Do we need demographic data to forecast the state of plant populations?

Andrew T. Tredennick[1]*, Mevin B. Hooten[2,3,4], and Peter B. Adler[1]

[1] Department of Wildland Resources and the Ecology Center, 5230 Old Main Hill, Utah State University, Logan, Utah 84322, USA; [2] U.S. Geological Survey, Colorado Cooperative Fish and Wildlife Research Unit, Department of Fish, Wildlife and Conservation Biology, Colorado State University, Fort Collins, CO 80523-1403, USA; [3] Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, CO 80523 USA; [4] Department of Statistics, Colorado State University, Fort Collins, CO 80523 USA

Summary

1. Rapid environmental change has generated growing interest in forecasts of future population trajectories. Traditional population models built with detailed demographic observations from one study site can address the impacts of environmental change at particular locations, but are difficult to scale up to the landscape and regional scales relevant to management decisions.

2. An alternative is to build models using population-level data which are much easier to collect than individual-level data over broad spatial scales. However, it is unknown whether models built using population-level data adequately capture the effects of density-dependence and environmental forcing that are necessary to generate skillful forecasts.

3. Here, we test the consequences of aggregating individual responses when forecasting the population states and trajectories of four perennial grass species in a semi-arid grassland in Montana,

*Corresponding author: E-mail: atredenn@gmail.com

USA. We parameterized two population models for each species, one based on individual-level data (survival, growth and recruitment) and one on population-level data (percent cover), and compared their forecasting skill and forecast horizons with and without the inclusion of climate covariates. For both models we used Bayesian ridge regression to weight the influence of climate covariates for optimal prediction.

4. In the absence of climate effects, we found no significant difference between the forecasting skill of models based on individual-level data and models based on population-level data. Climate effects were weak and only marginally increased forecasting skill. Increases in skill with climate covariates were similar between model types for three of the four species. For the fourth species, forecast accuracy of the individual-level model with climate covariates was significantly higher than the accuracy of the equivalent population-level model.

5. *Synthesis.* For our focal species at this particular location, and using our particular statistical models, demographic data was generally unnecessary to achieve skillful [ARE ANY OF THESE FORECASTS SKILLFUL? FOCUS INSTEAD ON MODEL DIFFERENCES] forecasts, though for certain species forecast skill can be gained by using demographic data linked to climate covariates. We conclude that models based on aggregated individual-level data offer a practical alternative to data-intensive demographic models when species do not respond strongly to interannual variation in weather, but when modeling species that do respond to climate drivers, demographically-based models can generate more skillful forecasts.

Key-words: forecasting, climate change, grassland, integral projection model, population model, statistical regularization, ridge regression

Introduction

Perhaps the greatest challenge for ecology in the 21st century is to forecast the impacts of environmental change (Clark et al. 2001, Petchey et al. 2015). Forecasts require sophisticated modeling approaches that fully account for uncertainty and variability in both ecological process and model parameters (Luo et al. 2011, but see Perretti et al. 2013). The increasing statistical sophistication of population models (Rees and Ellner 2009) makes them promising tools for predicting the impacts of environmental change on species persistence and abundance. But reconciling the scales at which population models are parameterized and the scales at which environmental changes play out remains a challenge (Clark et al. 2010, 2012, Freckleton et al. 2011, Queenborough et al. 2011). Most population models are built using demographic data from a single study site because tracking the fates of individuals is so difficult. The resulting models cannot be applied to the landscape and regional scales relevant to decision-making without information about how the fitted parameters respond to spatial variation in biotic and abiotic drivers (Sæther et al. 2007). The limited spatial extent of individual-level demographic datasets constrains our ability to use population models to address applied questions about the consequences of climate change.

Aggregate measures of population status, rather than individual performance, offer an intriguing alternative for modeling populations (Clark and Bjørnstad 2004, Freckleton et al. 2011). Population-level data cannot provide inference about demographic mechanisms, but might be sufficient for modeling future population states, especially since such data are feasible to collect across broad spatial extents (e.g., Queenborough et al. 2011). The choice between individual and population-level data involves a difficult trade-off: while individual-level data are necessary for mechanistic models, population-level data enable models that can be applied over greater spatial and temporal extents. An open question is how much forecasting skill is lost when we build models based on population rather than individual-level data.

To date, most empirical population modelers have relied on individual-level data, with few attempts to capitalize on population-level measures. An important exception was an effort by Taylor and Hastings (2004) to model the population growth rate of an invasive species to identify the best strategies for invasion control. They used a "density-structured" model where the state variable is a discrete density state rather than a continuous density measure. Such models do not require individual-level demographic data and can adequately describe population dynamics.

Building on Taylor and Hastings (2004), Freckleton et al. (2011) showed that density-structured models compare well to continuous models in theory, and Queenborough et al. (2011) provide empirical evidence that density-structured models are capable of reproducing population dynamics at landscape spatial scales, even if some precision is lost when compared to fully continuous models. However, previous tests of density-structured models have yet to assess their ability to forecast out-of-sample observations, and they have not included environmental covariates, which are necessary to forecast population responses to climate change.

Addressing climate change questions with models fit to population-level data is potentially problematic. Population growth (or decline) is the outcome of demographic processes such as survival, growth, and recruitment that occur at the level of individual plants. Climate can affect each demographic process in unique, potentially opposing, ways (Dalgleish et al. 2011). These unique climate responses may be difficult to resolve in statistical models based on population-level data where demographic processes are not identifiable. Futhermore, models based on aggregated data may reflect short-term effects of one vital rate more than others whose importance may only emerge over the long-term. For example, a one-year change in a plant species' cover or biomass might reflect individual growth or shrinkage, whereas the long-term trajectory of the population might be more influenced by recruitment. If important climate effects are missed because of the aggregation inherent in in population-level data, then population models built with such data will make uninformative or unreliable forecasts.

Here, we compare the forecasting skill of statistical and population models based on aggregated, population-level data with the skill of models based on individual-level data. We used a demographic dataset that tracks the fates of individual plants from four species over 14 years to build two kinds of single-species population models, traditional models using individual growth, survival, and recruitment data and alternative models based on basal cover. We simulated the models to answer two questions: (1) Can population models fit using aggregated individual-level data (percent cover) adequately capture density dependence to produce forecasts as skillful as those from models fit to demographic data? (2) Can population models fit using aggregated data adequately capture the influence of climate on population growth and, in turn, produce forecasts as skillful as those from models fit to demographic data? [THIS NEXT SENTENCE SEEMS OUT OF PLACE. CAN YOU MODIFY PREVIOUS PARAGRAPH TO ADDRESS DENSITY DEPENDENCE ALONG WITH CLIMATE EFFECTS? E.G. DENSITY DEPENDENCE MIGHT ACT MOST STRONGLY ON RECRUITMENT WHICH IS HARD TO DETECT IN A COVER BASED MODEL?] Both questions are motivated by the fact that the effects of intraspecific competition (density dependence) and interannual weather variability act at the level of the individual (Clark et al. 2011), meaning that models based on individual-level processes should generate more skillful forecasts.

Materials and Methods

Study site and data

Our demographic data come from a northern mixed grass prairie at the Fort Keogh Livestock and Range Research Laboratory near Miles City, Montana, USA ($46°$ 19' N, $105°$ 48' W). The dataset is available on Ecological Archives[1] (Anderson et al. 2011), and interested readers should refer to the metadata for a complete description. The site is 800 m above sea level and mean annual precipitation (1878-2009) is 334 mm, with most annual precipitation falling from April through September. The community is grass-dominated, and we focused on the four most abundant grass species: *Bouteloua gracilis* (BOGR), *Hesperostipa comata* (HECO), *Pascopyrum smithii* (PASM), and *Poa secunda* (POSE) (Fig. 1). *B. gracilis* is a warm-season perennial grass, whereas *H. comata*, *P. smithii*, and *P. secunda* are cool-season perennial grasses. All species typically begin growth in the early spring, reach maximum growth and flower in early to mid summer (May-June), and disperse seed in mid to late summer (July-September). [LAST SENTENCES GLOSSES OVER LIKELY BIG DIFFERENCES IN PHENOLOGY. JUST DESCRIBE TIMING OF GROWING SEASON INSTEAD, TO SET UP CLIMATE COVARIATES?]

[1] http://esapubs.org/archive/ecol/E092/143/

From 1932 to 1945 individual plants were identified and mapped annually in 44 1-m^2 quadrats using a pantograph. The quadrats were distributed among six pastures, each assigned a grazing treatment of light (1.24 ha/animal unit month), moderate (0.92 ha/aum), and heavy (0.76 ha/aum) stocking rates (two pastures per treatment). In this analysis we account for potential differences among the grazing treatments, but do not focus on grazing$\times$climate interactions. The annual maps of the quadrats were digitized and the fates of individual plants tracked and extracted using a computer program (Lauenroth and Adler 2008, Chu et al. 2014). The permanent quadrats have not been relocated, but their distribution in six different pastures means the data represent a broad spatial distribution for the study area. Daily climate data are available for the duration of the data collection period (1932 - 1945) from the Miles City airport, Wiley Field, 9 km from the study site.

We modeled each grass population based on two levels of data: individual and quadrat. The individual data is the "raw" data. For the quadrat-level, we data we simply sum individual basal cover for each quadrat by species. This is equivalent to a near-perfect census of quadrat percent cover because measurement error at the individual-level is small (Chu and Adler 2014). Based on these two datasets of 13 year-to-year transitions, we can compare population models built using individual-level data and aggregated, quadrat-level data. At the individual level we explicitly model three vital rates: growth, survival, and recruitment. At the quadrat level we model population growth as change in percent cover of quadrats with non-zero cover in year *t* and in year *t-1*, ignoring within-quadrat extirpation and colonization events because they are very rare in our time series ($N = 16$ and $N = 13$, respectively, across all species). Sample sizes for each species and vital rate model are shown in Table 1.

All R code and data necessary to reproduce our analysis is archived on GitHub as release v1.0[2] (http://github.com/atredennick/MicroMesoForecast/releases). That stable release will remain static as a record of this analysis, but subsequent versions may appear if we update this work. We

[2] *Note to reviewers*: so that v1.0 will be associated with the published version of the manuscript, we have released v0.2 to be associated with this review version.

have also deposited the v1.0 release on Dryad (*link here after acceptance*).

Stastical models of vital rates

At both levels of inference (individual and quadrat), the building blocks of our population models are vital rate regressions. For individual-level data, we fit regressions for survival, growth, and recruitment for each species. At the quadrat-level, we fit a single regression model for population growth. We describe the statistical models separately since they required different approaches. For both model types, we fit vital rate models with and without climate covariates. Models with climate effects contain five climate covariates that we chose *a priori* based on previous model selection efforts using these data (Chu et al. *in press*) and expert advice (Lance Vermeire, *personal communication*): "water year" precipitation at *t*-2 (lagppt); April through June precipitation at *t*-1 and *t* (ppt1 and ppt2, respectively) and April through June temperature at *t*-1 and *t* (TmeanSpr1 and TmeanSpr2, respectively), where *t*-1 to *t* is the transition of interest. We also include interactions among same-year climate covariates (e.g., ppt1 $\times$ TmeansSpr1), resulting in a total of seven climate covariates.

We fit all models using a hierarchical Bayesian approach. The models are fully descibed in Appendix A, so here we focus on the main process and the model likelihood. For the likelihood models, $\mathbf{y}^X$ is always the relevant vector of observations for vital rate *X* (*X* = *S*, *G*, *R*, or *P*) for survival, growth, recruitment, or population growth). For example, $\mathbf{y}^S$ is a vector of 0's and 1's indicating whether a genet survives from *t* to *t+1*, or not. All model parameters are species-specific, but we omit subscripts for species in model descriptions below to reduce visual clutter. For brevity, we only describe models with climate covariates included, but models without climate covariates are simply the models described below with the climate effects removed.

Vital rate models at the individual level We used logistic regression to model survival probability (s) of a genet:

$$\mathbf{y}^S \sim \text{Bernoulli}(\mathbf{s}) \tag{1}$$

$$\text{logit}[s(x, \mathbf{z}_t, w)] = \beta_{0,t} + \beta_{s,t}x + \beta_Q + \mathbf{z}_t'\boldsymbol{\beta}_c + \beta_{d,1}w + \beta_{d,2}(xw) \tag{2}$$

where x is the log of genet basal area, $\beta_{0,t}$ is a year specific intercept intercept, β_Q is the random effect of quadrat group location, $\beta_{s,t}$ is the year-specific slope parameter for size, **z** is a vector of *i* climate covariates specific to year *t*, $\boldsymbol{\beta}_c$ is a vector of fixed climate effects of length *i*, $\beta_{d,1}$ is the effect of intraspecific crowding experienced by the focal genet (w), and $\beta_{d,2}$ is a size by crowding (xw) interaction effect.

YOU NEED TO EXPLAIN HOW THE w's ARE CALCULATED. COPY TEXT FROM ADLER ET AL. 2010 or CHU & ADLER 2015.

We modeled growth as a Gaussian process describing genet size at time $t + 1$ as a function of size at t and climate covariates:

$$\mathbf{y}^G \sim \text{Normal}(\boldsymbol{\mu}, \sigma^2_{x,t}) \tag{3}$$

$$\mu(x, \mathbf{z}_t, w) = \beta_{0,t} + \beta_{s,t}x + \beta_Q + \mathbf{z}_t'\boldsymbol{\beta}_c + \beta_{d,1}w + \beta_{d,2}(xw) \tag{4}$$

where $\mu(x, \mathbf{z}_t, w)$ is log of predicted genet size at time *t* + 1, and all other parameters are as described for the survival regression. We capture non-constant error variance in growth by modeling the variance around the growth regression ($\sigma^2_{x,t}$) as a nonlinear function of predicted genet size:

$$\sigma^2_{x,t} = a \exp[b \times \mu(x, \mathbf{z}_t, w)] \tag{5}$$

where $\mu(x, \mathbf{z}_t, w)$ is log of predicted genet size predicted from the growth regression (Eq. 4), and *a* and *b* are constants.

Our data allows us to track new recruits, but we cannot assign a specific parent to new genets. Therefore, we model recruitment at the quadrat level. We assume the number of individuals, y^R, recruiting at time $t+1$ in quadrat q follows a negative binomial distribution:

$$y^R_{q,t+1} \sim \text{NegBin}(\lambda_{q,t+1}, \phi) \tag{6}$$

where λ is the mean intensity and ϕ is the size parameter. We define λ as a function of quadrat composition and climate in the previous year:

$$\lambda_{q,t+1} = C'_{q,t} \exp\left(\beta_{0,t} + \beta_Q + \mathbf{z}'_t \boldsymbol{\beta}_c + \beta_d \sqrt{C'_{q,t}}\right) \tag{7}$$

where C' is effective cover (cm^2) of the focal species in quadrat q, and all other terms are as in the survival and growth regressions. Effective cover is a mixture of observed cover (C) in the focal quadrat (q) and the mean cover across the entire group ($\bar{C}$) of Q quadrats in which q is located:

$$C'_{q,t} = pC_{q,t} + (1-p)\bar{C}_{Q,t} \tag{8}$$

where p is a mixing fraction between 0 and 1 that is estimated within the model.

Population model at the quadrat level The statistical approach used to model aggregated data depends on the type of data collected. We have percent cover data, which can easily be transformed to proportion data. An obvious choice for fitting a linear model to proportion data is beta regression because the support of the beta distribution is [0,1], not including true zeros or ones. However, when we used fitted model parameters from a beta regression in a quadrat-based population model, the simulated population tended toward 100% cover for all species. We therefore chose a modeling approach based on a truncated log-normal likelihood. The model for quadrat cover change from time t to $t+1$ is

$$\mathbf{y}^P \sim \text{LogNormal}(\mu(x, \mathbf{z}_t), \sigma^2)\text{T}[0,1] \tag{9}$$

$$\mu(x, \mathbf{z}_t) = \beta_{0,t} + \beta_{s,t}x + \beta_Q + \mathbf{z}'_t \boldsymbol{\beta}_c \tag{10}$$

where $\mu(x, \mathbf{z}_t)$ is the log of proportional cover in quadrat q at time $t + 1$, and all other parameters are as in the individual-level growth model (Eq. 4) except that *x* now represents log of proportional cover. The log normal likelihood includes a truncation (T[0,1]) to ensure that predicted values do not exceed 100% cover.

Model fitting and statistical regularization

Model fitting Our Bayesian approach to fitting the vital rate models required choosing appropriate priors for unknown parameters and deciding which, if any, of those priors should be hierarchical. We decided to fit models where all terms were fit by species. Within a species, we fit yearly size effects and yearly intercepts hierarchically where year-specific coefficients were drawn from global distributions representing the mean size effect and intercept. Quadrat random effects were also fit hierarchically, with quadrat offsets being drawn from distributions with mean zero and a shared variance term (independent Gaussian priors, Appendix A). Climate effects were modeled as independent covariates whose prior distributions were determined by statistical regularization (see **Statistical regularization: Bayesian ridge regression** below).

All of our analyses (model fitting and simulating) were conducted in R (R Core Team 2013). We used the 'No-U-Turn' MCMC sampler in Stan (Stan Development Team 2014a) to estimate the posterior distributions of model parameters using the package 'rstan' (Stan Development Team 2014b). We obtained posterior distributions for all model parameters from three parallel MCMC chains run for 1,000 iterations after discarding an initial 1,000 iterations. Such short MCMC chains may surprise readers more familiar with other MCMC samplers (i.e. JAGS or WinBUGS), but the Stan sampler is exceptionally efficient, which reduces the number of iterations needed to achieve convergence. We assessed convergence visually and made sure scale reduction factors for all parameters were less than 1.1. For the purposes of including parameter uncertainty in our population models, we saved the final 1,000 iterations from each of the three MCMC chains to be used as randomly drawn values during population simulation. We report the posterior mean, standard deviation, and 95% Bayesian Credible Intervals for every parameter of each model for each species in Appendix B.

Statistical regularization: Bayesian ridge regression For models with climate covariates, our objective is to model the response of our focal grass species to interannual variation in climate, even if those responses are weak. Therefore, we avoid selecting among models with all possible combinations of climate covariates, and instead use Bayesian ridge regression to regulate, or constrain, the posterior distributions of each climate covariate (Hooten and Hobbs 2015). Ridge regression is a specific application of statistical regularization that seeks to optimize model generality by trading off bias and variance. As the name implies, statistical regularization involves the use of a regulator that constrains an optimization. The natural regulator in a Bayesian application is the prior on the coefficient of interest. Each of our statistical models includes the effects of climate covariates via the term $\mathbf{z}_t'\boldsymbol{\beta}_c$ with prior $\boldsymbol{\beta}_c \sim \text{Normal}(\boldsymbol{\mu}_{\beta_c}, \sigma^2_{\beta_c})$. Since we standardized all climate covariates to have mean zero and variance one, we can set $\boldsymbol{\mu}_{\beta_c} = 0$ and let $\sigma^2_{\beta_c}$ serve as the regulator that can shrink covariates toward zero – the smaller the prior variance, the more the posteriors of $\boldsymbol{\beta}_c$ are shrunk toward zero, and the stronger the penalty (Hooten and Hobbs 2015).

To find the optimal penalty (i.e., optimal value of the hyperparameter $\sigma^2_{\beta_c}$), we fit each statistical model with a range of values for $\sigma^2_{\beta_c}$ and compared predictive scores from leave-one-year-out cross-validation. We performed the grid search over 24 values of $\sigma^2_{\beta_c}$, ranging from $\sigma^2_{\beta_c} = 0.01$ to $\sigma^2_{\beta_c} = 2.25$. For each statistical model and each species we fit $13 \times 24 = 312$ models (13 years to leave out for cross-validation and 24 values of $\sigma^2_{\beta_c}$) – a total of 4,992 models. We calculated the log pointwise predictive density (*lppd*) to score each model's ability to predict the left-out data. Thus, for training data y_{train} and held-out data y_{hold} at a given value of σ^2_θ across all MCMC samples $s = 1, 2, ..., S$ and all hold outs of data from year *t* to year *T*, and letting θ represent all unknowns, *lppd* is

$$\text{lppd}_{\text{CV}} = \sum_{t=1}^{T} \log_e \int [y_{t,hold}|\theta][\theta|y_{train}]d\theta, \tag{11}$$

and computed as

$$\sum_{t=1}^{T} \log_e \left(\frac{1}{S} \sum_{s=1}^{S} (y_{t,hold}|\theta_{ts}) \right). \tag{12}$$

We chose the optimal prior variance for each species-statistical model combination as the one that produced the highest *lppd* and then fit each species-statistical model combination using the full data set for each species and the optimal prior variance.

Population models

With the posterior distribution of the vital rate statistical models in hand, it is straightforward to simulate the population models. We used an Integral Projection Model (IPM) to model populations based on individual-level data (Ellner and Rees 2006) and a quadrat-based version of an individually-based model (Quadrat-Based Model, QBM) to model populations based on quadrat-level data. We describe each in turn.

Integral projection model We use a stochastic IPM (Rees and Ellner 2009) that includes the climate covariates from the vital rate statistical models. In all simulations we ignore the random year effects so that interannual variation is driven solely by climate. We fit the random year effects in the vital rate regressions to avoid over-attributing variation to climate covariates. Our IPM follows the specification of Chu and Adler (2015) where the population of species *j* is a density function $n(u_j, t)$ giving the density of sized-*u* genets at time *t*. Genet size is on the natural log scale, so that $n(u_j, t)du$ is the number of genets whose area (on the arithmetic scale) is between e^{u_j} and e^{u_j+du}. The density function for any size *v* at time $t + 1$ is

$$n(v_j, t+1) = \int_{L_j}^{U_j} k_j(v_j, u_j, \bar{w_j}(u_j)) n(u_j, t) \tag{13}$$

where $k_j(v_j, u_j, \bar{w_j})$ is the population kernel that describes all possible transitions from size u to v and $\bar{w_j}$ is a scalar representing the average intraspecific crowding experienced by a genet of size u_j and species j. The integral is evaluated over all possible sizes between predefined lower (*L*) and upper (*U*) size limits that extend beyond the range of observed genet sizes.

Since the IPM is spatially-implicit, we cannot calculate neighborhood crowding for specific genets (w_{ij}). Instead, we use an approximation ($\bar{w_j}$) that captures the essential features of neighborhood interactions (Adler et al. 2010). This approximation relies on a 'no-overlap' rule for conspecific genets to approximate the overdispersion of large genets in space (Adler et al. 2010).

The population kernel is defined as the joint contributions of survival (*S*), growth (*G*), and recruitment (*R*):

$$k_j(v_j, u_j, \bar{w_j}) = S_j(u_j, \bar{w_j}(u_j)) G_j(v_j, u_j, \bar{w_j}(u_j)) + R_j(v_j, u_j, \bar{w_j}), \tag{14}$$

which means we are calculating growth (*G*) for individuals that survive (*S*) from time *t* to *t+1* and adding in newly recruited (*R*) individuals of an average sized one-year-old genet for the focal species. Our stastical model for recruitment (*R*, described above) returns the number of new recruit produced per quadrat. Following previous work (Adler et al. 2012, Chu and Adler 2015), we assume that fecundity increases linearly with size ($R_j(v_j, u_j, \bar{w_j}) = e^{u_j} R_j(v_j, \bar{w_j})$) to incorporate the recruitment function in the spatially-implicit IPM.

We used random draws from the final 1,000 iterations, thinned by 10, from each of three MCMC chains to carry-through parameter uncertainty into our population models. At each time step, we randomly selected climate covariates from one of the 14 observed years. Then, we drew the full parameter set (climate effects and density-dependence fixed effects) from a randomly selected MCMC iteration. Relatively unimportant climate covariates (those that broadly overlap 0) will have little effect on the mean of the simulation results, but can contribute to their variation. Since our focus was on the contribution of density dependence and climate covariates to population states, we set the random year effects and the random group effects to zero.

Quad-based model To simulate our quad-based model (QBM), we simply iterate the quadrat-level statistical model (Eqs. 9-10). We use the same approach for drawing parameter values as described for the IPM. After drawing the appropriate parameter set, we calculate the mean response (log cover at *t+1* = μ_{t+1}) according to Eq. 10. We then make a random draw from a [0,1] truncated lognormal distribution with mean equal to μ_{t+1} from Eq. 10 and the variance estimate from the fitted model. We can then project the model forward by drawing a new parameter set (unique to climate year and MCMC iteration) at each timestep. As with the IPM, random year effects are ignored for all simulations.

Model validation

To test each model's ability to forecast population states, we made out-of-sample predictions using leave-one-year-out cross validation. For both levels of modeling and for models with and without climate covariates, we fit the vital rate models using observations from all years except one, and then used those fitted parameters in the population models to perform a one-step-ahead forecast for the year whose observations were withheld from model fitting. Within each observation year, several quadrats were sampled. We made predictions for each observed quadrat in each focal year, initializing each simulation with cover in the quadrat the previous year. Since we were making quadrat-specific predictions, we incorporated the group random effect on the intercept for both models. We repeated this procedure for all 13 observation years, making 100 one-step-ahead forecasts for each quadrat-year combination with parameter uncertainty included via random draw from the MCMC chain as described above. Random year effects were set to zero since year effects cannot be assigned to unobserved years.

This cross-validation procedure allowed us to compare accuracy and precision of the two modeling approaches (IPM versus QBM) with and without climate covariates. We first calculated the median predicted cover across the 100 simulations for each quadrat-year and then calculated forecast skill as the correlation (ρ) between forecasts and observations. We compared ρ between model types and within model types between models with and without climate covariates using a one-sided t test with adjusted degrees of freedom following Wilcox (2009) and standard errors calculated using the HC4 estimator of Cribari-Neto (2004). Statistical tests for differences between ρ were conducted using R functions from Ye et al. (2015).

Forecast horizons

An important feature of any forecasting model is the rate at which forecast skill declines as the time between an observation and a forecast increases; the so-called ecological forecast horizon (Petchey et al. 2015). To assess the forecast horizons of our models, we iniate the model with the population state at some time t and make sequential forecasts of the population at times $t + 1, t + 2, \ldots, t + T$ where T is the maximum number of years between the initial year and the final year of our observations. For example, if we initialize the model with percent cover in 1940, we are able to make five forecasts up to the year 1945. Models are not re-initialized with observations between years. Thus, in our current example, the model forecast for percent cover in 1941 has a forecast horizon of one year, the forecast in 1942 has a forecast horizon of two years, and so on. We ran these simulations for all model types (IPM with/without climate; QBM with/without climate) using mean parameter values for all possible initial years. Then, for a given forecast horizon, we averaged the correlation between forecasts and observations. Note that these forecasts are all made using in-sample data since we used model fits from the full data set. Nonetheless, these simulations offer insight into the differences between model forecast horizons.

Results

Both the IPM and QBM generated skillful one-step-ahead forecasts for out-of-sample observations, with an average correlation between predictions and observations (ρ) of 0.73 across all models and species (Fig. 2). IS THIS SKILLFUL FORECASTING OR IS THIS JUST STRONG AUTOCORRELATION IN COVER? Without climate covariates, the accuracy of forecasts from the IPM were not statistically greater than the accuracy of forecasts from the QBM (Fig. 2) and overall error was similar (mean absolute error; Fig. S1). With climate covariates, the best out-of-sample predictive model (highest *lppd*) for each species and vital rate typically resulted from highly constrained priors on the climate effects (Fig. S2). Thus, the posterior distributions of climate effects included in our models overlapped zero and generally shrunk toward zero, though for some species-vital rate combinations important effects (80% credible interval does not include zero) did emerge (Fig. 3).

Despite the weak climate effects, including climate covariates did increase the accuracy of forecasts for all species except *P. smithii* (Fig. 2). Increases in accuracy due to the inclusion of climate covariates were not significant ($P > 0.05$ for all comparisons of ρ between climate and no-climate forecasts within model types; Fig. 2). PREVIOUS SENTENCE READS AS INCONSISTENT WITH NEXT SENTENCE. However, the IPM with climate covariates for *P. secunda* did have significantly lower error than the IPM without climate covariates ($t_{(196)}$ = -1.84, P = 0.033; Fig. S2). In only one case were IPM forecasts significantly more accurate than the QBM (Fig. 2): forecast accuracy of *P. secunda* percent cover from an IPM with climate covariates was greater than the accuracy from the QBM with climate covariates ($t_{(195)}$ = 1.72, P = 0.043). WHAT IS THE DIFFERENCE BETWEEN ACCURACY (PREVIOUS SENTENCE) AND ERROR (NEXT SENTENCE)? Forecasts from IPMs with climate covariates had significantly lower error than equivalent QBM forecasts for *P. smithii* ($t_{(216)}$ = -3.49, P < 0.001) and *P. secunda* ($t_{(196)}$ = -1.83, P = 0.034) (Fig. S2). Results from all pairwise statistical tests are shown in Table S1.

The accuracy of both model's forecasts declined as the forecast horizon increased, but they did so at similar rates (Fig. 4). The only exception is for *P. secunda* where forecast accuracy appears to slightly increase with forecast horizon, after an initial decrease from a forecast horizon of one year (Fig. 4).

I TRIED TO BREAK THIS PARAGRAPH UP TO MAKE RESULTS EASIER TO FOLLOW

IS THIS NEXT PART OLD?

Discussion

Population models built using individual-level data allow inference on demographic processes, but they can only forecast future population states across the (typically limited) spatial extent of the observations. Population-level data are much easier to collect across broad spatial extents, so models built using such data offer an appealing alternative to traditional population models (Queenborough et al. 2011). However, density-structured models rely on the aggregation of individual-level data. Given that individuals, not populations, respond to intraspecific competition and weather (Clark et al. 2011), can models based on population-level metrics generate forecasts that are as skillful as those generated from models based on individual-level data? Are models based on population-level metrics as sensitive to climate forcing as models based on individual-level data?

Our comparison of a traditional, demographic population model without environmental forcing (the IPM) to an equivalent model inspired by density-structured models (the QBM) showed that, generally, IPM forecasts of out-of-sample plant population states were no more accurate than forecasts from the QBM (Fig. 2; 'no-climate' bars). We expected the IPM to out-perform the QBM since the IPM includes more mechanistic detail on the perennial plant life cycle, but this was not the case, at least when we ignored environmental forcing. Such a finding confirms theoretical (Freckleton et al. 2011) and empirical work (Queenborough et al. 2011) showing that density-structured models can be useful surrogates for demographic models when the goal is to estimate or forecast population states over large spatial extents. Likewise, the equivalency of forecast accuracy between the IPM and QBM held for all four species, when climate covariates are not included, extending the generality of previous findings (e.g., Taylor and Hastings 2004, Queenborough et al. 2011). THIS LAST SENTENCE SEEM REDUNDANT WITH PREVIOUS SENTENCES. . . AM I MISSING SOMETHING? COMBINE?

While the models did not differ in forecast accuracy when density-dependence was the only driver of population dynamics, we expected the inclusion of environmental forcing to reveal more differences between the models. We expected the IPM to outperform the QBM when we included climate covariates because interannual variation in weather can affect vital rates in different ways (Dalgleish et al. 2011). Thus, estimates of climate effects on plant population growth may be biased or non-identifiable when the underlying statistical model is fit using population-level data that integrates over the potentially unique climate responses of individual vital rates. However, we found that IPM forecast accuracy was only significantly higher than the QBM for one species, *P. secunda* (Fig. 2, ‘climate’ bars). This result likely stems from the relatively weak climate effects in the vital rate regressions (Fig. 3). If climate effects on vital rates are weak, then models based on percent cover or density should fair well compared to models based on individual-level data so long as density-dependence is adequately estimated (see Fig. 2 and Freckleton et al. 2011).

Our naive expectation was that a single, relatively important climate effect in a vital rate regression would cause a discernable difference between IPM and QBM forecast skill because individual-level responses may be less identifiable from percent cover data (Clark et al. 2012). We actually found the opposite. Survival or recruitment regressions for *B. gracilis*, *H. comata*, and *P. smithii* all included at least one relatively important climate effect, which we *a posterior* defined as a standardized coefficient whose 80% credible interval does not overlap zero (Fig. 3). Yet, only the forecast skill for *P. secunda* differed between the IPM and QBM (Fig. 2). Our explanation is that while at least one vital regression for the other three species included a relatively strong climate effect, only *P. secunda* had a vital rate regression with several smaller climate effects all trending in the same direction (Fig. 3). Thus, it appears that the estimation of several small but consistent effects, rather than a single relatively large effect, can lead to increased forecast skill.

The higher accuracy of the IPM with climate covariates for *P. secunda* highlights the advantage of contemporary model and variable selection approaches such as ridge regression and LASSO over techniques that would exclude "non-significant" effects from final models. Especially in a Bayesian forecasting framework, ridge regression allows researchers to include forcing variables that we know are changing (e.g., temperature and precipitation) in final models even if the estimated effects are relatively small or even deemed unimportant due to noisy data, out of which it is more difficult to estimate effects on interannual variation [REWERITE THIS SENTENCE, I DON'T KNOW HOW TO RESCUE IT CUZ I DON'T UNDERSTAND IT]. Likewise, if a species is truly unresponsive to a given climate variable, statistical regularization techniques will shrink the mean and variance of a covariate estimate to zero (Hooten and Hobbs 2015). Of course, no matter what model selection approach is adopted, a critical step is identifying the appropriate candidate covariates, which we attempted to do based on our knowledge of this semi-arid plant community. However, the climate covariates we chose required aggregating daily weather data over discrete time periods. It is possible that we did not choose the optimal time periods over which to aggregate. New methods using functional linear models (or splines) may offer a data-driven approach for identifying the appropriate time periods over which to aggregate to produce a tractable set of candidate climate variables (Teller et al. 2016).

We also expected IPM forecast accuracy to decline at a lower rate than the QBM as the forecast horizon increased. In principle, more mechanistic models should produce better predictions, especially under novel conditions (Evans 2012, Schindler and Hilborn 2015). In our case, the IPM explicitly models the influence of weather on recruitment and survival, effects that may be poorly represented in the QBM since recruitment and survival mainly affect small plants that contribute little to year-to-year changes in percent cover. Over time, of course, the addition and subtraction of small plants can have large effects on population growth, so explicitly modeling these effects could contribute to a longer forecast horizon. However, we found no evidence for a difference between the IPM and QBM forecast horizons (Fig. 4). Similar forecast horizons should be expected if both models adequately capture density dependence and environmental forcing is negligible, as is the case for our models (Fig. 2). In fact, only the forecast horizons for *P. secunda* noticeably differ (Fig. 4), and that is also the only species for which the IPM outperformed the QBM when we included climate effects.

In conclusion, we found that models based on individual-level demographic data generally failed to generate more skillful population forecasts than models based on population-level data. This finding runs counter to our expectations, but is consistent with recent theoretical (Freckleton et al. 2011) and empirical work (Queenborough et al. 2011). However, when we included climate covariates, we did achieve more skillful forecasts from the IPM for one species, *P. secunda*, that appears to respond consistently, but weakly, to several climate variables [THIS SENTENCE NEEDS WORK, BUT I AM GOING TO BED]. Thus, we conclude that models based on population-level data, rather than individual-level data, may be adequate for forecasting the state of plant populations for species that do not respond consistently or strongly to climate. Unfortunately, our analysis, where climate effects were relatively unimportant in vital rate regressions, did not allow us to sufficiently test our prediction that individual-level data is neccessary to generate skillful forecasts if different vital rates respond to climate in unique, potenially opposing, ways. Nonetheless, our results are encouraging for the use of easy-to-collect percent cover for forecasting the state of plant populations.

IS THE REST OLD? I STOPPED HERE.

Acknowledgments

This work was funded by the National Science Foundation through a Postdoctoral Research Fellowship in Biology to ATT (DBI-1400370) and a CAREER award to PBA (DEB-1054040). We thank the original mappers of the permanent quadrats in Montana and the digitizers in the Adler lab, without whom this work would not have been possible. Informal conversations with Stephen Ellner, Giles Hooker, Robin Snyder, and a series of meetings between the Adler and Weecology labs at USU sharpened our thinking. Brittany Teller provided comments that improved our manuscript. Compute, storage and other resources from the Division of Research Computing in the Office of Research and Graduate Studies at Utah State University are gratefully acknowledged.

Tables

Table 1: Description of data.

Species	Vital Rate Model	Num. Obs.	Num. Quadrats
B. gracilis	Growth	5670	29
	Survival	10102	33
	Recruitment	304	33
	Percent cover	281	29
H. comata	Growth	1990	16
	Survival	3257	18
	Recruitment	304	18
	Percent cover	171	17
P. smithii	Growth	8052	19
	Survival	11344	19
	Recruitment	304	19
	Percent cover	217	19
P. secunda	Growth	3018	18
	Survival	4650	18
	Recruitment	304	18
	Percent cover	197	18

Figures

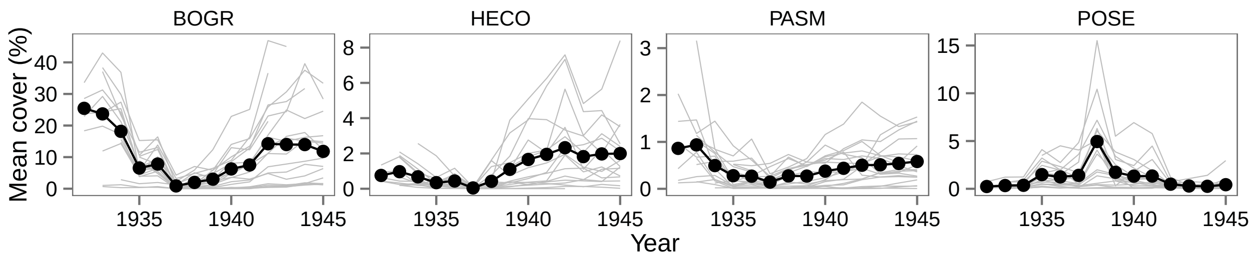
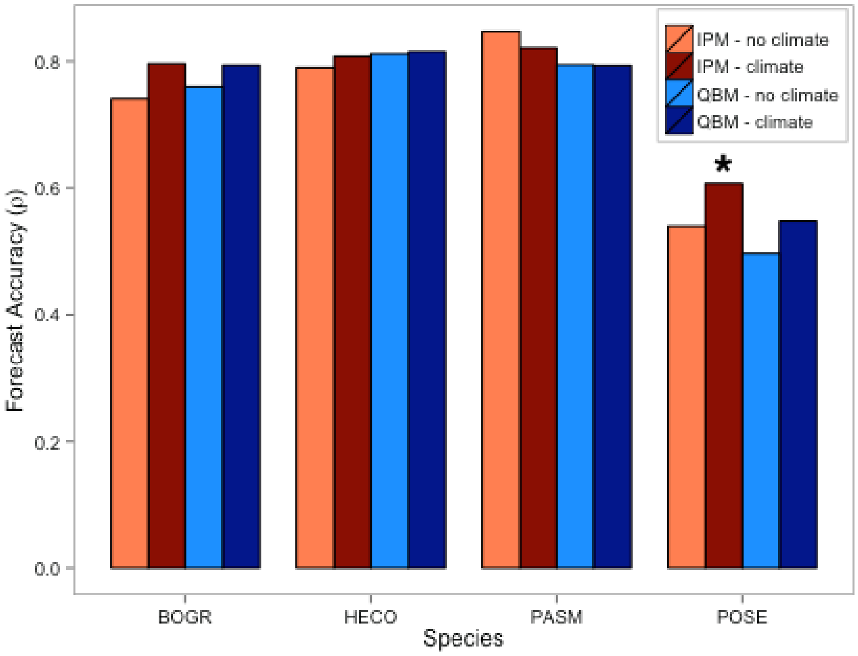


Figure 1: Time series of average percent cover over all quadrats for our four focal species: *Bouteloua gracilis* (BOGR), *Hesperostipa comata* (HECO), *Pascopyrum smithii* (PASM), and *Poa secunda* (POSE). Light grey lines show trajectories of individual quadrats. Note the different y-axis scales across panels.



Significance Tests

	BOGR	HECO	PASM	POSE
IPM - no climate vs. QBM - no climate	$P = 0.83$	$P = 0.68$	$P = 0.36$	$P = 0.14$
IPM - climate vs. QBM - climate	$P = 0.51$	$P = 0.61$	$P = 0.27$	***P = 0.043**
IPM - no climate vs. IPM - climate	$P = 0.22$	$P = 0.33$	$P = 0.81$	$P = 0.064$
QBM - no climate vs. QBM - climate	$P = 0.30$	$P = 0.48$	$P = 0.50$	$P = 0.053$

Figure 2: Comparison of one-step-ahead, out-of-sample forecast accuracy between the IPM and QBM models with and without the inclusion of climate covariates. Comparisons between equivalent IPM and QBM models indicate no significant difference in accuracy ($P > 0.05$ for all comparisons). Likewise, including climate covariates did not result in siginificantly higher forecast accuracy ($P > 0.05$ for all comparisons).

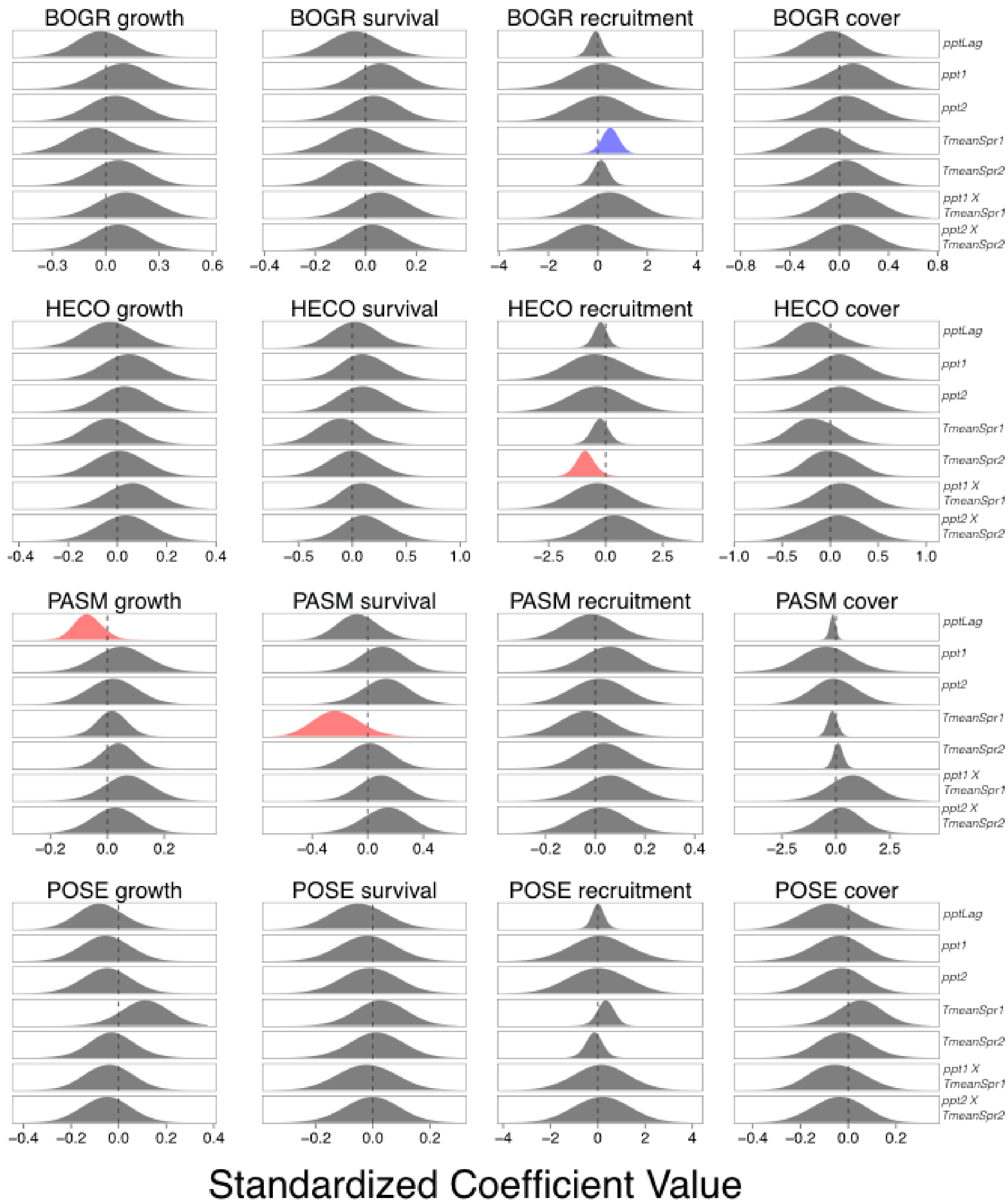


Figure 3: Posterior distributions of climate effects (β_C) for each species and vital rate statistical model. Since our priors were constrained via ridge-regression, we highlight climate effects whose 80% credible intervals do not overlap zero (red for negative coefficients, blue for positive coefficients). Kernel bandwidths of posterior densities were adjusted by a factor of 4 for visual clarity.

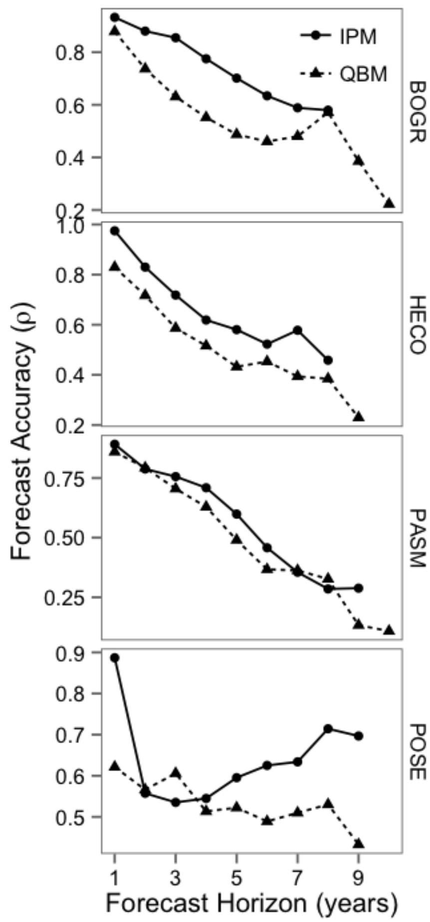


Figure 4: The forecast horizons for both models. Points show the average accuracy (ρ) across all forecasts at a given time horizon.

References

Adler, P. B., H. J. Dalgleish, and S. P. Ellner. 2012. Forecasting plant community impacts of climate variability and change: when do competitive interactions matter? Journal of Ecology 100:478–487.

Adler, P. B., S. P. Ellner, and J. M. Levine. 2010. Coexistence of perennial plants: An embarrassment of niches. Ecology Letters 13:1019–1029.

Anderson, J., L. Vermeire, and P. B. Adler. 2011. Fourteen years of mapped, permanent quadrats in a northern mixed prairie, USA. Ecology 92:1703.

Chu, C., and P. B. Adler. 2014. When should plant population models include age structure? Journal of Ecology 102:531–543.

Chu, C., and P. B. Adler. 2015. Large niche differences emerge at the recruitment stage to stabilize grassland coexistence. Ecological Monographs 85:373–392.

Chu, C., K. M. Havstad, N. Kaplan, W. K. Lauenroth, M. P. McClaran, D. P. Peters, L. T. Vermeire, and P. B. Adler. 2014. Life form influences survivorship patterns for 109 herbaceous perennials from six semi-arid ecosystems. Journal of Vegetation Science 25:947–954.

Clark, J. S., and O. N. Bjørnstad. 2004. Population time series: Process variability, observation errors, missing values, lags, and hidden states. Ecology 85:3140–3150.

Clark, J. S., D. M. Bell, M. H. Hersh, M. C. Kwit, E. Moran, C. Salk, A. Stine, D. Valle, and K. Zhu. 2011. Individual-scale variation, species-scale differences: Inference needed to understand diversity. Ecology Letters 14:1273–1287.

Clark, J. S., D. M. Bell, M. Kwit, A. Stine, B. Vierra, and K. Zhu. 2012. Individual-scale inference to anticipate climate-change vulnerability of biodiversity. Philosophical Transactions of the Royal Society B: Biological Sciences 367:236–246.

Clark, J. S., D. Bell, C. Chu, B. Courbaud, M. Dietze, M. Hersh, J. HilleRisLambers, I. Ibáñez, S. LaDeau, S. McMahon, J. Metcalf, J. Mohan, E. Moran, L. Pangle, S. Pearson, C. Salk, Z. Shen, D. Valle, and P. Wyckoff. 2010. High-dimensional coexistence based on individual variation: a synthesis of evidence. Ecological Monographs 80:569–608.

Clark, J. S., S. R. Carpenter, M. Barber, S. Collins, A. Dobson, J. A. Foley, D. M. Lodge, M. Pascual, R. Pielke, W. Pizer, C. Pringle, W. V. Reid, K. A. Rose, O. Sala, W. H. Schlesinger, D. H. Wall, and D. Wear. 2001. Ecological forecasts: an emerging imperative. Science (New York, N.Y.) 293:657–660.

Cribari-Neto, F. 2004. Asymptotic inference under heteroskedasticity of unknown form. Computational Statistics and Data Analysis 45:215–233.

Dalgleish, H. J., D. N. Koons, M. B. Hooten, C. A. Moffet, and P. B. Adler. 2011. Climate influences the demography of three dominant sagebrush steppe plants. Ecology 92:75–85.

Ellner, S. P., and M. Rees. 2006. Integral projection models for species with complex demography. The American naturalist 167:410–428.

Evans, M. R. 2012. Modelling ecological systems in a changing world. Philosophical transactions of the Royal Society of London. Series B, Biological sciences 367:181–190.

Freckleton, R. P., W. J. Sutherland, A. R. Watkinson, and S. A. Queenborough. 2011. Density-structured models for plant population dynamics. American Naturalist 177:1–17.

Hooten, M. B., and N. T. Hobbs. 2015. A guide to Bayesian model selection for ecologists. Ecological Monographs 85:3–28.

Lauenroth, W. K., and P. B. Adler. 2008. Demography of perennial grassland plants: Survival, life expectancy and life span. Journal of Ecology 96:1023–1032.

Luo, Y., K. Ogle, C. Tucker, S. Fei, C. Gao, S. LaDeau, J. S. Clark, and D. S. Schimel. 2011. Ecological forecasting and data assimilation in a data-rich era. Ecological Applications 21:1429–1442.

Perretti, C. T., S. B. Munch, and G. Sugihara. 2013. Model-free forecasting outperforms the correct mechanistic model for simulated and experimental data. Proceedings of the National Academy of Sciences of the United States of America 110:5253–5257.

Petchey, O. L., M. Pontarp, T. M. Massie, S. Kéfi, A. Ozgul, M. Weilenmann, G. M. Palamara, F. Altermatt, B. Matthews, J. M. Levine, D. Z. Childs, B. J. McGill, M. E. Schaepman, B. Schmid, P. Spaak, A. P. Beckerman, F. Pennekamp, and I. S. Pearse. 2015. The ecological forecast horizon, and examples of its uses and determinants. Ecology Letters 18:597–611.

Queenborough, S. A., K. M. Burnet, W. J. Sutherland, A. R. Watkinson, and R. P. Freckleton. 2011. From meso- to macroscale population dynamics: A new density-structured approach. Methods in Ecology and Evolution 2:289–302.

R Core Team. 2013. R: A language and environment for statistical computing.

Rees, M., and S. P. Ellner. 2009. Integral projection models for populations in temporally varying environments. Ecological Monographs 79:575–594.

Schindler, D. E., and R. Hilborn. 2015. Prediction, precaution, and policy under global change. Science 347:953–954.

Stan Development Team. 2014a. Stan: A C++ Library for Probability and Sampling, Version 2.5.0.

Stan Development Team. 2014b. Rstan: the R interface to Stan, Version 2.5.0.

Sæther, B. E., S. Engen, V. Grøtan, W. Fiedler, E. Matthysen, M. E. Visser, J. Wright, A. P. Møller, F. Adriaensen, H. Van Balen, D. Balmer, M. C. Mainwaring, R. H. McCleery, M. Pampus, and W. Winkel. 2007. The extended Moran effect and large-scale synchronous fluctuations in the size of great tit and blue tit populations. Journal of Animal Ecology 76:315–325.

Taylor, C. M., and A. Hastings. 2004. Finding optimal control strategies for invasive species: a density-structured model for Spartina alterniflora. Journal of Applied Ecology 41:1049–1057.

Teller, B. J., P. B. Adler, C. B. Edwards, G. Hooker, R. E. Snyder, and S. P. Ellner. 2016. Linking demography with drivers: climate and competition. Methods in Ecology and Evolution 7:171–183.

Wilcox, R. R. 2009. Comparing Pearson Correlations: Dealing with Heteroscedasticity and Non-normality. Communications in Statistics - Simulation and Computation 38:2220–2234.

Ye, H., R. J. Beamish, S. M. Glaser, S. C. H. Grant, C.-h. Hsieh, L. J. Richards, J. T. Schnute, and G. Sugihara. 2015. Equation-free mechanistic ecosystem forecasting using empirical dynamic modeling. Proceedings of the National Academy of Sciences 112:E1569–E1576.