Received: 16 October 2023 | Revised: 26 December 2023 | Accepted: 31 January 2024

DOI: 10.1002/ecy.4283

CONCEPTS & SYNTHESIS

ECOLOGY
ECOLOGICAL SOCIETY OF AMERICA

A century of statistical *Ecology*

Neil A. Gilbert[1,2] | **Bruna R. Amaral**[1,2] | **Olivia M. Smith**[1,2,3] | **Peter J. Williams**[1,2] | **Sydney Ceyzyk**[2] | **Samuel Ayebare**[1,2] | **Kayla L. Davis**[1,2] | **Wendy Leuenberger**[1,2] | **Jeffrey W. Doser**[1,2] | **Elise F. Zipkin**[1,2]

[1]Ecology, Evolution, and Behavior Program, Michigan State University, East Lansing, Michigan, USA

[2]Department of Integrative Biology, Michigan State University, East Lansing, Michigan, USA

[3]Center for Global Change and Earth Observations, Michigan State University, East Lansing, Michigan, USA

Correspondence
Neil A. Gilbert
Email: neil.allen.gilbert@gmail.com

Funding information
Division of Environmental Biology, Grant/Award Numbers: NSF DEB-2016347, NSF DEB-2213565; Directorate for Biological Sciences, Grant/Award Number: 2208894; Division of Biological Infrastructure, Grant/Award Number: NSF DBI-1954406

Handling Editor: Kathryn L. Cottingham

Abstract

As data and computing power have surged in recent decades, statistical modeling has become an important tool for understanding ecological patterns and processes. Statistical modeling in ecology faces two major challenges. First, ecological data may not conform to traditional methods, and second, professional ecologists often do not receive extensive statistical training. In response to these challenges, the journal *Ecology* has published many innovative statistical ecology papers that introduced novel modeling methods and provided accessible guides to statistical best practices. In this paper, we reflect on *Ecology*'s history and its role in the emergence of the subdiscipline of statistical ecology, which we define as the study of ecological systems using mathematical equations, probability, and empirical data. We showcase 36 influential statistical ecology papers that have been published in *Ecology* over the last century and, in so doing, comment on the evolution of the field. As data and computing power continue to increase, we anticipate continued growth in statistical ecology to tackle complex analyses and an expanding role for *Ecology* to publish innovative and influential papers, advancing the discipline and guiding practicing ecologists.

KEYWORDS
data, history of science, model selection, quantitative ecology, statistical ecology, uncertainty

ECOLOGY: AN IMPORTANT VENUE FOR STATISTICAL ECOLOGY

Ecology is a generalist journal. It publishes papers on diverse taxa across many biomes, addressing a wide variety of ecological questions. Papers focused on the development of ecological research methods have been central to *Ecology*'s niche, dating back to the journal's inception. For example, in its third year, *Ecology* published a critique of the Arrhenius formula, which models species–area relationships, focusing on experimental design considerations that explained why the formula failed to fit empirical data (Gleason, 1922). Of the methods papers appearing in *Ecology*, a notable subgenre is statistical ecology, particularly in recent decades. Our goal in this article is to highlight influential statistical ecology papers that have appeared in *Ecology* and, in so doing, trace the development of important concepts and methods in the analysis of ecological data. The 36 papers that we showcase appear in a collection of *Ecology* articles, highlighting a century of statistical innovations.

WHAT *IS* STATISTICAL ECOLOGY?

Statistical ecology is the study of ecological systems using mathematical equations, probability, and empirical data. Statistical ecology is closely related to the discipline of theoretical (or mathematical) ecology. Two hallmarks of statistical ecology set it apart from theoretical ecology. First, statistical ecology holds a data-first perspective. Analysis and interpretation of data—whether empirical or simulated—is at the forefront of statistical ecology research. This contrasts with theoretical modeling efforts in which the focus is to explore behaviors of complex systems, with application to data being of secondary importance. The second hallmark of statistical ecology is uncertainty. Statistical ecology is concerned with estimating or predicting some ecological quantity, as well as the uncertainty associated with that estimate or prediction. This contrasts with theoretical approaches, which often provide deterministic outputs of mathematical formulas where quantifying uncertainty is not a central focus (although stochastic mathematical models are an exception). All told, it is difficult to draw strict lines between subdisciplines, and thus there is ambiguity in the definition. However, for the purposes of this retrospective, we define statistical ecology as the empirical effort to explain or predict ecological phenomena using mathematical equations with associated measures of probabilistic uncertainty.

IDENTIFYING INNOVATIVE PAPERS ON STATISTICAL ECOLOGY

We performed a literature review to identify innovative statistical ecology papers that have appeared in *Ecology* throughout its 100-year history. On 15 February 2023, we downloaded the full records of papers published in *Ecology* from its inception to the February 2023 issue ($n = 17{,}589$) from Web of Science. We then screened titles and abstracts (or first pages when abstracts were not present) of all papers (removing nonresearch items such as errata, editorials, and book reviews) and classified whether or not each paper focused on statistical ecology. We used the data-first and uncertainty criteria to distinguish statistical ecology papers from theoretical and mathematical papers. We considered a "focus on statistical ecology" as a paper that described a novel modeling framework, extended previously described models, or provided synthesis or guidance on applying statistical methods; simply applying an established statistical method did not qualify. We identified 1447 statistical ecology papers, representing 8.8% of research papers published in *Ecology* in its first 103 years (Gilbert, 2024). Unsurprisingly, the proportion of statistical ecology papers in *Ecology* each year has increased through time (Figure 1). Fast upward growth of statistical ecology papers began around 1970, coinciding with advances in computing technology; the proportion peaked in the early 2000s and dropped after 2010, likely due to the growth of methods-focused journals such as *Methods in Ecology and Evolution*.

We used the number of citations to help identify innovative papers, reasoning that influential papers get more citations (Teplitskiy et al., 2022), though we acknowledge this approach may create gender bias in the selected papers due to gender homophily in citation behavior (Zhou et al., 2024). We retained statistical ecology papers that had at least as many citations as the 75th percentile of citations of all research articles in *Ecology* for that year, leaving us with 424 articles. This represents 29.3% of all statistical ecology papers ($n = 1447$) and 9.3% of all research articles above the 75th percentile of citations ($n = 4550$), indicating that papers focused on statistical ecology are slightly overrepresented in *Ecology*'s highly cited papers relative to their baseline prevalence. We subsequently read each paper to classify it into seven (nonmutually exclusive) themes (Figure 2): (1) models for individuals (12.0% of the 424 highly cited statistical ecology papers), (2) population models (37.3%), (3) methods for communities (30.5%), (4) methods for ecosystems (6.1%), (5) spatial methods (19.9%), (6) model selection and evaluation (5.2%), and (7) tools and best practices (17.0%). Note that the reported percentages do not sum to 100 since a paper could be classified into multiple categories.

We selected 36 papers to highlight in this collection, choosing to showcase several papers among the highest citation rank per category (Figure 3). We focused on papers that we believe have impacted the direction of statistical ecology as well as ecology more broadly. Since our objective is to relay a story of how statistical ecology has grown over the decades—rather than giving an overview of cutting-edge statistical methods—some of the work described herein no longer represents best practice in data analysis. Finally, we acknowledge that the authorship of these highly cited papers (particularly from the earlier years) reflect the exclusionary history of ecology and hope that, moving forward, statistical ecology will see increasing participation by more diverse ecologists (Beck et al., 2014; Gilbert, 2018; Martin, 2012; Whelan & Schimel, 2019).

THEME 1: MODELS FOR INDIVIDUALS (FIVE PAPERS)

Statistical ecology papers within this category seek to understand the movements and behaviors of individual

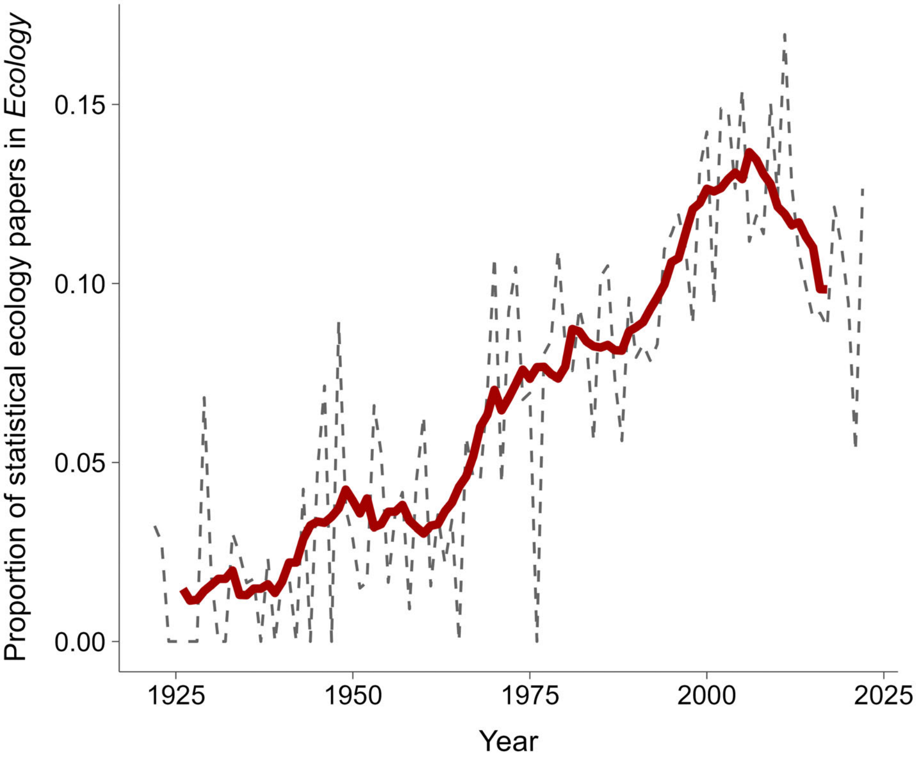

FIGURE 1 Time series of proportion of papers published in *Ecology* with a focus on statistical ecology from 1920 to 2022, the lifetime of the journal. Thick red line: 10-year moving average.

animals. With the advent of radio telemetry in the late 1950s and 1960s and the subsequent explosion of wildlife tracking efforts (Benson, 2010), researchers needed corresponding statistical methods to make sense of such data. Analysis objectives with tracking data include the estimation of home ranges, understanding habitat selection, and characterizing movement patterns. Methodological development in these areas remains very active, particularly given the blossoming of "high-throughput" movement data (Nathan et al., 2022) and miniaturization of tracking devices enabling application to taxa as small as insects (Knight et al., 2019).

Estimating home ranges of animals from location data at first seems trivial (just plot points on a map) but comes with many statistical challenges. **Worton (1989)** introduced a method for kernel smoothing of animal locations to characterize home ranges. This nonparametric method was an improvement over parametric methods because it accommodated multimodal distributions (i.e., noncircular home ranges). The paper also provided a cross-validation approach for selecting an appropriate smoothing parameter. Ensuing years saw continued advances in home range modeling, including **Fleming et al. (2015)**, who presented a method for estimating animals' home ranges by viewing autocorrelation in location data as information to be used, rather than a nuisance parameter. Traditional kernel density methods—as in Worton (1989)—assume independence of such data and as a result dramatically underestimate home range sizes and perform poorly when confronted with new data.

Evaluating resource selection is a major focus of studies that use data from individual animals. **Johnson (1980)** introduced a method for characterizing usage and availability (e.g., of foods or habitats) by ranking. This approach was innovative because it provided consistent estimates of preference and avoidance even when changing the "components" (foods, habitat types) included in the analysis and because it translated to existing methods of statistical hypothesis testing. In addition, the paper's discussion outlined the hierarchical nature of habitat selection (from a species' geographic range to the selection of food resources within an individual's home range), a framework that is now ubiquitous in animal ecology and has established the paper as a citation classic. **Northrup et al. (2013)** provided practical guidance for defining available locations when performing resource selection analyses. Using simulations, they demonstrated that spatial mismatch in availability (i.e., drawing random "available" locations from a spatial extent different from what the animal can access/use) leads to biased estimates of resource preference. Spatial autocorrelation in environmental variables exacerbated this problem. Because defining resource availability is crucial—but largely arbitrary—for such

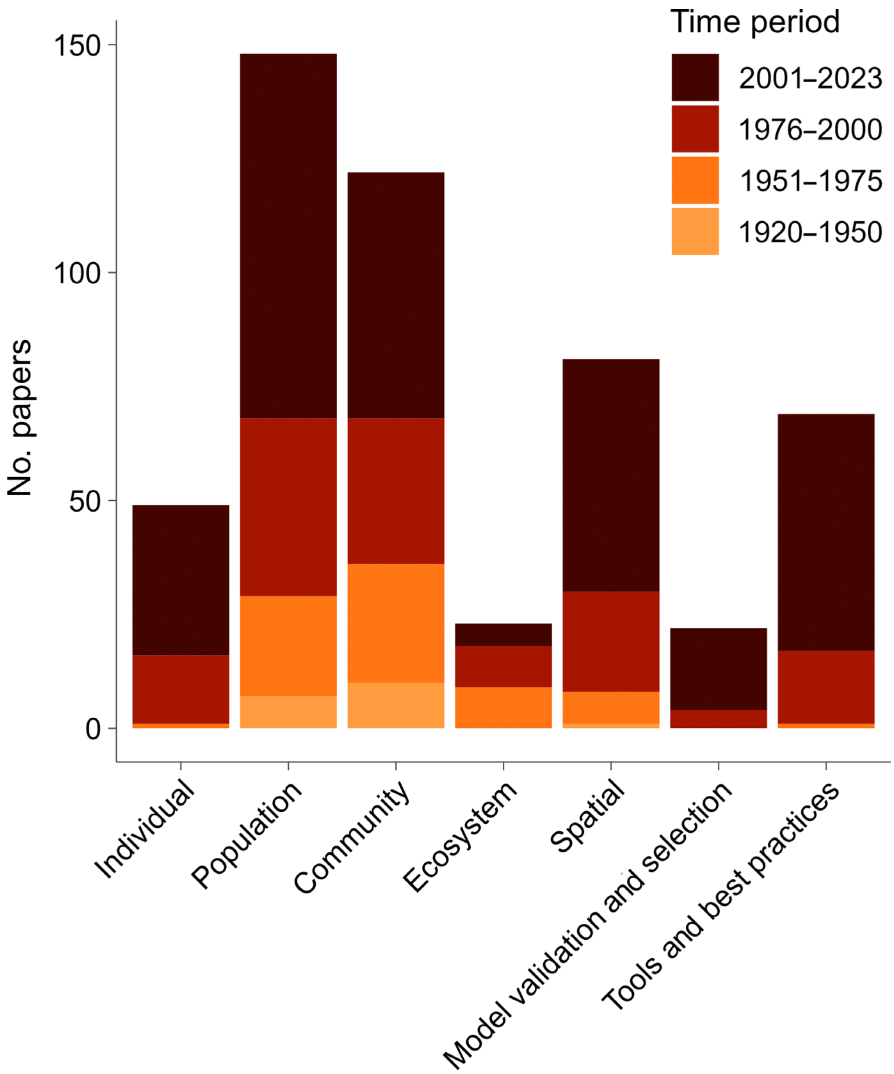

FIGURE 2 Prevalence of each theme by time period within sample of 424 highly cited statistical ecology papers reviewed in detail.

analyses, this paper is a valuable read for anyone performing resource selection studies.

Finally, movement ecology has emerged as a subdiscipline of ecology, unified by the view that characterizing an individual's movement can lend insight into physiological, behavioral, and population questions (Converse et al., 2022; Patterson et al., 2008). State-space models have been an important tool in this arena, as they attempt to distinguish ecological reality (an animal's movement) from the data collection or observation process. **Jonsen et al. (2005)** demonstrated how state-space modeling could be applied to location data to make inferences on animal movement as well as behavioral states. Important features of their model include (1) its ability to accommodate complex error structure (e.g., data from irregular time intervals) and (2) model-based filtering of extreme (erroneous) locations, removing the need for ad hoc filtering of location data. They applied their model to tracking data for seals in the North Atlantic, distinguishing between migrating and foraging behaviors.

THEME 2: MODELS FOR POPULATIONS (SIX PAPERS)

Central goals for population modeling include estimating species distributions, population abundance (or density), demographic parameters (e.g., survival), and population trends over time. Population modeling attracted considerable attention prior to and concurrent with the inception of *Ecology* with the work of scientists such as Malthus, Lotka, and Volterra. However, we believe that this work—and many population studies that appeared in *Ecology* in subsequent decades (e.g., De Bach & Smith, 1941)—are better classified as theoretical or mathematical modeling. Nevertheless, statistical models of populations saw many

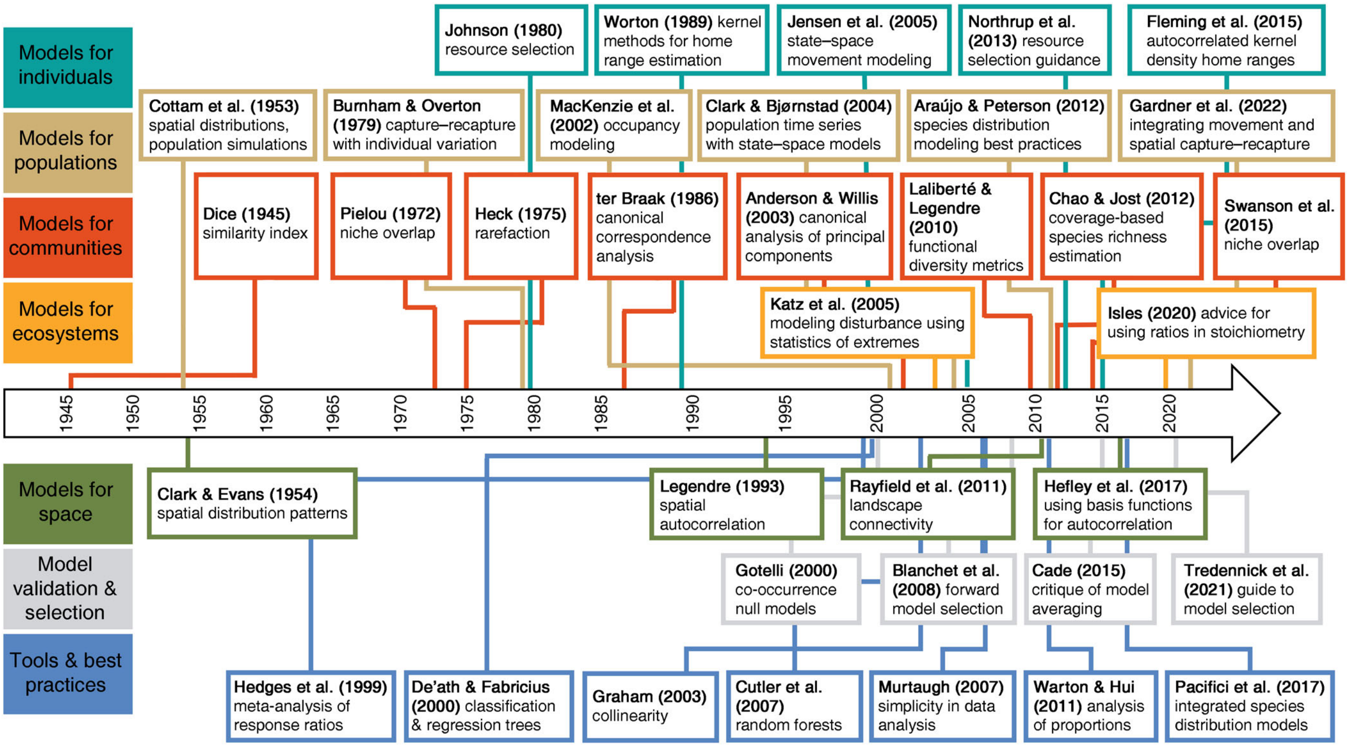

FIGURE 3 Timeline of 36 statistical ecology papers highlighted, classified into seven themes.

advances in the first several decades of *Ecology*'s existence. For example, **Cottam et al. (1953)** evaluated methods for assessing whether spatial patterns of plants are random or nonrandom, an important consideration when estimating population density. They found that accurate estimates could be made with multiple sampling methods (quadrats, distances between individuals), assuming that a suitable number of individuals was sampled. Perhaps the most notable aspect of the paper is the authors' use of a computer-simulated population of individuals, a considerable effort for the early 1950s. Indispensable for understanding and validating methods, data simulation remains a core component of statistical ecology today (DiRenzo et al., 2023). Unlike methods for plants, estimating population parameters for animals faces challenges of animal movement and imperfect detection. Capture–recapture modeling, broadly considered the gold standard of estimating population abundance as well as various demographic parameters for animals, has roots coinciding with *Ecology*'s beginning and has experienced extensive development over the last century (Seber & Schofield, 2019). For example, **Burnham and Overton (1979)** described an important advance in capture–recapture modeling by developing a method that accommodated variation in detectability among individual animals. Such detection heterogeneity causes systematic negative bias in estimates of abundance (Kéry & Royle, 2015); thus, this method bolsters abundance estimates in light of among-individual variation in "catchability," for example, due to idiosyncrasies among animals.

Since populations are collections of individuals, it is conceptually desirable to link models for individuals and their movements to population parameters. **Gardner et al. (2022)** accomplished this with an integrated animal movement and spatial capture–recapture model. The successors to capture–recapture, spatial capture–recapture models (Borchers & Efford, 2008; Royle & Young, 2008), estimate individual animals' activity centers from spatial patterns of repeated detections of individuals across traps and thus provide a powerful way to estimate population density and other demographic parameters. However, spatial capture–recapture treats animal movement and space use in a simplistic (and often unrealistic) fashion (Theng et al., 2022). Gardner et al. (2022) rectified this by injecting spatial capture–recapture with animal movement models (e.g., correlated random walks); their simulations demonstrated that animal movement parameters could be estimated with spatial capture–recapture data alone but that auxiliary telemetry data improved estimation of these parameters. The approach of Gardner et al. (2022) is a framework for linking animal movement to population dynamics and offers opportunities for further development (e.g., application to open populations, alternative movement models).

Not all populations can be directly captured or otherwise marked, so **MacKenzie et al. (2002)** made a valuable contribution by developing the occupancy model for "unmarked" animal populations. Borrowing ideas from the realm of capture–recapture, occupancy models recognize that the detection of individuals is imperfect, so such models rely on the detection histories of species across multiple sites (rather than detection histories of individuals within a site) to estimate detection probability and thereby provide more accurate estimates of species occurrence patterns and distributions. The original occupancy model has been adapted in a myriad of ways. Among papers published in *Ecology* alone, extensions include multiseason occupancy models to estimate site colonization and extinction rates (MacKenzie et al., 2003), false-positive models to account for species misidentifications (Chambert et al., 2015; Clare et al., 2021; Royle & Link, 2006), and models to estimate abundance from detection–nondetection data (Royle & Nichols, 2003).

Understanding population trends through time has long been a primary focus of population modeling efforts in mathematical ecology. **Clark and Bjørnstad (2004)** gave a perspective from statistical ecology by modeling population time series with state-space models. Like the state-space modeling approach for individual movement described by Jonsen et al. (2005), these models distinguish between an ecological process (population size) and an observation process (how those populations are surveyed). Moreover, state-space modeling of populations directly accommodates the temporal dependence between population sizes in adjacent time steps. While modeling tools for population time series are valuable in and of themselves, these state-space models are a key component of integrated population models, which have become important tools for estimating population sizes and demographic parameters (Saunders et al., 2019; Schaub & Kéry, 2021).

Finally, a common goal of population modeling is to estimate relationships between species' occurrence data and environmental patterns (e.g., climate, land cover, terrain); these correlations allow for projections of species range maps. Many seemingly disparate methods (e.g., occupancy models, random forests, maximum entropy) fall within this broad category of species distribution modeling. **Araújo and Peterson (2012)** provided a perspective on the assumptions and approaches of species distribution modeling, focusing on broad-scale efforts to map species distributions based on associations with climate variables. For example, they discuss the equilibrium assumption of species distribution models, which implies that species can be found in all locations with habitable climate conditions. This assumption simplifies the reality that species' distributions are also shaped by biotic interactions and dispersal, meaning that climate-only models may overestimate the distributions of species. Beyond assumptions, the authors discuss applications of species distribution modeling, including identification of suitable sites for translocations and assessing impacts of climate change on species' distributions.

THEME 3: MODELS FOR COMMUNITIES (EIGHT PAPERS)

Communities are ecology's core. Many influential papers describing the statistical analysis of communities have been published in *Ecology*. These fall into three broad categories: models focused on quantifying biodiversity, models to relate species composition to environmental patterns, and models to estimate the niche overlap of species.

Biodiversity's fundamental measure is species richness. However, species richness is difficult to observe, and thus many methods to estimate the number of species present in a given area have been suggested. Heck et al. (1975) offered a development in rarefaction, a method developed in the late 1960s to estimate species richness by fitting a curve to the number of detected species plotted against sample size. **Heck et al. (1975)** provided formulas to calculate the expected number of species as well as the variance in species richness and demonstrated how optimal sampling effort can be allocated based on applying the formulas to an initial exhaustive survey. **Chao and Jost (2012)** provided an important advance in estimating species richness based on sample coverage rather than sample size. Coverage—or "completeness"—is the proportion of the number of individuals in a community that pertain to species detected in the sample and, counterintuitively, can be estimated from data. This method is particularly relevant for applications that compare communities of differing diversity (e.g., temperate versus tropical communities). Approaches based on sample size give biased estimates of species richness in such cases because a given sized sample may fully characterize the low-diversity community but fail to comprehensively represent the high-diversity community.

Species richness is just one of many biodiversity variables that can be used to characterize communities. Functional diversity, or the diversity of trait combinations in a community, can be useful for understanding many ecological phenomena. Villéger et al. (2008) proposed three metrics to quantify functional diversity: functional richness, evenness, and divergence. Shortly after, **Laliberté and Legendre (2010)** proposed a new functional diversity metric (functional dispersion) to add to the three metrics proposed by Villéger et al. (2008). The authors also developed a way of accommodating qualitative traits and missing trait values when calculating functional diversity metrics. Finally, the paper described the FD R package, which provided a user-friendly means of calculating functional diversity metrics and vastly expanded the method's scope of application.

Community data collected over multiple locations represent high-dimensional data that may be difficult to interpret. Such data contain information on many species, and there exist many possible environmental variables (e.g., climate, land cover, pH) to relate to patterns of species composition. Ordination methods—which summarize the similarity and dissimilarity of multivariate ecological data—are as old as *Ecology* but saw many exciting developments in a community ecology context starting in the 1950s (ter Braak, 1995). **ter Braak (1986)** introduced canonical correspondence analysis, a massively influential ordination technique in community ecology. Briefly, this method combines an existing ordination method (correspondence analysis) with regression: An ordination of community abundance data is constrained by (related to) environmental variables, allowing associations between community composition and the environment to be robustly visualized and interpreted (Palmer, 1993). **Anderson and Willis (2003)** described another ordination advance that they called canonical analysis of principal components. Like canonical correspondence analysis, canonical analysis of principal components is a constrained ordination method—which use a priori hypotheses to structure the output of ordination—in contrast to unconstrained ordinations (e.g., principal component analysis or nonmetric multidimensional scaling), which do not account for grouping patterns within data when reducing dimensionality. The key innovation provided by Anderson and Willis (2003) is that their method can use any distance or dissimilarity measure (e.g., Euclidean distance, Bray–Curtis dissimilarity), providing greater flexibility and expanding interpretation opportunities, since different dissimilarity or distance metrics emphasize different aspects of multivariate data.

Additionally, community analyses may have the goal of understanding how species interact by measuring their co-occurrence or niche overlap. **Dice (1945)** described a method to quantify how likely species are to co-occur. In addition to describing an index that quantifies co-occurrence, the author described how to evaluate the co-occurrence of two species with a chi-squared test and discussed aspects of experimental design that affect the validity and interpretation of the index. Beyond its original use, the index (known as the Sørensen–Dice coefficient) is widely used to quantify the similarity between two samples or communities (e.g., beta diversity), and for this reason, Dice (1945) has the distinction of being *Ecology*'s most highly cited paper of all time. Similar in objectives, **Pielou (1972)** described a method to quantify niche width and niche overlap based on occurrence data of multiple species across multiple habitats. The motivating data set was aphid species occupying different goldenrod (*Solidago*) species (which represented the "habitats"), but the method is generalizable to any instance in which occurrence data for multiple species are collected across multiple replicates of multiple habitats. Finally, similar in philosophy to the previous two papers, **Swanson et al. (2015)** described a modern method of niche overlap estimation based on stable isotope ratios. This method extended existing isotopic methods beyond two dimensions (i.e., more than two isotopes), more fully approximating the concept of the niche as an "*n*-dimensional hypervolume" and calculated the probability (with uncertainty) that an individual of one species is found within the niche of another species.

THEME 4: MODELS FOR ECOSYSTEMS (TWO PAPERS)

Ecosystem ecology presents challenges for statistical analysis: Many factors—biotic and abiotic—affect processes of interest, often at multiple spatiotemporal scales and over spatial extents that are difficult (or impossible) to manipulate experimentally (but see Carpenter et al., 2001; Likens, 1985). Thus, ecosystem ecologists often rely on process-based or simulation models, defined as models that describe interactions between at least two ecosystem components (Geary et al., 2020; Krinner et al., 2005). While such models often incorporate empirical data to fit or constrain parameters, mathematical models and simulated ecological processes (e.g., predator–prey interactions via Lotka–Volterra equations) constitute the heart of these ecosystem models (Geary et al., 2020). Ecosystem ecologists have also developed and refined statistical models that are suited for the types of observational data collected in ecosystem settings.

Ecosystems are disproportionately impacted by rare—but extreme—disturbances such as floods, megafires, and heatwaves as compared to disturbances of lesser magnitude (Gaines & Denny, 1993). These rare events are challenging to model with traditional statistical machinery, which treats such events as more unusual than they are, hampering inference and prediction. **Katz et al. (2005)** described a solution: application of extreme value modeling in ecosystem ecology. In addition to parameters for the location (akin to the mean of the normal distribution) and scale (variation) of a distribution, the generalized extreme value distribution has a shape parameter that controls the tail of the distribution and accommodates distributions with "heavy" tails (implying that extreme values are less unexpected than a traditional distribution with a "light" tail). The authors modeled a 500-year time series of sediment yield from the high Arctic and showed that an extreme value distribution better fit the data as compared to traditional methods.

Understanding nutrient flows is a major goal in ecosystem ecology, and researchers commonly measure and analyze nutrient ratios such as nitrogen–phosphorus ratios in lakes. **Isles (2020)** gave practical advice on analyzing and interpreting such ratios. Untransformed ratios have statistical properties that produce pitfalls when analyzed—even with basic summary metrics such as averages. Such pitfalls are prevalent in the literature: Isles (2020) sampled 100 ecological stoichiometry papers and found analysis errors in 93%. The author demonstrated that simply log-transforming ratio data prior to analysis prevents these pitfalls and ensures more robust ecological interpretation of results.

THEME 5: MODELS FOR SPACE (FOUR PAPERS)

Ecology is inherently spatial, creating challenges as well as opportunities for analyses. Issues such as scale and spatial autocorrelation can challenge traditional analyses, but on the other hand, scaling of ecological phenomena is an important avenue of research, and inclusion of autocorrelation can improve predictive mapping and prevent overly precise estimates of uncertainty. Throughout its history, *Ecology* has published many influential papers focused on spatial modeling.

The locations of organisms in space are perhaps one of the most fundamental forms of biodiversity data; appropriate analyses of such data can address many ecological questions. **Clark and Evans (1954)** presented a method to analyze the spacing of individuals in a population. Their method was based on distance-to-nearest-neighbor data to determine whether individuals showed an aggregated, random, or uniform distribution and included a significance test accompanying these classifications. Because their method used distance-to-nearest-neighbor data, it was not sensitive to the effects of quadrat size, which had plagued earlier analyses. Clark and Evans (1954) applied their method to several plant data sets and reported aggregated distributions of prairie forbs and a more uniform than expected distribution of forest trees.

Spatial autocorrelation is prevalent in ecological data. **Legendre (1993)** gave an in-depth overview of spatial autocorrelation and how to accommodate it in ecological analyses. The paper showed that failing to account for spatial autocorrelation can lead to overestimating the influence of explanatory variables. The author demonstrated how autocorrelation could be described (e.g., variograms) and discussed appropriate statistical tests for autocorrelated data. Finally, the paper described model-based ways of accommodating spatial structure, focusing on two approaches: first, modeling spatial autocorrelation by including polynomials of the geographic coordinates of sampled locations, and second, via a geographic distance matrix among the sampled locations.

Expanding on the issue of autocorrelation, **Hefley et al. (2017)** provided a detailed explanation of accounting for autocorrelation (either spatial or temporal) in ecological models via basis functions, which are functions that transform a covariate. The polynomial approach described by Legendre (1993) is an example of a basis function; in the quadratic regression $\boldsymbol{y} = \beta_0\mathbf{x}^0 + \beta_1\mathbf{x}^1 + \beta_2\mathbf{x}^2$, the predictor variable $\mathbf{x}$ (a vector) is raised to the 0 power (the coefficient β_0 being the intercept), and then the first power (β_1 being the linear effect), and then the second power (β_2 being the quadratic effect). However, many other possible basis functions can be used (e.g., transformations based on the difference between each observation of the predictor variable and specified "knots," or anchor points along the x-axis), which means that many seemingly different spatial methods share the same basis function framework. Via empirical examples (quail population trends, spatial distribution of a forb), Hefley et al. (2017) showed how various basis-function approaches could provide similar results and improved inference relative to models that do not account for autocorrelated data.

Finally, conservation planning often has a spatial focus as connectivity between habitat patches is important for promoting persistence of species and associated ecosystem functions. Network theory provides important tools for quantifying and visualizing landscape connectivity. **Rayfield et al. (2011)** reviewed network methods for quantifying landscape connectivity and developed a framework for classifying network measures based on measures' connectivity property (e.g., area of connected habitats, redundancy of routes between habitats) and structural level (e.g., network-level vs. component-level). This framework is helpful for finding similarities among the many existing network measures and identifying categories with limited analysis options.

THEME 6: MODEL VALIDATION AND SELECTION (FOUR PAPERS)

Fitting a statistical model is just one step in an analysis. Model validation—or evaluating whether the model makes reasonable predictions—is recommended for certain types of analysis. For cases when multiple competing models are evaluated, some method of model selection is important for interpreting results. Model selection and validation are active areas of research in the statistical literature, and *Ecology* has published many papers on the topic.

The use of null models for model validation was pioneered by Connor and Simberloff (1979), who argued that species co-occurrence patterns previously hypothesized to be the result of competition (Diamond, 1975) could be due to random chance, as demonstrated by a null model of random species distributions. The debate that followed profoundly influenced how ecologists approach model design and hypothesis testing. Advancing this legacy of null modeling, **Gotelli (2000)** evaluated nine null model algorithms to determine whether observed species co-occurrence patterns (across sites or islands) statistically differed from randomly generated co-occurrence matrices. The author identified three of these algorithms as robust to falsely identifying co-occurrence patterns. These three algorithms preserved each species' occurrence frequency (i.e., the proportion of sites/islands in which it was recorded) when generating random co-occurrence matrices. The author noted that no single approach is a "silver bullet" and encouraged practitioners to compare their empirical data to multiple null modeling approaches.

Ecologists often have many possible explanatory variables to use in analyses; therefore, some method of selecting which variables to use is often needed. **Blanchet et al. (2008)** introduced a method of forward selection of explanatory variables, an algorithm to determine which predictors are meaningfully associated with a response variable. Their approach was to first evaluate a global model with all predictor variables. They recommended proceeding with forward selection only if the global model is significant, that is, if it fits the data better than a model without predictor variables. If the global model is indeed significant, forward selection (i.e., building up models with more variables) is to be performed until an added variable brings either the model's significance value or adjusted R^2 to a predetermined threshold. Blanchet et al.'s approach represented a considerable improvement over traditional forward selection, which is plagued with false positives (uninformative variables being considered important) and overestimation of the amount of variance explained. Forward selection is one of several so-called stepwise selection techniques, which have received criticism over the years (Whittingham et al., 2006). Among the alternative model-selection approaches, information-theoretic approaches such as AIC are perhaps the most popular (Burnham & Anderson, 2002).

When evaluating multiple competing models, averaging predictions across models is an attractive means of simplifying interpretation (Dormann et al., 2018). **Cade (2015)** critiqued the common practice of averaging regression coefficients across models, weighted by each model's AIC weight. While averaging predictions from multiple models is acceptable, Cade (2015) showed that model averaging of regression coefficients was generally not valid. Because predictor variables from observational studies are typically collinear, the scales of their coefficient estimates vary among models with different sets of predictors. This makes averaging coefficients across models a dangerous practice, particularly since model averaging might lend a false sense of security to practitioners that the model adequately accommodates uncertainty in covariate effects when in fact collinearity prevents reasonable interpretations. With the rigorous denouncement of model averaging ("The simple averaging of regression coefficients … should be discontinued immediately") and hundreds of citations in less than a decade, Cade (2015) has had a profound influence on the practice of multimodel inference in ecology.

As indicated by the previously mentioned papers, model selection and validation is not always straightforward. **Tredennick et al. (2021)** made the fundamental point that the proper model selection approach depends upon the purpose of the analysis. The authors differentiated exploration, inference, and prediction as the three primary purposes of a statistical analysis. Each effort comes with its own potential pitfalls and key statistical tools. For example, the authors noted that prediction-focused analyses might fail to validate predictive accuracy and recommended AIC and cross-validation as key statistical tools for such efforts. Careful reading of Tredennick et al. (2021) would benefit any researcher in the design stages of an ecological study.

THEME 7: TOOLS AND BEST PRACTICES (SEVEN PAPERS)

Many of the most influential statistical ecology papers published in *Ecology* provide syntheses about general statistical tools or best practices that transcend subdisciplines. Such papers are invaluable resources for the ecological community. Quantitative training for ecologists is often limited or outdated (Barraquand et al., 2014; Ellison & Dennis, 2010), and so practicing ecologists may not be aware of improper statistical habits or of new statistical tools that are appropriate for their data.

Real ecological data often do not neatly conform to the cookbook statistical analyses taught in biostatistics courses. **Warton and Hui (2011)** provided a guide for navigating one such scenario: proportional data. Historically, the arcsine transformation, followed by linear modeling, was commonly used for proportional data, but Warton and Hui (2011) argued that the transformation muddles interpretability and produces biologically impossible predictions (e.g., negative proportions). For cases where proportions are of the form "*x* out of *n*" (e.g., 12 out of 50 animals were infected), logistic regression is an accessible and robust alternative. For cases where proportions are not binomial (e.g., "75% of the leaf was damaged"), other forms of generalized linear (optionally mixed) models such as beta regression are appropriate. The article by Warton and Hui (2011) exemplifies a paper that identifies a problematic statistical practice and outlines accessible and appropriate alternatives.

Meta-analysis is a powerful way of synthesizing across studies but comes with statistical challenges. **Hedges et al. (1999)** described an effective meta-analysis method for ecological experiments: the response ratio, which is the ratio of the mean outcome for an experimental group to that of the control. The response ratio is especially appealing because it quantifies the proportionate change of experimental manipulations, allowing results to be compared across studies in which the response magnitudes vary dramatically. As described in our discussion of Isles (2020) in *Theme 4: Models for ecosystems*, ratios have statistical properties that produce analytic pitfalls. Recognizing this, Hedges et al. (1999) anchored their method on the analysis of the log of the response ratio. Their approach represented an important advance because it quantified uncertainty and provided confidence intervals for the summary of effects across multiple experiments.

In contrast to meta-analysis, which seeks to synthesize results that were produced across separate studies, data integration synthesizes across multiple data sets within individual models (Zipkin et al., 2019). **Pacifici et al. (2017)** described a framework for integrating multiple types of data in species distribution models. This practice of data integration—or bringing disparate data sets into a cohesive model—has grown in recent years and has several attractive attributes, including expanding the spatiotemporal scope of inference and increasing the precision of inference by nature of the larger volume of data applied to the problem. Pacifici et al. (2017) provided a helpful overview of integrated species distribution modeling and described several alternative integrated models to analyze data from the North American Breeding Bird Survey and eBird, two public-science programs with structured and semistructured data collection protocols, respectively.

Machine-learning methods have become important in ecology as the discipline has become increasingly data-intensive in recent decades. **De'ath and Fabricius (2000)** reviewed classification and regression trees, powerful and flexible methods to analyze complex ecological data. Some of the advantages of these methods include their ability to handle many different types of response variables (e.g., numeric and categorical), as well as missing values in response and explanatory variables. The basic premise of trees is to split the response data into similar groups based on explanatory variables, while keeping the tree reasonably small (i.e., not splitting the data into too many groups). **Cutler et al. (2007)** provided an ecologist's introduction to random forests, an advance over earlier machine-learning tools like classification trees. Random forests create many classification trees and subsequently combine predictions from the trees, resulting in high classification accuracy. In addition, random forests provide a measure of variable importance, allowing users to interpret which variables are most influential in differentiating groups. Random forests and other machine-learning methods have the advantage of being able to accommodate many predictor variables regardless of collinearity or the presence of complex interactions (thorns in the side of traditional regression-based approaches); they have the limitation of being primarily predictive tools rather than providing traditional statistical inference (e.g., determining whether a predictor variable has a statistically significant relationship with a response).

Even with the rise of machine learning, the varied goals of statistical analysis (Tredennick et al., 2021) imply that traditional statistical methods will continue to occupy a dominant role in the analysis of ecological data. **Graham (2003)** described solutions to a ubiquitous challenge for traditional regression: collinearity. The author first demonstrated via simulations that even modest collinearity produces significant issues with analysis, including inaccurate parameterization of models and failure to identify important explanatory variables. As mentioned in our discussion of Cade (2015) in *Theme 6: Model validation and selection*, collinearity also hamstrings efforts to average or otherwise interpret regression coefficients across candidate models. The paper then outlined approaches (beyond the default of dropping collinear variables) for facing collinearity, including residual and sequential regression, principal component analysis, and structural equation modeling. Given collinearity's prevalence in ecological data and the importance of multiple regression in ecological analyses, ecologists should be aware of collinearity and that no perfect solution to it exists. Finally, **Murtaugh (2007)** advocated for simplicity and interpretability in ecological analyses when possible. The author provided several empirical examples of relatively complex analyses (nested analysis of variance, linear mixed-effect modeling) that produced results similar to those of much simpler alternatives and argued that the more complex analyses were thus not strictly necessary and indeed might hamper interpretability. This is an important message for ecologists to hear, particularly given the runaway growth of computing power and the constellation of statistical packages available with which to analyze data.

DATA AND COMPUTING SHAPE STATISTICAL ECOLOGY

By definition, statistical ecology is a data-forward approach to understanding ecological issues. Thus, it is worth reflecting on how data have shaped the history of statistical ecology (Ellison & Dennis, 2010). In its early years, ecology was a largely descriptive science; many early papers in *Ecology* gave detailed descriptions and species lists of vegetation and animal communities (e.g., Adams et al., 1920; Rigg, 1922). This period of the early 20th century also saw the appearance and widespread uptake of standard frequentist statistical methods such as the *t*-test and analysis of variance, or ANOVA. These methods became analytic workhorses for ecological studies, as reflected in features of study design prevalent in subsequent decades (and to this day) such as quadrats, randomization, and experimental manipulations. However, such approaches to data collection cannot inform all of ecology's questions, particularly those focused on broader spatiotemporal scales (e.g., ecosystem ecology, biogeography). A data revolution in ecology was launched in the latter decades of the 20th century with the expansion of remote sensing data (Turner et al., 2003), animal tracking technology (Kays et al., 2015), automated camera and acoustic monitoring (Burton et al., 2015; Shonfield & Bayne, 2017), and public-science data (Dickinson et al., 2010).

Together with this data revolution, growth in computing power has played a key role in the expansion of statistical ecology in recent decades. After all, more data are of little help in addressing ecological questions if computational infrastructure is not available to store, manage, and analyze them. Improvements in computing technology in the 1970s and 1980s are reflected in the increased prevalence of statistical ecology papers in this period (Figure 1). By the turn of the 21st century, statistical tools that either did not exist or were infeasible to implement just a generation or two previously were being routinely used by practicing ecologists. As a telling example, logistic regression—today a ubiquitous method in ecology—was referred to as "nontraditional" by Trexler and Travis (1993). How will this trajectory continue?

THE FUTURE OF STATISTICAL ECOLOGY

We envision that the volume and resolution of ecological data will continue to grow and that improvements in metadata structures will allow data streams to be increasingly open and accessible. GPS tags will become smaller and cheaper and have longer lifespans; eDNA monitoring will expand and be collated into centralized biodiversity databases (Berry et al., 2021); and public-science projects like eBird (Sullivan et al., 2009) will increase in volume, accumulating vast spatiotemporal records. Given this future, we anticipate that machine-learning and artificial intelligence methods will become integral parts of the data collection process, for example, through automated classification of photo and audio data (Kahl et al., 2021; Tabak et al., 2019). We see a particular need for bridges between data sets produced by automated algorithms and the statistical models with which these data sets are analyzed. Rhinehart et al. (2022) exemplify this future by describing an occupancy modeling approach that uses continuous-score classifications produced by machine-learning algorithms from acoustic data sets, rather than the binary detection/nondetection data collected by human observers in traditional occupancy studies. Rather than replacing traditional statistical tools, we envision that machine learning will complement and expand the existing canon of statistical ecology tools. Finally, we hope that the community of statistical ecologists will become more diverse; historically, participation in ecology (such as Ecological Society of America membership) has not reflected gender and racial composition of the broader population (Beck et al., 2014), and we support practices that promote recruitment and retention of historically excluded groups in statistical ecology (Duc Bo Massey et al., 2021).

Whether a data set is small or large, whether collected automatically or by pencil on datasheets, we see an ongoing need for synthesis and practical guidance on many of the statistical choices facing ecologists. More data are not always better; some ecological questions are best addressed with small, hard-won data sets collected by human hands and eyes (Todman et al., 2023). As advised by Murtaugh (2007), simplicity is a virtue in statistical analysis. The breadth of statistical ecology research published by *Ecology* over its 100-year lifetime has pushed forward the scope and depth of ecological questions that can be asked and answered, exemplifying the very history of the field itself. There is no doubt that *Ecology* will continue to produce innovative papers that will help new generations in the practice of statistical ecology.

ACKNOWLEDGMENTS

This work was supported by National Science Foundation (NSF) DBI-1954406, NSF DEB-2213565, and NSF DEB-2016347. NG was supported by NSF Postdoctoral Research Fellowship in Biology (2208894). We thank Brian Cade and a second anonymous reviewer for comments that improved the manuscript.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data and code (Gilbert, 2024) are available in Zenodo at https://doi.org/10.5281/zenodo.10611823.

ORCID

Neil A. Gilbert https://orcid.org/0000-0003-0949-5612
Olivia M. Smith https://orcid.org/0000-0002-9404-0243
Jeffrey W. Doser https://orcid.org/0000-0002-8950-9895
Elise F. Zipkin https://orcid.org/0000-0003-4155-6139

REFERENCES

Adams, C. C., G. P. Burns, T. L. Hankinson, B. Moore, and N. Taylor. 1920. “Plants and Animals of Mount Marcy, New York, Part III.” *Ecology* 1: 274–287.

Anderson, M. J., and T. J. Willis. 2003. “Canonical Analysis of Principal Coordinates: A Useful Method of Constrained Ordination for Ecology.” *Ecology* 84: 511–525.

Araújo, M. B., and A. T. Peterson. 2012. “Uses and Misuses of Bioclimatic Envelope Modeling.” *Ecology* 93: 1527–39.

Barraquand, F., T. H. G. Ezard, P. S. Jørgensen, N. Zimmerman, S. Chamberlain, R. Salguero-Gómez, T. J. Curran, and T. Poisot. 2014. “Lack of Quantitative Training among Early-Career Ecologists: A Survey of the Problem and Potential Solutions.” *PeerJ* 2: e285.

Beck, C., K. Boersma, C. S. Tysor, and G. Middendorf. 2014. “Diversity at 100: Women and Underrepresented Minorities in the ESA.” *Frontiers in Ecology and the Environment* 12: 434–36.

Benson, E. 2010. *Wired Wilderness: Technologies of Tracking and the Making of Modern Wildlife*, 1st ed. Baltimore: John Hopkins University Press.

Berry, O., S. Jarman, A. Bissett, M. Hope, C. Paeper, C. Bessey, M. K. Schwartz, J. Hale, and M. Bunce. 2021. “Making Environmental DNA (eDNA) Biodiversity Records Globally Accessible.” *Environmental DNA* 3: 699–705.

Blanchet, F. G., P. Legendre, and D. Borcard. 2008. “Forward Selection of Explanatory Variables.” *Ecology* 89: 2623–32.

Borchers, D., and M. Efford. 2008. “Spatially Explicit Maximum Likelihood Methods for Capture-Recapture Studies.” *Biometrics* 64: 377–385.

Burnham, K. P., and D. R. Anderson, eds. 2002. *Model Selection and Multimodel Inference*. New York: Springer.

Burnham, K. P., and W. S. Overton. 1979. “Robust Estimation of Population Size when Capture Probabilities Vary among Animals.” *Ecology* 60: 927–936.

Burton, A. C., E. Neilson, D. Moreira, A. Ladle, R. Steenweg, J. T. Fisher, E. Bayne, and S. Boutin. 2015. “Wildlife Camera Trapping: A Review and Recommendations for Linking Surveys to Ecological Processes.” *Journal of Applied Ecology* 52: 675–685.

Cade, B. S. 2015. “Model Averaging and Muddled Multimodel Inferences.” *Ecology* 96: 2370–82.

Carpenter, S. R., J. J. Cole, J. R. Hodgson, J. F. Kitchell, M. L. Pace, D. Bade, K. L. Cottingham, T. E. Essington, J. N. Houser, and D. E. Schindler. 2001. “Trophic Cascades, Nutrients, and Lake Productivity: Whole-Lake Experiments.” *Ecological Monographs* 71: 163–186.

Chambert, T., D. A. W. Miller, and J. D. Nichols. 2015. “Modeling False Positive Detections in Species Occurrence Data under Different Study Designs.” *Ecology* 96: 332–39.

Chao, A., and L. Jost. 2012. “Coverage-Based Rarefaction and Extrapolation: Standardizing Samples by Completeness Rather than Size.” *Ecology* 93: 2533–47.

Clare, J. D. J., P. A. Townsend, and B. Zuckerberg. 2021. “Generalized Model-Based Solutions to False-Positive Error in Species Detection/Nondetection Data.” *Ecology* 102: e03241.

Clark, J. S., and O. N. Bjørnstad. 2004. “Population Time Series: Process Variability, Observation Errors, Missing Values, Lags, and Hidden States.” *Ecology* 85: 3140–50.

Clark, P. J., and F. C. Evans. 1954. “Distance to Nearest Neighbor as a Measure of Spatial Relationships in Populations.” *Ecology* 35: 445–453.

Connor, E. F., and D. Simberloff. 1979. “The Assembly of Species Communities: Chance or Competition?” *Ecology* 60: 1132–40.

Converse, S. J., B. T. McClintock, and P. B. Conn. 2022. “Special Feature: Linking Capture–Recapture and Movement.” *Ecology* 103: e3770.

Cottam, G., J. T. Curtis, and B. W. Hale. 1953. “Some Sampling Characteristics of a Population of Randomly Dispersed Individuals.” *Ecology* 34: 741–757.

Cutler, D. R., T. C. Edwards, Jr., K. H. Beard, A. Cutler, K. T. Hess, J. Gibson, and J. J. Lawler. 2007. “Random Forests for Classification in Ecology.” *Ecology* 88: 2783–92.

De Bach, P., and H. S. Smith. 1941. “Are Population Oscillations Inherent in the Host–Parasite Relation?” *Ecology* 22: 363–69.

De’ath, G., and K. E. Fabricius. 2000. “Classification and Regression Trees: A Powerful Yet Simple Technique for Ecological Data Analysis.” *Ecology* 81: 3178–92.

Diamond, J. M. 1975. “The Island Dilemma: Lessons of Modern Biogeographic Studies for the Design of Natural Reserves.” *Biological Conservation* 7: 129–146.

Dice, L. R. 1945. “Measures of the Amount of Ecologic Association between Species.” *Ecology* 26: 297–302.

Dickinson, J. L., B. Zuckerberg, and D. N. Bonter. 2010. “Citizen Science as an Ecological Research Tool: Challenges and Benefits.” *Annual Review of Ecology, Evolution, and Systematics* 41: 149–172.

DiRenzo, G. V., E. Hanks, and D. A. W. Miller. 2023. “A Practical Guide to Understanding and Validating Complex Models Using Data Simulations.” *Methods in Ecology and Evolution* 14: 203–217.

Dormann, C. F., J. M. Calabrese, G. Guillera-Arroita, E. Matechou, V. Bahn, K. Bartoń, C. M. Beale, et al. 2018. “Model Averaging in Ecology: A Review of Bayesian, Information-Theoretic, and Tactical Approaches for Predictive Inference.” *Ecological Monographs* 88: 485–504.

Duc Bo Massey, M., S. Arif, C. Albury, and V. A. Cluney. 2021. “Ecology and Evolutionary Biology Must Elevate BIPOC Scholars.” *Ecology Letters* 24: 913–19.

Ellison, A. M., and B. Dennis. 2010. “Paths to Statistical Fluency for Ecologists.” *Frontiers in Ecology and the Environment* 8: 362–370.

Fleming, C. H., W. F. Fagan, T. Mueller, K. A. Olson, P. Leimgruber, and J. M. Calabrese. 2015. “Rigorous Home Range Estimation with Movement Data: A New Autocorrelated Kernel Density Estimator.” *Ecology* 96: 1182–88.

Gaines, S. D., and M. W. Denny. 1993. “The Largest, Smallest, Highest, Lowest, Longest, and Shortest: Extremes in Ecology.” *Ecology* 74: 1677–92.

Gardner, B., B. T. McClintock, S. J. Converse, and N. J. Hostetter. 2022. "Integrated Animal Movement and Spatial Capture–Recapture Models: Simulation, Implementation, and Inference." *Ecology* 103: e3771.

Geary, W. L., M. Bode, T. S. Doherty, E. A. Fulton, D. G. Nimmo, A. I. T. Tulloch, V. J. D. Tulloch, and E. G. Ritchie. 2020. "A Guide to Ecosystem Models and their Environmental Applications." *Nature Ecology & Evolution* 4: 1459–71.

Gilbert, G. S. 2018. "Can 100 Must-Read Papers Also Reflect 'Who' Is Ecology?" *Nature Ecology & Evolution* 2: 203.

Gilbert, N. 2024. "n-a-gilbert/statistical_ecology: Publish." Zenodo. https://doi.org/10.5281/zenodo.10611823.

Gleason, H. A. 1922. "On the Relation between Species and Area." *Ecology* 3: 158–162.

Gotelli, N. J. 2000. "Null Model Analysis of Species Co-Occurrence Patterns." *Ecology* 81: 2606–21.

Graham, M. H. 2003. "Confronting Multicollinearity in Ecological Multiple Regression." *Ecology* 84: 2809–15.

Heck, K. L., G. van Belle, and D. Simberloff. 1975. "Explicit Calculation of the Rarefaction Diversity Measurement and the Determination of Sufficient Sample Size." *Ecology* 56: 1459–61.

Hedges, L. V., J. Gurevitch, and P. S. Curtis. 1999. "The Meta-Analysis of Response Ratios in Experimental Ecology." *Ecology* 80: 1150–56.

Hefley, T. J., K. M. Broms, B. M. Brost, F. E. Buderman, S. L. Kay, H. R. Scharf, J. R. Tipton, P. J. Williams, and M. B. Hooten. 2017. "The Basis Function Approach for Modeling Autocorrelation in Ecological Data." *Ecology* 98: 632–646.

Isles, P. D. F. 2020. "The Misuse of Ratios in Ecological Stoichiometry." *Ecology* 101: e03153.

Johnson, D. H. 1980. "The Comparison of Usage and Availability Measurements for Evaluating Resource Preference." *Ecology* 61: 65–71.

Jonsen, I. D., J. M. Flemming, and R. A. Myers. 2005. "Robust State–Space Modeling of Animal Movement Data." *Ecology* 86: 2874–80.

Kahl, S., C. M. Wood, M. Eibl, and H. Klinck. 2021. "BirdNET: A Deep Learning Solution for Avian Diversity Monitoring." *Ecological Informatics* 61: 101236.

Katz, R. W., G. S. Brush, and M. B. Parlange. 2005. "Statistics of Extremes: Modeling Ecological Disturbances." *Ecology* 86: 1124–34.

Kays, R., M. C. Crofoot, W. Jetz, and M. Wikelski. 2015. "Terrestrial Animal Tracking as an Eye on Life and Planet." *Science* 348: aaa2478.

Kéry, M., and J. A. Royle. 2015. *Applied Hierarchical Modeling in Ecology*, 1st ed. Amsterdam: Elsevier.

Knight, S. M., G. M. Pitman, D. T. T. Flockhart, and D. R. Norris. 2019. "Radio-Tracking Reveals how Wind and Temperature Influence the Pace of Daytime Insect Migration." *Biology Letters* 15: 20190327.

Krinner, G., N. Viovy, N. de Noblet-Ducoudré, J. Ogée, J. Polcher, P. Friedlingstein, P. Ciais, S. Sitch, and I. C. Prentice. 2005. "A Dynamic Global Vegetation Model for Studies of the Coupled Atmosphere-Biosphere System." *Global Biogeochemical Cycles* 19: GB1015.

Laliberté, E., and P. Legendre. 2010. "A Distance-Based Framework for Measuring Functional Diversity from Multiple Traits." *Ecology* 91: 299–305.

Legendre, P. 1993. "Spatial Autocorrelation: Trouble or New Paradigm?" *Ecology* 74: 1659–73.

Likens, G. E. 1985. "An Experimental Approach for the Study of Ecosystems: The Fifth Tansley Lecture." *Journal of Ecology* 73: 381–396.

MacKenzie, D. I., J. D. Nichols, J. E. Hines, M. G. Knutson, and A. B. Franklin. 2003. "Estimating Site Occupancy, Colonization, and Local Extinction when a Species Is Detected Imperfectly." *Ecology* 84: 2200–2207.

MacKenzie, D. I., J. D. Nichols, G. B. Lachman, S. Droege, J. A. Royle, and C. A. Langtimm. 2002. "Estimating Site Occupancy Rates when Detection Probabilities Are Less than One." *Ecology* 83: 2248–55.

Martin, L. J. 2012. "Where Are the Women in Ecology?" *Frontiers in Ecology and the Environment* 10: 177–78.

Murtaugh, P. A. 2007. "Simplicity and Complexity in Ecological Data Analysis." *Ecology* 88: 56–62.

Nathan, R., C. T. Monk, R. Arlinghaus, T. Adam, J. Alós, M. Assaf, H. Baktoft, et al. 2022. "Big-Data Approaches Lead to an Increased Understanding of the Ecology of Animal Movement." *Science* 375: eabg1780.

Northrup, J. M., M. B. Hooten, C. R. Anderson, Jr., and G. Wittemyer. 2013. "Practical Guidance on Characterizing Availability in Resource Selection Functions under a Use–Availability Design." *Ecology* 94: 1456–63.

Pacifici, K., B. J. Reich, D. A. W. Miller, B. Gardner, G. Stauffer, S. Singh, A. McKerrow, and J. A. Collazo. 2017. "Integrating Multiple Data Sources in Species Distribution Modeling: A Framework for Data Fusion." *Ecology* 98: 840–850.

Palmer, M. W. 1993. "Putting Things in Even Better Order: The Advantages of Canonical Correspondence Analysis." *Ecology* 74: 2215–30.

Patterson, T. A., L. Thomas, C. Wilcox, O. Ovaskainen, and J. Matthiopoulos. 2008. "State–Space Models of Individual Animal Movement." *Trends in Ecology & Evolution* 23: 87–94.

Pielou, E. C. 1972. "Niche Width and Niche Overlap: A Method for Measuring Them." *Ecology* 53: 687–692.

Rayfield, B., M.-J. Fortin, and A. Fall. 2011. "Connectivity for Conservation: A Framework to Classify Network Measures." *Ecology* 92: 847–858.

Rhinehart, T. A., D. Turek, and J. Kitzes. 2022. "A Continuous-Score Occupancy Model that Incorporates Uncertain Machine Learning Output from Autonomous Biodiversity Surveys." *Methods in Ecology and Evolution* 13: 1778–89.

Rigg, G. B. 1922. "The Sphagnum Bogs of Mazama Dome." *Ecology* 3: 321–24.

Royle, J. A., and W. A. Link. 2006. "Generalized Site Occupancy Models Allowing for False Positive and False Negative Errors." *Ecology* 87: 835–841.

Royle, J. A., and J. D. Nichols. 2003. "Estimating Abundance from Repeated Presence–Absence Data or Point Counts." *Ecology* 84: 777–790.

Royle, J. A., and K. V. Young. 2008. "A Hierarchical Model for Spatial Capture–Recapture Data." *Ecology* 89: 2281–89.

Saunders, S. P., M. T. Farr, A. D. Wright, C. A. Bahlai, J. W. Ribeiro, S. Rossman, A. L. Sussman, T. W. Arnold, and E. F. Zipkin. 2019. "Disentangling Data Discrepancies with Integrated Population Models." *Ecology* 100: e02714.

Schaub, M., and M. Kéry. 2021. *Integrated Population Models: Theory and Ecological Applications with R and JAGS*. London, UK: Academic Press.

Seber, G. A. F., and M. R. Schofield. 2019. “A Brief History of Capture–Recapture.” In *Capture-Recapture: Parameter Estimation for Open Animal Populations*, edited by G. A. F. Seber and M. R. Schofield, 1–11. Cham: Springer International Publishing.

Shonfield, J., and E. Bayne. 2017. “Autonomous Recording Units in Avian Ecological Research: Current Use and Future Applications.” *Avian Conservation and Ecology* 12: 14.

Sullivan, B. L., C. L. Wood, M. J. Iliff, R. E. Bonney, D. Fink, and S. Kelling. 2009. “eBird: A Citizen-Based Bird Observation Network in the Biological Sciences.” *Biological Conservation* 142: 2282–92.

Swanson, H. K., M. Lysy, M. Power, A. D. Stasko, J. D. Johnson, and J. D. Reist. 2015. “A New Probabilistic Method for Quantifying n-Dimensional Ecological Niches and Niche Overlap.” *Ecology* 96: 318–324.

Tabak, M. A., M. S. Norouzzadeh, D. W. Wolfson, S. J. Sweeney, K. C. Vercauteren, N. P. Snow, J. M. Halseth, et al. 2019. “Machine Learning to Classify Animal Species In Camera Trap Images: Applications in Ecology.” *Methods in Ecology and Evolution* 10: 585–590.

Teplitskiy, M., E. Duede, M. Menietti, and K. R. Lakhani. 2022. “How Status of Research Papers Affects the Way they Are Read and Cited.” *Research Policy* 51: 104484.

ter Braak, C. J. F. 1986. “Canonical Correspondence Analysis: A New Eigenvector Technique for Multivariate Direct Gradient Analysis.” *Ecology* 67: 1167–79.

ter Braak, C. J. F. 1995. “Ordination.” In *Data Analysis in Community and Landscape Ecology*, edited by C. J. F. ter Braak, O. F. R. van Tongeren, and R. H. G. Jongman, 91–173. Cambridge: Cambridge University Press.

Theng, M., C. Milleret, C. Bracis, P. Cassey, and S. Delean. 2022. “Confronting Spatial Capture–Recapture Models with Realistic Animal Movement Simulations.” *Ecology* 103: e3676.

Todman, L. C., A. Bush, and A. S. C. Hood. 2023. “‘Small Data’ for Big Insights in Ecology.” *Trends in Ecology & Evolution* 38: 615–622.

Tredennick, A. T., G. Hooker, S. P. Ellner, and P. B. Adler. 2021. “A Practical Guide to Selecting Models for Exploration, Inference, and Prediction in Ecology.” *Ecology* 102: e03336.

Trexler, J. C., and J. Travis. 1993. “Nontraditional Regression Analyses.” *Ecology* 74: 1629–37.

Turner, W., S. Spector, N. Gardiner, M. Fladeland, E. Sterling, and M. Steininger. 2003. “Remote Sensing for Biodiversity Science and Conservation.” *Trends in Ecology & Evolution* 18: 306–314.

Villéger, S., N. W. H. Mason, and D. Mouillot. 2008. “New Multidimensional Functional Diversity Indices for a Multifaceted Framework in Functional Ecology.” *Ecology* 89: 2290–2301.

Warton, D. I., and F. K. C. Hui. 2011. “The Arcsine Is Asinine: The Analysis of Proportions in Ecology.” *Ecology* 92: 3–10.

Whelan, A. M., and D. S. Schimel. 2019. “Authorship and Gender in ESA Journals.” *The Bulletin of the Ecological Society of America* 100: e01567.

Whittingham, M. J., P. A. Stephens, R. B. Bradbury, and R. P. Freckleton. 2006. “Why Do we Still Use Stepwise Modelling in Ecology and Behaviour?” *Journal of Animal Ecology* 75: 1182–89.

Worton, B. J. 1989. “Kernel Methods for Estimating the Utilization Distribution in Home-Range Studies.” *Ecology* 70: 164–68.

Zhou, S., S. Chai, and R. B. Freeman. 2024. “Gender Homophily: In-Group Citation Preferences and the Gender Disadvantage.” *Research Policy* 53: 104895.

Zipkin, E. F., B. D. Inouye, and S. R. Beissinger. 2019. “Innovations in Data Integration for Modeling Populations.” *Ecology* 100: e02713.

How to cite this article: Gilbert, Neil A., Bruna R. Amaral, Olivia M. Smith, Peter J. Williams, Sydney Ceyzyk, Samuel Ayebare, Kayla L. Davis, Wendy Leuenberger, Jeffrey W. Doser, and Elise F. Zipkin. 2024. “A Century of Statistical *Ecology*.” *Ecology* 105(6): e4283. https://doi.org/10.1002/ecy.4283