

Benchmarking novel approaches for modelling species range dynamics

DAMARIS ZURELL¹, WILFRIED THUILLER^{2,3}, JÖRN PAGEL⁴, JULIANO S. CABRAL^{5,6}, TAMARA MÜNKEMÜLLER^{2,3}, DOMINIQUE GRAVEL⁷, STEFAN DULLINGER⁸, SIGNE NORMAND⁹, KATJA H. SCHIFFERS^{2,3,10}, KARA A. MOORE¹¹ and NIKLAUS E. ZIMMERMANN^{1,12}

¹Dynamic Macroecology, Swiss Federal Research Institute WSL, Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland, ²Univ. Grenoble Alpes, Laboratoire d'Écologie Alpine (LECA), UMR-CNRS 5553 Université J. Fourier BP 53, F-38000 Grenoble, France, ³CNRS, Laboratoire d'Écologie Alpine (LECA), UMR-CNRS 5553 Université J. Fourier BP 53, F-38000 Grenoble, France, ⁴Institute of Landscape and Plant Ecology, University of Hohenheim, August-v.Hartmann-Str. 3, D-70599 Stuttgart, Germany, ⁵Biodiversity, Macroecology and Conservation Biogeography, University Göttingen, Büsgenweg 2, D-37077 Goettingen, Germany, ⁶Synthesis Centre of the German Centre for Integrative Biodiversity Research (iDiv), Deutscher Platz 5e, D-04103 Leipzig, Germany, ⁷Université de Québec à Rimouski, 300 Allée des Ursulines, Rimouski, Canada G5L 3A1, ⁸Department of Botany and Biodiversity Research, University of Vienna, Rennweg 14, 1030 Vienna, Austria, ⁹Section for Ecoinformatics and Biodiversity, Department of Bioscience, Aarhus University, Ny Munkegade 114, DK-8000 Aarhus C, Denmark, ¹⁰Senckenberg Biodiversity and Climate Research Centre (BiK-F), Georg Voigt-Straße 14-16, D-60325 Frankfurt (Main), Germany, ¹¹Center for Population Biology, University of California, Davis, One Shields Avenue, Davis, CA 95616, USA, ¹²Department of Environmental Systems Science, Swiss Federal Institute of Technology ETH, CH-8092 Zurich, Switzerland

Abstract

Increasing biodiversity loss due to climate change is one of the most vital challenges of the 21st century. To anticipate and mitigate biodiversity loss, models are needed that reliably project species' range dynamics and extinction risks. Recently, several new approaches to model range dynamics have been developed to supplement correlative species distribution models (SDMs), but applications clearly lag behind model development. Indeed, no comparative analysis has been performed to evaluate their performance. Here, we build on process-based, simulated data for benchmarking five range (dynamic) models of varying complexity including classical SDMs, SDMs coupled with simple dispersal or more complex population dynamic models (SDM hybrids), and a hierarchical Bayesian process-based dynamic range model (DRM). We specifically test the effects of demographic and community processes on model predictive performance. Under current climate, DRMs performed best, although only marginally. Under climate change, predictive performance varied considerably, with no clear winners. Yet, all range dynamic models improved predictions under climate change substantially compared to purely correlative SDMs, and the population dynamic models also predicted reasonable extinction risks for most scenarios. When benchmarking data were simulated with more complex demographic and community processes, simple SDM hybrids including only dispersal often proved most reliable. Finally, we found that structural decisions during model building can have great impact on model accuracy, but prior system knowledge on important processes can reduce these uncertainties considerably. Our results reassure the clear merit in using dynamic approaches for modelling species' response to climate change but also emphasize several needs for further model and data improvement. We propose and discuss perspectives for improving range projections through combination of multiple models and for making these approaches operational for large numbers of species.

Keywords: climate change, demographic models, dispersal, population viability, prediction, simulated data, species distribution models, virtual ecologist approach

Received 13 November 2015; revised version received 28 January 2016 and accepted 4 February 2016

Introduction

As climate change advances in its threat to biodiversity worldwide, reliable predictions of range dynamics are

needed to anticipate and mitigate potential impacts (Pereira *et al.*, 2010; Bellard *et al.*, 2012), and we have seen an upsurge of related methods and applications in recent years (Normand *et al.*, 2014; Lurgi *et al.*, 2015).

Range shifts are complex ecological processes driven by population dynamics and dispersal. These processes are co-determined by a plethora of other factors

Correspondence: Damaris Zurell, tel. +41 44 739 25 71, fax +41 44 7392 215, e-mail: damaris.zurell@wsl.ch

including changes to the abiotic and biotic environment (Sexton *et al.*, 2009). Adequately representing these interacting processes in an operational model and collecting data for reliably estimating the many parameters of such complex models is a formidable challenge even for a single species (Ehrlén & Morris, 2015) let alone for complex ecosystems. Older models relied on highly simplified conceptualizations where the abiotic environment is the essential driver of climate-induced range shifts ignoring any demographic processes involved in range shifts. These so-called species distribution models (SDM; Guisan & Thuiller, 2005; Guisan & Zimmermann, 2000) have reached high popularity for providing biodiversity scenarios under climate change, owing to the strong simplification of the represented processes and their relative ease to use. However, their use for climate change projections has been discussed controversially (Dormann *et al.*, 2012; Thuiller *et al.*, 2013) because SDMs assume that observed species' distributions are not affected by dispersal limitations (Svenning *et al.*, 2008) or source-sink dynamics (Holt *et al.*, 2005), and ignore complex transient dynamics during range shifting (Zurell *et al.*, 2009; Lawler *et al.*, 2013).

To address these issues, more mechanistic approaches of modelling range dynamics have been advocated (Thuiller *et al.*, 2008; Gallien *et al.*, 2010) and several frameworks have been developed (or revived) (Pereira *et al.*, 2010; Ehrlén & Morris, 2015; Lurgi *et al.*, 2015). While all of them attempt to overcome the limitations associated with SDMs, their ability to improve projections for species' range dynamics has never been compared systematically. The lack of such evaluation is likely due to the difficulty to get appropriate benchmark data, consisting of information on distribution, abundance and demography. Such complex datasets are rare and benchmarking may be hampered if not all processes occurring in these ecosystems are fully understood (Cheaib *et al.*, 2012). We propose to conduct benchmarking of new methods for modelling range dynamics using simulated community data, which allows controlling all relevant ecological processes driving species distribution and track transient dynamics in space and time ('virtual ecologist approach', Zurell *et al.*, 2010).

We compare five generic modelling frameworks for predicting range dynamics capable of (fast) calibration for any single species (Fig. 1). The choice of models was guided by our objectives to include frameworks that are representative of current approaches for predicting actual abundance (Ehrlén & Morris, 2015), that differ in their underlying assumptions and in the complexity of data and process detail needed to parameterize them, and for which we could find proficient users

to run the simulations for our study to assure a fair comparison. Due to a lack of common terminology, we refer to all five models (including SDMs) as range models and to those models that explicitly consider dynamic processes of range shifts (dispersal, population dynamics) as range dynamic models (Fig. 1). A subset of these models relies on SDMs to predict habitat suitability and infer demographic rates ('SDM hybrids'). One approach infers demographic rates directly from the data and models habitat suitability as an outcome of demographic processes (Fig. 1).

We explicitly focus our comparison of models on predicting range dynamics of single species. Virtual species, however, were simulated within a virtual community to imitate constraints on species distribution and abundance resulting from both abiotic factors and biotic interactions. We first simulated virtual communities using a dynamic, individual-based, multi-species simulation model, and then tracked these communities under climate change. Range models were calibrated using data from the output of the virtual community model. We tested the effects of different demographic (dispersal, source-sink dynamics) and community processes (single species, species sorting, neutral dynamics) as well as of environmental stochasticity on model predictive performance.

By comparing the performance of the different range models before and during climate change we aim at answering the following questions: (i) Do SDMs and different range dynamic models predict current species' distribution and abundance equally well? (ii) Do more complex, demography-based approaches consistently outperform SDMs under climate change? (iii) How is predictive model performance affected by prevailing demographic and community processes? (iv) What are the effects of structural decisions during model building? We found that more complex range dynamic models improved current range projections considerably compared to purely correlative SDMs. Under climate change, simpler range dynamic models often outperformed more complex models, especially when benchmarking data were simulated with more complex demographic and community processes. We discuss guidelines and perspectives for increasing the reliability of climate change-induced range predictions and for applying range dynamic models more widely in climate impact assessments.

Material and methods

Overview of range dynamic models and hypotheses

We compare a classical SDM and four different range dynamic models, three of which are SDM hybrids (Fig. 1,

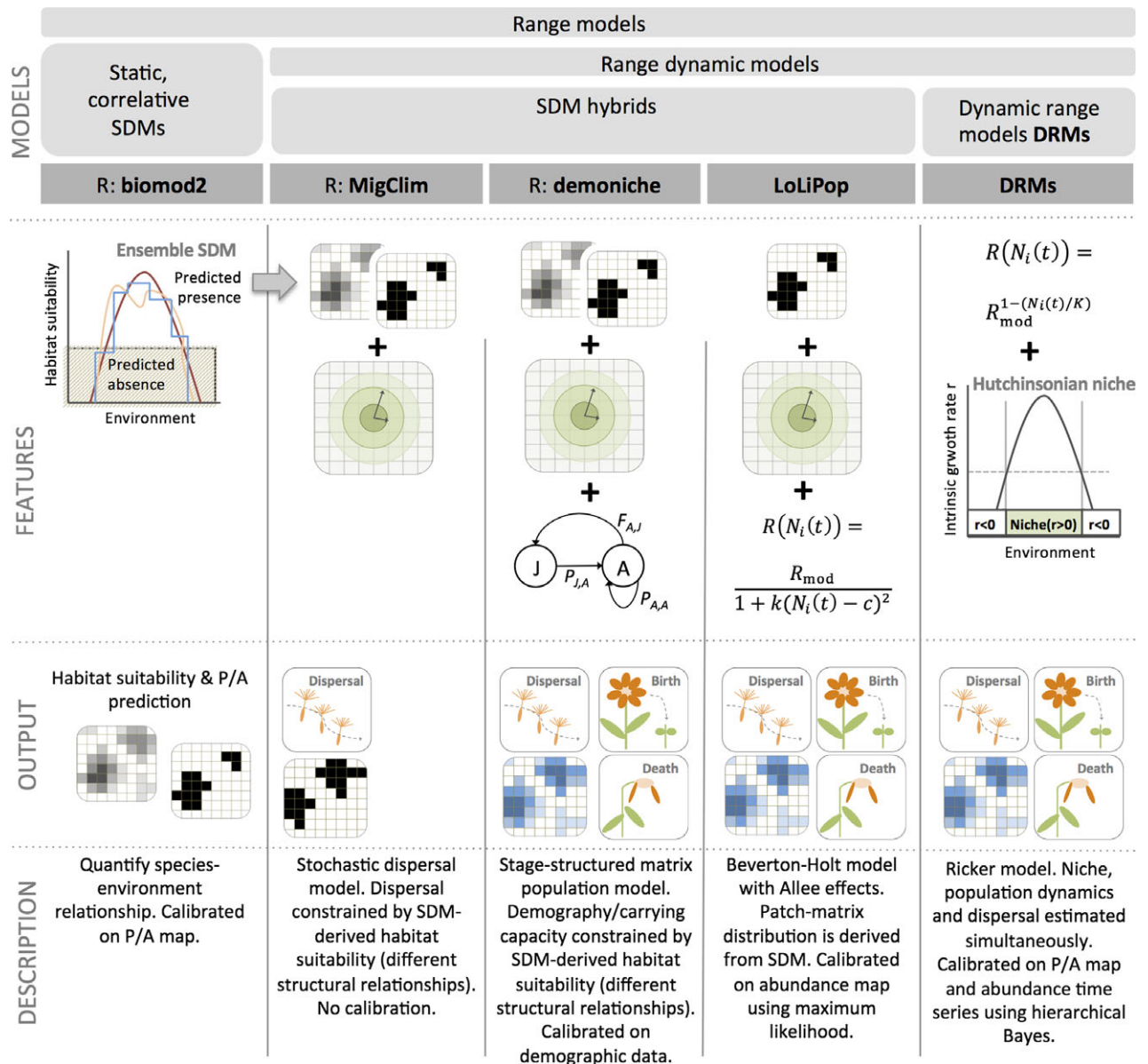


Fig. 1 Main characteristics of implemented range models used for model intercomparison. Habitat suitability maps (darker gray shades indicate higher suitability) or presence/absence maps (P/A, with black indicating presence) derived from SDMs serve as input to SDM hybrids. These are then linked with dispersal kernels and with a population model (except MigClim). DRMs infer the environment–demography relationship directly from the data and do not rely on SDMs, which is a major difference to DemoNiche and LoLiPop. Importantly, DemoNiche is calibrated on demographic data, while the demographic models in LoLiPop and DRM are calibrated on spatial abundance data with respective P/A maps and abundance time series. Different structural relationships may be assumed to link MigClim and DemoNiche to SDM-derived habitat suitability or P/A predictions. MigClim outputs P/A maps as predictions, while all other dynamic range models output abundance maps as predictions (darker blue shades indicate higher abundances). The prefix ‘R:’ indicates the available R package (R Development Core Team 2014).

Appendix S1 in Supporting Information). The simplest SDM hybrid, MigClim, supplements SDM predictions with distance-dependent colonization probabilities (Engler & Guisan, 2009; Normand *et al.*, 2013). However, local demographic processes including regeneration and mortality are not explicitly accounted for although these are crucial for

predicting population viability and spread rates. More complex SDM hybrids couple SDM-derived habitat suitability maps and population models (Keith *et al.*, 2008; Dullinger *et al.*, 2012; Zurell *et al.*, 2012b). These models can be calibrated with simple demographic information as we demonstrate with DemoNiche (Nenzén *et al.*, 2012) or can be

inversely fitted to abundance data as demonstrated by the application of LoLiPop (Cabral & Schurr, 2010).

The value of such SDM hybrids is debated, mainly because of potential circularity problems (Gallien *et al.*, 2010) and because the relation of SDM-derived habitat suitability to species demographic parameters remains unclear (Thuiller *et al.*, 2014). Dynamic range models (DRM) have been introduced to overcome these issues. They do not rely on SDM output and directly relate demographic rates to environmental factors and simultaneously estimate parameters of the population model and the demography-environment relationship from abundance and distribution data (Pagel & Schurr, 2012). We hypothesize that predictive performance under climate change will increase from left to right in Fig. 1 because (i) range dynamic models explicitly model the dispersal process and should hence outperform simple SDMs, (ii) population dynamic models (DemoNiche, LoLiPop, DRM) additionally model abundance dynamics from differently resolved demographic processes and should thus outperform MigClim that only simulates potential colonization, and (iii) DRMs jointly estimate the effects of dispersal and demography on distribution and abundance dynamics and should thus outperform SDM hybrids which may suffer from circularity problems.

Virtual species/communities

Simulation model. An individual-based, spatially explicit, stochastic model (IBM; Gravel *et al.*, 2006; Münkemüller *et al.*, 2012) was adapted to simulate the dynamics of a focal species and its co-occurring competitors in heterogeneous environments. The IBM is a cellular automaton in which each sub-cell is characterized by unique environmental conditions (temperature and soil moisture) and can support one sessile individual. In the following, we provide a simple overview over the main characteristics of the simulation model, while more details are given in Appendix S2.

The IBM's spatial resolution only allows individual counts, yet the range models work at the population level. We therefore implemented a hierarchy of two spatial scales so that the sessile individuals are interacting locally but that aggregations at larger scale (10×10 sub-cells) provide information on community composition and species abundance, which serve as input for the range model comparison. One time step corresponds to 1 year and the four main processes within one time step are large-scale, contagious disturbances at the coarse resolution, and then the local (sub-cell) processes of adult mortality, propagule supply, and recruitment of propagules to adult size (Fig. S1).

1. Large-scale contagious disturbances like fire or windthrow act at the coarse scale with an overall probability of 0.05, resulting in stochastic population dynamics and incomplete range filling. Disturbances were omitted for some scenarios (see Simulations), and are initiated in randomly selected cells at the coarse scale and spread to the eight nearest neighbours. In response to disturbance, individuals are killed with a probability of 0.9.

2. Within each time step, adults die with a probability of 0.1 and can be replaced by recruits from the local community or by immigrants.
3. Propagule supply is determined by offspring production and by propagule dispersal. Only adults that are at least one time step old can produce propagules. Offspring production rate is determined by the local (temperature and moisture) environment, and is described by a two-dimensional Gaussian function for the reproductive niche, with each species having a unique optimum. The Gaussian function is cut off at a threshold of 0.001 to obtain finite niche breadth. Dispersal is simulated using a negative exponential dispersal kernel where the rate parameter α determines the mean dispersal distance ($1/\alpha$).
4. Recruitment into empty space or by competitive replacement follows a lottery function of dispersal-driven propagule supply and interspecific differences in competitive ability. Competitive performance depends on the species' niche, in particular the probability λ of propagules to survive in the understory of adults prior to recruitment, thus mimicking simple resource competition. An individual's competitive performance (the probability of a propagule to establish and replace the resident) is proportional to the ratio between its survival probability λ and that of the resident (or between λ and a threshold of 0.1 in empty cells). The survival probability λ is a function of the local environment, and we assumed a two-dimensional Gaussian function for the survival niche (driven by soil moisture and temperature, see section 'Simulations and sampling' for more detail), which is equivalent to the reproductive niche, if not mentioned otherwise. The cells can stay empty, if the overall propagule rain is too small.

The species' fundamental niche is equivalent to the reproductive niche while the realized niche can be smaller due to interspecific competition (determined by the overlap of propagule survival niches of competing species), or can be larger due to source-sink dynamics, which occur in the IBM if the species' survival niche is wider than its reproductive niche.

Simulations and sampling. The virtual communities were simulated in artificial landscapes of 20×125 (coarse-scale) cells. Spatially auto-correlated patterns of soil moisture in the coarse-scale landscape were generated from two-dimensional fractal Brownian motion. Spatial variation in temperature was represented as a linear latitudinal gradient with added spatially auto-correlated noise. Sub-scale environmental heterogeneity at the scale of the 10×10 sub-cells was added as normally distributed noise to the coarse-scale cells' mean temperature and moisture. We simulated temporal variability in temperature by adding for each time step a temporally auto-correlated random deviate to temperature.

Simulations were initiated by randomly distributing the virtual species in their respective suitable habitats. First, simulations were run for 900 spin-up years under current environmental conditions and variability to ensure that species/communities were in (dynamic) equilibrium with the environment. After the end of the spin-up period (hereafter

referred to as year 0), climate change was initiated with a linear increase in temperature of 3°C over a period of 100 years. Soil moisture patterns were assumed to remain constant over the simulation period.

We ran scenarios for three different community types: (i) neutral dynamics, (ii) species sorting, and (iii) single species without any competitors. The first two community simulations consisted of nine interacting species. For the species-sorting scenarios, the niche optima of the eight co-occurring species in environmental space were symmetrically arranged around the focal species niche with a fixed radius (Table S1, Fig. S4). In the neutral scenarios, all species had equivalent niche optima and widths. All species within a community were assumed to have equal dispersal ability.

We ran four scenarios for each type of community dynamics:

1. SR: short-range dispersal without large-scale disturbances (with $\alpha = 1/\text{mean dispersal distance} = 0.05$).
2. LR: like SR but with long-range dispersal (LR; $\alpha = 0.1$).
3. SR + sinks: source-sink dynamics where the reproductive niche was narrower (by 1/3) than the survival niche resulting in the realized niche being larger than the fundamental niche (Fig. S4).
4. SR + cont: with large-scale contagious disturbances.

Overall, we ran 12 different scenarios as input for the subsequent range model intercomparison. To avoid increasing prevalence in the single species and in the source-sink scenarios, the fundamental niches for these scenarios were reduced (Table S1 and Fig. S4).

Sampling data were gathered at the coarse spatial scale. We assumed no detection errors and thus species data always represented a 'perfect' sampling of the virtual world. Different kinds of data were sampled according to the input needs of the range models (Fig. 1), including presence-absence and abundance data ($n = 500$, year 0), abundance time series ($n = 50$, years -10 to 10), and mean demographic rates (years -20 to 0 ; Appendix S1).

Calibration of range models

SDMs were fitted by relating the sampled presence/absence data from year 0 to the two environmental variables using the ensemble platform *biomod2* in R (Thuiller *et al.* 2009) and simple averaging was used to derive consensus predictions. Predicted habitat suitability was transformed to presence/absence predictions by applying a TSS-maximizing threshold (true skill statistic; Allouche *et al.*, 2006).

MigClim combines habitat suitability or presence/absence predictions from SDMs with a dispersal kernel to predict colonization probabilities (Engler & Guisan, 2009). Here, habitat suitability was rescaled to range between 0 and 1 and a sigmoidal relationship was used to relate habitat suitability to colonization probability. Other structural relationships were tested and are discussed in Appendices S1 and S4. As dispersal kernel we took the known dispersal kernel from the IBM.

DemoNiche is a stage-structured matrix population model that constrains demographic rates or carrying capacity by

habitat suitability (given by an SDM) and connects local populations by a dispersal kernel (Nenzén *et al.*, 2012). Thereby, choices have to be made regarding the demographic property (vital rates or carrying capacity) that should be constrained and the relationship to habitat suitability (linear, sigmoidal, threshold). Here, we evaluated predictive performance of different model configuration for year 0, and averaged predictions of the five best model configurations (see Appendix S1). Appendix S4 provides a more detailed discussion on effects of structural uncertainty. We used a box-constraint variable metric algorithm to calibrate the transition matrix by minimizing differences between observed and predicted mean demographic rates (Appendix S1). Calibrating the transition matrix was necessary because the induced functional relationship between habitat suitability and demographic transition probabilities implicitly assumes that demography is known for optimal conditions (habitat suitability = 1) while demographic rates are usually averaged across the population. As dispersal kernel we took the known dispersal kernel from the IBM.

LoLiPop simulates local population dynamics with different populations connected by two-dimensional dispersal kernels (Cabral & Schurr, 2010). Population dynamics can only take place on suitable cells (given by SDM). Here, local population dynamics followed a Beverton-Holt model extended with Allee effects (Cabral & Schurr, 2010). Demographic parameters were estimated from spatial abundance data from year 0 using maximum likelihood estimation. As dispersal kernel we took the known dispersal kernel from the IBM.

The DRM is based on a simple stochastic model of local population dynamics (Ricker model) coupled by a negative exponential dispersal kernel. The demography-environment relationship is formulated as a regression of intrinsic population growth rate against the environmental variables and the carrying capacity is assumed to likewise vary proportional to the growth rate. This demography-environment relationship, the proportionality between growth rates and carrying capacity as well as the mean dispersal distance are jointly estimated from presence/absence data from years -10 and 10 and from 50 randomly selected abundance time series (covering the same 20 years) using a hierarchical Bayesian modelling approach (Pagel & Schurr, 2012).

Assessing range model performance

Models were calibrated under equilibrium conditions and were then used to project species distribution and abundance under climate change. Predictive performance was evaluated against full known truth for each time step within the 100 years of climate change. We assessed accuracy of spatial predictions by first converting all predictions to binary maps and then calculating TSS (Allouche *et al.*, 2006). Correlations between observed and predicted abundance as well as predicted occurrence probabilities were assessed using Spearman's rank correlation coefficient ρ . Furthermore, we compared positions of range front, centre and rear edge by calculating quantiles (95%, 50% and 5%, respectively) of the observed and predicted latitudinal positions. Last, we

calculated relative deviation in total abundance (summed over all cells) as difference between predicted and observed abundance divided by observed abundance (except for SDMs and MigClim), and differences in relative abundance estimates, which is the relative decrease in observed and predicted abundance since year 0.

Results

Stochastic community IBM

The 12 different scenarios led to distinct spatial and temporal distribution of presence-absence and abundance of the focal species (Appendix S2). Generally, short-range dispersal resulted in stronger spatial clustering of populations. Differences in spatial clustering for long- and short-range dispersal were particularly pronounced in neutral communities and large-scale contagious disturbances resulted in even patchier spatial distributions. Under climate change, the focal species exhibited range shifts accompanied with distinct population decreases. Also, for all scenarios the focal species showed distinct time lags in range filling following climate change with dispersal limitations at the range front and persistence in unsuitable conditions at the rear range edge (Figs 2 and S6). Dispersal limitations were more pronounced in short-range dispersal scenarios and under biotic interactions. In the

species-sorting scenario, competing species were blocking the range front, whereas in neutral communities, dispersal success became more random due to strong spatial clustering. Overall, these diverse distributions provided a solid basis for comparing the predictive performance of SDMs and range dynamic models.

Range model performance under current conditions

For the observation period (year 0), DRMs best predicted the focal species' mean and maximum abundances along the temperature gradient (Figs 2, S7 and S8). SDMs often slightly overestimated occurrence probability at range margins, which propagated differently to the SDM hybrids. For example, LoLiPop predicted local abundances near range margins quite successfully but underestimated abundances at range centres, whereas DemoNiche overpredicted local abundances across nearly the whole range. Correspondingly, DRMs obtained highest TSS and Rho in year 0 for most scenarios although differences between range models were minor except that DemoNiche achieved much lower TSS scores while Rho scores were consistently high (Fig. 3). By contrast, LoLiPop and SDMs predicted range positions under equilibrium conditions best while DemoNiche and DRMs predicted slightly too large ranges. DRMs and LoLiPop approximated

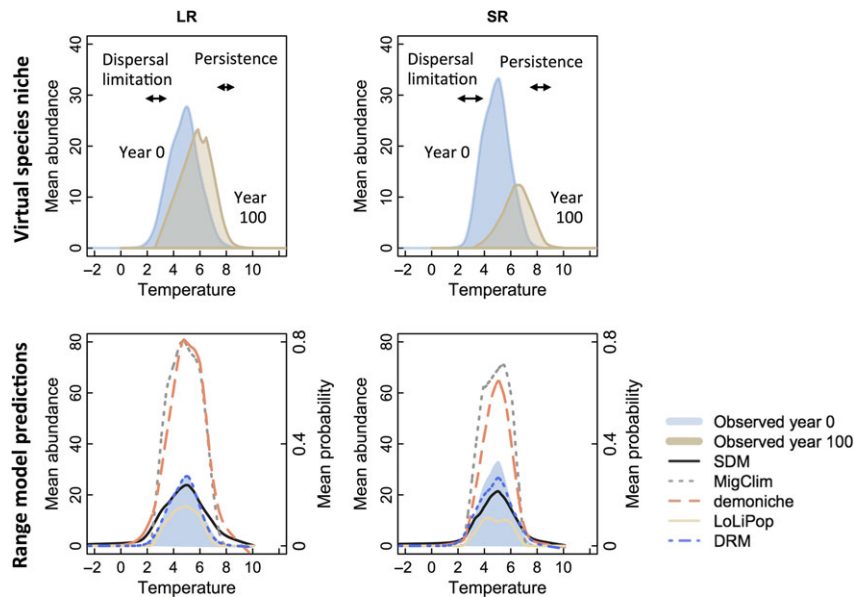


Fig. 2 Realized temperature range of focal species for the long-range (LR) and short-range (SR) dispersal variants of the species-sorting scenario as approximated by mean observed abundances along the temperature gradient. Top row shows the niche filled by the virtual species before (year 0) and after climate change (year 100). For both scenarios, realized ranges in year 100 exhibit lower overall abundances and a shift to the warmer end of the temperature tolerance (i.e. dispersal limitation at the cold front, persistence at the warm rear edge). Bottom row shows the corresponding predictions by the different range models for year 0.

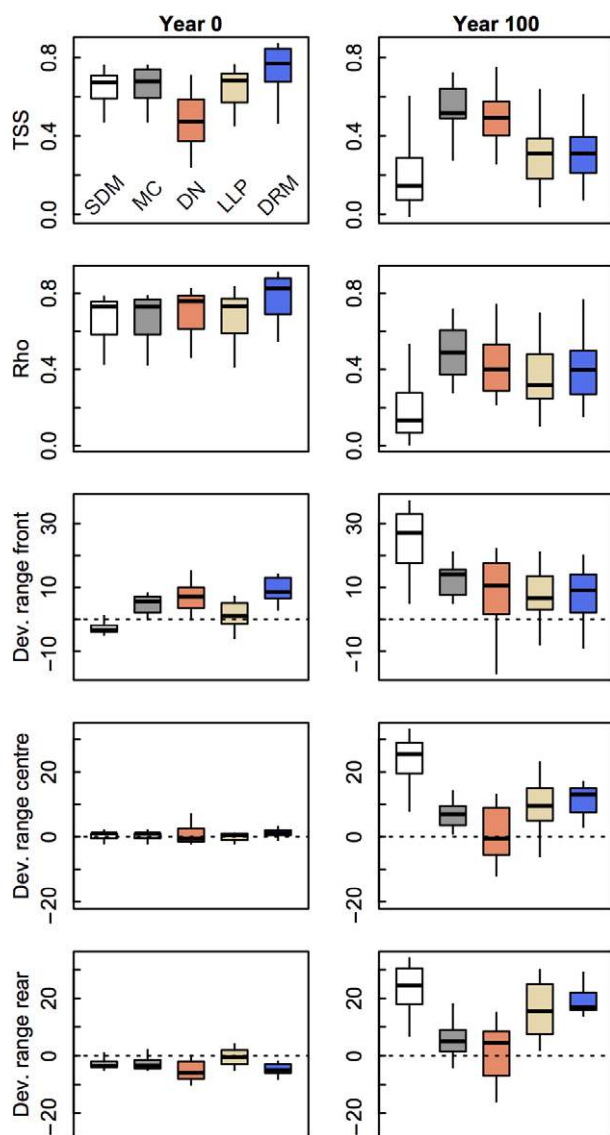


Fig. 3 Boxplots depicting performance of different range models over all scenarios. We show TSS, Spearman's rank correlation coefficient Rho and deviations from range front, centre and rear for the years 0 and 100. SDM: species distribution model, MC: MigClim, DN: DemoNiche, LLP: LoLiPop, DRM: dynamic range model.

total abundance best while DemoNiche generally overestimated abundances (Fig. 4).

Range model performance under a warming climate

Under climate change, prediction accuracy of SDMs decreased significantly because the range shifting potential was greatly overestimated by the full-dispersal SDMs (Fig. 3), especially under short-range dispersal (Figs 5 and 6). All range dynamic models were able to overcome these limitations to some extent, achieving

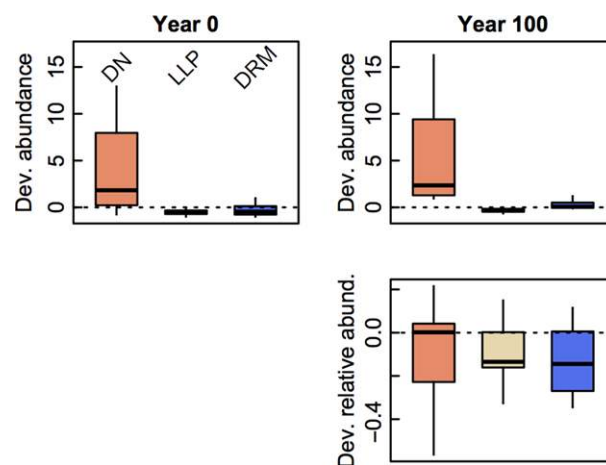


Fig. 4 Boxplots depicting abundance estimates of population models over all scenarios. Top row shows the factor of deviation in total abundance with positive and negative values referring to overestimation and underestimation of true abundance, respectively. Bottom row shows deviation in relative abundance with year 0 as reference. DN: DemoNiche, LLP: LoLiPop, DRM: dynamic range model.

higher TSS and Rho values than SDMs and generally showing less deviation between observed and predicted range margins (Figs 3, 5 and S11-S27). Here, DRMs were outperformed by SDM hybrids though often only marginally, with MigClim and DemoNiche showing considerably higher TSS scores, slightly higher Rho scores and smallest average deviations from range centre and rear edge. LoLiPop predicted lowest average deviations from range front. Again, total abundance was best predicted by DRMs and LoLiPop, and overestimated by DemoNiche. However, in terms of relative change in abundance, all population models produced more similar projections with a tendency towards underestimating the relative abundance and, thus, overestimating extinction risks (Figs 4 and S28-S29). On average, DemoNiche best predicted relative change in abundance.

Effects of demographic and community processes

Prediction accuracy (TSS and Rho) of all range models decreased most under short-range dispersal and for scenarios including biotic interactions (Fig. 6). Under neutral dynamics, which produced the most complex range-shifting dynamics, the simple MigClim generally performed best. Overestimation of future total abundance by DemoNiche was particularly strong under neutral dynamics. Here, LoLiPop predicted total abundance best, while there were no clear winners among the population models for the other cases. However, in some scenarios, LoLiPop estimated quite extreme demographic rates that resulted in over-compensatory

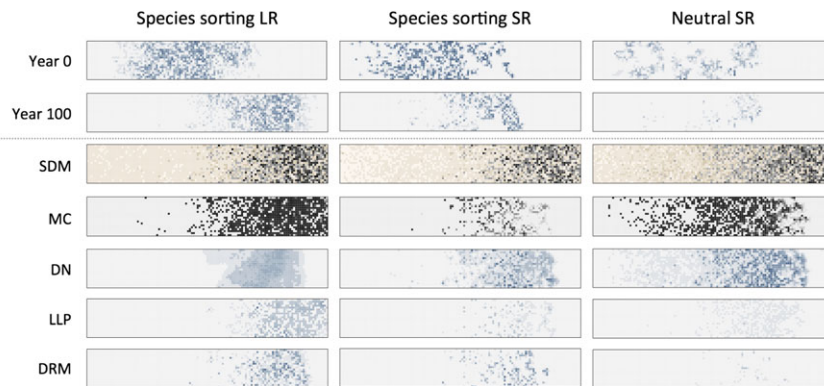


Fig. 5 Observed abundances of virtual species for years 0 and 100, and predictions of range models for year 100 after climate change for the species sorting scenario with long range dispersal (left) and short range dispersal (centre), and for the neutral dynamics scenario with short range dispersal (right). Abundances are presented in blue shading with darker colour indicating higher abundance. For SDMs, predicted habitat suitability is shown with sandy colours indicating suitability values that correspond to predicted absences, and grey shading indicating suitability values that correspond to predicted presences, with darker colours indicating higher habitat suitability. MigClim predicts colonised (in black) vs. uncolonised cells. Grey shading in MigClim indicates that these cells were not colonised in all replicate runs. SDM: species distribution model, MC: MigClim, DN: DemoNiche, LLP: LoLiPop, DRM: dynamic range model.

local population dynamics, most severely for the long-range dispersal, species sorting scenario (Fig. S29). Generally, underestimation of relative abundance and, thus, overestimation of extinction risks was more severe under long-range dispersal and, to a minor extent, under neutral dynamics.

Effects of structural uncertainties

Effects of structural uncertainty were very pronounced in all range dynamic models (Appendix S4). For example, assuming a linear relationship between SDM-derived habitat suitability and colonization probability in MigClim led to misleading predictions of near-complete colonization of the entire grid in long-range dispersal scenarios (Figs S30-S31). In DemoNiche, prior knowledge of the demography-environmental relationships helped reducing uncertainty considerably. Also, model configurations that achieved highest accuracies under equilibrium conditions usually remained among the best for climate change predictions. We found no clear differences between constraining demographic rates or carrying capacities by habitat suitability in DemoNiche, although the latter showed reduced accuracy in long-range dispersal scenarios (Figs S32-S34).

Discussion

We took the challenge of benchmarking state-of-the-art range models of varying complexity using a comprehensive set of simulated data that account for demographic and community processes. Based on the comparison across these benchmark data and diverse

model outputs, our initial questions can be answered as follows: (i) Under current climate, more complex range dynamic models like DRMs better fit the observed species distributions and abundances, although differences are small. (ii) Under climate change, all dynamic approaches improved predictions over simple SDMs. We could not, however, identify a single, best approach for making predictions. Predictions of absolute abundance differed markedly between population models while predictions of relative abundance were similarly accurate. (iii) Differences in dispersal ability and complex biotic interactions may introduce high uncertainty in range predictions, while the effects of source-sink dynamics and increased disturbance were minor. (iv) In all range dynamic models, structural decisions during model building can have great impact on model accuracy, but prior system knowledge on important processes can reduce these uncertainties considerably. Our results reassure the clear merit in using dynamic approaches for range predictions. But they also emphasize several needs, namely: (i) to compare and combine multiple model outcomes for better capturing the uncertainty associated with range predictions under climate change; (ii) to gather more and higher quality data on species' demography; (iii) to run preliminary tests with each demographic model in order to optimize the structural decisions and settings; and (iv) to incorporate assembly processes for better capturing the within-community dynamic processes and their constraints on range dynamics. In summary, implementing these insights will greatly help advancing our ability of predicting future range dynamics and making these approaches operational for larger numbers of species.

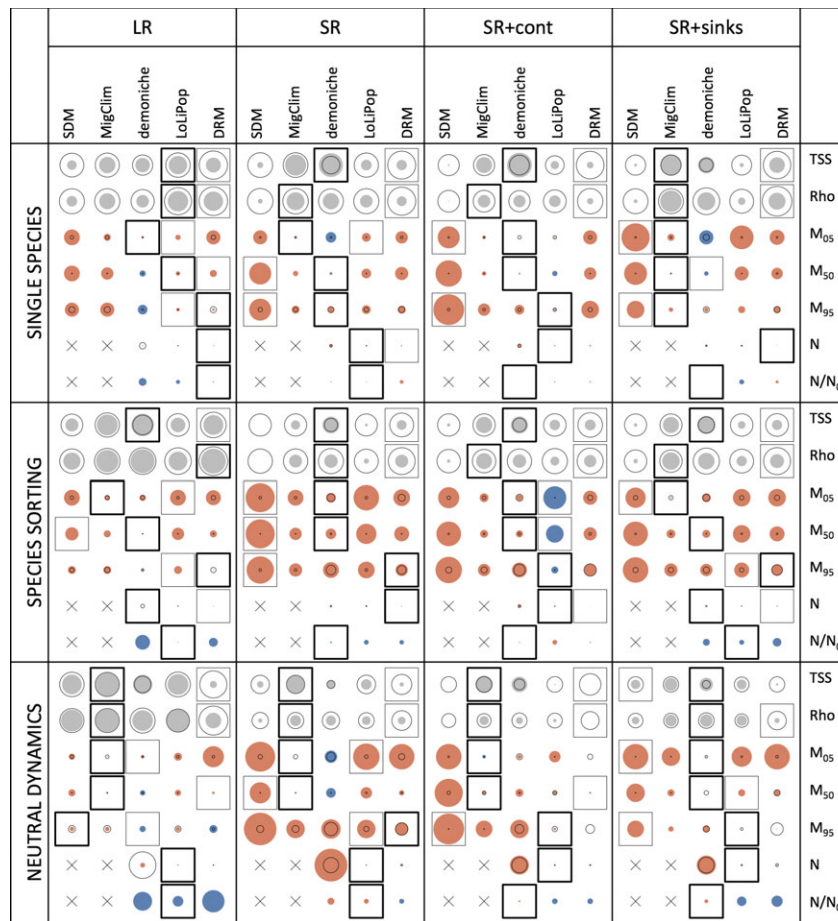


Fig. 6 Prediction accuracies (gray) and error rates (orange/blue) obtained for range models under different scenarios, with circle sizes being proportional to accuracy or error rate. The table compares measures before (open circles) and after climate change (filled circles). The more similar filled circles are to open circles, the less decrease in prediction accuracy or the less increase in error rates were observed over time. Squares indicate the best model for year 0 (thin lines) and year 100 (thick lines). Accuracy measures: TSS (true skill statistic) and Rho (Spearman's rank correlation coefficient). Error measures: M05, M50 and M95: absolute differences between observed and predicted range margins (5%, 50% and 95% quantiles of latitudinal distribution). N : factor by which predicted absolute abundance overestimates/underestimates observed abundance. N/N_0 : absolute difference between observed and predicted relative abundance. Orange indicates overestimation and blue underestimation. Maximum circle sizes correspond to TSS = 0.87, ρ = 0.89, M = 37 cells, N = 22, and N/N_0 = 57%.

Range dynamic models on a par for current climates

Under current climates, DRMs provided the best fit in most cases, although their advantages in prediction accuracies were generally small compared to other range dynamic models and also compared to SDMs, which were not consistently outperformed by the more dynamic approaches. DRMs jointly estimate the demography-environment relationship, dispersal and other demographic parameters and, thus, avoid possible circularities that might arise from the fitting steps for SDM hybrids (Gallien *et al.*, 2010; Pagel & Schurr, 2012). Yet, for the source-sink scenarios that could be assumed to be problematic for SDM hybrids while not for DRMs (Pagel & Schurr, 2012; Schurr *et al.*, 2012), we

did not find evidence that DRMs generally outperform hybrids.

Separate fitting steps in SDM hybrids of first fitting SDMs and then fitting the population model may lead to bias when species are not in equilibrium with their environment. However, if species are dispersal limited in some parts of their range but the entire niche in environmental space is well-covered by data, as is the case for our source-sink and contagious disturbance scenarios, circularity in SDM hybrids appear not to be a limiting problem. If, on the other hand, observed species' distributions are biased by dispersal limitations such that parts of the environmental niche are not filled, then DRMs may be the preferred choice, as this approach directly accounts for the (limited) dispersal in

the estimation of the species' niche and can thereby reduce the bias that may otherwise arise from the disequilibrium (Pagel & Schurr, 2012). Still, more research is needed to (i) develop robust approaches for assessing limitations in range filling *a priori* (Svenning & Skov, 2004) as an important model assumption of SDM hybrids, (ii) to assess how prevalent the phenomenon of incomplete range filling is in extant species, and (iii) to evaluate DRM predictive performance and their ability to accurately distinguish between environmental filtering and dispersal limitations in such situations.

Large variation in predictive performance under future climates

All range dynamic models tested here considerably improved climate change projections compared to SDMs, although relative model performance differed from those under current climates. Surprisingly, advantages of DRMs in predicting current ranges did not result in better predictions of future ranges compared to SDM hybrids. This may result from different calibration approaches, calibration data and process detail covered by the models. For example, the relatively poor performance of DRMs might arise to some extent from uncertainty in estimated dispersal rates while dispersal kernels were known for SDM hybrids. Also, DRMs overestimated the position of future trailing range edges (Figs 3 and 6). This likely originates from the fact that DRMs did not explicitly describe adult survival, which is independent from the environment in the IBM, but summarized all demographic processes in an environment-dependent population growth rate. Therefore, the transient persistence of populations when temperatures have become unfavourable at the rear edge (Fig. 2) was not accurately predicted. Consequently, the different SDM hybrids outperformed the DRM in almost all scenarios. Notably, also MigClim, the simplest approach that only supplemented SDMs with a dispersal kernel, showed spatial predictions that were, on average, similarly accurate as the more complex approaches.

The main advantage of the more complex approaches is that they also predict population dynamics and associated extinction risks. Both before and after climate change simulations, DemoNiche largely overestimated abundance, while LoLiPop and DRMs produced more reasonable estimates. Nevertheless, the relative decrease in abundance and associated extinction risks were often better approximated by DemoNiche than by the other two population dynamic models. This is likely due to the fact that DemoNiche uses demographic instead of abundance data for calibration, and that DemoNiche does not restrict abundance except when

habitat suitability was related to carrying capacity. Calibrating demographic parameters directly allows DemoNiche to more accurately predict population trajectories in many cases. In fact, calibration on abundance data led to estimation of partially unrealistic demographic rates in LoLiPop resulting in over-compensatory local population dynamics (although this could be avoided by choosing an alternative underlying population model; Cabral & Schurr, 2010). Although also calibrating on abundance data, this effect was not found in DRMs, as DRMs use abundance time series for calibration that inherently include information on demographic trajectories.

We conclude that (i) predictions of relative change in abundance are often more reliable than predictions of absolute change in abundance, a feature already known for population viability models (Beissinger & Westphal, 1998; Zurell *et al.*, 2012b), and that (ii) data on demography either through direct measurements or through abundance time series are indispensable for reliably calibrating population dynamic models (Schurr *et al.*, 2012). Our study confirms that all calibration approaches using demographic data and/or spatial or temporal abundance data can generate reasonably accurate predictions overall. However, models based solely on spatial abundance data may exhibit high uncertainty in future predictions and require careful testing of model structure. A constraining point is that accurate data on demography and spatial distribution and abundance are not available for many species and in high quality. Also, density is not easy to measure in many plant communities, where biomass or relative cover is preferentially recorded. Time series of relative cover may show strong inter-annual variability, which is not necessarily related to population processes (Boulangéat *et al.*, 2012). Thus, we not only need to increase our efforts in data collection but also in defining (more) meaningful response variables for population dynamic studies.

Effects of demographic and community processes

Generally, spatial prediction accuracies of all range models decreased when the focal species was interacting with other competitor species, especially under short-range dispersal (Fig. 6). In contrast, the effects of large-scale disturbances and source-sink dynamics on model accuracy were comparably low. None of the tested range dynamics models account for biotic interactions and hence they all experienced difficulties in these particular cases. More work is thus needed to incorporate assembly processes in such dynamic models (Boulangéat *et al.*, 2012; Cabral & Kreft, 2012; Mokany *et al.*, 2012). Therefore, caution is advised with

these models when biotic interactions are highly stochastic as was shown in the neutral scenarios. Although the assumption of neutral community dynamics is much debated for temperate ecosystems, observed spatial distributions are often astoundingly consistent with neutrality even if driven by different mechanisms (Bell, 2005). Our results indicate that whenever the ecosystem under study is strongly affected by biotic or environmental stochasticity, simpler range models such as MigClim may be preferred over more complex population models although prediction of population dynamics and abundances cannot be retrieved from such models.

We additionally stress that although predictive performances of all models decreased stronger in the short-range compared to long-range dispersal scenarios, this does not imply that predictions are generally more reliable for long-distance dispersers. Our results need to be judged with some caution and with respect to the virtual simulation model setup, in which all species of the community had the same dispersal ability and recruitment was proportional to the amount of propagule rain. This is, of course, a simplifying assumption and we will likely observe even more complex community dynamics if species vary in dispersal ability (Cabral & Kreft, 2012). It is, thus, reasonable to assume that long-distance dispersers may also experience substantial migration limitation from competitors.

Uncertainty through structural decisions in range dynamic models

Most applications of range dynamic models ignore uncertainty in model structure and their effect on prediction uncertainty (but see Cabral & Schurr, 2010). Our results clearly show that structural decisions in model building are crucial and may strongly affect prediction accuracies. Range dynamic models necessarily simplify the species–environment relationship as well as the colonization and extinction processes and these simplifying assumptions may lead to large uncertainties (Appendix S4). Important structural decisions in range dynamic models relate to differences in the relationship between habitat suitability and demographic rates, which has hitherto little empirical support (Thuiller *et al.*, 2010, 2014; McGill, 2012) and should receive more attention in the future. The different range dynamic models may strongly differ in how variation in the environment-dependent demographic rate drives variation in (i) local abundance or carrying capacity, (ii) persistence, and (iii) propagule dispersal and associated colonization success. This aspect becomes even more important, when such models are applied in a climate change context, where different

sources of uncertainty need to be considered in order to make meaningful projections with sufficient attention given to the sources of uncertainties.

Such structural aspects are rarely considered in dynamic range predictions, but we strongly advise to do so and to assess to what degree prediction accuracies vary under different model structures. In DemoNiche, for example, using habitat suitability to constrain carrying capacity led to more realistic abundance estimates in some scenarios, but caused higher uncertainty and erroneous predictions of extinction in other scenarios. Constraining vital rates seemed more robust in that respect, but only if appropriate vital rates were selected according to prior knowledge of the species' ecology. Overall, we found that the DemoNiche configurations that achieved highest accuracy under equilibrium conditions usually remained among the best during climate change as well. Accordingly, the best model configurations under current climate could be used for making consensus forecasts. Alternatively, approximate Bayesian computation might be employed to optimize the structural link between the habitat suitability and demography in SDM hybrids given the data (Hartig *et al.*, 2011). This idea is similar to the information criterion based approach used by Cabral & Schurr (2010) in LoLiPop in order to select among different underlying population models including or not Allee effects and overlapping generations. Notably, the environment-demography relationships in LoLiPop and DRMs, in particular more differentiated environmental responses of different demographic processes, should also be explored more thoroughly in the future, although such assessment will require larger computational efforts. Nevertheless, we want to stress that the underlying idea of DRMs of jointly estimating the different constraints on the niche, namely environmental limitation, population dynamics and dispersal, is better integrated with ecological theory than hybrid approaches (for discussion see Schurr *et al.*, 2012).

Limitations and extensions

Clearly, even a model intercomparison using simulated data can by no means be exhaustive (Zurell *et al.*, 2010). Our choice of assembly processes, dispersal and other demographic processes, and landscape structure represents only one possible implementation and is still highly simplified in many aspects. Nevertheless, such approach allowed for generating consistent benchmarking data, producing an array of important demographic and community processes that are known to affect species' range limits. Therefore our results provide insight into comparative model performance in a wide subset of potential

cases. Clearly, subsequent comparison on field data and evaluation of other complicating factors such as sampling effort or bias, unequal dispersal abilities and unequal competitive ability of heterospecifics among others will be additionally informative. Also, the robustness of range dynamic models under novel environments (Williams & Jackson, 2007; Zurell *et al.*, 2012a), changing collinearity structures in environmental predictors (Dormann *et al.*, 2013) or arbitrary scale decisions (Thuiller *et al.*, 2015) should be tested in the future.

Moreover, although we aimed for a representative selection of current modelling frameworks for predicting range dynamics and actual abundance (Ehrlén & Morris, 2015), our study could only include a subset of available software applications. Lurgi *et al.* (2015) provide an extensive review on available computer platforms for predicting population- or individual-based range dynamics, which vary in the accommodated detail of demographic processes and complexity in species' lifecycles. In comparison to some other platforms, for example RAMAS (Akçakaya, 2000) that has been used rather widely in conservation context (cf. Fordham *et al.*, 2013) and more recently also for predicting range dynamics (e.g., Keith *et al.*, 2008; Anderson *et al.*, 2009; Pearson *et al.*, 2014; Swab *et al.*, 2015), DemoNiche, as our example of a stage-structured matrix population model, allows only a relatively low level of complexity in the species' modelled life history. As the life cycle of the simulated species was likewise simple, this is unlikely to have limited the performance in our model. However, for other applications that demand a higher level of detail in the species' life cycle or in environmental drivers, other approaches, such as RAMAS, might be preferable (Lurgi *et al.*, 2015).

Summary

Our range dynamic model intercomparison yielded no clear winners or losers. While all range dynamic models show clear benefits over simple SDMs, we cannot provide simple suggestions which model framework to choose for any single application. Currently, model choice will depend to a large extent on data availability and on prior knowledge on species' ecology (Lurgi *et al.*, 2015). For example, we currently lack spatiotemporal abundance data or solid demographic information for many taxa, which clearly limits model choice. As far as possible given data limitation, we generally advise a comparison of predictions from multiple models for improved understanding of model behaviour and prediction uncertainty (Cheaib *et al.*, 2012). Thereby, great discrepancies between model predictions may indicate that we missed important ecological

mechanisms. More efforts are needed to better understand the underlying mechanisms and its calibration in range dynamic models. This is specifically true for the interplay between demographic rates and biotic interactions in communities. Also, uncertainty through structural decisions should be assessed more routinely, and important model assumptions of range models should be verified *a priori*, for example the degree of range filling. Clearly, broader application of range dynamic models is limited by data and by computational efforts. Computation times are still quite high for DRMs compared to SDM hybrids. However, data availability is more crucial. We have shown that different kinds of calibration data (abundance, demographic rates) can be utilized, and that also SDM hybrids can be calibrated in a (semi-) automated way. Still, more efforts should be given to collecting longer-term and large-scale data on abundance and demography. Only such consistent data basis can ensure wide application of range dynamic models for climate impact assessment. Future studies should further focus on evaluating the effects of sampling effort and sampling bias, and of other complicating processes such as asymmetric competition or niche evolution on prediction accuracy of range dynamic models.

Acknowledgements

We thank Laure Gallien, Frank Schurr, Boris Schröder, Jane Elith and Antoine Guisan for helpful discussions. This study arose from two workshops entitled 'Advancing concepts and models of species range dynamics: understanding and disentangling processes across scales' in 2012–2013. Funding was provided by the Danish Council for Independent Research – Natural Sciences (grant no. 10-085056 to SN). DZ acknowledges funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement no. 624958. WT, TM and KS received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) Grant Agreement no. 281422 (TEEM-BIO). The LECA is part of Labex OSUG@2020 (ANR10 LABX56). JP acknowledges financial support from the German Research Foundation (DFG, grant SCHU 2259/5-1). JSC acknowledges financial support from German Research Foundation (DFG, via SA-21331 and the German Initiative of Excellence). DG was supported by a NSERC-Strategic grant. SD acknowledges support from the Austrian Panel on Climate Change (project B175127, SPEC-ADAPT). KS acknowledges support from the EU COST Action 'HarmBio' (ES1101). NEZ acknowledges support from the Swiss National Science Foundation SNSF (grant: 31003A_149508/1).

References

- Akçakaya HR (2000) Viability analyses with habitat-based metapopulation models. *Population Ecology*, **42**, 45–53.
- Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology*, **43**, 1223–1232.

- Anderson BJ, Akçakaya HR, Araújo MB, Fordham DA, Martinez-Meyer E, Thuiller W, Brook BW (2009) Dynamics of range margins for metapopulations under climate change. *Proceedings of the Royal Society. B, Biological sciences*, **276**, 1415–1420.
- Beissinger SR, Westphal MI (1998) On the use of demographic models of population viability in endangered species management. *Journal of Wildlife Management*, **62**, 821–841.
- Bell G (2005) The co-distribution of species in relation to the neutral theory of community ecology. *Ecology*, **86**, 1757–1770.
- Bellard C, Bertelsmeier C, Leadley P, Thuiller W, Courchamp F (2012) Impacts of climate change on the future of biodiversity. *Ecology Letters*, **15**, 365–377.
- Boullangeat I, Gravel D, Thuiller W (2012) Accounting for dispersal and biotic interactions to disentangle the drivers of species distributions and their abundances. *Ecology Letters*, **15**, 584–593.
- Cabral JS, Kreft H (2012) Linking ecological niche, community ecology and biogeography: insights from a mechanistic niche model. *Journal of Biogeography*, **39**, 2212–2224.
- Cabral JS, Schurr FM (2010) Estimating demographic models for the range dynamics of plant species. *Global Ecology and Biogeography*, **19**, 85–97.
- Cheab A, Badeau V, Boe J *et al.* (2012) Climate change impacts on tree ranges: model intercomparison facilitates understanding and quantification of uncertainty. *Ecology Letters*, **15**, 533–544.
- Dormann CF, Schymanski SJ, Cabral J *et al.* (2012) Correlation and process in species distribution models: bridging a dichotomy. *Journal of Biogeography*, **39**, 2119–2131.
- Dormann CF, Elith J, Bacher S *et al.* (2013) Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, **36**, 27–46.
- Dullinger S, Gatttringer A, Thuiller W *et al.* (2012) Extinction debt of high-mountain plants under twenty-first-century climate change. *Nature Climate Change*, **2**, 619–622.
- Ehrlich J, Morris WF (2015) Predicting changes in the distribution and abundance of species under environmental change. *Ecology Letters*, **18**, 303–314.
- Engler R, Guisan A (2009) MigClim: predicting plant distribution and dispersal in a changing climate. *Diversity and Distributions*, **15**, 590–601.
- Fordham DA, Akçakaya HR, Araújo MB, Keith DA, Brook BW (2013) Tools for integrating range change, extinction risk and climate change information into conservation management. *Ecography*, **36**, 956–964.
- Gallien L, Münkemüller T, Albert CH, Boullangeat I, Thuiller W (2010) Predicting potential distributions of invasive species: where to go from here? *Diversity and Distributions*, **16**, 331–342.
- Gravel D, Canham CD, Beaudet M, Messier C (2006) Reconciling niche and neutrality: the continuum hypothesis. *Ecology Letters*, **9**, 399–409.
- Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, **8**, 993–1009.
- Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecological Modelling*, **135**, 147–186.
- Hartig F, Calabrese JM, Reineking B, Wiegand T, Huth A (2011) Statistical inference for stochastic simulation models - theory and application. *Ecology Letters*, **14**, 816–827.
- Holt RD, Keitt TH, Lewis MA, Maurer BA, Taper ML (2005) Theoretical models of species' borders: single species approaches. *Oikos*, **108**, 18–27.
- Keith DA, Akçakaya HR, Thuiller W *et al.* (2008) Predicting extinction risks under climate change: coupling stochastic population models with dynamic bioclimatic habitat models. *Biology Letters*, **4**, 560–563.
- Lawler JJ, Ruesch AS, Olden JD, Mcrae BH (2013) Projected climate-driven faunal movement routes. *Ecology Letters*, **16**, 1014–1022.
- Lurgi M, Brook BW, Saltré F, Fordham DA (2015) Modelling range dynamics under global change: which framework and why? *Methods in Ecology and Evolution*, **6**, 247–256.
- McGill BJ (2012) Trees are rarely most abundant where they grow best. *Plant Ecology*, **5**, 46–51.
- Mokany K, Harwood TD, Williams KJ, Ferrier S (2012) Dynamic macroecology and the future for biodiversity. *Global Change Biology*, **18**, 3149–3159.
- Münkemüller T, De Bello F, Meynard CN *et al.* (2012) From diversity indices to community assembly processes: a test with simulated data. *Ecography*, **35**, 468–480.
- Nenzén HK, Swab RM, Keith DA, Araújo MB (2012) demoniche – an R-package for simulating spatially-explicit population dynamics. *Ecography*, **35**, 577–580.
- Normand S, Randin C, Ohlemüller R *et al.* (2013) A greener Greenland? Climatic potential and long-term constraints on future expansions of trees and shrubs. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, **368**, 20120479.
- Normand S, Zimmermann NE, Schurr FM, Lischke H (2014) Demography as the basis for understanding and predicting range dynamics. *Ecography*, **37**, 1149–1154.
- Pagel J, Schurr FM (2012) Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. *Global Ecology and Biogeography*, **21**, 293–304.
- Pearson RG, Stanton JC, Shoemaker KT *et al.* (2014) Life history and spatial traits predict extinction risk due to climate change. *Nature Climate Change*, **4**, 217–221.
- Pereira HM, Leadley PW, Proença V *et al.* (2010) Scenarios for global biodiversity in the 21st century. *Science*, **330**, 1496–1501.
- R Development Core Team (2014) *R: a language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. Available at: <https://www.R-project.org/>. (accessed 20 June 2014).
- Schurr FM, Pagel J, Cabral JS *et al.* (2012) How to understand species' niches and range dynamics: a demographic research agenda for biogeography. *Journal of Biogeography*, **39**, 2146–2162.
- Sexton JP, Mcintyre PJ, Angert AL, Rice KJ (2009) Evolution and ecology of species range limits. *Annual Review of Ecology, Evolution and Systematics*, **40**, 415–436.
- Svenning J-C, Skov F (2004) Limited filling of the potential range in European tree species. *Ecology Letters*, **7**, 565–573.
- Svenning J-C, Normand S, Skov F (2008) Postglacial dispersal limitation of widespread forest plant species in nemoral Europe. *Ecography*, **31**, 316–326.
- Swab RM, Regan HM, Matthies D, Becker U, Bruun HH (2015) The role of demography, intra-species variation, and species distribution models in species' projections under climate change. *Ecography*, **38**, 221–230.
- Thuiller W, Albert C, Araújo MB *et al.* (2008) Predicting global change impacts on plant species' distributions: future challenges. *Perspectives in Plant Ecology, Evolution and Systematics*, **9**, 137–152.
- Thuiller W, Lafourcade B, Engler R, Araújo MB (2009) BIOMOD - a platform for ensemble forecasting of species distributions. *Ecography*, **32**, 369–373.
- Thuiller W, Albert CH, Dubuis A, Randin C, Guisan A (2010) Variation in habitat suitability does not always relate to variation in species' plant functional traits. *Biology Letters*, **6**, 120–123.
- Thuiller W, Münkemüller T, Lavergne S, Mouillot D, Mouquet N, Schiffers K, Gravel D (2013) A road map for integrating eco-evolutionary processes into biodiversity models. *Ecology Letters*, **16**, 94–105.
- Thuiller W, Münkemüller T, Schiffers KH *et al.* (2014) Does probability of occurrence relate to population dynamics? *Ecography*, **37**, 1155–1166.
- Thuiller W, Pollock LJ, Gueguen M, Münkemüller T (2015) From species distributions to meta-communities. *Ecology Letters*, **18**, 1321–1328.
- Williams JW, Jackson ST (2007) Novel climates, no-analog communities, and ecological surprises. *Frontiers in Ecology and the Environment*, **5**, 475–482.
- Zurell D, Jeltsch F, Dormann CF, Schröder B (2009) Static species distribution models in dynamically changing systems: how good can predictions really be? *Ecography*, **32**, 733–744.
- Zurell D, Berger U, Cabral JS *et al.* (2010) The virtual ecologist approach: simulating data and observers. *Oikos*, **119**, 622–635.
- Zurell D, Elith J, Schröder B (2012a) Predicting to new environments: tools for visualising model behaviour and impacts on mapped distributions. *Diversity and Distributions*, **18**, 628–634.
- Zurell D, Grimm V, Rossmannith E, Zbinden N, Zimmermann NE, Schröder B (2012b) Uncertainty in predictions of range dynamics: black grouse climbing the Swiss Alps. *Ecography*, **35**, 590–603.

Supporting Information

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

Appendix S1. Range models descriptions.

Appendix S2. Details and analysis of virtual community model.

Appendix S3. Model comparison under climate change.

Appendix S4. Structural uncertainty in range dynamic models.

Table S1. IBM parameters

Figure S1. Relationship between habitat suitability and MigClim invasion probability.

Figure S2. Overview of IBM scheduling.

Figure S3. Observed abundance for different scenarios.

Figure S4. Observed abundance along niche axes for different scenarios.

Figure S5. Boxplots of observed local abundance distribution for focal species in different scenarios.

Figure S6. Observed temperature niche of focal species before and after climate change.

Figure S7. Predicted temperature niche of focal species approximated by maximum abundances along temperature gradient.

Figure S8. Predicted temperature niche of focal species approximated by mean abundances along temperature gradient.

Figure S9. Time series of sensitivity obtained for SDMs for different scenarios.

Figure S10. Time series of specificity obtained for SDMs for different scenarios.

Figures S11–S22. Maps of observed and predicted abundances before and after climate change for different scenarios.

Figures S23–S29. Time series of prediction accuracy (TSS, Spearman's Rho), range margins, total and relative abundance.

Figure S30. Maps of predicted colonization by different MigClim configurations.

Figure S31. True skill statistic for different MigClim configurations.

Figure S32. Spearman's Rho for different DemoNiche configurations.

Figure S33. Total abundance predicted for different DemoNiche configurations.

Figure S34. Relative abundance predicted for different DemoNiche configurations.

Figure S35. Spearman's Rho for different DRM configurations.

Figure S36. Total abundance predicted for different DRM configurations.