Do we need detailed demographic data to forecast population responses to climate change?

Andrew T. Tredennick and Peter B. Adler

Andrew T. Tredennick (atredenn@gmail.com), Department of Wildland Resources and the Ecology Center, Utah State University, Logan, UT

Peter B. Adler, Department of Wildland Resources and the Ecology Center, Utah State University, Logan, UT

Abstract

Keywords: forecasting, climate change, grassland, integral projection model

Introduction

Population models are important tools for predicting the impacts of environmental change on species. 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). The major hurdle is that most population models, at least for plant species, are built using data from small, localized plots because parameterizing traditional population models requires tracking the fates of individuals. These models are difficult to scale up from the micro to meso-scales because the fitted parameters do not fully represent the spatial variation present at scales beyond that at which the data are collected (Sæther et al. 2007). At the same time, most demographic data is collected over short time spans. For example, the most common study duration in the COMPADRE matrix population model database is 4 years and only a few exceed 10 years (Salguero-Gómez et al. 2015). The constrained spatio-temporal extent of most demographic datasets reflects

the difficulty of collecting such data, but those constraints limit our ability to extrapolate population models. Thus, our ability to use population models to predict the consequences of climate change is limited when we rely on individual-level data.

Aggregate measures of individual plant performance, such as those typically collected as part of large-scale census efforts, offer an alternative to detailed demographic data for modeling populations (Clark and Bjørnstad 2004, Freckleton et al. 2011). Such population-level data will never match the precision of individual-level data, but it is more feasible to attain a broad coverage sample when collecting coarse-scale data. This presents a difficult trade-off: on the one hand, individual-level data leads to more reliable models; on the other hand, population-level data leads to models that will produce less precise predictions but can be applied over greater spatial and temporal extents. An open question is how well models based on population-level data compare to models based on individual-level data.

To date, relatively few studies have tried to model populations based on data other than detailed individual-level data. An important exception is an effort by Taylor and Hastings (2004) to model the population growth rate of an invasive species to investigate 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. Building on this work, Freckleton et al. (2011) showed that density-structured models compare well to continuous models in theory, and Queenborough et al. (2011) showed the application of such methods in a study on arable weeds. In particular, Queenborough et al. (2011) provide empirical evidence that density-structured models are capable of reproducing population dynamics, even if some precision is lost when compared to fully continuous models. Thus, population models based on coarse, population-level data show promise for producing ecological forecasts at landscape and regional scales (Queenborough et al. 2011). However, none of these models included environmental covariates.

Basing population models on aggregated individual-level data in a climate change context is

hampered by the fact that it is individuals that respond to climate, not populations (Clark et al. 2012). This fact puts us in uneasy proximity to an "ecological fallacy" where one deduces inference on the individual from statistical inference on the group (Piantadosi et al. 1988). For example, individual plants may respond positively to precipitation but a negative trend is observed at the population level due to increased competition among plants as they grow larger and consume more resources. Thus, it is important to ask the question: Can aggregated data be used to detect climate signals of the same sign and magnitude as individual-level data? If not, then building population models with climate covariates on aggregated data will lead to incorrect forecasts.

Here, we test the assumption that statistical and population models based on aggregated data can detect climate signals as wells as models based on individual-level data. We use a unique demographic dataset that tracks the fates of individual plants from four species over 14 years to build single-species population models, since those are often used tools for ecological forecasts and climate vulnerability assessments. We first fit population models with interannual variation in vital rates explained, in part, by climate covariates. We then perturb the climate covariates to test the sensitivities of species to climate change. By doing these analyses using both individual and aggregated forms of the same data, we provide a rigorous test of our hypothesis. We find that...

Materials and Methods

Study site and data

Our demographic data comes from the Fort Keogh Livestock and Range Research Laboratory in eastern Montana's northern mixed prairie near Miles City, Montana, USA (46° 19' N, 105° 48' W). The dataset is freely available on Ecological Archives (Anderson et al. 2011), and interested readers should refer to the metadata therein for a complete description. The site is about 800 m above sea level and mean annual precipitation (1878-2009) is 334 mm, with

most annual precipitation falling from April through September. The site is grass dominated and, for the purposes of our study, we focus on the four most abundant graminoid species: *Bouteloua gracilis* (BOGR), *Hesperostipa comata* (HECO), *Pascopyrum smithii* (PASM), and *Poa secunda* (POSE).

From 1932 to 1945 individual plants were identified and mapped annually in 44 1-m^2 quadrats using a pantograph. The quadrats were distributed in 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×climate interactions. The annual maps of the quadrats were digitized and the fates of individual plants tracked and extracted using a computer program. Daily climate data, which we aggregated into climate variables of interest, 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.

In this paper, we model populations 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 areal cover for each quadrat by species. This is equivalent to a perfect census of quadrat percent cover, so we do not need to consider measurement error. Based on these two datasets we can compare population models built using individual level data and aggregated quadrat level data.

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 models for survival, growth, and recruitment of new individuals for each species. At the quadrat level we fit analogous models of extinction probability, percent cover increase/decrease, and quadrat colonization for each species. We describe the statistical models separately since fitting the models required

different approaches at the individual and quadrat levels. All models contain four climate covariate that we chose *a priori*: fall through spring precipitation at t-1 and t-2 (ppt1 and ppt2, respectively) and mean spring temperature at t-1 and t-2 (TmeanSpr1 and TmeanSpr2, respectively), where t is the observation year.

We fit all models using a hierarchical Bayesian approach, which we describe in more detail below. However, for each vital rate statistical model we also define the likelihood model we use. For the likelihood models, Y is always the relevant vector of observations (e.g., whether a genet survived [1] or not [0] from year t to $t+1$).

Vital rate models: individual level We used logistic regression to model survival probability (S) of genet i from species j in quadrat group Q from time t to $t+1$:

$$\text{logit}(S_{ijQ,t}) = \gamma^S_{j,t} + \phi^S_{jQ} + \beta^S_{j,t} x_{ij,t} + \omega^S_j w_{ij,t} + \theta^S_{jk} C_{k,t} + \varepsilon^S_t \quad (1)$$

$$Y^S_{ijQ,t} \sim \text{Bernoulli}(S_{ijQ,t}) \quad (2)$$

where $x_{ij,t}$ is the log of genet size, $\gamma^S_{j,t}$ is a year-specific intercept, $\beta^S_{j,t}$ is the year-specific slope parameter for size, ϕ^S_{jQ} is the random effect of quadrat group location, and θ^S_k is the fixed parameter for the effect of the kth climate covariate at time t ($C_{k,t}$). We include density-dependence by estimating the effect of crowding on the focal individual by other individuals of the same species. ω is the effect of crowding and $w_{t,Q}$ is the crowding experienced by the focal individual at time t in quadrat group Q.

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

$$x_{ijQ,t+1} = \gamma^G_{j,t} + \phi^G_{jQ} + \beta^G_{j,t} x_{ij,t} + \omega^G_j w_{ij,t} + \theta^G_{jk} C_{k,t} \tag{3}$$

$$Y^G_{ijQ,t} \sim \text{Normal}(x_{ijQ,t+1}, \sigma_j) \tag{4}$$

where x is log genet size and all other parameters are as described for the survival regression.

Our data allows us to track new recruits, but we cannot assign a specific parent to new genets. So, for recruitment, we work at the quadrat level and model the number of new individuals of species j in quadrat q recruiting at time $t+1$ as a function of quadrat "effective cover" (A') in the previous year (t). Effective cover is a mixture of observed cover (A) in the focal quadrat (q) and the mean cover across the entire group ($\bar{A}$) of Q quadrats in which q is located:

$$A'_{jq,t} = p_j A_{jq,t} + (1 - p_j)\bar{A}_{jQ,t} \tag{5}$$

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

We assume the number of individuals, Y^R, recruiting at time $t+1$ follows a negative binomial distribution:

$$Y^R_{jq,t+1} \sim \text{NegBin}(\lambda_{jq,t+1}, \zeta) \tag{6}$$

where λ is the mean intensity and ζ is the size parameter. We define λ as:

$$\lambda_{jq,t+1} = A'_{jq,t} e^{(\gamma^R_{j,t} + \phi^R_{jQ} + \theta^R_{jk} C_{k,t} + \omega^R \sqrt{A'_{q,t}})} \tag{7}$$

where A' is effective cover (cm^2) of species j in quadrat q and all other terms are as in the survival and growth regressions.

Vital rate models: quadrat level At the quadrat level we defined three vital rates:

1. Probability of extirpation (S): the probability that, for a given species, a particular quadrat will go from non-zero cover at time t to zero cover at time $t+1$.
2. Cover change (G): the change in percent cover from time t to $t+1$ for a given species within a particular quadrat.
3. Probability of colonization (R): the probability that, for a give species, a particular quadrat will go from zero cover at time t to non-zero cover at time $t+1$.

We retain the abbreviations S, G, and R from the analogous processes at the individual level. The vital rate models at the quadrat level are all based on quadrat proportional cover. Also, at the quadrat level we do not need to explicitly include a density dependent term. Since we are modeling proportional cover, we essentially get density-dependence for "free" when proportional cover is included as a covariate. That is, density-dependence emerges because quadrats with low cover generally increase in cover whereas quadrats with high cover generally decrease in cover.

We modeled the probability of extirpation of species j in quadrat q from time t to $t+1$ as:

$$\text{logit}(S_{jq,t}) = \gamma_j^S + \phi_{jQ}^S + \beta_j^S x_{jq,t} + \theta_{jk}^S C_{k,t} \tag{8}$$

$$Y_{jq,t}^S \sim \text{Bernoulli}(S_{jq,t}) \tag{9}$$

where all parameters are as in Eq. 1 except that x is now quadrat proportional cover. Note, however, that we do not include year random effects on the intercept γ or the slope term β for quadrat proportional cover. The quadrat data is inherently more sparse than the individual data from which it is aggregated, and this is especially evident when modeling rare events like extirpation and colonization. Thus, when we tried to fit random year effects, those terms did not converge.

We modeled quadrat cover change (G) from time t to $t+1$ as:

$$\text{logit}(x_{jq,t+1}) = \gamma^G_{j,t} + \phi^G_{jQ} + \beta^G_{j,t} x_{jq,t} + \theta^S_{jk} C_{k,t} \tag{10}$$

where, in this case, we do include random year effects on the intercept γ and the slope term β. For cover change we had enough data for those terms to converge. Note that our model for quadrat cover change uses a logit transformation to link the expected cover at $t+1$ ($x_{jq,t}$) to the linear predictors. We do so because during model fitting we use a beta likelihood since the data, proportional cover, is beta distributed. The beta likelihood requires shape (ρ) and rate (η) parameters that can be calculated using moment-matching:

$$\rho_{jq,t+1} = x_{jq,t+1}\tau_j \tag{11}$$

$$\eta_{jq,t+1} = (1 - x_{jq,t+1})\tau_j \tag{12}$$

with likelihood:

$$Y^G_{jq,t+1} \sim \text{Beta}(\rho_{jq,t+1}, \eta_{jq,t+1}). \tag{13}$$

Finally, we modeled probability of colonization in quadrat q by species j from time t to $t+1$ as:

$$\text{logit}(R_{jq,t}) = \gamma^R_j + \phi^R_{jQ} + \theta^R_{jk} C_{k,t} \tag{14}$$

$$Y^R_{jq,t} \sim \text{Bernoulli}(R_{jq,t}). \tag{15}$$

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 prior should be hierarchical. We decided to fit models where all terms except climate covariates were fit by species, while the climate covariates were fit hierarchically where species-specific coefficients were drawn from a shared 'global' coefficient distribution. We did so for two reasons: (1) the four focal species are all perennial grasses that we expect to respond similarly to climate covariates, and (2) convergence of climate effects at the quadrat level was much easier to achieve when we modeled these terms hierarchically, allowing them to "share" statistical strength via partial pooling (Gelman and Hill 2007). So, climate effects were modeled as:

$$\theta_{jk} \sim \text{Normal}(\bar{\theta_k}, \sigma_k) \tag{16}$$

where $\bar{\theta_k}$ is the interspecific effect of the kth climate covariate.

We used uninformative priors for all unknown parameters, specifically:

$$\boldsymbol{\gamma}, \boldsymbol{\beta}, \bar{\boldsymbol{\theta}} \sim \text{Normal}(0, 1e^{-6}) \tag{17}$$

$$\boldsymbol{\phi} \sim \text{Normal}(0, \boldsymbol{\sigma}_{\boldsymbol{\phi}}) \tag{18}$$

$$\boldsymbol{\sigma}_{\boldsymbol{\phi}} \sim e^{(\text{Gamma}(2,0.5))} \tag{19}$$

$$\boldsymbol{\sigma}_{\boldsymbol{\theta}}, \boldsymbol{\sigma}_{\boldsymbol{\gamma}}, \boldsymbol{\sigma}_{\boldsymbol{\beta}}, \boldsymbol{\tau}, \boldsymbol{\zeta} \sim \text{Gamma}(0.001, 0.001) \tag{20}$$

All of our analyses (model fitting and simulating) were conducted in R (Team 2013). We used the MCMC sampler in JAGS (Plummer 2003) to estimate the posterior distributions of model parameters and the package 'r2jags' (Su and Yajima 2012) to connect R to JAGS.

We obtained posterior distributions for all model parameters from three parallel MCMC chains run for 50,000 iterations, after discarding an initial 50,000 iterations. We assessed convergence visually and using the Gelman and Rubin (1992) diagnostic in the R package 'coda' (Plummer et al. 2006). Scale reduction factors for all parameters were less than 1.02, indicating convergence. For the purposes of introducing stochasticity in our population models, we saved the final 1,000 iterations from each chain for all parameters to be used as randomly drawn values during population simulation.

We assessed the statistical importance of the climate covariates included the final vital rate regressions by comparing the residual deviance of models with climate covariates and temporal random effects, climate covariates only, and temporal random effects only. When a model includes climate covariates, this comparison shows the relative contribution of the climate covariates in explaining the total interannual variability (Adler et al. 2012).

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 and an quadrat based version of an individually-based model (Quadrat-Based Model, QBM) to model populations based on quadrat level data. Both models take the general form:

$$N_{t+1} = S \times G + R. \tag{21}$$

So, at each time step in a simulation, we use the survival regression to determine if each genet lives or not (if each quadrat remains occupied or not), the growth regression to determine size changes of surviving individuals (cover change of occupied quadrats), and the recruitment regression to determine the number of new recruits (if a quadrat is colonized or not). We first use one-step-ahead forecasts to assess each model's ability to reproduce observed cover

changes. Then we use the models to analyze the effect of potential climate changes on population size over long time scales.

We used random draws from the final 1,000 iterations from each of 3 MCMC chains to introduce stochasticity into our population models. At each time step, we first randomly selected climate covariates from one of the 14 observed years and also randomly drew one set of random year effects for each vital rate regression. Climate years and random year effects were drawn independently because they were uncorrelated. Then, we drew the full parameter set (specific random year effects, climate effects, fixed effects) from a randomly selected MCMC iteration. We selected random year effects and the climate year independently.

Results

Climate effects on vital rates

At the individual level we fit vital rate models that included climate covariates and random year effects (full model), just climate covariates (climate model), and no climate or random year effects (constant model). We used the metric from Adler et al. (2012) to quantify the contribution of climate covariates to explaining interannual variability in vital rates based on comparisons of model deviance: (Climate model - Constant Model)/(Full model - Constant model). At the quadrat level, we could only fit random year effects for the growth regression. This is, in fact, one of the limitation of census-based data: local extirpation and colonization events are likely to be rare, making it difficult to fit complex statistical models. So, in this section we focus on growth only.

Climate covariates improved the growth regressions at both levels, with a slightly greater contribution at the individual level (Figure 1). Parameter estimates were generally consistent in terms of sign and magnitude between the individual level and the quadrat level for each species and climate covariate (Figure 2). There is no clear pattern suggesting that parameters

estimated using individual level data are more certain, as evidenced by credible interval widths, than estimates based on quadrat data.

Climate effects on population dynamics

Discussion

We sought to test the assumption that the sensitivities of plant populations to climate variables can be detected equally well using either individual level data or population level data. This is an important question to answer because population models are key tools for predicting the consequences of global climate change. However, they can be of limited use when built on data from a small subset of a population in space or time. If population level data (i.e., some aggregated form of individual level data) can be used to detect climate effects on population dynamics, then we would have a cheaper and easier option for data collection over relatively large temporal and spatial extents (*e.g.* Freckleton et al. 2011).

1. Climate detection
2. Limitations:

- Sample size at quadrat level
- are these species inherently sensitive to climate (compare to Chu and Adler in review – say why we chose Montana site – most sensitive to climate variables)
- we know recruitment is important, and this is something that may be easily missed at quad level

3. Implications

Figures

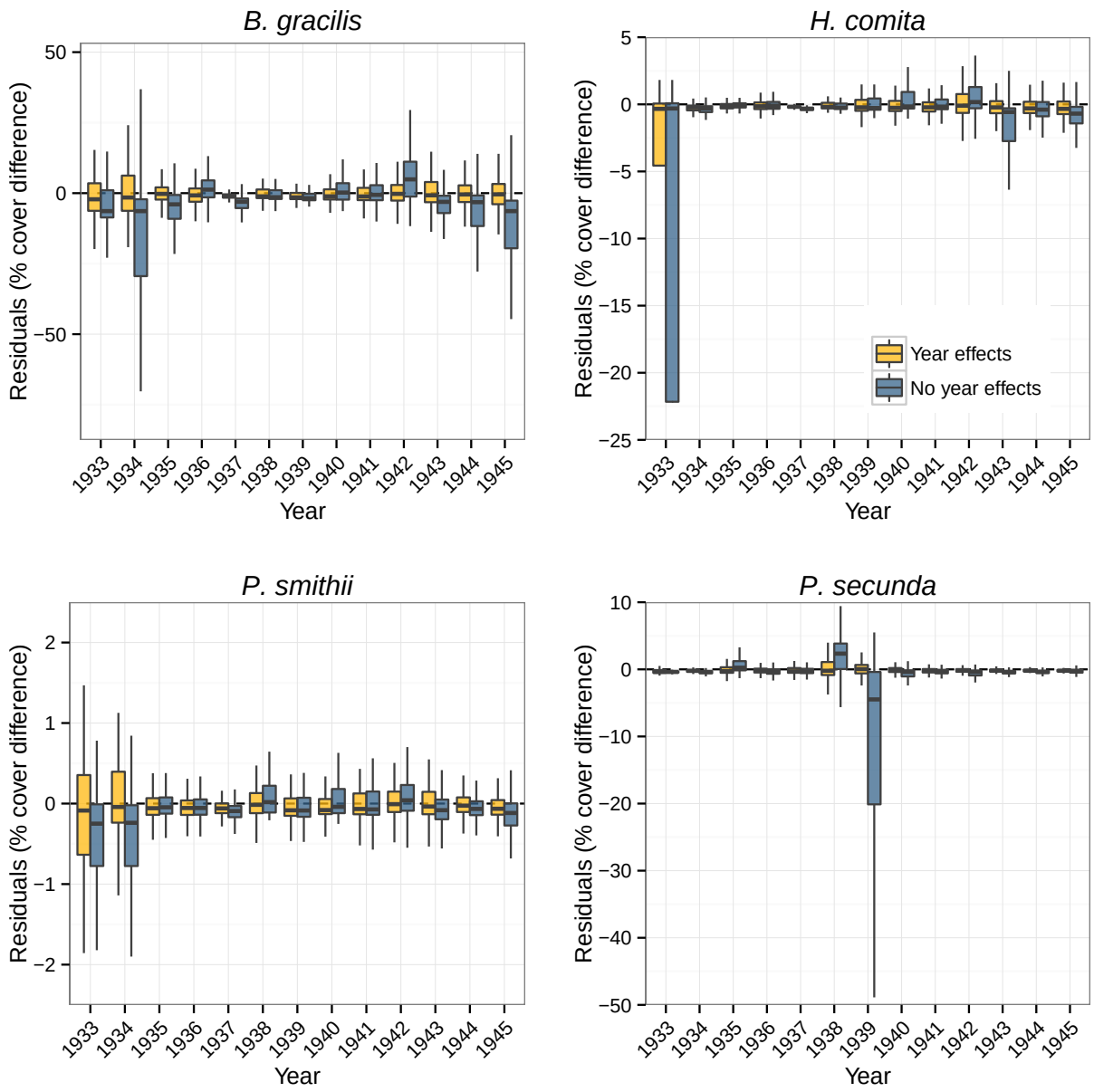

Figure 1: Boxplots of model residuals for one-step-ahead forecasts at each observation year. Each one-step forecast was simulated `r nSims` times. Note that the y-axes vary across panels.

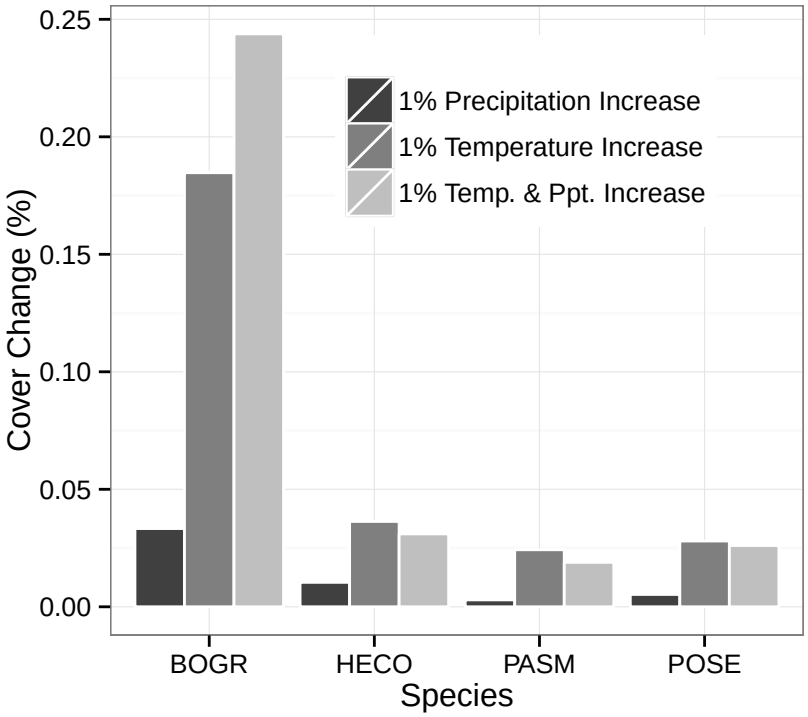

Figure 2: Simulated changes species' median cover caused by a 1% increase in observed precipitation or temperature. We used the median of simulated results because the average was highly influence by some extreme outliers.

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.

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

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. Kwit, A. Stine, B. Vierra, and K. Zhu. 2012. Individual-scale inference to anticipate climate-change vulnerability of biodiversity.

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.

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.

Gelman, A., and J. Hill. 2007. Data analysis using regression and multilevel/hierarchical models. Page xxii, 625p.

Gelman, A., and D. B. Rubin. 1992. Inference from Iterative Simulation Using Multiple Sequences.

Piantadosi, S., D. P. Byar, and S. B. Green. 1988. The Ecological Fallacy. American Journal of Epidemiology 127:893–904.

Plummer, M. 2003. JAGS: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling. Pages 20–22 *in* Proceedings of the 3rd international workshop on distributed statistical computing (dSC 2003). march.

Plummer, M., N. Best, K. Cowles, and K. Vines. 2006. CODA: Convergence Diagnosis and Output Analysis for MCMC. R News 6:7–11.

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.

Salguero-Gómez, R., O. R. Jones, C. R. Archer, Y. M. Buckley, J. Che-Castaldo, H. Caswell, D. Hodgson, A. Scheuerlein, D. A. Conde, E. Brinks, H. de Buhr, C. Farack, F. Gottschalk, A. Hartmann, A. Henning, G. Hoppe, G. Römer, J. Runge, T. Ruoff, J. Wille, S. Zeh, R. Davison, D. Vieregg, A. Baudisch, R. Altwegg, F. Colchero, M. Dong, H. de Kroon, J.-D. Lebreton, C. J. E. Metcalf, M. M. Neel, I. M. Parker, T. Takada, T. Valverde, L. A. Vélez-Espino, G. M. Wardle, M. Franco, and J. W. Vaupel. 2015. The compadrePlant Matrix Database: an open online repository for plant demography. Journal of Ecology 103:202–218.

Su, Y., and M. Yajima. 2012. R2jags: A Package for Running jags from R. http://CRAN. R-project. org/package= R2jags.

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.

Team, R. 2013. R Development Core Team. R: A Language and Environment for Statistical Computing.