

MACROECOLOGICAL
METHODS



Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics

Jörn Pagel^{1,2*} and Frank M. Schurr¹

¹Plant Ecology and Nature Conservation, Institute of Biochemistry and Biology, University of Potsdam, Maulbeerallee 2, 14469 Potsdam, Germany, ²Earth System Analysis, Potsdam Institute for Climate Impact Research (PIK), Telegraphenberg, 14412 Potsdam, Germany

ABSTRACT

Aim The study and prediction of species–environment relationships is currently mainly based on species distribution models. These purely correlative models neglect spatial population dynamics and assume that species distributions are in equilibrium with their environment. This causes biased estimates of species niches and handicaps forecasts of range dynamics under environmental change. Here we aim to develop an approach that statistically estimates process-based models of range dynamics from data on species distributions and permits a more comprehensive quantification of forecast uncertainties.

Innovation We present an approach for the statistical estimation of process-based dynamic range models (DRMs) that integrate Hutchinson's niche concept with spatial population dynamics. In a hierarchical Bayesian framework the environmental response of demographic rates, local population dynamics and dispersal are estimated conditional upon each other while accounting for various sources of uncertainty. The method thus: (1) jointly infers species niches and spatiotemporal population dynamics from occurrence and abundance data, and (2) provides fully probabilistic forecasts of future range dynamics under environmental change. In a simulation study, we investigate the performance of DRMs for a variety of scenarios that differ in both ecological dynamics and the data used for model estimation.

Main conclusions Our results demonstrate the importance of considering dynamic aspects in the collection and analysis of biodiversity data. In combination with informative data, the presented framework has the potential to markedly improve the quantification of ecological niches, the process-based understanding of range dynamics and the forecasting of species responses to environmental change. It thereby strengthens links between biogeography, population biology and theoretical and applied ecology.

Keywords

Biogeography, ecological forecasts, global change, hierarchical Bayesian statistics, long-distance dispersal, niche theory, process-based model, range shifts, spatial demography, species distribution modelling.

*Correspondence: Jörn Pagel, Plant Ecology and Nature Conservation, Institute of Biochemistry and Biology, University of Potsdam, Maulbeerallee 2, 14469 Potsdam, Germany. E-mail: jpagel@uni-potsdam.de

INTRODUCTION

The estimation of geographical species ranges via ecological niche modelling is a very active field of research. In particular, niche modelling is used to forecast the effects of environmental change on biodiversity (Fischlin *et al.*, 2007) and to guide conservation planning (Rodríguez *et al.*, 2007). The majority of studies that predict the impacts of environmental change on species distributions use statistical species distribution models (SDMs). SDMs relate current species distributions to selected

environmental variables (Guisan & Zimmermann, 2000) and their widespread use has been fostered by easy to use statistical techniques and the growing availability of environmental and biogeographical data (Guisan & Thuiller, 2005). However, the usefulness of SDMs for predicting range shifts and extinction risks and for guiding policy making and planning has been questioned (Pearson & Dawson, 2003; Araújo & Guisan, 2006).

As purely phenomenological models, SDMs do not represent the dynamic processes that underlie the formation of species ranges. Their application thus assumes that species distributions

are at equilibrium with environmental conditions (Guisan & Thuiller, 2005). However, many species distributions currently deviate from equilibrium (Svenning & Skov, 2004; Araújo & Pearson, 2005; Schurr *et al.*, 2007), and limited ability to spread will further increase these deviations under future climate change (Loarie *et al.*, 2009). Moreover, SDMs ignore spatial population dynamics which cause mismatches between species niches and distributions (Holt, 2009). To overcome these shortcomings, various authors suggested that SDMs should incorporate dynamic processes (Guisan & Thuiller, 2005; Araújo & Guisan, 2006; Thuiller *et al.*, 2008) and should better integrate ecological theory (Guisan & Thuiller, 2005; Araújo & Guisan, 2006; Kearney, 2006).

Species ranges ultimately arise from the interplay of birth, death and dispersal in space (Holt *et al.*, 2005). Hence, species ranges do not simply reflect niche limitations but are also shaped by demographic stochasticity, environmental variability, dispersal limitations and source–sink dynamics (Holt *et al.*, 2005). Nonetheless, recent theoretical contributions emphasize the potential of Hutchinson's niche concept for understanding species ranges (Soberón, 2007; Holt, 2009). Hutchinson (1978) described a species' niche by the response of demographic rates to the environment and defined it as the subset of environmental states where the intrinsic population growth rate is positive. Models that combine these demographic response functions with population dynamics and dispersal have been successfully used in theoretical studies of the link between niches and species ranges (Pulliam, 2000; Holt *et al.*, 2005).

These theoretical model concepts have not yet been used to forecast range dynamics. Moreover, dynamic process-based models of species ranges are generally rare (Thuiller *et al.*, 2008), mainly because they are difficult to parameterize (Morin & Thuiller, 2009). Even for well-studied species, parameters are usually derived from different sources in a piecemeal approach that inadequately accounts for parameter uncertainties (Thomas *et al.*, 2005). Further sources of uncertainty in modeling population dynamics include process variation (demographic and environmental stochasticity), observation errors and model uncertainty (Calder *et al.*, 2003). A full assessment and clear communication of uncertainty is crucial for ecological forecasts (Clark & Gelfand, 2006), especially when forecasts are used for climate impact assessment or conservation planning (Schröter *et al.*, 2005).

Given the importance of assessing uncertainties and the challenges of parameterizing process-based models, the value of dynamic niche modelling approaches for forecasting future ranges depends on the ability to link these models to available biogeographical data in a rigorous statistical framework. Hierarchical Bayesian statistics provides a general framework for consistently addressing uncertainty in complex systems (Clark, 2005; Clark & Gelfand, 2006). Efficient computational methods have already been used to estimate state-space models of population dynamics (e.g. Calder *et al.*, 2003; Buckland *et al.*, 2004; Thomas *et al.*, 2005) and also to study spatiotemporal dynamics (e.g. Hooten *et al.*, 2007). In the context of niche modelling, hierarchical Bayesian methods were previously used to extend

static SDMs, for example by the consideration of spatially correlated random effects and irregular sampling intensity (Latimer *et al.*, 2006) or environmental variation at multiple spatial scales (Diez & Pulliam, 2007).

Here we present a novel approach for understanding and predicting species niches and distributions by the statistical estimation of process-based models that describe spatial population dynamics with demographic rates varying in response to environmental conditions. We use a hierarchical Bayesian framework to estimate demographic response functions, local population dynamics and dispersal processes conditional upon each other while accounting for various sources of uncertainty. These dynamic range models (DRMs) thus aim to: (1) infer species niches and spatiotemporal population dynamics from distribution data, and (2) provide fully probabilistic forecasts of future range dynamics under environmental change. After formulating a general process-based model of range dynamics we apply a basic implementation to simulated data. To elucidate the potential and limitations of the novel framework we compare scenarios that differ not only in ecological characteristics of the study species but also in the quality and quantity of data used for model estimation. Based thereon, we furthermore discuss the practicability of the DRM approach for real-world applications and its implications for future monitoring.

A DYNAMIC RANGE MODEL

A dynamic concept of species ranges dissects the link between environmental conditions and species distribution data into a hierarchy of three processes: (1) environmental conditions affect the demographic rates of local populations; (2) spatiotemporal variation in local population dynamics and dispersal determine spatiotemporal abundance distributions; (3) these abundance distributions are sampled to obtain different types of distributional data (Fig. 1a). To build a predictive model one has to jointly estimate all parameters and the related uncertainties at each process level. In a Bayesian approach, this is expressed as the joint posterior distribution of parameters conditional on the data, $p(\text{parameters}|\text{data})$. The hierarchical approach estimates this complex joint distribution via a sequence of conditional distributions (Clark, 2005). For our DRM, factorization of the posterior yields three submodels (Fig. 1b) that correspond to the three process levels:

$$\begin{aligned}
 p(\text{parameters}|\text{data}) \propto & \overbrace{p(\text{data}|\text{population states, data parameters})}^{\text{Observer model}} \\
 & \times \overbrace{p(\text{population states}|\text{demographic rates, process parameters})}^{\text{Population dynamics model}} \\
 & \times \overbrace{p(\text{demographic rates}|\text{niche parameters})}^{\text{Niche model}} \times \overbrace{p(\text{parameters})}^{\text{Priors for parameters}}
 \end{aligned} \quad (1)$$

The *niche model* links demographic rates to environmental variables. Based on heterogeneous demographic rates and dispersal parameters, a grid-based *population dynamics model* then describes spatiotemporal distributions of population abundances. Abundances on this grid are not perfectly known but are

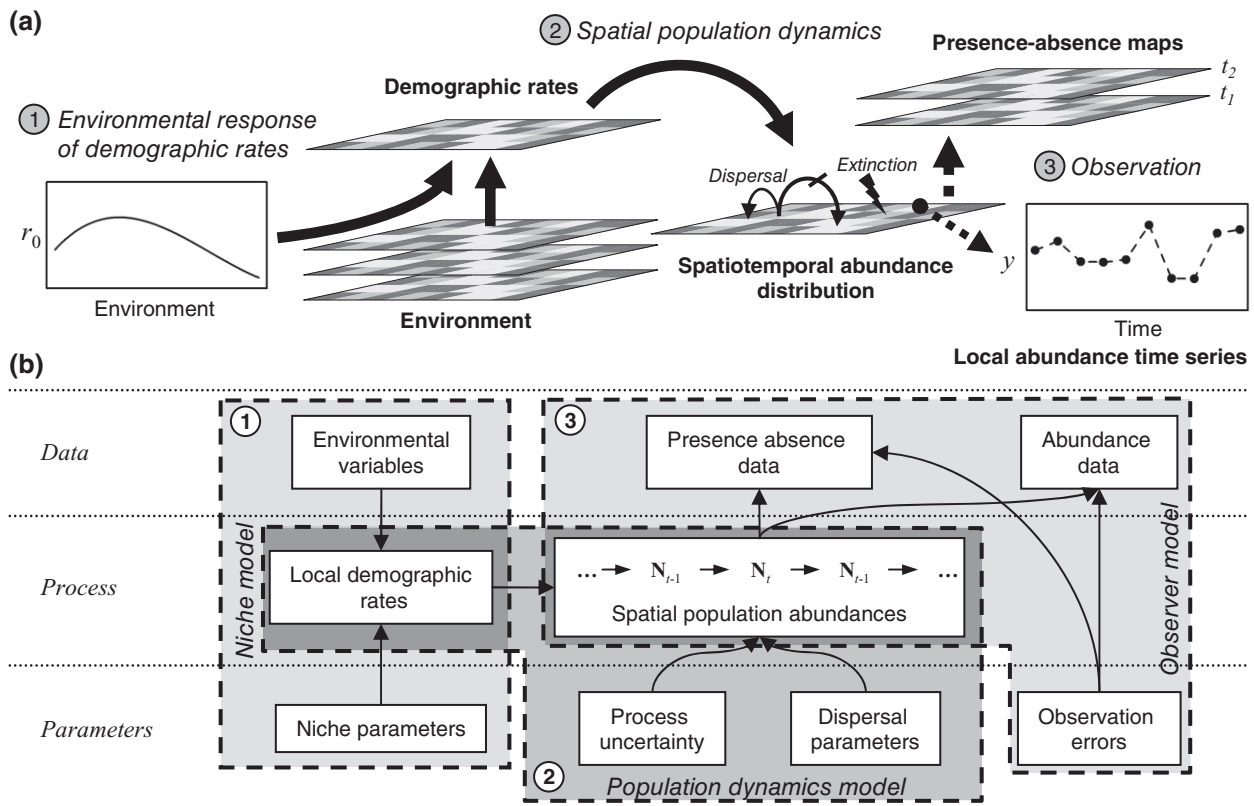


Figure 1 Concept of a predictive dynamic range model (DRM). (a) A hierarchy of three processes describes the link between environmental conditions and species distribution data: (1) environmental conditions affect demographic rates of local populations; (2) variation in local demography together with dispersal determines spatiotemporal abundance distributions; (3) these abundance distributions are sampled to obtain different types of data. (b) Graph of the hierarchical Bayesian model (following Clark, 2005), which accordingly describes conditional relationships between data, processes and parameters by three submodels. The hierarchical model structure thereby accounts for uncertainties arising from unexplained variance in the environmental response, demographic stochasticity in the population model as well as observation errors.

modelled as latent state variables. Thus an *observer model* links these actual abundances to different types of observed data such as presence–absence maps or time series of local abundance. In the following, we describe basic models for each level of the DRM framework and derive the respective conditional distributions of model parameters.

Niche model

A simple *niche model* describes the environmental response of the intrinsic population growth rate r . At the typical resolution of SDM applications (grid cells of 1–100 km²), environmental variables characterize the scenopoetic environment (Hutchinson, 1978), and resource-mediated interactions are hardly considered explicitly (Soberón, 2007). Consequently, inter-specific interactions are implicit in this niche description, which therefore characterizes Hutchinson's realized niche. The environmental response function of r is modelled as a linear combination of the effects of static landscape variables X and dynamic climatic variables C , so that $r = g(\beta, X) + h(\gamma, C) + \epsilon$,

where β and γ are parameter vectors and ϵ denotes iid normal errors $\epsilon \sim N(0, \sigma_r^2)$. These errors account for unexplained environmental effects and variability in growth rates (demographic heterogeneity). The likelihood of r across sites i and years t is thus

$$p(r|\beta, \gamma, \sigma_r) = N(r|\mu, \sigma_r^2) \text{ with } \mu_i^t = g(\beta, X_i) + h(\gamma, C_i^t). \quad (2)$$

Population dynamics model

We model spatiotemporal population dynamics by representing local population dynamics and extinction within grid cells as well as dispersal between cells. As a simple deterministic description of population dynamics within cells we use the Ricker model $N^{t+1} = N^t \exp(r - hN^t)$. At a time step of 1 year, the Ricker model describes the population dynamics of species with non-overlapping annual generations (such as many insects or annual plants). However, extension to other life histories is possible. With the parameterization in terms of intrinsic growth rate r and competition intensity h , the carrying capacity ($K = r/h$)

depends on the same factors that also influence intrinsic growth (Holt *et al.*, 1997). Local extinction can arise from variability in growth rates (see Niche model) and from demographic stochasticity represented by Poisson errors on local abundance.

Dispersal is described by fat-tailed dispersal kernels based on mixture distributions (Higgins & Richardson, 1999). Thus, we dissect dispersal into short-distance dispersal over scales smaller than the extent of grid cells and long-distance dispersal between grid cells. Long-distance dispersal of a fraction f_{LDD} of dispersal units is described by an exponential kernel $f(r) = 1/\alpha \exp(-r/\alpha)$ with mean dispersal distance α . To obtain dispersal probabilities $P_{j \rightarrow i}(f_{LDD}, \alpha)$ between spatially discrete cells, this dispersal kernel is integrated over both the cell of origin j and the target cell i .

Combining the stochastic Ricker model and the dispersal kernel, the likelihood of population abundances N can be formulated conditional on parameters of the Ricker model (r, h), the dispersal kernel (f_{LDD}, α) and on initial abundances N^0 :

$$p(N|N^0, r, h, f_{LDD}, \alpha, \sigma_p) = \prod_i \text{Poisson}(N_i^t | \lambda_i^t) \quad (3)$$

with expectation of the Poisson

$$\lambda_i^t = \tilde{N}_i^{t-1} \exp(r_i^{t-1} - h \tilde{N}_i^{t-1} + \varepsilon_i^t), \quad (4)$$

and post-dispersal abundances

$$\tilde{N}_i^t = \sum_j P_{j \rightarrow i}(f_{LDD}, \alpha) N_j^t. \quad (5)$$

Note that iid normal errors $\varepsilon \sim N(0, \sigma_p^2)$ account for misspecifications of the demographic model.

Observer model

The sampling of either abundance time series A or presence-absence records P is modelled as a binomial process with an independent detection probability π for each individual within a site (Dorazio *et al.*, 2006). Typically, per-individual detection probabilities for the sampling of abundance (π_A) will be higher than those for the sampling of presence-absence (π_P). For the latter, the likelihood of a presence record is given by the probability to observe at least one out of N individuals, which is $\psi = 1 - (1 - \pi_P)^N$. With Θ and Ω denoting the subset of sites and years for which abundance and presence-absence data exist, respectively, the likelihood of the data given the population states is:

$$p(A, P|N, \pi_A, \pi_P) = \prod_{\Theta} \text{Binomial}(A_i^t | N_i^t, \pi_A) \times \prod_{\Omega} \text{Bernoulli}(P_i^t | \psi_i^t). \quad (6)$$

Bayesian parameter estimation

Based on the submodels described above the joint posterior distribution of all model parameters is

$$p(\beta, \gamma, \sigma_r, h, f_{LDD}, \alpha, \sigma_p, N, \pi_A, \pi_P | A, P) \propto \underbrace{p(A, P|N, \pi_A, \pi_P)}_{\text{Observer model}} \times \underbrace{p(N|N^0, r, h, f_{LDD}, \alpha, \sigma_p)}_{\text{Population dynamics model}} \times \underbrace{p(r|\beta, \gamma, \sigma_r)}_{\text{Niche model}} \times \underbrace{p(h, \sigma_p)}_{\text{Priors for parameters}} p(f_{LDD}, \alpha) p(N^0) p(\pi_A, \pi_P) \quad (7)$$

Samples from this joint posterior distribution are generated by a Markov chain Monte Carlo (MCMC) algorithm. We use a Metropolis-within-Gibbs scheme, which jointly updates parameters using a combination of delayed rejection and adaptive Metropolis samplers (DRAM; Haario *et al.*, 2006), whereas elements of the high-dimensional state matrix N are updated in component-wise Metropolis-Hastings steps (for details see Appendix S1 in Supporting Information).

VIRTUAL CASE STUDY

To assess the potential of the DRM framework and its applicability to real species, it is necessary to evaluate its performance for species with different range dynamics and its dependence on the data available for model estimation. Such a comprehensive assessment is best achieved by a ‘virtual ecologist’ approach (Zurell *et al.*, 2010). Following this approach, we first simulated range dynamics under environmental change for different hypothetical species, and simulated a virtual ecologist who collected distribution data for these species. The sampled data were then used to estimate posteriors of DRM parameters. Subsequently, we assessed the DRM framework by comparing these parameter estimates and the resulting forecasts of future ranges with the known ‘truth’ and with predictions of a static SDM.

Simulation of range dynamics

Range dynamics were simulated on an artificial landscape of 20×40 grid cells. Environmental variation within this grid was described by two static ‘landscape’ variables (X_1 and X_2) and temperature as a single dynamic ‘climate’ variable. The two static variables are spatially autocorrelated (generated from two-dimensional fractal Brownian motion). Annual temperatures were calculated from a time series of global mean temperature that increases linearly after an initial spin-up period (Fig. 2a). These global temperatures were then regionalized by adding a linear latitudinal gradient and an auto-correlated spatial random effect.

On this dynamic landscape we simulated spatiotemporal population dynamics for four ecological scenarios: a K -strategist and an r -strategist under equilibrium and non-equilibrium conditions, respectively. The K -strategist has relatively low population densities, moderate intrinsic growth rates r , and limited dispersal ability. In contrast, the r -strategist has high population densities, high and variable r , and high rates of long-distance dispersal. For both life-history types, r increases linearly with both landscape variables and has a hump-shaped response to temperature. Global warming over 75 years causes the temperature optimum of r to shift from the lower third of the landscape

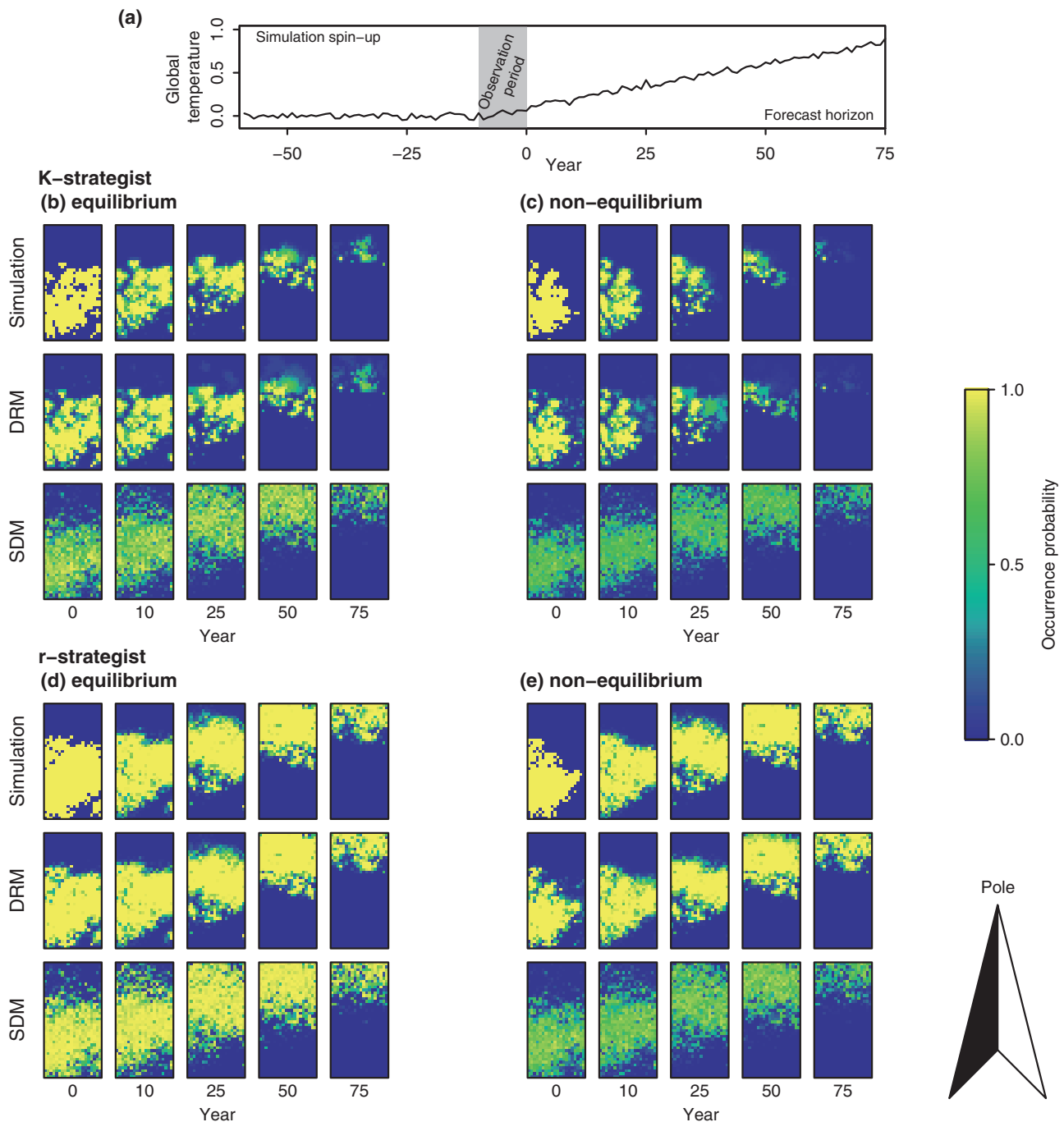


Figure 2 Simulation and forecasts of range dynamics. (a) Climate change scenario and scheduling of the simulation study. After a spin-up period of 50 years, virtual data are sampled in an observation period which spans the first 10 years of an increase of global mean temperature. Temperature rise then continues throughout the forecasting horizon of 75 years. (b)–(e) Maps of ‘true’ versus predicted ranges for four ecological scenarios. For each scenario, the top row shows the spatiotemporal dynamics of species occurrence probability for repeated simulations of the ‘true’ model. These ‘true’ dynamics are compared with forecasts of the dynamic range model (DRM) and the static species distribution model (SDM).

during the spin-up period to the upper boundary. Non-equilibrium conditions were created by restricting the species’ initial distribution and reducing the duration of the spin-up simulation so that the species does not fill its entire potential range at the time of data collection. The range dynamics of these

four scenarios are typical of common, widespread species (equilibrium *r*-strategist, Fig. 2b), exotic invasives (non-equilibrium *r*-strategist, Fig. 2c), rare, widespread species (equilibrium *K*-strategist, Fig. 2d) and rare species limited by post-glacial migration (non-equilibrium *K*-strategist, Fig. 2e). Parameter

values for the simulations of range dynamics are given in Table S1.

Data

For all four ecological scenarios, the virtual ecologist collected data during the first 10 years of temperature rise, that is from year -10 to year 0 (Fig. 2a). Our standard data scenario consists of presence–absence maps that cover the entire landscape at year -10 and year 0 , and of time series of annual abundance for a subset of 30 sites. To simulate observation errors, both presence–absence and abundance data were sampled from binomial distributions (see Observer model), with different per-individual detection probabilities representing differences in accuracy between presence–absence ($\pi_P = 0.1$) and abundance data ($\pi_A = 0.9$).

Prior information

Bayesian analyses can incorporate prior knowledge on certain model parameters. Here, we assumed that the virtual ecologist has some prior knowledge of dispersal ability and the accuracy of species distribution data. Hence we used moderately informative priors for the fraction of long-distance dispersal (f_{LDD}) and for the mean dispersal distance α (cf. Fig. 3 and Appendix S1). Such prior information might for instance stem from mechanistic dispersal models (Nathan *et al.*, 2008).

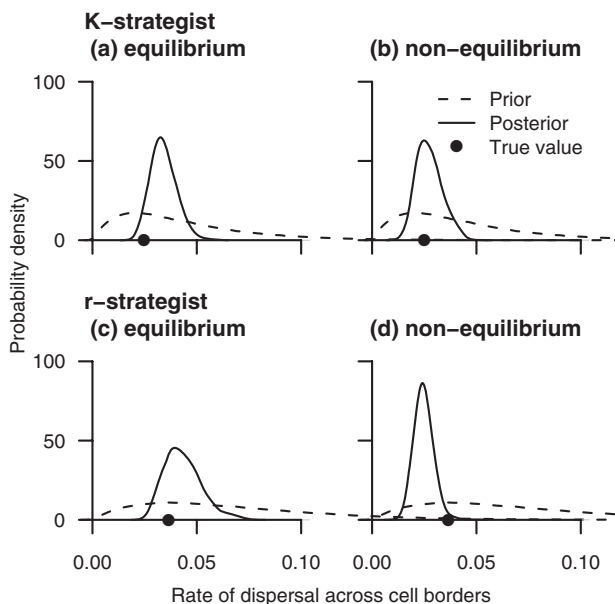


Figure 3 Estimates of long-distance dispersal rates for the four ecological scenarios. Mean dispersal distance α and fraction of long-distance dispersal f_{LDD} are aggregated to calculate the probability that a dispersing individual leaves the cell of origin. Dots indicate the 'true' value, solid lines give posterior estimates and dashed lines show the prior used for model estimation. Posterior distributions are displayed as Gaussian kernel density estimates (bandwidth = 0.002).

Prior information on detection probabilities was important for model estimation since our data scenarios do not entail replicate surveys of the same site in the same year. Such replicate surveys are considered to be a pre-requisite for the estimation of observation errors (Dorazio *et al.*, 2006). However, informative priors for detection probabilities can often be derived from studies on similar species (cf. Hooten *et al.*, 2007) or the uncertainties communicated with census data (cf. Clark & Bjørnstad, 2004). Thus, we assumed that the virtual ecologist has some information on data quality and formulated informative priors for per-individual observation probabilities π_A and π_P as beta distributions with expectation at the true value p and variance $p(1 - p)/100$. For all other parameters we specified weakly or non-informative prior distributions (see Appendix S1).

Model estimation

To estimate DRMs, we ran two independent chains of the MCMC sampler for 100,000 iteration steps, discarded the first 50,000 samples as 'burn-in' and thinned the remaining samples by storing every 50th iteration step. Convergence of the chains within the first 50,000 iterations was checked by calculating the scale reduction factor of Gelman & Rubin (1992). Although the general model posterior (equation 7) could be simplified for the virtual study (see Appendix S1) in order to facilitate computation, sampling was still computationally expensive and each iteration of a single MCMC chain took about 35 s (on a Xeon 3 GHz CPU).

For comparison, we additionally estimated static SDMs from the true occurrences in the last year of the sampling period (year 0). We chose a binomial generalized linear model with a logit link because its functional form corresponds to the demographic response function used in the simulation. If spatial population dynamics were absent and range filling was complete, this SDM would thus yield unbiased niche estimates. To convert predicted probabilities of presence into presence–absence maps, we calibrated the SDM by adjusting the threshold in order to minimize the difference between sensitivity and specificity (Jiménez-Valverde & Lobo, 2007). This calibration accounts for differences in prevalence (or range size) between the scenarios. Note that we did not intend an exhaustive comparison of DRMs and SDM methods. Instead, the SDM results serve as examples of what can typically be achieved with static models.

Estimation of species niches and dispersal abilities

DRMs estimate a species' niche as the set of environments for which mean intrinsic population growth rate r is positive. For SDMs, the niche is commonly defined as the set of environments for which presence probability is above the calibration threshold. For all ecological scenarios, the DRM accurately estimated niche limits, whereas the SDM consistently produced biased niche estimates (Fig. 4). This consistent bias arose because the SDM neglects source–sink dynamics (Pulliam, 2000) which

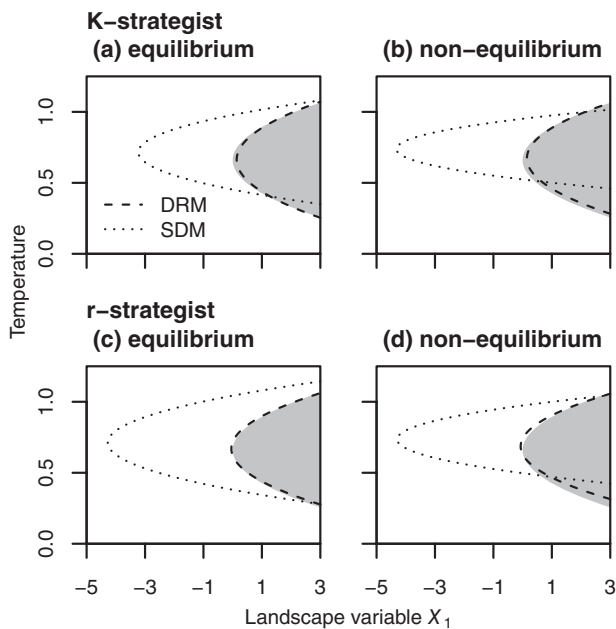


Figure 4 'True' versus estimated niches for the four ecological scenarios. The 'true' niche (grey areas) is the set of environments in which mean intrinsic growth rate is positive in the simulation. Lines show niche borders estimated by the dynamic range model (DRM, dashed) and the species distribution model (SDM, dotted). For the DRM, these borders are calculated from the posterior medians of niche parameters. The non-varied environmental variable is set to its mean across the landscape ($X_2 = 0$).

caused it to underestimate the effects of landscape relative to climate and thus to overestimate the width of the landscape niche in good climates (Fig. 4). For the non-equilibrium scenarios, the SDM's climatic niche estimates were additionally biased towards higher temperatures (Fig. 4b, d). This is because in the non-equilibrium scenarios the species fill less of its high-latitude potential range (Fig. 2). In contrast, the DRM niche estimates were not affected by these non-equilibrium conditions (Fig. 4b, d).

DRMs estimate not only the niche but also the rate of long-distance dispersal between grid cells. For three ecological scenarios, posterior estimates of long-distance dispersal rates were markedly improved compared with the prior distribution (Fig. 3a–c). However, for the *r*-strategist in non-equilibrium conditions the DRM clearly underestimated long-distance dispersal (Fig. 3d). In a supplementary analysis we found that this bias disappears when the DRM estimation uses 'perfect' presence–absence data that are observed without error. This suggests that the bias arises from uncertainty about the existence of small undetected outlier populations beyond the leading edge of the true range. If such outlier populations existed, the observed range expansion in the sampling period could be explained with lower rates of long-distance dispersal.

Forecasts of range dynamics

DRM forecasts of future ranges were generated by stochastic forward simulations of the niche model and the population dynamics model using samples from the joint posterior of parameters and population states N in the last year of the sampling period. For comparison, we also projected the SDM's niche estimates into the future. To quantify the predictive accuracy of DRM and SDM forecasts, we evaluated predicted future range sizes and Cohen's kappa coefficient. Cohen's kappa is a frequently used measure of the accuracy of presence–absence predictions that considers both commission and omission errors (Fielding & Bell, 1997). It ranges from 1 to 0, where 1 denotes perfect prediction and values below 0.4 indicate poor prediction (Monserud & Leemans, 1992).

Maps of predicted ranges (Fig. 2), predicted range sizes and Cohen's kappa (Fig. 5) show clear differences in the performance of DRMs and SDMs for the four ecological scenarios. The DRM correctly predicted the *K*-strategist's limited ability to track climate change (Fig. 2b, c) and the resulting range contractions (Fig. 5a, b). The greatest uncertainties in range predictions occurred at the leading edge for the equilibrium scenario (Fig. 2b) and in the initially unsaturated range for the non-equilibrium scenario (Fig. 2c). For the non-equilibrium scenario, these uncertainties caused some decrease in predictive accuracy although predictions of total range size were unbiased (Fig. 5b). In contrast, the SDM, which neglects migration limitation, largely overestimated future range sizes and had a lower predictive accuracy for both scenarios (Fig. 5a, b).

Future range dynamics of the *r*-strategist are less migration-limited than those of the *K*-strategist (Fig. 2d, e). Moreover, migration limitation is restricted to a transient phase before year 50 when the leading edge of the range reaches the upper boundary of the landscape. For the equilibrium *r*-strategist, the DRM correctly predicted the resulting range size reduction in this transient phase, whereas the SDM again overestimated these range sizes and had lower prediction accuracy (Fig. 5c). Yet, for the non-equilibrium *r*-strategist, the DRM's underestimation of long-distance dispersal (Fig. 3d) caused an underestimation of transient range sizes (Fig. 5d). Surprisingly, the SDM yielded almost unbiased range size forecasts for this scenario (Fig. 5d). However, this apparent absence of a bias is really the consequence of two opposing biases: the SDM's underestimation of climatic niche width (Fig. 4d) was largely compensated by its neglect of incomplete range filling in the first years (Fig. 2e). Hence, also for the non-equilibrium *r*-strategist Cohen's kappa shows that the DRM had consistently higher prediction accuracy than the SDM (Fig. 5d).

Effects of data type and quality

To investigate what kinds of data are most useful for estimating DRMs, we varied scenarios of data availability for the equilibrium *K*-strategist. To assess the effect of data type (presence–absence maps versus abundance time series), we modified the reference data scenario (which comprises abundance time series

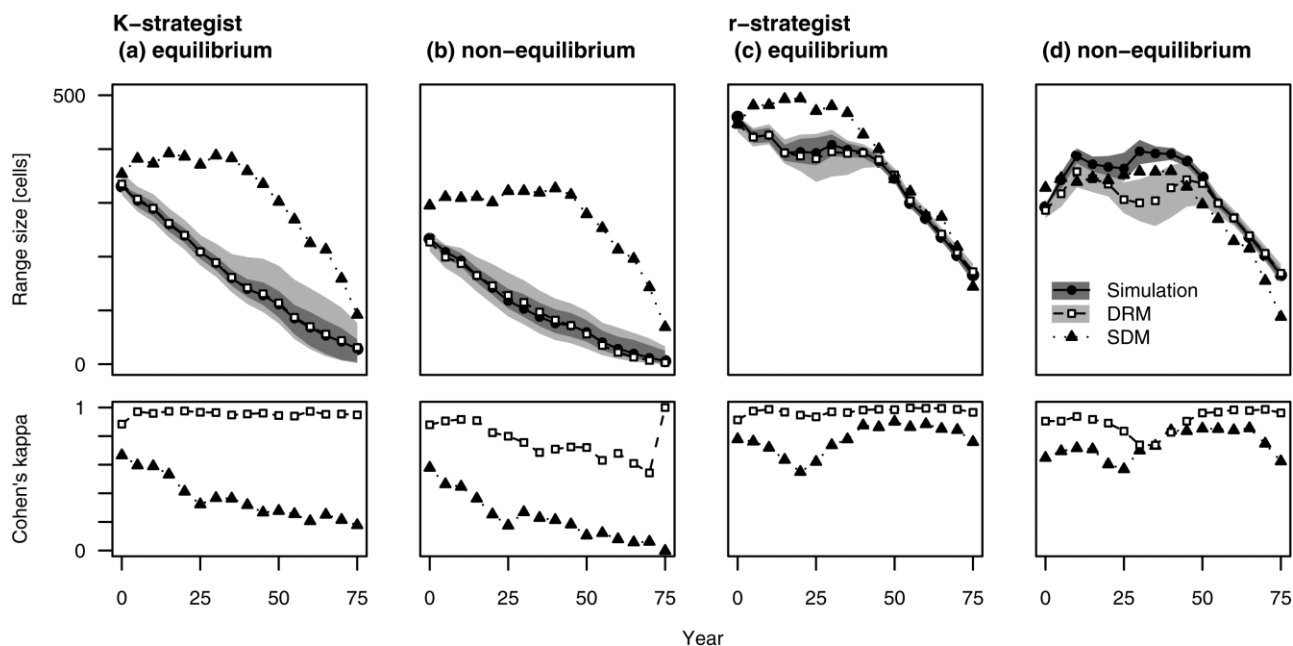


Figure 5 Top panel: forecasts of future range size from the dynamic range model (DRM) compared with the simulated 'truth' and the projections of the static species distribution model (SDM). Shaded areas depict the 95% central interval of range sizes from the stochastic simulation (dark grey) and from the DRM forecasts (light grey). Bottom panel: prediction accuracy of the DRM and the SDM measured by Cohen's kappa.

and two presence-absence maps, denoted as '2PA+TS', by either omitting the presence-absence map at the end of the sampling period (scenario '1PA+TS'), omitting both presence-absence maps and using only abundance time series (scenario 'TS') or omitting the abundance data and using only the two presence-absence maps (scenario '2PA'). For each of the resulting four data type scenarios, we additionally assessed the effect of data quality by comparing scenarios with high detection probabilities ($\pi_A = 0.9$, $\pi_P = 0.1$, as above) and lower detection probabilities ($\pi_A = 0.5$, $\pi_P = 0.01$).

For all but one combination of data type and quality, the 95% credibility interval of DRM range size forecasts comprised the simulated truth (Fig. 6, the only exception being the '2PA' scenario with low data quality, see below). For the scenarios of *high* data quality, the omission of one or two presence-absence maps (scenarios '1PA+TS' and 'TS', respectively) sequentially increased both uncertainties and biases in forecast range sizes and decreased predictive accuracy (Fig. 6). On the contrary, the scenario without abundance time series ('2PA') did not perform substantially worse than the reference. We suspect that this somewhat surprising result is rather particular to our virtual study, since the spatiotemporal pattern of occurrence provides information about spatial population dynamics only if the data are precise *and* the model captures the real dynamics well.

For data of low quality the accuracy of range forecasts was reduced as expected (Fig. 6). Yet this reduction was particularly strong if the data only comprised two presence-absence maps ('2PA'). Additionally, the bias in range size forecasts also increased strongly for the scenario with abundance time series

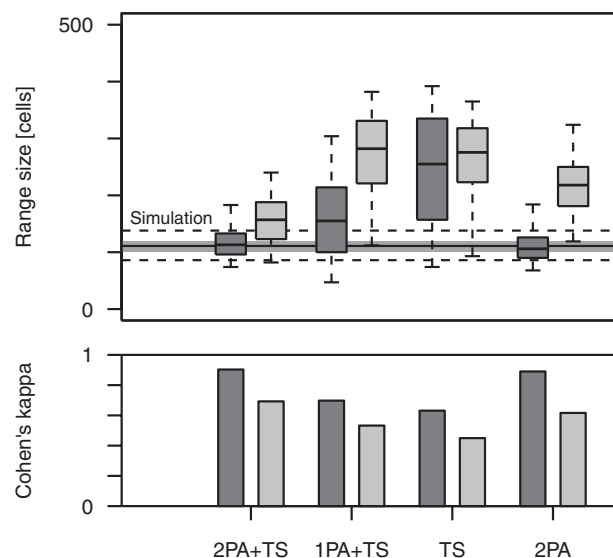


Figure 6 Effects of data type and quality on dynamic range model (DRM) forecasts. Data scenarios represent different combinations of presence-absence maps (PA) and abundance time series (TS). For each data scenario, we estimated DRMs with data sets subject to small and large observation errors (dark and light grey, respectively). Top: boxplots of forecast range size in year 50 after last data collection. Boxes comprise the inter-quartile range and whiskers span the 95% credibility interval of forecast range sizes. For the stochastic simulation of the 'true' model the figure accordingly shows the median (solid horizontal line), the inter-quartile range (grey bar) and the 95% central interval (dashed lines) of range sizes in year 50. Bottom: prediction accuracy of the DRM range forecasts for year 50 measured by Cohen's kappa.

and only one presence–absence map ('1PA+TS'). In summary, reliable range forecasts for low data quality thus required abundance time series and two presence–absence maps.

DISCUSSION

In the previous sections we have developed and tested a framework for statistically estimating dynamic models of species ranges (DRMs). These process-based DRMs describe how ecological niches and spatial population dynamics interact to determine range dynamics, and the hierarchical Bayesian framework enables one to statistically fit these process-based models to data on species occurrence and abundance (Fig. 1). In the following we discuss implications of the new framework for forecasts of range shifts under environmental change, for the monitoring of species responses to environmental change and for the quantification of ecological niches and spatial population dynamics.

Process-based forecasts of range dynamics

Our simulation study showed that DRMs can be parameterized from a combination of presence–absence maps and population censuses for a limited number of sites. This yields accurate niche estimates (Fig. 4) and predictions of future range dynamics (Fig. 5) for virtual study species that differ strongly in demographic properties and the degree to which their distribution is in equilibrium with the environment (Fig. 2). The crucial importance of a dynamic view of species ranges is especially apparent when migration limitation causes disequilibrium between species distributions and the environment (Figs 2 & 5). Yet, SDMs produce biased niche estimates even under equilibrium conditions (Fig. 4). Such biases in SDM niche estimates also affect recently developed 'hybrid models' of range dynamics that combine habitat distributions (estimated by phenomenological SDMs) with dynamic population models (Keith *et al.*, 2008; Thuiller *et al.*, 2008). This is because these hybrid approaches confound the relationship between niche and distribution by simulating range dynamics on top of a habitat model that is inferred from a static view on species distributions. Moreover, the demographic parameters of hybrid models are typically derived from various sources in a piecemeal approach (but see Cabral & Schurr, 2010), which prevents a comprehensive treatment of uncertainty (Clark & Gelfand, 2006). Hence, the application of hybrid models for the prediction of range shifts remains problematic (Thuiller *et al.*, 2008), and conservation managers and decision makers should treat their forecasts with caution. While we demonstrated the general possibility of linking process-based range models to data in a rigorous framework, we admit that computational efforts are much higher than for the estimation of hybrid models and SDMs. Appreciating this pragmatic point of view, we propose to compare these different approaches in controlled virtual systems as well as with real data in order to assess when the discrepancies are most severe.

Data requirements and monitoring of species distributions

The virtual simulation study provides a basis for identifying which data are needed to estimate DRMs for real species. The comparison of niche estimates and forecasts of range dynamics for exemplary data scenarios (Fig. 6) indicates that the combination of repeated presence–absence maps with abundance time series for some sites may not only allow accurate predictions but is also most robust to a decrease in data quality. On regional to continental scales, long-term data sets that combine presence records and abundance estimates currently exist for a limited set of taxa such as British butterflies (Pollard & Yates, 1993) and North American birds (Pardieck & Sauer, 2007). Still, we are optimistic that more data suitable for estimating DRMs will become available in the future. Recent initiatives like the Global Biodiversity Observation Network of the Group on Earth Observation (GEO BON) also call for a 'hierarchical sampling approach' that combines large amounts of relatively simple data like presence records with some more detailed surveys such as abundance monitoring (Scholes *et al.*, 2008). Our findings underline the need for such initiatives to improve forecasts of range shifts for a broader range of species. However, the identification of optimal sampling schemes for the estimation of DRMs requires further studies that will also have to consider financial constraints (Field *et al.*, 2005).

Reliable information about past species distributions can be particularly important for DRM estimation in non-equilibrium situations. This became apparent in our study of the non-equilibrium *r*-strategist, where uncertainty about the species' absence beyond the leading edge of the true range negatively biased estimates of long-distance dispersal rates (Fig. 3b) and consequently forecasts of future range size (Fig. 5d). Such sensitivity to accurate knowledge about initial conditions is not a flaw of our method but is well known from studies of non-equilibrium (population) dynamics, for example from the estimation of spatial metapopulation models (Moilanen, 2000). More specifically the role of outlier populations versus long-distance dispersal is also discussed for post-glacial tree migrations. In this field, understanding has recently been enhanced by phylogeographic analyses (Petit *et al.*, 2008). In the context of DRM estimation, the future integration of phylogeographic data could thus be another way to reduce forecast uncertainty that results from inaccurate knowledge about past distributions – besides a better a priori quantification of long-distance dispersal (Nathan *et al.*, 2008).

Limitations and possible extensions

In this study we considered fairly simple DRMs to study the role of spatial population dynamics for estimates of range dynamics and ecological niches. These simple DRMs subsume many of the processes that shape range dynamics in stochastic variance terms rather than describing these processes explicitly. However, we believe that the process-based framework presented here is a useful starting point for developing more realistic DRMs. In fact

the decision of where to let 'stochasticity stand in for complexity' and where to resolve this complexity more explicitly is a central issue in the construction of complex statistical models (Clark & Gelfand, 2006). The basic process-based DRM concept (Fig. 1) offers entry points for more complexity at different levels. For example, the variance term of our simple niche model could partly be replaced by submodels that describe how realized niches are altered by biotic interactions and niche evolution (Holt, 2009). Spatial random effects for growth rates could account for the effect of unobserved environmental variables, disentangling their effects from the autocorrelating effects of spatial population dynamics (cf. Latimer *et al.*, 2006; Dormann *et al.*, 2007). Population models other than the basic Ricker model permit the application to species with different life cycles and the consideration of Allee effects that might be important for range dynamics (e.g. Holt, 2009; Cabral & Schurr, 2010). A further extension would be to break down population growth rate into its component processes such as fecundity and survival at different life stages (Clark, 2003). This would not just enable the specification of more realistic (e.g. nonlinear) environmental response functions, but would also facilitate the use of prior information from a mechanistic understanding of species responses to environmental variation (e.g. from biophysical models; Kearney & Porter, 2009) or from studies that parameterize environmental response curves based on demographic field measurements (e.g. Dullinger *et al.*, 2004). Finally, existing approaches could be adapted to consider environmental effects on detection probabilities (e.g. Kéry *et al.*, 2009) or environmental and evolutionary effects on dispersal (e.g. Ovaskainen *et al.*, 2008; Phillips *et al.*, 2008; Kuparinen *et al.*, 2009).

Clearly, the introduction of additional processes and parameters increases computational challenges to the fitting of DRMs. However, hardware capacities are constantly advancing, and efficient Bayesian computation is an active research field (Clark & Gelfand, 2006). Keeping track of these developments and facing technical challenges is highly worthwhile in order to integrate more ecological processes into predictive models of range dynamics.

Efforts at increasing the realism of DRMs have to go hand in hand with the identification of the data needed to parameterize these models. A valuable source of information could be knowledge about species responses to their environment which is obtained from small-scale experiments and observations. Yet, since such small-scale environmental responses are not necessarily transferable to the spatial and temporal scales of DRMs, we need to better understand the scaling of demographic response functions (cf. Morin & Lechowicz, 2008). Here we see great potential for the combination of DRM analyses – which estimate the environmental response of demographic quantities at large scales – with small-scale measurements of these demographic quantities in different environments. This combination can serve to test whether there are general rules for the scaling of demographic responses. By facilitating the use of small-scale population biological knowledge as prior information for DRMs, the identification of such rules would in turn benefit range forecasts.

CONCLUSIONS

The potential of the presented approach for predictive models of range dynamics stems from the statistical framework that dissects species–environment relationships into spatiotemporal population dynamics and environmental response functions of demographic rates. This has the twofold benefit of: (1) inferring species niches from distribution data while accounting for spatiotemporal population dynamics, and (2) producing fully probabilistic forecasts of future range dynamics under environmental change. We are optimistic that such predictive process-based range models will improve forecasts of the impacts of environmental change on biodiversity and will thereby strengthen the scientific basis of conservation planning. By estimating niches and range dynamics from the combination of species distribution data with models founded in ecological theory, the presented framework furthermore strengthens links between biogeography and population biology, theoretical and applied ecology.

ACKNOWLEDGEMENTS

For helpful discussion we thank Wolfgang Cramer as well as Bob O'Hara, Otso Ovaskainen, Barbara Anderson and other members of the UKPopNet (NERC R8-H12-01 and English Nature) working group 'Bayesian distribution models: dynamics, processes and projections'. We are grateful for comments of two anonymous referees that helped to improve the manuscript. J.P. acknowledges financial support from the University of Potsdam Graduate Initiative on Ecological Modelling (UPGradE), the German Federal Agency for Nature Conservation and the European Commission 7th Framework Programme for the Environment (projects ALARM, GOCE-CT-2003-506675, and CarboExtreme, 226701). F.S. furthermore acknowledges support from the European Union through Marie Curie Transfer of Knowledge Project FEMMES (MTKD-CT-2006-042261).

REFERENCES

- Araújo, M.B. & Guisan, A. (2006) Five (or so) challenges for species distribution modelling. *Journal of Biogeography*, **33**, 1677–1688.
- Araújo, M.B. & Pearson, R.G. (2005) Equilibrium of species' distributions with climate. *Ecography*, **28**, 693–695.
- Buckland, S.T., Newman, K.B., Thomas, L. & Koesters, N.B. (2004) State-space models for the dynamics of wild animal populations. *Ecological Modelling*, **171**, 157–175.
- Cabral, J.S. & Schurr, F.M. (2010) Estimating demographic models for the range dynamics of plant species. *Global Ecology and Biogeography*, **19**, 85–97.
- Calder, C., Lavine, M., Müller, P. & Clark, J.S. (2003) Incorporating multiple sources of stochasticity into dynamic population models. *Ecology*, **84**, 1395–1402.
- Clark, J.S. (2003) Uncertainty and variability in demography and population growth: a hierarchical approach. *Ecology*, **84**, 1370–1381.

- Clark, J.S. (2005) Why environmental scientists are becoming Bayesians. *Ecology Letters*, **8**, 2–14.
- Clark, J.S. & Bjørnstad, O.N. (2004) Population time series: process variability, observation errors, missing values, lags, and hidden states. *Ecology*, **85**, 3140–3150.
- Clark, J.S. & Gelfand, A.E. (2006) A future for models and data in environmental science. *Trends in Ecology and Evolution*, **21**, 375–380.
- Diez, J.M. & Pulliam, H.R. (2007) Hierarchical analysis of species distributions and abundance across environmental gradients. *Ecology*, **88**, 3144–3152.
- Dorazio, R.M., Royle, J.A., Söderström, B. & Glimskär, A. (2006) Estimating species richness and accumulation by modeling species occurrence and detectability. *Ecology*, **87**, 842–854.
- Dormann, C.F., McPherson, J.M., Araújo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissling, W.D., Kühn, I., Ohlemüller, R., Peres-Neto, P.R., Reineking, B., Schröder, B., Schurr, F.M. & Wilson, R. (2007) Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, **30**, 609–628.
- Dullinger, S., Dirnböck, T. & Grabherr, G. (2004) Modelling climate change-driven treeline shifts: relative effects of temperature increase, dispersal and invasibility. *Journal of Ecology*, **92**, 241–252.
- Field, S.A., Tyre, A.J. & Possingham, H.P. (2005) Optimizing allocation of monitoring effort under economic and observational constraints. *Journal of Wildlife Management*, **69**, 473–482.
- Fielding, A.H. & Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24**, 38–49.
- Fischlin, A., Midgley, G.F., Price, J.T., Leemans, R., Gopal, B., Turley, C., Rounsevell, M.D.A., Dube, O.P., Tarazona, J. & Velichko, A.A. (2007) Ecosystems, their properties, goods, and services. *Climate change 2007: impacts, adaptation and vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (ed. by M.L. Parry, O.F. Canziani, J.P. Palutikof, P.J. van der Linden and C.E. Hanson), pp. 211–272. Cambridge University Press, Cambridge, UK.
- Gelman, A. & Rubin, D.B. (1992) Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457–472.
- Guisan, A. & Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.
- Guisan, A. & Zimmermann, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.
- Haario, H., Laine, M., Mira, A. & Saksman, E. (2006) DRAM: efficient adaptive MCMC. *Statistics and Computing*, **16**, 339–354.
- Higgins, S.I. & Richardson, D.M. (1999) Predicting plant migration rates in a changing world: the role of long-distance dispersal. *The American Naturalist*, **153**, 464–475.
- Holt, R.D. (2009) Bringing the Hutchinsonian niche into the 21st century: ecological and evolutionary perspectives. *Proceedings of the National Academy of Sciences USA*, **106**, 19659–19665.
- Holt, R.D., Lawton, J.H., Gaston, K.J. & Blackburn, T.M. (1997) On the relationship between range size and local abundance: back to basics. *Oikos*, **78**, 183–190.
- Holt, R.D., Keitt, T.H., Lewis, M.A., Maurer, B.A. & Taper, M.L. (2005) Theoretical models of species' borders: single species approaches. *Oikos*, **108**, 18–27.
- Hooten, M.B., Wikle, C.K., Dorazio, R.M. & Royle, J.A. (2007) Hierarchical spatiotemporal matrix models for characterizing invasions. *Biometrics*, **63**, 558–567.
- Hutchinson, G.E. (1978) *An introduction to population ecology*. Yale University Press, New Haven, CT.
- Jiménez-Valverde, A. & Lobo, J.M. (2007) Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecologica – International Journal of Ecology*, **31**, 361–369.
- Kearney, M. (2006) Habitat, environment and niche: what are we modelling? *Oikos*, **115**, 186–191.
- Kearney, M. & Porter, W. (2009) Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters*, **12**, 334–350.
- Keith, D.A., Akçakaya, H.R., Thuiller, W., Midgley, G.F., Pearson, R.G., Phillips, S.J., Regan, H.M., Araújo, M.B. & Rebelo, T.G. (2008) Predicting extinction risks under climate change: coupling stochastic population models with dynamic bioclimatic habitat models. *Biology Letters*, **4**, 560–563.
- Kéry, M., Dorazio, R.M., Soldaat, L., van Strien, A., Zuiderwijk, A. & Royle, J.A. (2009) Trend estimation in populations with imperfect detection. *Journal of Applied Ecology*, **46**, 1163–1172.
- Kuparinen, A., Katul, G., Nathan, R. & Schurr, F.M. (2009) Increases in air temperature can promote wind-driven dispersal and spread of plants. *Proceedings of the Royal Society B: Biological Sciences*, **276**, 3081–3087.
- Latimer, A.M., Wu, S.S., Gelfand, A.E. & Silander, J.A. (2006) Building statistical models to analyze species distributions. *Ecological Applications*, **16**, 33–50.
- Loarie, S.R., Duffy, P.B., Hamilton, H., Asner, G.P., Field, C.B. & Ackerly, D.D. (2009) The velocity of climate change. *Nature*, **462**, 1052–1055.
- Moilanen, A. (2000) The equilibrium assumption in estimating the parameters of metapopulation models. *Journal of Animal Ecology*, **69**, 143–153.
- Monserud, R.A. & Leemans, R. (1992) Comparing global vegetation maps with the kappa-statistic. *Ecological Modelling*, **62**, 275–293.
- Morin, X. & Lechowicz, M.J. (2008) Contemporary perspectives on the niche that can improve models of species range shifts under climate change. *Biology Letters*, **4**, 573–576.
- Morin, X. & Thuiller, W. (2009) Comparing niche- and process-based models to reduce prediction uncertainty in species range shifts under climate change. *Ecology*, **90**, 1301–1313.
- Nathan, R., Schurr, F.M., Spiegel, O., Steinitz, O., Trakhtenbrot, A. & Tsoar, A. (2008) Mechanisms of long-distance seed dispersal. *Trends in Ecology and Evolution*, **23**, 638–647.

- Ovaskainen, O., Rekola, H., Meyke, E. & Arjas, E. (2008) Bayesian methods for analyzing movements in heterogeneous landscapes from mark-recapture data. *Ecology*, **89**, 542–554.
- Pardieck, K.L. & Sauer, J.R. (2007) The 1999–2003 summary of the North American Breeding Bird Survey. *Bird Populations*, **8**, 28–45.
- Pearson, R.G. & Dawson, T.P. (2003) Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*, **12**, 361–371.
- Petit, R.J., Hu, F.S. & Dick, C.W. (2008) Forests of the past: a window to future changes. *Science*, **320**, 1450–1452.
- Phillips, B.L., Brown, G.P., Travis, J.M.J. & Shine, R. (2008) Reid's paradox revisited: the evolution of dispersal in range-shifting populations. *The American Naturalist*, **172**, S34–S48.
- Pollard, E. & Yates, T.J. (1993) *Monitoring butterflies for ecology and conservation*. Chapman and Hall, London.
- Pulliam, H.R. (2000) On the relationship between niche and distribution. *Ecology Letters*, **3**, 349–361.
- Rodríguez, J.P., Brotons, L., Bustamante, J. & Seoane, J. (2007) The application of predictive modelling of species distribution to biodiversity conservation. *Diversity and Distributions*, **13**, 243–251.
- Scholes, R.J., Mace, G.M., Turner, W., Geller, G.N., Jürgens, N., Larigauderie, A., Muchoney, D., Walther, B.A. & Mooney, H.A. (2008) Toward a global biodiversity observing system. *Science*, **321**, 1044–1045.
- Schröter, D., Cramer, W., Leemans, R. *et al.* (2005) Ecosystem service supply and vulnerability to global change in Europe. *Science*, **310**, 1333–1337.
- Schurr, F.M., Midgley, G.F., Rebelo, A.G., Reeves, G., Poschlod, P. & Higgins, S.I. (2007) Colonization and persistence ability explain the extent to which plant species fill their potential range. *Global Ecology and Biogeography*, **16**, 449–459.
- Soberón, J. (2007) Grinnellian and Eltonian niches and geographic distributions of species. *Ecology Letters*, **10**, 1115–1123.
- Svenning, J.C. & Skov, F. (2004) Limited filling of the potential range in European tree species. *Ecology Letters*, **7**, 565–573.
- Thomas, L., Buckland, S.T., Newman, K.B. & Harwood, J. (2005) A unified framework for modelling wildlife population dynamics. *Australian and New Zealand Journal of Statistics*, **47**, 19–34.
- Thuiller, W., Albert, C., Araújo, M.B., Berry, P.M., Cabeza, M., Guisan, A., Hickler, T., Midgley, G.F., Paterson, J., Schurr, F.M., Sykes, M.T. & Zimmermann, N.E. (2008) Predicting global change impacts on plant species' distributions: future challenges. *Perspectives in Plant Ecology, Evolution and Systematics*, **9**, 137–152.
- Zurell, D., Berger, U., Cabral, J.S., Jeltsch, F., Meynard, C.N., Münkemüller, T., Nehrbass, N., Pagel, J., Reineking, B., Schröder, B. & Grimm, V. (2010) The virtual ecologist approach: simulating data and observers. *Oikos*, **119**, 622–635.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Table S1 Parameter values of the ecological scenarios.

Appendix S1 Details on the use of MCMC methods for DRM parameter estimation.

As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be reorganized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

BIOSKETCHES

Jörn Pagel is interested in the modelling of spatial population dynamics and in computational statistical methods to link demographic models to data.

Frank Schurr is interested in developing a mechanistic understanding of how range dynamics arise from demographic processes.

Editor: Antoine Guisan