On using integral projection models to generate demographically driven predictions of species' distributions: development and validation using sparse data

Cory Merow, Andrew M. Latimer, Adam M. Wilson, Sean M. McMahon, Anthony G. Rebelo and John A. Silander Jr

C. Merow (cory.merow@gmail.com) and S. M. McMahon, Smithsonian Environmental Research Center, Edgewater, ML 21037, USA. – J. A. Silander Jr and CM, Dept of Ecology and Evolutionary Biology, Univ. of Connecticut, Storrs, CT 06269, USA. – A. M. Latimer, Dept of Plant Sciences, Univ. of California Davis, Davis, CA 95616, USA. – A. M. Wilson, Dept of Ecology and Evolutionary Biology, Yale Univ., New Haven, CT 06520, USA. – A. G. Rebelo, South African National Biodiversity Inst., Cape Town 0184, South Africa.

Knowledge of species' geographic distributions is critical for understanding and forecasting population dynamics, responses to environmental change, biodiversity patterns, and conservation planning. While many suggestive correlative occurrence models have been used to these ends, progress lies in understanding the underlying population biology that generates patterns of range dynamics. Here, we show how to use a limited quantity of demographic data to produce demographic distribution models (DDMs) using integral projection models for size-structured populations. By modeling survival, growth, and fecundity using regression, integral projection models can interpolate across missing size data and environmental conditions to compensate for limited data. To accommodate the uncertainty associated with limited data and model assumptions, we use Bayesian models to propagate uncertainty through all stages of model development to predictions. DDMs have a number of strengths: 1) DDMs allow a mechanistic understanding of spatial occurrence patterns; 2) DDMs can predict spatial and temporal variation in local population dynamics; 3) DDMs can facilitate extrapolation under altered environmental conditions because one can evaluate the consequences for individual vital rates. To illustrate these features, we construct DDMs for an overstory perennial shrub in the Proteaceae family in the Cape Floristic Region of South Africa. We find that the species' population growth rate is limited most strongly by adult survival throughout the range and by individual growth in higher rainfall regions. While the models predict higher population growth rates in the core of the range under projected climates for 2050, they also suggest that the species faces a threat along arid range margins from the interaction of more frequent fire and drying climate. The results (and uncertainties) are helpful for prioritizing additional sampling of particular demographic parameters along these gradients to iteratively refine projections. In the appendices, we provide fully functional R code to perform all analyses.

Knowledge of species' geographic distributions is critical for understanding and forecasting population dynamics, responses to environmental change, biodiversity patterns and the impacts of conservation plans. Using distribution data is challenging, however, because distributions reflect the combined result of many processes – e.g. demography, dispersal, biotic interactions, behavior, historical biogeography – that interact to produce observed spatial and temporal patterns. Most geographic distribution information derives from occurrence data (presences and sometimes absences), rather than on information describing specific ecological processes. This lack of direct information about processes has strongly limited our ability to build and validate mechanistic models

that would allow us to better understand and predict population responses to environmental change (but see, Morin et al. 2008, Morin and Thuiller 2009).

Demographic processes such as survival, ontogenetic growth, and reproduction are the biological foundation of distributional patterns and combine to define the Hutchinsonian niche (Pulliam 2000, Holt 2009, Pagel and Schurr 2012), i.e. the set of conditions where population growth is nonnegative in the absence of immigration. Here, we describe methods to build environmentally dependent demographic distribution models (DDMs) using an integral projection modeling (IPM) approach, with relatively sparse demographic data. Modeling these demographic processes directly facilitates mechanistic explanations and predictions of species' distributions and range-wide population dynamics, while clarifying the roles of important environmental factors. We use this approach to infer the demographic processes that restrict population growth at range margins and assess how population dynamics across the

species' range will respond to potential environmental change. The reason such 'mechanistic niche models' have not been generally developed is that they have proven too difficult to parameterize – directly surveying multiple demographic processes across the entire range of a species would require prohibitive effort (Holt 2009). While this IPM approach would work best with large, spatially extensive data sets, it can still provide useful insights when data for different life stages are imperfect and spatially and temporally mismatched.

Use of demographic information to model species' distributions necessarily involves simplifications of the underlying demographic processes. For example, some studies have been successful in relating measurements of individuals or populations to environmental factors and using these relationships to project the potential distribution of the species (Crozier and Dwyer 2006, Kearney and Porter 2009, Chapman et al. 2014). When a single easily measurable environmental factor has a strong and consistent effect on a sensitive demographic rate – like lethal temperature for mollusks (Helmuth et al. 2006) or development temperature for butterflies (Buckley et al. 2011) – this approach can set biologically based limits on species occurrence. However, the processes modeled in such physiologically based studies do not explicitly consider the demographic pathway through which these abiotic mechanisms affect individual performance. Such models may omit the complexities of tradeoffs in resource allocation manifested in survival, growth, and reproduction that are necessary for persistence across different environments. To capture these demographic effects, we need to work on scaling up measurements of individual performance to demographic rates and consequently population dynamics (Clark et al. 2011).

Previous attempts to understand range-wide demographic responses can be categorized as those with either 1) more detailed models that require a large amount of individual-level data and are computationally challenging to project over large areas or many species or 2) less detailed models that are feasible to study at landscape scales (see review and references in Snell et al. 2014). The first approach is represented by forest gap models (Kohyama 1992, reviewed by Bugmann 2001) and state space models (Clark et al. 2010), which simulate individual-level demography based on abiotic and biotic conditions, typically at stand scales. A variety of models exemplify the second, landscape-level approach. Perhaps the least data-intensive approach is to couple species distribution models with local population projection models (Keith et al. 2008), however such models rely on strong assumptions about the relationship between occurrence probability and demographic rates that are difficult to evaluate. Upscaled forest gap models, such as TreeMig (Lischke et al. 2006), include biotic interactions, dispersal and describe demographic responses to climate across large spatial extents, but require simulation and large amounts of data. As another example, Vanderwel et al. (2013) models demographic rates of functional groups of eastern North American trees as a function of climate and projects population dynamics using cohort-based simulations to gain the flexibility to model range wide population predictions. This approach has the greatest similarity to the models presented here, with the exception that our models include heterogeneity among individuals within a cohort. Our framework represents a compromise between these approaches, in that IPMs

capture some aspects of individual-level demographic variation (size-structure) and explicit response to climate using relatively simple size-structured models and require fewer data than individual-based modeling approaches. IPMs use individual state variables at higher resolution than most other approaches and have a number of associated analytic tools (see below). This simplicity provides a computationally efficient framework, which facilitates Bayesian uncertainty analyses and enables landscape scale predictions. Ultimately, the choice of approach should depend on available data and sufficient modeling detail to capture spatial patterns of interest.

Scaling up from individual responses to demographic rates presents major challenges. Crone et al. (2013) have shown that population projection models have often proven to be poor predictors of species dynamics over time or space, in part because they fail to capture how spatio-temporal environmental variation affects demographic parameters. To make DDMs more reliable, we need ways to: 1) reduce the quantity of data needed to parameterize them; 2) synthesize disparate demographic data to parameterize and use models of population dynamics; and 3) utilize demographic data from contrasting environments across the species' ranges. In this paper, we argue that IPMs can help achieve these goals, based on their reliance on regression models, and present a detailed case study using these methods to build a DDM for the South African shrub species *Protea repens* (the common sugarbush).

Despite the challenges involved, there are important reasons to work toward these goals of applying DDMs to species' distributions: 1) they aid in understanding the mechanisms driving distributional patterns; 2) they can predict spatial and temporal variation in local population statistics such as population growth rate, sensitivities/ elasticities, stage passage times (e.g. to reproduction); 3) they can improve our ability to project population-level patterns to new locations or environmental conditions (cf. Huntley et al. 2010, Schurr et al. 2012b, Crone et al. 2013). Taken together, these advantages also allow us to bridge multiple biological scales: intraspecific trait variation, life history strategies, local population dynamics, landscape scale population dynamics, and range dynamics. Landscape ecologists and biogeographers generally focus on the population-level patterns at larger geographic extents, but currently do not have an accessible means of benefiting from the insights of population biology. By scaling demographic processes from the individual to regional level, biologists can check that their understanding and model predictions match observed patterns at each level of biological organization, leading to greater confidence about inference and projection. When the processes in the model fail to explain major patterns at a particular scale, we can focus research on measuring demographic transitions identified as the most critical for driving patterns at that scale.

Integral projection models (IPMs; Easterling et al. 2000) are size-structured demographic models. Unlike population projection models derived from stage-based matrices (cf. Caswell 2001), IPMs use regression models to relate continuous state variables, such as individual size, to explanatory covariates, such as environmental conditions, to predict vital rates (Table 1; Easterling et al. 2000, Ellner and Rees 2006, Merow et al. 2014). For example, we used logistic regression to relate individual survival to individual size, winter temperature, and mean annual precipitation. Typically,

Table 1. Regression coefficients for all vital rate models. Environmental predictors have been standardized. Columns correspond to posterior mean parameter values and lower (upper) 95 percent credible interval bounds. Models were chosen by backward stepwise DIC selection described in Supplementary material Appendix A.

	Mean	Lower-95% CI	Upper-95% CI
Seedling survival probability			
Intercept	−2.597	−3.680	−1.376
Size	4.744	2.807	6.357
Minimum July temperature	−0.362	−0.603	−0.101
Mean annual precipitation	0.188	0.014	0.363
Adult survival probability			
Intercept	3.406	2.782	4.010
Size	−0.695	−0.889	−0.491
Minimum July temperature	0.154	−0.034	0.327
Mean annual precipitation	0.608	0.414	0.846
Growth			
Intercept	0.182	0.167	0.198
% High fertility soil	−0.030	−0.051	−0.013
% High fertility soil^2	0.056	0.018	0.085
Winter soil moisture days	0.036	0.015	0.053
Winter soil moisture days^2	−0.015	−0.024	−0.005
% Acidic soil	−0.002	−0.009	0.007
% Acidic soil^2	−0.035	−0.051	−0.019
Minimum July temperature	0.018	0.011	0.026
Summer soil moisture days	−0.003	−0.012	0.006
Summer soil moisture days^2	−0.030	−0.037	−0.022
Flowering probability			
Intercept	−2.629	−3.755	−1.503
Size	3.627	2.583	4.584
Seedheads/individual			
Intercept	1.495	1.211	1.824
Size	0.622	0.466	0.788
Minimum July temperature	−0.272	−0.401	−0.156
% Acidic soil	0.347	0.163	0.540
% Acidic soil^2	−0.428	−0.652	−0.236
Seeds/seedhead			
Intercept	4.156	4.075	4.241
Germination probability			
Intercept	0.010	0.009	0.012
Offspring size			
Intercept	0.163	0.137	0.193
% Acidic soil	−0.020	−0.035	−0.002
% High fertility soil	−0.054	−0.091	−0.019
% High fertility soil^2	0.065	0.012	0.117
Minimum July temperature	0.041	0.024	0.056
Summer soil moisture days	−0.001	−0.012	0.011
Summer soil moisture days^2	−0.035	−0.051	−0.020

observations of vital rates for some sizes or environments will be unavailable. Regressions allow IPMs to bridge these missing values through interpolating and extrapolating the fitted regression function. IPMs can then be used to evaluate the implications of those projections. For example, in the survival model mentioned above, one can predict the survival probability for any combination of plant size, winter temperature, and mean annual precipitation, whether or not the particular combination has been observed. Such projections mean that robust inference depends on ensuring that regressions make sensible predictions when extrapolating (at least qualitatively).

DDMs rely on predictions from the vital rate regressions across a landscape where spatially extensive data are available on the environmental covariates. From predicted vital rate

models, one can construct IPM kernels and therefore predict population dynamics at new locations, so long as the values of the environmental covariates are known for those locations. The critical step for building DDMs is thus not the use of IPMs per se (i.e. continuously varying state variables), but the regression of demographic parameters along environmental gradients. Unstructured population models, matrix projection models or individual-based models could also employ regression of their relevant parameters on environmental covariates to project population dynamics across a landscape, although this is not common (but see Doak and Morris 2010, Vanderwel et al. 2013).

We used Bayesian regressions models and focus here on their value for handling missing or sparse data while accounting for uncertainty in model predictions (Clark 2005), which is rarely quantified in population projection models (Crone et al. 2013). Though many biologists have valuable demographic data for their study organisms, some aspects of size structure, life history or environmental response are often poorly measured or exhibit inexplicable variability. Such data gaps make demographic modeling challenging and limit inference on population biology. Nonetheless, building models with incomplete data sets or, perhaps relying in part on complementary expert knowledge, can give new insights into environmental responses (Cabral et al. 2013, Crone et al. 2013) while highlighting priorities for further data collection. Accurately describing the uncertainty that results from incomplete data, a key strength of Bayesian modeling (Clark and Bjørnstad 2004), is crucial for determining whether predictions depend heavily on assumptions rather than data, and to what extent the data narrow the range of feasible predictions. With large data sets that cover the entire life history and the species' distribution, Bayesian IPMs can be used to project population dynamics across a species' range. With sparser data sets from fewer populations, we can still use these models to describe the studied populations and extract qualitative insights about range dynamics, generate mechanistic hypotheses about their drivers, and target future data collection by identifying the critical or least understood stages/sizes.

As a case study for the use of DDMs, we build models for *Protea repens*, a common, often abundant, and widely distributed overstory shrub of the Mediterranean climate, fynbos shrubland biome in the Cape Floristic Region (CFR) of South Africa. Proteaceae in the CFR are an example of an iconic, biodiverse suite of species and have been projected to be vulnerable to climate change (Midgley et al. 2003, Schurr et al. 2012a, Cabral et al. 2013). The fynbos is a fire prone system with median fire return times typically ranging from 10 to 21 yr (van Wilgen et al. 2010). On average across the region, fire frequency has decreased by about four years over the past three decades, due in part to changing climate (Wilson et al. 2010). Thus dynamic models are critical for understanding species' response to environmental change in this system, above and beyond existing static, correlative species distribution models (Midgley et al. 2003, Keith et al. 2008, Franklin 2010, Cabral et al. 2013).

Understanding the factors that control the distribution and abundance of Proteas represents a fairly common and important type of modeling challenge. To manage fynbos ecosystems for resilience, it is important to understand

demographic vulnerabilities and their variation during ontogeny to make fire management decisions that avoid bottlenecks for population growth. Furthermore, mapping demographic patterns is critical to provide spatially explicit, reserve-specific management advice (e.g. on the consequences of different fire return times for population growth, cf. Fig. 4a, b). There are strong environmental gradients in the region so it is vital that these conservation decisions explicitly account for population responses to different environments (Thuiller et al. 2006, Latimer et al. 2009, Wilson and Silander 2013).

In this paper, we describe data collection, fitting, and projection for environmentally dependent IPMs, and demonstrate how these can be used to predict a variety of demographic attributes of species' geographic distributions. We illustrate models built from demographic data collected during a single visit to each of 121 populations that cover a wide range of environmental variation but do not span the entire range of the species. We demonstrate that these models can produce insights about population growth rates, sensitivities, and conservation-relevant metrics such as the probability of persistence across a species' range. In the appendices, we provide fully functional R code (R Core Team) to perform all analyses that can be readily adapted for other data sets.

Methods

Integral projection models

To illustrate the attributes of IPMs that facilitate modeling range dynamics, we present a very brief introduction to their construction and refer readers to reviews in Ellner and Rees (2006), Coulson (2012) and Merow et al. (2014) for further details. As is typical in IPMs, we use individual size as the state variable, here measured as the total length along the central stem. Demographic transitions (survival, growth and fecundity) are described by the kernel function, $K(z'|z,\boldsymbol{x})$, where z' denotes the size at time $t + 1$, z denotes the size at time t, and $\boldsymbol{x}$ is a vector of environmental covariates at a particular location. Projecting the size distribution of individuals in the population at t, given by $n_t(z)$, to $t + 1$, given by $n_{t+1}(z')$, is performed by:

$$n_{t+1}(z') = \int_\Omega K(z'|z,\boldsymbol{x})\, n_t(z)\, dz \qquad (1)$$

where Ω denotes the set of all possible sizes. The kernel can be decomposed into a growth/survival subkernel, $P(z'|z,\boldsymbol{x})$, and a fecundity subkernel, $F(z'|z,\boldsymbol{x})$, as $K(z'|z,\boldsymbol{x}) = P(z'|z,\boldsymbol{x}) + F(z'|z,\boldsymbol{x})$. The subkernels $F(z'|z,\boldsymbol{x})$ and $P(z'|z,\boldsymbol{x})$ can be further decomposed into functions specific to the species' life history and which can be estimated from regression. For example, here we write

$$P(z'|z,\boldsymbol{x}) = s(z,\boldsymbol{x})\, g(z'|z,\boldsymbol{x}) \qquad (2)$$

where $s(z)$ is the probability of survival as a function of individual size and $g(z'|z,\mathbf{x})$ is the probability of growing from size z to size z' during one time step. We similarly decompose fecundity into flowering probability, $p_{flower}(z)$,

number of seedheads/individual, $f_{seedhead}(z,\boldsymbol{x})$, number of seeds/seedhead, f_{seed}, recruitment probability, $p_{recruit}$, and the offspring size distribution $f_{recruit\,size}(z',\boldsymbol{x})$, as,

$$F(z'|z,\boldsymbol{x}) = p_{flower}(z)\, f_{seedhead}(z,\boldsymbol{x})\, f_{seed}\, p_{recruit}\, f_{recruit\,size}(z',\boldsymbol{x}) \qquad (3)$$

Each of the vital rate functions is then estimated through a regression on the state variable, size, and environmental covariates, as appropriate (Table 1). Although the state variable and environmental covariates enter the vital rate regressions in the same way, they are used differently in the IPM. A single value for each environmental covariate, corresponding to the location that the IPM represents, must be supplied in order to build the IPM kernel. In the model, all individuals in a location 'experience' the same values of the environmental covariates. In contrast, the state variable differs among individuals within a population during ontogeny. The details of each regression are described below.

Data collection

When fires occur in fynbos shrubland, virtually all *P. repens* adults in burned areas are killed, as they lack the capacity to resprout (Rebelo 2001). Their seeds are stored in serotinous seedheads (cones) that open after fire, and seedlings recruit exclusively in the autumn following fire (Holmes and Newton 2004). This fire-driven life history presents sampling challenges, because data on recruitment rates can be collected only in areas that have recently burned, while information on growth rates, adult mortality and fecundity can be collected only in areas that have not recently burned. In our sampling campaign, we sampled 121 sites (38 for growth and 27 of those for fecundity, 63 for mortality, and 20 for seedling recruitment) across the geographic and environmental range of the species (Fig. 1). At growth sites, we measured the length of the central stem of 15 randomly selected plants along a transect across the population. Given the time since the populations last burned (estimated in the field and validated with data from CapeNature fire records; Wilson et al. 2010), we calculated average annual growth for each individual as stem length divided by plant age. Stem length ranged from 0.03 to 4.4 m. Plant height is approximately 86% of total stem length due to branching architecture. Because *P. repens* typically adds one internode segment to the central stem each year, we took the length of the basal internode segment as a measure of offspring size.

At 27 of the 38 plots where growth was measured, the number of seedheads was counted for each individual. For two plants at each location, we collected three seedheads from each of the last three years of production (when available) and counted the number of seeds in each. We assumed that seeds on reproductive adults become capable of producing recruits in the year following production, and remain viable for two years, so that the aerial seed bank includes up to three crops of seeds (Musil 1991). At the mortality sites, we surveyed four randomly located areas of approximately 100 m² each (except where populations were too small), in which we counted the number of live and dead plants. Plants that had died within the previous year were easily identified as those with brown leaves remaining on the stem,
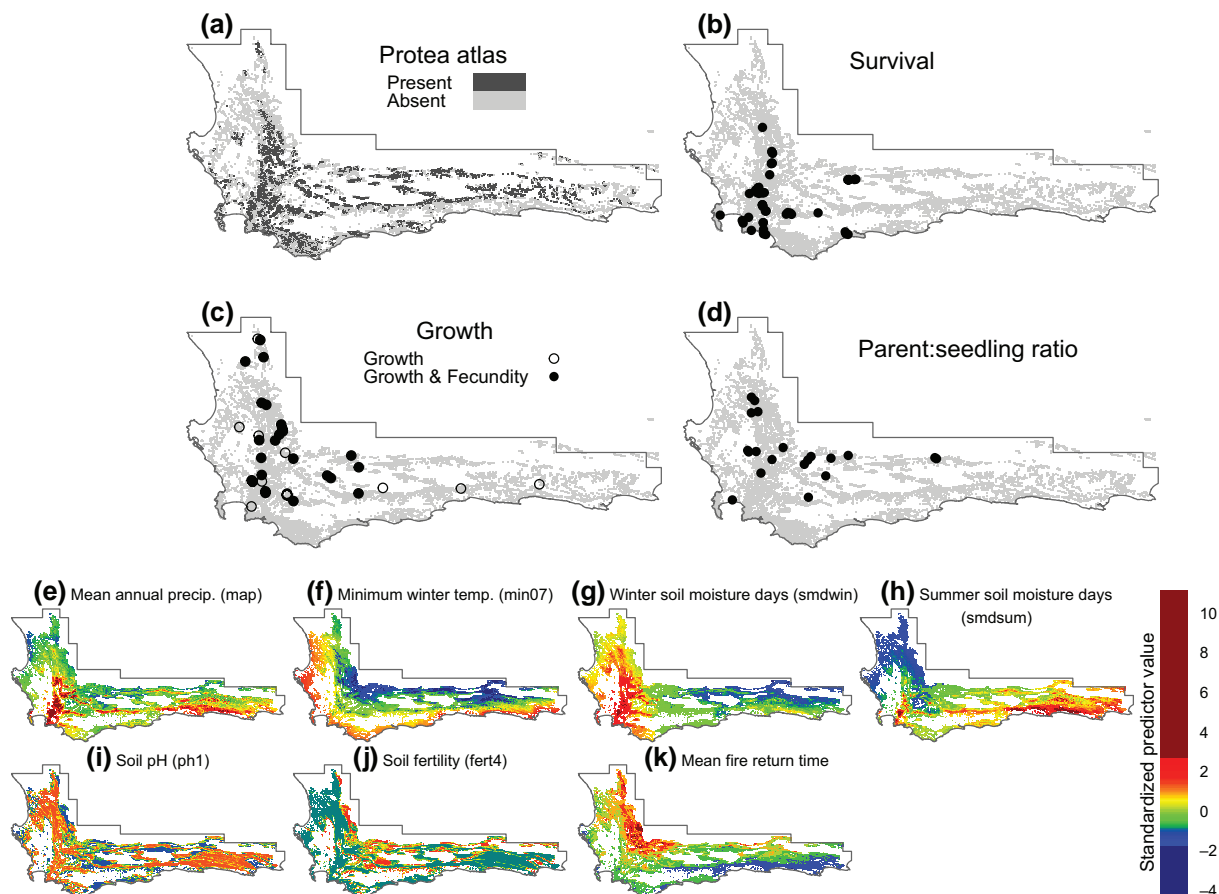
Figure 1. Sampling locations for demographic data (a–d) and the candidate environmental covariates at 1-minute resolution (1.55 × 1.85 km) used in vital rate regressions (e–k). The Cape Floristic Region is bordered by the Atlantic and Indian Oceans to the west and south (respectively) and the Great Karoo Desert to the north.

enabling us to capture mortality over a single year (cf. Kobe et al 1995). At the seedling recruitment sites, we counted the number of dead parent plants and live seedlings in a 2 m wide belt transect across the population. The data were collected in short field visits during 2008–2011, in late summer to early fall, after conclusion of the summer establishment and growth period following preceding fires. Note that size was not measured with the mortality or parent–seedling ratio data.

Occurrence and abundance data across the region were used to assess model predictions. The Protea Atlas database derives from a major citizen science initiative to record occurrence and ordinal abundance data (0, 1–10, 100, 1000 individuals) for all ~330 Proteaceae species throughout the Cape Floristic Region (Goldblatt and Manning 2002) and contains more than 250 000 species records over ~90 000 km² (Rebelo 2002). The Protea Atlas data represent community surveys of Proteaceae for an area with diameter up to 500 m. We aligned these surveys with the 1′ square grid imposed on the landscape by the resolution of our environmental data (described below) and summarized the data to determine abundance, presence, and absence, while minimizing the potentially false absences due to incomplete sampling of a grid cell (Fig. 1; details in Supplementary material Appendix A). We used these data (3985 presences and 2338 absences for *P. repens*) to evaluate predictions of population growth rate (λ), assuming that presence (absence) locations should

have λ > 1 (< 1). We expected that observed abundance should have positive correlation with λ (see Discussion).

Environmental data were summarized at 1′ × 1′ (1.55 × 1.85 km) resolution and include both climatic factors (averaged from 1950–2000) and edaphic factors that have proven useful for explaining occurrence patterns in the fynbos (Richards et al. 1997, Schulze 1997, Latimer et al. 2006). After removing predictors with correlation higher than 0.5, we were left with the following: minimum July temperature, number of summer soil moisture days (summer SMD; the number of summer days above a soil moisture threshold; Supplementary material Appendix A), number of winter soil moisture days (winter SMD), proportion acidic soil, and proportion high fertility soil (Fig. 1; Supplementary material Appendix A). Mean fire return times for each pixel were estimated using a survival model between observed fire return times derived from satellite and field data over 1980–2010 (Wilson et al. 2010, de Klerk et al. 2012) and satellite-derived post-fire ecosystem recovery trajectories (Supplementary material Appendix B).

Regression models

The regressions between the various vital rates (survival, growth, and fecundity; Table 1) and the environmental data were all fit in a Bayesian framework and 1000 posterior

samples of the regression coefficients were saved in order to propagate the uncertainties from this step through the population projections (explained below). Posterior samples were thinned appropriately after an initial burn in period to remove autocorrelation in the MCMC chains. Regression models were built using MCMCglmm (Hadfield 2010), JAGS (Plummer 2003) and MCMCpack (Martin et al. 2011) in R ver. 3.0.2 (R Core Team).

Growth

We modeled average interannual growth using linear regression assuming normal residuals. This average growth increment was added to the size at time t to predict the size at time $t + 1$ (providing $g(z'|z,\mathbf{x})$ in Eq. 2). This model assumes that length of the central stem is a linear function of time since fire. This approximation is feasible for interfire intervals we consider, producing individuals approximately 3 m tall after 25 yr (a long interval), which is consistent with the sizes we observed in the field (estimate obtained from growth model intercept (0.182) $\times$ 25 yr fire interval $\times$ correlation between central stem length and plant height (0.86)). While smoothing growth in this way reduces temporal resolution, it allows us to avoid remeasuring individuals each year, while providing a more stable estimate of individual growth rates for comparisons among sites along environmental gradients. The variance of the residuals was taken as the predicted variance in annual growth.

Fecundity

The transition from adult in the year of fire to seedling in the year following fire is challenging to parameterize because this transition includes several processes operating in turn: flowering, seed production, recruitment, and first-year survival and growth. While potentially limiting our ability to make robust predictions, this challenge gives an opportunity to evaluate how well we can do even when some demographic transitions are difficult to observe and data are sparse.

Flowering probability ($p_{flower}(z)$ in Eq. (3)) was modeled using logistic regression with an indicator for flowering as the response variable and size as a predictor. Too few observations were available to determine environmental dependence of flowering probability. We modeled the total number of seedheads per plant ($f_{seedhead}(z)$ in Eq. (3)) using a Poisson regression with size and environmental covariates as predictors. From field collection of seedheads we obtained an estimate of seeds per mature seedhead using Poisson regression (f_{seed} in Eq. (3)).

To estimate effective recruitment rates ($p_{recruit}$ in Eq. (3)), we used parent-seedling ratio data at the three available sites that were surveyed in the year immediately after fire. Using the seedhead model, the number of observed parents, and their estimated sizes based on the growth model ($g(z'|z,\mathbf{x})$ above), we predicted the total number of seeds. By dividing the observed number of seedlings by the predicted number of seeds, we estimated the effective recruitment probability. The posterior distribution of recruitment probability incorporated uncertainty in predicted parent size and, conditional on parent size, the number of seeds produced, by sampling parameters from the posterior distributions of the growth and seedhead models to be used in the prediction. Data were not available to identify how variation along environmental

gradients in these components contribute to overall recruitment rates hence we use a constant recruitment probability for all locations.

Finally, we predicted the distribution of seedling sizes ($f_{recruit size}(z')$ in Eq. (3)). We regressed the length of the first stem segment (*P. repens* typically adds one segment per year) on environmental covariates using linear regression, assuming normal residuals. The variance of the residuals was taken as the predicted variance of seedling sizes.

Survival

Our survival models similarly provide an illustration of how Bayesian IPMs can employ sparse data. To estimate survival probability across all sizes ($s(z)$ in Eq. (2)), we used two complementary data sets that related to different life stages. The first consisted of mortality plots, in which the number of living and dead adults/juveniles were recorded, and the second consisted of parent-seedling ratios from recently burned (1–5 yr) populations. Because the size of these individuals was not recorded, we imputed the size of individuals in both data sets using our fitted regressions for the initial size of seedlings and their subsequent growth. Parent size was predicted in the year before mortality using the time since fire and the predicted average interannual growth from the growth model above (i.e. offspring size + (time since fire – 1) $\times$ (average interannual growth)). An analogous procedure for predicting seedling mortality from parent-seedling data, which also involved predicting parent sizes, the number of seeds, recruitment, and offspring size is described in detail in Supplementary material Appendix A. Using a Bayesian framework made it easier to incorporate uncertainty into the model, since within the model, we could sequentially predict each unobserved quantity from the fitted regressions above. We propagated parameter uncertainty from the models for growth, fecundity and seedling initial size by making 100 random draws from the posterior distribution of those regression coefficients, then refitting the survival models for each set of coefficients. These imputed sizes, along with environmental covariates, were used as predictors in a logistic regression, using a binary survival response.

In spite of the many assumptions associated with this model, the limited available empirical information is broadly consistent with our estimation. In a field experiment involving broad scale transplantation of seeds and seedlings of four *Protea* species with similar life history and seed traits, the median first year survival rate was 0.12 (Latimer et al. 2009). Our model predicts median (95% CI) seedlings survival across the entire region of 16.5% (13.3%, 34.2%) in their first year after establishment.

Predicting vital rates

A number of demographically informative quantities and maps can be calculated directly from the vital rate regressions. For example, we mapped the predicted value of each vital rate (for individuals of a particular size) across the landscape to look for similarities and differences in their response to environment (Fig. 2). We mapped the time until 90% of individuals in a cell become reproductive, a quantity used in fynbos management to determine optimal timing of controlled burns (de Klerk et al. 2007). To do this, we used the flowering probability model to predict the size at
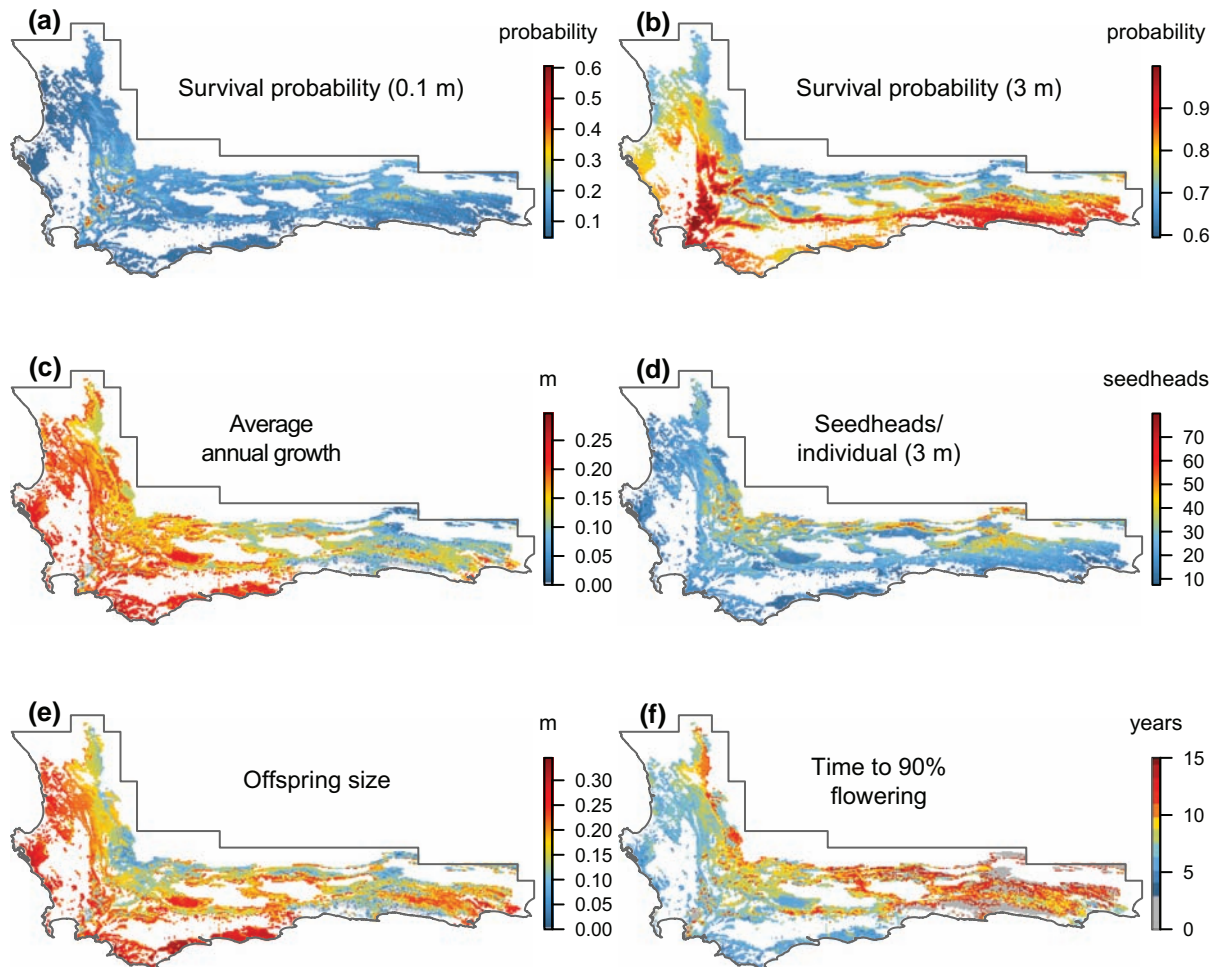
Figure 2. Maps of predicted vital rates. (a) Survival probability of 0.1 m individuals; (b) survival probability of 3 m individuals; (c) mean annual growth increment; (d) number of seedheads for 3 m individuals; (e) offspring size; (f) time until 90% probability of flowering.

which there is a 90% chance of flowering, and then used the seedling size distribution model, in conjunction with the growth model, to predict the amount of time needed to reach this size (i.e. time = [(90% flowering size − offspring size)/interannual growth] + 1).

Population inference

Population growth rate

The vital rate regressions were combined to build growth/survival and fecundity subkernels according to Eq. (2–3). We predicted asymptotic population growth rates based on the mean fire return time using an annual periodic projection model (cf. Caswell 2001). *Protea repens* very rarely recruits in non-fire years and it does not survive most fires, so a generation corresponds to one fire interval. For a cell with a given set of environmental covariates and mean fire interval, we combined the appropriate number of cell-specific growth/survival subkernels (the same for each non-fire year) with a single fecundity subkernel (fire year). Integration was performed using the midpoint rule by discretizing the subkernels at 100×100 cell resolution (Easterling et al. 2000). Eigen-analyses were used to extract population growth rates and associated sensitivities/elasticities (Easterling et al. 2000) for

different combinations of environmental variables and fire regimes. For example, for an 18-yr fire interval, 17 growth/survival subkernels were combined with a single fecundity subkernel to describe the dynamics during a single generation. For simplicity, the dominant eigenvalue, corresponding to the asymptotic population growth rate, was converted to an annual scale (the periodic model describes growth among generations) by raising it to the power of 1/(fire return time). This conversion enabled us to compare growth in cells with different expected fire return times on the same scale.

A number of informative statistics related to λ are readily calculated from our models. By examining locations where $\lambda > 1$, based on the observed fire regimes, we can characterize the Hutchinsonian niche. By mapping λ, we can predict the species' range. The accuracy of these predictions can be assessed using presence/absence and abundance data that were independently collected. We calculated the percentage of 2338 presence and 3985 absences correctly predicted based on whether predicted $\lambda > 1$. We also calculated the AUC (area under the ROC curve) for a threshold-independent measure of prediction accuracy (Fielding and Bell 1997). Finally, we calculated the mean predicted value of λ for each abundance class ([0], [1–10], [10–100], [100–1000], [1000, ∞]).

1173

Sensitivity/elasticity analysis

We performed two types of sensitivity analysis with respect to λ, which were mapped across the region. First, we calculated parameter elasticities by perturbing regression coefficients by $\delta = +1\%$. These were calculated at site i for parameter j as $e_{ij} = [(\lambda_{perturbed} - \lambda_{fitted})/(\delta \times \beta_j)] \times (\beta_j/\lambda_{fitted})$. We interpret these elasticities to understand how different life history transitions contribute differentially to population growth. Second, to accommodate the fact that the same environmental predictors appear in multiple vital rate regressions, which makes interpretation of the coefficients in terms of λ challenging, we also explored the sensitivity of λ to perturbation in the environmental covariates. All environmental predictors were standardized, making these sensitivities comparable. The sensitivities were calculated at site i for covariate j as $s_{ij} = (\lambda_{perturbed} - \lambda_{fitted})/(x_j + 0.1)$. By changing, e.g. the mean July temperature by $+ 0.1$, we could map the implications for *P. repens*, integrated over all components of the model affected by mean July (winter) temperature. We interpret these sensitivities to describe the environmental conditions that limit *P. repens*' distribution, while accounting for the impact across all aspects of life history. The sensitivities can also be interpreted to qualitatively describe anticipated response to environmental change.

Uncertainty analysis

A key advantage of using a Bayesian modeling framework is that it enables us to propagate uncertainty through each stage (e.g. growth regression) of a model to 'downstream' stages (e.g. from predicted parent size to survival regression parameters to λ). This accumulation of uncertainty through the model then allows us to explore in more depth the patterns of uncertainty in results and predictions via the posterior distributions of population statistics (here, λ).

We plotted the interquartile range of λ in each cell to understand where knowledge is lacking. Regions with a large interquartile range have environmental conditions with uncertain suitability, which can be used to target future data collection (and model improvement). The posterior distribution of λ was also used to estimate the probability that $\lambda > 1$ (indicating likely long term population sustainability) by tallying the number of posterior samples that were greater than 1. This metric incorporates the uncertainty to identify regions that the model confidently predicts could support sustainable populations. In contrast to occurrence models, which estimate the probability of presence in a grid cell, we directly estimate the probability that the environment is suitable using the underlying demographic processes.

Scenario-based projections

We predicted λ under three different fire and climate scenarios. The regional mean fire return time over 1975–2000 of 18.75 yr is estimated to have decreased 4 yr compared to 1951–1975 due to warming and drying associated with climate change in this region (Wilson et al. 2010). First, we estimated the impact of future climate change by further reductions in fire return-time (an additional four years) and predicting λ under this scenario. Second, we predicted the consequences of the fire regime returning to its previous pattern by increasing the fire return time by 4 yr. Third, we combined the reduced fire time projection with a future climate scenario. For illustration, we increased winter temperature by 1 degree and decreased precipitation-based metrics by 10%, based on the conservative RCP4.5 scenario for mid-century, median multimodel predicted change (IPCC-WGI 2013). We assumed that a 10% decrease in precipitation would translate into a 10% reduction in summer and winter soil moisture days, although a more thorough, spatially explicit downscaling would be ideal. In this third case, we compared the present and future values of sensitivity/elasticity of λ to parameter values and environmental conditions (discussed above) to understand how vulnerability may alter under climate change.

Results

In this section, we report and interpret our findings on the demography of *P. repens*. Discussion of the more general implications of our case study for building DDMs appears in Discussion.

Vital rate regressions

Survival and fecundity regressions showed a strong dependence on size, indicating the importance of using a size structured demographic model (Table 1; see partial dependence plots, with fitting data, in Supplementary material Appendix A). Survival had a positive dependence on size for smaller plants (Table 1; Supplementary material Appendix A, Fig. A4). This trend is not surprising; seedlings become less susceptible to drought mortality by building a root system in their first year or so (Manders and Smith 1992). In contrast, adult survival was high, peaking near 94% for 1 m individuals, and an approximately constant function of size at moderate to wetter locations (over the relevant size range; Supplementary material Appendix A, Fig. A7) as expected given the well-developed root systems of larger plants (Rebelo unpubl.). The negative size dependence was more apparent at drier locations (Table 1; Fig. 2b) where the cumulative effect of water resource limitation or higher likelihood of droughts over an individual's lifetime is likely more pronounced (Bond 1980). As expected based on the sensitivity of a less extensive root system to moisture (Midgley 1988), seedling survival was also positively correlated with mean annual precipitation with slightly lower values in coastal areas (Table 1; Fig. 2a).

Growth and offspring size patterns were driven largely by a unimodal response to the number of summer SMD (Supplementary material Appendix A, Fig. A2, A11), indicating that *P. repens* distribution is potentially limited by both regions that are too dry and too wet (e.g. Fig. 2c, e, 5a). In general, growth was predicted to be higher in the western half of the range, compared to the east, driven primarily by poor performance under lower winter rainfall and higher summer rainfall (Fig. 2c, 5a, Supplementary material Appendix A, Fig. A2).

Seed production exhibits a qualitatively different spatial pattern than that of growth and survival, with higher values near the Great Karoo desert where large adult mortality was highest (Table 1, Fig. 2d). Flowering probability was strongly size dependent, with 50% of individuals flowering

by size 0.73 m and 90% of individuals flowering by size 1.35 m (Supplementary material Appendix A, Fig. A13). Edaphic factors had only a minor effect across all vital rate regressions. The spatial pattern of the time after which 90% of individuals have flowered (Fig. 2f) is inversely related to pattern for growth (Fig. 2c), with longer times in the east and toward arid interior regions.

We estimated a constant value (1.1%) of effective recruitment probability based on data from three locations, which we used across the entire landscape. It is challenging to validate this estimate because the literature on recruitment rates in Proteas in general, and *P. repens* specifically, is scattered with no consistent basis for estimation across different stages or time periods of observation, under different experimental treatments (controlled environments or in the field) (Witkowski 1991, Mustart et al. 2012), at different spatial and temporal environmental conditions in the field (Bond 1980, Musil 1991, Mustart and Cowling 1993, Maze and Bond 1996), or across observations at different times after fire (Bond 1984, Heelemann et al. 2008). In any case, the use of parent-seedling data to estimate recruitment implicitly incorporates all the sources of seed loss that operate in any given population (e.g. density dependence, presence of pathogens, granivory, dispersal to unsuitable locations, etc.; cf. Bond 1980) and hence it is not surprising that our estimation of effective recruitment probability is somewhat lower than under controlled conditions (Witkowski 1991, Mustart et al. 2012).

Population statistics

Population growth rate

Our DDM broadly predicted the spatial pattern of occurrence data that we used for model evaluation. Based on a threshold of $\lambda > 1$ (positive population growth rate) to indicate a prediction of presence, we correctly predicted 71% of (2338) presences and 70% of (3985) absences (Fig. 3). Alternative thresholds (to $\lambda = 1$) for determining accuracy of occurrence predictions would not have substantially improved predictions (Supplementary material Appendix A, Fig. A18), indicating that our model is reasonably well calibrated. AUC was 0.79, indicating that a randomly chosen occurrence point would be correctly classified 79% of the time. Of the 121 locations where demographic data were collected, 80% were predicted to have $\lambda > 1$. The mean value of λ was less than 1 at absence locations and increased strongly with observed abundance as expected (Fig. 3f). Given possible source-sink dynamics, that populations are often not at equilibrium, and that factors other than environmental suitability can strongly affect species' distributions (Svenning and Skov 2004, Latimer et al. 2006), we would not expect a perfect match between population growth rate and observed presence/absence data, but this level of qualitative agreement across thousands of sample sites is encouraging.

Our predictions exhibit some systematic bias, based on comparison with occurrence records. In particular, we over-predict *P. repens'* range in the northwest and under-predict
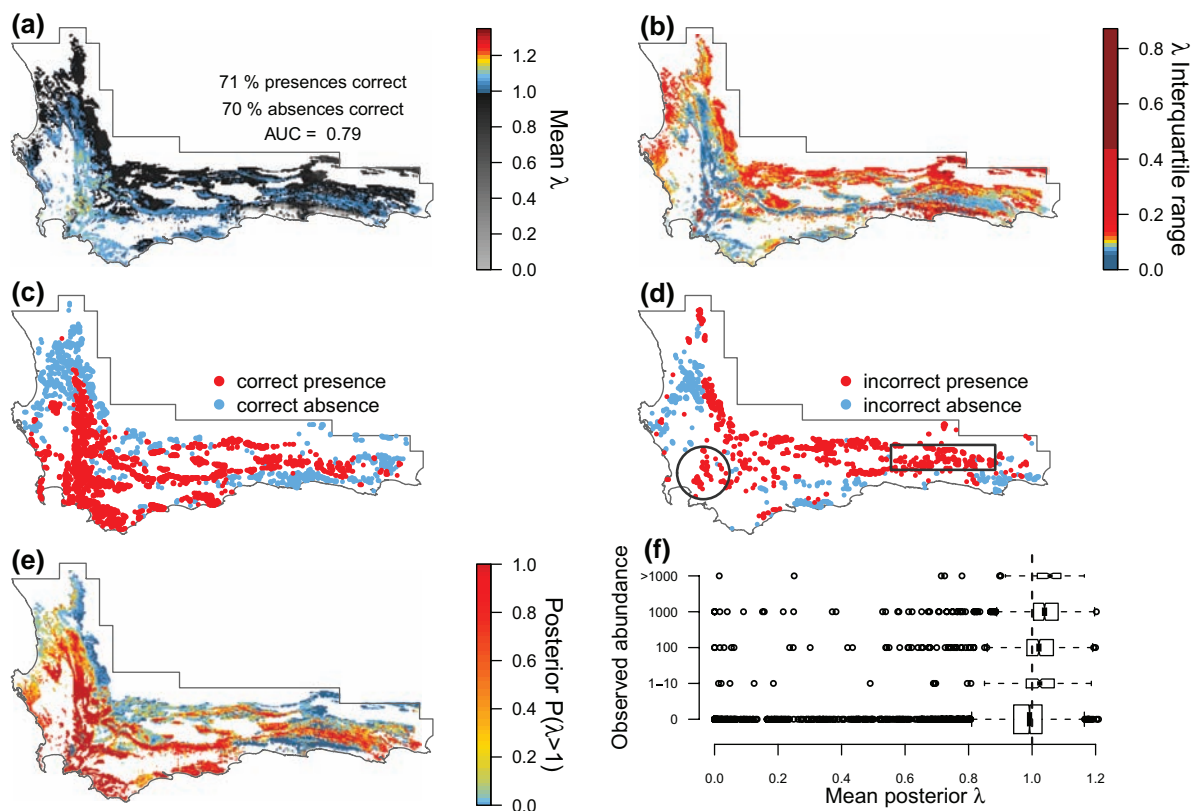

Figure 3. Predicted population growth rate (λ) and model evaluation. (a) Mean λ and (b) interquartile range of λ. (c–d) Evaluation of (a) using presence/absence data. Note that we focus on explaining prediction error in the main text in the southeast (boxed region) and the southwest (circled region), collectively referred to as the high summer SMD (summer soil moisture days). (e) Posterior probability that $\lambda > 1$, representing a viable population. (f) Evaluation of (a) using ordinal abundance data.

its distribution along the border with the Karoo desert in the eastern half of the Cape Floristic Region (Fig. 3d). Over prediction in the northwest is driven largely by high adult survival there, based on the observation that adult survival has both high values (Fig. 2b) and elasticities (Fig. 6e). While growth is also high in this region (Fig. 2c), elasticity for the growth intercept is low (Fig. 6a), suggesting that variation in individual growth rates is not driving prediction in this region. Winter temperature has a positive coefficient and elasticity value in the northwest in the adult survival model, suggesting that *P. repens* may have a weaker response to winter temperature than fitted in our model. This expectation is corroborated by under prediction in the mountains bordering the southern and western edges of the Great Karoo desert (Fig. 3d). There, adult survival is low (Fig. 2b) and most strongly associated with low winter temperatures (Supplementary material Appendix A, Fig. A7). Given that our survival data did not span the extremes of the winter temperature gradient (Fig. 1), it is not surprising that some bias exists in the fitted model. Overall, the evaluation data suggest that *P. repens* generally has a weaker response to winter temperature than our DDM estimates; it seems likely that colder locations do not strongly inhibit survival (higher recruitment is also reported; Holmes and Newton 2004) and warmer locations do not substantially enhance it.

Sensitivity/elasticity analysis

Elasticity analyses of intercept and size slope parameters are shown in Fig. 6. The intercept of the adult survival model had the largest elasticity, particularly in arid regions bordering the Great Karoo Desert where adult survival was low (Fig. 2b), indicating that populations there were limited by the ability of individuals to survive until reproduction. Many of the largest elasticity values are seen in the high summer SMD regions where our model incorrectly predicts absence: high-elevation areas in the southwest (circle in Fig. 3d covering the Boland and Hottentot-Holland mountains), and the

southeast (box in Fig. 3d). This indicates that our models predict that *P. repens* performs more poorly under increased precipitation than it actually does.

Sensitivity analysis to current environmental conditions showed strong responses to seasonal precipitation availability, primarily along the margins of the predicted distribution (Fig. 7). Increasing summer SMD reduced λ in the high summer SMD regions where our predictions were most errant, which we refer to collectively as the high summer SMD regions. Inspection of the elasticities related to summer SMD (Supplementary material Appendix A, Fig. A23) and partial dependence plots (Fig. 5a, Supplementary material Appendix A, Fig. A2, A10) reveal that the reduction in λ derives from strong unimodal responses to summer SMD in conditions where we extrapolate (summer SMD > 1.5). That is, a sharp decline in growth and offspring size was predicted in regions with relatively higher summer rainfall, which reduced λ. Observed presences in these regions (Fig. 3d) suggest that the true response to SDMSUM does not decline so sharply at higher values (see Discussion, section Population growth rate).

Uncertainty analysis

Spatially varying confidence in the predictions was apparent (Fig. 3b). The average interquartile range of λ across all pixels in the region was 0.13 (95% HPD interval 0.04–0.32). In parts of the region where $\lambda \geq 1$, however, the interquartile range was even narrower (0.04–0.13), indicating higher predictive confidence in areas that are environmentally suitable for the species. The highest uncertainties are in the projected unsuitable ($\lambda < 1$) high summer SMD regions, where few demographic data have been collected (Fig. 1). Uncertainty is high there because the model predicts a sharply declining response to very high summer SMD (solid line in Fig. 5a; see below for discussion of an alternative 'clamped' model that avoids this sharp decline) with high uncertainty about the magnitude of this response (above values of
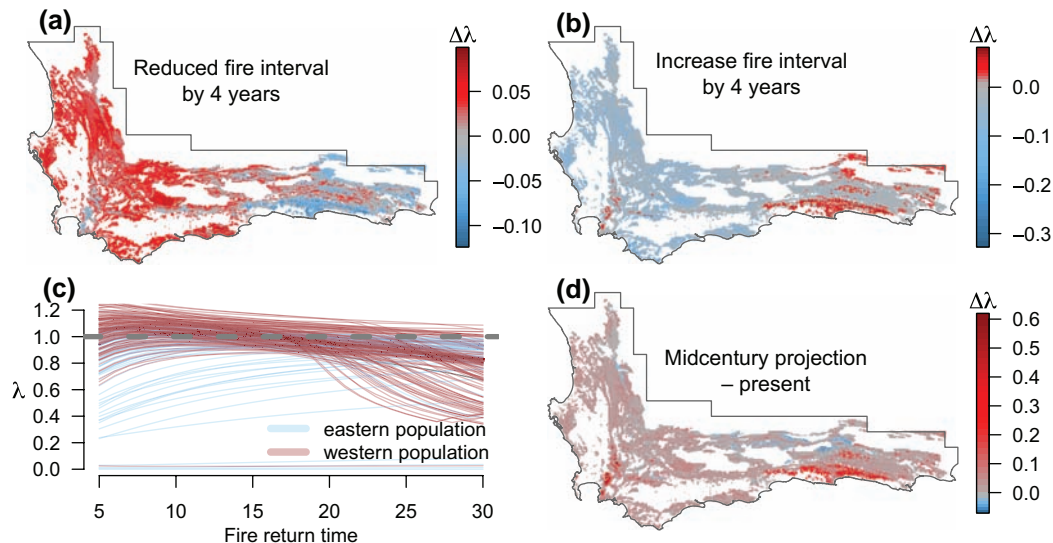

Figure 4. Projections under different scenarios. (a–b) Reducing (increasing) the observed fire return time by 4 yr. $\Delta\lambda$ is calculated as the predicted value in each scenario minus the present day predictions in (c) Variation of mean λ as a function of fire return time at a random sample (meant to reflect the spectrum of responses) of 200 pixels on the landscape. The horizontal dashed line indicates $\lambda = 1$. (d) The difference between present day predictions (in Fig. 3a) and projections under future climate change scenario with temperature increased by 1 degree and precipitation reduced by 10%.

1.5 in Fig. 5a). Because summer soil moisture has high elasticity (Supplementary material Appendix A, Fig. A23h, i), this results in low but uncertain values of λ.

The posterior probability of positive population growth ($P(\lambda \geq 1)$) is a useful metric for assessing confidence that *P. repens* could survive in a location, given all the uncertainties in the underlying environmental relationships. Compared to the observed abundances in the Protea Atlas evaluation dataset and previous abundance modeling (Chakraborty et al. 2011), $P(\lambda \geq 1)$ captured the overall pattern of site suitability (Fig. 3e). Grid cells with absences had relatively low values of $P(\lambda \geq 1)$ (0.44 ± 0.41 (SD)), while cells with higher observed abundance were predicted to have larger $P(\lambda \geq 1)$: 1–10 individuals (0.63 ± 0.40 (SD)), 10–100 individuals (0.63 ± 0.39 (SD)), 100–1000 (0.74 ± 0.35 (SD)), and > 1000 (0.82 ± 0.31 (SD)). These patterns corroborate those found for mean λ predictions in Fig. 3f.

Scenario-based projections

Longer fire return times decreased λ throughout much of the range except in the high summer SMD regions (Fig. 4b). The decreasing relationship between λ and fire return time in the majority of cells is apparent in Fig. 4c and is driven largely by the adult survival. In the model, adult mortality increases with size (and consequently age; Supplementary material Appendix A, Fig. A7) which means that longer fire return times allows more individuals to die without reproducing. Since the species relies on an aerial serotinous seed bank, the main threat to future persistence of the species in most areas appears to be long fire return intervals resulting in adult death and the loss of the seed bank (cf. Bond 1980). Sensitivity to decreased fire return intervals is similarly driven by adult survival patterns (Fig. 4a).

The projected rates of population growth for *P. repens* increased under midcentury scenarios for temperature, precipitation and fire regime changes. The amount of suitable habitat in the east also increased (Fig. 4d). The largest increases are observed in the high summer SMD regions, because the predicted future decline in precipitation shifts these populations toward more favorable summer SMD patterns closer to the maximum of individual growth in Fig. 5a. Comparing the parameter elasticities between present (Fig. 6, Supplementary material Appendix A, Fig. A22–A23) and future (Supplementary material Appendix A, Fig. A25–A26) shows a reduced elasticity to intercepts and size-based slopes in the future, indicating that the more favorable, dryer future conditions buffered against population declines across the core and the wetter parts of the range of the species. Similarly, models predicted a reduced elasticity in the future to parameters describing *P. repens'* response to temperature and precipitation, particularly in colder and wetter regions. In contrast, at the drier interior margin of the range, the model projects a drop in population growth rate by mid-century (blue in Fig. 4d). These lower population growth rates decline further if fire return times continue to decrease, since frequent fire would then kill adults before they have produced many seeds; this fire–climate interaction would be a concern primarily in the east, along the northern boarder with the Great Karoo Desert, where growth is projected to be lower (also observed by Kraaij et al. 2013).
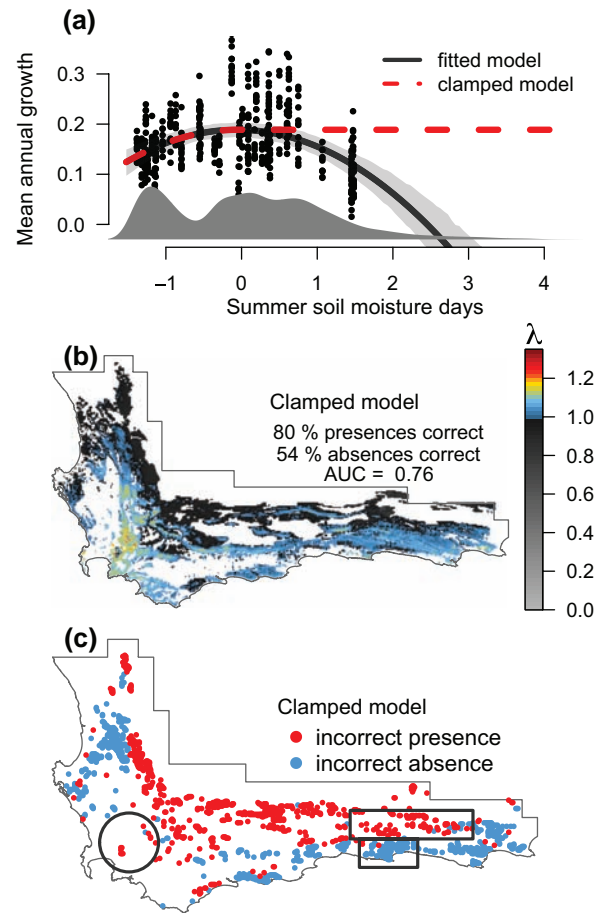

Figure 5. Modeling experiments to understand the role of summer precipitation on growth and consequently λ. (a) Growth data (black dots) and partial dependence plot of the fitted model (black line; 95% credible interval in light grey) on summer SMD, a primary environmental gradient driving differences in λ. Dark grey density function along the horizontal axis shows the distribution of summer SMD across the entire region, highlighting that extrapolation is necessary for values > 1.5. The dashed line indicates a modeling experiment in which we clamped the (black) response curve at is maximum value to reflect our expectation that growth does not decline with increasing water availability. The results of this experiment, in terms of λ and predicted occurrence patterns, are shown in (b) and (c), respectively.

Discussion

DDMs are valuable for gaining an understanding of the demographic drivers of species' ranges and their relationship to environmental variation. For large data sets and vital rate regressions with good fit, DDMs should enable robust projections of range-wide population-level patterns. When data are sparse and numerous assumptions are required to develop models, DDMs allow us to study the implications of our existing knowledge for range-wide population-level patterns and evaluate the consequences of different assumptions or hypotheses. Moreover, these models address many of the limitations of current demographic and distribution models and follow suggestions in recent reviews on biogeography and forecasting population dynamics (Schurr et al. 2012b, Crone et al. 2013). Below, we illustrate how our
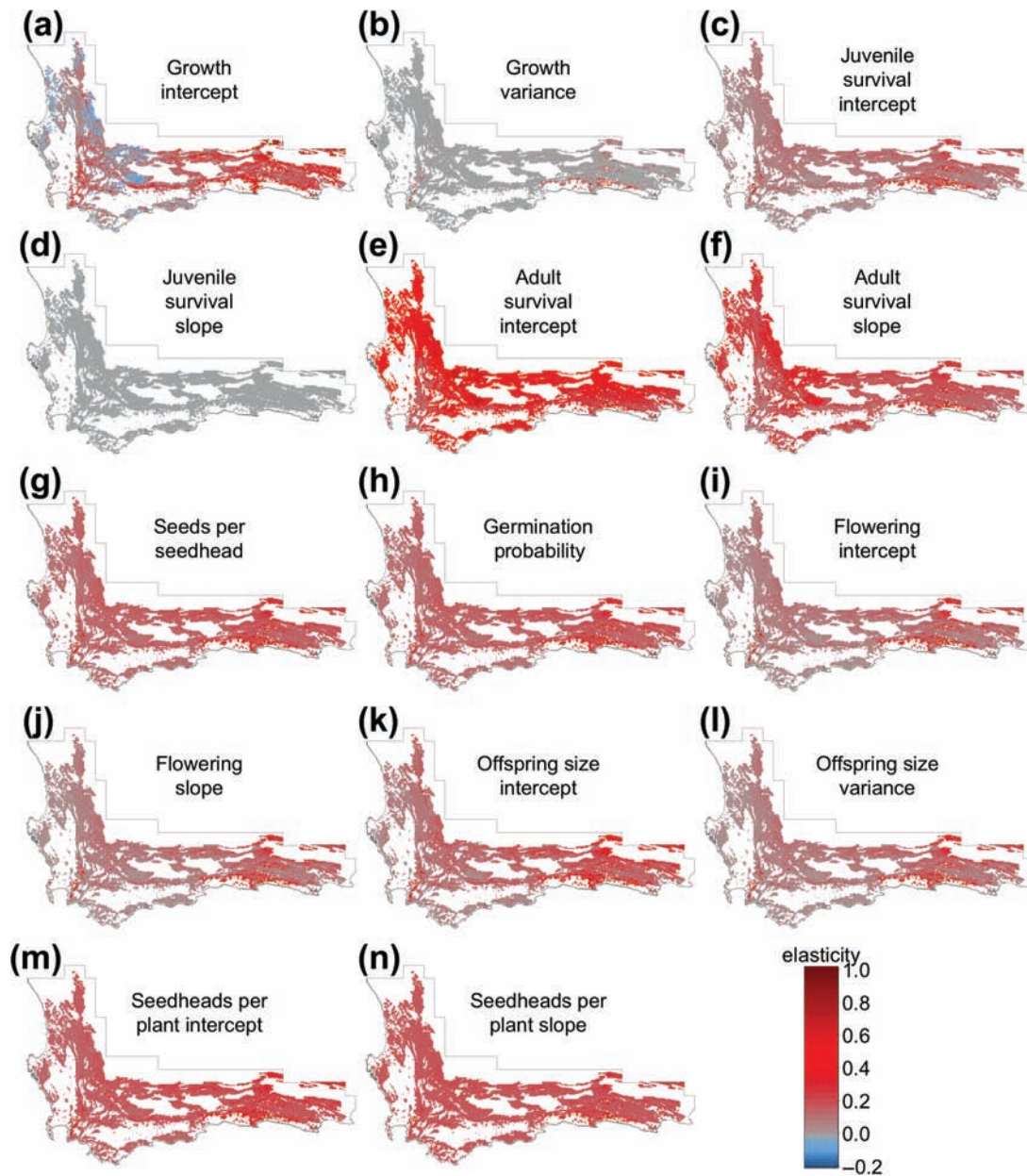
Figure 6. Elasticity of λ to regression coefficients for intercepts and size-related parameters. Elasticities for environmental predictors are shown in Supplementary material Appendix A, Fig. A21.

models for *P. repens* offer generally applicable insights, while identifying critical knowledge gaps to prioritize for further study.

The overall accuracy of our predictions is encouraging for predicting species' ranges from demography, given the limitations of available data and the challenges of modeling complex population dynamics. We were able to capture general spatial patterns in occurrence and abundance in spite of the high spatial and temporal variation reported for various life history stages of *P. repens* as well as other *Protea* species (Bond 1984, Midgley 1988, Witkowski 1991, Maze and Bond 1996, Holmes and Newton 2004, Higgins et al. 2008, Mustart et al. 2012, Kraaij et al. 2013, Nottebrock et al. 2013). Since climate variables are resolved only at 1 min, sub-cell resolution of habitat conditions that otherwise

might yield higher resolution predictions in topographically heterogeneous landscapes are unknown. There will inevitably be recruitment limitation at many spatial scales (within cells, among neighboring cells and distant cells) (cf. Latimer et al. 2009). Furthermore, because our DDMs do not use 'absence' data for model fitting, the regressions of demographic parameters on environment are not constrained by information from unsuitable sites and require extrapolation to unsuitable regions. The occurrence patterns used for evaluation are also imperfect; source-sink dynamics can lead to 'false' presences (Pulliam 2000, Schurr et al. 2012b) because a few consecutive favorable years are often all that is need to for plants to establish (seedlings typically reaching the water table typically within the first year; Manders and Smith 1992), even though in an average year they would not
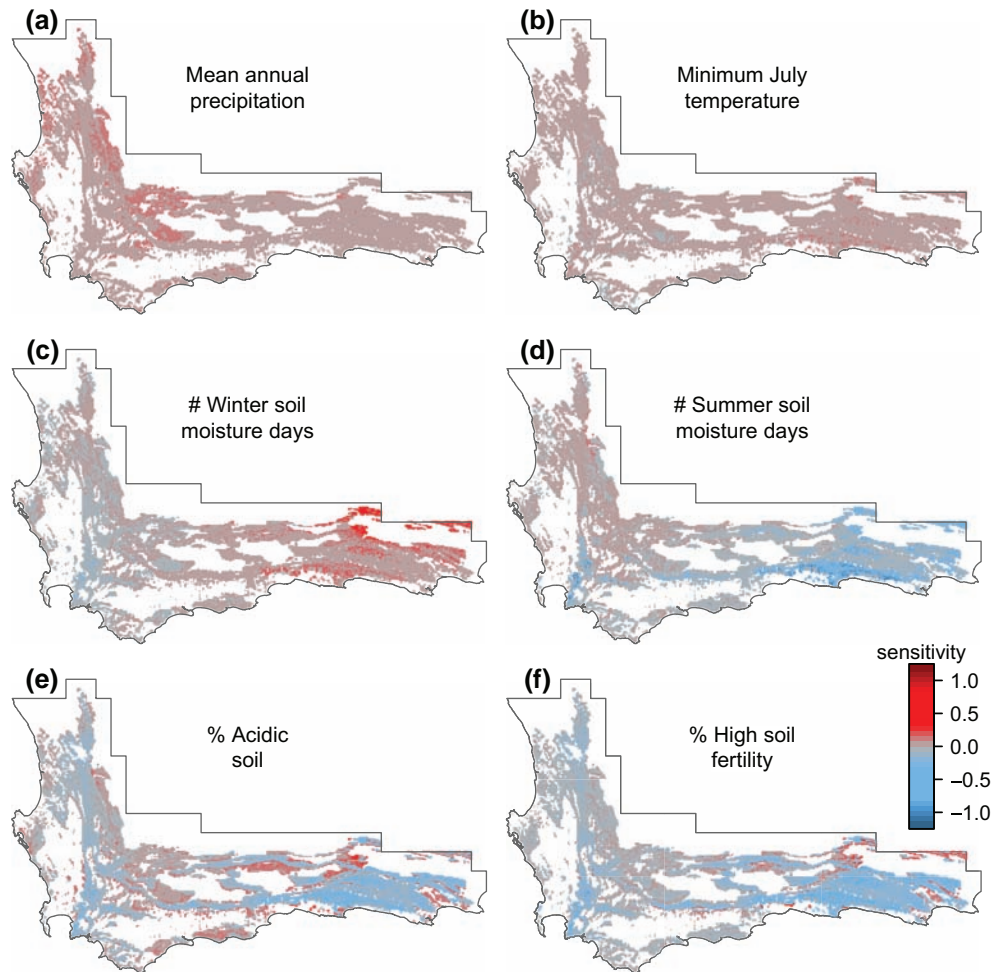
Figure 7. Sensitivity of λ to environmental values. Sensitivity was calculated by increasing the (standardized) values of environmental predictors by 0.1. This sensitivity incorporates the effect of each predictor across all vital rate models (Fig. 2) simultaneously.

be expected to persist. Furthermore, absences do not indicate that a cell was searched completely and that *P. repens* was not found, but simply that it was not found at a site where one or more surveys occurred (Gelfand et al. 2005). Taken together, our ability to capture broad scale demographically calibrated population patterns across environmentally heterogeneous landscapes with a limited number of stratified life history stage observation sites supports optimism about these methods.

Regressions

An important component of our DDMs for *P. repens* was our use of limited, existing data on the life cycle to parameterize a full IPM. Our ability to work around limited data relied in part on particular attributes of *Protea* life history, although other taxonomic groups may be modeled under similar assumptions. It will be extremely rare to have high-resolution data on all life-stage transitions across the full range of environmental conditions covered by a species' range, especially for long-lived individuals. Indeed, it will often be necessary to adopt rapid, one-shot sampling strategies to be able to measure demographic rates at replicate sites across multiple environmental gradients. Coverage of many populations will

be necessary to capture environmental dependencies and to reduce the chance of mistaking idiosyncrasies of particular populations for broader patterns. For these reasons, strategies for using sub-optimal data will be essential for the further development of DDMs. Ideal growth data includes sequential observations of the same individuals over many years. However, average growth increments might be sufficient in some cases for species whose age, current size, and allometry can be determined. Ideal survival data might consist of sequential observations of marked individuals whose growth was measured. Capture–mark–recapture data may be available for some species to estimate survival probabilities even if size-structure is unavailable. Ideal fecundity data might consist of observations of the number of offspring produced by each individual, recruitment rates (if applicable), and the resultant sizes of the offspring. However, parent-seedling ratios may provide enough information to describe effective fecundity if it is not critical to break recruitment into its various component processes leading from parents in one period to offspring in the next. Clearly, the usefulness of different data sources for DDMs with requires further study.

Many insights are possible simply from mapping vital rates. In our example, the spatial patterns of growth, offspring size, seedling survival, adult survival and flowering

probability are relatively similar, with favorable values in the mountainous regions of the interior Cape Floristic Region (typically bounded by the coast and arid desert areas; Fig. 2a–c, e, f). In contrast, seedhead production is most favorable in mountainous areas closer to the Great Karoo desert to the north and east of the CFR (Fig. 2d). Our DDM predicts many fewer populations with $\lambda > 1$ in the region where seedhead production is high. Taken together, the fecundity, survival, and population growth rate patterns suggest multiple competing explanatory hypotheses for low values of λ in this region that could be evaluated with further study: 1) higher fecundity trades off with lower survival and longer times to flowering in more arid regions to sustain populations (Fig. 2); 2) fecundity is even higher in arid regions than our model predicts, which would account for our under prediction of viable populations in this regions (Fig. 3a); and 3) although the fecundity predictions are reasonable, under prediction of viable populations in arid regions occurs because adult survival is under predicted (Fig. 2b). This line of inquiry illustrates how mechanistic hypotheses can emerge from the understanding gained from DDMs and their successive refinement. Simulations exploring these hypotheses can help evaluate them and guide further, targeted data collection. Simulations (Supplementary material Appendix A, Fig. A28) suggest hypothesis 3) is more likely than 1) and 2) because increasing the intercept of adult survival by one standard deviation improves predictions of occurrence (based on $\lambda > 1$) in this region, while in contrast, an increase of three standard deviations in the intercept of seedhead production model is needed to achieve similar results (Supplementary material Appendix A, Fig. A28).

Population growth rate

In this case study, the DDM provided some clear mechanistic insights about *P. repens* demography and biogeography. Distributional patterns are driven largely by the best locations for adult survival, and depend less on whether the species can grow rapidly or have high reproduction there (see Fig. 2 for comparing vital rate maps to predicted λ in Fig 3a). Summer precipitation has the largest coefficients in the regression for adult survival (Table 1), with the result that populations in the most arid areas risk having most adults die before fire provides a new recruitment opportunity (cf. Fig. 1k, Supplementary material Appendix A, Fig. A7; Bond 1980). Flowering time is substantially longer in the east (which may reflect the local ecotypic variation found by Mustart et al. [2012] and Kraaij et al. [2013]), so if fire frequency were to further increase, populations there might become threatened. At the same time, the higher temperatures and frequent fires expected in the future are predicted to increase population growth rates in the core of the range through their effects on growth, especially in the coolest and wettest areas (Fig. 4a, d; Table 1).

Adult survival patterns drive many of the predictions; it generally shows the highest elasticities (Fig. 6), making it a more important influence on population growth rates than expert opinion would suggest (<www.proteaatlas.org.za/mortality.htm>). Previous experimental work on similar species tended to show that recruitment rates (including first-year survival) most often had the highest sensitivities (Latimer et al. 2009). But in those models, we assumed a constant and relatively high rate of adult survival, whereas the extensive field surveys in this study revealed an unexpectedly high variance in adult mortality rates. The DDMs show that when adult survival is lower and more variable, this vital rate tends to drive population growth and persistence. More consistent with our previous intuition is that population growth rates are most sensitive to reductions in precipitation at drier sites, based on unimodal response of growth to summer and winter soil moisture availability (Table 1; Fig. 5a, Supplementary material Appendix A, Fig. A2).

Errors in our predictions also lead to insights. First, our models predict that populations are unsustainable in the cool and very wet Boland-Hottentots Holland Mountains (circled region in Fig. 3d), although many presences and no absences have been observed there (Fig. 3c, d). In order to build an IPM in this region, regression models were extrapolated three standard deviations beyond the range of the data along the upper end of the summer SMD gradient, corresponding to an extremely high rainfall region (up to 3000 mm yr^{-1}; Fig. 5a). The strong unimodal response to summer SMD apparently overestimated the steep decline in growth in response to very high summer moisture availability. While a decline in growth toward the lower end of the moisture gradient is expected due to water limitation, we have no reason to expect lower performance in *P. repens* with higher rainfall (with the exception of unsuitable wetland soils; cf. Seiben et al. 2004). This suggests that favorable modifications to the vital rate models (e.g. increased intercept or size-slope parameters) would improve predictive performance there. To explore the implications of these expectations and the consequences of extrapolating to high summer SMD values, we clamped the growth and (closely related) offspring size models at their maximum values (dashed line in Fig. 5a) and calculated λ. Prediction of presences were drastically improved in the high summer SMD regions (compare Fig. 5b, c, vs Fig. 3a, d), though there were still some incorrect predictions along the coast (smaller grey rectangle in Fig. 5c vs Fig. 3d). In any case, it is clear that accurately characterizing *P. repens'* response to the full range of summer moisture levels and exploring other predictors to describe its absence from the eastern coastal region will require further data collection and investigation of potential ecotypic differentiation in this part of the range (cf. Mustart et al. 2012, Kraaij et al. 2013). This type of modeling experiment highlights how DDMs facilitate an understanding of the link between specific ecological processes and environmental gradients when extrapolating models.

A potentially important omission from our model for *P. repens* is density dependence. Literature on the influence of density dependence in Protea populations is mixed: while there is some indication that it is most apparent in seedhead production with an Allee effect (Maze and Bond 1996, Nottebrock et al. 2013), others have found that seedling survival density dependent effects may be weak (Bond 1980, Midgley 1988, Latimer et al. 2009), and yet others have found negative density dependence of fecundity and seedling parent ratios in Proteaceae (Bond et al. 1984, 1995, Esler and Cowling 1990). In part, the variable evidence for density-dependence reflects a fire-dominated system where,

unlike many other systems, fires continually reset population structure every 17 yr, on average. Our existing models may be sufficient to implicitly incorporate density dependence if it occurs primarily during the first few years after fire; the parent-seedling ratios that we use to parameterize the seedling survival model implicitly include the outcome of whatever density dependence may occur during establishment and maturation. More data is needed to determine the consequence of density dependence among adults. In any case, our models appear to yield reasonable predictions without explicitly accounting for density dependence.

Limitations of DDMs

Though the goal of this paper is to show that DDMs can provide useful insights even in the absence of large data sets, they do have some limitations. The regressions described here, which combine data across sites and years, require sampling across the full range of the relevant environmental gradients with sufficient density of samples to accurately estimate the variation in demographic responses. There are also some limitations to the relatively simple models that we impose to take advantage of the IPM framework: although we model temporal dynamics, we do not explicitly incorporate dispersal to fully estimate spatio-temporal population dynamics (cf. Pagel and Schurr 2012) for proteas. Hence our predictions can be interpreted as short-term projections or as potential distributions, in the absence of dispersal limitation. In principal, however, our demographic models could readily be coupled with a dispersal model (cf. Merow et al. 2011, Pagel and Schurr 2012, Cabral et al. 2013) such as the secondary wind dispersal model described by Schurr et al. (2005). Under certain simplifying assumptions about spatial structure, spatially explicit IPMs can be used (Jongejans et al. 2011). Environmental stochasticity is also omitted from our models, although many tools exist for to accommodate it in matrix projection models that are readily adapted to IPMs (cf. Caswell 2001, Metcalf et al. 2008). In our case study, the negative influence of stochastic variation in fire return times would likely decrease our estimates of λ because serotinous nonsprouters like *P. repens* are adversely affected if fires arrive before reproductive maturity and their fecundity may decline if fire intervals are very long (Bond 1980). Our models also omit biotic interactions, although these could in principle be included following methods used in other IPM studies: e.g. competition (Adler et al. 2010, 2011); disease (Bruno et al. 2011); seed predation (Dahlgren and Ehrlén 2011).

Conclusions

Until more environmentally stratified demographic data become available, interpolation, imputation, and expert opinion will play a key role in linking demography to geographic distributions. The regression models used by IPMs are valuable for interpolation, while Bayesian models facilitate imputation, integration of prior information, and propagating uncertainties through various stages of the analysis. Using these tools, we can determine the most sensitive attributes of life history and environmental dependence to better structure demographic sampling schemes. When demographic data are missing, occurrence data, which are often more readily available, can be used to explore different demographic assumptions in attempts to reproduce observed occurrence patterns. This 'inverse modeling' strategy can help in exploring how qualitative differences in species' distributions relate to different demographic mechanisms. As predictions rely more heavily on assumptions, inference should shift from quantitative predictions to qualitative ones, such as the effect of positive versus negative relationships on population persistence. Insights from these explorations can provide proof of concept for sampling designs focused on different parts of geographic or environmental space for future demographic studies.

DDMs represent a next step beyond the correlative occurrence models that are typically used to understand or predict geographic distributions. This advance comes at the expense of greater data requirements; however we argue that process-based modeling, even driven by expert opinion and limited data, can provide valuable insights into population and range dynamics that are not possible with occurrence models. DDMs can predict temporal dynamics and facilitate understanding of the relationship between specific demographic processes and environment, which is particularly critical when extrapolating. DDMs can potentially predict many more population statistics than those illustrated here, including responses to stochastic environments, size distributions, passage times to critical life history events, life expectancies, reproductive rates, etc. (cf. Caswell 2001). DDMs can take advantage of valuable existing occurrence data sets to evaluate model assumptions and predictions. Finally, if further data collection is an option, we hypothesize that even short term environmentally stratified demographic data are likely to be more insightful than a large number of occurrence records collected with the same amount of human and financial resources. While fully parameterized process-based models of spatiotemporal population dynamics remain data hungry and elusive, DDMs represent a manageable step in the right direction.

References

Adler, P. B. et al. 2010. Coexistence of perennial plants: an embarrassment of niches. – Ecol. Lett. 13: 1019–1029.

Adler, P. B. et al. 2011. Forecasting plant community impacts of climate variability and change: when do competitive interactions matter? – J. Ecol. 100: 478–487.

Bond, W. 1980. Fire and senescent Fynbos in the Swartberg, southern Cape. – S. Afr. For. J. 114: 68–71.

Bond, W. J. 1984. Fire survival of Cape Proteaceae-influence of fire season and seed predators. – Plant Ecol. 56: 65–74.

Bond, W. J. et al. 1984. Variation in seedling recruitment of Cape Proteaceae after fire. – J. Ecol. 72: 209–221.

Bond, W. J. et al. 1995. Fire life histories and the seeds of chaos. – Ecoscience 2: 252–260.

Bruno, J. et al. 2011. Impacts of aspergillosis on sea fan coral demography: modeling a moving target. – Ecol. Monogr. 81: 123–139.

Buckley, L. B. et al. 2011. Does including physiology improve species distribution model predictions of responses to recent climate change? – Ecol. Lett. 92: 2214–2221.

Bugmann, H. 2001. A review of forest gap models. – Clim. Change 51: 259–305.

Cabral, J. S. et al. 2013. Impacts of past habitat loss and future climate change on the range dynamics of South African Proteaceae. – Divers. Distrib. 19: 363–376.

Caswell, H. 2001. Matrix population models: construction, analysis, and interpretation. – Sinauer.

Chakraborty, A. et al. 2011. Point pattern modelling for degraded presence-only data over large regions. – J. R. Stat. Soc. C 60: 757–776.

Chapman, D. S. et al. 2014. Phenology predicts the native and invasive range limits of common ragweed. – Global Change Biol. 20: 192–202.

Clark, J. 2005. Why environmental scientists are becoming Bayesians. – Ecol. Lett. 8: 2–14.

Clark, J. and Bjørnstad, O. 2004. Population time series: process variability, observation errors, missing values, lags, and hidden states. – Ecol. Lett. 85: 3140–3150.

Clark, J. S. et al. 2010. High-dimensional coexistence based on individual variation: a synthesis of evidence. – Ecol. Monogr. 80: 569–608.

Clark, J. S. et al. 2011. Individual-scale inference to anticipate climate-change vulnerability of biodiversity. – Phil. Trans. R. Soc. B 367: 2865–2873.

Coulson, T. 2012. Integral projections models, their construction and use in posing hypotheses in ecology. – Oikos 121: 1337–1350.

Crone, E. et al. 2013. Ability of matrix models to explain the past and predict the future of plant populations. – Conserv. Biol. 27: 968–978.

Crozier, L. and Dwyer, G. 2006. Combining population-dynamic and ecophysiological models to predict climate-induced insect range shifts. – Am. Nat. 167: 853–866.

Dahlgren, J. P. and Ehrlén, J. 2011. Incorporating environmental change over succession in an integral projection model of population dynamics of a forest herb. – Oikos 120: 1183–1190.

de Klerk, H. M. et al. 2007. Monitoring manual: ecological fire management. – CapeNature Scientific Services.

de Klerk, H. M. et al. 2012. Evaluation of satellite-derived burned area products for the fynbos, a Mediterranean shrubland. – Int. J. Wildland Fire 21: 36–47.

Doak, D. F. and Morris, W. F. 2010. Demographic compensation and tipping points in climate-induced range shifts. – Nature 467: 959–962.

Easterling, M. R. et al. 2000. Size-specific sensitivity: applying a new structured population model. – Ecology 81: 694–708.

Ellner, S. and Rees, M. 2006. Integral projection models for species with complex demography. – Am. Nat. 167: 410–428.

Esler, K. J. and Cowling, R. M. 1990. Effects of density on the reproductive output of *Protea lepidocarpodendron*. – S. Afr. J. Bot. 56: 29–33.

Fielding, A. and Bell, J. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. – Environ. Conserv. 24: 38–49.

Franklin, J. 2010. Moving beyond static species distribution models in support of conservation biogeography. – Divers. Distrib. 16: 321–330.

Gelfand, A. et al. 2005. Modelling species diversity through species level hierarchical modelling. – J. R. Stat. Soc. C 54: 1–20.

Goldblatt, P. and Manning, J. 2002. Plant diversity of the Cape Region of southern Africa. – Ann. Missouri Bot. Gard. 89: 281–302.

Hadfield, J. D. 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. – J. Stat. Softw. 33: 1–22.

Heelemann, S. et al. 2008. Fire season effects on the recruitment of non-sprouting serotinous Proteaceae in the eastern (bimodal rainfall) fynbos biome, South Africa. – Austral Ecol. 33: 119–127.

Helmuth, B. et al. 2006. Mosaic patterns of thermal stress in the rocky intertidal zone: the implications for climate change. – Ecol. Monogr. 76: 461–479.

Higgins, S. I. et al. 2008. Costs of persistence and the spread of competing seeders and sprouters. – J. Ecol. 96: 679–686.

Holmes, P. M. and Newton, R. J. 2004. Patterns of seed persistence in South African fynbos. – Plant Ecol. 172: 143–158.

Holt, R. D. 2009. Bringing the Hutchinsonian niche into the 21st century: ecological and evolutionary perspectives. – Proc. Natl Acad. Sci. USA 106: 19659–19665.

Huntley, B. et al. 2010. Beyond bioclimatic envelopes: dynamic species' range and abundance modelling in the context of climatic change. – Ecography 33: 621–626.

IPCC-WGI 2013. Climate change 2013: the physical science basis. Annex I: atlas of global and regional climate projections. – Final Draft Underlying Scientific-Technical Assessment. Assessment Report AR5, Geneva, WMO, <www.climate-change2013.org/>.

Jongejans, E. et al. 2011. Importance of individual and environmental variation for invasive species spread: a spatial integral projection model. – Ecol. Lett. 92: 86–97.

Kearney, M. and Porter, W. 2009. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. – Ecol. Lett. 12: 334–350.

Keith, D. A. et al. 2008. Predicting extinction risks under climate change: coupling stochastic population models with dynamic bioclimatic habitat models. – Biol. Lett. 4: 560–563.

Kobe, R. K. et al. 1995. Juvenile tree survivorship as a component of shade tolerance. – Ecol. Appl. 5: 517.

Kohyama, T. 1992. Density–size dynamics of trees simulated by a one-sided competition multi-species model of rain forest stands. – Ann. Bot. 70: 452–460.

Kraaij, T. et al. 2013. Proteaceae juvenile periods and post–fire recruitment as indicators of minimum fire return interval in eastern coastal fynbos. – Appl. Veg. Sci. 16: 84–94.

Latimer, A. M. et al. 2006. Building statistical models to analyze species distributions. – Ecol. Appl. 16: 33–50.

Latimer, A. M. et al. 2009. Experimental biogeography: the role of environmental gradients in high geographic diversity in Cape Proteaceae. – Oecologia 160: 151–162.

Lischke, H. et al. 2006. TreeMig: a forest-landscape model for simulating spatio-temporal patterns from stand to landscape scale. – Ecol. Model. 199: 409–420.

Manders, P. T. and Smith, R. E. 1992. Effects of artificially established depth to water-table gradients and soil typ on the growth of Cape fynbos and forest plants. – S. Afr. J. Bot. 58: 195–201.

Martin, A. D. et al. 2011. MCMCpack: Markov Chain Monte Carlo in R. – J. Stat. Softw. 42: 1–21.

Maze, K. E. and Bond, W. J. 1996. Are Protea populations seed limited? Implications for wildflower harvesting in Cape fynbos. – Austral Ecol. 21: 96–105.

Merow, C. et al. 2011. Developing dynamic mechanistic species distribution models: predicting bird-mediated spread of invasive plants across northeastern North America. – Am Nat 178: 30–43.

Merow, C. et al. 2014. Advancing population ecology with integral projection models: a practical guide. – Methods Ecol. Evol. 5: 99–110.

Metcalf, C. et al. 2008. Evolution of flowering decisions in a stochastic, density-dependent environment. – Proc. Natl Acad. Sci. USA 105: 10466–10470.

Midgley, G. F. et al. 2003. Developing regional and species-level assessments of climate change impacts on biodiversity in the Cape Floristic Region. – Biol. Conserv. 112: 87–97.

Midgley, J. J. 1988. Mortality of Cape Proteaceae seedlings during their first summer. – S. Afr. For. J. 145: 9–12.

Morin, X. and Thuiller, W. 2009. Comparing niche- and process-based models to reduce prediction uncertainty in species range shifts under climate change. – Ecology 90: 1301–1313.

Morin, X. et al. 2008. Tree species range shifts at a continental scale: new predictive insights from a process-based model. – J. Ecol. 96: 784–794.

Musil, C. 1991. Seed bank dynamics in sand plain lowland fynbos. – S. Afr. J. Bot. 57: 131–142.

Mustart, P. J. and Cowling, R. M. 1993. The role of regeneration stages in the distribution of edaphically restricted fynbos Proteaceae. – Ecology 74: 1490–1499.

Mustart, P. J. et al. 2012. Wide variation in post-emergence desiccation tolerance of seedlings of fynbos proteoid shrubs. – S. Afr. J. Bot. 80: 110–117.

Nottebrock, H. et al. 2013. Effects of intraspecific and community density on the lifetime fecundity of long-lived shrubs. – Perspect. Plant Ecol. Evol. Syst. 15: 150–161.

Pagel, J. and Schurr, F. M. 2012. Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. – Global Ecol. Biogeogr. 21: 293–304.

Plummer, M. 2003. JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. – Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), 20–22 March.

Pulliam, H. R. 2000. On the relationship between niche and distribution. – Ecol. Lett. 3: 349–361.

Rebelo, T. 2001. A field guide to the Proteas of southern Africa. – Fernwood Press.

Rebelo, T. 2002. The Protea Atlas Project – technical report. – <www.proteaatlas.org.za/history.htm>.

Richards, M. B. et al. 1997. Soil factors and competition as determinants of the distribution of six fynbos Proteaceae species. – Oikos 79: 394–406.

Schulze, R. 1997. South African atlas of agrohyrdology and climatology. – Tech. Rep., Report TT82/96, Water Research Commission, Pretoria, South Africa.

Schurr, F. et al. 2005. A mechanistic model for secondary seed dispersal by wind and its experimental validation. – J. Ecol. 93: 1017–1028.

Schurr, F. M. et al. 2012a. Fynbos Proteaceae as model organisms for biodiversity research and conservation. – S. Afr. J. Sci. 108: 1–4.

Schurr, F. M. et al. 2012b. How to understand species' niches and range dynamics: a demographic research agenda for biogeography. – J. Biogeogr. 39: 2146–2162.

Seiben, E. et al. 2004. Vegetation of high-altitude fens and restio marshlands of the Hottentots Holland Mountains, western Cape, South Africa. – Bothalia 34: 141–153.

Snell, R. et al. 2014. Using dynamic vegetation models to simulate plant range shifts. – Ecography 37: 1184–1197.

Svenning, J.-C. and Skov, F. 2004. Limited filling of the potential range in European tree species. – Ecol. Lett. 7: 565–573.

Thuiller, W. et al. 2006. Predicting patterns of plant species richness in megadiverse South Africa. – Ecography 29: 733–744.

van Wilgen, B. W. et al. 2010. Fire management in Mediterranean-climate shrublands: a case study from the Cape fynbos, South Africa. – J. Appl. Ecol. 47: 631–638.

Vanderwel, M. C. et al. 2013. Climate-related variation in mortality and recruitment determine regional forest-type distributions. – Global Ecol. Biogeogr. 22: 1192–1203.

Wilson, A. M. and Silander, J. A. 2013. Estimating uncertainty in daily weather interpolations: a Bayesian framework for developing climate surfaces. – Int. J. Climatol. doi: 10.1002/joc.3859

Wilson, A. M. et al. 2010. A hierarchical Bayesian model of wildfire in a Mediterranean biodiversity hotspot: implications of weather variability and global circulation. – Ecol. Model. 221: 106–112.

Witkowski, E. 1991. Growth and competition between seedlings of *Protea repens* (L.) L. and the alien invasive, *Acacia saligna* (Labill.) Wendl. in relation to nutrient availability. – Funct. Ecol. 5: 101–110.