Christian Stratton
christianstratton@montana.edu

Research statement

John Tukey once said, “The best thing about being a statistician is that you get to play in everyone’s backyard.” This quote is an adage I have embraced through my research by developing statistical methodology to answer scientific questions posed by many collaborators in the environmental and ecological sciences. My collaborative work has collectively established my research identity in environmental statistics, broadly working on Bayesian statistical methodology for species distribution models (SDMs). The diverse statistical methods for species distribution modeling have also exposed me to related research fields, including Bayesian computation, spatial statistics, data integration techniques, and Bayesian nonparametrics. Beyond the environmental and ecological sciences, I have research interests and experience in statistics education. Below, I describe my contributions to these fields and discuss future research directions.

Completed work

My initial exposure to statistical research began during my master’s program when I developed a computationally efficient Markov chain Monte Carlo (MCMC) method for estimating multi-scale occupancy models. This effort resulted from a collaboration with U.S. Geological Survey (USGS) scientists interested in using environmental DNA to assess the presence of a parasite leading to massive fish kills in the Yellowstone River. To assist timely assessments of the presence of the parasite along the Yellowstone River, I developed an R package that implemented a data augmentation strategy allowing for fully Bayesian estimation of multi-scale occupancy models via a Gibbs sampler, resulting in significantly faster model estimation than pre-existing methods. This work is published in *Methods in Ecology and Evolution*, and is provided to the public via the `msocc` `R` package.

Following the collaboration with USGS scientists on the Yellowstone River, I worked with National Park Service (NPS) scientists to assess the status and trend of sagebrush steppe communities in Grand Teton National Park. My role in this work entailed using simulation to determine the best distance-based ordination techniques for the sagebrush data and implementing those techniques on multiple large data sets. I then constructed an R Shiny web application that allowed dissemination of this work among NPS scientists. This work is published as a peer-reviewed natural resource report for the National Park Service.

During my doctoral program, I worked with Dr. Jennifer Green in statistics education, expanding my research interests beyond environmental and ecological statistics. For this project, I developed an R Shiny web application to facilitate the teaching of statistical power to undergraduate and graduate-level statistics students. The web application allows for visualization of a sample statistic’s power function for various population distributions and sample statistics, with additional knobs enabling changes in the significance level, sample size, and alternative hypothesis. The application also allows students to select points on the power function and visualize the implied sampling distribution under the null hypothesis and true value of the parameter, facilitating understanding of the relationship between statistical power and sampling distributions. Following development, I implemented the web application alongside a guided activity in the undergraduate and graduate-level mathematical statistics at Montana State University. The details of the web application and its implementation are published in *Technology Innovations in Statistics Education*.

As I transitioned into my post-doctoral appointment, my focus again shifted to the environmental and ecological sciences as I investigated the influence of priors and calibration data on multi-species, false-positive count detection modeling frameworks. Researchers often rely on calibration data to estimate the classification probabilities when using data prone to misclassification in species distribution models. When calibration data are unavailable, informative Bayesian priors are often used instead. My work, published in *Methods in Ecology and Evolution*, investigated the influence of calibration data and priors on parameter estimation in the multi-species count detection framework and provided recommendations practitioners.

Current work

As a post-doctoral researcher working with Dr. Kathryn Irvine at the U.S. Geological Survey, my current research predominantly concerns the development of statistical methodology to support the North American Bat Monitoring Program (NABat). The NABat program is a multi-agency, multi-discipline, collaborative monitoring program that aims to assess the status and trend of bat populations in North America, with particular emphasis on evaluating the response of bat populations to stressors such as disease and wind energy. Recently, my research in this area has concerned the development of a modeling framework to assess the impact of a wildlife disease called White-nose Syndrome (WNS) on North American bat populations. Before my involvement, this assessment was typically made by first modeling the disease spread to obtain some summary of disease occurrence across the landscape. Then, a measure of species distribution is regressed on that disease summary, resulting in an estimate of the impact of the disease. However, this approach does not account for the uncertainty present in estimating the disease occurrence, resulting in overly precise estimates of the impact of the disease on the species distribution. A potential solution to this problem is to jointly model the two data sources in a single spatially misaligned regression model, allowing for appropriate uncertainty propagation throughout the model. This work is under review in the *Journal of Agricultural, Biological, and Environmental Statistics* and was presented at the Joint Statistical Meetings in 2023.

Another avenue of active research in support of NABat is the development of data integration techniques for ecological data. Modern wildlife monitoring programs often rely on multiple data sources to infer species distribution and abundance. For example, when providing analytic support for listing three bat species under the Endangered Species Act, we were required to submit analyses of multiple data sources, including stationary acoustic data, mobile acoustic data, capture data, and roost count data. Recently, there has been interest in integrating multiple data sources under a single unified framework. Previous attempts at data integration in ecology typically rely on defining a multi-state observation process that accommodates the various data sources. My current research focuses on developing a single hierarchical model, motivated by a spatial point process, that can accommodate multiple ecological data sources. This model was recently implemented for a regulatory clearance process for tri-colored bats. The manuscript detailing this framework is currently in preparation.

Outside of my work on behalf of NABat, I am currently working to improve ordination of vegetation communities in Grand Teton National Park as a follow-up to previous work with the National Park Service. Traditional ordination techniques rely upon a multi-step, distance-based process: first, the community data are converted to a dissimilarity matrix, where each entry describes the dissimilarity between sample locations based on some metric. Then, the distance matrix is projected into a lower dimension, often two, where the results are finally clustered based on their proximity in the lower dimension space. This multi-step process requires several subjective decisions (choice of dissimilarity index, projection technique, number of dimensions, number of clusters, etc.) that cannot be formally assessed, as the distance-based framework does not present a likelihood. To address these shortcomings, I have developed a hierarchical, model-based ordination framework that uses a Dirichlet process mixture to perform the clustering. This framework allows for probabilistic inferences about the number of clusters in Grand Teton National Park and provides tools for model comparison. This work is currently in review in *Methods in Ecology and Evolution.* and was presented at the Joint Statistical Meetings in 2022.

Future work

My future work will continue researching statistical methodology for the environmental and ecological sciences through four additional avenues: 1) continued development of methodology to accommodate the idiosyncrasies of acoustic data, 2) further development of the joint spatial modeling framework to better reflect the observation process, 3) inclusion of additional streams of data in the data integration framework, including roost count information, and 4) advancement of Bayesian computational methods.

First, I will continue to develop statistical methodology that leverages more of the acoustic data pipeline for species distribution models. Current analytic frameworks remove audio recordings that are not assigned a single species label before analysis, resulting in a large volume of data being discarded before analysis, potentially biasing resulting inferences. In the next three to five years, I plan to develop modeling frameworks that accommodate the downgrading of acoustic recordings to multi-species labels and frequency groups, reducing the burden to collect data on monitoring programs that use acoustic data.

Second, I will further development of the spatially misaligned regression model developed for the WNS project. The first outstanding question I will investigate is that of optimal sampling design. Previous investigations into optimal designs for spatial modeling typically focus on guidance for selecting sample locations to measure one variable. For the joint modeling framework, it would be valuable to consider how to distribute effort for both the acoustic monitoring and disease surveillance data sources. Remaining opportunities for further development include adding complexity to the disease observation process to accommodate false-positive detections and adding temporal dynamics to the disease process.

Third, I plan to further develop data integration techniques for ecological monitoring programs. Specifically, I will investigate how to incorporate roost counts into the integrated model currently constructed for the NABat program. This problem is nuanced, as the roost count locations represent a spatial point process that is prone to observation error and misaligned with the current monitoring effort. Beyond the work related to the NABat program, data integration is broadly of interest for a wide variety of taxa due to the advent of citizen science data sources. My future work will explore how to integrate citizen science data sources with design-based monitoring programs to obtain unbiased inferences about species distribution.

Finally, I will generally seek to develop computational improvements for a variety of ecological models. I have received frequent communication from researchers across the world about the computational value of the `msocc` package, and computational improvement to existing ecological models is highly desired by practicing ecologists. As increasingly large and complex data becomes more accessible, the need for scalable analytic methods increases. My future work in this area will seek to make computationally efficient Bayesian techniques more accessible to practitioners. For example, by incorporating data augmentation strategies, the regression coefficients in the spatially misaligned regression model developed for the WNS project could be estimated via a Gibbs sampler. Additionally, the model is not well suited for large spatial domains. I plan to explore potential solutions to this problem, possibly considering locally approximated or nearest neighbor Gaussian processes.

To John Tukey's point, my time in the backyard of ecologists has given me the opportunity to explore a number of different research topics, all unified under the development of statistical methodology for the environmental and ecological sciences. Beyond the environmental and ecological sciences, I have interest and experience in statistics education. However, as I look to begin the next chapter in my life as statistics faculty, I am excited by the opportunity to forge new collaborative relationships with the diverse and talented faculty at [**SCHOOL**] and continue to explore backyards.