

Introduction: Allocating limited conservation resources requires identifying regions of high importance for biodiversity both now and in the future as species shift in response to global change. Current biodiversity-based efforts to prioritize regions for conservation focus almost exclusively on biodiversity maps derived from geographic range map data for one or more taxa. These data are available for many taxa at continental to global scales, but have two major weaknesses: 1) they are static, which makes it difficult to evaluate how biodiversity patterns will respond to changes in climate and land use; and 2) they do not directly reflect biodiversity at the high resolution needed for conservation¹. It is therefore imperative to provide an alternative method for quantifying and predicting biodiversity patterns, by using dynamic data that are more closely linked to conditions in local communities.

Due to extensive government and citizen science efforts, survey data are increasingly available to fill this need. In contrast to range maps, surveys provide direct observations of species richness (number of species) or abundance (number of individuals) in the field. Comparisons of range map and survey data show large disparities in biodiversity estimates². For example, my previous work has shown that these disparities are reflected and magnified in large-scale conservation prioritizations. Specifically, locations of hotspots determined from both data types have as little as 15% overlap in location, emphasizing the need to leverage survey data to inform conservation analyses.

In order to use survey data for comprehensive conservation analyses, we need a means to fill gaps in data availability across spatial and taxonomic extent. Current practices focus on filling spatial gaps with species distribution models (SDM), which use environmental data and known occurrences to predict species occurrence in unsurveyed areas. For multi-species assemblages, this approach simply stacks SDMs for individual species, predicting each species' occurrence independently³. Because the assumption of independence is rarely if ever met in real communities, subsequent predictions yield unrealistic results that consistently over predict the number of species in an area⁴. New advances in joint species distribution models (JSDM) seeks to address this failure by predicting assemblages as a whole while accounting for the co-occurrence of individuals within the assemblage. While JSDMs succeed in providing more realistic assemblages across space and time⁴⁻⁶, JSDMs' improved representation of ecology also present an unique opportunity to address taxonomic holes in survey data through multi-taxa prediction. My project begins where current JSDM research leaves off, by developing an approach to predict assemblages not only across space but across taxa.

Research Plan: I will test the hypothesis that JSDMs can use one taxon to inform the prediction of another with a latent variable modeling approach to predict communities of birds and butterflies. Though multiple statistical approaches to JSDMs exist⁴⁻⁶, latent variable model's ability to predict multiple response variables (species) with a reduced number of parameters make it especially suited to the task of predicting multi-taxa communities⁷. Latent variable models use unobserved ("latent") predictors to account for unmeasured environmental or biotic variables⁷. Taxa inhabiting similar scales and ecological spaces, such as birds and butterflies, are likely to respond to similar latent variables. This approach therefore provides a framework to use data rich birds to inform the prediction of data poor butterflies⁸. My project will consist of three phases: data preparation, model development, and model evaluation.

Phase 1: The first step in model creation is to ensure all data are on the same scale. I will divide the landscape into a grid and treat each 1° cell as a separate community, with information on species occurrence for birds and butterflies within each cell, and corresponding climate data for that location. Data will be compiled from a number of large-scale survey efforts in North

America including the North American Breeding Bird Survey (BBS) and eBird for birds and the North American Butterfly count for butterflies. I will use PRISM climate data for the climatic predictor variables. Cells will be split into testing and training data sets in a checkerboard pattern⁷, with separate data sets for birds and butterflies drawn from cells with complete information for that taxon.

Phase 2: I will then fit a suite of latent variable models to the bird training data and PRISM data using the R package lavaan⁹. Each model will include a number of latent variables ranging from 1-5 to address the trade off between model complexity and more accurate representation of unmeasured variables.

Phase 3: I will select the model with the best predictive ability for birds and evaluate its ability to predict butterflies. I will use cross-validation to determine the model best able to predict bird assemblages in the testing data. The latent variables from the chosen model will then be used in a model for butterfly prediction, built based on butterfly training cells. I will then evaluate the butterfly model for its ability to predict the butterfly testing data using cross-validation.

One of the limitations of latent variable models is that potential poor prediction of butterflies by the bird latent variables could be attributed to a number of ecological scenarios: birds and butterflies could be 1) responding to different environmental variables, 2) responding to the same variables but in different ways, 3) not responding strongly to environmental variables. I will therefore also validate my model against simulated bird and butterfly assemblages for each of those three scenarios to gain ecological insight into prediction shortfalls. This will inform improvement of bird and butterfly models through the inclusion of covariates that can control for the relationships occurring in each scenario.

Intellectual Merit: This will be the first JSDM that predicts assemblages for multiple taxa, an essential methodological task for applications in conservation and management. It would provide an important illustration of the plausibility of using information rich survey data for large-scale biodiversity analyses. Even potential failure of bird latent variables to predict butterfly assemblages offers a clear opportunity to understand the underlying ecological scenario at play, and inform better prediction in the future.

Broader Impacts: Results of this project would be directly relevant to conservation and management action. Predictions from the successful model will be used to create comprehensive maps of bird and butterfly assemblages for North America. Maps for individual species such as the Monarch butterfly (*Danaus plexippus*), a species being considered for listing under the Endangered Species act and of special concern for government and conservation agencies, will also be created. Products of this research, including maps and software created, will be made publicly available for use by citizen science efforts and conservation decision makers through the ROpenSci project. In tandem with this project, I will continue to use my unique perspective as a female conservation ecologist using advanced quantitative methods to encourage these skills at multiple levels. First, I will develop curriculum for introducing undergraduate ecologists to the power of basic quantitative approaches as a member of the Berkeley Institute for Data Science Education and Training working group. Second, I will organize workshops for middle school girls illustrating the many possible career paths in math and science through my position on the planning committee of the Berkeley Expanding Your Horizons network.

References: 1. Hurlbert, A. H. & Jetz, W. *PNAS*. 104, 13384–13389 (2007). 2. Hurlbert, A. H. & White, E. P. *Ecol. Lett.* 8, 319–327 (2005). 3. Thuiller, W. et al. *T. Ecol. Lett.* 18, 1321–1328 (2015). 4. Clark, J. S. et al. *Ecol. Appl.* 24, 990–999 (2014). 5. Harris, D. J. *Methods Ecol. Evol.* 6, 465–473 (2015). 6. Pollock, L. J. et al. *Methods Ecol. Evol.* 5, 397–406 (2014). 7. Warton, D. I. et al. *Trends Ecol. Evol.* 30, 766–779 (2015). 8. Thorson, J. T. et al. *Methods Ecol. Evol.* 6, 627–637 (2015). 9. Rosseel, Y. *Stat. Softw.* 48, (2012).