

# 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](mailto: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 well 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<sup>2</sup> 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.

### Statistical 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,  $\mathbf{Y}$  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  $t + 1$ :

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

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

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 parameter for size,  $\phi_{jQ}^S$  is the random effect of quadrat group location, and  $\theta_k^S$  is the fixed parameter for the effect of the  $k$ th 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.  $\omega$  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:

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

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

117 where  $x$  is log genet size and all other parameters are as described for the survival regression.  
 118 Our data allows us to track new recruits, but we cannot assign a specific parent to new genets.  
 119 So, for recruitment, we work at the quadrat level and model the number of new individuals of  
 120 species  $j$  in quadrat  $q$  recruiting at time  $t + 1$  as a function of quadrat “effective cover” ( $A'$ ) in  
 121 the previous year ( $t$ ). Effective cover is a mixture of observed cover ( $A$ ) in the focal quadrat  
 122 ( $q$ ) and the mean cover across the entire group ( $\bar{A}$ ) of  $Q$  quadrats in which  $q$  is located:

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

123 where  $p$  is a mixing fraction between 0 and 1 that is estimated within the model.  
 124 We assume the number of individuals,  $Y^R$ , recruiting at time  $t + 1$  follows a negative binomial  
 125 distribution:

$$Y_{jq,t+1}^R \sim \text{NegBin}(\lambda_{jq,t+1}, \zeta) \quad (6)$$

126 where  $\lambda$  is the mean intensity and  $\zeta$  is the size parameter. We define  $\lambda$  as:

$$\lambda_{jq,t+1} = A'_{jq,t} e^{(\gamma_{j,t}^R + \phi_{jQ}^R + \theta_{jk}^R C_{k,t} + \omega_j^R \sqrt{A'_{q,t}})} \quad (7)$$

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

## 129 **Vital rate models: quadrat level**

130 The statistical approach used to model vital rates  
 131 using aggregated data depends on the type of data collected. In our case, and as is often  
 132 the case with census data, we have percent cover data (which can easily be transformed to  
 133 proportion data, of course). We first considered fitting three vital rate models analagous  
 134 to those we fit at the individual level: one for probability of extirpation within a quadrat  
 135 (analagous to survival), one for cover change within a quadrat (analagous to growth), and  
 136 one for probability of colonization within a quadrat (analagous to recruitment). However,  
 137 within-quadrat extirpation and colonization events were rare in our time series ( $N = 9$  and  
 138  $N = 10$ , respectively). Given the broad spatial distribution of the quadrats we are studying,  
 139 it is safe to assume that these events are in fact rare enough to be ignored for our purposes.  
 140 So we constrained our statistical modeling of vital rates at the population level to change in  
 141 percent cover within quadrats. For most of the remaining discussion of statistical modeling  
 142 we refer to proportion data, which is simply percent data divided by 100.

143 An obvious choice for fitting a linear model to proportion data is beta regression because the  
 144 support of the beta distribution is  $[0,1]$ , not including true zeros or ones. However, when we  
 145 used fitted model parameters from a beta regression in quadrat-based population model the  
 146 simulated population tended toward 100% cover for all species. We therefore chose a more  
 147 constrained modeling approach based on a truncated log-normal likelihood. The model for  
 quadrat cover change ( $G$ ) from time  $t$  to  $t + 1$

$$x_{jq,t+1} = \gamma_{j,t}^G + \phi_{jQ}^G + \beta_{j,t}^G x_{jq,t} + \theta_{jk}^S C_{k,t} \quad (8)$$

$$Y_{jq,t+1}^G \sim \text{LogNormal}(x_{jq,t+1}, \tau j) T[0, 1] \quad (9)$$

148 where  $x_{jq,t}$  is the log of species'  $j$  proportional cover in quadrat  $q$  at time  $t$  and all other  
 149 parameters are as in the individual-level growth model (Eq. #). The log normal likelihood  
 150 includes a truncation ( $T[0,1]$ ) to ensure that predicted values do not exceed 100% cover.

## 151 Model fitting

152 Our Bayesian approach to fitting the vital rate models required choosing appropriate priors  
 153 for unknown parameters and deciding which, if any, of those priors should be hierarchical.  
 154 We decided to fit models where all terms except climate covariates were fit by species, while  
 155 the climate covariates were fit hierarchically with species-specific coefficients drawn from a  
 156 shared ‘global’ coefficient distribution. We did so for two reasons: (1) the four focal species  
 157 are all perennial grasses that we expect to respond similarly to climate covariates, and (2)  
 158 convergence of climate effects at the quadrat level was much easier to achieve when we  
 159 modeled these terms hierarchically, allowing them to “share” statistical strength via partial  
 160 pooling (Gelman and Hill 2007). So, climate effects were modeled as:

$$\theta_{jk} \sim \text{Normal}(\bar{\theta}_k, \sigma_k) \quad (10)$$

161 where  $\bar{\theta}_k$  is the interspecific effect of the  $k$ th climate covariate.

162 We used uninformative priors for all unknown parameters, specifically:

$$\gamma, \beta, \bar{\theta} \sim \text{Normal}(0, 1e^{-6}) \quad (11)$$

$$\phi \sim \text{Normal}(0, \sigma_\phi) \quad (12)$$

$$\sigma_\phi \sim e^{(\text{Gamma}(2, 0.5))} \quad (13)$$

$$\sigma_\theta, \sigma_\gamma, \sigma_\beta, \tau, \zeta \sim \text{Gamma}(0.001, 0.001) \quad (14)$$

163 All of our analyses (model fitting and simulating) were conducted in R (Team 2013). We  
 164 used the MCMC sampler in JAGS (Plummer 2003) to estimate the posterior distributions  
 165 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 1.01, indicating convergence. For the purposes of introducing stochasticity in our population models, we saved the final 1,000 iterations from each of the three MCMC chains 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. 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:

$$N_{t+1} = S \times G + R. \quad (15)$$

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 three 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

## 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

Model	Species	Year Structure	Mean	SD	Min	Max
IPM	BOGR	Full	-0.24	5.78	-46.52	43.05
IPM	BOGR	Climate	-0.28	11.40	-230.10	47.81
IPM	HECO	Full	-0.29	0.95	-10.99	4.33
IPM	HECO	Climate	0.14	1.05	-8.65	6.67
IPM	PASM	Full	-0.04	0.30	-4.29	1.04
IPM	PASM	Climate	0.05	0.20	-2.04	0.99
IPM	POSE	Full	-0.58	1.65	-30.73	7.34
IPM	POSE	Climate	0.23	1.51	-6.08	14.32
QBM	BOGR	Full	-2.06	12.45	-81.48	52.88
QBM	BOGR	Climate	-1.86	13.61	-94.71	54.33
QBM	HECO	Full	-0.25	2.28	-68.53	7.84
QBM	HECO	Climate	-0.47	3.49	-94.57	7.91
QBM	PASM	Full	-0.04	0.50	-15.13	1.56
QBM	PASM	Climate	-0.05	0.53	-9.69	1.67
QBM	POSE	Full	-0.05	2.07	-58.34	13.97
QBM	POSE	Climate	-0.01	2.47	-73.14	15.18

Table 1: Table of residual summary statistics.



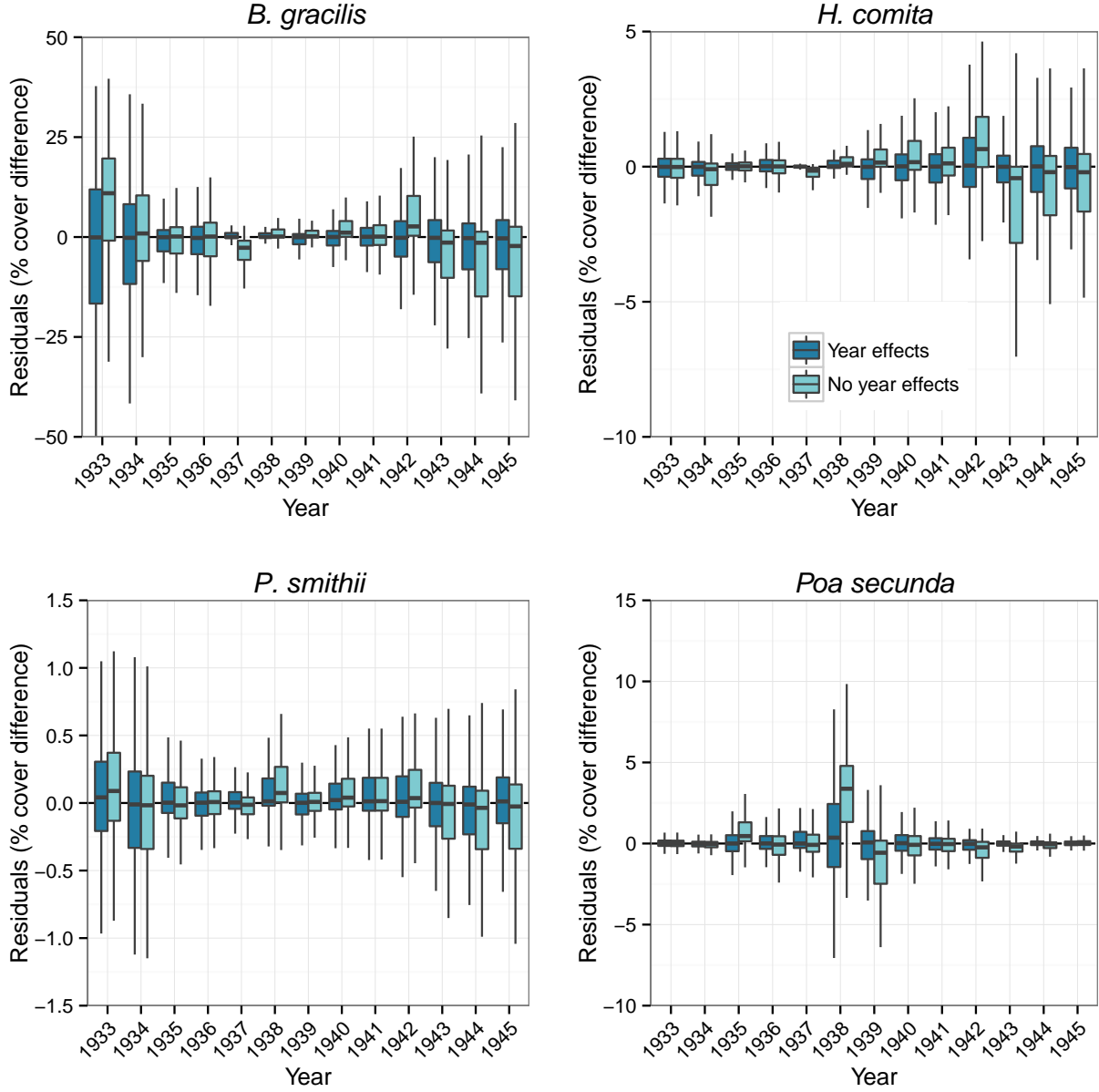


Figure 1: Boxplots of model residuals for one-step-ahead forecasts at each observation year for each quadrat. Residuals represent the difference in observed and predicted cover change from year  $t - 1$  to year  $t$ . Each one-step forecast was simulated 100 times for each quadrat. Note that the y-axes vary across panels.

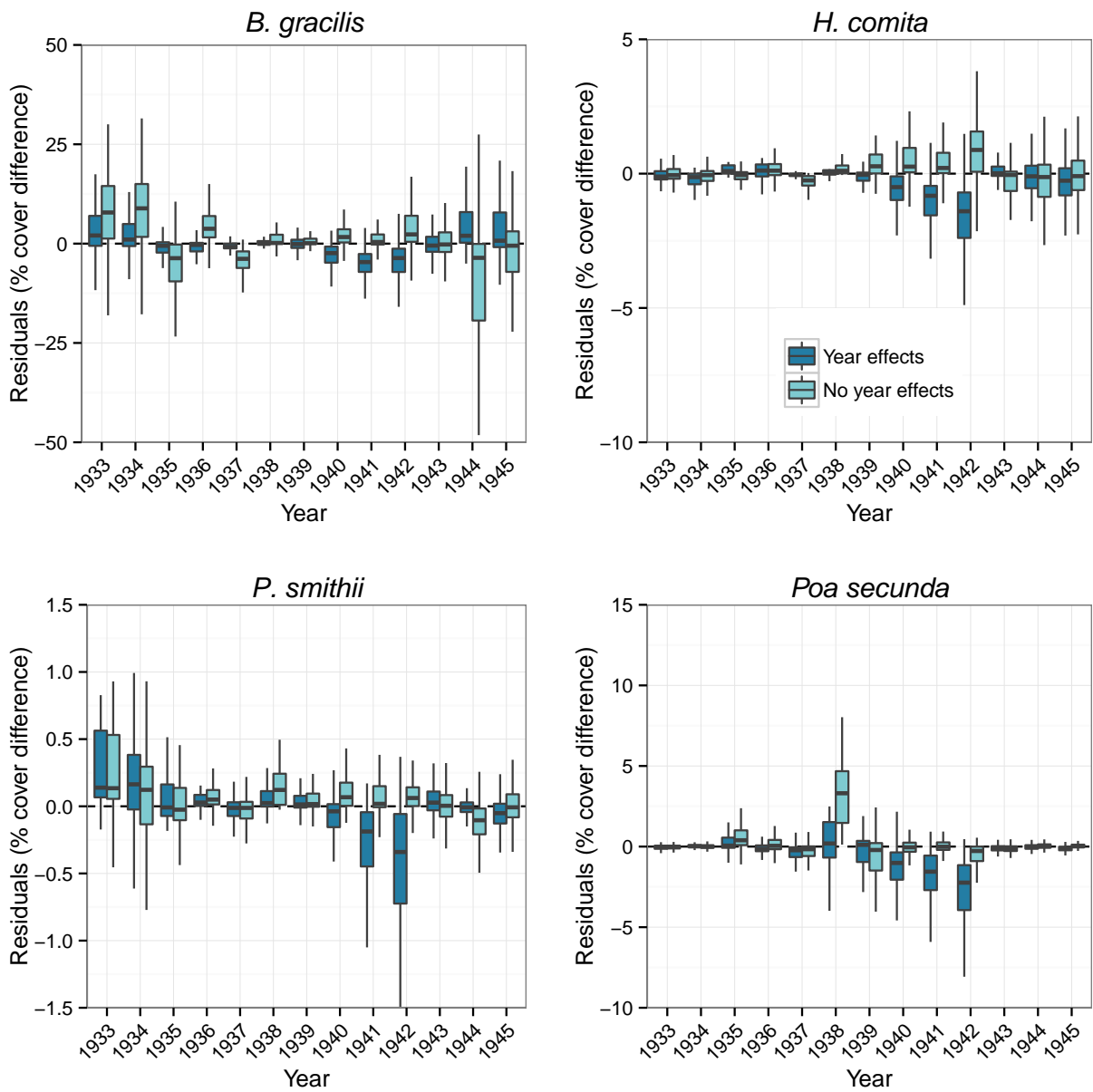


Figure 2: One step forecasts from IPM version.

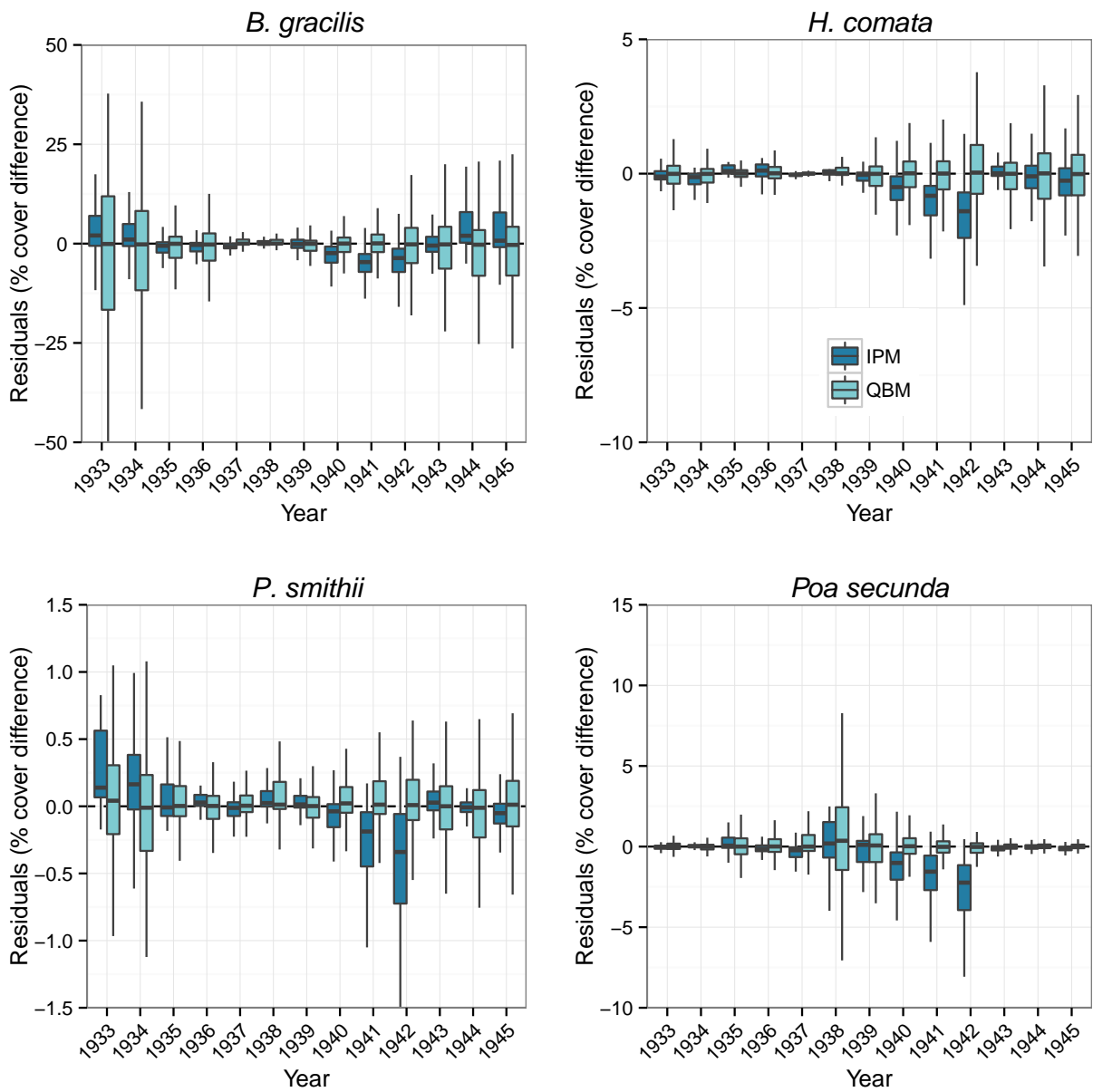


Figure 3: One step forecasts from IPM version.

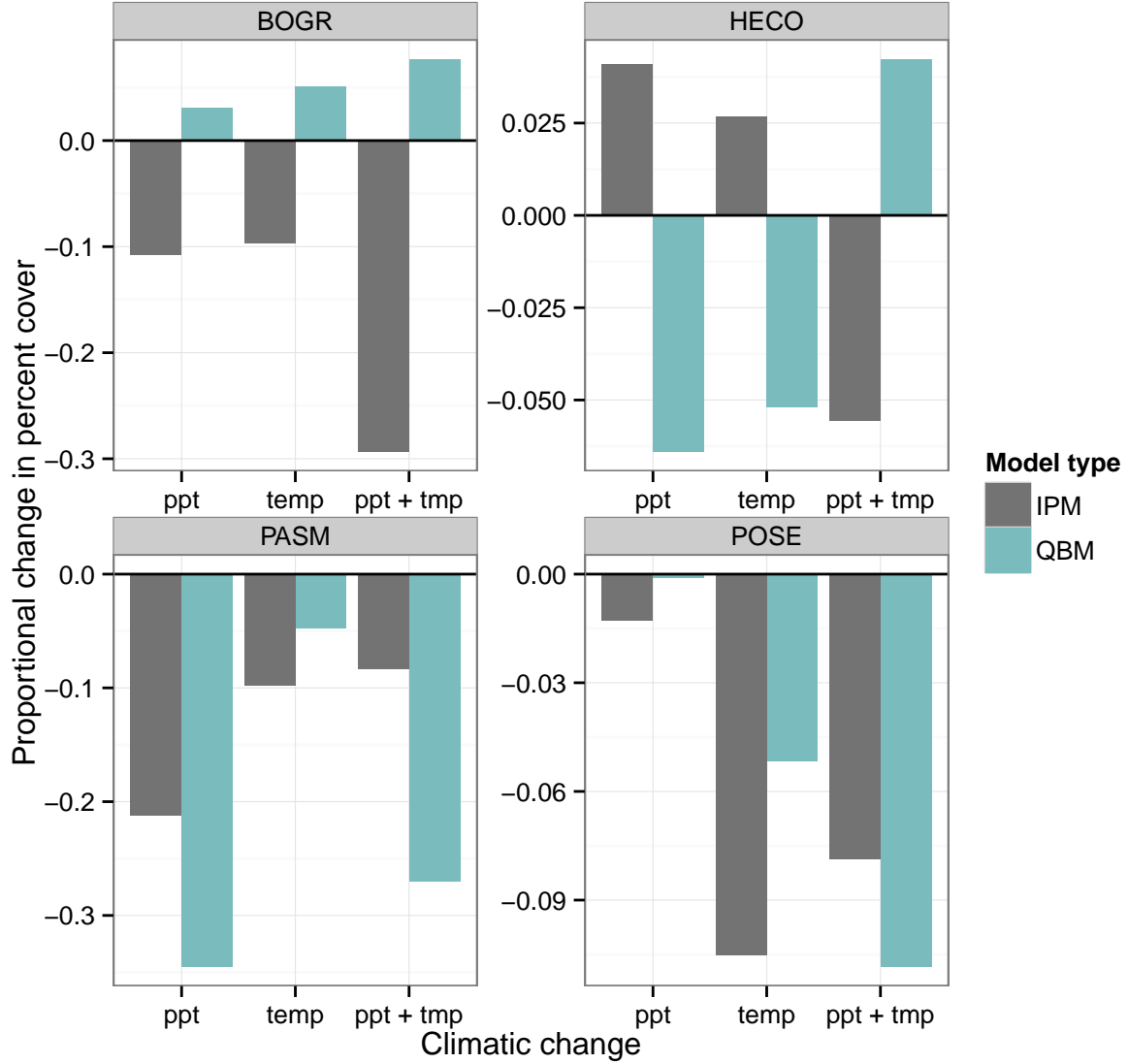


Figure 4: Proportional change in species' mean cover caused by a 1% increase in observed precipitation, temperature, or both as predicted by the individual-based IPM and the aggregate-based QBM. Note that the change for POSE due to a precipitation increase predicted by the QBM is almost zero and so does not show up in the figure. Labels on x-axis refer to: 'ppt' = 1% increase in mean precipitation; 'temp' = 1% increase in mean temperature; 'ppt + temp' = 1% increase in mean precipitation and mean temperature.



## References

- 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.

237 Queenborough, S. A., K. M. Burnet, W. J. Sutherland, A. R. Watkinson, and R. P. Freckleton.  
 238 2011. From meso- to macroscale population dynamics: A new density-structured approach.  
 239 *Methods in Ecology and Evolution* 2:289–302.

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

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

249 Sæther, B. E., S. Engen, V. Grøtan, W. Fiedler, E. Matthysen, M. E. Visser, J. Wright, A.  
 250 P. Møller, F. Adriaensen, H. Van Balen, D. Balmer, M. C. Mainwaring, R. H. McCleery, M.  
 251 Pampus, and W. Winkel. 2007. The extended Moran effect and large-scale synchronous  
 252 fluctuations in the size of great tit and blue tit populations. *Journal of Animal Ecology*  
 253 76:315–325.

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

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