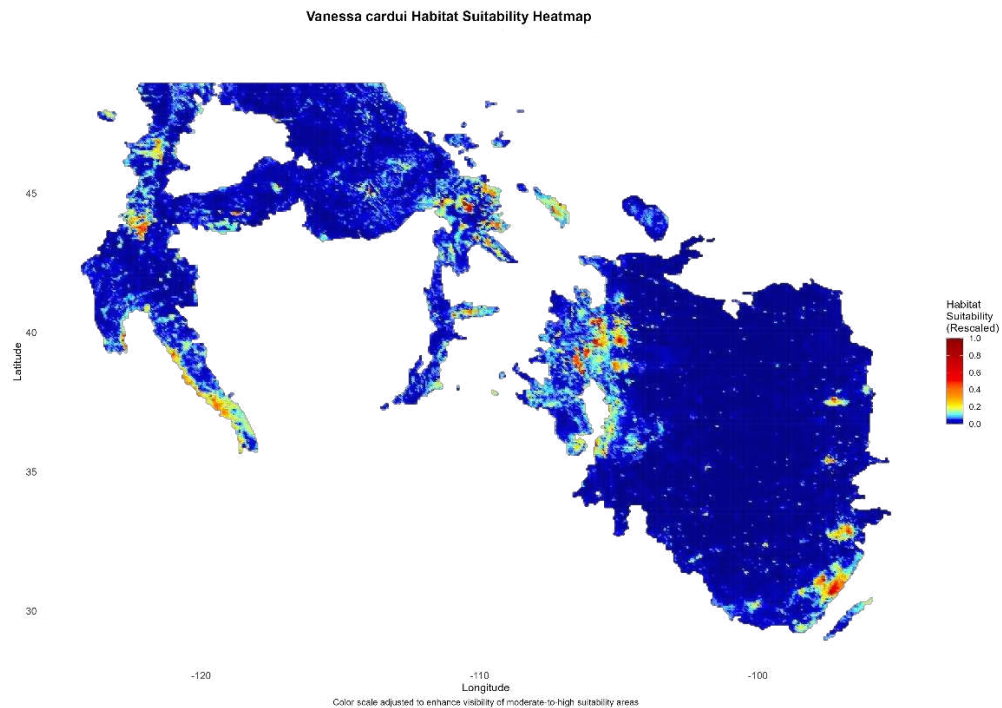
Model evaluation metrics indicated strong predictive performance. The Random Forest model achieved an AUC of 0.968, a True Skill Statistic of 0.895, a Kappa of 0.924, an out-of-bag error of 1.27%, and a test accuracy of 97.93%.

**Table 6:** Evaluation metrics for the Random Forest model of *Vanessa cardui* species distribution, including AUC, TSS, Cohen’s Kappa, out-of-bag error, and test accuracy.

<b>Metric</b>	<b>Value</b>
<b>AUC</b>	0.968
<b>TSS</b>	0.895
<b>Kappa</b>	0.924
<b>OOB Error (%)</b>	1.27
<b>Test Accuracy (%)</b>	97.93

Between early (2008-2014) and late (2016-2022) periods, a map of species distribution revealed persistence of *Vanessa cardui* occurrence across the Western Cordillera and South Central Semi-Arid Prairies (Figure 9). Much of the study area shows low suitability (0.0-0.2), with suitable habitat representing a small proportion of

the total landscape. Additionally, some suitable areas overlap densely populated regions (such as California's Central Valley, Denver metropolitan area, and the Texas Triangle which encompasses cities like Dallas and Houston), while others occur in less developed regions (e.g., Yellowstone National Park).



**Figure 9:** Species distribution model predictions for *Vanessa cardui* across western North America. The rescaled species distribution heatmap demonstrates predicted occurrence probability from Random Forest modeling, with values compressed (0-0.2) and expanded (0.2-1.0) to enhance visualization of moderate to high suitability areas. Red areas indicate highest species distribution (0.8-1.0), while blue areas represent lowest species distribution (0.0-0.2).

## Discussion

The Random Forest models effectively captured major distributional dynamics of *Vanessa cardui* across the Western Cordillera and South Central Semi-Arid Prairies, with high accuracy and discriminatory power underscoring the strength of the framework. However, the resulting systematic residual patterns highlight challenges in predicting rare outcomes. To better interpret these systematic errors and disentangle true ecological signals from site and data specific influences, we examined residual correlations in relation to population changes.

The correlation analysis of population change residuals provides compatible insights into these patterns. The weaker residual correlations suggest that raw abundance values partly reflect site-level density differences, inflating apparent climate-population links. This agrees with recent work showing that abundance indices often misrepresent true population size (Numminen et al., 2023). Residuals help correct this bias, isolating subtler and more ecologically meaningful climate relationships.

The residual-based correlations align with ecological expectations for *Vanessa cardui*. A positive response to temperature seasonality and diurnal variation is typical for migratory butterflies, as marked temporal variability generates windows of suitable conditions such as nectar resources and host plant availability (Hu et al., 2021; Saldivar et al., 2022; Stefanescu et al., 2017). Negative associations with warmer baselines suggest that consistently warm conditions may constrain growth, likely via phenological mismatches, reduced diapause cues, or physiological costs (Huang et al., 2024; von Schmalensee et al., 2024; Granato et al., 2024). Residuals also indicate that maximum heat is less limiting than abundance patterns suggested, while thermal stability plays a

minor role for this thermally flexible, migratory species. Overall, *Vanessa cardui* appears to perform better in regions where temperature fluctuates substantially around cooler baselines, favoring environments with pronounced climatic variability over consistently warm or thermally stable areas such as tropical lowlands (Mesler & Mabry, 2024; Saldivar et al., 2022; Stefanescu et al., 2017). These patterns suggest that migratory life history strategies are closely tied to environments with high seasonality, relationships that become clearer once abundance-related biases are removed.

Mapping these ecological tendencies geographically, the combined heatmap (2008–2014 and 2016–2022) revealed distinct spatial patterns: some suitable areas overlap densely populated regions (e.g., California’s Central Valley, Denver, Dallas, Houston, Oklahoma City), while many others occur in less developed landscapes (e.g., Yellowstone National Park and the Sangre de Cristo Mountains). These hotspots in remote mountainous and prairie regions suggest monitoring captured ecologically representative habitats, but the concentration of records near urban centers indicates that spatial sampling bias likely remains a limitation affecting predictions in less surveyed regions. Variable importance analyses show that both climatic factors (i.e., precipitation, temperature) and anthropogenic variables (i.e., impervious surface coverage, road density) shaped suitability, along with soil characteristics. The apparent positive association with impervious surfaces likely reflects how human-modified landscapes create the microclimatic and resource heterogeneity that aligns with this species’ preference for variable environments, through ornamental plantings, green spaces, and edge habitats (Schueller et al., 2023; Kc, 2025; Nason, 2021) not a direct benefit of impervious surfaces themselves.

These results emphasize the importance of maintaining connectivity across the broader landscape to support migratory movements, with persistent high suitability areas providing clear conservation priorities. Areas of declining suitability may benefit from restoration or management to mitigate anthropogenic pressures, particularly where climate remains suitable, but connectivity may have been compromised. Important limitations must also be acknowledged. The analysis does not directly model abundance or density, and occurrence records reflect variation in reporting effort and detection probability in addition to potential spatial sampling bias. Some of the systematic residual patterns indicate reduced reliability in predicting rare outcomes, which is particularly relevant for conservation focused on range margins or refugia.

Overall, this study demonstrates how methodological rigor can sharpen ecological understanding and improve conservation outcomes. By employing residual analysis to correct for abundance-related biases, we uncovered climate-population relationships that raw occurrence data obscured: specifically, that *Vanessa cardui* thrives in environments with high thermal variability rather than consistently warm or stable conditions. This insight, made visible through diagnostic approaches, has direct conservation implications. It suggests that maintaining landscape connectivity across regions with pronounced seasonality should be prioritized over focusing solely on warm, thermally stable areas that superficially appear suitable based on uncorrected abundance patterns. Furthermore, the identification of persistent high-suitability corridors in both remote mountainous terrain and human-modified landscapes underscores the need for conservation strategies that span the rural-urban gradient. As climate change potentially reduces thermal variability in some regions, while increasing

it in others, this framework provides a replicable approach for identifying refugia and migration corridors for other wide-ranging species. Ultimately, these findings illustrate that rigorous analytical frameworks--particularly those that account for observational biases--are essential for translating species distribution models into actionable conservation priorities that reflect true ecological requirements rather than artifacts of sampling design.

## REFERENCES

- Abbott, C. H. (1951). A quantitative study of the migration of the painted lady butterfly, *Vanessa cardui* L. *Ecology*, 32(2), 155-171.
- Ahlström, A., Raupach, M. R., Schurgers, G., Smith, B., Arneth, A., Jung, M., ... & Zeng, N. (2015). The dominant role of semi-arid ecosystems in the trend and variability of the land CO<sub>2</sub> sink. *Science*, 348(6237), 895-899.
- Bauer, D. M., & Wing, I. S. (2010). Economic consequences of pollinator declines: a synthesis. *Agricultural and Resource Economics Review*, 39(3), 368-383.
- Bartel, R. A., Oberhauser, K. S., De Roode, J. C., & Altizer, S. M. (2011). Monarch butterfly migration and parasite transmission in eastern North America. *Ecology*, 92(2), 342-351.
- Belsky, J., & Joshi, N. K. (2018). Assessing role of major drivers in recent decline of monarch butterfly population in North America. *Frontiers in Environmental Science*, 6, 86.
- Birch, C. P., Oom, S. P., & Beecham, J. A. (2007). Rectangular and hexagonal grids used for observation, experiment and simulation in ecology. *Ecological Modelling*, 206(3-4), 347-359.
- Bonebrake, T. C., Ponisio, L. C., Boggs, C. L., & Ehrlich, P. R. (2010). More than just indicators: a review of tropical butterfly ecology and conservation. *Biological Conservation*, 143(8), 1831-1841.
- Boyes, D. H., Evans, D. M., Fox, R., Parsons, M. S., & Pocock, M. J. (2021). Street lighting has detrimental impacts on local insect populations. *Science Advances*, 7(35), eabi8322.
- Boyle, J. H., Dalgleish, H. J., & Puzey, J. R. (2019). Monarch butterfly and milkweed declines substantially predate the use of genetically modified crops. *Proceedings of the National Academy of Sciences*, 116(8), 3006-3011.
- Bradford, J. B., Schlaepfer, D. R., Lauenroth, W. K., & Palmquist, K. A. (2020). Robust ecological drought projections for drylands in the 21st century. *Global Change Biology*, 26(7), 3906-3919.
- Bristow, L. V., Grundel, R., Dzurisin, J. D., Wu, G. C., Li, Y., Hildreth, A., & Hellmann, J. J. (2024). Warming experiments test the temperature sensitivity of an endangered butterfly across life history stages. *Journal of Insect Conservation*, 28(1), 1-13.
- Brower, L. P., Brower, J. V. Z., & Cranston, F. P. (1968). Courtship behavior of the queen butterfly, *Danaus gilippus berenice* (Cramer). *Zoologica*, 53(1), 1-39.
- Brower, L. P., Fink, L. S., & Walford, P. (2006). Fueling the fall migration of the monarch butterfly. *Integrative and Comparative Biology*, 46(6), 1123-1142.

- Brower, L. P., Taylor, O. R., Williams, E. H., Slayback, D. A., Zubieta, R. R., & Ramirez, M. I. (2012). Decline of monarch butterflies overwintering in Mexico: is the migratory phenomenon at risk?. *Insect Conservation and Diversity*, 5(2), 95-100.
- CaraDonna, P. J., Iler, A. M., & Inouye, D. W. (2014). Shifts in flowering phenology reshape a subalpine plant community. *Proceedings of the National Academy of Sciences*, 111(13), 4916-4921.
- Cowie, R. H., Bouchet, P., & Fontaine, B. (2022). The Sixth Mass Extinction: fact, fiction or speculation?. *Biological Reviews*, 97(2), 640-663.
- Crewe, T. L., & McCracken, J. D. (2015). Long-term trends in the number of monarch butterflies (Lepidoptera: Nymphalidae) counted on fall migration at Long Point, Ontario, Canada (1995-2014). *Annals of the Entomological Society of America*, 108(5), 707-717.
- Crossley, M. S., Smith, O. M., Berry, L. L., Phillips-Cosio, R., Glassberg, J., Holman, K. M., ... & Snyder, W. E. (2021). Recent climate change is creating hotspots of butterfly increase and decline across North America. *Global Change Biology*, 27(12), 2702-2714.
- Dalton, R. M., Underwood, N. C., Inouye, D. W., Soulé, M. E., & Inouye, B. D. (2023). Long-term declines in insect abundance and biomass in a subalpine habitat. *Ecosphere*, 14(8), e4620.
- Dell, D. E. N. N. I. S., Sparks, T. H., & Dennis, R. L. (2005). Climate change and the effect of increasing spring temperatures on emergence dates of the butterfly *Apatura iris* (Lepidoptera: Nymphalidae). *European Journal of Entomology*, 102(2), 161-167.
- DeVries, P. J. (1987). *The butterflies of Costa Rica and their natural history: Papilionidae, Pieridae, Nymphalidae*.
- Dingle, H., Zalucki, M. P., Rochester, W. A., & Armijo-Prewitt, T. (2005). Distribution of the monarch butterfly, *Danaus plexippus* (L.) (Lepidoptera: Nymphalidae), in western North America. *Biological Journal of the Linnean Society*, 85(4), 491-500.
- Ehrlich, P. R., & Raven, P. H. (1964). Butterflies and plants: A study in coevolution. *Evolution*, 18(4), 586-608. <https://doi.org/10.2307/2406212>
- Feinsinger, P. (1983). Coevolution and pollination.
- Fiedler, A. K., Landis, D. A., & Arduser, M. (2012). Rapid shift in pollinator communities following invasive species removal. *Restoration Ecology*, 20(5), 593-602.
- Forister, M. L., McCall, A. C., Sanders, N. J., Fordyce, J. A., Thorne, J. H., O'Brien, J., ... & Shapiro, A. M. (2010). Compounded effects of climate change and habitat alteration shift patterns of butterfly diversity. *Proceedings of the National Academy of Sciences*, 107(5), 2088-2092. <https://doi.org/10.1073/pnas.0909686107>

- Fox, R., Brereton, T. M., Asher, J., August, T. A., Botham, M. S., Bourn, N. A. D., Cruickshanks, K. L., Bulman, C. R., Ellis, S., Harrower, C. A., Middlebrook, I., Noble, D. G., Powney, G. D., Randle, Z., Warren, M. S., & Roy, D. B. (2015). *The State of the UK's Butterflies 2015*. Butterfly Conservation and the Centre for Ecology & Hydrology.
- Fox, R., Dennis, E. B., Purdy, K. M., Middlebrook, I., Roy, D. B., Noble, D. G., Botham, M. S., & Bourn, N. A. D. (2023). *The State of the UK's Butterflies 2022*. Butterfly Conservation.
- Flockhart, D. T., Brower, L. P., Ramirez, M. I., Hobson, K. A., Wassenaar, L. I., Altizer, S., & Norris, D. R. (2017). Regional climate on the breeding grounds predicts variation in the natal origin of monarch butterflies overwintering in Mexico over 38 years. *Global Change Biology*, 23(7), 2565-2576.
- Furlong, M. J., & Zalucki, M. P. (2017). Climate change and biological control: the consequences of increasing temperatures on host-parasitoid interactions. *Current Opinion in Insect Science*, 20, 39-44.
- GBIF.org (11 December 2023) GBIF Occurrence Download  
<https://doi.org/10.15468/dl.nrgwxj>
- Gibbons, S. M., Lekberg, Y., Mummey, D. L., Sangwan, N., Ramsey, P. W., & Gilbert, J. A. (2017). Invasive plants rapidly reshape soil properties in a grassland ecosystem. *mSystems*, 2(2), 10-1128.
- Granato, C., Campera, M., & Bulbert, M. (2024). Sensitivity of *Vanessa cardui* to temperature variations: A cost-effective experiment for environmental education. *Insects*, 15(4), 221.
- Hald-Mortensen, C. (2023). The Main Drivers of Biodiversity Loss: A Brief Overview. *Journal of Ecology and Natural Resources*, 7(3), 000346.
- Hernandez, K., & Bowers, M. D. (2024). Food and time: dietary plasticity of different sources of a generalist insect herbivore. *Journal of Insect Science*, 24(2), 15.
- Hu, G., Stefanescu, C., Oliver, T. H., Roy, D. B., Brereton, T., Van Swaay, C., ... & Chapman, J. W. (2021). Environmental drivers of annual population fluctuations in a trans-Saharan insect migrant. *Proceedings of the National Academy of Sciences*, 118(26), e2102762118.
- Inamine, H., Ellner, S. P., Springer, J. P., & Agrawal, A. A. (2016). Linking the continental migratory cycle of the monarch butterfly to understand its population decline. *Oikos*, 125(8), 1081-1091.
- Inouye, D. W. (2022). Climate change and phenology. *Wiley Interdisciplinary Reviews: Climate Change*, 13(3), e764.

- Jepsen, S., Berglundh, T., Genco, R., Aass, A. M., Demirel, K., Derks, J., ... & Zitzmann, N. U. (2015). Primary prevention of peri-implantitis: Managing peri-implant mucositis. *Journal of Clinical Periodontology*, *42*, S152-S157.
- Júnior, G. D. B. F., & Diniz, I. R. (2015). Temporal dynamics of fruit-feeding butterflies (Lepidoptera: Nymphalidae) in two habitats in a seasonal Brazilian environment. *Florida Entomologist*, *98*(4), 1207-1216.
- Kc, S. (2025). An Untapped and Undocumented Butterfly Diversity in a Rapidly Urbanizing and Fragmenting Forest Habitat in Pokhara, Nepal: First Checklist and Implications for Conservation and Ecotourism. *Ecology and Evolution*, *15*(8), e71937.
- Kogan, F. (2023). The IPCC Reports on Global Warming and Land Changes. In *Remote Sensing Land Surface Changes: The 1981-2020 Intensive Global Warming* (pp. 67-79). Cham: Springer International Publishing.
- Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*, 1-26.
- Kulmatiski, A., & Beard, K. H. (2011). Long-term plant growth legacies overwhelm short-term plant growth effects on soil microbial community structure. *Soil Biology and Biochemistry*, *43*(4), 823-830.
- Lampert, T., Müters, S., Stolzenberg, H., Kroll, L. E., & KiGGS Study Group. (2014). Messung des sozioökonomischen Status in der KiGGS-Studie. *Bundesgesundheitsblatt-Gesundheitsforschung-Gesundheitsschutz*, *57*(7), 762-770.
- Liang, H., He, Y. D., Theodorou, P., & Yang, C. F. (2023). The effects of urbanization on pollinators and pollination: A meta-analysis. *Ecology Letters*, *26*(9), 1629-1642.
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18-22.
- Malcolm, S. B. (2018). Anthropogenic impacts on mortality and population viability of the monarch butterfly. *Annual Review of Entomology*, *63*, 277-302.
- Markevich, A., Jones, S. R., Markevich, C. G., & Seastedt, T. R. (2023). Cheatgrass, Mammals, Birds, Butterflies, and Wildfire: A Study of Ecosystem Interactions.
- McKinney, M. L. (2006). Urbanization as a major cause of biotic homogenization. *Biological Conservation*, *127*(3), 247-260.
- Menchetti, M., Guéguen, M., & Talavera, G. (2019). Spatio-temporal ecological niche modelling of multigenerational insect migrations. *Proceedings of the Royal Society B*, *286*(1910), 20191583.
- Mesler, S. P., & Mabry, K. E. (2024). Effects of temperature experienced across life stages on morphology and flight behavior of painted lady butterflies (*Vanessa cardui*). *Movement Ecology*, *12*(1), 76.

- Middleton, A. D., Kauffman, M. J., McWhirter, D. E., Cook, J. G., Cook, R. C., Nelson, A. A., ... & Klaver, R. W. (2013). Animal migration amid shifting patterns of phenology and predation: lessons from a Yellowstone elk herd. *Ecology*, *94*(6), 1245-1256.
- Mirzabaev, A., Stringer, L. C., Benjaminsen, T. A., Gonzalez, P., Harris, R., Jafari, M., Stevens, N., Tirado, C. M., & Zakieldean, S. (2022). Cross-Chapter Paper 3: Deserts, semiarid areas and desertification. In H.-O. Pörtner, D. C. Roberts, M. Tignor, E. S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Lösschke, V. Möller, A. Okem, & B. Rama (Eds.), *Climate change 2022: Impacts, adaptation and vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp. 2195-2231). Cambridge University Press.  
<https://doi.org/10.1017/9781009325844.020>
- Mirzabaev, A., Wu, J., Evans, J., García-Oliva, F., Hussein, I. A., Iqbal, M. H., ... & Weltz, M. (2019). Desertification.
- Mitchell, R. J., Irwin, R. E., Flanagan, R. J., & Karron, J. D. (2009). Ecology and evolution of plant-pollinator interactions. *Annals of Botany*, *103*(9), 1355-1363.
- Muchoney, N. D., Bowers, M. D., Carper, A. L., Mason, P. A., Teglas, M. B., & Smilanich, A. M. (2022). Use of an exotic host plant shifts immunity, chemical defense, and viral burden in wild populations of a specialist insect herbivore. *Ecology and Evolution*, *12*(3), e8723.
- Myers, M. C., Mason, J. T., Hokschi, B. J., Cambardella, C. A., & Pfrimmer, J. D. (2015). Birds and butterflies respond to soil-induced habitat heterogeneity in experimental plantings of tallgrass prairie species managed as agroenergy crops in Iowa, USA. *Journal of Applied Ecology*, *52*(5), 1176-1187.
- Nacua, A. E., Clemente, K. J., Macalalad, E. P., Deocariz, C. C., & Galvez, M. C. (2020). Urban diversity of butterflies as a biological indicator of an air quality in Manila, Philippines. *Asian Journal of Conservation Biology*, *9*(2), 315-321.
- Nason, Lindsay D., "Saving butterflies in the city: gardens as conservation spaces in urban landscapes." (2021). *Electronic Theses and Dissertations*. Paper 3643.
- Numminen, E., Jokinen, M., Lindén, A., & Vanhatalo, J. (2023). Species ecology can bias population estimates. *Biological Conservation*, *283*, 110115.
- Öckinger, E., Schweiger, O., Crist, T. O., Debinski, D. M., Krauss, J., Kuussaari, M., ... & Bommarco, R. (2010). Life-history traits predict species responses to habitat area and isolation: A cross-continental synthesis. *Ecology Letters*, *13*(8), 969-979.  
<https://doi.org/10.1111/j.1461-0248.2010.01487.x>
- Omernik, J. M. (2004). Perspectives on the nature and definition of ecological regions. *Environmental Management*, *34*(Suppl 1), S27-S38.
- Osborne, A., Longden, M., Bourke, D., & Coulthard, E. (2022). Bringing back the Manchester Argus *Coenonympha tullia* ssp. *davus* (Fabricius, 1777): Quantifying

- the habitat resource requirements to inform the successful reintroduction of a specialist peatland butterfly. *Ecological Solutions and Evidence*, 3(2), e12147.
- Pardikes, N. A., Shapiro, A. M., Dyer, L. A., & Forister, M. L. (2015). Global weather and local butterflies: variable responses to a large-scale climate pattern along an elevational gradient. *Ecology*, 96(11), 2891-2901.
- Parlin, A. F., Stratton, S. M., & Guerra, P. A. (2022). Oriented migratory flight at night: Consequences of nighttime light pollution for monarch butterflies. *iScience*, 25(5).
- Paschke, M. W., Redente, E. F., Klein, D. A., & Smith, L. (2005). Integrated control and assessment of knapweed and cheatgrass on Department of Defense installations. 2003 annual report. SERDP CS1145. SERDP Support Office, HydroGeoLogic Inc, 1155.
- Pepin, N. C., Arnone, E., Gobiet, A., Haslinger, K., Kotlarski, S., Notarnicola, C., ... & Adler, C. (2022). Climate changes and their elevational patterns in the mountains of the world. *Reviews of Geophysics*, 60(1), e2020RG000730.
- Peterson, E. M., Shaw, K. R., & Smith, P. N. (2019). Toxicity of agrochemicals among larval painted lady butterflies (*Vanessa cardui*). *Environmental Toxicology and Chemistry*, 38(12), 2629-2636.
- Pievani, T. (2014). The sixth mass extinction: Anthropocene and the human impact on biodiversity. *Rendiconti Lincei*, 25(1), 85-93.
- Pleasants, J. M., & Oberhauser, K. S. (2013). Milkweed loss in agricultural fields because of herbicide use: effect on the monarch butterfly population. *Insect Conservation and Diversity*, 6(2), 135-144.
- Pörtner, H. O., Roberts, D. C., Masson-Delmotte, V., Zhai, P., Tignor, M., Poloczanska, E., & Weyer, N. M. (2019). The ocean and cryosphere in a changing climate. *IPCC special report on the ocean and cryosphere in a changing climate*, 1155, 10-1017.
- Potts, S. G., Ngo, H. T., Biesmeijer, J. C., Breeze, T. D., Dicks, L. V., Garibaldi, L. A., ... & Vanbergen, A. (2016). *The assessment report of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services on pollinators, pollination and food production*.
- Pritekel, C., Whittemore-Olson, A., Snow, N., & Moore, J. C. (2006). Impacts from invasive plant species and their control on the plant community and belowground ecosystem at Rocky Mountain National Park, USA. *Applied Soil Ecology*, 32(1), 132-141.
- Thogmartin, W. E., Diffendorfer, J. E., López-Hoffman, L., Oberhauser, K., Pleasants, J., Semmens, B. X., ... & Wiederholt, R. (2017). Density estimates of monarch butterflies overwintering in central Mexico. *PeerJ*, 5, e3221.

- Rahbek, C., Borregaard, M. K., Colwell, R. K., Dalsgaard, B. O., Holt, B. G., Morueta-Holme, N., ... & Fjeldså, J. (2019). Humboldt's enigma: What causes global patterns of mountain biodiversity?. *Science*, *365*(6458), 1108-1113.
- Reppert, S. M., & de Roode, J. C. (2018). Demystifying monarch butterfly migration. *Current Biology*, *28*(17), R1009-R1022.
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J. C., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*(1), 77.
- Romme, W. H. (1982). Fire and landscape diversity in subalpine forests of Yellowstone National Park. *Ecological Monographs*, *52*(2), 199-221.
- Saldivar, J. L. A., Romero, A. N., & Wilson Rankin, E. E. (2022). Community science reveals high diversity of nectaring plants visited by painted lady butterflies (Lepidoptera: Nymphalidae) in California sage scrub. *Environmental Entomology*, *51*(6), 1141-1149.
- Sánchez-Bayo, F., & Wyckhuys, K. A. (2019). Worldwide decline of the entomofauna: A review of its drivers. *Biological Conservation*, *232*, 8-27.
- Saraf, K. K., & Vijaykumar, K. (2020). An overview on Butterflies significance. *International Journal of Entomology Research*, *5*(5), 97-104.
- Schirmel, J., Bundschuh, M., Entling, M. H., Kowarik, I., & Buchholz, S. (2016). Impacts of invasive plants on resident animals across ecosystems, taxa, and feeding types: a global assessment. *Global Change Biology*, *22*(2), 594-603.
- Schueller, S. K., Li, Z., Bliss, Z., Roake, R., & Weiler, B. (2023). How informed design can make a difference: supporting insect pollinators in cities. *Land*, *12*(7), 1289.
- Schweiger, O., Dormann, C. F., Bailey, D., & Frenzel, M. (2006). Occurrence pattern of *Pararge aegeria* (Lepidoptera: Nymphalidae) with respect to local habitat suitability, climate and landscape structure. *Landscape Ecology*, *21*, 989-1001.
- Scott, J. A. (1992). *The butterflies of North America: a natural history and field guide*. Stanford University Press.
- Shukla, I., Gaynor, K. M., Worm, B., & Darimont, C. T. (2023). The diversity of animals identified as keystone species. *Ecology and Evolution*, *13*(10), e10561.
- Sims, S. R. (2017). *Speyeria* (Lepidoptera: Nymphalidae) Conservation. *Insects*, *8*(2), 45.
- Souther, S. K., Sandor, M. E., Sample, M., Gabrielson, S., & Aslan, C. E. (2024). Bee and butterfly records indicate diversity losses in western and southern North America, but extensive knowledge gaps remain. *PLoS ONE*, *19*(5), e0289742.
- Spaniol, R. L., Mendonca Jr, M. D. S., Hartz, S. M., Iserhard, C. A., & Stevens, M. (2020). Discolouring the Amazon Rainforest: how deforestation is affecting butterfly coloration. *Biodiversity and Conservation*, *29*(9), 2821-2838.

- Steen, V. A., Tingley, M. W., Paton, P. W., & Elphick, C. S. (2021). Spatial thinning and class balancing: Key choices lead to variation in the performance of species distribution models with citizen science data. *Methods in Ecology and Evolution*, 12(2), 216-226.
- Stefanescu, C., Alarcón, M., & Àvila, A. (2007). Migration of the painted lady butterfly, *Vanessa cardui*, to north-eastern Spain is aided by African wind currents. *Journal of Animal Ecology*, 888-898.
- Stefanescu, C., PUIG-MONTSERRAT, X. A. V. I. E. R., Samraoui, B., Izquierdo, R., Ubach, A., & Arrizabalaga, A. (2017). Back to Africa: autumn migration of the painted lady butterfly *Vanessa cardui* is timed to coincide with an increase in resource availability. *Ecological Entomology*, 42(6), 737-747.
- Stefanescu, C., Ubach, A., & Wiklund, C. (2021). Timing of mating, reproductive status and resource availability in relation to migration in the painted lady butterfly. *Animal Behaviour*, 172, 145-153.
- Sunde, J., Franzén, M., Betzholtz, P. E., Francioli, Y., Pettersson, L. B., Pöyry, J., ... & Forsman, A. (2023). Century-long butterfly range expansions in northern Europe depend on climate, land use and species traits. *Communications Biology*, 6(1), 601.
- Talavera, G., & Vila, R. (2017). Discovery of mass migration and breeding of the painted lady butterfly *Vanessa cardui* in the Sub-Saharan: the Europe-Africa migration revisited. *Biological Journal of the Linnean Society*, 120(2), 274-285.
- Thomas, J. A. (2005). Monitoring change in the abundance and distribution of insects using butterflies and other indicator groups. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1454), 339-357.  
<https://doi.org/10.1098/rstb.2004.1585>
- Tüzün, N., Op de Beeck, L., Oliarinony, R., Van Dievel, M., & Stoks, R. (2018). Warming under seminatural outdoor conditions in the larval stage negatively affects insect flight performance. *Biology Letters*, 14(5), 20180121.
- Vanbergen, A. J., & Insect Pollinators Initiative. (2013). Threats to an ecosystem service: pressures on pollinators. *Frontiers in Ecology and the Environment*, 11(5), 251-259.
- von Schmalensee, L., Süess, P., Roberts, K. T., Gotthard, K., & Lehmann, P. (2024). A quantitative model of temperature-dependent diapause progression. *Proceedings of the National Academy of Sciences*, 121(36), e2407057121.
- Wagner, D. L. (2020). Insect declines in the Anthropocene. *Annual Review of Entomology*, 65(1), 457-480.
- Warren, M. S., Maes, D., van Swaay, C. A., Goffart, P., Van Dyck, H., Bourn, N. A., ... & Ellis, S. (2021). The decline of butterflies in Europe: Problems, significance, and possible solutions. *Proceedings of the National Academy of Sciences*, 118(2), e2002551117.

- Wei, F., Xie, T., Su, C., He, B., Shu, Z., Zhang, Y., ... & Hao, J. (2024). Stability and Assembly Mechanisms of Butterfly Communities across Environmental Gradients of a Subtropical Mountain. *Insects*, 15(4), 230.
- Wepprich, T., Adrion, J. R., Ries, L., Wiedmann, J., & Haddad, N. M. (2019). Butterfly abundance declines over 20 years of systematic monitoring in Ohio, USA. *PLoS ONE*, 14(7), e0216270.
- Westerling, A. L., Turner, M. G., Smithwick, E. A., Romme, W. H., & Ryan, M. G. (2011). Continued warming could transform Greater Yellowstone fire regimes by mid-21st century. *Proceedings of the National Academy of Sciences*, 108(32), 13165-13170.
- Wickham, H., Chang, W., & Wickham, M. H. (2016). Package 'ggplot2'. Create elegant data visualisations using the grammar of graphics. Version, 2(1), 1-189.
- Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., ... & van Den Brand, T. (2007). ggplot2: Create elegant data visualisations using the grammar of graphics. (*No Title*).