

**Assessing Predictive Ability of Species-Area Relationship Models**

by

Philip J. Yoo

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

**Bachelor of Science with Honours in Biology-Psychology**

Supervisor(s): Dr. Jeff Houlahan, Department of Biology

This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit: <http://creativecommons.org/licenses/by/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

THE UNIVERSITY OF NEW BRUNSWICK

SAINT JOHN

April, 2022

©Philip Yoo, 2022

## **ABSTRACT**

The species-area relationship (SAR) is a well-established concept but there is still limited understanding of its predictive ability. I investigated the predictive ability of four SAR models (species/area, log-species/log-area, species/log-area, and log-species/area) to new data using linear regression models. I collected 84 SAR datasets and broke them into training and test sets. For each training set I estimated the slope and intercept for each of the 4 SAR models and then used these estimates to predict species richness in each of the 84 test sets. The predictive ability of SAR models for 68 of the 84 was more accurate than the mean. Whether I was assessing the transferability of SAR models in space (i.e., a model built using data from one geographical location in order to predict onto another location) or non-spatially (i.e., model built using randomized sites to predict richness to sites nearby) the predictive ability was similar.

## **ACKNOWLEDGEMENTS**

I would like to thank Dr. Jeff Houlahan, my supervisor, for giving me the opportunity to do research under his wing and for providing me with all the necessary feedback and guidance throughout this project.

## **STATEMENT OF RESEARCH CONTRIBUTION**

I began researching and collecting data on species-area relationship during the summer of 2021. In October I wrapped up data collection and in December 2021 I began building and testing models. I completed the testing for all 84 SAR models in January 2022 and submitted the final draft in March 2022.

## Table of Contents

ABSTRACT.....	ii
DEDICATION.....	<b>Error! Bookmark not defined.</b>
ACKNOWLEDGEMENTS.....	iii
STATEMENT OF RESEARCH CONTRIBUTION.....	iv
Table of Contents.....	<b>Error! Bookmark not defined.</b>
List of Tables.....	vii
List of Figures.....	viii
List of Symbols, Nomenclature or Abbreviations.....	x
Introduction.....	<b>Error! Bookmark not defined.</b>
Overview of species-area relationship.....	1
Plausible Causes of SAR.....	2
Complete knowledge.....	3
Methods.....	3
Data collection + criteria.....	4
The four SAR Models.....	5
Null Model.....	6
Predictive Ability.....	6
Testing for Complete Understanding.....	9
Results.....	11
All 84 datasets.....	11
Spatial datasets.....	15
Non-spatial datasets.....	17

Discussion.....	21
Predictive Ability of SAR models.....	21
Comparison of the four SAR models.....	22
Overall complete understanding.....	23
Study Limitations.....	23
Conclusion.....	24
References.....	25
Appendix.....	<b>Error!</b>
<b>Bookmark not defined.</b>	<b>7</b>

## List of Tables

Table 1: The most accurately predicting model for each dataset (84 total) and information such as taxonomic group, R2 value, prediction error (model and mean), prediction improvement of SA model from mean, number of sites per dataset, mean species richness per dataset, mean area per dataset and slope coefficient.....37

## List of Figures

Fig.1: Study Area in southern Brazil conducted by Rolon et al. (2008), comprising a total of 15 sampled wetlands for macrophyte richness. The red lines indicate the separation of the training (1-11) and testing (12-15) datasets, with sites 1-11 predicting to sites 12-15. ....8

Fig. 2. Visualization of total knowledge in SA models; left hand side is the null model (mean or median) with a prediction error of 8 (representing ‘zero knowledge’); middle represents the training model with a prediction error of 5; right hand side with prediction error of 0.....9

Fig 3: Prediction error for the most accurately predicting SAR model across 84 datasets.....19

Fig 4: The overall predictive improvement across all 84 datasets (spatial (1-48) + non-spatial (49-84) combined) .....20

Fig 5: Relationship between R2 on the training dataset and training dataset prediction error across all 84 datasets.  
.....22

Fig 6. Relationship between total number of sites (per dataset) and predictive improvement for most accurately predicting species-area models across 84 datasets.....23

Fig. 7. Prediction errors across 48 datasets (spatially extrapolated) from the four species-area relationship models which were selected from the most accurately predicting model per dataset..... 24

Fig 8. Relationship between R2 on the training dataset (spatial) and training dataset prediction error.....25

Fig 9. Relationship between total number of sites (per dataset) and predictive improvement for most accurately predicting species-area models for spatial data..... 26



Fig 10. Prediction errors across 36 datasets (non- spatially extrapolated) from the four species-  
area relationship models which were selected from the most accurately predicting model per  
dataset.....27

Fig 11. Relationship between R2 on the training dataset (non-spatial) and training dataset  
prediction error. .... 28

Fig 12. Relationship between total number of sites (per dataset) and predictive improvement for  
most accurately predicting species-area models for non-spatial data.....29

## **List of Symbols, Nomenclature or Abbreviations**

SAR: Species-area relationship

SA: Species-area

PE: prediction error

## Introduction

The causes and consequences of biological diversity and more particularly species richness have been intensively studied in ecology (He & Legendre, 2002). Numerous studies reveal a generally positive relationship between species richness and ecosystem functions (Cardinale et al., 2006). For instance, there is evidence that increasing species richness may enhance the aggregate rate of resource uptake (Yachi & Loreau, 1999); species diversity thereby mediates community stability through compensating interactions to environmental fluctuations among species (McNaughton, 1977). However, the key biological or physical processes that influence species richness are not well understood (He & Legendre, 2002). One potentially important driver of species richness is the amount of available habitat (i.e., area).

The species-area relationship (SARs) is one of the most intensely studied patterns in ecology (Ladle & Whittaker, 2011). The empirical evidence shows that, across habitats and taxonomic groups, as patch size increases so does species richness (Lomolino, 2000). The positive species-area relationship has been found in a variety of patches/sites, including lakes (Zhao & Zhou, 2018), forest patches (De Camargo & Currie, 2015), and wetlands (Benassi et al., 2007). In addition, it has been found across many taxonomic groups such as birds (Benassi et al., 2007), mammals (Brunet & Medellin, 2001), and plants (Li & Baiser, 2018). Several causal mechanisms have been proposed for SARs (Connor & McCoy, 1979) including habitat diversity, passive sampling (Connor & McCoy, 1979) and extinction and colonization rates (MacArthur & Wilson, 1967).

Three plausible explanations for the positive species-area relationship are (i) extinction/colonization, (ii) habitat diversity and (iii) resource concentration. The theory of

island biogeography asserts that species richness is determined by extinction and colonization rates because large areas have lower extinction rates and higher colonization rates (MacArthur & Wilson, 1967). The habitat diversity hypothesis states that large patches contain more habitats than small patches, which allows large patches to support species requiring multiple habitats and to support a wider range of species that require a particular habitat (Williams, 1964). The resource concentration hypothesis proposes that large patches will support more individuals and therefore more species because they contain large amounts of resources (e.g., areas of high plant density). These proposed mechanisms are not mutually exclusive and may work in combination to cause positive SARs (Connor & McCoy, 1979). In recent decades, ecologists have applied SARs to wildlife conservation.

Conservation biologists frequently rely on species-area relationships to predict changes in species diversity under scenarios of habitat loss (Shafer, 1990). A study by Brooks et al. (1997) used the species-area relationship to predict the number of bird species that would go globally extinct following deforestation on islands; they found that the predicted number of extinctions from deforestation compared to the number of highly threatened species (i.e., species near extinction) to be similar. Another study by Ulrich (2005) applied the SAR approach using the grid model to predict species loss of south European butterflies after habitat destruction; it was found that the SAR works reasonably well at predicting species numbers for the butterflies at high degrees of habitat loss. Although the SAR have been proven to be useful, there are still uncertainties around its predictive ability to new data.

Ecological models contain all of our ecological knowledge and the amount of knowledge in a model can only be assessed by its predictive ability (Houlahan et al., 2017). Hundreds of species-area models have been fit but few, if any, have been applied to new datasets. This

implies that we know little about how much understanding/knowledge is contained in our species-area models.

In theory, a model that perfectly describes the relationship between each driver and the phenomenon of interest, should make perfect predictions (Houlahan et al., 2017). Therefore, models with perfect knowledge are free of prediction errors (Houlahan et al., 2017). The null model on the other hand excludes biological mechanisms and functions as a simple statistical descriptor for detecting a pattern (Gotelli & McGill, 2006). Since the null model contain zero mechanistic knowledge, this model represents zero knowledge with respect to species-richness. The scientific progress can be measured as the (Null prediction error – SAR model prediction error)/Null prediction error) to quantify knowledge in models.

The objective of this study is to evaluate the predictive ability of species-area models to new datasets. The predictive ability of species-area models will provide estimates of how much knowledge these models contain and how much progress ecologists are making towards understanding diversity and species richness.

## **Methods**

### **Data search and inclusion criteria**

I collected eighty-four species-area datasets using Google Scholar and EBSCO and targeted specific databases such as Wildlife and Ecology Studies Worldwide, Biological & Agricultural Index Plus, and BIOSIS. Within these databases, I targeted specific journals including Ecology, Journal of Biogeography, Journal of Applied Ecology, and Conservation Biology. At the beginning of data collection, I used broad combinations of keywords including

“species richness/diversity,” “species-area relationships (SARs)”, and “species-area model”.

However, my search terms refined over time to capture as many studies as possible. The majority of my refined terms were related to species-area relationships for specific taxonomic groups such as “species-area relationship and plants,” “avian richness and landscape structure”, “diversity of marine invertebrates and area”, etc. For a paper to meet my inclusion criteria, it must include species richness and area data for >10 sites.

### **Data Extraction**

I extracted primary and ancillary variables. If a primary variable was missing or there were 10 or fewer patches, the study was excluded:

*Primary:*

- Species richness data.
- Patch area (with units).
- The taxonomic group (e.g., small mammals, amphibians)

*Ancillary:* (Note: the study is to be kept even if one or more of these variables are missing)

- Spatial coordinates (longitude + latitude or UTM coordinates)
- The patch type (e.g., grassland).
- The country or continent where the study took place.
- Duration of study.
- The citation for the paper the data came from.

Data were obtained from tables in the article or from data repositories identified in the article. Otherwise, data was extracted from figures using ImageJ software. If spatial coordinates were not provided in the dataset, it was noted whether a map was provided in the paper to be manually extracted in the later phases of data collection.

## **Models**

I fitted species-area and null models for predicting species richness.

### ***Species-area models***

There are several approaches to modelling the relationship between species richness and patch area.

I modelled the linear relationship between species and area in four ways:

1.  $SR = a + b * AREA$
2.  $LogSR = a + b * LogAREA$
3.  $LogSR = a + b * AREA$
4.  $SR = a + b * LogAREA$

Where: SR = species richness, AREA = patch size. LogSR and LogAREA = the natural log transformations of species richness and patch size, respectively. Each of these models identifies patch size (i.e., AREA) as a causal driver of species richness and thus, have the potential to contain knowledge about the causes of variability in species richness. If area causes variability in species richness we would expect species-area models to make better predictions than a null model.

### Null models:

I defined the null model in two ways – as the mean or the median of the training dataset. Since the null model is the mean/median of the training set, this model implies no causal mechanism causing fluctuations in species richness and is therefore, defined as containing ‘zero-knowledge.’ If patch size does not contribute to variability in species richness, we would expect predictions using these null models to be as good or better than predictions using species-area models.

### **Predictive Ability**

Predictive ability can only be assessed by making predictions to ‘new’ data (i.e., data that were not used to estimate the slope and intercept parameters for the species-area models). This implies separating the data into a training set that will be used to estimate the intercept and slope parameters and a test set that will be used for making predictions. I can then compare the test set predicted values to observed values to estimate predictive ability

### Training and test sets

For each of the 84 datasets I used 75% of sites as training data and the remaining 25% set aside for testing. I used the training set to estimate parameters (intercept and slope for the species-area model and the mean/median value for the null model) and the test set to estimate the predictive performance of the training model. Separate models were built for each dataset. The 75 to 25 split was chosen as this ratio is standard practice in machine learning (Anifowose et al., 2017). I also extracted the statistical significance of the slope estimates and the overall model fit (i.e., adjusted  $R^2$ ) for the species-area models.



### *Spatial extrapolation*

Where geographic coordinates for each patch were provided, I split the training and testing data based on spatial coordinates (see Fig. 1). Splitting the dataset into testing and training dataset using spatial coordinates allows for extrapolation in space. This allowed me to assess how well the information contained in species-area models transferred beyond the spatial boundaries where the model was created. For example, the dataset in Fig. 1 contained 15 sites so I used 11 sites as training data and 4 sites as test data. The training set included sites 1-11 while the test set included sites 12-15. I chose sites 1-11 for training because they were the 11 most southerly sites and sites 12-15 for testing because they were the most northerly. Thus, none of the test sites fell within the spatial boundaries of the sites used to build the species-area models. When spatial coordinates were not provided in the paper, I separated the dataset into testing and training sets at random in a 75-25 split. In those cases, it is likely that test sites fell within the spatial boundaries of the training set, and this implies that I was not testing the spatial transferability of the model beyond the area where the model was built. Forty-eight of the datasets had spatial coordinates (i.e., spatial) and 36 did not (i.e., non-spatial).

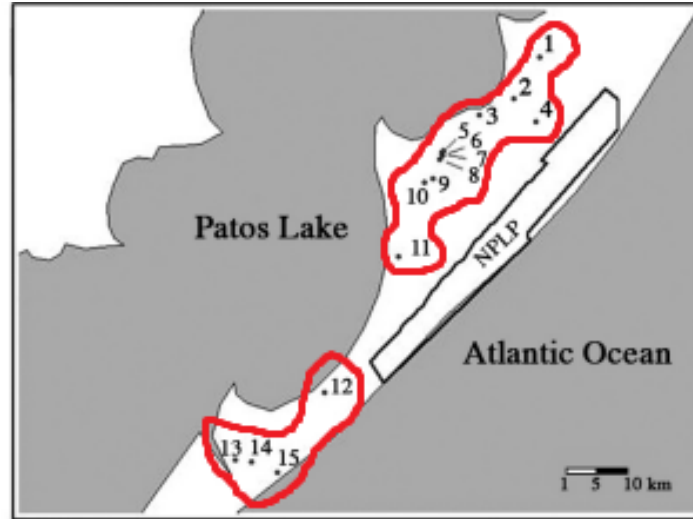


Fig.1. Study Area in southern Brazil conducted by Rolon et al. (2008), comprising a total of 15 sampled wetlands for macrophyte richness. The red lines indicate the separation of the training (1-11) and testing (12-15) datasets, with sites 1-11 predicting to sites 12-15.

#### *Fit to training data and predictive ability*

Model fit to training data might reasonably be expected to be related to the predictive ability of a model. The coefficient of determination ( $R^2$ ) is one commonly used index of model fit. In addition, more data is expected to lead to better parameter estimates and therefore better models. To test these hypotheses, I examined the relationship between model fit to the data ( $R^2$ ) and predictive ability and sample size of dataset and predictive ability.

#### **Prediction Error and Understanding**

The prediction error for species-area and null models is the absolute value of the difference between the observed and predicted error values:

$$|\text{Observed error} - \text{Predicted error}|$$

Predictions with the null model represent predictions with ‘zero knowledge’ while perfect knowledge implies zero prediction error. The total knowledge in each species-area model has been quantified across datasets as

$$\text{Model Knowledge} = \frac{(\text{NullError} - \text{SAError})}{\text{NullError}}$$

Where: NullError = the null prediction error and SAError = ‘best’ species area model prediction error.

This allowed for quantification of total knowledge/understanding in each of the 84 species-area models.

The total understanding contained in SA models were assessed based on predictive ability relative to the null model (i.e., predictive improvement). In the hypothetical scenario presented in Fig. 2. (below), the null model has a prediction error of 8 and the training model has a prediction error of 5. The right-hand side of Fig. 6 (prediction error of 0) is ‘complete understanding’ of SAR models, as a model free of error will make perfect predictions. In this case, the training model made better predictions than the null model as a prediction error of 5 is closer to complete knowledge (i.e., prediction error of 0) than 8. Thus, the SA model is 3/8th of the way closer to complete knowledge than the null model. Note that predictive improvement can be negative if the null makes better predictions than the best SA-model.

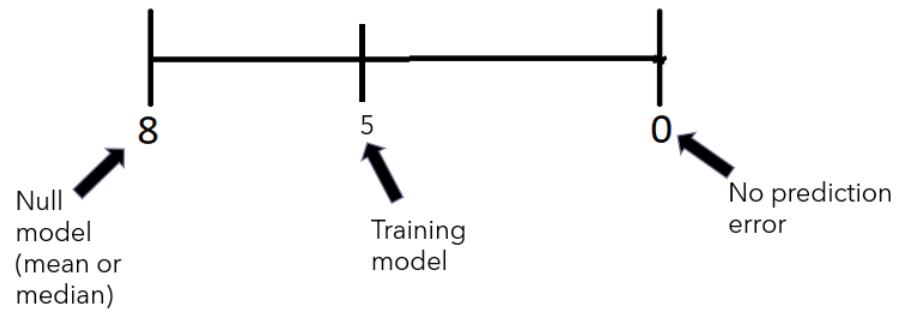


Fig. 2. Visualization of total knowledge in SA models; left hand side is the null model (mean or median) with a prediction error of 8 (representing 'zero knowledge'); middle represents the training model with a prediction error of 5; right hand side with prediction error of 0.

## Results

### ALL DATASETS

#### Data Description

There were 84 species-area datasets including plants, birds, mammals, amphibians, and invertebrates from 27 different countries and 3 different continents. The 84 datasets were from a variety of habitat including lake islands, forest patches, wetlands, forests, parks, ponds, etc. The mean number of sites was 42.84 and ranged from 10 and 242 (Appendix 1). The mean species richness was 22.82 and ranged from 1.63 to 161.26 (Appendix 1). The mean area was 291.61 ha and ranged from 0.04 ha to 17844 ha (Appendix 1).

#### Prediction Error: SAR versus Null models

Across the 84 datasets, 68 of the 84 models were found to better predict species-richness than the mean (Fig. 3). The mean prediction errors using the best SAR model was 5.67643 and ranged between 0.09947 and 47.7399 (Appendix 1). The mean prediction error using the best null model was 7.64153 and ranged from 0.30151 to 45.75 (Appendix 1). For the best SAR models, 37 were model 4 (log SR-area), 19 were model 2 (log SR-log area), 14 were model 1 (SR-area) and 14 were model 3 (SR-log area) (Appendix 1). The mean predictive improvement was 0.27091 and ranged from -1.3721 to 0.9215 (Fig 4.).

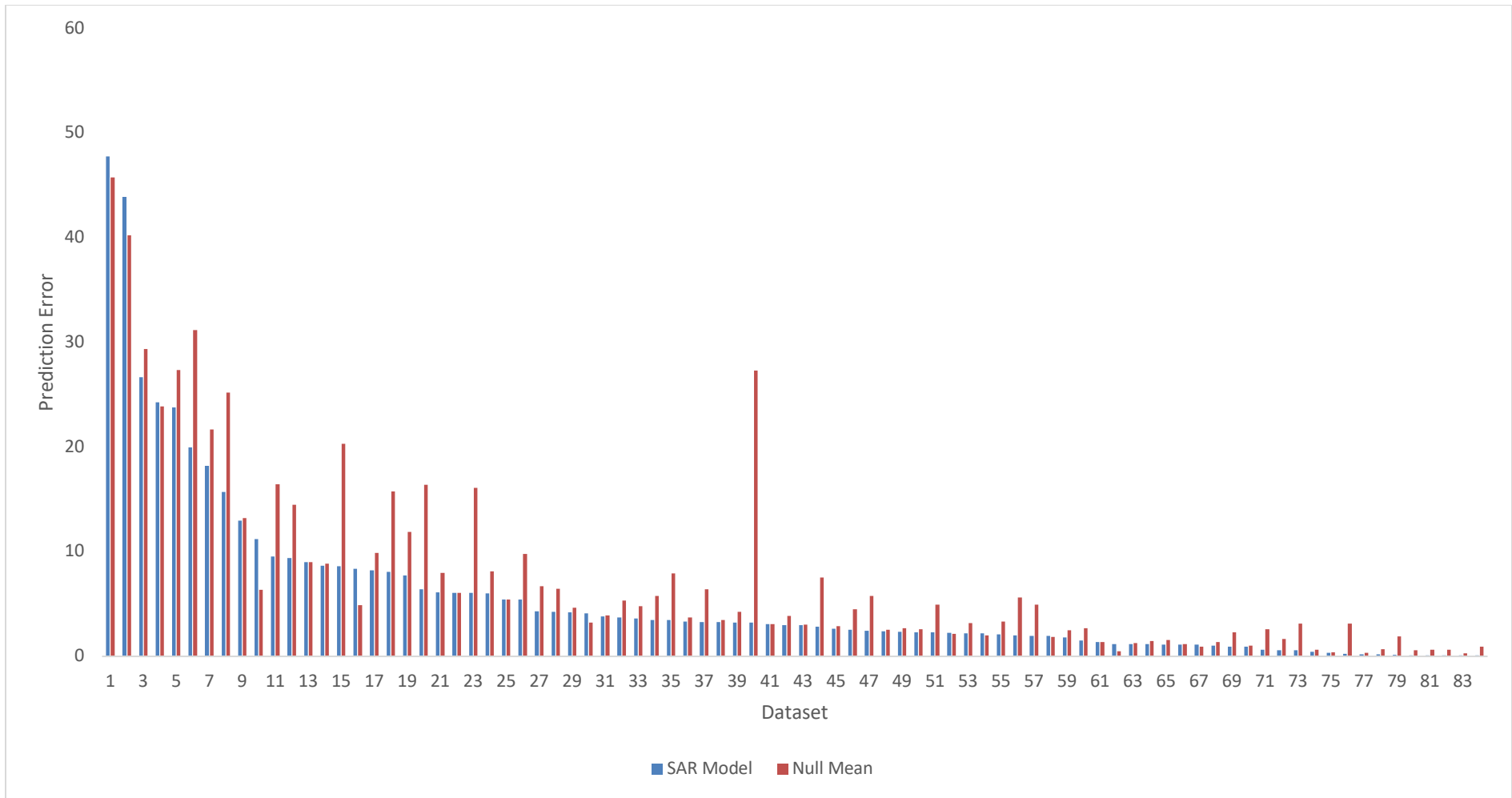


Fig 3: Prediction error for the most accurately predicting SAR model across 84 datasets.

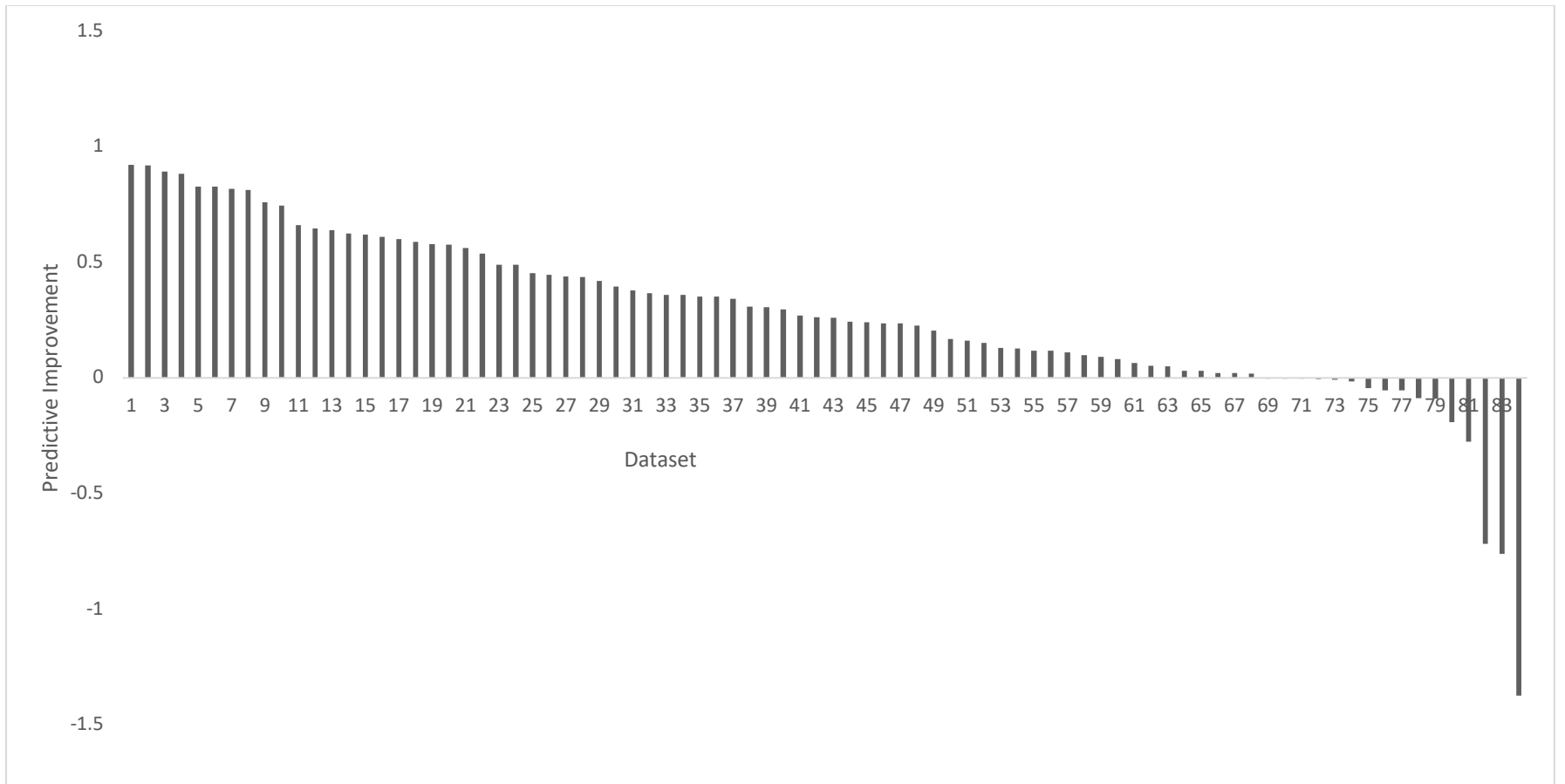


Fig 4: The overall predictive improvement across all 84 datasets (spatial (1-48) + non-spatial (49-84) combined)

### Best Model Fit to Training Set

Model 4 (SR/ Log-Area) had the best fit to the training set (i.e., best estimated the parameters (intercept and slope) for species-area models). Model 1 had the worst fit to the training set.

### R<sup>2</sup> versus Predictive Improvement (84 datasets)

I found a statistically significant but weak positive relationship between R<sup>2</sup> and predictive improvement (Fig. 5). The mean R<sup>2</sup> was 0.4879 and ranged from -0.1017 and 0.9954 (Fig 5.).

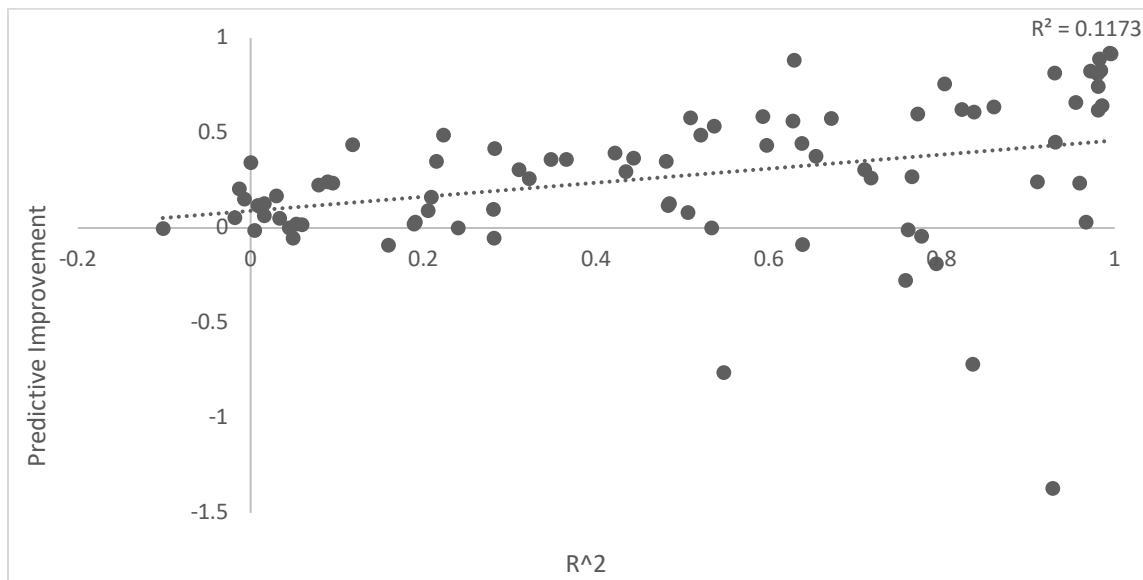


Fig 5: Relationship between R<sup>2</sup> on the training dataset and predictive improvement across all 84 datasets.

### Number of Sites versus Predictive Improvement (84 datasets)

The relationship between the total number of sites in a dataset and predictive improvement of the species-area models (Fig. 6) was not statistically significant. The observed relationship was very weak and negative (Fig. 6).



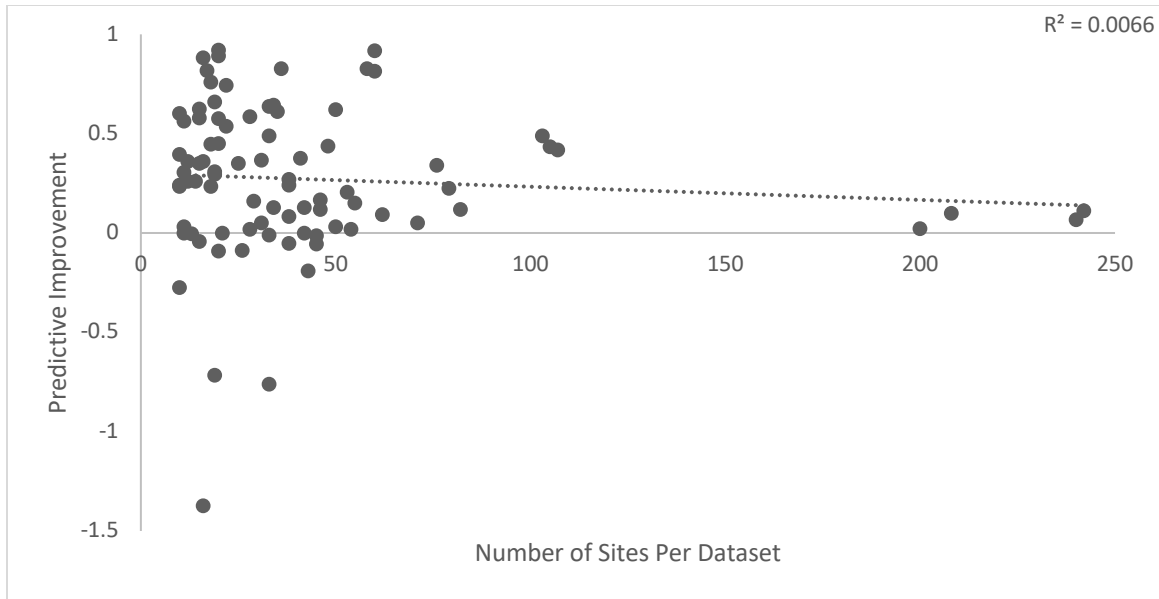


Fig 6. Relationship between total number of sites (per dataset) and predictive improvement for ‘best’ species-area models across 84 datasets.

## SPATIAL DATASETS

### Data Description

Forty-eight out of the 84 datasets were spatially extrapolated (Fig. 7) and these datasets included plants, birds, mammals, amphibians, and invertebrates from 18 different countries and 1 continent. The 48 datasets were from a variety of habitat including lake islands, forest patches, wetlands, forests, parks, ponds, etc. The mean number of sites was 46.7 ranged from 10 and 242 (Appendix 1). The mean species richness was 21.43 and ranged from 2.81 and 161.26 (Appendix 1). The mean area was 74.91 ha ranged from 0.0004 ha and 76.43 ha (Appendix 1).

### Model vs Mean- Spatial Extrapolation

The mean prediction error for SA model was 4.1859 and ranged from 0.09946 and 26.6901. The best SAR model was the log species richness – area model for 24 datasets, log species richness – log area model for 11 datasets, species richness – log area for seven datasets and species richness – area for 6 datasets (Fig. 7). The mean predictive improvement was 0.3457 and ranged from -0.7162 and 0.9215.

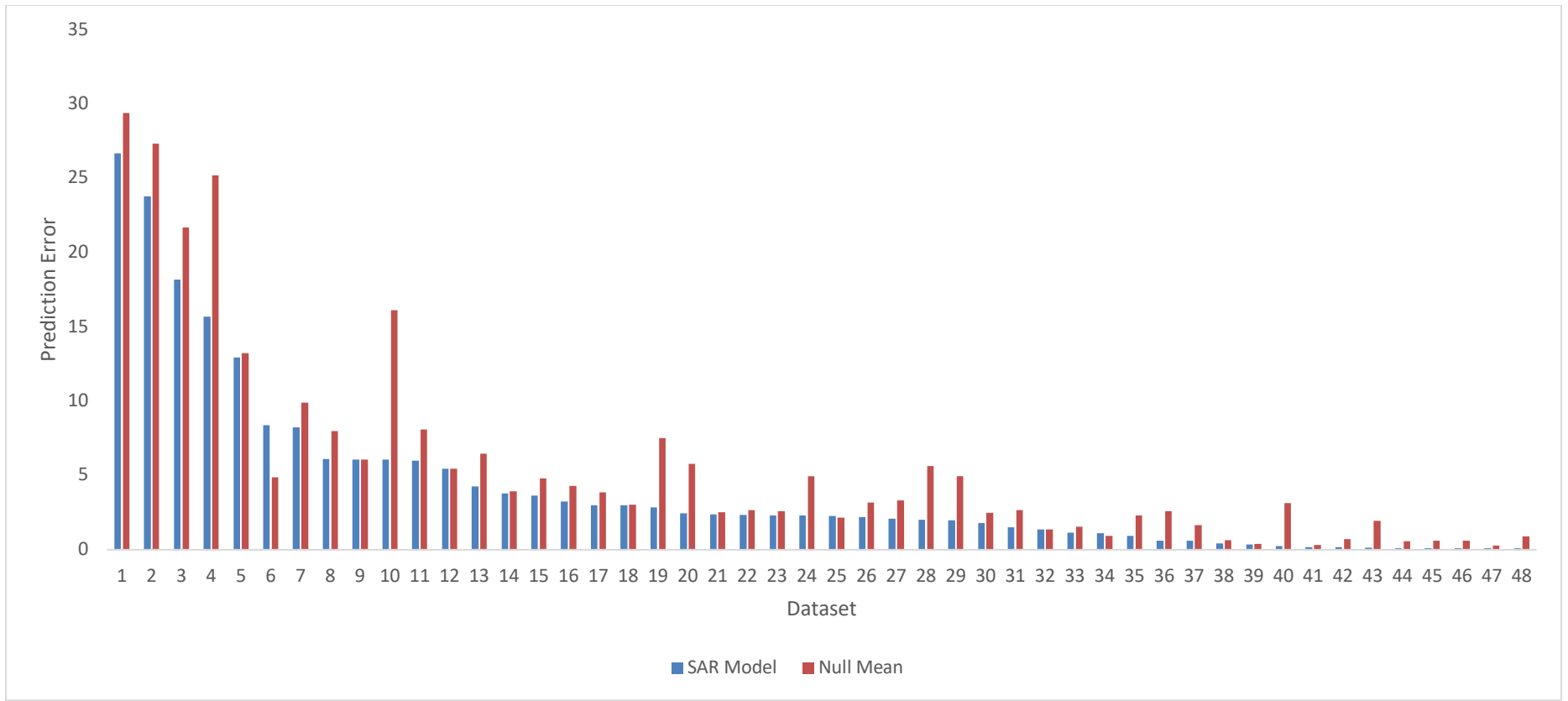


Fig. 7. Prediction errors across 48 datasets (spatially extrapolated) from the four species-area richness models which were selected from the most accurately predicting model per dataset.

## **R<sup>2</sup> versus Prediction Error- Spatial Extrapolation**

The model fit to the training data (i.e., residual R<sup>2</sup>) in the spatial datasets provided moderate evidence of its predictive ability to the test set (Fig. 8). The mean R<sup>2</sup> was 0.5628 and ranged from -0.01844 and 0.9954. When plotted, the relationship between R<sup>2</sup> and predictive improvement was found to be a moderate positive trend (R<sup>2</sup> = 0.2812; Fig. 8).

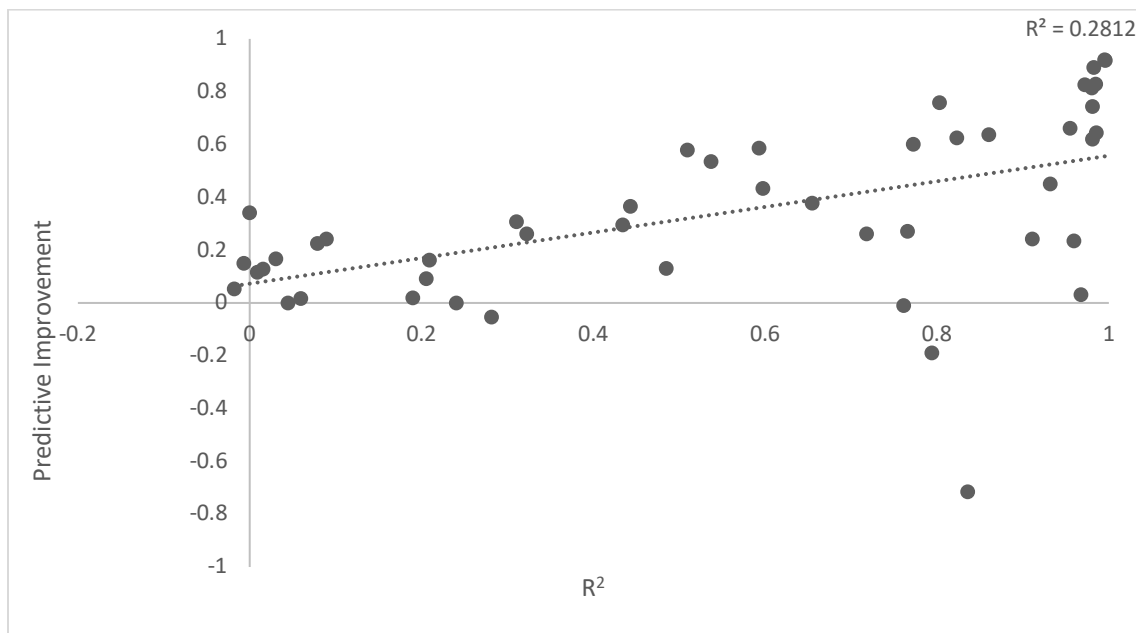


Fig 8. Relationship between R<sup>2</sup> on the training dataset (spatial) and predictive improvement

## **Number of Sites versus Predictive Improvement – Spatial Extrapolation**

There was no statistically significant relationship between the total number of sites in a dataset and predictive improvement of the species-area models for spatial data (Fig. 9).

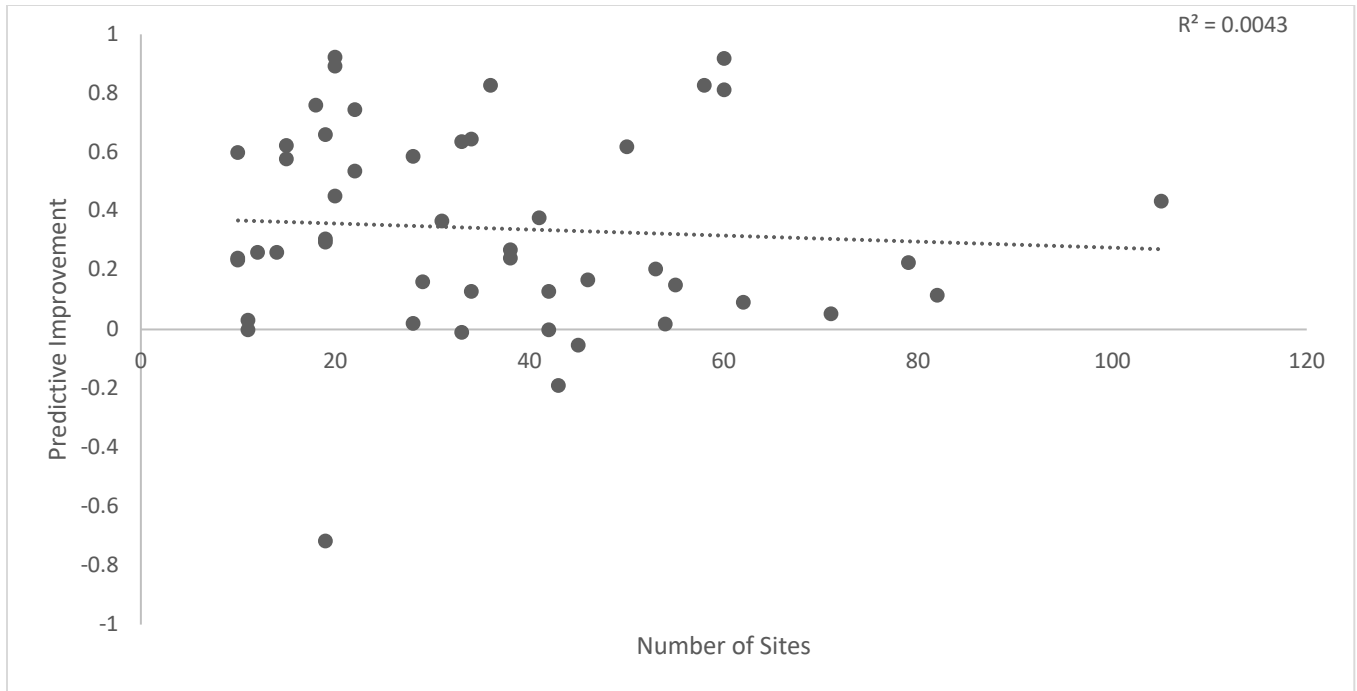


Fig 9. Relationship between total number of sites (per dataset) and predictive improvement for ‘best’ species-area models for spatial data.

## NON-SPATIAL DATASETS

### Data Description

36 out of the 84 datasets were non-spatially extrapolated (Fig. 10) and these datasets included plants, birds, mammals, amphibians, and invertebrates from 9 different countries and 2 continents. The 36 datasets were from a variety of habitat including lake islands, forest patches, wetlands, forests, parks, ponds, etc. The mean number of sites was 43.5 ranged from 10 and 208 (Appendix 1). The mean species richness was 45.06 and ranged from 1.63 and 115.06 (Appendix 1). The mean area was 50.60 ranged from 0.035 ha and 17844 ha (Appendix 1).

## Model vs Mean- Non-Spatial Extrapolation

The mean prediction error for SA models was 7.6637 ranged from 0.57 to 47.73. Model 4 predicting most accurately in 13 datasets, followed by models 1 and 2, each predicting most accurately in 8 datasets, and lastly, model 3 predicting accurately in 7 datasets (Fig. 10). The mean predictive improvement was 0.1711 and ranged from -1.3721 and 0.8814 (Fig. 10).

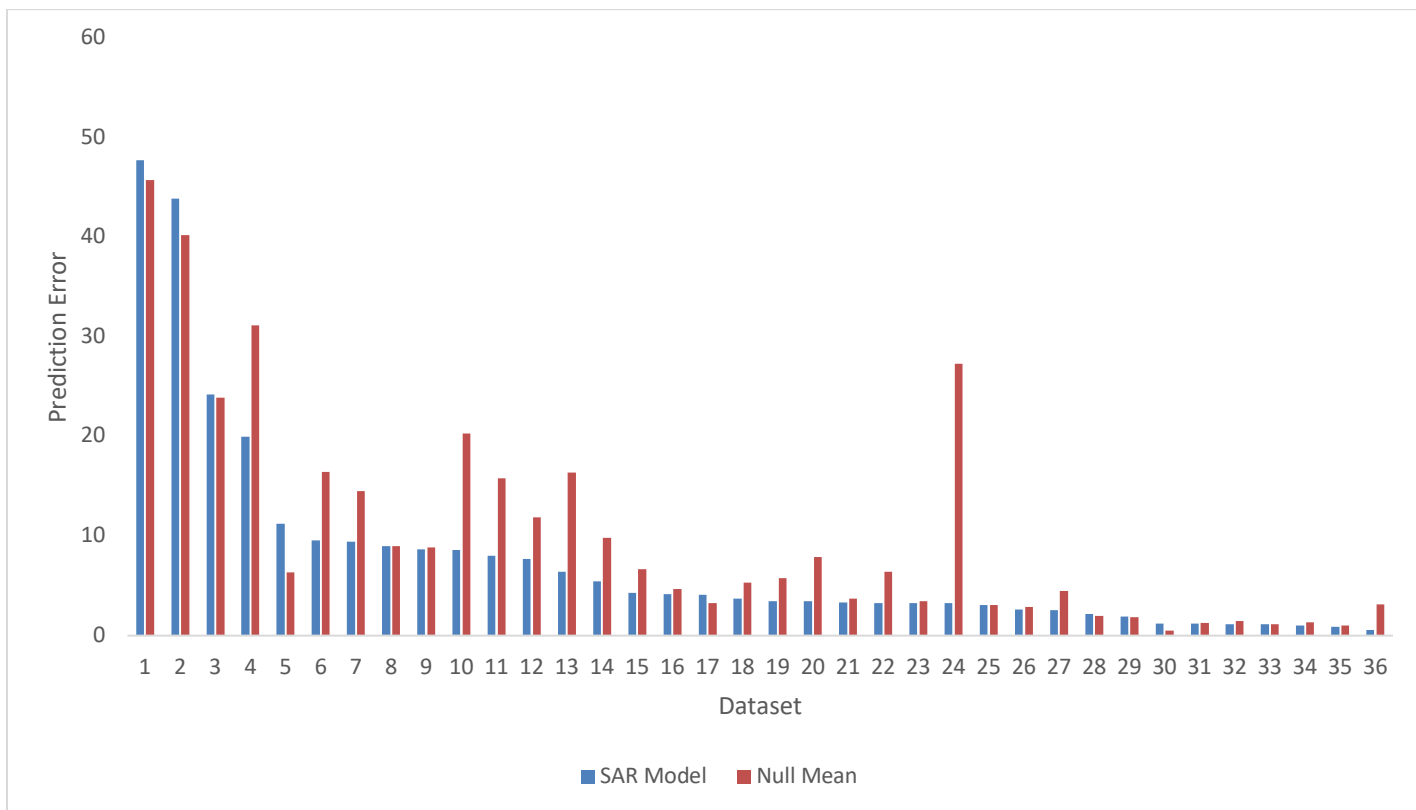


Fig 10. Prediction errors across 36 datasets (non- spatially extrapolated) from the four species-area richness models which were selected from the most accurately predicting model per dataset.

## R<sup>2</sup> versus Prediction Error- Non-Spatial Extrapolation

The model fit to the training data (i.e.,  $R^2$ ) in the non-spatial datasets provided no evidence of its predictive ability to the test set (Fig. 11).

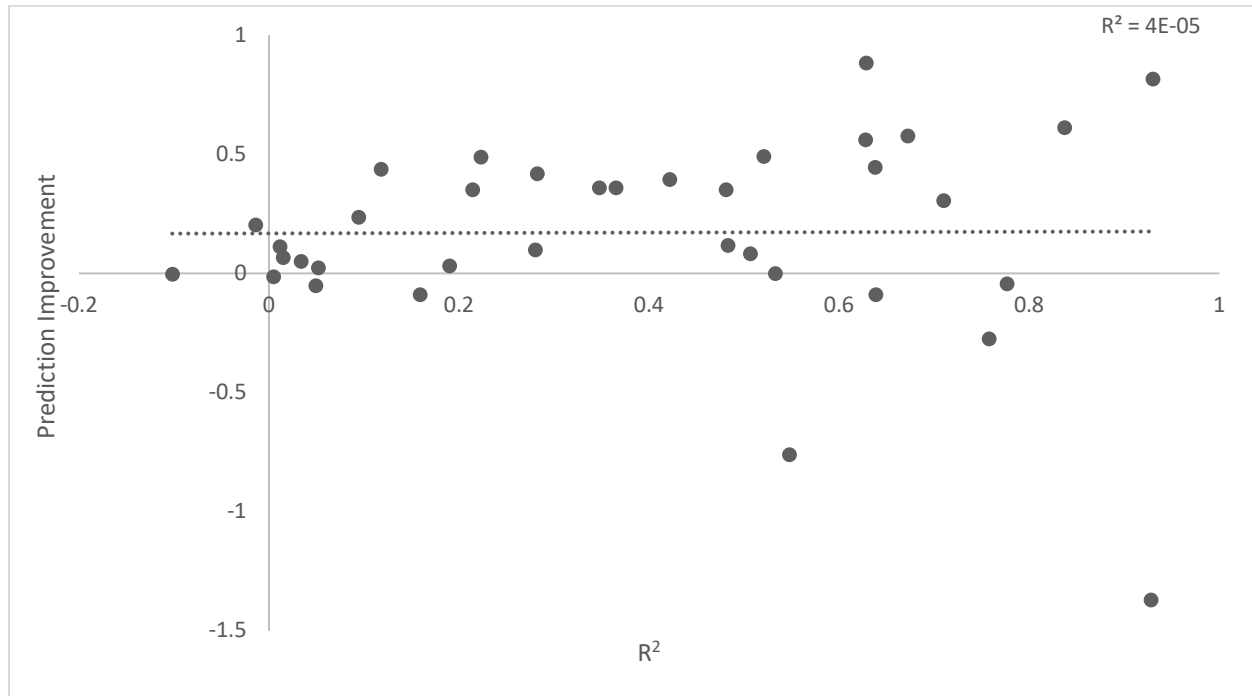


Fig 11. Relationship between  $R^2$  on the training dataset (non-spatial) and predictive improvement.

### Number of Sites and Predictive Improvement

There was no relationship between the total number of sites in a dataset and predictive improvement of the species-area models for non-spatial data (Fig. 12).

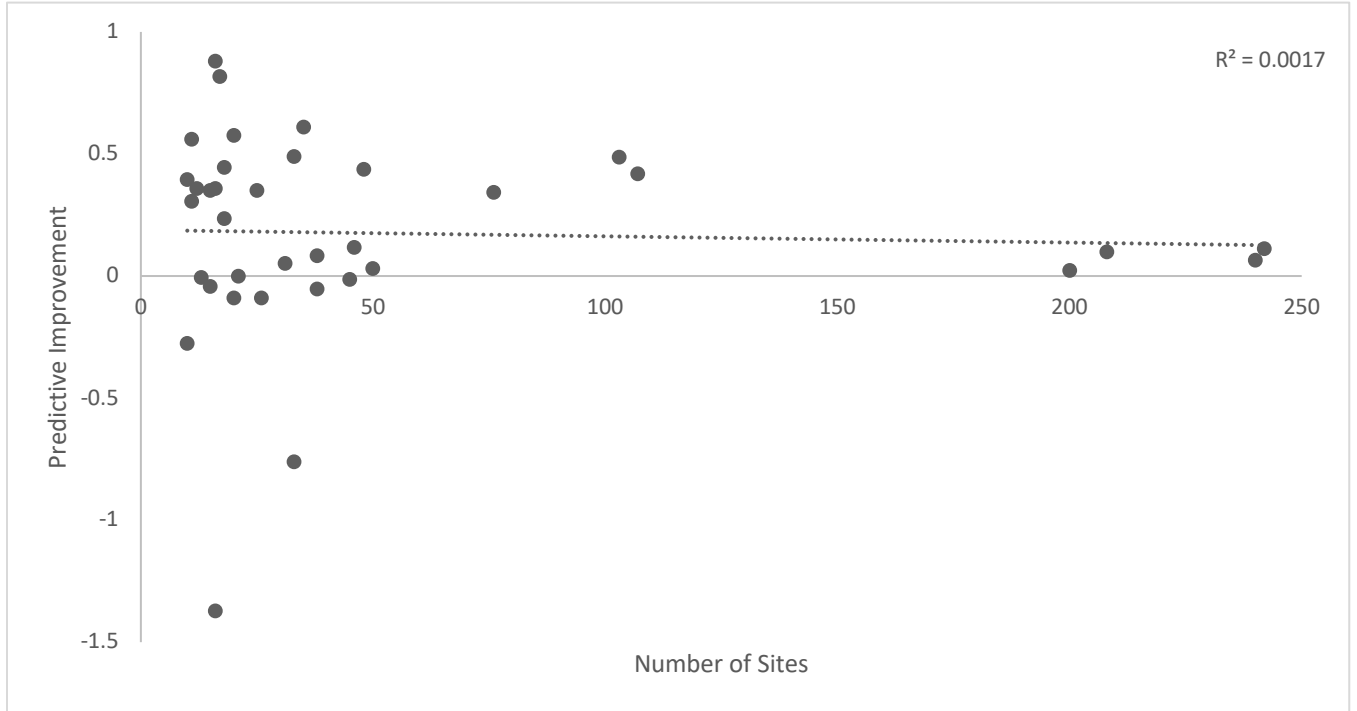


Fig 12. Relationship between total number of sites (per dataset) and predictive improvement for most accurately predicting species-area models for non-spatial data.



## Discussion

I found that (1) species-area models tend to outperform null models for the entire dataset and for both spatial and non-spatial datasets, (2) model 4 (SR/Log AREA) fit to new data was most consistent across datasets, (3) there is weak to moderate evidence for the model fit to the training data (i.e., residual  $R^2$ ) across datasets.

### Predictive Ability of SAR Models

Studies have assessed the predictive ability of SAR models for a variety of purposes. For instance, Grelle et al. (2005) accurately predicted number of species that are likely to go extinct given the loss of specific habitats using SAR models for mammals, birds, reptiles, and amphibians. They found that the species-area model accurately estimated the number of endemic mammals and birds listed as endangered but overestimated for amphibians and reptiles and concluded that birds and mammals endemic to the Atlantic Forest seemed to be more negatively affected by deforestation than amphibians and reptiles.

Ferreira & Pellegrini (2019) also investigated the predictive ability of SAR models by assessing the impact of reduced cave area on species-richness over-time. These researchers estimated the effects of ground area lost in a limestone cave on species-richness of invertebrate fauna in Southeastern Brazil over the course of 2 years and found that species-area models accurately predicted species-richness in the cave community following reduction of area.

### The Four SAR Models

Connor & McCoy (1979) fit each of the four SAR models (i.e., untransformed model, log/log model, species/log-area model, and log-species/area model), to 100 species-area datasets.

In the instance where two or more models fit equally well, there would be more than one ‘best model’ (i.e., strongest fit to data), therefore, the total sum of best models exceeded 100. Of these, 35 were best fit by the untransformed model, 43 were best fit by the log/log model, 14 were best fit by the log-species/area model, and 27 were best fit by the species/log-area model (Connor & McCoy, 1979). Thus, it was found that the log/log models fit most frequently to the datasets as this model turns any monotonic function into a straight line (Preston, 1962). My results are not in agreement with Connor & McCoy’s (1979) as I showed that the species/log-area model fit more frequently to new data than the log/log model and the untransformed and the species/log-area models combined were the best for only 1/3 of datasets.

Other ecological models have been tested in the past to predict to new data. A study by Wenger and Olden (2012) assessed the transferability of distribution models for brook trout (*Salvelinus fontinalis* Mitchell) and brown trout (*Salmo trutta* Linnaeus) in western United States. The cross-validation technique was applied in this study which is when data is split based on non-random factors which in this case was split based on geographically distinct subsets (Wenger & Olden, 2012). They found that traditional linear models had greater transferability than other machine-learning techniques (e.g., random forests and artificial neural networks). Another study by Bahn & McGill (2012) tested predictive ability of distribution models for abundance of bird species by utilizing a multitude of testing measures to address different levels of information such as regression  $R^2$  for absolute abundance of birds, squared correlation coefficient  $r^2$  for relative abundance and AUC/Somer’s D for presence/absence. They found that various factors lead to lower predictive performance such as higher levels of independence between testing and training sets and increasing information content (e.g., identifying presence/absence, relative abundance, and absolute abundance) (Bahn & McGill, 2013). Despite

the fact that transferability has been assessed for various ecological models in the past, it has yet to be tested for SAR models.

### **Predictive Ability and Complete Understanding**

As mentioned earlier, ecological models contain all of our knowledge of ecological systems, and the only way to assess the total amount of knowledge/understanding in a model is by testing its predictive ability (Houlahan et al., 2017). Despite this, the predictive ability of ecological models is rarely assessed (Consonni et al., 2010; Ellis et al., 2012; Randin et al., 2006; Townsend et al., 2007). This is the case with SAR models as this relationship has been referred to as “one of community ecology’s few genuine law” (Schoener 1976, p. 629), yet, its transferability has not been tested, which implies we do not know much how information is contained in these models. Based on the results of my study however, the SAR models had better transferability than the null models in most cases, which tells us that species-area models are capturing something true about how the world works. Therefore, it would be useful to assess the transferability of SAR models with a much larger dataset to assess complete knowledge.

### **Study Limitations and Future Studies**

At the beginning of my study, the main goal was to use SAR models to extrapolate in space. During my data collection, I was able to extract 84 SAR datasets, but I was only able to collect spatial coordinates (longitude + latitude, UTM, etc.) for 48 of the 84 datasets. Therefore, I was only able to extrapolate in space for 48 of the datasets and non-spatially for 36 datasets. This was a limiting factor in my study as there is potential for more variability when splitting data into 75% training and 25% testing for non-spatial data than for spatial data. Due to this variability, it

would have been ideal to split the training set for non-spatial data multiple times per dataset and then take the average prediction error values, however, I was not able to complete this due to limiting time.

Although there is no baseline for what sufficient data is, a larger dataset will allow detection of more complex interactions and relationships than a smaller dataset (Kuhn & Johnson, 2013). There are many species-area data dispersed throughout literature and in data archives; the conclusions would be stronger if we collected a larger database of species-area relationships. I found little to no relationship between number of sites and predictive improvement, but a larger database of species-area relationships may allow us to detect key covariates of predictive performance. Therefore, it would be important to keep in mind next time to not only collect a larger quantity of datasets, but also increase the threshold of sites per dataset to allow for larger training sets to improve parameter estimates and also, larger test sets to improve the estimates of predictive ability (Dobbin & Simon, 2011).

### **Conclusion**

I conclude that (1) in the majority of the cases, species-area models outperform null models for the entire dataset and for both spatial and non-spatial datasets, (2) compared to the mean, species-area relationship model 4 (i.e., species/log-area model) was most accurate at spatially transferring to new data than other SA models (3) there is weak to moderate evidence that the model fit to the training data (i.e., residual  $R^2$ ) tells us something about how well the model will predict to new data.

## References

- Abarca-Arenas, L. G., Franco-López, J., Valero-Pacheco, E., & González-Gándara, C. Species-Area and Trophic Level of the Fish Community of the Veracruz Coast, Mexico. *Estudios Científicos en el Corredor Arrecifal del Suroeste del Golfo de México*, 247.
- Andrade, R. D., & Marini, M. Â. (2002). Bird species richness in natural forest patches in southeast Brazil. *Lundiana: International Journal of Biodiversity*, 3(2), 141-149.
- Angermeier, P. L., & Schlosser, I. J. (1989). Species-area relationship for stream fishes. *Ecology*, 70(5), 1450-1462.
- Anifowose, F., Khoukhi, A., & Abdulraheem, A. (2017). Investigating the effect of training–testing data stratification on the performance of soft computing techniques: an experimental study. *Journal of Experimental & Theoretical Artificial Intelligence*, 29(3), 517-535.
- Ault, T. R., & Johnson, C. R. (1998). Spatial Variation in Fish Species Richness on Coral Reefs: Habitat Fragmentation and Stochastic Structuring Processes. *Oikos*, 82(2), 354.  
doi:10.2307/3546976
- Bahn, V., & McGill, B. J. (2013). Testing the predictive performance of distribution models. *Oikos*, 122(3), 321-331. <https://doi.org/10.1111/j.1600-0706.2012.00299.x>
- Bellamy, P. E., Hinsley, S. A., & Newton, I. (1996). Factors Influencing Bird Species Numbers in Small Woods in South-East England. *The Journal of Applied Ecology*, 33(2), 249.  
doi:10.2307/2404747
- Benassi, G., Battisti, C., & Luiselli, L. (2007). Area effect on bird species richness of an archipelago of wetland fragments in Central Italy. *Community Ecology*, 8(2), 229–237.

- Benedick, S., Hill, J. K., Mustaffa, N., Chey, V. K., Maryati, M., Searle, J. B., ... & Hamer, K. C. (2006). Impacts of rain forest fragmentation on butterflies in northern Borneo: species richness, turnover and the value of small fragments. *Journal of Applied Ecology*, *43*(5), 967-977.
- Benitez-Malvido, J., & Martinez-Ramos, M. (2003). Impact of Forest Fragmentation on Understory Plant Species Richness in Amazonia. *Conservation Biology*, *17*(2), 389–400. doi:10.1046/j.1523-1739.2003.01120.x
- Bila, K., Kuras, T., Sipos, J., & Kindlmann, P. (2013). Lepidopteran species richness of alpine sites in the High Sudetes Mts.: effect of area and isolation. *Journal of insect conservation*, *17*(2), 257-267.
- Boecklen, W. J. (1986). Effects of Habitat Heterogeneity on the Species-Area Relationships of Forest Birds. *Journal of Biogeography*, *13*(1), 59. doi:10.2307/2844849
- Brooks, T. M., Pimm, S. L., & Collar, N. J. (1997). Deforestation predicts the number of threatened birds in insular Southeast Asia: La extensión de las deforestaciones predice el número de aves amenazadas de extinción en las islas del sureste de Asia. *Conservation Biology*, *11*(2), 382-394.
- Brose, U. (2008). Relative importance of isolation, area and habitat heterogeneity for vascular plant species richness of temporary wetlands in east-German farmland. *Ecography*, *24*(6), 722–730. doi:10.1111/j.1600-0587.2001.tb00533.x
- Brunet, A. K., & Medellín, R. A. (2001). The species–area relationship in bat assemblages of tropical caves. *Journal of Mammalogy*, *82*(4), 1114-1122.
- Cardinale, B. J. et al. (2006). Effects of biodiversity on the functioning of trophic groups and ecosystems. *Nature* 443: 989992.

- Chamberlain, D. E., Gough, S., Vaughan, H., Vickery, J. A., & Appleton, G. F. (2007). Determinants of bird species richness in public green spaces. *Bird Study*, *54*(1), 87–97. doi:10.1080/00063650709461460
- Diver, K. (2010). Beyond Area and Isolation: Forest Dynamics and Change in Plant Species Richness Patterns on Islands. *Annals of the Association of American Geographers*, *100*(5), 1110–1127. doi:10.1080/00045608.2010.517738
- Connor, E. F., & McCoy, E. D. (1979). The statistics and biology of the species-area relationship. *The American Naturalist*, *113*(6), 791–833.
- Consonni, V., Ballabio, D., & Todeschini, R. (2010). Evaluation of model predictive ability by external validation techniques. *Journal of Chemometrics*, *24*(3–4), 194–201. <https://doi.org/10.1002/cem.1290>
- De Camargo, R., X., & Currie, D. J. (2015). An empirical investigation of why species-area relationships overestimate species losses. *Ecology*, *96*(5), 1253.
- De Lacy, P., & Shackleton, C. M. (2017). Woody plant species richness, composition and structure in urban sacred sites, Grahamstown, South Africa. *Urban Ecosystems*, *20*(5), 1169–1179. doi:10.1007/s11252-017-0669-y
- Dobbin, K. K., & Simon, R. M. (2011). Optimally splitting cases for training and testing high dimensional classifiers. *BMC medical genomics*, *4*(1), 1–8. <https://doi.org/10.1186/1755-8794-4-31>
- Dodson, S. I., & Silva-Briano, D. (1996). Crustacean zooplankton species richness and associations in reservoirs and ponds of Aguascalientes State, Mexico. *Hydrobiologia*, *325*(2), 163–172. doi:10.1007/bf00028277

Dos Anjos, L., & Bocon, R. (1999). Bird communities in natural forest patches in southern Brazil. *The Wilson Bulletin*, 397-414.

Eckmann, R. (1995). Fish species richness in lakes of the northeastern lowlands in Germany. *Ecology of Freshwater Fish*, 4(2), 62–69. doi:10.1111/j.1600-0633.1995.tb00118.x

Elliott, L. H., Igl, L. D., & Johnson, D. H. (2019). The relative importance of wetland area versus habitat heterogeneity for promoting species richness and abundance of wetland birds in the Prairie Pothole Region, USA. *The Condor*, 122(1). doi:10.1093/condor/duz060

Ellis, N., Smith, S. J., & Pitcher, C. R. (2012). Gradient forests: calculating importance gradients on physical predictors. *Ecology*, 93(1), 156–168. <https://doi.org/10.1890/11-0252.1>

Ferenc, M., Sedláček, O., Fuchs, R., Dinetti, M., Fraissinet, M., & Storch, D. (2013). Are cities different? Patterns of species richness and beta diversity of urban bird communities and regional species assemblages in Europe. *Global Ecology and Biogeography*, 23(4), 479–489. doi:10.1111/geb.12130

Ferreira, R. L., & Pellegrini, T. G. (2019). Species-area model predicting diversity loss in an artificially flooded cave in Brazil. *International Journal of Speleology*, 48(2), 4.

Figueroa, J. A., Castro, S. A., Reyes, M., & Teillier, S. (2018). Urban park area and age determine the richness of native and exotic plants in parks of a Latin American city: Santiago as a case study. *Urban Ecosystems*, 21(4), 645–655. doi:10.1007/s11252-018-0743-0

Flaspohler, D. J., Giardina, C. P., Asner, G. P., Hart, P., Price, J., Lyons, C. K., & Castaneda, X. (2010). Long-term effects of fragmentation and fragment properties on bird species richness in Hawaiian forests. *Biological Conservation*, 143(2), 280–288. doi:10.1016/j.biocon.2009.10.009

Gotelli, N. J., & McGill, B. J. (2006). Null versus neutral models: what's the difference?. *Ecography*, 29(5), 793-800.



Grelle, C. E. V., Alves, M. A. S., Bergallo, H. G., Geise, L., Rocha, C. F. D., Van Sluys, M., & Caramaschi, U. (2005). Prediction of threatened tetrapods based on the species–area relationship in Atlantic Forest, Brazil. *Journal of Zoology*, 265(4), 359-364.

Guadagnin, D. L., Maltchik, L., & Fonseca, C. R. (2009). Species–area relationship of Neotropical waterbird assemblages in remnant wetlands: looking at the mechanisms. *Diversity and Distributions*, 15(2), 319-327.

Haig, A. R., Matthes, U., & Larson, D. W. (2000). Effects of natural habitat fragmentation on the species richness, diversity, and composition of cliff vegetation. *Canadian journal of botany*, 78(6), 786-797.

Hattori, A., & Shibuno, T. (2009). The effect of patch reef size on fish species richness in a shallow coral reef shore zone where territorial herbivores are abundant. *Ecological Research*, 25(2), 457–468. doi:10.1007/s11284-009-0675-5

Hasegawa, K., Mori, T., & Yamazaki, C. (2016). Density-dependent effects of non-native brown trout *Salmo trutta* on the species-area relationship in stream fish assemblages. *Journal of Fish Biology*, 90(1), 370–383. doi:10.1111/jfb.13185

He, F., & Legendre, P. (2002). Species diversity patterns derived from species–area models. *Ecology*, 83(5), 1185-1198.

Helden, A. J., & Leather, S. R. (2004). Biodiversity on urban roundabouts—Hemiptera, management and the species–area relationship. *Basic and applied Ecology*, 5(4), 367-377.

Helm, A., Hanski, I., & Partel, M. (2005). Slow response of plant species richness to habitat loss and fragmentation. *Ecology Letters*, 0(0), 051109031307003. doi:10.1111/j.1461-0248.2005.00841.x

Hobbs, E. R. (1988). Species richness of urban forest patches and implications for urban landscape diversity. *Landscape Ecology*, 1(3), 141–152. doi:10.1007/bf00162740

Holyoak, G. A., Turk, S. M., & Cameron, R. A. D. (2005). Land molluscs on the Isles of Scilly: Species richness and composition related to area. *Journal of Conchology*, 38(5), 549-560.

Honnay, O., Endels, P., Vereecken, H., & Hermy, M. (1999). The role of patch area and habitat diversity in explaining native plant species richness in disturbed suburban forest patches in northern Belgium. Biodiversity Research. *Diversity Distributions*, 5(4), 129–141. doi:10.1046/j.1472-4642.1999.00047.x

Houlahan, J. E., Keddy, P. A., Makkay, K., & Findlay, C. S. (2006). The effects of adjacent land use on wetland species richness and community composition. *Wetlands*, 26(1), 79–96. doi:10.1672/0277-5212(2006)26[79:teoalu]2.0.co;2

Houlahan, J. E., McKinney, S. T., Anderson, T. M., & McGill, B. J. (2017). The priority of prediction in ecological understanding. *Oikos*, 126(1), 1-7.

Huang, Y., Zhao, Y., Li, S., & von Gadow, K. (2015). The Effects of habitat area, vegetation structure and insect richness on breeding bird populations in Beijing urban parks. *Urban Forestry & Urban Greening*, 14(4), 1027–1039. doi:10.1016/j.ufug.2015.09.010

Jackson, J. (2005). Is there a relationship between herbaceous species richness and buffel grass (*Cenchrus ciliaris*)?. *Austral Ecology*, 30(5), 505-517.

Khaine, I., Woo, S. Y., Kang, H., Kwak, M., Je, S. M., You, H., ... Kim, J. (2017). Species Diversity, Stand Structure, and Species Distribution across a Precipitation Gradient in Tropical Forests in Myanmar. *Forests*, 8(8), 282. doi:10.3390/f8080282

Kimmerer, R. W., & Driscoll, M. J. (2000). Bryophyte species richness on insular boulder habitats: the effect of area, isolation, and microsite diversity. *Bryologist*, 748-756.

- Kuhn, M., & Johnson, K. (2013). Over-fitting and model tuning. *In Applied predictive modeling* (pp. 61-92). Springer, New York, NY.
- Ladle, R.J. & Whittaker, R.J. (2011) Conservation biogeography. John Wiley & Sons, *Oxford*.
- Lawes, M. J., Lamb, B. C. C., & Boudreau, S. (2005). Area- but no edge-effect on woody seedling abundance and species richness in old Afromontane forest fragments. *Journal of Vegetation Science*, 16(4), 363. doi:10.1658/1100-9233(2005)016[0363:abneow]2.0.co;2
- Lawesson, J. E., de Blust, G., Grashof, C., Firbank, L., Honnay, O., Hermy, M., ... Jensen, L. M. (1998). Species diversity and area-relationships in Danish beech forests. *Forest Ecology and Management*, 106(2-3), 235–245. doi:10.1016/s0378-1127(97)00316-2
- Leather, S. R. (1986). Insect Species Richness of the British Rosaceae: The Importance of Host Range, Plant Architecture, Age of Establishment, Taxonomic Isolation and Species-Area Relationships. *The Journal of Animal Ecology*, 55(3), 841.
- Lee, M.-B., & Carroll, J. P. (2018). Effects of patch size and basal area on avian taxonomic and functional diversity in pine forests: Implication for the influence of habitat quality on the species-area relationship. *Ecology and Evolution*, 8(14), 6909–6920. doi:10.1002/ece3.4208
- Li, D., Monahan, W. B., & Baiser, B. (2018). Species richness and phylogenetic diversity of native and non-native species respond differently to area and environmental factors. *Diversity and Distributions*, 24(6), 853-864.
- Lucey, J. M., Tawatao, N., Senior, M. J. M., Chey, V. K., Benedick, S., Hamer, K. C., ... Hill, J. K. (2014). Tropical forest fragments contribute to species richness in adjacent oil palm plantations. *Biological Conservation*, 169, 268–276. doi:10.1016/j.biocon.2013.11.014

- Liu, Z., Li, X., Yan, Q., & Wu, J. (2007). Species richness and vegetation pattern in interdune lowlands of an active dune field in Inner Mongolia, China. *Biological Conservation*, 140(1-2), 29–39. doi:10.1016/j.biocon.2007.07.030
- Lizee, M.-H., Tatoni, T., & Deschamps-Cottin, M. (2015). Nested patterns in urban butterfly species assemblages: respective roles of plot management, park layout and landscape features. *Urban Ecosystems*, 19(1), 205–224. doi:10.1007/s11252-015-0501-5
- Lomolino, M. V. (2000). Ecology's most general, yet protean pattern: the species-area relationship. *Journal of Biogeography*, 17-26.
- Longato, L. O., Ferreira, I. E. de P., & Perbiche-Neves, G. (2018). Relationship between zooplankton richness and area in Brazilian lakes: comparing natural and artificial lakes and trends. *Acta Limnologica Brasiliensia*, 30(0). doi:10.1590/s2179-975x1518
- Ma, M., Tarmi, S., & Helenius, J. (2002). Revisiting the species–area relationship in a semi-natural habitat: floral richness in agricultural buffer zones in Finland. *Agriculture, ecosystems & environment*, 89(1-2), 137-148.
- MacArthur, R. H., & E. O. Wilson. (1963). An equilibrium theory of insular zoogeography. *Evolution* 17: 373-387. --, & --. 1967. The theory of island biogeography. Princeton, New Jersey, Princeton Univ. Press.
- MacArthur, R. H., and E. O. Wilson. (1967). The theory of island biogeography. *Princeton Univ. Press, Princeton*, N.J. 203 p
- Maltchik, L., Lanés, L. E. K., Stenert, C., & Medeiros, E. S. F. (2010). Species-area relationship and environmental predictors of fish communities in coastal freshwater wetlands of southern Brazil. *Environmental Biology of Fishes*, 88(1), 25–35. doi:10.1007/s10641-010-9614-0

- Matthews, J. W., Tessene, P. A., Wiesbrook, S. M., & Zercher, B. W. (2005). Effect of area and isolation on species richness and indices of Floristic Quality in Illinois, USA wetlands. *Wetlands*, 25(3), 607–615. doi:10.1672/0277-5212(2005)025[0607:eoaaio]2.0.co;2
- Mason, C. F. (2001). Woodland area, species turnover and the conservation of bird assemblages in lowland England. *Biodiversity and Conservation*, 10(4), 495–510.  
doi:10.1023/a:1016606410892
- McNaughton, S. J. (1977). Diversity and stability of ecological communities: a comment on the role of empiricism in ecology. *The American Naturalist*, 111(979), 515-525.
- Natuhara, Y., & Imai, C. (1999). Prediction of species richness of breeding birds by landscape-level factors of urban woods in Osaka Prefecture, Japan. *Biodiversity and Conservation*, 8(2), 239–253. doi:10.1023/a:1008869410668
- Nicolas, D., Lobry, J., Lepage, M., Sautour, B., Le Pape, O., Cabral, H., ... Boët, P. (2010). Fish under influence: A macroecological analysis of relations between fish species richness and environmental gradients among European tidal estuaries. *Estuarine, Coastal and Shelf Science*, 86(1), 137–147. doi:10.1016/j.ecss.2009.11.006
- Oberdorff, T., Guilbert, E., & Lucchetta, J.-C. (1993). Patterns of fish species richness in the Seine River basin, France. *Hydrobiologia*, 259(3), 157–167. doi:10.1007/bf00006595
- Oertli, B., Joye, D. A., Castella, E., Juge, R., Cambin, D., & Lachavanne, J.-B. (2002). Does size matter? The relationship between pond area and biodiversity. *Biological Conservation*, 104(1), 59–70. doi:10.1016/s0006-3207(01)00154-9
- Oliver, A. J., Hong-Wa, C., Devonshire, J., Olea, K. R., Rivas, G. F., & Gahl, M. K. (2011). Avifauna richness enhanced in large, isolated urban parks. *Landscape and Urban Planning*, 102(4), 215–225. doi:10.1016/j.landurbplan.2011.04.007

- Öster, M., Cousins, S. A. O., & Eriksson, O. (2007). Size and heterogeneity rather than landscape context determine plant species richness in semi-natural grasslands. *Journal of Vegetation Science*, *18*(6), 859–868. doi:10.1111/j.1654-1103.2007.tb02602.x
- Park, C.-R., & Lee, W.-S. (2000). Relationship between species composition and area in breeding birds of urban woods in Seoul, Korea. *Landscape and Urban Planning*, *51*(1), 29–36. doi:10.1016/s0169-2046(00)00094-3
- Peay, K. G., Bruns, T. D., Kennedy, P. G., Bergemann, S. E., & Garbelotto, M. (2007). A strong species?area relationship for eukaryotic soil microbes: island size matters for ectomycorrhizal fungi. *Ecology Letters*, *10*(6), 470–480.
- Preston, F.W. 1962: The canonical distribution of commonness and rarity. Part I. *Ecology* *43*, 185–215
- Proença, V., & Pereira, H. M. (2013). Species–area models to assess biodiversity change in multi-habitat landscapes: The importance of species habitat affinity. *Basic and Applied Ecology*, *14*(2), 102–114.
- Querner, P., Milasowszky, N., Zulka, K. P., Abensperg-Traun, M., Willner, W., Sauberer, N., ... & Zechmeister, H. G. (2018). Habitat structure, quality and landscape predict species richness and communities of Collembola in dry grasslands in Austria. *Insects*, *9*(3), 81.
- Rolon, A. S., Lacerda, T., Maltchik, L., & Guadagnin, D. L. (2008). Influence of area, habitat and water chemistry on richness and composition of macrophyte assemblages in southern Brazilian wetlands. *Journal of Vegetation Science*, *19*(2), 221–228.
- Schoener, T. W (1976) The species-area relationship within archipelagoes: models and evidence from island birds. *Proceedings of XVI International Ornithological Congress*, *6*, 629-642.

- Shafer, C.L. (1990). Nature Reserves. *Island Theory and Conservation Practice*. Smithsonian Institution Press, Washington, DC.
- Sillen, B., & Solbreck, C. (1977). Effects of Area and Habitat Diversity on Bird Species Richness in Lakes. *Ornis Scandinavica*, 8(2), 185. doi:10.2307/3676103
- Simberloff, D. S. (1970). Taxonomic diversities of island biotas. *Evolution* 24: 23-47. --. 1972. Models in biogeography. In *Models in paleobiology* (T. J. M. Schopf, Ed.). San Francisco, Freeman, Cooper and Co. 1976. Experimental zoogeography of islands: effects of island size. *Ecology* 57: 629-648.
- Smith, L. A., & Chow-Fraser, P. (2010). Implications of the Species-Area Relationship on Sampling Effort for Marsh Birds in Southern Ontario. *Wetlands*, 30(3), 553–563. doi:10.1007/s13157-010-0048-4
- Townsend Peterson, A., Papeş, M., & Eaton, M. (2007). Transferability and model evaluation in ecological niche modeling: a comparison of GARP and Maxent. *Ecography*, 30(4), 550–560. <https://doi.org/10.1111/j.09067590.2007.05102.x>
- Triantis, K. A., Vardinoyannis, K., & Mylonas, M. (2005). Area and habitat relationships in island land snail faunas: an Aegean case study exploring the choros model. *Records of the Western Australian Museum*, 68, 133-141.
- Ulrich, W. (2005). Predicting species numbers using species–area and endemics–area relations. *Biodiversity & Conservation*, 14(14), 3351-3362.
- Usher, M. B., Brown, A. C., & Bedford, S. E. (1992). Plant species richness in farm woodlands. *Forestry: An International Journal of Forest Research*, 65(1), 1-13.

Wenger, S. J., & Olden, J. D. (2012). Assessing transferability of ecological models: an underappreciated aspect of statistical validation. *Methods in Ecology and Evolution*, 3(2), 260-267. <https://doi.org/10.1111/j.2041-210X.2011.00170.x>

Williams, C.B. (1964): Patterns in the balance of nature and related problems in quantitative biology. New York: Academic Press.

Yachi, S., & Loreau, M. (1999). Biodiversity and ecosystem productivity in a fluctuating environment: the insurance hypothesis. *Proceedings of the National Academy of Sciences*, 96(4), 1463-1468.

Yamaguchi, T. (2004). Influence of urbanization on ant distribution in parks of Tokyo and Chiba City, Japan I. Analysis of ant species richness. *Ecological Research*, 19(2), 209–216.  
[doi:10.1111/j.1440-1703.2003.00625.x](https://doi.org/10.1111/j.1440-1703.2003.00625.x)

Yamasaki, K., Chiba, S., & Nagahama, H. (2000). Anisotropic shape of islands and species richness of land snail fauna of the Ryukyus. *Tropics*, 10: 93-101.  
<https://doi.org/10.3759/tropics.10.93>

Zhao, J.-M., & Zhou, L.-Z. (2018). Area, isolation, disturbance and age effects on species richness of summer waterbirds in post-mining subsidence lakes, Anhui, China. *Avian Research*, 9(1).



## Appendix

**Table 1: The most accurately predicting model for each dataset (84 total) and information such as taxonomic group, R<sup>2</sup> value, prediction error (model and mean), prediction improvement of SA model from mean, number of sites per dataset, mean species richness per dataset, mean area per dataset and slope coefficient.**

Source	Mean SR	Mean Area	Country/ Continent	Best SAR Model	Taxonomy	R <sup>2</sup>	SAR Model PE	Null Mean	Prediction Improvement	# Sites	Slope Coefficient
Abarca- Arenas et al.	115.06	178.44 km <sup>2</sup>	Mexico	4	Fish	0.7768	47.73991	45.75	-0.0435	15	0.5819
Oliver et al. (2011)	68.80	1214.80 ha	U.S.	1	Birds	0.1591	43.89514	40.24	-0.0908	20	44.249
Lawesson et al. (1998)	107.30	74.29 ha	Denmark	1	Plants	0.2051	26.6901	29.375	0.0914	62	93.2114
De Camargo et al. (2015)	97.11	56 ha	Canada	3	Birds	0.00485	24.25237	23.89405	-0.01499	45	4.56092

---

Houlahan et al. (2006)	161.26	60.52 ha	Canada	1	Amphibians	0.12956	23.7959	27.3379	0.12956	42	136.9177
Leather, S. R. (1986)	28.03	2.37 km <sup>2</sup>	Britain	4	Plants	0.3655	19.9609	31.1617	0.35944	12	-36.68
Hobbs, E. R. (1988)	3.43	86.08 ha	U.S.	2	Plants	0.2091	18.1787	21.6703	0.16112	29	3.9557
Ferenc et al. (2013)	1.92	2.19 km <sup>2</sup>	Japan	2	Birds	0.6544	15.6824	25.1978	0.37763	41	3.68798
De Lacy & Shackleton (2017)	9.51	1.22 ha	South Africa	4	Birds	0.1893	12.9506	13.2143	0.01995	28	9.0838
Usher et al. (1992)	30.78	3.40 ha	England	1	Plants	0.5478	11.2277	6.37888	-0.7601	33	23.6596
Matthews et al. (2005)	31.67	1.031 ha	U.S	2	Plants	0.2826	9.56312	16.4369	0.41819	107	3.61351

---

---

Rolon et al. (2008)	28.86	9.12 ha	Brazil	4	Macrophyte	0.2147	9.41154	14.5	0.35093	15	8.173
Haig et al. (2000)	20.14	18577.57 m <sup>2</sup>	Canada	3	Plants	0.5331	9.00667	9	-0.0007	21	2.70E+00
Longato et al. (2018)	17.14	88.43 km <sup>2</sup>	Brazil	4	Microcrustacean	0.05233	8.66504	8.85874	0.02187	200	13.3117
山崎和仁 (2000)	26.60	207.43 km <sup>2</sup>	Japan	2	Mollusks	0.6723	8.62946	20.32	0.57532	20	1.05129
Benitez- Malvido & Martinez- Ramos (2003)	14.12	4.93 m <sup>2</sup>	Brazil	3	Plants	0.8356	8.38422	4.88507	-0.7163	19	1.646683
Ma & Helenius (2002)	45.17	621.52 km <sup>2</sup>	Finland	4	Plants	0.02983	8.23531	9.88726	0.16708	46	33.136

---

---

Ault et al. (1998)	3.42	32.32	Australia	2	Fish	0.5209	8.04827	15.775	0.48981	33	3.20235
Liu et al. (2007)	25.58	2.55 ha	China	1	Plants	0.4811	7.70939	11.8877	0.35148	25	12.9643
Diver (2010)	56.45	4.96 ha	Canada	4	Plants	0.8375	6.39281	16.406	0.61034	35	55.84
Khaine et al. (2017)	54.66	14958.14 m <sup>2</sup>	Myanmar	4	Plants	0.9593	6.11256	7.99808	0.23575	10	-197.126
Khaine et al. (2017)	59.86	13706.58 m <sup>2</sup>	Myanmar	1	Plants	0.04415	6.07866	6.07078	-0.0013	11	3.53E+01
Lizee et al. (2015)	35.27	59.82 ha	France	4	Insect	0.8229	6.05703	16.1045	0.62389	15	-18.396
Dos Anjos & Bocon (1999)	69.58	80.20	Brazil	3	Birds	0.322	5.99807	8.11111	0.26051	12	4.1669501

---

---

Guadagnin et al. (2009)	18.97	3254.59 ha	Brazil	1	Birds	0.2403	5.44442	5.43906	-0.001	42	1.71E+01
Helden & Leather (2004)	8.88	8819.94 m <sup>2</sup>	England	4	Plants	0.6377	5.43624	9.80769	0.44572	18	-36.68
Leather (1986)	57.90	705.26 km <sup>2</sup>	Britain	2	Plants	0.3477	4.28977	6.69663	0.35941	16	-0.8665
Oertli et al. (2002)	9.43	8216.51 m <sup>2</sup>	Switzerland	3	Coleoptera	-4.39E- 05	4.25205	6.46432	0.34223	76	2.17E+00
Chamberlain et al. (2007)	18.15	11.76 ha	England	1	Birds	0.2805	4.18919	4.64831	0.09877	208	14.86242
Bila et al. (2013)	7.4	650.87 ha	Czech Republic	2	Insect	0.758	4.13254	3.2381	-0.2762	10	-0.11272
Peay et al. (2007)	12.34	12698.67 m <sup>2</sup>	U.S	4	Ectomycorrhizal fungi	0.7103	3.71054	5.34589	0.30591	11	-0.8218

---

---

Benassi et al (2007)	11.37	7.88 ha	Italy	2	Birds	0.9106	3.62831	4.78648	0.24197	10	1.62109
Lucey et al. (2014)	26.98	3.92 ha	Malaysia	2	Insects	0.4215	3.48982	5.76857	0.39503	10	2.4246
Holyoak et al. (2005)	30.81	3.15 km <sup>2</sup>	England	4	Mollusks	0.6276	3.47237	7.91667	0.56138	11	23.911
Li et al. (2018)	9.73	8566.32 m <sup>2</sup>	U.S	1	Birds	0.01137	3.31293	3.72759	0.11124	242	9.417134
Boecklen (1986)	15.90	7.79 ha	N/A	3	Birds	0.2228	3.28767	6.42425	0.48824	103	2.432183
Li et al. (2018)	13.66	6019.12 m <sup>2</sup>	U.S	3	Plant	0.01531	3.25897	3.48736	0.06549	240	2.22E+00
Zhao & Zhou (2018)	6.19	85.77 km <sup>2</sup>	China	3	Birds	0.08882	3.24822	4.28004	0.24108	38	1.8243079
Leather (1986)	9.86	2.58	Britain	3	Plants	0.6288	3.23749	27.3101	0.88145	16	1.3640297

---

---

Maltchik et al. (2010)	4.76	10.01 ha	Brazil	4	Fish	-0.1017	3.09795	3.08333	-0.0047	13	3.8147
Oertli et al. (2002)	9.77	8393.54 m <sup>2</sup>	Switzerland	4	Odonata	0.07824	2.99697	3.87298	0.22618	79	3.7167
Oertli et al. (2002)	5.73	11026.19 m <sup>2</sup>	Switzerland	4	Gastropod	0.05945	2.97971	3.03338	0.01769	54	2.5556
Honnay et al. (1999)	76.95	764327.79 m <sup>2</sup>	Belgium	4	Plants	0.9809	2.85855	7.5125	0.61949	50	-119.6659
Hattori & Shibuno (2009)	6.72	3.94 m <sup>2</sup>	Japan	4	Fish	0.5063	2.6429	2.88024	0.0824	38	4.9442
Figueroa et al. (2018)	13.10	10.56 ha	Chile	2	Plants	0.118	2.53928	4.51798	0.43796	48	1.5084
Jackson (2005)	15.70	16.91 m <sup>2</sup>	Australia	1	Plants	0.5088	2.44261	5.78909	0.57807	15	8.1844

---

---

Oertli et al. (2002)	2.30	10021.35 m <sup>2</sup>	Switzerland	2	Sphaeriidae	-0.01844	2.38698	2.51863	0.05227	71	0.562
Helm et al. (2005)	47.78	3.12 km <sup>2</sup>	Estonia	3	Plants	0.01558	2.34295	2.68685	0.128	34	3.82228
Lee & Carroll (2018)	10.43	630.19 m <sup>2</sup>	U.S.	3	Plants	0.008491	2.30838	2.61364	0.1168	82	2.33E+00
Oberdorff et al. (1993)	13.40	42.63 ha	England	4	Insects	0.5364	2.29109	4.9375	0.53598	22	1.5626
Yamaguchi (2004)	7.63	5724.53 m <sup>2</sup>	Japan	4	Insects	0.2814	2.2648	2.14868	-0.054	45	-0.7483
Brunet & Medellín (2001)	3.31	0.00040 km <sup>2</sup>	Mexico	2	Mammals	0.3106	2.21018	3.19231	0.30765	19	4.6361
Huang et al. (2015)	9.65	22.29 ha	China	3	Birds	0.6383	2.17716	2	-0.0886	26	1.67464

---



---

Sillen & Solbreck (1977)	26.61	7.03 ha	Sweden	4	Birds	0.4431	2.10666	3.32609	0.36663	31	5.79987
Elliott et al. (2019)	9.82	2683.74 ha	U.S.	2	Birds	0.9853	2.00635	5.64689	0.6447	34	1.834773
Benedick et al. (2006)	38.03	3.43 ha	Malaysia	1	Insects	0.772	1.97999	4.95457	0.60037	10	12.047
Kimmerer et al. (2000)	9.20	16.06 m <sup>2</sup>	U.S	1	Bryophyte	0.04913	1.94258	1.84537	-0.0527	38	8.64E+00
Park et al. (2000)	10.73	73.01 ha	S. Korea	2	Birds	0.7651	1.81199	2.48225	0.27002	38	1.60348
Bellamy et al. (1996)	6.32	1.97 ha	England	4	Birds	0.5972	1.51179	2.67422	0.43468	105	6.189
Mason (2001)	18.04	8.64 ha	England	2	Birds	0.7611	1.36441	1.35213	-0.0091	33	2.42724

---

---

Eckmann (1995)	10.12	2.24 km <sup>2</sup>	Germany	4	Fish	0.928	1.18605	0.5	-1.3721	16	10.715
Lawes et al. (2005)	3.53	27.62 m <sup>2</sup>	Africa	2	Plants	0.03363	1.18532	1.24891	0.05092	31	1.19597
Oertli et al. (2002)	5.66	8868.77 m <sup>2</sup>	Switzerland	3	Amphibians	-0.01364	1.16324	1.46158	0.20412	53	1.723237
Andrade & Marini (2002)	5.35	0.42 ha	Brazil	3	Bird	0.7178	1.14454	1.55	0.26159	14	0.8867
Querner et al. (2018)	5	12866.64 m <sup>2</sup>	Arthropod	1	Insects	0.190144	1.12521	1.16008	0.03006	50	4.88E+00
Yamaguchi (2004)	11.72	4388.61 m <sup>2</sup>	Japan	4	Insects	0.7934	1.11956	0.9407	-0.1901	43	-8.0124
Helden & Leather (2004)	24.77	8819.94 m <sup>2</sup>	England	1	Plants	0.09478	1.03453	1.35385	0.23586	18	5.24E+00

---

---

Natuhara & Imai (1999)	8.14	14.06 ha	Japan	4	Birds	0.5925	0.94971	2.3	0.58708	28	4.8145
Hasegawa et al. (2016)	4.47	353.70 m <sup>2</sup>	Japan	4	Fish	0.4831	0.90457	1.02542	0.11785	46	-2.4037
Flaspohler et al. (2010)	4.66	9.38 ha	Hawaii	4	Birds	8.03E-01	0.62202	2.58462	0.75934	18	4.385
Benitez- Malvido & Martinez- Ramos (2003)	12.20	18.65 m <sup>2</sup>	Brazil	4	Plants	0.86	0.60719	1.67436	0.63736	33	-0.7631
Nicolas et al. (2010)	13.25	24848.34 m <sup>2</sup>	France	4	Fish	0.9303	0.57544	3.13794	0.81662	17	-42.9449

---

---

Smith & Chow- Fraser (2010)	10.31	409.47 ha	Canada	2	Birds	0.4341	0.44262	0.62778	0.29493	19	2.22089
Brose (2008)	3	6.03 m <sup>2</sup>	Germany	2	Plants	-0.00716	0.34025	0.40041	0.15025	55	0.8962
Benitez- Malvido & Martinez- Ramos (2003)	9.61	28.50 m <sup>2</sup>	Brazil	4	Plants	0.9954	0.25771	3.13651	0.91784	60	-13.4567
Benitez- Malvido & Martinez- Ramos (2003)	5.27	31.31 m, <sup>2</sup>	Brazil	4	Plants	0.9313	0.18453	0.33652	0.45166	20	0.40603

---

---

Benitez- Malvido & Martinez- Ramos (2003)	5.89	17.34 m <sup>2</sup>	Brazil	2	Plants	0.9945	0.15124	1.92892	0.92159	20	1.524741
Benitez- Malvido & Martinez- Ramos (2003)	3.29	30.51 m <sup>2</sup>	Brazil	4	Plants	0.9802	0.10714	0.57313	0.81306	60	0.16842
Benitez- Malvido & Martinez- Ramos (2003)	2.29	28.93 m <sup>2</sup>	Brazil	4	Plants	0.972	0.10685	0.61826	0.82717	58	-0.20955

---

---

Benitez- Malvido & Martinez- Ramos (2003)	2.81	18.34 m <sup>2</sup>	Brazil	4	Plants	0.9841	0.10543	0.6122	0.82779	36	0.16842
Benitez- Malvido & Martinez- Ramos (2003)	1.63	10.97 m <sup>2</sup>	Brazil	4	Plants	0.9548	0.1023	0.30151	0.66071	19	0.16345
Benitez- Malvido & Martinez- Ramos (2003)	3.10	23.12 m <sup>2</sup>	Brazil	4	Plants	0.9825	0.09947	0.91492	0.89128	20	0.31561

---