Combining mesocosm and field experiments to predict invasive plant performance: a hierarchical Bayesian approach

CHRIS H. WILSON, T. TREVOR CAUGHLIN, DAVID J. CIVITELLO, AND S. LUKE FLORY

1School of Natural Resources and Environment, University of Florida, Gainesville, Florida 32611 USA
2Department of Biology, University of Florida, Gainesville, Florida 32611 USA
3Department of Integrative Biology, University of South Florida, Tampa, Florida 33620 USA
4Agronomy Department, University of Florida, Gainesville, Florida 32611 USA

Abstract. Invasive plant fecundity underlies propagule pressure and ultimately range expansion. Predicting fecundity across large spatial scales, from regions to landscapes, is critical for understanding invasion dynamics and optimizing management. However, to accurately predict fecundity and other demographic processes, improved models that scale individual plant responses to abiotic drivers across heterogeneous environments are needed. Here we combine two experimental data sets to predict fecundity of a widespread and problematic invasive grass over large spatial scales. First, we analyzed seed production as a function of plant biomass in a small-scale mesocosm experiment with manipulated light levels. Then, in a field introduction experiment, we tracked plant performance across 21 common garden sites that differed widely in available light and other factors. We jointly analyzed these data using a Bayesian hierarchical model (BHM) framework to predict fecundity as a function of light in the field. Our analysis reveals that the invasive species is likely to produce sufficient seed to overwhelm establishment resistance, even in deeply shaded environments, and is likely seed-limited across much of its range. Finally, we extend this framework to address the general problem of how to scale up plant demographic processes and analyze the factors that control plant distribution and abundance at large scales.

Key words: Bayesian hierarchical models; ecological niche; fecundity; invasive species; mechanistic models; Microstegium vimineum; plant demography; scaling up; seed production; species distribution models; stiltgrass.

INTRODUCTION

Fecundity is a key demographic parameter underlying propagule pressure (Holle and Simberloff 2005) and, along with dispersal, is critical for establishment success and range expansion. Given the interactive threats to biodiversity from climate change, habitat loss, and invasive species (Brook et al. 2008), there is growing urgency to predict the demographic processes underlying species distributions, including fecundity. One current approach to modeling plant range expansion is to estimate the relationship between plant size and demographic parameters, such as fecundity, and then link plant size to environmental covariates via regression (Diez et al. 2014, Merow et al. 2014). On its own, this correlative approach does not explicitly account for the mechanisms linking abiotic factors to plant biomass production, and therefore may be vulnerable to cause-effect errors when attempting to predict fecundity across diverse habitats that encompass significant variability in abiotic resources. Moreover, directly measuring fecundity in the field is often logistically difficult and ethically questionable, particularly for nonnative invasive plant species (Flory et al. 2011). Thus, in order to model invasive plant range expansion or success across heterogeneous environments, we propose an approach to modeling seed production based on an abiotic factor that is manipulated in controlled experiments and readily measured in the field.

Many abiotic and biotic factors interact to determine demographic processes such as seed production, making it difficult to disentangle the relationship between any given abiotic factor and an observed species’ distribution pattern. Additionally, predictive models must account for environmental heterogeneity in order to scale across habitats. Previous approaches to predicting species distributions as a function of abiotic factors have included controlled manipulative experiments to delineate the “fundamental niche” of a species (Chase and Leibold 2003). However, such studies are logistically limited to few abiotic factors (Clancy and King 1993, Hooper et al. 2008) and, maybe more importantly, they ignore (by definition) biotic interactions and processes. Thus, their predictive power in the field is limited (Schurr et al. 2012). On the other hand, correlative predictive models such as “bioclimate envelopes” (Kearney and Porter 2009) tacitly incorporate both biotic and abiotic interactions but do not explicitly
model demographic processes such as seed production (Gallien et al. 2010, Merow et al. 2011). Thus, such models cannot identify which mechanisms drive patterns in distribution or abundance (Guisan and Thuiller 2005). More recent niche modeling work has used an integral projection model (IPM) framework to link demographic processes and environmental covariates, while also accounting for population variability (Diez et al. 2014). Since observational correlations of abundance with environmental variables have been shown to generate unreliable predictions when extrapolated from one geographic range to another (Bahn and McGill 2013), it may be preferable to link demographic responses to abiotic factors through experimental introductions that eliminate confounding effects of dispersal, and other biotic interactions. Thus, to accurately predict the performance of invasive species at particular habitats or sites, and their potential to spread at the landscape scale, a hybrid approach that combines the causal insight of manipulative approaches with the ecological realism of performance data under natural conditions is needed (Kearney and Porter 2009).

Although the need to address spatial scale has long been discussed in ecology (Levin 1992), analytical methods for combining data sets from different scales and contexts into quantitative, predictive models have been lacking. Bayesian hierarchical models (BHMs) can facilitate the integration of ecological data sets by offering significant flexibility in model specification (Gelman and Hill 2007). Moreover, by explicitly modeling variance at all hierarchical levels, BHMs incorporate uncertainty from all parts of the model, in contrast to predictive models built on point estimates (Cressie et al. 2009, Gelman et al. 2013, Ogle et al. 2014). BHMs have been used to infer seed production as a function of fire frequency (Evans et al. 2010) and forest type (Caughlin et al. 2014) using observational data and linear correlations. However, this approach could be greatly strengthened by incorporating manipulative experiments and nonlinear relationships that reflect key mechanisms driving demographic performance. BHMs provide an intuitive and useful context for this integration by allowing specification of informative priors based on accurate manipulative experiments and by fully modeling variance at all hierarchical levels of the model (Gelman et al. 2013), thus quantifying the distribution of site-to-site variability.

Here we demonstrate how field and mesocosm data sets can be linked with BHMs to generate predictions for the field performance of an aggressive and widespread invasive plant species. Using Microstegium vimineum (Trin.) A. Camus (stiltgrass), a nonnative grass distributed throughout forests of the eastern United States (Fairbrothers and Gray 1972) as a model system, we combine data from a mesocosm experiment where we manipulated light availability and measured M. vimineum biomass and seed production with data from experimental introductions of M. vimineum across heterogeneous field sites, where light availability was the strongest covariate of biomass production (Flory et al. 2011). Because we measured available light at each of the field introduction sites, these two data sets are, in combination, well suited to scaling predictions of M. vimineum performance as a function of light. Therefore, we combined these data using BHMs to mechanistically link biomass production to available light in the field, while simultaneously using the estimated size-fecundity relationship to predict seed production as a function of biomass. We then used our models to predict relative performance across habitats within our model species’ geographic range to determine first, how much M. vimineum seed production varies with available light in the field, and second, if M. vimineum seed production is likely to provide adequate propagule pressure to colonize and sustain invasions in forest interior habitats given realistic establishment resistance.

**METHODS**

**Mesocosm experiment**

To determine how light availability affects *M. vimineum* performance and fecundity, we established an outdoor mesocosm experiment in June 2010. We sowed 20 seeds of *M. vimineum* (source: Big Oaks National Wildlife Refuge, Madison, Indiana, USA) into two-gallon (7.57 L) plastic containers filled with composted greenhouse soil. These three containers were then placed into each of eight different light treatments that ranged from 1.59% to 100% full sun (AccuPAR Linear PAR/LAI ceptometer, Decagon Devices, Pullman, Washington, USA). Light treatments were replicated three times (24 total tents, 72 total containers). We constructed each shade tent out of polyvinyl chloride (PVC) piping, forming 1 × 1 m cubes that we covered with one to four layers of various types of shade cloth to achieve a range of light conditions. Uncovered tent frames were defined as the full sun treatment and we calculated relative light availability as the ratio of light intensity in a treatment to the full sun treatment. We observed no herbivory on the plants during the course of the experiment. Seed germination was tabulated after three weeks and seedlings thinned to three per container. We watered containers to capacity twice per day using automatic drip irrigation and supplied plants with a 20-20-20 (NPK) fertilizer solution once per month over the course of the experiment.

*M. vimineum* seed production begins in early October in southern Indiana and continues until temperatures are below freezing. Because seeds immediately drop from plants once they are mature, we lined each shade treatment structure with water-permeable landscape cloth to capture seed as it dropped. Since *M. vimineum* has a mixed mating system (produces both chasmogamous [CH] and cleistogamous [CL] seed), we counted all seed that dropped into the landscape fabric or remained attached to exposed seed heads as CH and dissected all...
M. vimineum stems to quantify CL seed production (Cheplick 2008). In late November, after all seed had fallen from the plants, aboveground biomass was harvested, dried at 60°C to constant mass, and weighed.

Field introduction experiment

To evaluate the performance of M. vimineum under natural environmental conditions, we established a field introduction experiment at the Griffy Woods (39°11′17″ N, 86°30′18″ W) and Bayles Road (39°13′9″ N, 86°32′29″ W) properties of the Indiana University Research and Teaching Preserve in May 2009 (Flory et al. 2011). We selected 21 introduction sites across a wide range of habitats including shaded bottomland forest, mowed open fields, dry forested ridge tops, forested stream banks, and lowland and upland forest edge habitat (Fig. 1). Environmental conditions at the sites, including light availability, soil characteristics, and plant communities varied widely (Flory et al. 2011). No introduction sites had invasive M. vimineum in the immediate area but populations of M. vimineum generally occurred within 1 km.

In April 2009, we germinated seed and grew seedlings in the Indiana University greenhouses from 10 M. vimineum populations collected from 10 different states in the invasive U.S. range (Flory et al. 2011). We grew seedlings for four weeks in individual cell packs. Then, we transplanted two seedlings from each of the 10 populations into each of the 21 experimental introduction sites on a random 5 × 8 grid with 0.5-m spacing. At planting time, the transplanted seedlings were 5–10 cm tall and were approximately uniform in height across populations. We assumed that seedlings that died within one week (~15/440; <4%) had died from transplant shock and we therefore replaced them. Although seed germination and seedling establishment is often an important key life history transition (Moles and Westoby 2004), we planted seedlings to prevent accidental release of novel genotypes into natural areas.

In late August 2009, three months after planting, all plants were harvested at ground level, dried at 60°C to constant mass, and weighed. Belowground biomass was not harvested because at most common gardens it was not possible to separate roots of M. vimineum from co-occurring resident species. Plants were deliberately harvested before flowering and seed set to prevent genetic mixing and escape into surrounding habitats.

Modeling

Overall, the aim of our modeling was to combine the size-dependent fecundity relationship that we quantified in the mesocosm experiment with plant growth data across realistic environmental gradients from the field introduction. We use BHMs to reconcile these data and to generate predictions for seed production of individual plants in unobserved sites that span the range of 1–100% available light. These predictions account for uncertainty in all parts of the model and are useful for answering ecological questions about which factors control M. vimineum distribution across its introduced range.

Mesocosm experiment.—To quantify the effects of light on M. vimineum growth in the mesocosm experiment we assumed that light intensity, \( L \), had a saturating effect on biomass production, \( B \), and used these data to parameterize a Michaelis-Menton (MM) function

\[
B = B_{\max} \times \frac{L}{(h + L)}
\]

This function has two parameters, an asymptotic biomass \( B_{\max} \) and a half-saturation constant \( h \), which represents the light intensity that yields half of \( B_{\max} \). We treated biomass as a gamma-distributed random variable because biomass data are positive only, often right skewed, and have variance proportional to the mean (Bolker 2008, Gelman et al. 2013). Thus, for the \( i \)th observation, \( Y_i \), we parameterized the gamma distribution in terms of a mean \( \mu_i \) and a dispersion constant \( k \)

\[
Y_i \sim \text{Gamma}(\mu_i, k)
\]

\[
\mu_i = \frac{B_{\max} \times L_i}{(h + L_i)}
\]
where $L_i$ is the light level associated with the $i$th biomass observation. We sampled a 95% credible interval from the posterior predictive distribution for biomass and compared it to the field introduction data to assess whether the results of the mesocosm could predict biomass in the field introductions (Fig. 2b).

Next, to understand the size-dependent reproduction of $M. vimineum$, we modeled total (CH + CL) seed production as a simple power law function of biomass with a negative binomial error distribution

$$S_i \sim \text{Neg.Bin}(u_i, k)$$

$$u_i = c \times X_i^d$$

where $S_i$ is the estimated number of seeds produced per individual, and $X_i$ is the biomass of the $i$th individual, and the parameters to be estimated are the constant $c$, the power-law exponent $d$, and the over-dispersion parameter $k$.

**Field introduction experiment.**—Next, we built a multilevel model to predict biomass production as a function of light while accounting for variance within and across sites for the field introduction experiment. Because we measured environmental covariates at the site level, we fit the relationship to light at the site level and modeled individual biomass observations as gamma distributed within a site. The basic model is

$$Y_{ij} \sim \text{Gamma}(u_{j1}, k_1)$$

$$u_{j1} \sim \text{Gamma}(u_{j2}, k_2)$$

$$u_{j2} = B_{\text{max}} \times L_j / (h + L_j)$$

where $i$ is the number of individual biomass observations ($n = 381$), and $j$ is the number of sites ($n = 21$), $Y_{ij}$ is the $i$th biomass observation in site $j$, and $L_j$ represents the $j$th value of available light. The parameters to be estimated are $B_{\text{max}}$, $h$, $k_1$, and $k_2$. In essence, this model estimates the parameters of the MM function globally, with individuals nested within sites treated as random effects (for further details about model development see Appendices B and E). Because we had already parameterized the MM function with the mesocosm data, we used the posterior sampling distribution of the MM parameters from the mesocosm model to set informative priors for the multi-level biomass model. We then compared the posterior sampling distribution of MM parameters from the multi-level model to their priors in order to assess the information contributed by the field introduction data. In order to analyze the relative importance of among-site variability compared to within-site variability, we examined the 95% credible intervals of the site-level dispersion constant $k_2$ and individual-level dispersion constant $k_1$. Specifically, we created a test statistic ($k_{\text{test}} = k_2/k_1$) in order to evaluate the probability that the site-level dispersion term was higher than the individual level dispersion term. Given the estimates of all parameters from our data, we predicted the biomass of unobserved individuals nested within 100 unobserved sites, ranging from 1% to 100% available light. We then sampled the posterior distribution for each of these predictions in order to define the probability contour graphs in Fig. 3a.

In order to verify that the fit of our hierarchical biomass model to data (Fig. 3a) was not overly sensitive to the inclusion of data from any particular site, we generated a cross-validated posterior predictive distribution (see Appendix D: Fig. D1). Since we modeled biomass observations as exchangeable within a site (given a known light level), we chose to hold out entire sites and then predict their mean biomass with a model absent their data, a form of leave-one-out cross-validation modified for a hierarchical model such as ours (see also Marshall and Spiegelhalter 2003). We compared the mean squared error (MSE) for the cross-validated predictive distribution to the regular posterior
predictive distribution in order to assess sensitivity to inclusion of the site being predicted.

Finally, we integrated the hierarchical biomass and size–fecundity seed models to predict seed production as a function of light in the field. The full model takes the biomass predictions (from Eq. 6) for an individual nested within a site of a given light level ($Y_{ij}$; Fig. 3a) and propagates them together with the seed production model, Eqs. 4 and 5 (Fig. 3b), in order to generate posterior predictions for seed production as a function of available light (Fig. 3c). Because BHMs estimate all parameters conditional upon one another, these predictions correctly propagate the uncertainty in both parts of the model (Cressie et al. 2009). Therefore, the posterior predictive intervals represent the probabilities that the seed production of a specific individual plant growing within a site of a certain light level will fall within a given range. In particular, we analyzed the posterior seed production predictions at sites of very low light (2%, 5%, and 10% available light) compared to a site of high available light (80%). We calculated the probability that an individual plant’s seed production would exceed 10 or 40 seeds, a range in which prior work has shown that propagule pressure overwhelms forest litter-layer resistance to seedling recruitment (Warren et al. 2012).

We analyzed all models in R (R version 3.0.2; R Core Development Team 2013) and JAGS (version 3.4.0) using the R2JAGS package. We fit each model via Gibbs sampling Markov chain Monte Carlo (MCMC) methods. In each case, we ran five independent Markov chains for 12,000 iterations, with a 2,000 iteration burn-in period and then thinned by 50 iterations, used the $\hat{r}$-hat $< 1.1.$ criterion to assess MCMC convergence, and visually checked for well-blended chains (Gelman et al. 2013).

**RESULTS**

In both experiments, biomass increased nonlinearly with light availability such that above ~50% light availability further increases in light are not associated consistently with greater biomass production. In the mesocosm biomass model, the posterior distribution of the asymptote parameter is $18.2 \pm 2.3$ g (dry mass; mean ± SD) and the half-saturation constant posterior is $40.4 \pm 10.5$% available light (Table 1). As expected, seed production is highly correlated with biomass (Fig. 3b). At peak biomass observed in the mesocosm experiment (~30 g), seed production per plant is between 6000–7000 seeds.

On its own, the results of the mesocosm experiment do not adequately predict biomass in the field introductions. Several sites fall outside the 95% posterior predictive interval altogether and many more have at least several points outside the interval (Fig. 2b). Thus, among-site variation, which the mesocosm experiment cannot assess, clearly influences *M. vimineum* performance.
The hierarchical biomass model, which explicitly accounts for among-site variation, fit the field introduction data much better (Fig. 3a). In addition, the posterior sampling distributions of the MM parameters have less uncertainty than in the mesocosm model (Table 1). Thus, the field introduction data provide additional information for understanding the underlying biomass–light-availability relationship, in addition to accounting for among-site variability. Moreover, the cross-validated mean square error (MSE; 29.21) was very close to the standard posterior predictive distribution MSE (28.61), confirming that our model is not sensitive to the inclusion/exclusion of the particular site being predicted (see Appendix D: Fig. D1).

Our analysis suggests that among-site variability in *M. vimineum* biomass production is at least as large as within-site variability. The 95\% credible interval of the site-level dispersion constant $k_2$ is [1.472; 5.989] and the individual-level dispersion constant $k_1$ is [1.61; 2.23]. Our test statistic result was that ($p(k_{test} > 1)$) = 0.878 (Appendix A: Fig. A1). Thus, averaging over the uncertainty in both estimates (Gelman et al. 2013) there is an 87.8\% probability that the site-level dispersion exceeds the individual-level dispersion. This result is consistent with our observation that incorporating both levels of variance into our predictions (via the hierarchical model structure) significantly enhanced model fit to real data (Fig. 3a, compared to Fig. 2b).

The hierarchical fecundity model predicts that, in the field, seed production will have a similar saturating response to light availability as does biomass (Fig. 3c). In full sun, mean fecundity exceeds 2000 seeds/plant. At all levels, the distribution of predictions has a positive skew. Therefore, even at low light levels there is a small probability of large seed production that significantly exceeds the mean.

Our predictions of seed production are qualitatively distinct in sites of very low vs. high light (Fig. 4 and Table 2). At low light levels, there is a significant probability that an individual plant will not produce seed. However, even at 2\% light, there is a 2\% probability that seed production will exceed 2000. Although there is a decent probability that an individual plant in a site with 2\% light availability will not produce seeds (0.673), this probability is reduced by 50\% when light availability is at 5\% (0.366; Table 2).

### TABLE 1. Michaelis-Menton (MM) parameter estimates from sampling distributions of mesocosm model (used as informative priors in the hierarchical model) and field introduction model (the posterior sampling distribution).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Prior Mean</th>
<th>Precision</th>
<th>Posterior Mean</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymptote</td>
<td>18.2</td>
<td>1/(2.3^2)</td>
<td>17.4</td>
<td>1/(1.9^2)</td>
</tr>
<tr>
<td>Half saturation</td>
<td>40.4</td>
<td>1/(10.5^2)</td>
<td>41.6</td>
<td>1/(8.7^2)</td>
</tr>
</tbody>
</table>

**DISCUSSION**

To predict invasive plant range expansion and performance across variable habitats it is necessary to understand how plant demographic processes respond to environmental factors at various spatial scales (Diez and Pulliam 2007). Here we linked light availability across field sites to plant productivity in order to predict
a specific demographic process (seed production) that is otherwise difficult to measure in the field, particularly for invasive species. First we used a manipulative experiment to understand the fundamental plant response to an abiotic driver (Chase and Leibold 2003). Then, to generate predictions that scale across habitats and account for environmental heterogeneity, we extended this mechanistic insight to large spatial scales by combining our mesocosm experiment with field introduction data using a BHM framework.

> **Table 2.** Probability of seed production of a single plant equaling zero seeds, greater than 10 seeds, or greater than 40 seeds. These are calculated by sampling the posterior predictive distribution of the hierarchical fecundity model (Fig. 3).

<table>
<thead>
<tr>
<th>Light (%)</th>
<th>P[seeds = 0]</th>
<th>P[seeds &gt; 10]</th>
<th>P[seeds &gt; 40]</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.683</td>
<td>0.283</td>
<td>0.248</td>
</tr>
<tr>
<td>5</td>
<td>0.403</td>
<td>0.542</td>
<td>0.478</td>
</tr>
<tr>
<td>10</td>
<td>0.203</td>
<td>0.745</td>
<td>0.691</td>
</tr>
<tr>
<td>80</td>
<td>0.006</td>
<td>0.990</td>
<td>0.985</td>
</tr>
</tbody>
</table>

Our results indicate that *M. vimineum* seed production in the field is strongly linked to light availability, despite including uncertainty from both individual and site-level variability. However, even in low light habitats, our model suggests that seed production is likely high enough to overwhelm establishment resistance (Warren et al. 2012). This finding suggests that the distribution of our model species within its geographic range is not likely to be limited by light availability alone, although its abundance will certainly be greater in higher light habitats (e.g., open fields) than in deeply shaded forest interiors (Cole and Weltzin 2005, Flory 2010). For sites with intermediate or high light availability (30–100% light), our model predicts that individuals can produce approximately 2000 seeds on average. Germination of *M. vimineum* seeds is consistently high at all light levels, and seed viability from plants grown in various light environments is similarly high (as seen in Appendix C: Figs. C1 and C2) and we expect seed production to translate relatively consistently into germinating seedlings (Warren et al. 2011). Therefore, we infer that range expansion of *M. vimineum* is limited by seed arrival, because newly established stands are unlikely to fail due to inadequate local seed production and propagule pressure.

Our model suggests that the variability among experimental introduction sites was at least as important, and likely more so, than variability within sites (probability = 0.878). Thus, when scaling across sites, the relationship between focal abiotic factors and demographic processes is complicated by numerous environmental factors. Accounting for the resulting variability is central to the task of predicting distribution in the field, not merely a nuisance factor. For instance, ignoring the stochasticity introduced by these additional factors and their interactions, equivalent to fixing the among-site variability in the model at zero, would lead to artificially precise prediction intervals. Therefore, by including among-site variability in our hierarchical model, estimated from our heterogeneous field data, we can scale up demographic predictions from mesocosm experiments to the field.

Another approach would have been to include additional environmental covariates, alongside available light, to try and account for some of the among-site variance. For instance, soil resource supply is a major determinant of plant biomass production. Therefore, we recommend that future mesocosm experiments include a cross-factorial manipulation of an integrative soil variable (e.g., average soil moisture supply) alongside available light in order to improve predictive accuracy and biological realism (Tilman 1982). Using a BHM framework, there are several ways that a soil variable could be integrated into the mechanistic function for predicting biomass as a function of light. For instance, the asymptotic biomass ($B_{\text{max}}$) could be allowed to vary for each site based on a regression with soil moisture supply, using informative priors from the cross-factorial manipulation. Such an expansion of the experimental design and model framework would account for more of the among-site variance than our present single-factor model, but at the expense of increased logistical and model complexity.

Validation of our model predictions is complicated by the main consideration that motivated model development in the first place: it is logistically and ethically challenging to accurately quantify seed production in the field. Nonetheless, sub-components of the model can be validated with future data collection, or by approximation with procedures such as cross-validation. Although leave-one-out cross-validation is often computationally expensive, in hierarchical models one option is to iteratively hold out and predict groups rather than individual data-points (Marshall and Spiegelhalter 2003). Our cross-validated predictive distribution for site biomass means (as seen in Appendix D: Fig. D1) is basically the same as our regular posterior predictive distribution (Fig. 3b): half of our observed site means fall outside the cross-validated 50% predictive confidence interval, and one observed mean fell outside the cross-validated 95% interval, which are about the expected numbers for each interval.

One limitation of our model is the need to extrapolate the performance of plants transplanted individually to natural stand dynamics in which density-dependent effects operate. For instance, previous work has established that as *M. vimineum* stand density increases, individual biomass decreases, although stand-level biomass remains stable or even increases slightly at higher densities (Cheplick and Fox 2011). However, we argue that our individual-level predictions are ideal for addressing the spread and establishment phases of the invasion process, rather than for providing insight into dense, long-established stands. Moreover, a distinct advantage of generating clear-cut individual-level pre-
dictions is that we can infer a role for micro-site variability, such as canopy variability in forest interiors, in maintaining local populations.

Although our model indicates that small populations may not persist in very low-light forest interiors (e.g., <2–5% light), there is always a small probability of high seed producing individuals within an established stand. Generalizing to forest interiors, we speculate that M. vimineum is likely exploiting light gaps and microsite variability in light availability (Horton and Neufeld 1998). Interestingly, our results suggest that micro-site variability in light may influence whether M. vimineum is limited by seed arrival (seed limitation) or growth and survival (establishment limitation) after seeds have arrived (Clark et al. 2007). Given the differences in predictions between 2%, 5%, and 10% light, it seems likely that a few plants may be capable of subsidizing seed production for an entire stand where many individuals produce little if any seed. Conversely, given the dramatic increases in seed production with increasing light, our models suggest that disturbances that increase light availability in forest understories (e.g., timber harvests or natural blowdowns) are likely to strongly promote M. vimineum abundance.

Our modeling framework has direct applications to analyzing invasive and native plant species range expansion and performance. For instance, Harris et al. (2011) fit empirical relationships between fecundity and habitat type (varying in type and cover of canopy) for Rhododendron ponticum, and used these estimates of fecundity in an individual-based model to predict speed of range expansion. BHMs help generalize this approach because they facilitate robust predictions for species whose seed production, unlike R. ponticum, is difficult to measure in the field. Moreover, as Kearney and Porter (2009) point out, correlational approaches linking habitat variables and demographic processes are inappropriate for invasive species undergoing range expansion because the species’ distribution is far from equilibrium. For invasive species, it may be especially important to experimentally manipulate abiotic factors in order to understand the fundamental plant responses to these drivers. Our modeling framework provides a way to scale responses to experimental manipulations to the field.

The approach we implemented here could be applied more broadly to analyzing the ecological niche of plant species. For instance, in analyzing the large-scale distributional limits of a species based on temperature, one could parameterize a thermal response curve under controlled conditions, and then use those estimates as informative priors to fit field data with a hierarchical Bayesian model for field observations. We strongly advocate the use of mechanistic functions, that describe process-based relationships between abiotic factors and biotic processes (Kearney et al. 2008). This should help minimize predictive errors that arise when predicting species’ distribution with correlative linear models that often include arbitrary polynomial terms (Bahn and McGill 2013). In general, a hierarchical model structure will facilitate a flexible analysis of variability across many levels of biological organization including individuals, sites, and regions. Thus, such models are better suited to predict across a geographic range, or when extrapolated into new ranges. Moreover, the posterior predictive distributions from BHMs automatically and efficiently incorporate estimation uncertainty from all parts of the model (Cressie et al. 2009, Gelman et al. 2013), eliminating the need to devise potentially complicated bootstrapping algorithms to quantify predictive uncertainty.

In conclusion, we scaled up predictions of a key plant demographic process (seed production) by combining two data sets: a mesocosm experiment with manipulated light availability where we could accurately measure seed production and a field introduction experiment that captured the complex interactions of abiotic variables in the field, thereby adding ecological realism. Our use of experimental data to parameterize nonlinear functions that reflect biological mechanisms should minimize the cause-effect errors that plague correlative distribution approaches (Kearney and Porter 2009, Bahn and McGill 2013). Thus, we do not expect our model predictions to disintegrate when extrapolated to new geographic ranges (Bahn and McGill 2013), but there is the possibility that the geographic location of our experiments (Indiana, USA) has introduced subtle limitations to our predictive power. Overall, we suggest that the influence of abiotic factors on demographic parameters be studied both in logistically manageable manipulative experiments that clarify underlying relationships, and in field studies that realistically account for natural variability and interactions of abiotic variables. BHMs offer wide latitude for combining these sources of information to make intuitive, quantitative predictions.

ACKNOWLEDGMENTS

We thank Jeremy Lichstein for consultation and critical feedback on our hierarchical model and Keith Clay and Furong Long for assistance with the field experiment. This project was funded in part by the University of Florida School of Natural Resources and the Environment. Statement of authorship: S. L. Flory and D. J. Civitello designed experiments and collected data, C. H. Wilson and T. T. Caughlin performed modeling work, C. H. Wilson wrote the first draft of the manuscript, all contributed to writing and revising.

LITERATURE CITED


SUPPLEMENTAL MATERIAL

Ecological Archives

Appendices A–E and the Supplement are available online: http://dx.doi.org/10.1890/14-0797.1.sm