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

Andrew T. Tredennick[*,a], Peter B. Adler[a]

[a]*Department of Wildland Resources and the Ecology Center, Utah State University, Logan, UT 84322, USA*

Abstract

Forecasts of populations under future climate change are constrained by the spatial and temporal scales at which we can gather data. For plants, ecologists often require individual level data to build population models; data that is difficult and expensive to collect at spatial scales greater than one square meter. Models built using such data are difficult to scale-up to spatial scales relevant to land management decisions and policy intervention. A possible route forward are density-structured models based on coarse-scale census data, which are much easier to collect over large spatial extents. However, to use coarse-scale data in a climate change context requires the assumption that the climate signals in population dynamics identified using individual-level data can also be identified when individual-level data are averaged over. Here we use a longterm dataset to model population dynamics using individual-level and plot-level data.

Key words: ecological forecasts, population models, integral projection model, climate change, grassland,

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 (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). Thus, our ability to

[*]Corresponding author
  Email address: `atredenn(at)gmail.com` (Andrew T. Tredennick)

December 21, 2014

17 use population models to predict the consequences of climate change is limited when we rely on

18 individual-level data.

19 In contrast to individual-level data, population-level data arising from large-scale census efforts

20 Recently, Freckleton et al. (2011), building on work by Taylor and Hastings (2004), have proposed

21 density-structured population models that focus on the transition of populations among discrete

22 states, rather than the traditional approach of modeling the transitions of individuals. Such an

23 approach could be extremely valuable because the data needed to parameterize density-structured

24 population models is much easier, and less costly, to collect.

25 For example, using a density-structured approach, one could build a population model using a

26 time series of annual plot-based censuses of species percent cover. However, a major assumption

27 of the density-structured approach is that the aggregate dynamics of the population observed at

28 coarse spatial resolution faithfully represent, and correspond to, the fates of individual plants.

29 In other words, using a density-structured approach requires a leap of faith that important

30 covariates (e.g., climate variables) at the level of the individual are captured adequately at the

31 population level. If we seek to forecast the impacts of climate change on plant populations, then

32 clearly this assumption requires testing.

Materials and Methods

34 *Study site and data*

35 Our demographic data comes from the Fort Keogh Livestock and Range Research Laboratory

36 in eastern Montana's northern mixed prairie near Miles City, Montana, USA (46 deg. 19' N,

37 105 deg 48' W). The dataset is freely available on Ecological Archives (CITE), and interested

38 readers should refer to the metadata therin for a complete description. The site is about 800

39 m above sea level and mean annual precipitation (1878-2009) is 334 mm, with most annual

40 precipitaion falling from April through September (76). The site is grass dominated and, for the

41 purposes of our study, we focus on the four most abundant graminoid species: *Bouteloua gracilis*,

42 *Hesperostipa comata*, *Pascopyrum smithii*, and *Poa secunda*.

43 From 1932 to 1945 individual plants were identified and mapped annualy in 44 1-m2 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 analagous models of extinction probability, percent cover increase/decrease, and quadrat colonization for each species. We describe the statistical models seprately 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 $t + 1$).

3

71 *Vital rate models: individual level.* We used logistic regression to model survival probability (S)

72 of genet i from species j in quadrat group Q from time t to $t+1$:

$$\text{logit}(S_{ijQ,t}) = \gamma_{j,t}^S + \phi_{jQ}^S + \beta_{j,t}^S x_{ij,t} + \omega_j^S w_{ij,t} + \theta_{jk}^S C_{k,t} + \varepsilon_t^S \tag{1}$$

$$Y_{ijQ,t}^S \sim \text{Bernoulli}(S_{ijQ,t}) \tag{2}$$

73 where $x_{ij,t}$ is the log of genet size, $\gamma_{j,t}^S$ is a year-specific intercept, $\beta_{j,t}^S$ is the year-specific slope

74 parameter for size, ϕ_{jQ}^S is the random effect of quadrat group location, and θ_k^S is the fixed

75 parameter for the effect of the kth climate covariate at time t ($C_{k,t}$). We include density-

76 dependence by estimating the effect of crowding on the focal individual by other individuals of

77 the same species. ω is the effect of crowding and $w_{t,Q}$ is the crowding experienced by the focal

78 individual at time t in quadrat group Q.

79 We modeled growth as gaussian process describing genet size at time $t+1$ as a function of size

80 at t and climate covariates:

$$x_{ijQ,t+1} = \gamma_{j,t}^G + \phi_{jQ}^G + \beta_{j,t}^G x_{ij,t} + \omega_j^G w_{ij,t} + \theta_{jk}^G C_{k,t} \tag{3}$$

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

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

82 Our data allows us to track new recruits, but we cannot assign a specific parent to new genets.

83 So, for recruitment, we work at the quadrat level and model the number of new individuals of

84 species j in quadrat q recruiting at time $t+1$ as a function of quadrat "effective cover" (A') in

85 the previous year (t). Effective cover is a mixture of observed cover (A) in the focal quadrat (q)

86 and the mean cover across the entire group ($\bar{A}$) of Q quadrats in which q is located:

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

87 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 analagous 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 gerenally 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:

$$\text{logit}(S_{jq,t}) = \gamma^S_j + \phi^S_{jQ} + \beta^S_j x_{jq,t} + \theta^S_{jk} C_{k,t} \tag{8}$$

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

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:

$$\text{logit}(x_{jq,t+1}) = \gamma_{j,t}^{G} + \phi_{jQ}^{G} + \beta_{j,t}^{G} x_{jq,t} + \theta_{jk}^{S} C_{k,t} \tag{10}$$

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 caclulated 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:

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

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

$$\text{logit}(R_{jq,t}) = \gamma_{j}^{R} + \phi_{jQ}^{R} + \theta_{jk}^{R} C_{k,t} \tag{14}$$

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

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 heierarchically 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 perrennial 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 strenght via partial pooling (Gelman and Hill 2007). So, climate effects were modeled as:

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

$$\bar{\theta}_k \sim \text{Normal}(0, 1e-6) \tag{17}$$

We used uninformative priors for all unknown parameters, specifically:

$$\boldsymbol{\gamma}, \boldsymbol{\beta} \sim \text{Normal}(0, 1e-6) \tag{18}$$

$$\boldsymbol{\phi} \sim \text{Normal}(0, \boldsymbol{\sigma}_\phi) \tag{19}$$

$$\boldsymbol{\sigma}_\phi \sim e(\text{Gamma}(2, 0.5)) \tag{20}$$

$$\boldsymbol{\sigma}_\theta, \boldsymbol{\sigma}_\gamma, \boldsymbol{\sigma}_\beta \sim \text{Gamma}(0.001, 0.001) \tag{21}$$

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

7

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.

Population models

With the posterior distribution of the vital rate statistical models in hand, it is straightforward
to simulate the population models.

Results

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
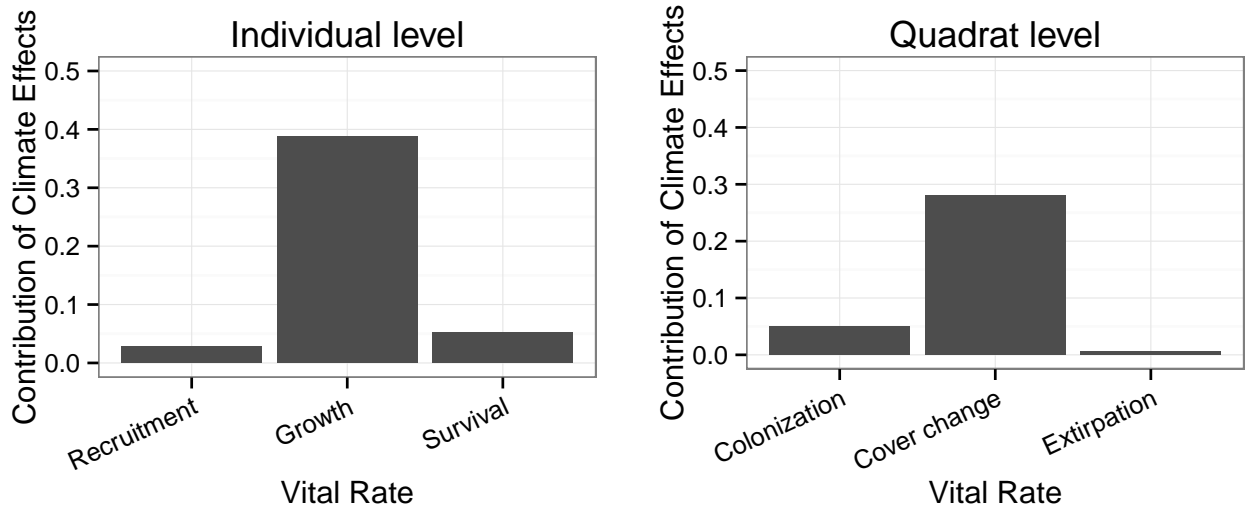explaining the total interannual variability (Adler et al. 2012).

Figure 1: The proportion of interannual variability in vital rates explained by the climate covariates. The
contribution for growth is defined as: (Climate model - Constant Model)/(Full model - Constant model). The
contribution for survival and colonization, where we could not estimate a full model with year random effects at
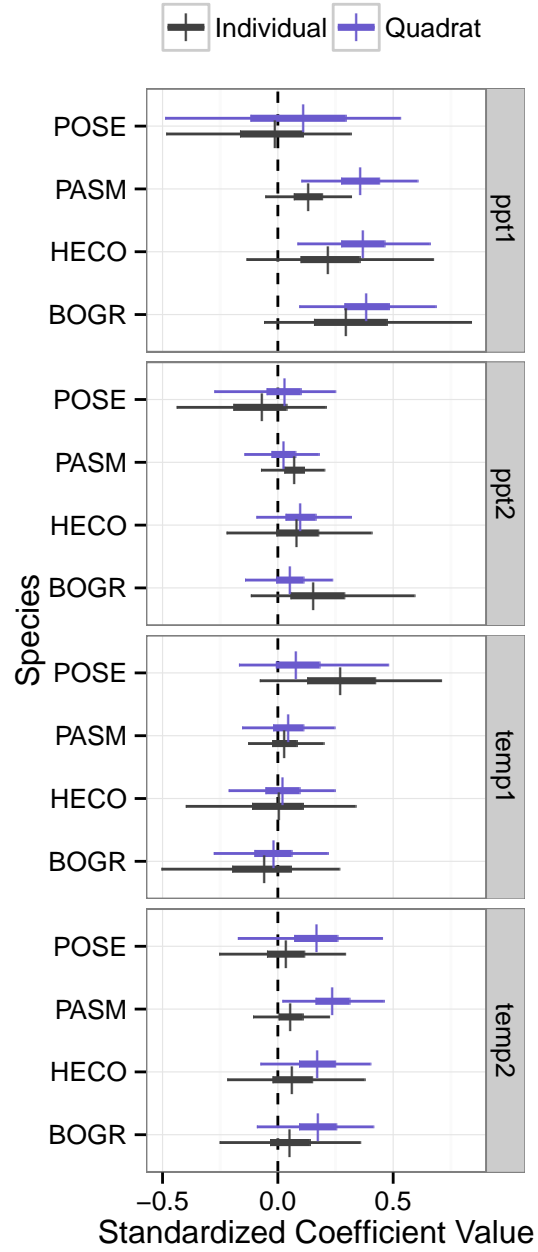the quadrat level, is defined as: (Constant Model - Climate Model)/Constant Model.

8

Figure 2: Posterior means (vertical ticks), 75% credible intervals (heavy lines), and 95% credible intervals (light lines) of climate effects on growth at both levels of inferences. The dashed vertical line is at 0, indicating no effect. Horizontal line at 0 indicates perfect agreement between mean observed cover in that year and the model predictions.
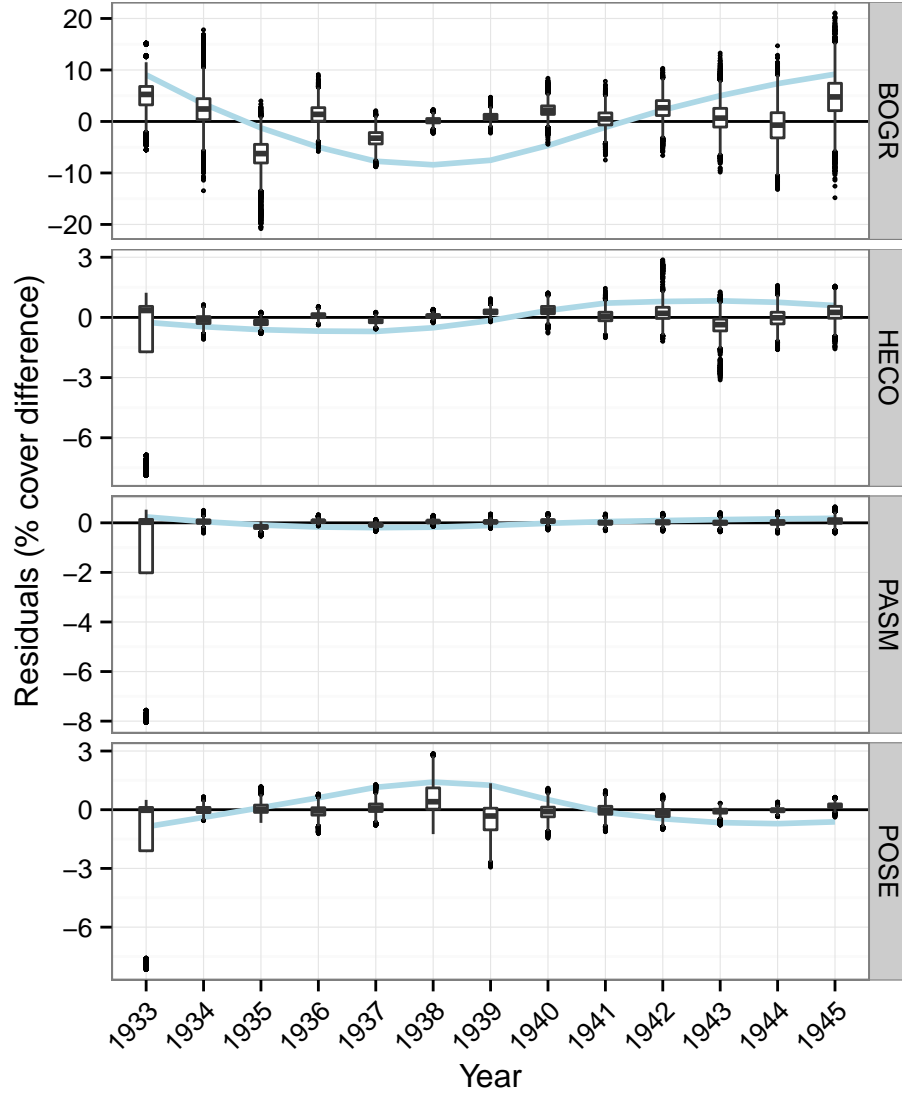
Figure 3: Boxplots of model residuals for one-step-ahead forecasts at each observation year. Each one-step forecast was simulated `r nSims` times. Note that the y-axes vary across panels. The light blue line shows the difference between the observed-year percent cover and the average cover observed across all years. The models tend to underpredict and perform poorly when observed cover in a given year is a large deviant from the mean.

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.

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.

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.

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.

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

179 Computing.