Received: 31 July 2023 | Accepted: 14 March 2024

DOI: 10.1111/1365-2664.14670

Journal of Applied Ecology BRITISH ECOLOGICAL SOCIETY

RESEARCH ARTICLE

Leveraging relationships between species abundances to improve predictions and inform conservation

C. Lane Scher[1] | Sarah M. Roberts[2] | Kevin P. Krause[3] | James S. Clark[1,4]

[1]Nicholas School of the Environment, Duke University, Durham, North Carolina, USA

[2]Department of Earth, Marine, and Environmental Science, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA

[3]U.S. Geological Survey, Eastern Ecological Science Center, Kearneysville, West Virginia, USA

[4]Department of Statistical Science, Duke University, Durham, North Carolina, USA

Correspondence
C. Lane Scher
Email: lane.scher@duke.edu

Present address
Kevin P. Krause, Minnesota Department of Natural Resources, St. Paul, Minnesota 55155, USA

Funding information
NASA Headquarters, Grant/Award Number: AIST18-0063; Programme d'Investissement d'Avenir under project FORBIC, Grant/Award Number: 18-MPGA-0004; National Science Foundation, Grant/Award Number: DEB-1754443, DEB-2211764 and DGE-1644868; Belmont Forum, Grant/Award Number: 854976; U.S. Geological Survey's Ecosystems Mission Area

Handling Editor: Debbie Russell

Abstract

1. Many management and conservation contexts can benefit from understanding relationships between species abundances, which can be used to improve predictions of species occurrence and abundance.
2. We present conditional prediction as a tool to capture information about species abundances via residual covariance between species. From a fitted joint species distribution model, this framework produces a species coefficient matrix that contains relationships between species abundances. The species coefficients allow co-observed species to be treated as a second set of predictors supplementing covariates in the model to improve prediction. We use simulations to demonstrate the potential benefits and limitations of conditional prediction across data types and species covariance before applying conditional prediction to two management contexts with real data.
3. Simulations demonstrate that conditional prediction provides the largest benefits to continuous data and when there is residual covariance between many species.
4. In our first application, we show that conditioning on other species improves in-sample and out-of-sample predictions of fish and invertebrate species, including Atlantic cod. In our second application, we show that the species coefficient matrix can be used to identify bird species at risk of nest parasitism by Brown-headed Cowbirds.
5. *Synthesis and applications*. We present guidelines for using conditional prediction, which can help understand relationships between species abundances, improve predictions and inform conservation in a variety of contexts.

KEYWORDS
Atlantic cod, Brown-headed Cowbird, fisheries, joint species distribution model, prediction

1 | INTRODUCTION

Informed management and conservation strategies depend on knowledge of species abundances and the habitat conditions that they respond to (Bean et al., 2014). Multiple species can be affected by many of the same habitat variables, so the observed abundances of one or more species can provide information on the abundances of others. Identifying the relationships between species abundances might be useful to anticipate those at risk from competitors, disease or invasive species. We implement *conditional* *prediction* to integrate observable environmental variables with information available from the abundances of other species to improve predictions of distribution across space. To predict *focal species* of interest conditioned on counts of *incidental species*, we jointly fit the distribution of all species to measurable variation in the environment. Residual covariance between species quantifies co-abundance that is not explained by environmental variables in the model; together, the fitted environmental variables and residual species covariance are the basis for prediction. This framework applies the classic equations for the conditional multivariate normal distributions to ecological modelling to improve predictions of species distributions. Using simulated data, where true species abundances and the variables that explain those abundances are known, as well as data from actual monitoring networks, we offer guidance on how conditional prediction can sharpen our understanding of patterns in species abundance. We show that conditional prediction requires residual covariance between species and the potential benefit from conditioning depends on the noise in the data and the type of data (e.g. presence–absence, discrete counts, continuous abundance). We demonstrate the broad applicability of conditional prediction to management contexts using two examples from distinct ecological communities: fish and invertebrate populations to inform harvesting practices and a nest parasite (Brown-headed Cowbird) and its host species to improve management.

Conditional prediction uses information from co-observed species while accounting for uncertainty in observations. Because species are recorded together as part of the same observations (e.g. fish returns or bird point counts), models must allow for uncertainty in the counts of all species (Tang et al., 2021). The counts for one species cannot be used to directly predict counts for another (e.g. Zhang et al. (2020)); doing so assumes that the counts of incidental species have no observation error, and only the focal species counts are random (Figure 1a). Modelling species as a joint response treats all species as random, with observation error, and accounts for co-dependence between species (Figure 1b). The fitted joint model is the basis for conditional prediction (Figure 1c).

The potential benefit of conditional prediction over traditional prediction (i.e. predictions that are based solely on species–environment relationships) comes from the residual covariance between species. Because the number of variables that can be measured at all sites is limited, these variables often explain small fractions of the variation in species abundance data. Similarly, environmental variables in the model may not explain species interactions. Therefore, the residual covariance matrix captures additional information. Conditional prediction uses residual covariance to add unmeasured information about the environment to predictions (Figure 1c). Whereas previous efforts are limited to species traits (Seyednasrollah & Clark, 2020) or presence or absence of an incidental species (Wilkinson et al., 2020), we use incidental species to directly inform the abundances of others via residual covariance.

It is important to note that although residual covariance can occur between species that directly interact, it does not quantify the strength of competition, predation or mutualism (Blanchet et al., 2020; Clark et al., 2017; Poggiato et al., 2021; Zurell et al., 2018) For example, positive residual covariance does not reveal a mutualism but rather indicates that a pair of species occur together beyond what is explained in the model. Although residual covariance may arise from species interactions, the residual covariance itself does not identify or quantify the relationships between two species, that is, whether they are parasite and host, predator and prey, competitors or have no interaction at all (Clark et al., 2020).

Balancing the potential benefits of conditional prediction is its sensitivity to noise in the data. Traditional prediction only depends on observed counts indirectly: noisy observations are propagated to parameter estimates, which integrate variation over a potentially large number of observations. These parameter estimates are then used for prediction. Prediction error is large if the sample size is

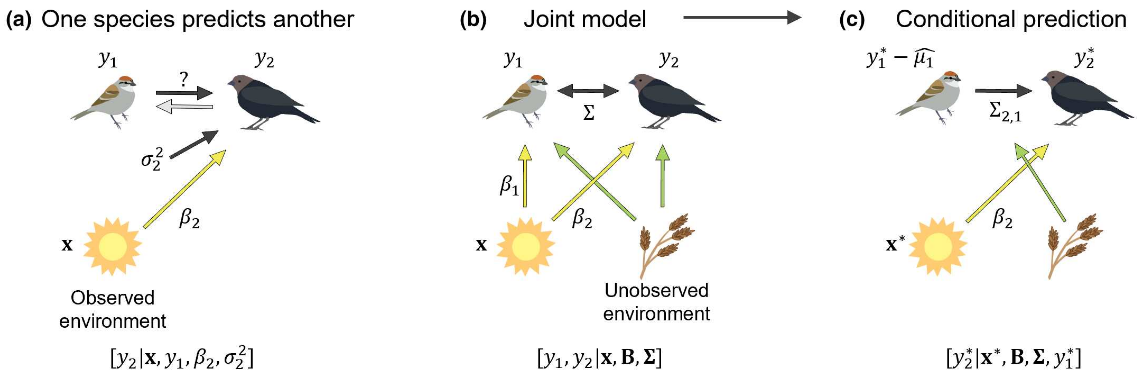

FIGURE 1 (a) Using abundance of species y_1 to predict y_2 incorrectly assumes that the observation error σ_2^2 only affects y_2. Reversing predictor and response species (predicting species y_1 from species y_2) suffers the reverse problem. The model identifies the relationship between the observed environment x and y_2, β_2. (b) A joint attribute model admits the observation errors in both and the unobserved variables contribute to residual covariance $\mathbf{\Sigma}$. The relationships between the observed environment x and both y_1 and y_2, β_1 and β_2 are identified. (c) Using the fitted joint model, conditional prediction for a new environment $\mathbf{x}^*$ exploits information on unobserved variables through the residual covariance $\mathbf{\Sigma}_{2,1}$. Predictive distributions are shown at bottom, where conditional prediction integrates over the posterior distribution for parameters $[\mathbf{B}, \mathbf{\Sigma} \,|\, \mathbf{X}, \mathbf{Y}]$.

small and the residual variance is large (Tang et al., 2023). By contrast, when a prediction is conditioned on counts of other species from a single location (i.e. a single observation), the raw count directly impacts prediction error. Thus, while conditioning adds information from other species, it also introduces noise from observed values in a way that traditional prediction does not. We examine how noise affects the benefits of conditional prediction by simulating species data of different types that contain different amounts of noise (Table 1). In presence–absence data, the lack of abundance information (thousands are no different than one) may contribute to uncertainty and therefore noise. Conversely, in continuous abundance data, such as counts per effort, tree basal area, or population density, abundances are continuous but with discrete zeroes. For discrete counts, uncertainty can depend on the effort associated with each observation, such as plot area, where high effort tends to reduce noise.

In our first case study, we apply conditional prediction to monitoring of marine fish and invertebrate biomass in the Gulf of Maine. The Gulf of Maine is a biologically rich and interconnected ecosystem that supports a variety of fisheries. The region is also warming faster than 96.2% of the world's oceans (Pershing et al., 2021), leading to shifts in the geographic distribution of fish and invertebrate species. These distribution shifts combined with overharvesting can lead to population declines: Atlantic cod (*Gadus morhua*), for example, has been depleted to less than 5% of historic biomass levels (Sguotti et al., 2019). As species composition changes, managers must re-examine historical quotas and fishing grounds. To protect cod populations, fisheries that target functionally similar species, such as the haddock fishery, are required to close if too many cod are caught as bycatch (H.R.2617, 117th Congress, 2022). The current rarity of cod makes it difficult to estimate their distributions and abundances using single-species assessments alone. Therefore, predictions conditioned on functionally similar species (in terms of habitat use and trophic level; e.g. demersal large-bodied consumers), which are likely to co-occur with cod, as well as known prey from gut content studies could help manage this vulnerable stock. In this example, conditioning fish and invertebrate species on each other improves both in-sample and out-of-sample biomass predictions. We highlight the improved predictions of Atlantic cod, whose management under climate change relies heavily on population assessments and distribution maps typically generated from relationships between cod and environmental conditions alone (Pershing et al., 2015).

In a second case study, we use conditional prediction to understand relationships between the abundances of Brown-headed Cowbirds (*Molothrus ater*), which are obligate nest parasites, and their host species. Brown-headed Cowbirds rely on breeding birds of other species to incubate their eggs and raise their young, often at the expense of the host's own reproductive success (Lorenzana & Sealy, 1999). Cowbirds have been reported to parasitize over 200 species (Friedmann, 1963), with varying effects on host populations. Management of endangered species, such as Kirtland's Warblers (*Setophaga kirtlandii*) and Least Bell's Vireos (*Vireo bellii pusillus*), has involved population control of cowbirds to improve reproductive success (Cooper et al., 2019; Mayfield, 1961). The frequency of nest parasitism on a given host species depends on both the relative abundances of host species in the community (Barber & Martin, 1997) and the abundance of cowbirds (Cooper et al., 2019). Therefore, monitoring the impact of cowbirds on host species population dynamics requires understanding the extent to which breeding individuals of cowbirds and host species overlap at a fine spatial scale. We fit a model with cowbirds and 16 of their frequent host species and predict abundances of each species conditioned on combinations of incidental species to evaluate the benefits of conditional prediction. We then identify host species that are positively associated with cowbird abundance as evidence that they may be at risk of nest parasitism.

In both case studies demonstrated here, we expect covariance between species from a variety of sources that suggest benefits from conditioning. Unmeasured environmental conditions, such as food supply, may result in residual covariance. For example, grass seeds are a large part of the diets of both Brown-headed Cowbirds and Chipping Sparrows (*Spizella passerina*). This food source is not measured (and therefore cannot be used as a covariate in a joint model), so the effects of food will show up as positive residual covariance—both species may be more abundant where the unobserved food resource is abundant. Therefore, an abundance of cowbirds provides indirect evidence for the unobserved shared resource. A variety of behaviours and species interactions may also produce residual covariance. Small-bodied pelagic fish species, such as Atlantic herring (*Clupea harengus*), could be informed by their generalist demersal predators (Ng et al., 2021). Cowbirds are generalist nest parasites; their abundances might be associated with abundances of their host species. Both fishes and birds can aggregate in multispecies groups (Greenberg, 2001; Parrish et al., 2002), another source of covariance. In addition to dependence between species abundances, there can also be dependence in their observation errors. The presence of a rare species can cause an observer to overlook or undercount a common species (Scher & Clark, 2023). The many ways that species can covary, beyond the variables that can be measured, suggest potential improvement from conditional prediction.

We illustrate the potential utility of conditional prediction using simulation and two applications. We use a Generalized Joint Attribute Model (GJAM), which admits multiple data types, including counts, continuous abundance, and composition data (Bachelot et al., 2018; Clark et al., 2017; Wang et al., 2019), although the framework we demonstrate can be applied broadly across joint Species Distribution Models (SDMs). The gjam R package (Clark et al., 2017) contains functions for conditional prediction; we demonstrate the workflow in a tutorial (rpubs.com/clanescher/conditionalprediction). Using simulated data, we examine the effect of data type and residual covariance on conditional prediction. Next, we evaluate applications to Atlantic cod management and management of species parasitized by Brown-headed Cowbirds. Our applications focus on species abundance (continuous biomass and discrete counts, respectively), but the simulations show that the concepts can apply to other data types, such as ordinal data (Schliep et al., 2018). Together, simulations and case studies provide guidance for the utility of conditional prediction to inform conservation by improving predictions and understanding species relationships.

2 | MATERIALS AND METHODS

2.1 | Model structure

Conditional prediction of one or more focal species extends traditional Bayesian prediction to exploit information from incidental species that are observed in the same set of observations (Figure 1b). Consider a vector $\mathbf{y}_i$ of observed abundances of S species for observations $i = 1, \dots, n$. Residual covariance in these responses (i.e. after accounting for a model mean) is estimated as a $S \times S$ covariance matrix $\boldsymbol{\Sigma}$,

$$\begin{aligned} \mathbf{y}_i &= \mathbf{B}'\mathbf{x}_i + \boldsymbol{\epsilon}_i \\ \boldsymbol{\epsilon}_i &\sim MVN(0, \Sigma) \end{aligned} \tag{1}$$

where $\mathbf{x}_i$ is a length-Q design vector of predictors, $\mathbf{B}$ is a $Q \times S$ matrix of coefficients, and $\boldsymbol{\varepsilon}_i$ is a length-S random vector of residuals (Clark et al., 2014; Ovaskainen et al., 2017; Poggiato et al., 2021; Pollock et al., 2014). Traditional prediction combines the distribution of data in eqn:MVN with the posterior distribution of parameters to generate a predictive multivariate normal (MVN) distribution for a response vector $\mathbf{y}^*$ (see Appendix 1 in Supporting Information).

Conditional prediction builds on Bayesian prediction by translating residual covariance into a species coefficient matrix $\mathbf{A}$, which allows the observed incidental species to be treated as a supplemental set of predictors. Without loss of generality, order the S species held in response vector $\mathbf{y}_i$ as the F focal species that we wish to conditionally predict $\mathcal{F} = (1, \dots, F)$ followed by I incidental species that we condition on $\mathcal{I} = (1, \dots, I)$. In other words, $\mathbf{y}_i = \left(\mathbf{y}'_{\mathcal{F},i}, \mathbf{y}'_{\mathcal{I},i}\right)'$, and $\mathbf{B} = \left(\mathbf{B}_{\mathcal{F}}, \mathbf{B}_{\mathcal{I}}\right)$. The $F \times I$ species coefficient matrix $\mathbf{A}$ comes from the residual covariance,

$$\mathbf{A} = \Sigma^{-1}_{\mathcal{I},\mathcal{I}} \Sigma_{\mathcal{I},\mathcal{F}} \tag{2}$$

Building on traditional prediction, conditional prediction for the focal species includes a supplement from the species coefficient matrix, which scales residual variance in the remaining I species,

$$\boldsymbol{\mu}_{\mathcal{F}} = \mathbf{X}\mathbf{B}_{\mathcal{F}} \tag{3}$$

$$\boldsymbol{\mu}_{\mathcal{F}|\mathcal{I}} = \mathbf{X}\mathbf{B}_{\mathcal{F}} + \mathbf{R}_{\mathcal{I}}\mathbf{A} \tag{4}$$

The second term in Equation (4) adds information from the I incidental species, which multiplies the $n \times I$ matrix of residuals $\mathbf{R}_{\mathcal{I}} = \mathbf{Y}_{\mathcal{I}} - \boldsymbol{\mu}_{\mathcal{I}}$. These are the classic equations for the conditional multivariate normal distributions and can be applied to any multivariate GLM.

We implement the conditional prediction framework using a generalized joint attribute model (GJAM) via the R package gjam (Clark et al., 2017). The model has the form of Equation (1). The covariance matrix Σ is a $S \times S$ covariance matrix, which is the basis for conditional prediction. GJAM can accommodate many types of response data; here, we demonstrate its use with continuous (CON), continuous abundance (CA), discrete abundance (DA) and presence-absence (PA). GJAM uses censoring (see Appendix 1) to avoid the use of non-linear transformation. Therefore, parameters can be interpreted on the same scale as the data. GJAM incorporates effort to set the uncertainty for each observation. In a GLM, effort can enter the model as an offset or covariate. Sample R code to conduct simple conditional prediction analyses using simulated and real data with GJAM can be found at rpubs.com/clanescher/conditionalprediction. Here, conditional prediction was implemented with the functions `gjamConditionalParameters` and `gjamPredict` in the GJAM package.

We compare traditional and conditional prediction by quantifying the reduction in root mean squared prediction error (RMSPE) for conditional prediction relative to traditional prediction. Predictions of presence-absence data are evaluated with the Brier score rather than RMSPE (Gneiting & Raftery, 2007). Positive values indicate improvement from conditioning.

2.2 | Simulation to understand model behaviour

Data were generated and analysed with the R package GJAM (Clark et al., 2017). Each simulated dataset contained 10 species that vary in their residual covariances with other species. For all species, there are five predictors with $n = 1000$ observations. Data $(\mathbf{X}, \mathbf{Y})$ and parameters $(\mathbf{B}, \boldsymbol{\Sigma})$ were randomly generated with residual covariance constrained to be positive definite.

To generate the data, we simulated 10 sets of continuous (CON) response data in which there is residual covariance between all species. We also simulated 10 sets of CON response data where there is residual covariance between only two species, that is, for the eight remaining species, all signal is captured by the covariates.

To allow comparisons across data types, we converted each continuous data set to three additional data types: continuous abundance (CA—negative values are zero), discrete abundance (DA—rounding to the nearest nonnegative integer) and presence–absence (PA—greater than zero or not; Table 1). Because these additional data types are generated from the original continuous data, all data types have the same residual covariance matrix and variance component ratio. For more information, see Appendix 2.

TABLE 1 Data types examined in simulations.

Data type	Abbreviation	Observed values	Examples
Continuous	CON	(−∞, ∞)	Body temperature
Continuous abundance	CA	[0, ∞]	Population density, biomass, basal area, nutrient concentration
Discrete abundance	DA	{0, 1, 2, …}	Counts
Presence-absence	PA	[0, 1]	Species or binary trait

We iteratively treated each simulated species as the focal species and conditioned it on all other species in the model. We aggregated percent improvement in RMSPE across species and replicates. We then compared these metrics across data types (CON, CA, DA, PA) and number of incidental species that have residual covariance with the focal species. In simulated datasets with residual covariance between all species, each focal species had residual covariance with nine incidental species. In the simulated data sets with residual covariance between two species, each focal species had residual covariance with one or zero incidental species.

We considered simulation examples in which both covariates and residual covariance could be meaningful because this is the situation most relevant to application. Variable selection typically includes predictor variables up to the point where added variables no longer contribute meaningful variation. However, in ecological data, residual variance is still typically substantial.

2.3 | Fish community data

We used biomass data from the National Oceanic and Atmospheric Administration (NOAA) Northeast Fishery Science Center (NEFSC) U.S. NES bottom trawl survey, which collects data on over 250 fish and invertebrate species (Center, 2020). We used data from the years 1998–2020, excluding years before 1998 due to many missing values for chlorophyll. To account for the seasonal migrations of many of the studied species, we modelled just the fall seasons (September through December) and included only regions that were consistently sampled in the fall. We modelled the fall because it is the feeding season so we expect interspecific co-occurrence to be more evident than during the breeding season (Zemeckis et al., 2017). Data were downloaded using the trawldata package in R (Batt, 2018). To reduce the influence of large outliers, we use the square root of biomass (measured as species-specific kg/tow) as the response variable.

We selected the 24 most common species, which were present in at least 150 tows, and account for environmental variation in the model using bottom and surface ocean temperature, bottom salinity, depth, sediment size, chlorophyll A concentration (Chla) and month. Depth, temperature and salinity were measured in situ during trawl surveys, with missing values augmented using the data-assimilative HYbrid Coordinate Ocean Model (HYCOM) daily and then monthly data (Chassignet et al., 2007). Sediment size (grain size in mm) was obtained from The Nature Conservancy's Northwest Atlantic Marine Ecoregional Assessment (Greene et al., 2010). Chla was obtained from the MODIS satellite (monthly rasters from 2003 to 2019) on a monthly time step (NASA Goddard Space Flight Center, O. B. P. G. Ocean Ecology The Laboratory, 2018a), with missing values filled using the SeaWIFS satellite (1998 to 2009) (NASA Goddard Space Flight Center, O. B. P. G. Ocean Ecology The Laboratory, 2018b). We examined the benefits of conditioning for in-sample and out-of-sample predictions of each species. We focused on predictions of Atlantic cod to help inform resource managers and their decision-making.

2.4 | Bird community data

Bird abundance data came from the Breeding Bird Survey (BBS, Pardieck et al., 2020) conducted between 2015 and 2020 in Bird Conservation Regions 12, 22 and 23, that is, part of the Midwest region of the United States, which contains the entire breeding range of the Kirtland's Warbler. A BBS observation consists of 50 point counts of 3 min each along a preselected route, for a total of 150 min per observation. We selected 17 species to model: Brown-headed Cowbird and 16 species whose nests it frequently parasitizes (Friedmann, 1963). Environmental variation was represented by temperature and precipitation site means and annual anomalies taken over the previous 12 months (Thornton et al., 2020), elevation (Hollister et al., 2021) and land cover (National Land Cover Database Dewitz (2019)). Land cover was included as the percent of seven land cover categories (developed, forest, herbaceous, planted, shrub, water, wetlands) within 10 km of the BBS route.

We tested how the benefit of conditional prediction is related to the number of incidental species conditioned on. We alternately selected focal species, and we conditioned on groups of incidental species ranging in size from one to 16 species. To determine whether incidental species provide different benefits to prediction (i.e. some incidental species improve predictions more than others), we predicted each focal species conditionally on different combinations of incidental species. For each group size, we selected 10 random combinations of incidental species to condition on (except the group of 16 incidental species, for which there is only one possible combination).

We also examined the species coefficient matrix **A** (Equation (10) in Appendix 1) to understand relationships between species abundances. The relationships quantified by the species coefficient matrix indicate the amount of information that the residual abundance of each species adds to predictions of each of the others. Specifically, we considered the information contributed by cowbird abundance to predictions of host species, and the information contributed by host species abundance to predictions of cowbirds.

Because this was a modelling study, it did not require any ethical approval.

3 | RESULTS

3.1 | Simulation

Simulations demonstrated that the benefit of conditional prediction depends on residual covariance and data type. Improvement from conditioning was greatest when there was residual covariance between the focal species and many incidental species (Figure 2a). When there was no residual covariance with any incidental species, conditional prediction was equivalent to traditional prediction for all data types (Figure 2c). Data types that contain more information (e.g. continuous [CON] and continuous abundance [CA]) benefited from conditioning more than other data types. Discrete abundance (DA) data generally benefited from conditioning, but the amount of improvement was more variable than for other data types (Figure 2a). Conditioning caused little to no improvement for presence–absence (PA) data, regardless of residual covariance. Within each replicate, all data types are generated from the same residual covariance matrix, data and parameters, so the variance component ratio is the same across data types and does not impact comparisons of results across data types.

3.2 | Conditionally predicting biomass of overfished species

Conditioning on incidental fish species biomass improved in- and out-of-sample prediction across species. Conditioning provided a similar benefit to in- and out-of-sample predictions (Figure 3a), but there was variation in the magnitude of improvement across species (Figure 3b). For example, Atlantic cod was among the species that benefited most from conditioning, with 20.4% and 17.9% improvement in RMSPE for in-sample and out-of-sample predictions, respectively. Conversely, conditioning offered little improvement for Summer Flounder (2.5% and 2.8% change in- and out-of-sample, respectively). Species in Figure 3b are ordered by abundance; there is no relationship between species abundance and the benefit gained from conditional prediction.

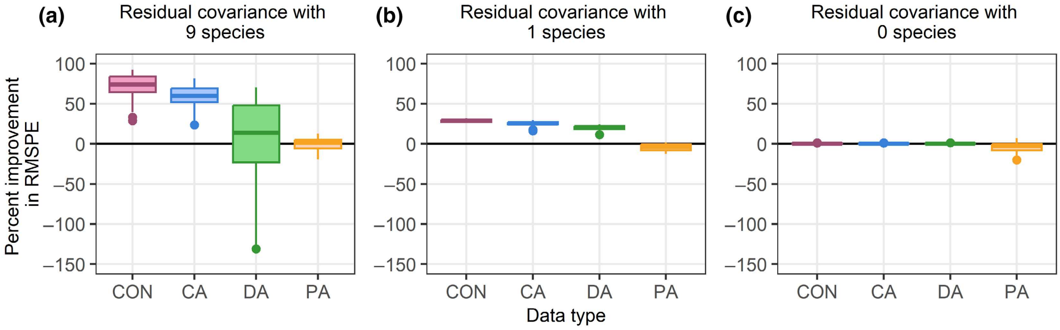

FIGURE 2 Percent improvement in RMSPE across data types for simulated data. Focal species are grouped by the number of incidental species with which they have residual covariance. From the models with residual covariance between all species, focal species have residual covariance with nine incidental species (a). From the models with residual covariance between only two species, focal species may have residual covariance with one incidental species (b) or zero incidental species (c). Data types shown are continuous (CON), continuous abundance (CA), discrete abundance (DA) and presence–absence (PA). The upper and lower hinges correspond to the first and third quartiles, and the horizontal line represents the median. Values more than 1.5 times the distance between the first and third quartiles beyond the hinges are considered outliers and plotted individually.

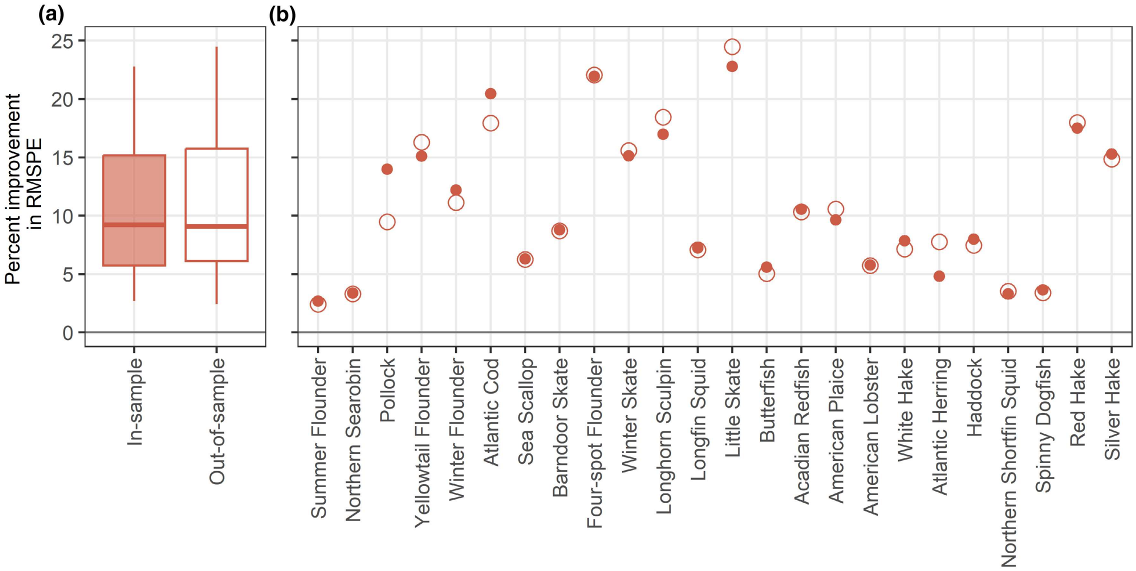

FIGURE 3 The percent improvement in RMSPE for in-sample (filled) and out-of-sample (open) prediction across fish species (a) and for each fish species (b). Species in b are ordered by abundance, with rare species on the left. In (a), the upper and lower hinges correspond to the first and third quartiles; and the horizontal line represents the median.

Conditioning predictions of Atlantic cod biomass on incidental species biomass captured more local variation than traditional prediction. Traditional prediction of Atlantic cod predicted values close to the mean, resulting in overprediction in areas with low counts and underprediction in areas with high counts (Figure 4b). Conditioning revealed variation in biomass that more closely matched observed values (Figure 4c). In our study area, conditional prediction severely overpredicts biomass in one region that had zero observed cod biomass. The overpredicted region is directly south of a region with high observed cod biomass which conditional prediction estimates accurately. The high biomass predicted in the southern region may result from a similar composition of incidental species in these two regions.

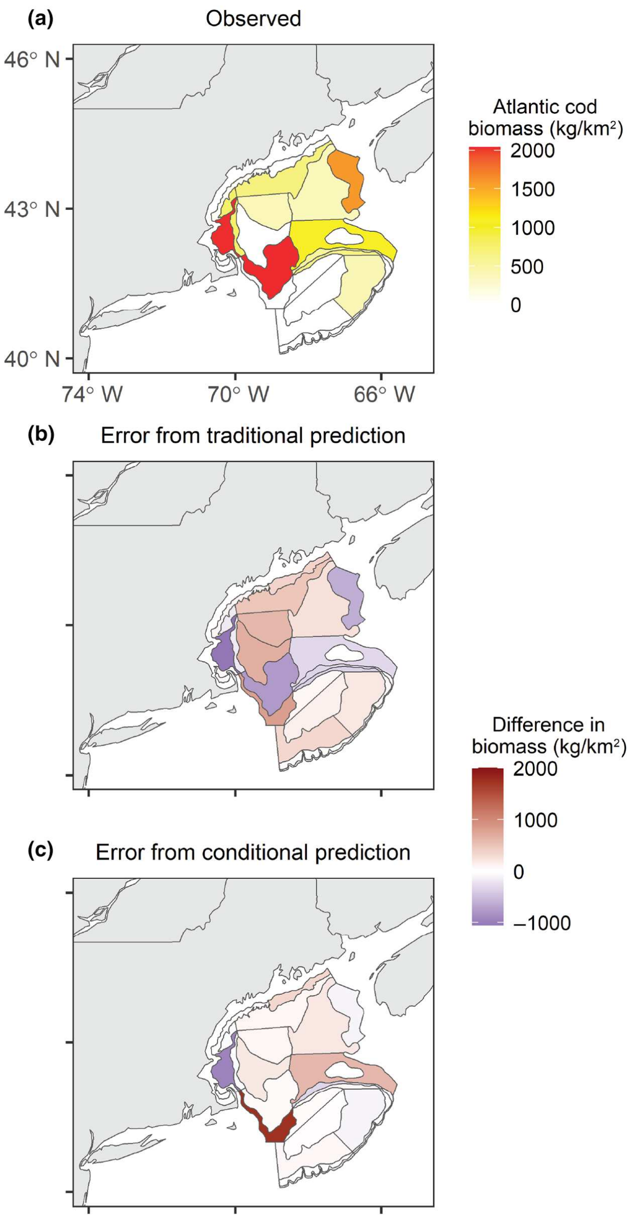

FIGURE 4 Maps of Atlantic cod observed biomass per region $\left(\text{kg}/\text{km}^2\right)$ (a) and difference between observed biomass and predicted biomass for traditional prediction (b) and conditional prediction (c). Maps show data from 2012 (Center, 2020). Base from Natural Earth 1:10,000,000-scale digital data.

3.3 | An obligate nest parasite and its hosts

Conditioning improved predictions of all bird species. Improvement in RMSPE was greatest when conditioning on many species rather than few (Figure 5). For most species, RMSPE improved by up to 8.3%–24.0%. RMSPE for American Redstart (*Setophaga ruticilla*), however, improved by only 2.4% even when conditioned on all incidental species. Interestingly, for a given number of incidental species conditioned on, the different combinations of incidental species caused little variation in RMSPE improvement (Figure 5, shaded areas indicate one standard error). As with the previous example, there is no clear pattern between a species' abundance and the benefit it gains from conditioning.

Using the species coefficient matrix **A**, we examined the relationships between species abundances and the improvement in conditional prediction from each species. Conditional predictions of cowbird abundance are informed primarily by residual abundance of three species: Kirtland's Warbler and Yellow-breasted Chat (*Icteria virens*) indicate fewer cowbirds, whereas Bell's Vireo (*Vireo bellii*) indicates more cowbirds than expected (Figure 6). Using traditional prediction, our model expects eight cowbirds at locations where Kirtland's Warblers are observed; conditional prediction expects fewer than five cowbirds at these locations, which is closer to the observed counts (Figure S1).

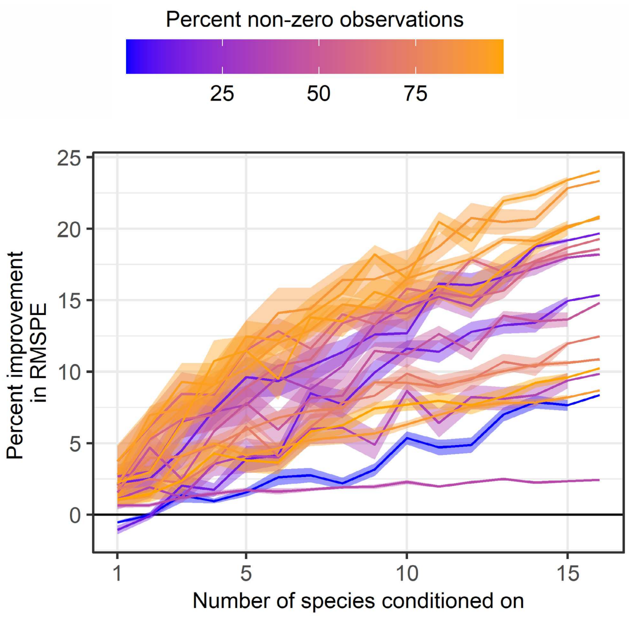

FIGURE 5 Percent improvement in RMSPE by conditioning on incidental species for each bird species in the model. Each line represents a focal species. Line colour indicates the percent of observations in which the species was seen (rare species are blue; common species are orange). The shaded areas show one standard error from the mean. Note that there is no standard error when conditioning on 16 species because there is only one combination of 16 species to condition on. RMSPE, root mean squared prediction error.

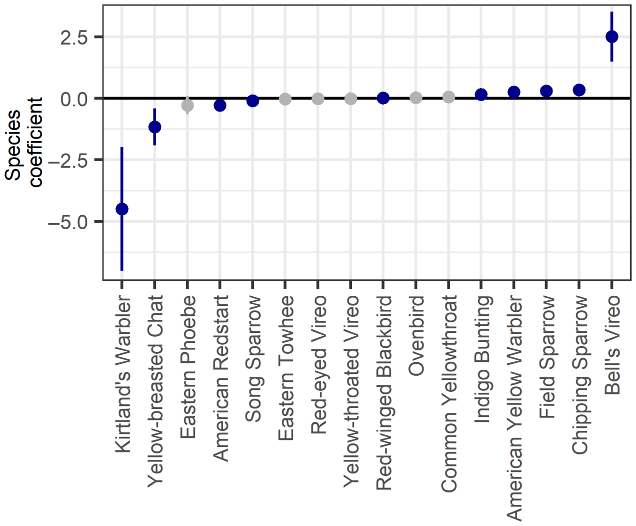

FIGURE 6 Species coefficients indicating the information gained from host species abundance about cowbird abundance. Points represent the mean estimate, and lines show one standard error from the mean. Blue lines represent standard errors that do not cross zero; grey lines are standard errors that do cross zero.

Our model expects 19–23 cowbirds where Bell's Vireos are observed using traditional prediction, but 22–28 using conditional prediction, also closer to the observed counts. The species coefficient matrix also showed that conditioning on cowbird abundance does not substantially improve predictions of each host species (Figure S2b). Although the standard errors for the estimates of most coefficients do not cross zero (Figure S2b, blue points), the magnitude of these coefficients is much smaller than the coefficients representing improvement in predictions of cowbirds provided by hosts (Figure S2a).

4 | DISCUSSION

Conditioning on observations of jointly modelled incidental species improved predictions across simulated data, continuous measures of fish biomass and discrete bird counts in all cases except simulated presence–absence data. Conditioning improved RMSPE by between 5% and 25% for most fish (Figure 3) and bird (Figure 5) species. The fish case study demonstrated that conditioning can improve out-of-sample predictions as much as in-sample predictions (Figure 3). From the bird application, we found that improvement in RMSPE of the focal species increases with the number of incidental species (Figure 5). Interestingly, the identities of the species conditioned on did not substantially influence the amount of improvement from conditioning (Figure 5).

The improved accuracy of conditional predictions holds promise for management and conservation efforts. Management of Atlantic cod involves allocating quotas based on historic population levels and assessing their spatial distribution using SDMs. Despite a longtime push for ecosystem-based fisheries management that incorporates species dependencies and environmental influences, progress towards ecosystem-based fisheries management in practice in the northeastern United States has remained slow (but see Townsend et al. (2019)). We provide one example of how incorporating species dependencies in spatially explicit models of cod biomass can benefit conservation efforts. When allocating quota and assessing fishing grounds, managers and fishers could draw information from both the environment and other species. This information can be particularly useful in cases of bycatch management, where the unintended catch of a rare or protected species, such as cod or sea turtles, can close an otherwise sustainable fishery (Santora, 2003). Our maps of Atlantic cod abundance, predicted conditionally on other species, can more accurately identify areas where unintended catch of Atlantic cod is likely, reducing the chances of future fisheries closures.

Predicting species conditionally on others can address many practical conservation challenges. There is a strong incentive for fishing ships to under-report bycatch species (Davies et al., 2009). Conditioning predictions of bycatch on target species returns, which are likely more accurate, might provide better estimates of bycatch than what is reported (Zhang et al., 2020). Further, the analysis demonstrated here can be developed into a tool to allow fishers to make decisions in real time. Fishers can predict bycatch using information about the environment and conditioning on the individuals already caught during a particular fishing trip. Such a tool would allow fishers to decide to move on before the bycatch of a rare species closes a fishery for an entire season (Beutel et al., 2008).

In addition to improving predictions, the species coefficient matrix can improve understanding of the relationships between species abundances. This matrix is derived from the covariance matrix (Equation 10 in Appendix 1), but is asymmetric: coefficients indicate how the residual abundance of species A informs prediction of the residual abundance of species B and vice versa. The bird case study demonstrates this asymmetry. Residual abundance of cowbirds contributed little information to the prediction of host species (Figure S2b), whereas residual abundance of three host species informed predictions of cowbird abundance (Figure S2a). This asymmetry may be related to the degree of specialization by a species. For example, Kirtland's Warblers are extreme habitat specialists, breeding only in young Jack Pine stands (Donner et al., 2008). Their presence, therefore, indicates a very specific local habitat. By comparison, cowbirds are generalists, whose presence provides little specific information about habitat conditions. It is important to note that the species coefficient matrix asymmetry does not represent species interactions. For example, Kirtland's Warblers do not depredate cowbirds, even though the coefficient is negative, but rather where Kirtland's Warblers are more abundant, cowbirds are less abundant.

The relationships between species abundances that are captured by the species coefficient matrix can help inform management strategies. The relationships between host species and cowbird abundance, for example, indicate the amount that each host is potentially exposed to cowbird parasitism. For 13 of 16 host species, there were weak relationships between residual cowbird abundance and residual host species abundance suggesting that these host species are not particularly vulnerable to nest parasitism. Two cases where there are strong relationships between cowbirds and host species demonstrate how these relationships can inform management. First, we found that where there are Kirtland's Warblers there are fewer cowbirds than expected from the model covariates (Figure 6). This relationship is likely driven by management programs that removed cowbirds from Kirtland's Warblers breeding areas, including our study area. Interestingly, we identified this pattern using data from 2015 to 2020, when cowbird trapping was reduced and eliminated (Cooper et al., 2019). Our results support recent findings that cowbird abundances have remained low in Kirtland's Warblers breeding areas even after removal programs ended (Cooper et al., 2019). We suggest that this analysis be repeated periodically using current data to quickly determine whether cowbird abundances are increasing and removals should resume. Conversely, our results indicated that where there are Bell's Vireos, there are more cowbirds than expected from the model covariates (Figure 6a). This result, combined with the fact that Bell's Vireo nests are parasitized by cowbirds at high rates (up to 54% in our study region, Mumford, 1952) suggests that Bell's Vireos may be at high risk of nest parasitism and their reproductive success should be monitored. Interestingly, the federally endangered Least Bell's Vireo subspecies, which occurs in Southwestern California outside our study area, has been managed with cowbird removals that have successfully reduced parasitism and increased reproductive success (Kus & Whitfield, 2005).

There are limitations to the benefit of conditional prediction, which are apparent in our simulations and case studies. First, the type of response data influences the improvement that comes from conditioning (Figure 2). Continuous and continuous abundance data benefit most. In our simulations, conditioning provided no benefit for presence–absence (PA) data. In PA data, all information is collapsed into zeros and ones, so species abundances (again, limited to (0, 1)) have limited information to offer about others. This pattern likely extends across cases regardless of species ecology and spatial scale. Second, residual covariance between species is required to improve predictions. There is more improvement in predictions of a focal species when it has residual covariance with many incidental species than with few species. Finally, our case studies demonstrate that the benefits of conditioning vary across focal species. Although conditioning leads to almost a 25% improvement in RMSPE for many species, others benefit less. We cannot decipher from our results why some species benefit more than others. It is possible that traits of the focal species, and whether they are similar to incidental species, influence the benefit from conditioning, but a trait-based assessment could support understanding of these patterns.

The conditional prediction framework has broad application beyond what is demonstrated here. Previous studies have shown the utility of conditional prediction beyond species, to traits (Seyednasrollah & Clark, 2020) and life stages (Qiu et al., 2021). The benefits of conditional prediction to out-of-sample predictions (Figure 3) open various avenues for applications. After fitting a model with data from an entire community of interest, surveys of a subset of those species could be used to predict the rest of the community. This framework could expand the utility of partial surveys and reduce the necessity of conducting full-community surveys. Conditional prediction could also aid in understanding the future distributions of invasive or reintroduced species. A model fit with data from a species' established region could be used to predict its future distribution conditionally on the community in its expanded range.

Here, we demonstrate conditional prediction using a generalized joint attribute model (GJAM), but the framework can be applied to a variety of models and situations. Using the classic equations for the conditional multivariate normal distributions as described here, conditional prediction can be used with any multivariate GLM that produces a residual covariance matrix. Both frequentist and Bayesian multivariate GLMs are suitable, regardless of the method of Bayesian approximation (e.g. MCMC, INLA). Previous attempts at conditional prediction also demonstrated that conditioning improves predictions, but those were limited to predicting occurrence conditionally on occurrence based on probability theory (Wilkinson et al., 2020). The equations presented here make use of the residual covariance matrix to allow prediction of abundance conditionally on residual abundance.

There are additional extensions of conditional prediction that may be possible with further testing. Our examples condition on measured abundance or occurrence of incidental species, but this framework can likely be used to condition on predicted abundance or occurrence, allowing for conditional prediction in regions where neither the focal species nor the incidental species were sampled. Although we use data sets collected at a single spatiotemporal scale, exploring whether a focal species can be predicted at a higher spatial or temporal resolution than it was observed by conditioning on an incidental species that was observed at a high resolution could provide additional insight. While this was not explicitly tested in our analysis, it is an exciting avenue for future research.

Through simulations and two applications to real data, we show that conditional prediction can inform conservation and management by improving accuracy of predictions and illuminating relationships between species abundances. Conditional prediction expands the utility of jSDMs by extracting information from the residual covariance matrix to predict species conditionally on other species. Further, the species coefficient matrix used for conditional prediction provides information about relationships between species abundances. We present guidelines on how this framework should be applied across data types and species abundance. Researchers can evaluate the suitability of conditional prediction for their system by comparing RMSPE of traditional predictions with RMSPE of conditional predictions, and by assessing whether the species coefficient matrix holds reasonable values.

AUTHOR CONTRIBUTIONS

C. Lane Scher, Sarah M. Roberts, James S. Clark and Kevin P. Krause conceived the ideas and designed methodology; C. Lane Scher and Sarah M. Roberts assembled the data; C. Lane Scher, Sarah M. Roberts and James S. Clark analysed the data; C. Lane Scher led the writing of the manuscript. C. Lane Scher, Sarah M. Roberts, James S. Clark and Kevin P. Krause contributed critically to the drafts and gave final approval for publication. Our study used publicly available data. The authorship team is broadly representative of the regions where the studies are based.

ACKNOWLEDGEMENTS

We thank Paul McLaughlin and two anonymous reviewers for comments on the manuscript. The project was funded by the National Science Foundation Graduate Research Fellowship Grant

DGE-1644868 to C.L.S. and DEB-1754443 and DEB-2211764 grants to J.S.C., who was further supported by the Belmont Forum grant 1854976, NASA grant AIST18-0063 and the Programme d'Investissement d'Avenir under project FORBIC (18-MPGA-0004) (*Make Our Planet Great Again*). K.P.K. was supported by the U.S. Geological Survey's Ecosystems Mission Area. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data used for this analysis are publicly available. Species abundance data were taken from the r package trawlData (Batt, 2018; Center, 2020) and the Breeding Bird Survey data portal (Pardieck et al., 2020). Environmental data were taken from a variety of publicly available sources that are detailed in the manuscript. All code needed to conduct the analysis is available on GitHub (github.com/lanescher/conditional-prediction) and Zenodo (https://doi.org/10.5281/zenodo.10839346) (Scher et al., 2024).

ORCID

C. Lane Scher https://orcid.org/0000-0003-3689-5769
Sarah M. Roberts https://orcid.org/0000-0001-5441-3458
Kevin P. Krause https://orcid.org/0000-0002-0255-7027

REFERENCES

Bachelot, B., Uriarte, M., Muscarella, R., Forero-Montaña, J., Thompson, J., McGuire, K., Zimmerman, J., Swenson, N. G., & Clark, J. S. (2018). Associations among arbuscular mycorrhizal fungi and seedlings are predicted to change with tree successional status. *Ecology*, *99*, 607–620. https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1002/ecy.2122

Barber, D. R., & Martin, T. E. (1997). Influence of alternate host densities on Brown-headed Cowbird parasitism rates in black-capped vireos. *The Condor*, *99*, 595–604. https://academic.oup.com/condor/article/99/3/595/5124463

Batt, R. (2018). Trawl Data Package: A compendium of North American bottom trawl survey data, documentation, and related tools and resources. https://github.com/rBatt/trawlData/

Bean, W. T., Prugh, L. R., Stafford, R., Butterfield, H. S., Westphal, M., & Brashares, J. S. (2014). Species distribution models of an endangered rodent offer conflicting measures of habitat quality at multiple scales. *Journal of Applied Ecology*, *51*, 1116–1125.

Beutel, D., Skrobe, L., Castro, K., Ruhle, P., Sr., Ruhle, P., Jr., O'Grady, J., & Knight, J. (2008). Bycatch reduction in the Northeast USA directed haddock bottom trawl fishery. *Fisheries Research*, *94*, 190–198.

Blanchet, F. G., Cazelles, K., & Gravel, D. (2020). Co-occurrence is not evidence of ecological interactions. *Ecology Letters*, *23*, 1050–1063.

Center, N. F. S. (2020). Fall Bottom Trawl Survey. https://www.fisheries.noaa.gov/inport/item/22560

Chassignet, E. P., Hurlburt, H. E., Smedstad, O. M., Halliwell, G. R., Hogan, P. J., Wallcraft, A. J., Baraille, R., & Bleck, R. (2007). The HYCOM (hybrid coordinate ocean model) data assimilative system. *Journal of Marine Systems*, *65*, 60–83.

Clark, J. S., Gelfand, A. E., Woodall, C. W., & Zhu, K. (2014). More than the sum of the parts: Forest climate response from joint species distribution models. *Ecological Applications*, *24*, 990–999. https://doi.org/10.1890/13-1015.1

Clark, J. S., Nemergut, D., Seyednasrollah, B., Turner, P. J., & Zhang, S. (2017). Generalized joint attribute modeling for biodiversity analysis: Median-zero, multivariate, multifarious data. *Ecological Monographs*, *87*, 34–56. https://doi.org/10.1002/ecm.1241

Clark, J. S., Scher, C. L., & Swift, M. (2020). The emergent interactions that govern biodiversity change. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(29), 17074–17083. https://www.pnas.org/content/pnas/early/2020/07/02/2003852117.full.pdf

Cooper, N. W., Rushing, C. S., & Marra, P. P. (2019). Reducing the conservation reliance of the endangered Kirtland's warbler through adaptive management. *The Journal of Wildlife Management*, *83*, 1297–1305. https://doi.org/10.1002/jwmg.21714

Davies, R., Cripps, S., Nickson, A., & Porter, G. (2009). Defining and estimating global marine fisheries bycatch. *Marine Policy*, *33*, 661–672.

Dewitz, J. (2019). National land cover database (NLCD) 2016 products: U.S. Geological Survey data release.

Donner, D. M., Probst, J. R., & Ribic, C. A. (2008). Influence of habitat amount, arrangement, and use on population trend estimates of male Kirtland's warblers. *Landscape Ecology*, *23*, 467–480.

Friedmann, H. (1963). Host relations of the parasitic cowbirds. *Bulletin of the United States National Museum.*, *233*, 1–276.

Gneiting, T., & Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, *102*, 359–378. https://doi.org/10.1198/016214506000001437

Greenberg, R. (2001). Birds of many feathers: The formation and structure of mixed-species flocks of forest birds. On the move: How and why animals travel in groups.

Greene, J., Anderson, M., Odell, J., & Steinberg, N. (2010). *The Northwest Atlantic marine ecoregional assessment: Species, habitats and ecosystems*. Phase one (p. 230). The Nature Conservancy, Eastern US Division.

H.R.2617, 117th Congress. (2022). *Magnuson-stevens fishery conservation and management act* (pp. 94–265). Public Law.

Hollister, J., Shah, T., Robitaille, A. L., Beck, M. W., & Johnson, M. (2021). Elevatr: Access elevation data from various APIs. https://github.com/jhollist/elevatr/

Kus, B. E., & Whitfield, M. J. (2005). Parasitism, productivity, and population growth: Response of least Bell's vireos (*Vireo bellii pusillus*) and southwestern willow flycatchers (*Empidonax traillii extimus*) to cowbird (Molothrus spp.) control. *Ornithological Monographs*, *57*, 16–27.

Lorenzana, J. C., & Sealy, S. G. (1999). A meta-analysis of the impact of parasitism by the Brown-headed Cowbird on its hosts. *Studies in Avian Biology*, *18*, 241–253.

Mayfield, H. (1961). Cowbird parasitism and the population of the Kirtland's warbler. *Evolution*, *15*, 174.

Mumford, R. E. (1952). Bell's vireo in Indiana. *The Wilson Bulletin*, *64*, 224–233. https://www.jstor.org/stable/4158133

NASA Goddard Space Flight Center, O. B. P. G. Ocean Ecology The Laboratory. (2018a). Moderate-resolution imaging spectroradiometer (MODIS) aqua L0 data.

NASA Goddard Space Flight Center, O. B. P. G. Ocean Ecology The Laboratory. (2018b). Sea-viewing Wide Field-of-view Sensor (SeaWiFS) Data.

Ng, E. L., Deroba, J. J., Essington, T. E., Grüss, A., Smith, B. E., & Thorson, J. T. (2021). Predator stomach contents can provide accurate indices of prey biomass. *ICES Journal of Marine Science*, *78*, 1146–1159.

Ovaskainen, O., Tikhonov, G., Dunson, D., Grotan, V., Engen, S., Saether, B. E., & Abrego, N. (2017). How are species interactions structured in species-rich communities? A new method for analysing time-series data. *Proceedings of the Royal Society B: Biological Sciences*, *284*, 20170768.

Pardieck, K. L., Ziolkowski, D. J., Jr., Lutmerding, M., Aponte, V. I., & Hudson, M.-A. R. (2020). North American Breeding Bird Survey

Dataset 1966–2019: U.S. Geological Survey data release. https://doi.org/10.5066/P9J6QUF6

Parrish, J. K., Viscido, S. V., & Grunbaum, D. (2002). Self-organized fish schools: An examination of emergent properties. *The Biological Bulletin*, *202*, 296–305.

Pershing, A. J., Alexander, M. A., Brady, D. C., Brickman, D., Curchitser, E. N., Diamond, A. W., McClenachan, L., Mills, K. E., Nichols, O. C., Pendleton, D. E., Record, N. R., Scott, J. D., Staudinger, M. D., & Wang, Y. (2021). Climate impacts on the Gulf of Maine ecosystem: A review of observed and expected changes in 2050 from rising temperatures. *Elementa: Science of the Anthropocene*, *9*, 00076.

Pershing, A. J., Alexander, M. A., Hernandez, C. M., Kerr, L. A., Le Bris, A., Mills, K. E., Nye, J. A., Record, N. R., Scannell, H. A., Scott, J. D., Sherwood, G. D., & Thomas, A. C. (2015). Slow adaptation in the face of rapid warming leads to collapse of the Gulf of Maine cod fishery. *Science*, *350*, 809–812.

Poggiato, G., Münkemüller, T., Bystrova, D., Arbel, J., Clark, J. S., & Thuiller, W. (2021). On the interpretations of joint modeling in community ecology. *Trends in Ecology & Evolution*, *36*, 391–401. https://www.sciencedirect.com/science/article/pii/S0169534721000045

Pollock, L., Tingley, R., Morris, W., Golding, N., O'Hara, R., Parris, K., Vesk, P., & McCarthy, M. (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*, *5*, 397–406.

Qiu, T., Sharma, S., Woodall, C., & Clark, J. S. (2021). Niche shifts from trees to fecundity to recruitment that determine species response to climate change. *Frontiers in Ecology and Evolution*, *9*, 1–12.

Santora, C. (2003). Management of turtle bycatch: Can endangered species be protected while minimizing socioeconomic impacts? *Coastal Management*, *31*, 423–434.

Scher, C. L., & Clark, J. S. (2023). Species traits and observer behaviors that bias data assimilation and how to accommodate them. *Ecological Applications*, *33*, e2815. https://doi.org/10.1002/eap.2815

Scher, C. L., Roberts, S. M., Krause, K. P., & Clark, J. S. (2024). Code from: Leveraging relationships between species abundances to improve predictions and inform conservation (v1.0.0). Zenodo. https://doi.org/10.5281/zenodo.10839346

Schliep, E. M., Lany, N. K., Zarnetske, P. L., Schaeffer, R. N., Orians, C. M., Orwig, D. A., & Preisser, E. L. (2018). Joint species distribution modelling for spatio-temporal occurrence and ordinal abundance data. *Global Ecology and Biogeography*, *27*, 142–155. https://doi.org/10.1111/geb.12666

Seyednasrollah, B., & Clark, J. S. (2020). Where resource-acquisitive species are located: The role of habitat heterogeneity. *Geophysical Research Letters*, *47*, e2020GL087626 https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020GL087626

Sguotti, C., Otto, S. A., Frelat, R., Langbehn, T. J., Ryberg, M. P., Lindegren, M., Durant, J. M., Stenseth, N. C., & Möllmann, C. (2019). Catastrophic dynamics limit Atlantic cod recovery. *Proceedings of the Royal Society B: Biological Sciences*, *286*, 20182877.

Tang, B., Clark, J. S., & Gelfand, A. E. (2021). Modeling spatially biased citizen science effort through the eBird database. *Environmental and Ecological Statistics*, *28*, 609–630.

Tang, B., Kamakura, R. P., Barnett, D. T., & Clark, J. S. (2023). Learning from monitoring networks: Few-large vs. many-small plots and multi-scale analysis. *Frontiers in Evolution and Ecology*, *11*, 1–12. https://www.frontiersin.org/articles/10.3389/fevo.2023.1114569

Thornton, M., Shrestha, R., Wei, Y., Thornton, P., Kao, S., & Wilson, B. (2020). Daymet: Daily surface weather data on a 1-km grid for North America, version 4. https://daac.ornl.gov/cgi-bin/dsviewer.pl?ds_id=1840

Townsend, H., Harvey, C. J., deReynier, Y., Davis, D., Zador, S. G., Gaichas, S., Weijerman, M., Hazen, E. L., & Kaplan, I. C. (2019). Progress on implementing ecosystem-based fisheries management in the United States through the use of ecosystem models and analysis. *Frontiers in Marine Science*, *6*, 641.

Wang, Z., Juarez, D. L., Pan, J.-F., Blinebry, S. K., Gronniger, J., Clark, J. S., Johnson, Z. I., & Hunt, D. E. (2019). Microbial communities across nearshore to offshore coastal transects are primarily shaped by distance and temperature. *Environmental Microbiology*, *21*, 3862–3872. https://sfamjournals.onlinelibrary.wiley.com/doi/abs/10.1111/1462-2920.14734

Wilkinson, D. P., Golding, N., Guillera-Arroita, G., Tingley, R., & McCarthy, M. A. (2020). Defining and evaluating predictions of joint species distribution models. *Methods in Ecology and Evolution*, *12*, 394–404. https://doi.org/10.1111/2041-210X.13518

Zemeckis, D. R., Liu, C., Cowles, G. W., Dean, M. J., Hoffman, W. S., Martins, D., & Cadrin, S. X. (2017). Seasonal movements and connectivity of an Atlantic cod (*Gadus morhua*) spawning component in the western Gulf of Maine. *ICES Journal of Marine Science*, *74*, 1780–1796.

Zhang, C., Chen, Y., Xu, B., Xue, Y., & Ren, Y. (2020). Improving prediction of rare species' distribution from community data. *Scientific Reports*, *10*, 12230. https://doi.org/10.1038/s41598-020-69157-x

Zurell, D., Pollock, L. J., & Thuiller, W. (2018). Do joint species distribution models reliably detect interspecific interactions from co-occurrence data in homogenous environments? *Ecography*, *41*, 1812–1819.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Figure S1. Predicted abundance of Brown-headed Cowbird using traditional prediction (red) and conditional prediction (yellow) at all observations, observations where Bell's Vireo is present, and observations where Kirtland's Warbler is present. Conditional predictions tend to be closer than traditional predictions to the observed counts.

Figure S2. Species coefficients indicating the information gained from host species abundance about cowbird abundance (a) and from cowbird abundance about host species abundance (b). Points represent the mean estimate, and lines show one standard error from the mean. Blue lines represent standard errors that do not cross zero; grey lines are standard errors that do cross zero.

Appendix 1. Additional information about model structure, count data and effort, and quantifying the conditional effect.

Appendix 2. Additional information about simulated data.

How to cite this article: Scher, C. L., Roberts, S. M., Krause, K. P., & Clark, J. S. (2024). Leveraging relationships between species abundances to improve predictions and inform conservation. *Journal of Applied Ecology*, *61*, 1662–1672. https://doi.org/10.1111/1365-2664.14670