Multi-species dependencies improve forecasts of population dynamics in a long-term monitoring study

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: n.clark@uq.edu.au

Short title: Multivariate ecological forecasting

Total word count (including references): 4997

Number of words in Abstract: 146

Number of References: 34

Number of Tables: 0

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

Forecasts of community dynamics are essential for the management of biodiversity. Theory suggests these predictions can be improved by leveraging multi-species dependencies to improve models, but empirical support for this is lacking. We test whether models that learn from multiple species, both to estimate nonlinear environmental effects and temporal dependence, improve forecasts for a semi-arid rodent community. Using Dynamic Generalized Additive Models, we analyze monthly captures for nine rodents over 25 years. We find strong evidence that multi-species dependencies improve performance, as models that captured these effects gave superior predictions. These models also provide novel insights, in our case by quantifying how changes in abundance for some species can have delayed, nonlinear effects on others while also uncovering important lagged effects of environmental drivers. We show that multivariate models are useful not only to improve 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 ecosystems is a global challenge (Intergovernmental Science - Policy Platform on Biodiversity and Ecosystem Services 2019). The need for explicit predictions has sparked a renewed emphasis on ecological forecasting (Lewis et al. 2023) including the prediction of population dynamics (Ward et al. 2014, Johnson-Bice et al. 2021). But this is a difficult task, particularly because ecological forecasting is often a multivariate problem. Monitoring studies typically gather data for multiple species, and managers often require community predictions to guide decisions (Clark et al. 2001, Greenville et al. 2016). A major focus of empirical and theoretical research seeks to understand how we can improve predictions using models that leverage data from multiple species (Holmes et al. 2014, Powell-Romero et al. 2023).

Forecasting species abundances is inherently difficult because many processes influence population dynamics (Johnson-Bice et al. 2021, Clark and Wells 2023). Population dynamics often exhibit complex responses to drivers (including non-linear responses and lags; Cárdenas et al. 2021), temporal autocorrelation (Ives et al. 2003), and data complexities (e.g., irregular sampling, observation errors, missing samples, and over dispersed counts; Clark and Wells 2023). Species' relationships are also important. The way a species responds to change results from direct influences of the environment on its ability to find resources and reproduce (Heske et al. 1994) and the effects of other species on its abundance and vital rates (Ives et al. 2003). Inference on both processes can be informed using data from multiple species, which should theoretically improve near-term predictions. For example, we expect some species to exhibit shared responses to environmental factors (Christensen et al. 2018), and we gain more precise inferences by learning these responses hierarchically. Importantly, these responses do not need to be assumed to be linear, as Pedersen and colleagues (2019) recently showed how Hierarchical Generalized Additive Models (HGAMs) can estimate multiple non-linear relationships in a single joint model. In addition to shared environmental responses, we also expect direct biotic interactions to play important roles (Heske et

al. 1994, Ernest and Brown 2001, Bledsoe and Ernest 2019). Models that impose multispecies temporal dependence offer a solution to explore these interactions and use them to make more informed predictions (Ives et al. 2003, Greenville et al. 2016). However, while we postulate that capturing these sources of multispecies dependence will improve population-level forecasts, empirical tests of this hypothesis are rare.

We evaluate whether models that incorporate multi-species relationships can improve near-term population forecasts. Using multi-species data from a long-term monitoring study, we build models that learn species' shared environmental responses and temporal interactions to make inference about factors that relate to community dynamics. Comparisons with simpler models are used to test whether the incorporating biotic dependence structures improves predictions. We show that this approach to modelling population abundances provides a more detailed understanding of the drivers of ecological community change as well as more accurate forecasts.

MATERIALS AND METHODS

Rodent capture data

Our data come from the Portal Project, a long-term monitoring study of a desert rodent community near Portal, Arizona (Brown 1998, Ernest et al. 2020). Sampling covers 24 experimental plots (50m x 50m), each containing 49 baited traps that are opened following the lunar monthly cycle (Ernest et al. 2020). The data exhibit many of the complexities that confront population forecasting. These include imperfect detection, missing observations (~5% on average; Dumandan et al. 2023b) and over dispersed discrete counts that include many zeros. Environmental drivers are known to exhibit lagged and nonlinear impacts on rodent population dynamics (Brown and Ernest 2002). The rodents at Portal compete for resources in complex ways, and these biotic interactions are postulated to have important consequences for population dynamics (Lima et al. 2008, Bledsoe and Ernest 2019).

We used the *portalr* R package (Christensen et al. 2019) to extract trapping records from the Portal data (version 3.134.0; downloaded October 2022; https://doi.org/10.5281/zenodo.7255488). The design uses three experimental treatments with varying levels of rodent exclusion (Ernest et al. 2020), but our study focused on long-term control plots (that allow free movement of all rodents) for the period December 1996 – August 2022. We excluded transient species, observed in < 10% of trapping sessions, to focus on species with the most influence on community dynamics (Figure 1).

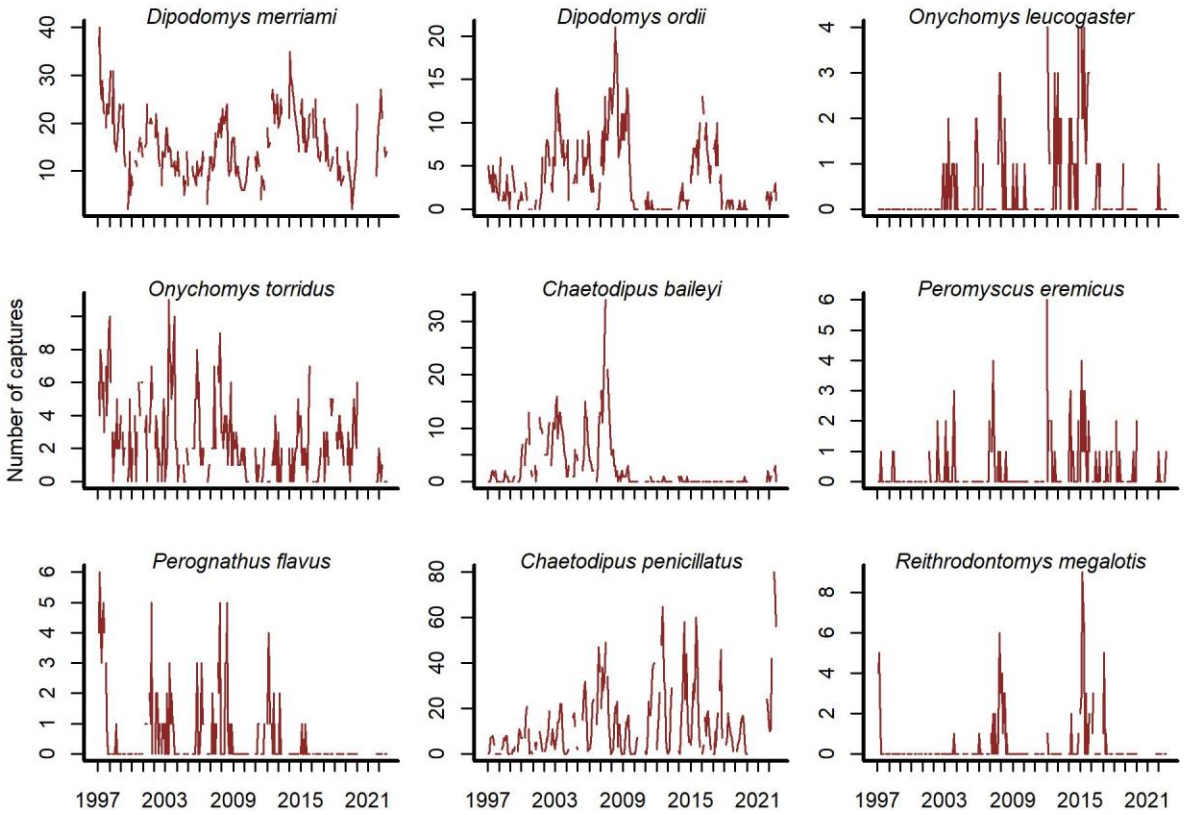

Figure 1: Rodent captures from December 1996 to August 2022. Counts are total captures across long-term control plots. Blanks are missing values.

Covariate measurements

Rodents at Portal 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 USGS Earth Resources Observation and Science Center; https://www.usgs.gov/centers/eros). Measurements were converted to monthly averages and extracted from one year before the start of captures (from January 1995) to calculate lagged and moving average versions.

Model description

There were several aspects we needed to consider in our model. Total captures showed both short- and long-term fluctuations (Figure S1). Captures for individual species undulated over multi-annual cycles and were positively autocorrelated at lags up to 20 months (Figures S2 and S3). To test whether multi-species information improves model performance, we needed to model these dynamics using multivariate dependence structures. We also needed to leverage community information to estimate species' time-delayed responses to vegetation and temperature variation. Rodent captures were modelled as $Poisson$ observations of a latent state that was composed of a hierarchical GAM component (to capture shared environmental responses) and a multivariate dynamic component to capture multispecies dependence. The full description for this model, which we abbreviate to ***GAM-VAR*** (Figure 2).

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

GAM component

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

Random slopes for species' responses to NDVI moving average

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

Distributed lag for community response to lagged minimum temperature

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

Distributed lags for species' responses to lagged minimum temperature

Multivariate dynamic component

$$A \in P(\mathbb{R})$$
$$P \sim Normal(0, 0.67)$$

Vector autoregression for lagged multispecies dependence

$$\sigma \sim Beta(10, 10)$$
$$C \sim LKJcorr(2)$$

Temporally correlated process errors

Figure 2: Definition of the ***GAM-VAR*** model. Coloured boxes highlight the five main components of the latent state model.

The GAM component consisted of hierarchical NDVI and minimum temperature effects that were informed by theory and exploration of covariate time series (shown in 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 hierarchical slopes. Responses to temperature were estimated using hierarchical distributed lags in which nonlinear effects of minimum temperature varied smoothly with increasing lag. To encourage multi-species learning, we included a shared community-level response $f_{global}(Mintemp, lag)$ and species-level deviations $f_{species[i]}(Mintemp, lag)$. These allowed each species to show a different temperature response from the community average, 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 matrix of autoregressive coefficients. Diagonal entries of A described density-dependence,

or the effect of a species' state (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 state at time $t-1$ on the current state 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 mapped the constrained A matrix to unconstrained partial autocorrelations P. Process errors were allowed to be contemporaneously dependent to capture any unmodelled correlations. Priors are shown in Figure 2 and described in the accompanying R code.

Do multi-species dependencies improve predictions?

We tested whether learning from multiple species improved our model's predictions using 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).

Benchmarking against simpler models is also useful for forecast evaluation (Simonis et al. 2021). 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 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) using the *cmdstanr* interface (Gabry and Češnovar 2021). Stan's superior 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 2020) 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 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 systematic variation (Figures S6 – S7). In-sample performances differed among models, with models that leveraged multi-species information producing higher ELPD's than 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 large uncertainty about the magnitude of this difference (Table S1).

Forecast performance also differed among models, with more complex multi-species models again tending to outperform 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; 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 ***GAM-VAR*** model estimated large, positive autoregressive coefficients for most species (diagonal entries in Figure S9). It also relied strongly on information from multiple species by estimating many non-zero cross-dependence effects (off-diagonal entries in Figure S9) and process error correlations (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 (Figure S11). It was these multispecies effects that enabled the ***GAM-VAR*** to produce more accurate forecasts. For example, Ord's kangaroo rat (*D. ordii*) and silky pocket mouse (*P. flavus*) had negative cross-dependencies, 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 (Figure S12). Next, we use simulations to interpret the multi-species effects that allowed the ***GAM-VAR*** to outperform simpler models.

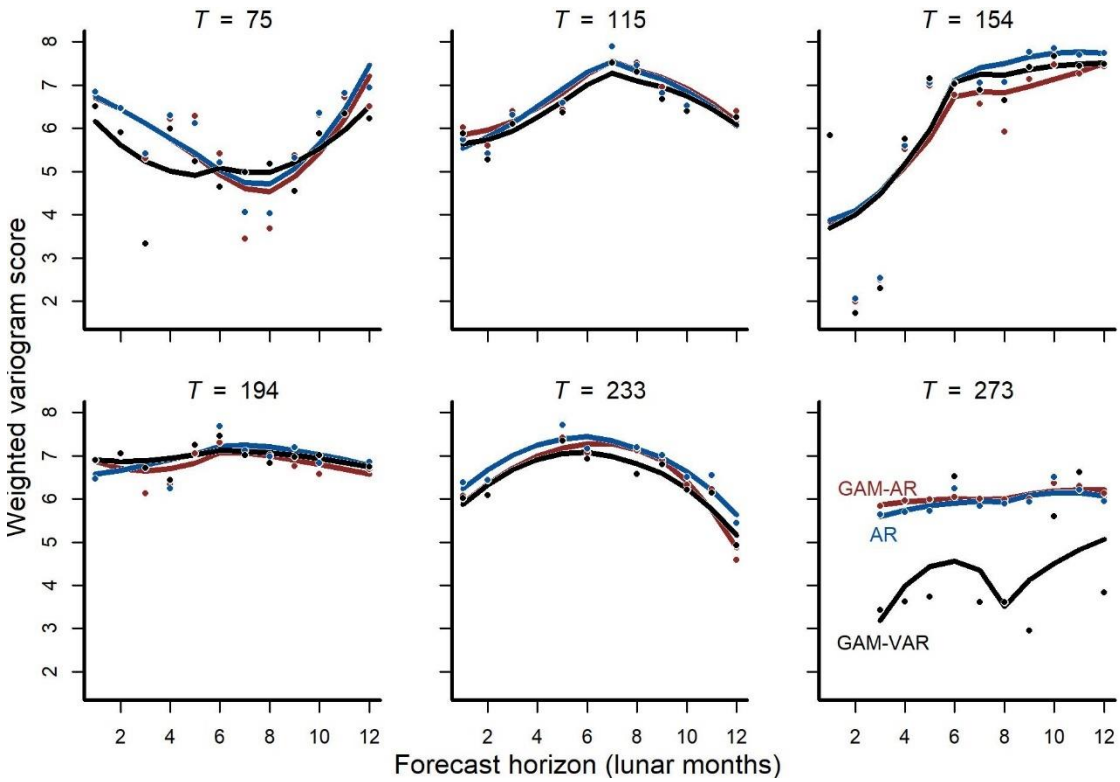

Figure 3: Forecast performances for three of the competing models (we do not show metrics for the ***GAM-AR no pooling*** model as they were not distinguishable from the ***GAM-AR*** metrics). Y-axis shows the log of the weighted variogram score, a scoring rule that penalizes multivariate forecasts if they are not well calibrated and do not capture inter-series correlations (lower scores preferred). 12-step ahead predictions were evaluated over six evenly spaced origins. Points show individual scores. Lines show Loess smoothed trend lines. Missing points indicate a missing observation.

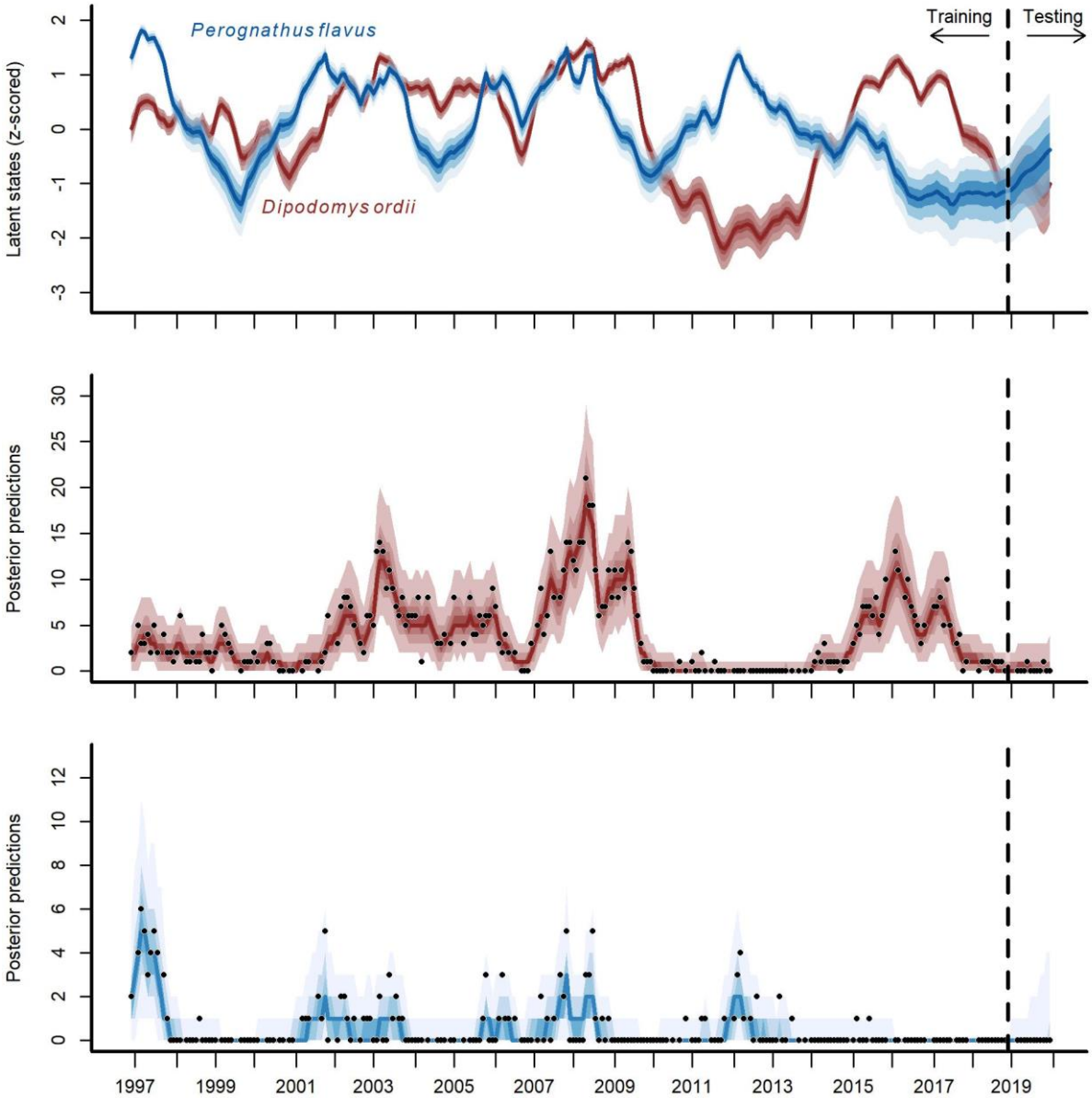

Figure 4: Posterior state estimates (top panel) and posterior predictions (bottom two panels) from the ***GAM-VAR*** model for Ord's kangaroo rat (*Dipodomys ordii*; in red) and silky pocket mouse (*Perognathus flavus*; in blue) for training and testing periods (demarked by the dashed line). State estimates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

Species relationships provide new 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 (Figure S13). But interpreting this cross-dependence is difficult because VAR effects provide only a marginal view into a 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 (Figure S14).

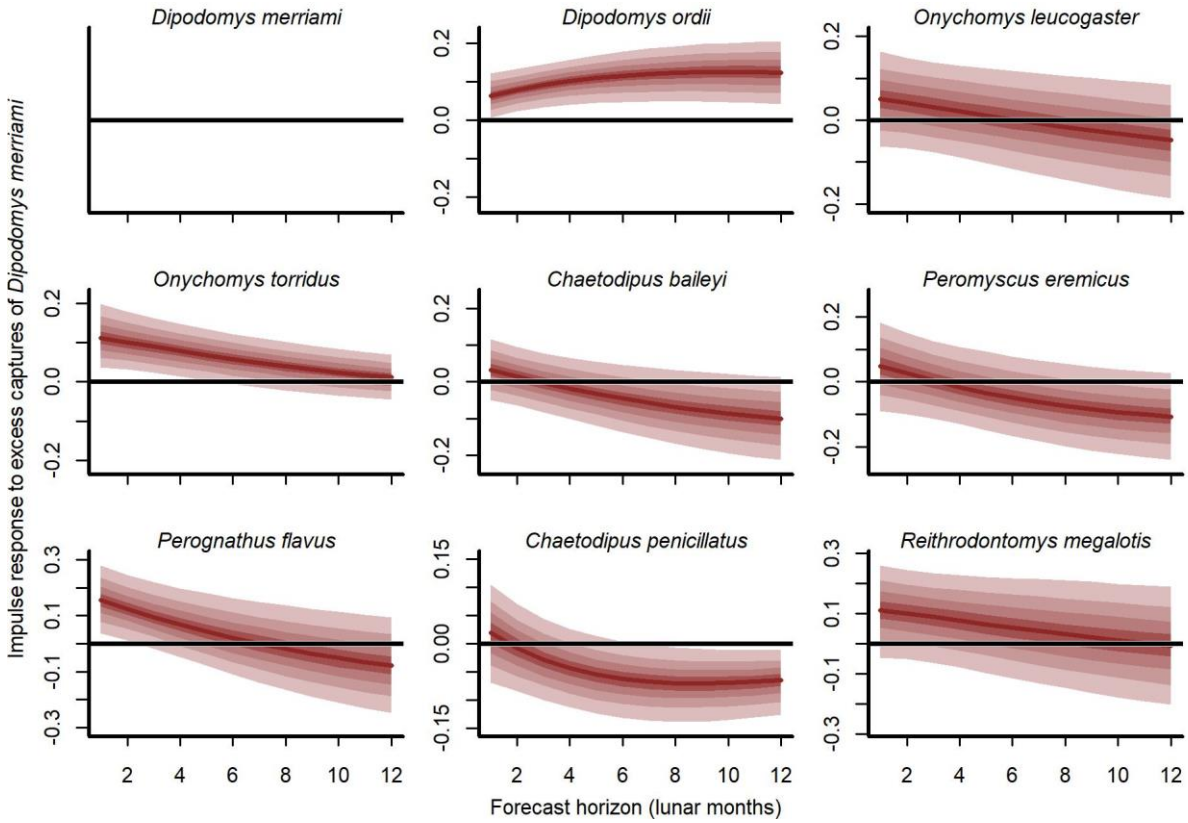

Figure 5: Impulse responses showing how mean captures (μ, on the log scale) are expected to change over the next six months if three additional *D. merriami* individuals are captured. Ribbons show quantiles (90th, 60th, 40th and 20th). Dark red lines show posterior medians.

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 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 Meriam's kangaroo rat (*D. merriami*). For these species, the model expected increases in ~70% of simulations and decreases in ~30% (Figure 6). While conclusions were

generally similar when using the ***GAM-AR no pooling*** model, which did not leverage multi-species learning, estimates were much more variable (Figure S15).

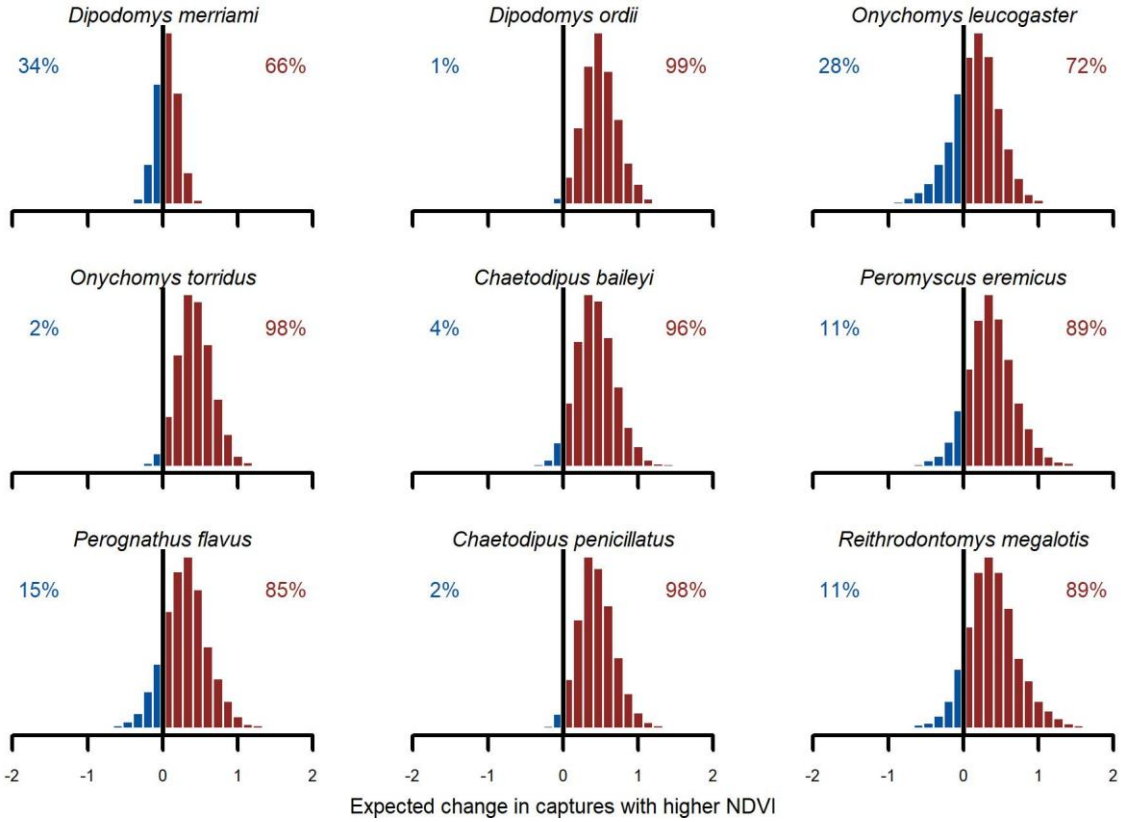

Figure 6: Posterior NDVI contrasts from the ***GAM-VAR*** model. Histograms illustrate how expected captures would change if the z-scored NDVI 12-month moving average changed from a relatively low value (-0.50) to a relatively high value (0.50). Numbers indicate the proportion of probability mass at or below zero (in blue) vs above zero (in red).

We interpreted minimum temperature distributed lag effects by simulating functions for each species using temperatures from 1997 (Figure S15). There was 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 expected more captures in spring (March – May) and fewer in late summer /

autumn (July – October). But the shapes of these responses varied. Kangaroo rats (*D. merriami* and *D. ordii*) had smoother shapes that decreased gradually from mid-summer to mid-winter, while the Southern grasshopper mouse (*O. torridus*) was expected to show higher captures in late autumn / early winter (Figure S16). The five species that relied solely on the shared 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 obvious. There was not enough information to learn nonlinear functions for these species, with the model 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 (Hampton et al. 2013, Holmes et al. 2014). Our results show that incorporating relationships between species to estimate temporal dependence, and to learn non-linear associations with environmental drivers, yields better model fits and more accurate 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. Models that describe biological complexity, both through nonlinear covariate functions and multi-species dependence, are useful for asking targeted questions about drivers of change.

Leveraging relationships between species

Our analyses show why models that target multi-species effects should be prioritized to study 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 relationships between species the model could better capture both shared responses to environmental factors (e.g., a wet year is good for most species) and direct temporal effects (e.g., competition for seeds). Simulations provided deeper insights into multi-species population dynamics. Simulated responses to sudden impulses were often delayed and nonlinear, with cascading impacts lasting up to six months. This type of simulation can have many broad uses in applied ecology, for example by assessing which species have the strongest cascading effects, what changes we can expect from management actions targeting key species, and how these effects relate to regime transitions (Ives et al. 2003).

Learning hierarchical nonlinear effects from community data

Our model captured linear, nonlinear, and lagged responses to environmental and climatic covariates that were informed by all species at once. While we cannot interpret these estimates as causal, recovered relationships were consistent with operational models of the system. Positive associations between capture rates and a 12-month moving average of NDVI were expected because rodents at Portal depend on plants for food and resources (Ernest et al. 2000, Brown and Ernest 2002). But there were interesting patterns of variation in these effects. The strongest positive association was shown by Ord's kangaroo rat (*D. ordii*), a species that 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 with more severe droughts, in part due to a preference for foraging habitat with less vegetation (Cárdenas et al. 2021).

Our study shows how distributed lag effects can be learned hierarchically and provides useful insights into delayed responses to temperature for rodents at Portal. Most species showed higher captures when temperatures were low 3 – 4 months prior, suggesting increases begin during

mid to late spring when resources become available. But others showed increases during cooler months in autumn and winter. Asynchronous phenology, where species show different reproductive timing in the presence of competition, has been reported at Portal (Dumandan et al. 2023b) and may be one reason why seasonal patterns differed among species. Comparing temperature responses on control vs experimental plots may help to determine if these competitive forces play a role in seasonal capture variation.

Despite the large number of observations in our data, estimates of environmental responses were more precise when they were learned from all species jointly. This strongly supports the need for hierarchical models in ecology. But while hierarchical intercepts and slopes are common, there has been less emphasis on hierarchical nonlinear functions (but see Pedersen et al. 2019). Open access to 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.

Future directions

As ecosystems undergo shifts in species composition (e.g. Dornelas et al. 2014), integrating species relationships into forecasting models may become increasingly important. Recent work leveraging the Portal Project's long-term experimental manipulations shows that single-species models can perform poorly when transferred to novel biotic contexts (i.e., between controls and manipulations Dumandan et al. 2023a). This suggests that in ecosystems undergoing biotic reorganization single-species models may be insufficient for predicting future populations, even if such models perform well under current conditions. By considering species relationships, multi-species models like the ***GAM-VAR*** have the potential to transfer better into situations where composition is changing – an assumption that can be tested using the experimental plots at Portal.

More broadly, we hope that estimating multi-species dependencies will result in more accurate forecasts, inspire new questions, and lead to an improved understanding of the factors that govern ecological community dynamics. The combination of long-term monitoring, manipulative experiments, and models that capture real-world biological complexities provide a framework for advancing both our understanding of ecological systems and our ability to forecast their 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).

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 $\boldsymbol{A}$). Off-diagonals represent cross-dependencies. For example, the entry in $\boldsymbol{A}[\mathbf{1},\mathbf{2}]$ 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 $\boldsymbol{C}$). 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 ($\boldsymbol{\Sigma_{VAR}}$) 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

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.

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**.

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. M. Yenni, H. Ye, J. L. Simonis, E. K. Bledsoe, R. Diaz, S. D. Taylor, E. P. White, and S. Ernest. 2019. 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., and K. Wells. 2023. Dynamic generalised additive models (DGAMs) for forecasting discrete ecological time series. Methods in Ecology and Evolution **14**:771-784.

Dornelas, M., N. J. Gotelli, B. McGill, H. Shimadzu, F. Moyes, C. Sievers, and A. E. Magurran. 2014. Assemblage time series reveal biodiversity change but not systematic loss. Science **344**:296-299.

Dumandan, P. K. T., J. L. Simonis, G. M. Yenni, S. K. M. Ernest, and E. P. White. 2023a. 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. 2023b. 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.

Ernest, S. M., J. H. Brown, and R. R. Parmenter. 2000. Rodents, plants, and precipitation: spatial and temporal dynamics of consumers and resources. Oikos **88**:470-482.

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

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.

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.

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., 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.

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.

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.

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.

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.

R Core Team. 2020. R: A language and environment for statistical computing. R Development Core Team, Vienna, Austria.

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.

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.

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.