Beyond single-species forecasts: leveraging multispecies models to improve predictions and navigate the dynamics of ecological predictability

Nicholas J. Clark[1,2*], S. K. Morgan Ernest[3], Henry Senyondo[3], Juniper L. Simonis[3,4], Ethan P. White[3], Glenda M. Yenni[3], K. A. N. K. Karunarathna[1,2,5]

[1] School of Veterinary Science, Faculty of Science, The University of Queensland, Queensland 4343, Australia

[2] UQ Spatial Epidemiology Laboratory, School of Veterinary Science, The University of Queensland, Gatton, Queensland 4343, Australia

[3] Wildlife Ecology and Conservation, University of Florida, Gainesville, Florida 32611, USA

[4] DAPPER Stats, 3519 NE 15th Avenue, Suite 467, Portland, Oregon 97212, USA

[5] Department of Mathematics, Faculty of Science, Eastern University, Sri Lanka

***Corresponding author:** Nicholas J Clark, n.clark@uq.edu.au

Short title: Multivariate ecological forecasting

Total word count (including references): 6622

Number of words in Abstract: 149

Number of References: 56

Number of Tables: 1

Number of Figures: 6

Authorship:

NJC, SKM and EPW designed the study; SKM and GMY oversaw field data collection; HS, JLS and EPW designed cyberinfrastructure to store and access data; NJC and KANKK performed statistical analyses and visualizations; NJC drafted the manuscript and all authors contributed to revisions.

Data availability statement:

All data from the Portal Project is publicly archived to Zenodo (Ernest et al. 2023) and is available in processed forms using the portalr R package (Christensen et al. 2019b). R code to reproduce analyses is provided in Supplementary materials and in a GitHub repository (https://github.com/nicholasjclark/portal_VAR). R code will be permanently archived on Zenodo on acceptance of the manuscript.

ABSTRACT

Forecasting the responses of natural populations to environmental change is a key priority in the management of ecological systems. This is challenging because the dynamics of multispecies ecological communities are influenced by many factors. Populations can exhibit complex, nonlinear responses to environmental change, often over multiple temporal lags. In addition, biotic interactions, and other sources of multi-species dependence, are major contributors to patterns of population variation. Theory suggests that near-term ecological forecasts of population abundances can be improved by modelling these dependencies, but empirical support for this idea is lacking. We test whether models that learn from multiple species, both to estimate nonlinear environmental effects and temporal interactions, improve ecological forecasts for a semi-arid rodent community. Using Dynamic Generalized Additive Models, we analyze monthly captures for nine rodents over 25 years. Model comparisons provide strong evidence that multi-species dependencies improve performance, as models that captured these effects gave superior predictions than models that ignored them. We show changes in abundance for some species can have delayed, nonlinear effects on others, and that lagged effects of temperature and vegetation greenness are key drivers of change. Our findings highlight that multivariate models are useful not only to improve near-term ecological forecasts but also to ask targeted questions about community dynamics.

KEYWORDS

Biotic interactions, Community dynamics, Ecological forecasting, Generalized additive model, Stan, State-Space model

INTRODUCTION

Predicting the impacts of environmental change on ecosystem function and biodiversity is a global challenge (Clark et al. 2001, Intergovernmental Science - Policy Platform on Biodiversity and Ecosystem Services 2019, Fredston et al. 2023). Explicit predictions are needed to guide ecological management decisions, inform monitoring programs, and perform scenario planning (Lindenmayer et al. 2012, Tulloch et al. 2020). This has led to a growing emphasis on the importance of near-term ecological forecasting to encourage greater reliance on ecological time series data, and on suitable models that can handle the complexities of these data, to generate quantitative forecasts that can be harnessed to guide management decisions (Dietze et al. 2018, Lewis et al. 2023, Karunarathna et al. 2024). The applications of ecological forecasting are broad, including the prediction of soil microbiome compositions (Averill et al. 2021), carbon cycle dynamics (Dietze et al. 2014) and species' population dynamics (Ward et al. 2014, White et al. 2019, Johnson-Bice et al. 2021).

Forecasts for species population dynamics are especially crucial for conservation planning, stock assessments and other ecological management priorities. However, these forecasts typically focus on only a single species at a time (Quinn 2003, Simonis et al. 2021, Lewis et al. 2022) or on aggregate measures such as species richness, biomass or diversity (Algar et al. 2009, Tonkin et al. 2017, Clark et al. 2020). This is problematic because key applications of population dynamics forecasts, including changes in ecosystem function and biodiversity loss, are rarely single-species issues (Lindenmayer et al. 2012, Greenville et al. 2016). In addition, because species differ in their niche requirements, ecosystems containing multiple species of interest may require managers to balance competing needs not only between human and ecosystem requirements, but also among different species (e.g., Romañach et al. 2022). Finally, species population dynamics are known to be related to one another due to both direct interactions between species (e.g., competition) and because species respond to shared environmental drivers (Volterra 1931, Warton et al. 2015, Ovaskainen et al. 2017). These associations between the dynamics of different species has resulted in extensive research into multivariate population dynamics models (Ives et al. 2003, Ward et al. 2010, Bunin 2017, Ward et al. 2022, Paniw et al. 2023). Leveraging these associations could potentially result in more accurate forecasts from multispecies models and better-informed scenario planning, including approaches to predicting the impacts of species extinctions or the potential spread of invasive species (Ibáñez et al. 2009).

However, despite the potential advantages of multispecies dynamic models, their implementation is still rare in ecological applications in general and in population forecasting specifically. A recent review on near-term ecological forecasts found that only XX% used multispecies models to generate and evaluate forecasts (Lewis et al. 2022), in line with an earlier review of population dynamics models for informing marine reserve design which found that only 1 of 34 studies considered multi-species dynamics (Gerber et al. 2003). One area of ecological modelling that has embraced multi-species approaches is Joint Species Distribution Models (JSDMs), which leverage spatial patterns to predict the distribution of species in space and time (Clark et al. 2016, Thorson et al. 2016, Norberg et al. 2019, Tobler et al. 2019, Powell-Romero et al. 2023). While many of these models only consider spatial data, recent advances have included time-series structures in the form of multivariate autoregressive components (Ovaskainen et al. 2017, Abrego et al. 2021).

The rarity of multispecies population dynamic forecasting is likely due in part to the increased computational complexity and statistical knowledge needed to formulate multivariate population dynamic models that incorporate real world complexities in ecological data (Karp et al. 2023). Forecasting the abundances of multiple species is particularly difficult, for several reasons. Many biological and physiological processes influence population dynamics (Quinn 2003, Hampton et al. 2013), and species often exhibit complex responses to external drivers (including non-linear responses and lags; Cárdenas et al. 2021, Karunarathna et al. 2024). Moreover, temporal autocorrelation is often prevalent in abundance time series data (due to population processes; Ives et al. 2010), which can be difficult to address in ecological models. Finally, because monitoring wildlife is challenging, data complexities (e.g., irregular sampling intervals, observation errors, missing samples, and overdispersed discrete counts with meaningful lower and/or upper bounds) bring additional challenges into an already complicated modelling environment (Clark and Wells 2023). In combination, these issues often make population time-series data unsuitable for traditional modelling approaches such as regression or simple time series models. Managers may also have differing needs for forecasts, ranging from predicting the most accurate near-term population sizes to exploring potential responses to differing management scenarios (Clark et al. 2001, Lindenmayer et al. 2012, Moustahfid et al. 2021, Lewis et al. 2023).

While forecasting multispecies population dynamics is challenging, recent work suggests that it has the potential to provide valuable insights for forecast applications. Incorporating other species – either by including lagged observations of other species as predictors in single-species models (Abrego et al. 2021, Daugaard et al. 2022) or by building temporal JSDMs with multispecies autoregressive terms (Hampton et al. 2013, Ovaskainen et al. 2017) can result in improved predictions compared to single-species models. Experimentally induced changes in the abundance of competitors also show that altered species interactions can impact the accuracy of single-species forecasts (Dumandan et al. 2024). While suggestive that multi-species modelling is important, none of these existing approaches incorporates the full suite of important real world complexities – data issues, non-linear responses to environmental drivers, temporal autocorrelation – that are necessary for many ecological forecast applications. Consequently, validation of multispecies forecasts, and comparisons against forecasts from simpler single species models, are rare and have generally been limited to either in-sample predictive measures (Sandal et al. 2022) or one-step ahead correlation measures (Ovaskainen et al. 2017, Abrego et al. 2021). Most forecast applications typically require predicting multiple time steps into the future to assess near-term management needs or responses to likely future scenarios (i.e. loss of important species, shifts in important drivers). The combination of a limited exploration of the utility of multispecies models for ecological time series applications and the need to incorporate more complex modeling structures constitutes a major gap in our ability to tackle realistic forecasting applications.

Here we evaluate whether models that incorporate multi-species relationships can improve near-term population forecasts using data from a long-term ecological monitoring study where there is evidence of both direct biotic interactions between species (Heske et al. 1994, Ernest and Brown 2001, Lima et al. 2008, Bledsoe and Ernest 2019, Christensen et al. 2019a) and shared responses to environmental factors (Christensen et al. 2018). Using the framework of Dynamic Generalized Additive Models developed by Clark and Wells (2023), we build a series of models that learn species' shared environmental responses and temporal interactions to make inference about environmental and biotic factors that relate to community dynamics. Our models highlight how several key challenges can be tackled when modelling the dynamics of multiple species, including how to estimate environmental effects that change nonlinearly over increasing lags, how to capture temporal autocorrelation, and how to estimate lagged temporal dependencies among species. We then test whether the incorporation of these biotic dependence structures improves forecasts over multiple near-term timescales (up to 12 months) using penalized in-sample performance criteria and out-of-sample forecast metrics. We also demonstrate how these models can be used to perform perturbation experiments for assessing community responses to shifts in key species abundances and to changes in environmental drivers with shared species responses. Because these multi-species dynamic models integrate both species interactions and complex environmental dependencies, we show that they can provide a deeper understanding of ecological dynamics while generating more accurate forecasts and predictions for scenario planning. These models are broadly applicable to time-series data, providing a versatile tool for conducting time-series based forecasting to meet the wide-ranging needs of both basic and applied research.

MATERIALS AND METHODS

We first describe the study system to outline why it is suitable for testing whether multi-species models lead to better ecological forecasts compared to single-species models. Second, we describe our full dynamic model, from which we can make inferences about the processes that drive community dynamics. Third, we describe how we compare this model to simpler models in an iterative forecasting exercise to ask whether models that include multi-species dependencies (a) improve in-sample fits to the observed data and (b) provide better out-of-sample near-term predictions.

Rodent capture data

Our data come from the Portal Project, a long-term monitoring study of a desert rodent community (Brown 1998, Ernest et al. 2020) that has been undergoing active forecasting since 2016 (White et al. 2019). The Portal Project is based in the Chihuahuan Desert near Portal, Arizona. The sampling design includes 24 experimental plots (50m x 50m), each containing a grid of 49 baited traps (Brown 1998, Ernest et al. 2020). The design uses three experimental treatments. In control plots (N = 10), holes in the fence are large enough to allow free access for all rodents. Full rodent removal plots (N = 6) have fences with no holes. Kangaroo rat exclosures (N = 8) have fences with holes to allow passage of all rodents except kangaroo rats (*Dipodomys* genus). Investigators close holes during trapping to ensure all captured rodents are residents. Trapping follows the lunar monthly cycle, but weather and other disruptions result in missing observations (~5% on average; Dumandan et al. 2023).

The Portal dataset exhibits many of the complexities that confront population forecasting. These include observation errors due to imperfect detection, missing samples due to weather or other issues (e.g., global pandemics), and over dispersed discrete counts for many species (20 rodent species) that include large numbers of zeros and upper bounds set by the number of traps. Environmental drivers, including temperature and measures of primary production, exhibit lagged and nonlinear impacts on rodent breeding, activity rates, and population dynamics (Brown and Ernest 2002). Moreover, the rodent species at Portal are known to compete for resources in complex ways, and these biotic interactions are postulated to have important consequences for variation in population dynamics. In other words, the Portal Project exhibits all the complexities that make the ecological forecasting of species populations particularly difficult, making it an ideal real-world test case for exploring whether multi-species models can provide better near-term predictions than single species models.

Open-source software exists to access the Portal Project data (Christensen et al. 2019b, Simonis et al. 2022). We used the *portalr package* to extract trapping records from the Portal data (version 3.134.0; downloaded October 2022; https://doi.org/10.5281/zenodo.7255488). Our study focused on rodent captures from the long-term control plots for the period December 1996 – August 2022. The data has records for 20 rodent species, but some are rarely captured. We excluded species if they were observed in < 10% of trapping sessions. We did this to focus inferences on species with the most influence on community dynamics. Each observation was a vector of total captures on long-term control plots for the nine remaining species (Figure 1).

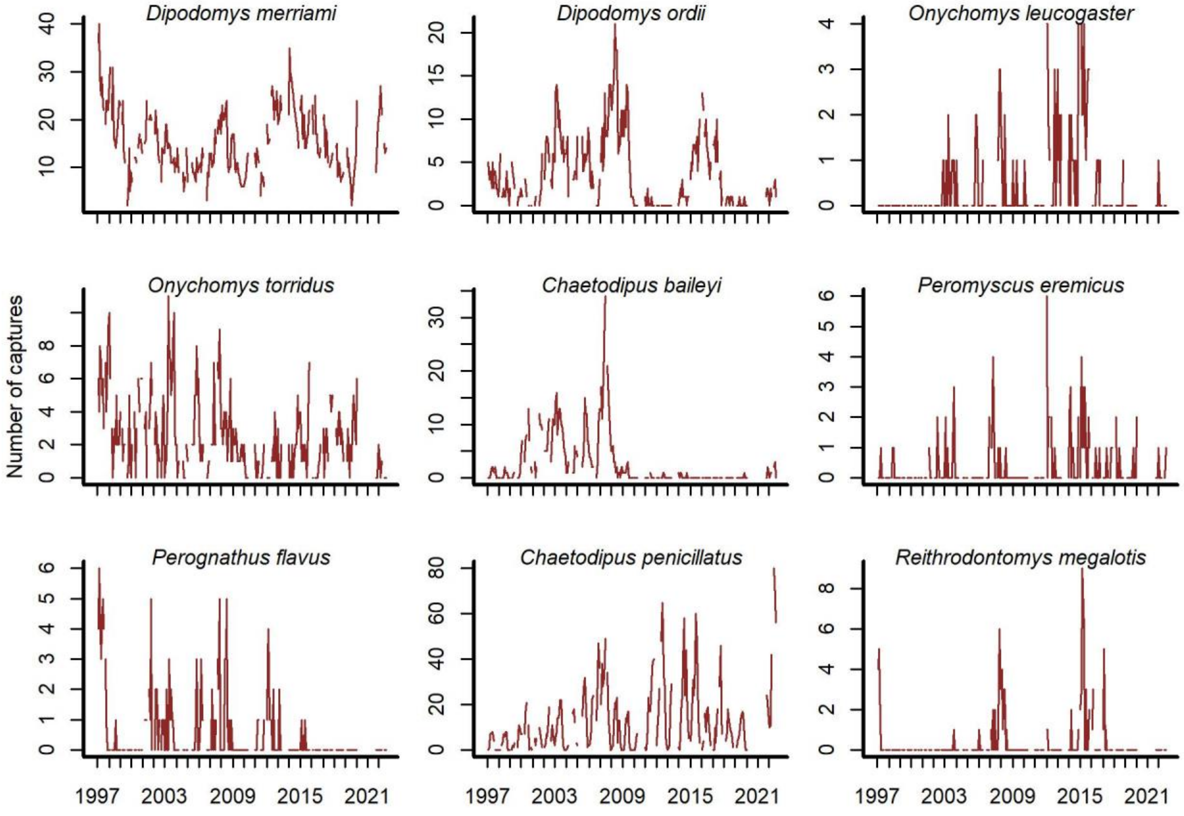


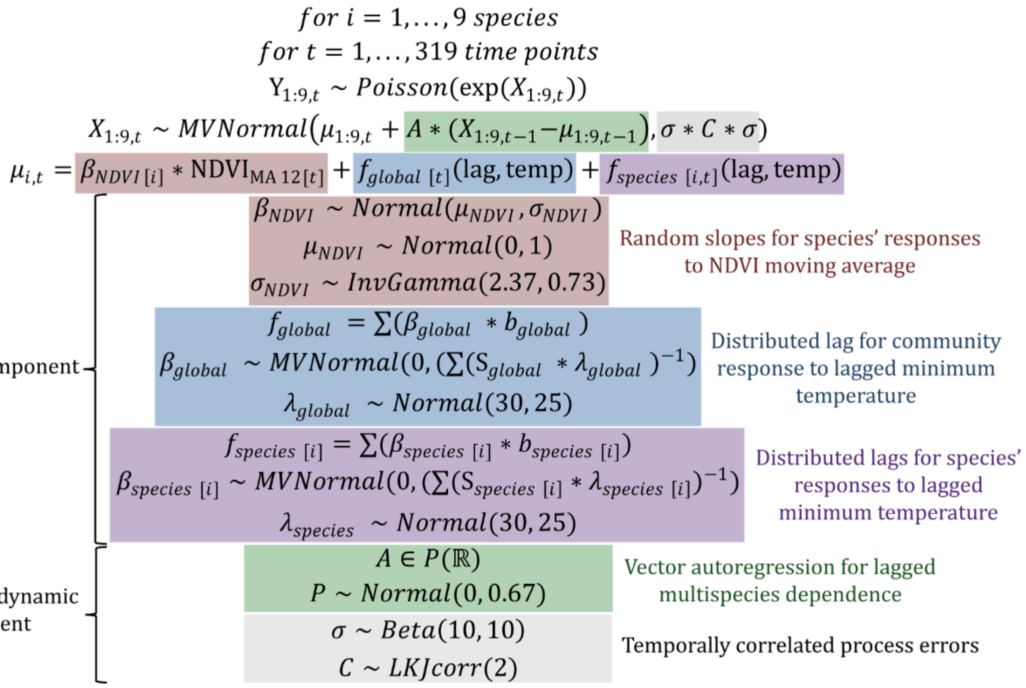
Figure 1: Rodent capture data from the Portal Project for the period December 1996 to August 2022. Counts are total captures across long-term control plots. Blanks are missing values.

Covariate measurements

Rodent populations at Portal, and the associated number of captures recorded during sampling, depend on environmental conditions that reflect resource availability and seasonal breeding signals. We therefore modelled species' responses to environmental variation using minimum temperature and the Normalized Difference Vegetation Index (NDVI) as covariates. Hourly air temperature (°C) is recorded by an automated weather station, while Landsat images are used to calculate NDVI (accessed from the US Geological Service Earth Resources Observation and Science Center; https://www.usgs.gov/centers/eros). Measurements for both covariates were converted to monthly averages. We extracted covariate data from one year before the start of captures (from January 1995) so we could calculate lagged and moving average versions. See Ernest et al. (2020) for details.

Model description

There were several aspects of the data we needed to consider when designing our model. Total rodent captures showed both short- and long-term fluctuations (Supplementary Figure S1). Captures for individual species also undulated over multi-annual cycles and were positively autocorrelated at lags up to 20 months (Supplementary Figures S2 and S3). To test whether multi-species information improves model performance, we needed to model these dynamics using a multivariate dependence structure. Second, we needed to leverage community information to estimate each species' time-delayed response to variation in vegetation and temperature. Because species' responses to environmental change in this system are expected to be nonlinear (Brown and Ernest 2002), we used splines to model these responses. Rodent captures were modelled as $Poisson$ observations of a latent state model that was composed of a hierarchical GAM component (to capture shared environmental responses) and a multivariate dynamic vector autoregressive component to capture multispecies dependence. The full description for this model, which we abbreviate to ***GAM-VAR***, is shown in Figure 2.



$$
\begin{aligned}
&for\ i = 1, \dots, 9\ species \\
&for\ t = 1, \dots, 319\ time\ points \\
&\mathrm{Y}_{1:9,t} \sim Poisson(\exp(X_{1:9,t})) \\
&X_{1:9,t} \sim MVNormal(\mu_{1:9,t} + A * (X_{1:9,t-1} - \mu_{1:9,t-1}), \sigma * C * \sigma) \\
&\mu_{i,t} = \beta_{NDVI\,[i]} * \mathrm{NDVI}_{\mathrm{MA\,12}[t]} + f_{global\ [t]}(\mathrm{lag}, \mathrm{temp}) + f_{species\ [i,t]}(\mathrm{lag}, \mathrm{temp})
\end{aligned}
$$

$$
\begin{aligned}
\beta_{NDVI} &\sim Normal(\mu_{NDVI}, \sigma_{NDVI}) \\
\mu_{NDVI} &\sim Normal(0, 1) \\
\sigma_{NDVI} &\sim InvGamma(2.37, 0.73)
\end{aligned}
$$

$$
\begin{aligned}
f_{global} &= \sum(\beta_{global} * b_{global}) \\
\beta_{global} &\sim MVNormal(0, (\sum(\mathrm{S}_{global} * \lambda_{global})^{-1}) \\
\lambda_{global} &\sim Normal(30, 25)
\end{aligned}
$$

$$
\begin{aligned}
f_{species\ [i]} &= \sum(\beta_{species\ [i]} * b_{species\ [i]}) \\
\beta_{species\ [i]} &\sim MVNormal(0, (\sum(\mathrm{S}_{species\ [i]} * \lambda_{species\ [i]})^{-1}) \\
\lambda_{species} &\sim Normal(30, 25)
\end{aligned}
$$

$$
\begin{aligned}
A &\in P(\mathbb{R}) \\
P &\sim Normal(0, 0.67) \\
\sigma &\sim Beta(10, 10) \\
C &\sim LKJcorr(2)
\end{aligned}
$$

Figure 2: Model definition and priors for the ***GAM-VAR*** model. Coloured boxes highlight the five main components of the latent state model (*X*).

The GAM component of the model consisted of hierarchical NDVI and minimum temperature effects. The structural forms of these functions were informed by theory and exploration of covariate time series (shown in Supplementary Figures S4-5). We used a 12-month moving average of NDVI ($NDVI_{MA12}$) because we expected rodents to respond gradually to vegetation change. Our model assumed linear effects of $NDVI_{MA12}$, equivalent to a hierarchical slopes model. The partial pooling properties of this model allowed us to regularize weakly informed slopes toward a community average. Responses to temperature were estimated using a hierarchical distributed lag model in which nonlinear effects of minimum temperature varied smoothly with increasing lag. These effects were constructed as tensor products of four cubic basis functions for lag and three thin plate basis functions for minimum temperature. To allow our model to capitalize on multi-species learning, we included a shared community-level response $f_{global}(Mintemp, lag)$ and species-level deviations $f_{species[i]}(Mintemp, lag)$. The sum of these effects allowed each species to show a different temperature response from the wider community, but only if there was information in the data to support such a deviation. We used lags of up to six months in the past.

A vector autoregression (VAR) of order 1 captured lagged multispecies dependence, where A was a 9 x 9 matrix of autoregressive coefficients. Diagonal entries of A described density-dependence, or the effect of a species' dynamic process (at time t) on its own lagged values (at $t - 1$). Off-diagonals represented cross-dependencies that could provide useful biological insights into interspecific interactions. For example, the entry in A**[2,3]** described the effect of species 3's dynamic state at time $t-1$ on the current state estimate for species 2 (at time t). To encourage stability and prevent forecast variance from increasing indefinitely, we enforced stationarity following methods described in Heaps (2023). Briefly, a multistep process was used to map the constrained A matrix to unconstrained partial autocorrelations P. Process errors were allowed to be contemporaneously dependent to capture any unmodelled correlations. Priors for all model components are shown in Figure 2 and described in detail in the accompanying R code.

Evaluating whether multi-species dependencies improve prediction performance

We formally tested whether learning from multiple species improved our model's predictions using prediction-based model comparisons. To do so, we estimated a series of benchmark models that acted as natural simplifications of the ***GAM-VAR*** by eliminating multi-species components in a stepwise manner. The first benchmark model used the same HGAM linear predictor as the ***GAM-VAR*** but replaced the multi-species VAR(1) dynamics with an AR(1) process. This model (called ***GAM-AR*** in subsequent sections) eliminated the covariances and temporal cross-dependencies among species' latent states, allowing us to ask whether the multivariate dynamic component was supported for improving model fit. Next, we further simplified the ***GAM-AR*** by removing the hierarchical environmental response functions from the linear predictor. This forced the model to learn environmental responses for each species without using information from other species in the data (***GAM-AR no pooling***). The final benchmark, referred to as ***AR***, also used independent AR(1) states but removed the GAM component entirely. Because this model only learned from past observations, comparisons against it helped us understand how covariates impacted predictions and inferences. Each candidate model was trained on all observations (through August 2022, N = 319 timepoints). Models were then compared using Pareto-smoothed importance sampling leave-one-out cross-validation (PSIS-LOO), a method that reweights posterior draws to estimate leave-one-out pointwise prediction accuracy using Estimated Log Predictive Density (ELPD) values (Vehtari et al. 2017).

To adequately evaluate competing forecast models, it is also necessary to perform out-of-sample validation (Harris et al. 2018, Clark et al. 2022, Lewis et al. 2022). This is particularly important because LOO-CV is designed to ask how models would generalize to new observations within the training window. This metric does not evaluate a time series model’s ability to forecast, as information from future timepoints is used to influence predictions for previous time points. To evaluate forecasts in a way that respected the temporal nature of our forecasting exercise, we used exact leave-future-out cross-validation in an iterative expanding window framework. Models were re-trained on the first 273 time points (~22 years), with the subsequent 12 time points (through November 2019; selected to avoid a large sampling gap due to the COVID-19 pandemic) used to evaluate forecasts. This allowed us to gauge how models might perform in a forecast scenario, but it only provided a single comparison. To further scrutinize models, we retrained models on the first 75, 115, 154, 194, and 233 observations, and evaluated the subsequent 12 observations in each cross-validation fold. All forecast comparisons used an evenly weighted combination of two proper multivariate scoring rules. We chose the variogram score, which penalizes distributions that do not adequately capture correlations in test observations, and the energy score, which ignores correlations but penalizes forecasts if they are not well-calibrated (Scheuerer and Hamill 2015).

Estimation

We estimated posterior distributions with Hamiltonian Monte Carlo in Stan (Carpenter et al. 2017, Stan Development Team 2022), specifically the *cmdstanr* interface (Gabry and Češnovar 2021). Stan’s algorithms provide state-of-the-art diagnostics for probabilistic models (Betancourt 2017). For example, Hamiltonian Markov chains diverged when attempting to estimate minimum temperature deviations for some species in the ***GAM-VAR***. Our data were not informative enough to learn how, or even if, these species responded to temperature change in ways that differed from the community response. Stan’s diagnostics guided us to a model that could be reliably estimated, which included deviation functions for the four most frequently captured species (*D. ordii*, *D. merriami*, *Onychomys torridus* and *C. penicillatus*). Posteriors were processed in R 4.3.1 (R Core Team 2023) with the *mvgam* R package (Clark and Wells 2023). Traceplots, rank normalized split-$\hat{R}$ and effective sample sizes interrogated convergence of four parallel chains. Each chain was run for 500 warmup and 1600 sampling iterations. R code to replicate all analyses and produce Figures is included in the Supplementary materials and will be permanently archived on Zenodo on acceptance.

RESULTS

Modeling relationships among species improves prediction performance

Our data included total captures for nine rodent species over 319 time points. All models showed adequate convergence and posterior exploration, and randomized quantile residuals showed no obvious evidence of unmodelled temporal or systematic variation (Supplementary Figures S6 – S7). However, in-sample performances differed, with models that leveraged multi-species information producing higher ELPD scores compared to simpler models (Table 1). This was the case for all stepwise comparisons apart from one: although the ***GAM-AR***, which used partial pooling to learn species' environmental responses, was favoured over the simpler ***GAM-AR no pooling***, overlapping ELPD standard errors suggested there was still large uncertainty about the magnitude of this difference (Table 1).

Table 1: Approximate Pareto-smoothed importance sampling leave-one-out cross-validation (PSIS-LOO) was used to compute the Estimated Log Predictive Density (ELPD) of competing models. A higher ELPD indicates a model is expected to generalize better to new data within the training window.

Model	ELPD difference	SE of ELPD difference
GAM-VAR	0.0	0.0

GAM-AR	-15.5	8.3
GAM-AR no pooling	-22.1	7.0
AR	-74.3	12.6

We also found that forecast performance differed among models, with more complex multi-species models again tending to score higher for forecast performance than simpler models. Forecasts from the multi-species ***GAM-VAR*** were the most accurate when considering all validation points in aggregate and for 4 / 6 cross-validation folds (Figure 3; Supplementary Figure S8). The ***GAM-AR*** and ***GAM-AR no pooling*** models gave similar predictions and effectively tied for second in forecast performance, giving the most accurate forecasts in 2 / 6 cross-validation folds (Figure 3). The simplest ***AR*** model gave the worst forecasts.

The multi-species ***GAM-VAR*** model estimated large, positive autoregressive coefficients for most species (diagonal entries in Supplementary Figure S9). It also relied strongly on information from multiple species by estimating many non-zero cross-dependence effects (off-diagonal entries in Supplementary Figure S9) and process error correlations (Supplementary Figure S10), which provided structure that the model leveraged to accurately simulate historical dynamics. The model recovered multiple notable transitions observed in the time-series including a major shift in community composition around 2000 following the establishment of Bailey's pocket mouse *C. baileyi*, and a second restructuring that happened following a drought in 2008 – 09 (Supplementary Figure S11). It was these multispecies effects that enabled the ***GAM-VAR*** to produce more accurate forecasts compared to the benchmarks. For example, Ord’s kangaroo rat (*D. ordii*) and silky pocket mouse (*P. flavus*) had negative cross-dependencies in the ***GAM-VAR***, providing structure that the model used to make predictions (Figure 4). The benchmarks, which ignored this structure, produced smoother, less synchronous trends and wider uncertainties (Supplementary Figure S12). In the following sections, we use simulations to briefly interpret each of the multi-species effects that allowed the ***GAM-VAR*** to outperform simpler models.

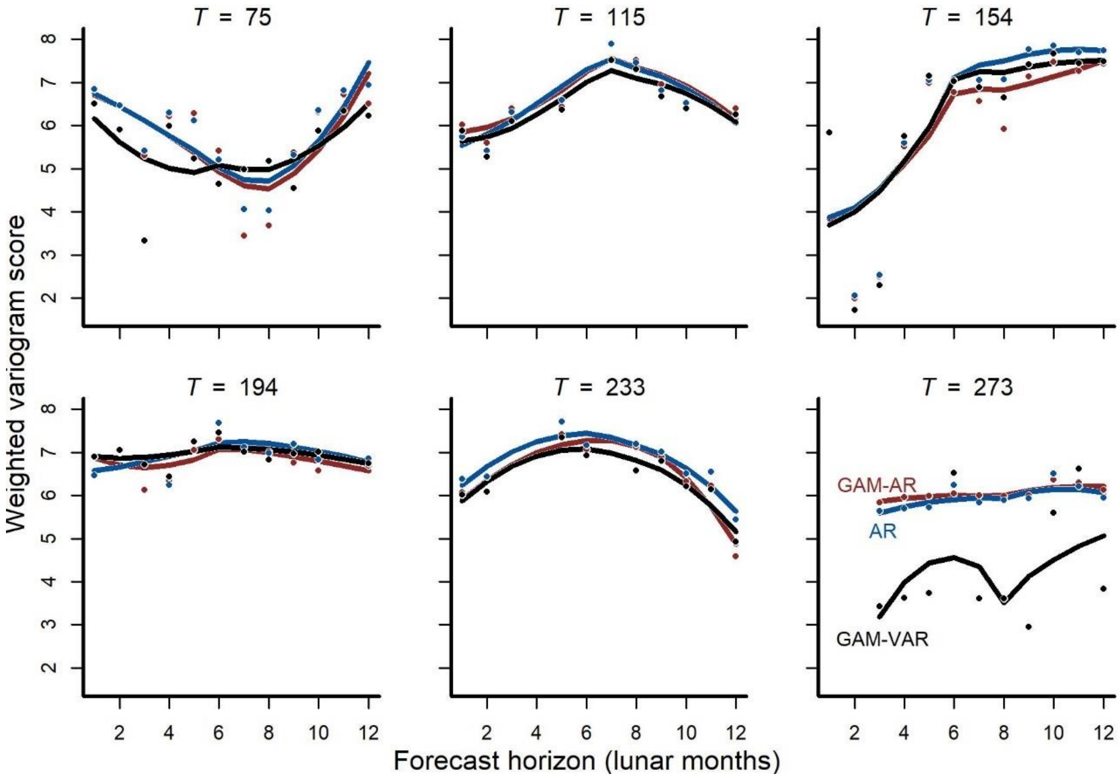
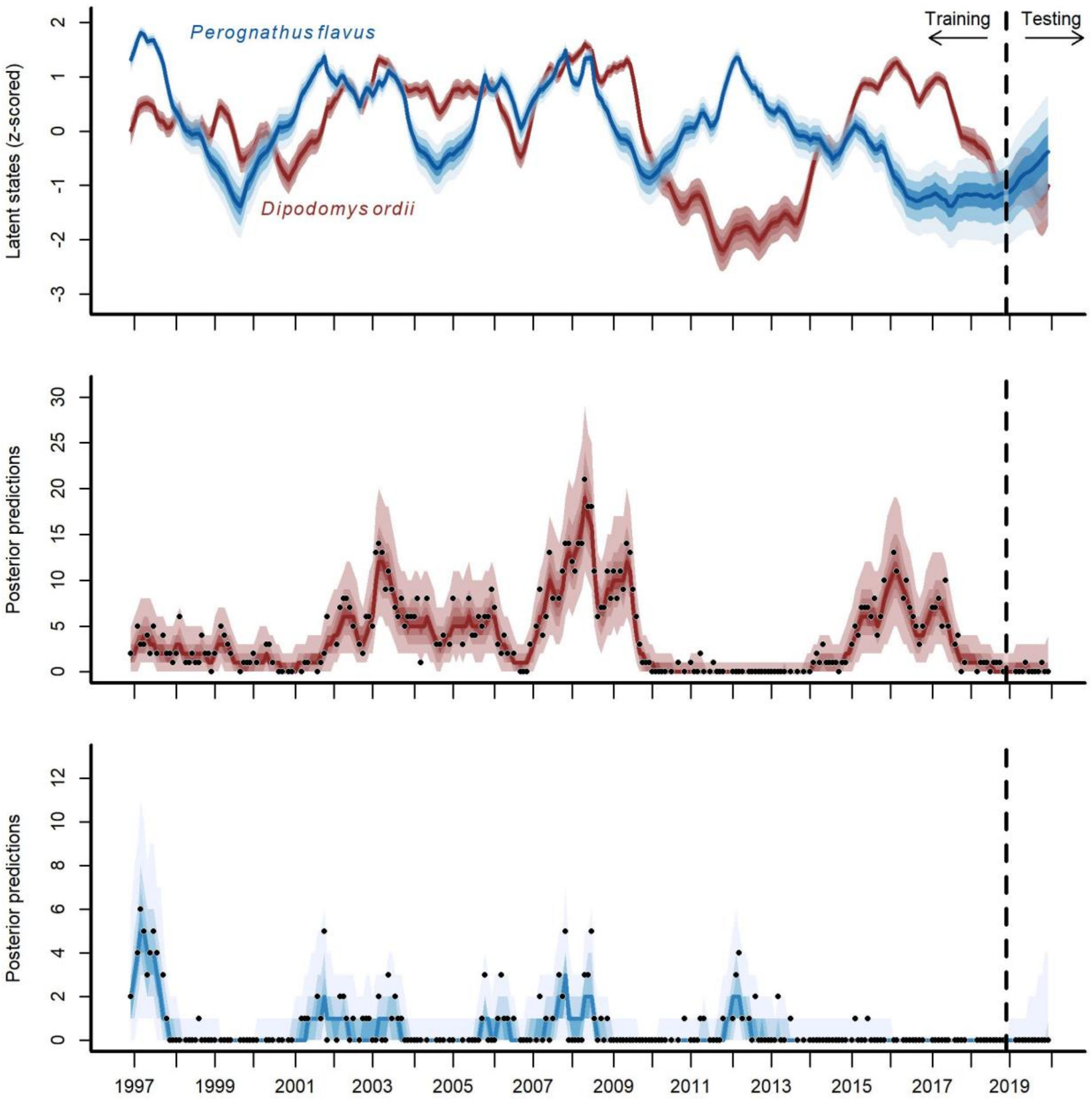


Figure 3: Cross-validation forecast performances for three of the competing models (we do not show metrics for the ***GAM-AR no pooling*** model as they were not clearly distinguishable from the ***GAM-AR*** metrics). Y-axis shows the log of the weighted variogram score, a scoring rule that penalizes multivariate forecasts if they are not well calibrated and do not capture inter-series correlations in observed data (lower scores are preferred). 12-step ahead predictions were evaluated over a sequence of six evenly spaced origins. Points show individual forecast scores, with lower scores indicating a better forecast. Lines show Loess smoothed trend lines. Missing points indicate that sampling did not occur at the time point for that horizon.



358

359 **Figure 4:** Posterior latent state estimates (top panel) and posterior predictions (bottom two panels) from the ***GAM-VAR*** model for Ord's kangaroo rat (*Dipodomys ordii*; in red) and silky pocket mouse (*Perognathus flavus*; in blue) for the training and testing periods (demarked by the vertical dashed line). State estimates were scaled to unit variance for comparisons. Ribbon shading shows posterior empirical quantiles (90th, 60th, 40th and 20th). Dark lines show posterior medians. Points show observations.

365

366 **Modeling relationships among species provides unique insights into community dynamics**

Our cross-validation metrics strongly favoured the ***GAM-VAR*** and suggested that the multivariate dynamic component was a particularly important driver of increased performance. Estimates of process error were larger for the benchmarks than the ***GAM-VAR*** for nearly all species (Supplementary Figure S13), suggesting this model used additional information from multi-species cross-dependencies to produce better predictions. But interpreting this cross-dependence is difficult because VAR effects provide only a marginal view into the complex network of conditional associations. We used impulse response functions (Lütkepohl 1990) to better understand the model. These functions involve simulating an 'impulse' in captures for one species and then evaluating how predicted captures for other species changed over the next six months (Figure 5). Following a simulated impulse of three extra captures for Merriam's kangaroo rat (*D. merriami*), the model expected some initial increases (due to the correlated process errors) followed by declines in captures for most of the other species (Figure 5). The shapes of these declines varied by species. Captures for the two pocket mouse species (*C. baileyi* and *C. penicillatus*) showed more immediate declines, while the two grasshopper mouse species (*O. leucogaster* and *O. torridus*) declined more gradually (Figure 5). In contrast, the other kangaroo rat species (*D. ordii*) was expected to increase following a *D. merriami* pulse (Figure 5). Different effects were expected when changing the focal species (Supplementary Figure S14)

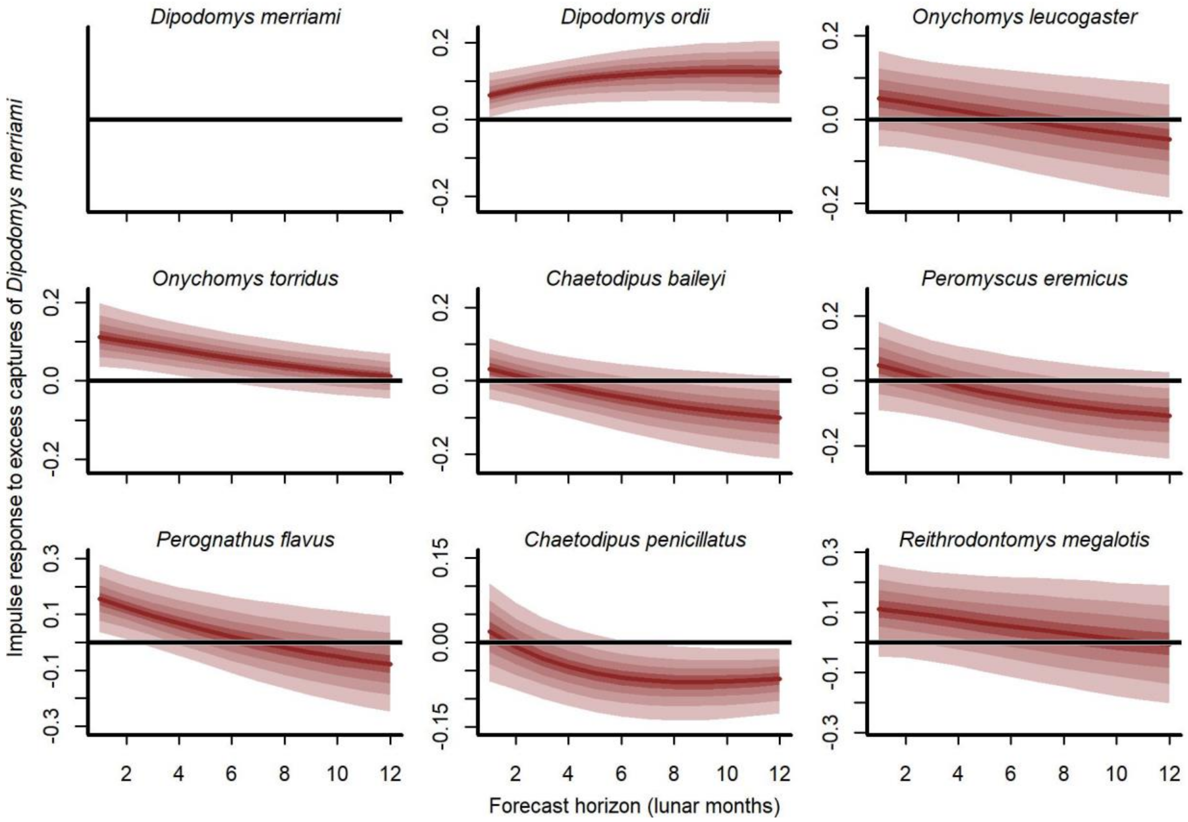


Figure 5: Expected responses to a simulated pulse in captures of Merriam's kangaroo rat (*D. merriami*). Ribbon plots show how mean captures (μ, on the log scale) are expected to change over the next six months if three additional *D. merriami* individuals are captured. Ribbons show posterior empirical quantiles (90th, 60th, 40th and 20th). Dark red lines show posterior medians.

Positive NDVI associations and hierarchical minimum temperature effects

We found broad support for positive $NDVI_{MA12}$ associations (Figure 6). Conditional simulations, which asked how rodents might respond if moved from a relatively dry/brown vegetation state to a relatively moist/green vegetation state, gave higher probability to increased captures in the moist/green scenario for all species. But uncertainties about this effect varied. Greatest increases were expected for Ord's kangaroo rat (*D. ordii*), Western harvest mouse (*R. megalotis*) and cactus mouse (*Peromyscus eremicus*). The model was less confident about the direction of effect for Northern grasshopper mouse (*O. leucogaster*) and for one of the most dominant species in the study, Meriam’s kangaroo rat (*D. merriami*). For these species, the model expected a increases in ~70% of simulations and decreases in ~30% (Figure 6). While primary conclusions were generally similar when using the ***GAM-AR no pooling*** model, which did not leverage multi-species learning, the estimates of these contrasts were much more variable (Figure S15).

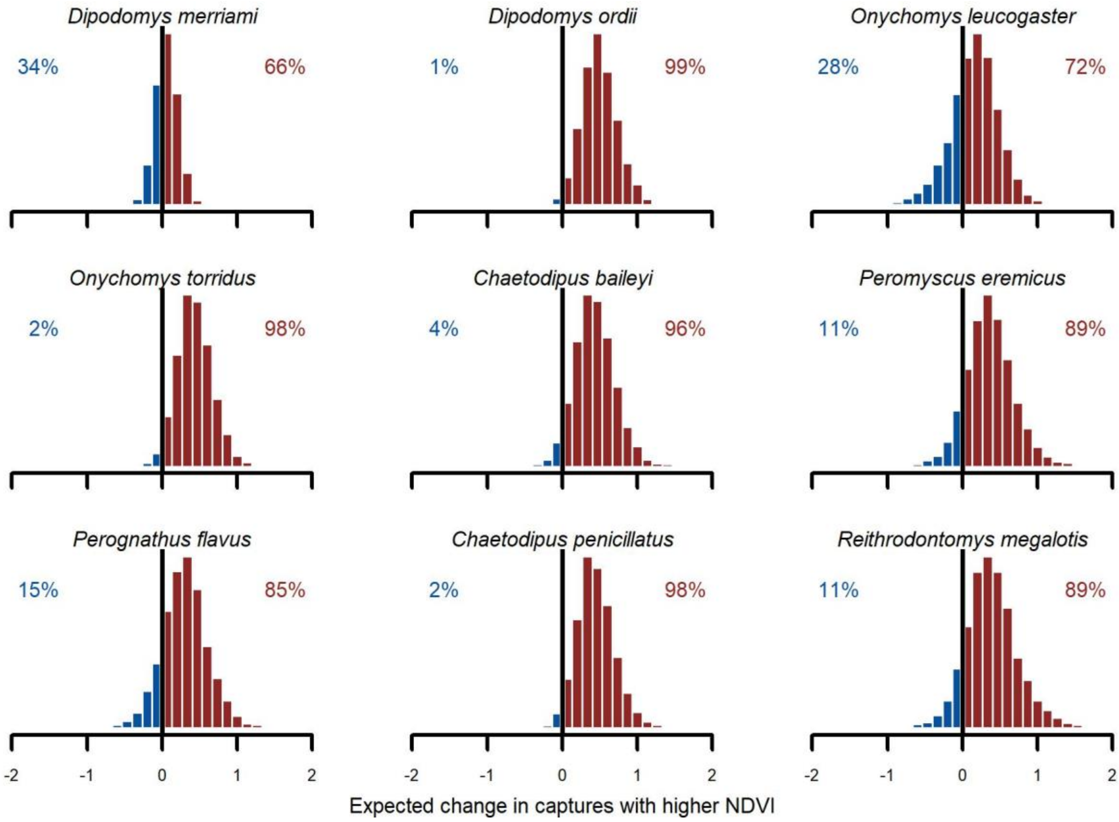


Figure 6: Posterior NDVI contrasts from the hierarchical slopes component of the ***GAM-VAR*** model. Histograms illustrate how much the expected number of captures, $exp(\mu)$, would change if the z-scored NDVI 12-month moving average ($NDVI_{MA12}$) changed from a relatively low value (-0.50) to a relatively high value (0.50). Numbers in each plot indicate the proportion of probability mass at or below zero (in blue) vs above zero (in red).

Interpreting minimum temperature distributed lag effects also required simulation. We visualized 1,000 simulated functions for each species using temperatures from 1997 (Figure S15). There was large uncertainty in function shapes for all species except desert pocket mouse (*C. penicillatus*). Captures for this species were expected to increase from May to October and decrease sharply in winter. For seven of the other eight species, the model generally expected more captures in spring (March – May) and fewer in late summer / autumn (July – October). But the shapes of these responses varied. The two kangaroo rats (*D. merriami* and *D. ordii*) had smoother shapes that decreased gradually from mid-summer to mid-winter. But the model expected *D. ordii* captures to peak slightly later (May as opposed to March for *D. merriami*). The Southern grasshopper mouse (*O. torridus*) was the only species that was expected to show higher captures in late autumn / early winter (Figure S16). The five species that relied solely on the global function (*O. leucogaster*, *C. baileyi*, *P. eremicus*, *P. flavus* and *R. megalotis*) were expected to show tighter spring peaks and autumn troughs. When simulating from the ***GAM-AR no pooling*** model, the lack of multi-species learning was immediately obvious. There was not enough information to learn nonlinear distributed lag functions for these five species, with the model instead estimating flat functions centred on zero for all five species (Figure S17).

DISCUSSION

Understanding and predicting change in species abundances requires models that capture realistic biotic structure and address data complexities to produce near-term ecological forecasts (Hampton et al. 2013, Holmes et al. 2014). Our results show that incorporating relationships between species to estimate their lagged dependence, and to learn their potentially non-linear associations with environmental drivers, yields more accurate in-sample and out-of-sample predictions. In addition to improved forecasts, incorporating these multi-species complexities provides opportunities for interpretation that are not possible with simpler models. Our dynamic VAR process uncovered biotic structure representing a cascading network of relationships within the community. Captures for all species increased with higher NDVI and responded nonlinearly to temperature change, but the shapes and magnitudes of these responses differed across species. Our results show that models that describe biological complexity, both through nonlinear covariate functions and multi-species dependence, are useful both for generating more accurate near-term forecasts and for asking targeted questions about drivers of ecological change (Ives et al. 2003, Greenville et al. 2016, Ovaskainen et al. 2017, Pedersen et al. 2019).

Leveraging relationships between species for ecological forecasting

Our analyses show why models that target multi-species effects in both their environmental responses and their direct biotic interactions should be strongly considered when studying community dynamics. The ***GAM-VAR***'s process variance estimates were smaller than those from the benchmarks because it used more information from the data. By learning about the relationships between species the model could better capture both shared responses to environmental factors (e.g., a wet year in the desert is good for most species) and direct temporal effects (e.g., competition for seeds). These relationships between species can allow forecasts for less commonly observed species to borrow strength from more common species, yielding better overall predictions for future population abundances. But like other multivariate autoregressive models (Ives et al. 2003, Holmes et al. 2014, Hannaford et al. 2023) the VAR parameters of the ***GAM-VAR*** should not be interpreted as a species interaction matrix, because these relationships can result from multiple sources (i.e., shared environmental responses and direct interactions). While the parameters are not interpretable as direct interactions, this approach does make it possible to gain a more detailed understanding of population dynamics. Conducting simulations from this type of model allows exploring which species have the strongest cascading effects, what changes might we expect if management increases or decreases abundance for target species, and how these effects relate to regime transitions. This approach also makes it possible to partition variance among observation error, environmental responses, and multispecies dependence to guide future efforts to improve ecological forecasting. In our study, forecasts were dominated by uncertainty in the dynamic process model, but using a vector autoregressive process allowed us to dissect this uncertainty in meaningful ways (Lütkepohl 1990, Ives et al. 2003). Simulated responses to sudden impulses in captures were often delayed and nonlinear. Despite the restriction to a VAR of lag of one month, these responses resulted in cascading changes that lasted up to six months.

Learning hierarchical nonlinear effects from community data

Our model captured linear, nonlinear, and lagged responses to environmental and climatic covariates that were informed by data from all species at once. We found positive linear associations between capture rates and a 12-month moving average of NDVI. This response was expected because the rodents at Portal depend on plants for food and other resources (Brown and Ernest 2002, Ernest et al. 2020) and NDVI measures vegetation greenness in the landscape. Within this overarching community pattern there were interesting patterns of variation in these responses among species. The strongest positive association was shown by Ord's kangaroo rat (*D. ordii*), a species that field evidence suggests consumes and harvests grasses (Kerley et al. 1997). In contrast, Merriam’s kangaroo rat (*D. merriami*) showed weaker associations with NDVI. This species has been predicted to increase in prevalence in the study region with more severe droughts, in part due to a preference for more open foraging habitat with less vegetation (Cárdenas et al. 2021).

Distributed lag functions determine the best combination of lags for environment covariates but are not commonly used in ecology (but see Ogle et al. 2015, Wells et al. 2016). Our study shows how these effects can be learned hierarchically and provides useful insights into delayed responses to temperature change for rodents at Portal. Most species showed higher captures when minimum temperatures were low 3 – 4 months prior, suggesting increases begin during mid to late spring when resources such as seeds become available. But others, such as Merriam’s kangaroo rat and Southern grasshopper mouse, showed increases during cooler months in autumn and winter.

Asynchronous phenology, where species show different reproductive timing, is sometimes expected in competitive communities (Carter and Rudolf 2022). Analysis of individual reproductive status in different biotic contexts suggests that some species shift their reproductive timing in the presence of strong competitors in the Portal system (Dumandan et al. 2023). Do these competitive forces play a role in seasonal capture variation over the long-term? Comparing temperature responses on control vs experimental plots would be one interesting way to tackle this question.

Interestingly, despite the relatively large number of observations our data contained for each species, estimates of environmental responses were still more precise and informative when using hierarchical models (which use partial pooling) as opposed to a no-pooling model that only considers species' effects in isolation. While hierarchical intercepts and slopes are commonly used in ecological models, there has been less emphasis on hierarchical nonlinear functions (but see Pedersen et al. 2019). Open access to new software that makes it easy to construct and estimate these types of functions, such as the *mvgam* R package that we used here (Clark and Wells 2023), should improve their uptake in ecological forecasting exercises.

But despite the power of hierarchical environmental effects to improve predictions, we cannot interpret environmental response estimates as directly causal for several reasons. First, we know NDVI is not a perfect measure of changes in seed production. Second, it is likely that changes to NDVI and minimum temperature are both related to other unmeasured environmental drivers that may also influence rodent abundance. Major storms, the El Niño Southern Oscillation and other factors that influence moisture levels can all influence temperature and vegetation change (Sun and Kafatos 2007). These other drivers could act as unmeasured confounds, biasing estimates in a causal inference framework (McElreath 2020).

Future directions

Two additional steps would be useful to fully assess the value of multi-species models for ecological forecasting, both in this system and more broadly as an ecological application. First, a more diverse suite of candidate models could be estimated to determine how forecasts could be combined into an ensemble to provide the best predictions in situations where prediction accuracy is the primary goal (Clark et al. 2022, Powell-Romero et al. 2023). This could be especially useful for detecting changes in the system. For example, ***GAM-VAR*** gave better forecasts in most cross-validation tests, but its performance was slightly worse than the simpler ***GAM-AR*** when the training window stopped just prior to a major restructuring of rodent abundances that was taking place in response to a drought. Second, determining which models are best for true forecasting requires evaluating forecasts in the presence of uncertainty in future covariate values. In this study we were hindcasting and therefore used the actual observed environmental measurements for the period reserved for model evaluation. Fortunately, the system is undergoing active forecasting involving a suite of simpler models and leveraging actual forecasts for environmental covariates (White et al. 2019, Simonis et al. 2022). A natural next step for this work is to compare the performance of the ***GAM-VAR*** model to simpler models both using hindcasting with observed covariates and when making true forecasts relying on predictions instead of observations for NDVI and minimum temperature.

The Portal Project also provides a unique opportunity to disentangle the combined influence of shared environmental responses and direct species interactions in driving observed relationships between species. The site includes a long-term experimental manipulation where kangaroo rats (*Dipodomys* species) are excluded from some plots. Recent work shows that single species forecasting models for *C. baileyi* do not transfer well between the control plots and this experimental manipulation, likely due to the different competitive environment experience in the absence of the behaviorially dominant kangaroo rats (Dumandan et al. 2023). Multi-species models like the GAM-VAR have the potential to transfer better in situations where one or more species are removed from the system by accurately predicting the response of the other species to this removal. Therefore, a key next step in evaluating the potential strengths of our models is to determine if they can more effectively transfer to make accurate predictions on the plots with the experimentally manipulated species composition. More broadly, we hope that the ability to estimate multi-species dependence and species-level variation in nonlinear environmental responses result in more accurate forecasts, inspire new questions, and lead to an improved understanding of the factors that govern ecological community dynamics.

ACKNOWLEDGEMENTS

We thank the many volunteers for their help during fieldwork to generate Portal data. This study was supported by an ARC DECRA Fellowship to N.J. Clark (DE210101439) and USDA National Institute of Food and Agriculture, Hatch Project FLA-WEC-005983 (Ernest) and FLA-WEC-005944 (White). The Portal Project has been funded nearly continuously since 1977 by the National Science Foundation, most recently by DEB-1929730 to S. K. M. Ernest and E.P. White. Development of portal software packages is supported by this NSF grant, NSF grant DEB-1622425 to S. K. M. Ernest, and the Gordon and Betty Moore Foundation's Data-Driven Discovery Initiative through Grant GBMF4563 to E. P. White.

SUPPORTING INFORMATION CAPTIONS

Figure S1: Total rodent captures from the Portal Project for the period December 1996 to August 2022. Counts represent total captures for nine species across long-term control plots, sampled at each cycle of the lunar moon. Blanks represent missing values.

Figure S2: Autocorrelation functions of rodent capture time series in the Portal Project. Dashed lines show values beyond which the autocorrelations are considered significantly different from zero.

Figure S3: Histograms of rodent capture time series in the Portal Project. Counts represent total captures across long-term control plots, sampled at each cycle of the lunar moon.

Figure S4: Seasonal and Trend decomposition using Loess smoothing (STL) applied to observed minimum temperature time series for the period December 1996 – August 2022. The top panel shows the raw time series. The middle plot shows the estimated long-term trend (calculated using a Loess regression to the de-seasoned time series). The bottom plot shows the time-varying estimate of seasonality (calculated using a Loess regression that smooths across years). STL components were estimated using the *msts()* function in the *forecast* R package (Hyndman and Khandakar 2008).

Figure S5: Top panel: observed Normalized Difference Vegetation Index (NDVI) time series for the period December 1996 – August 2022, with obvious seasonal fluctuations. Bottom panel: a 12-month moving average that represents smooth, gradual changes in NDVI at the study site.

Figure S6: Autocorrelation functions of randomized quantile residuals from the ***GAM-VAR*** model. Ribbon shading shows posterior empirical quantiles (90th, 60th, 40th and 20th). Dark red lines show posterior medians. Dashed lines show values beyond which the autocorrelations would be considered significantly different from zero in a Frequentist paradigm.

Figure S7: Normal quantile-quantile plots of randomized quantile residuals from the ***GAM-VAR*** model. Ribbon shading shows posterior empirical quantiles (90th, 60th, 40th and 20th). Dark lines show posterior medians.

Figure S8: Posterior predictions from the ***GAM-VAR*** model for the training and testing periods (demarked by the vertical dashed line). Latent state estimates were scaled to unit variance for comparisons. Ribbon shading shows posterior empirical quantiles (90th, 60th, 40th and 20th). Dark lines show posterior medians. Points show observations.

Figure S9: Posterior distributions of vector autoregressive coefficients (matrix A). Off-diagonals represent cross-dependencies. For example, the entry in $A[1,2]$ captures the effect of **DO**'s state at time $t-1$ on the current state estimate for **DM** (at time t). Diagonals (with grey shading) represent autoregressive coefficients (the effect of a species' state at time $t-1$on its own state at time t). Colours indicate the proportion of probability mass at or below zero (in blue) vs above zero (in red). **DO**, *Dipodomys merriami*; **DO**, *Dipodomys ordii*; **OL**, *Onychomys leucogaster*; **OT**, *Onychomys torridus*; **PB**, *Chaetodipus baileyi*; **PE**, *Peromyscus eremicus*; **PF**, *Perognathus flavus*; **PP**, *Chaetodipus penicillatus*; **RM**, *Reithrodontomys megalotis*.

Figure S10: Posterior distributions for process error correlations (matrix C). Colours indicate the proportion of probability mass at or below zero (in blue) vs above zero (in red). **DO**, *Dipodomys merriami*; **DO**, *Dipodomys ordii*; **OL**, *Onychomys leucogaster*; **OT**, *Onychomys torridus*; **PB**, *Chaetodipus baileyi*; **PE**, *Peromyscus eremicus*; **PF**, *Perognathus flavus*; **PP**, *Chaetodipus penicillatus*; **RM**, *Reithrodontomys megalotis*.

Figure S11: Simulated rodent communities. Using the ***GAM-VAR*** model's posterior predictive distribution, we simulated communities of 200 individuals at different timepoints to investigate how well the model captured known community transitions. Colours represent different species

Figure S12: Posterior trend estimates from three competing models for Ord's kangaroo rat (*Dipodomys ordii*; in red) and silky pocket mouse (*Perognathus flavus*; in blue). Trends were scaled to unit variance for comparisons. Ribbon shading shows posterior empirical quantiles (90th, 60th, 40th and 20th). Dark lines show posterior medians.

Figure S13: Posterior estimates of trend standard deviations from the three competing models. Estimates are the square root of diagonal parameters from the trend covariance matrix (Σ_{VAR}) for the ***GAM-VAR*** (black), ***GAM-AR*** (red) and ***AR*** (blue).

Figure S14: Expected responses to a pulse in captures of the desert pocket mouse (*Chaetodipus penicillatus*). Ribbon plots show how mean captures (μ, on the log scale) are expected to change over the next six months if three additional *C. penicillatus* individuals are captured. Ribbon shading shows posterior empirical quantiles (90th, 60th, 40th and 20th). Dark red lines show posterior medians.

Figure S15: Posterior NDVI contrasts from the independent slopes component of the ***GAM-AR no pooling*** model. Histograms illustrate how much the expected number of captures, $exp(\mu)$, would change if the z-scored NDVI 12-month moving average ($NDVI_{MA12}$) changed from a relatively low value (-0.50) to a relatively high value (0.50). Numbers in each plot indicate the proportion of probability mass at or below zero (in blue) vs above zero (in red).

Figure S16: Conditional distributed lag minimum temperature functions from the hierarchical smooth component of the ***GAM-VAR*** model, using temperatures observed in 1997. All other effects were ignored. Functions for *O. leucogaster*, *C. baileyi*, *P. eremicus*, *P. flavus* and *R. megalotis* were drawn solely from the global function. Functions for other species were the sum of the global function and a species-specific deviation function. Estimates were scaled to unit variance for comparisons. Ribbons show posterior empirical quantiles (90th, 60th, 40th and 20th). Dark red lines show posterior medians.

Figure S17: Conditional distributed lag minimum temperature functions from the independent smooth component of the ***GAM-AR no pooling*** model, using temperatures observed in 1997. All other effects were ignored. Functions for *O. leucogaster*, *C. baileyi*, *P. eremicus*, *P. flavus* and *R.*

megalotis were drawn solely from the global function. Functions for other species were the sum of the global function and a species-specific deviation function. Estimates were scaled to unit variance for comparisons. Ribbons show posterior empirical quantiles (90th, 60th, 40th and 20th). Dark red lines show posterior medians.

REFERENCES

Abrego, N., T. Roslin, T. Huotari, Y. Ji, N. M. Schmidt, J. Wang, D. W. Yu, and O. Ovaskainen. 2021. Accounting for species interactions is necessary for predicting how arctic arthropod communities respond to climate change. Ecography **44**:885-896.

Algar, A. C., H. M. Kharouba, E. R. Young, and J. T. Kerr. 2009. Predicting the future of species diversity: macroecological theory, climate change, and direct tests of alternative forecasting methods. Ecography **32**:22-33.

Averill, C., Z. R. Werbin, K. F. Atherton, J. M. Bhatnagar, and M. C. Dietze. 2021. Soil microbiome predictability increases with spatial and taxonomic scale. Nature Ecology & Evolution **5**:747-756.

Betancourt, M. 2017. A conceptual introduction to Hamiltonian Monte Carlo. arXiv preprint arXiv:1701.02434.

Bledsoe, E. K., and S. M. Ernest. 2019. Temporal changes in species composition affect a ubiquitous species' use of habitat patches. Ecology **100**:e02869.

Brown, J. H. 1998. The desert granivory experiments at portal. Pages 71-95 *in* W. J. Resetarits and J. Bernardo, editors. Experimental Ecology. Oxford University Press, Oxford, UK.

Brown, J. H., and S. M. Ernest. 2002. Rain and rodents: complex dynamics of desert consumers: although water is the primary limiting resource in desert ecosystems, the relationship between rodent population dynamics and precipitation is complex and nonlinear. BioScience **52**:979-987.

Bunin, G. 2017. Ecological communities with Lotka-Volterra dynamics. Physical Review E **95**:042414.

Cárdenas, P. A., E. Christensen, S. K. M. Ernest, D. C. Lightfoot, R. L. Schooley, P. Stapp, and J. A. Rudgers. 2021. Declines in rodent abundance and diversity track regional climate variability in North American drylands. Global Change Biology **27**:4005-4023.

Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell. 2017. Stan: A probabilistic programming language. Journal of Statistical Software **76**.

Carter, S. K., and V. H. Rudolf. 2022. Exploring conditions that strengthen or weaken the ecological and evolutionary consequences of phenological synchrony. The American Naturalist **200**:E189-E206.

Christensen, E. M., D. J. Harris, and S. Ernest. 2018. Long-term community change through multiple rapid transitions in a desert rodent community. Ecology **99**:1523-1529.

Christensen, E. M., G. L. Simpson, and S. Ernest. 2019a. Established rodent community delays recovery of dominant competitor following experimental disturbance. Proceedings of the Royal Society B: Biological Sciences **286**:20192269.

Christensen, E. M., G. M. Yenni, H. Ye, J. L. Simonis, E. K. Bledsoe, R. Diaz, S. D. Taylor, E. P. White, and S. Ernest. 2019b. portalr: an R package for summarizing and using the Portal Project Data. Journal of Open Source Software **4**:1098.

Clark, J. S., S. R. Carpenter, M. Barber, S. Collins, A. Dobson, J. A. Foley, D. M. Lodge, M. Pascual, R. Pielke, and W. Pizer. 2001. Ecological forecasts: an emerging imperative. Science **293**:657-660.

Clark, N. J., J. T. Kerry, and C. I. Fraser. 2020. Rapid winter warming could disrupt coastal marine fish community structure. Nature Climate Change:DOI: 10.1038/s41558-41020-40838-41555.

Clark, N. J., T. Proboste, G. Weerasinghe, and R. J. Soares Magalhães. 2022. Near-term forecasting of companion animal tick paralysis incidence: An iterative ensemble model. PLoS Computational Biology **18**:e1009874.

Clark, N. J., and K. Wells. 2023. Dynamic generalised additive models (DGAMs) for forecasting discrete ecological time series. Methods in Ecology and Evolution **14**:771-784.

Clark, N. J., K. Wells, D. Dimitrov, and S. M. Clegg. 2016. Co-infections and environmental conditions drive the distributions of blood parasites in wild birds. Journal of Animal Ecology **85**:1461-1470.

Daugaard, U., S. B. Munch, D. Inauen, F. Pennekamp, and O. L. Petchey. 2022. Forecasting in the face of ecological complexity: Number and strength of species interactions determine forecast skill in ecological communities. Ecology Letters **25**:1974-1985.

Dietze, M. C., A. Fox, L. M. Beck-Johnson, J. L. Betancourt, M. B. Hooten, C. S. Jarnevich, T. H. Keitt, M. A. Kenney, C. M. Laney, and L. G. Larsen. 2018. Iterative near-term ecological forecasting: Needs, opportunities, and challenges. Proceedings of the National Academy of Sciences **115**:1424-1432.

Dietze, M. C., S. P. Serbin, C. Davidson, A. R. Desai, X. Feng, R. Kelly, R. Kooper, D. LeBauer, J. Mantooth, and K. McHenry. 2014. A quantitative assessment of a terrestrial biosphere model's data needs across North American biomes. Journal of Geophysical Research: Biogeosciences **119**:286-300.

Dumandan, P. K. T., J. L. Simonis, G. M. Yenni, S. K. M. Ernest, and E. P. White. 2024. Transferability of ecological forecasting models to novel biotic conditions in a long-term experimental study. bioRxiv:2023.2011.2001.565145.

Dumandan, P. K. T., G. M. Yenni, and M. Ernest. 2023. Shifts in competitive structures can drive variation in species phenology. Ecology **104**:e4160.

Ernest, S., and J. H. Brown. 2001. Delayed compensation for missing keystone species by colonization. Science **292**:101-104.

Ernest, S., G. M. Yenni, G. Allington, E. K. Bledsoe, E. M. Christensen, R. M. Diaz, K. Geluso, J. R. Goheen, Q. Guo, E. Heske, D. Kelt, J. M. Meiners, J. Munger, C. Restrepo, D. A. Samson, M. R. Schutzenhofer, M. Skupski, S. R. Supp, K. M. Thibault, S. D. Taylor, E. P. White, H. Ye, D. W. Davidson, J. H. Brown, and T. J. Valone. 2020. The Portal Project: a long-term study of a Chihuahuan desert ecosystem. bioRxiv:332783.

Fredston, A. L., W. W. L. Cheung, T. L. Frölicher, Z. J. Kitchel, A. A. Maureaud, J. T. Thorson, A. Auber, B. Mérigot, J. Palacios-Abrantes, M. L. D. Palomares, L. Pecuchet, N. L. Shackell, and M. L. Pinsky. 2023. Marine heatwaves are not a dominant driver of change in demersal fishes. Nature **621**:324-329.

Gabry, J., and R. Češnovar. 2021. Cmdstanr: R interface to ’CmdStan’. https://mc-stan.org/cmdstanr.

Gerber, L. R., L. W. Botsford, A. Hastings, H. P. Possingham, S. D. Gaines, S. R. Palumbi, and S. Andelman. 2003. Population models for marine reserve design: a retrospective and prospective synthesis. Ecological Applications **13**:47-64.

Greenville, A. C., G. M. Wardle, V. Nguyen, and C. R. Dickman. 2016. Population dynamics of desert mammals: similarities and contrasts within a multispecies assemblage. Ecosphere **7**:e01343.

Hampton, S. E., E. E. Holmes, L. P. Scheef, M. D. Scheuerell, S. L. Katz, D. E. Pendleton, and E. J. Ward. 2013. Quantifying effects of abiotic and biotic drivers on community dynamics with multivariate autoregressive (MAR) models. Ecology **94**:2663-2669.

Hannaford, N. E., S. E. Heaps, T. M. Nye, T. P. Curtis, B. Allen, A. Golightly, and D. J. Wilkinson. 2023. A sparse Bayesian hierarchical vector autoregressive model for microbial dynamics in a wastewater treatment plant. Computational Statistics & Data Analysis **179**:107659.

Harris, D. J., S. D. Taylor, and E. P. White. 2018. Forecasting biodiversity in breeding birds using best practices. PeerJ **6**:e4278.

Heaps, S. E. 2023. Enforcing stationarity through the prior in vector autoregressions. Journal of Computational and Graphical Statistics **32**:74-83.

Heske, E. J., J. H. Brown, and S. Mistry. 1994. Long-term experimental study of a Chihuahuan Desert rodent community: 13 years of competition. Ecology **75**:438-445.

Holmes, E., E. Ward, and M. Scheuerell. 2014. Analysis of multivariate time-series using the MARSS package. NOAA Fisheries, Northwest Fisheries Science Center **2725**:98112.

Ibáñez, I., J. A. Silander Jr, A. M. Wilson, N. LaFleur, N. Tanaka, and I. Tsuyama. 2009. Multivariate forecasts of potential distributions of invasive plant species. Ecological Applications **19**:359-375.

Intergovernmental Science - Policy Platform on Biodiversity and Ecosystem Services. 2019. Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Bonn, Germany.

Ives, A. R., K. C. Abbott, and N. L. Ziebarth. 2010. Analysis of ecological time series with ARMA (p, q) models. Ecology **91**:858-871.

Ives, A. R., B. Dennis, K. L. Cottingham, and S. R. Carpenter. 2003. Estimating community stability and ecological interactions from time-series data. Ecological Monographs **73**:301-330.

Johnson-Bice, S. M., J. M. Ferguson, J. D. Erb, T. D. Gable, and S. K. Windels. 2021. Ecological forecasts reveal limitations of common model selection methods: predicting changes in beaver colony densities. Ecological Applications **31**:e02198.

Karp, M. A., J. S. Link, M. Grezlik, S. Cadrin, G. Fay, P. Lynch, H. Townsend, R. D. Methot, G. D. Adams, and K. Blackhart. 2023. Increasing the uptake of multispecies models in fisheries management. ICES Journal of Marine Science **80**:243-257.

Karunarathna, K. A. N. K., K. Wells, and N. J. Clark. 2024. Modelling nonlinear responses of a desert rodent species to environmental change with hierarchical dynamic generalized additive models. Ecological Modelling **490**:110648.

Kerley, G. I., W. G. Whitford, and F. R. Kay. 1997. Mechanisms for the keystone status of kangaroo rats: graminivory rather than granivory? Oecologia **111**:422-428.

Lewis, A. S., C. R. Rollinson, A. J. Allyn, J. Ashander, S. Brodie, C. B. Brookson, E. Collins, M. C. Dietze, A. S. Gallinat, and N. Juvigny-Khenafou. 2023. The power of forecasts to advance ecological theory. Methods in Ecology and Evolution **14**:746-756.

Lewis, A. S. L., W. M. Woelmer, H. L. Wander, D. W. Howard, J. W. Smith, R. P. McClure, M. E. Lofton, N. W. Hammond, R. S. Corrigan, R. Q. Thomas, and C. C. Carey. 2022. Increased adoption of best practices in ecological forecasting enables comparisons of forecastability. Ecological Applications **32**:e2500.

Lima, M., S. M. Ernest, J. H. Brown, A. Belgrano, and N. C. Stenseth. 2008. Chihuahuan Desert kangaroo rats: nonlinear effects of population dynamics, competition, and rainfall. Ecology **89**:2594-2603.

Lindenmayer, D. B., G. E. Likens, A. Andersen, D. Bowman, C. M. Bull, E. Burns, C. R. Dickman, A. A. Hoffmann, D. A. Keith, and M. J. Liddell. 2012. Value of long-term ecological studies. Austral Ecology **37**:745-757.

Lütkepohl, H. 1990. Asymptotic distributions of impulse response functions and forecast error variance decompositions of vector autoregressive models. The review of economics and statistics:116-125.

McElreath, R. 2020. Statistical rethinking: A Bayesian course with examples in R and Stan. 2 edition. Chapman and Hall/CRC.

Moustahfid, H., L. C. Hendrickson, A. Arkhipkin, G. J. Pierce, A. Gangopadhyay, H. Kidokoro, U. Markaida, C. Nigmatullin, W. H. Sauer, and P. Jereb. 2021. Ecological-fishery forecasting of squid stock dynamics under climate variability and change: review, challenges, and recommendations. Reviews in Fisheries Science & Aquaculture **29**:682-705.

Norberg, A., N. Abrego, F. G. Blanchet, F. R. Adler, B. J. Anderson, J. Anttila, M. B. Araújo, T. Dallas, D. Dunson, J. Elith, S. D. Foster, R. Fox, J. Franklin, W. Godsoe, A. Guisan, B. O'Hara, N. A. Hill, R.

D. Holt, F. K. C. Hui, M. Husby, J. A. Kålås, A. Lehikoinen, M. Luoto, H. K. Mod, G. Newell, I. Renner, T. Roslin, J. Soininen, W. Thuiller, J. Vanhatalo, D. Warton, M. White, N. E. Zimmermann, D. Gravel, and O. Ovaskainen. 2019. A comprehensive evaluation of predictive performance of 33 species distribution models at species and community levels. Ecological Monographs **89**:e01370.

Ogle, K., J. J. Barber, G. A. Barron-Gafford, L. P. Bentley, J. M. Young, T. E. Huxman, M. E. Loik, and D. T. Tissue. 2015. Quantifying ecological memory in plant and ecosystem processes. Ecology Letters **18**:221-235.

Ovaskainen, O., G. Tikhonov, D. Dunson, V. Grøtan, S. Engen, B.-E. Sæther, and N. Abrego. 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.

Paniw, M., D. García-Callejas, F. Lloret, R. D. Bassar, J. Travis, and O. Godoy. 2023. Pathways to global-change effects on biodiversity: new opportunities for dynamically forecasting demography and species interactions. Proceedings of the Royal Society B **290**:20221494.

Pedersen, E. J., D. L. Miller, G. L. Simpson, and N. Ross. 2019. Hierarchical generalized additive models in ecology: an introduction with mgcv. PeerJ **7**:e6876.

Powell-Romero, F., N. M. Fountain-Jones, A. Norberg, and N. J. Clark. 2023. Improving the predictability and interpretability of co-occurrence modelling through feature-based joint species distribution ensembles. Methods in Ecology and Evolution **14**:146-164.

Quinn, T. J. 2003. Ruminations on the development and future of population dynamics models in fisheries. Natural Resource Modeling **16**:341-392.

R Core Team. 2023. R: A language and environment for statistical computing. R Development Core Team, Vienna, Austria.

Romañach, S. S., S. M. Haider, C. Hackett, M. McKelvy, and L. G. Pearlstine. 2022. Managing multiple species with conflicting needs in the Greater Everglades. Ecological Indicators **136**:108669.

Sandal, L., V. Grøtan, B.-E. Sæther, R. P. Freckleton, D. G. Noble, and O. Ovaskainen. 2022. Effects of density, species interactions, and environmental stochasticity on the dynamics of British bird communities. Ecology **103**:e3731.

Scheuerer, M., and T. M. Hamill. 2015. Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. Monthly Weather Review **143**:1321-1334.

Simonis, J. L., E. P. White, and S. K. M. Ernest. 2021. Evaluating probabilistic ecological forecasts. Ecology **102**:e03431.

Simonis, J. L., G. M. Yenni, E. K. Bledsoe, E. M. Christensen, H. Senyondo, S. D. Taylor, H. Ye, E. P. White, and S. M. Ernest. 2022. portalcasting: Supporting automated forecasting of rodent populations. Journal of Open Source Software **7**:3220.

Stan Development Team. 2022. Stan Modeling Language Users Guide and Reference Manual, Version 2.26.1. https://mc-stan.org.

Sun, D., and M. Kafatos. 2007. Note on the NDVI-LST relationship and the use of temperature-related drought indices over North America. Geophysical Research Letters **34**.

Thorson, J. T., J. N. Ianelli, E. A. Larsen, L. Ries, M. D. Scheuerell, C. Szuwalski, and E. F. Zipkin. 2016. Joint dynamic species distribution models: a tool for community ordination and spatio-temporal monitoring. Global Ecology and Biogeography **25**:1144-1158.

Tobler, M. W., M. Kéry, F. K. Hui, G. Guillera-Arroita, P. Knaus, and T. Sattler. 2019. Joint species distribution models with species correlations and imperfect detection. Ecology **100**:e02754.

Tonkin, J. D., M. T. Bogan, N. Bonada, B. Rios-Touma, and D. A. Lytle. 2017. Seasonality and predictability shape temporal species diversity. Ecology **98**:1201-1216.

Tulloch, A. I., V. Hagger, and A. C. Greenville. 2020. Ecological forecasts to inform near-term management of threats to biodiversity. Global Change Biology **26**:5816-5828.

Vehtari, A., A. Gelman, and J. Gabry. 2017. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Statistics and computing **27**:1413-1432.

Volterra, V. 1931. Variations and fluctuations of the number of individuals in animal species living together. Animal ecology:412-433.

Ward, E. J., H. Chirakkal, M. González-Suárez, D. Aurioles-Gamboa, E. E. Holmes, and L. Gerber. 2010. Inferring spatial structure from time-series data: using multivariate state-space models to detect metapopulation structure of California sea lions in the Gulf of California, Mexico. Journal of Applied Ecology **47**:47-56.

Ward, E. J., E. E. Holmes, J. T. Thorson, and B. Collen. 2014. Complexity is costly: a meta-analysis of parametric and non-parametric methods for short-term population forecasting. Oikos **123**:652-661.

Ward, E. J., K. Marshall, and M. D. Scheuerell. 2022. Regularizing priors for Bayesian VAR applications to large ecological datasets. PeerJ **10**:e14332.

Warton, D. I., F. G. Blanchet, R. B. O'Hara, O. Ovaskainen, S. Taskinen, S. C. Walker, and F. K. Hui. 2015. So many variables: joint modeling in community ecology. Trends in Ecology & Evolution **30**:766-779.

Wells, K., R. B. O'Hara, B. D. Cooke, G. J. Mutze, T. A. A. Prowse, and D. A. Fordham. 2016. Environmental effects and individual body condition drive seasonal fecundity of rabbits: identifying acute and lagged processes. Oecologia **181**:853-864.

White, E. P., G. M. Yenni, S. D. Taylor, E. M. Christensen, E. K. Bledsoe, J. L. Simonis, and S. Ernest. 2019. Developing an automated iterative near-term forecasting system for an ecological study. Methods in Ecology and Evolution **10**:332-344.