

Supporting Information for:

Benchmarking novel approaches for modelling species range dynamics

Damaris Zurell*, Wilfried Thuiller, Jörn Pagel, Juliano S. Cabral, Tamara Münkemüller, Dominique Gravel, Stefan Dullinger, Signe Normand, Katja H. Schiffers, Kara A. Moore and Niklaus E. Zimmermann

* Corresponding author, Email: damaris.zurell@wsl.ch

Contents

For better overview, we provide in-text tables and figures within single Appendix chapters.

Supporting information:	Pages
Appendix S1: Range models descriptions	3
Appendix S2: Details and analysis of virtual community model	16
Appendix S3: Model comparison under climate change	24
Appendix S4: Structural uncertainty in range dynamic models	46
Tables:	
Table S1. IBM parameters	19
Figures:	
Figure S1. Relationship between habitat suitability and MigClim invasion probability.	7
Figure S2. Overview of IBM scheduling.	17
Figure S3. Observed abundance for different scenarios.	21
Figure S4. Observed abundance along niche axes for different scenarios.	22
Figure S5. Boxplots of observed local abundance distribution for focal species in different scenarios.	22
Figure S6. Observed temperature niche of focal species before and after climate change.	23

Figure S7. Predicted temperature niche of focal species approximated by maximum abundances along temperature gradient.	25
Figure S8. Predicted temperature niche of focal species approximated by mean abundances along temperature gradient.	26
Figure S9. Time series of sensitivity obtained for SDMs for different scenarios.	26
Figure S10. Time series of specificity obtained for SDMs for different scenarios.	27
Figures S11-S22. Maps of observed and predicted abundances before and after climate change for different scenarios.	27-34
Figures S23-S29. Time series of prediction accuracy (TSS, Spearman's Rho), range margins, total and relative abundance.	39-45
Figure S30. Maps of predicted colonization by different MigClim configurations.	47
Figure S31. True skill statistic for different MigClim configurations.	48
Figure S32. Spearman's Rho for different DemoNiche configurations.	50
Figure S33. Total abundance predicted for different DemoNiche configurations.	51
Figure S34. Relative abundance predicted for different DemoNiche configurations.	52
Figure S35. Spearman's Rho for different DRM configurations.	53
Figure S36. Total abundance predicted for different DRM configurations.	54

Appendix S1. Range models descriptions

In the following, we provide more detailed descriptions for the different range models and their specific implementations and calibrations.

Appendix S1:	Pages
Species distribution models - SDM	4
Dispersal model - MigClim	5
Matrix population model - DemoNiche	7
Maximum log-likelihood population model – LoLiPop	9
Dynamic range models - DRM	11
Figures:	
Figure S1. Relationship between habitat suitability and MigClim invasion probability	7

Species distribution models - SDM

FULL NAME OF METHOD/ ABBREVIATION:

Ensemble species distribution models (SDMs)

SOFTWARE USED:

R package biomod2 (Thuiller *et al.* 2014a)

KEY REFERENCES:

Guisan and Zimmermann (2000), Guisan and Thuiller (2005), Thuiller *et al.* (2009)

EXAMPLES OF IMPLEMENTATION FOR PREDICTION OF RANGE DYNAMICS:

Pearman *et al.* (2008), Zimmermann *et al.* (2009), Thuiller *et al.* (2011)

BRIEF DESCRIPTION:

Biomod2 is an R package that allows multiple calibrations of a large range of statistical species distribution models (SDMs), to test them on the same ground and to combine them in an ensemble forecast using different approaches. Biomod2 takes presence-absence of species and associated environmental information on the same sampled points. Data are generally split into two sub-datasets, one for calibration and one for evaluation. Selected models are calibrated on the calibration data and further evaluated on the evaluation data using predictive performance measure like the area under the receiver operating characteristic curve (AUC) or the True Skill Statistics. The split sampling procedure may be repeated several times. Models' predictions maybe then be combined in various ways, after discounting the models' prediction with a poor fit. Simple averaging, weighted averaging, or committee averaging are amongst the possible ways of combining the outputs.

DATA INPUT:

Presence-absence data of species and predictor variables

OUTPUT:

ensemble forecast (based on simple averaging) of probability of occurrence and corresponding binary prediction (based on TSS-maximising threshold)

SETTINGS:

Eight different SDM were conjointly run on the same data with the following settings:

- Generalised linear model (GLM). Predictors were included as linear and quadratic predictors. Stepwise selection based on AIC was selected.
- Generalised additive model (GAM). Degree of smoothing was automatically selected through generalised cross-validation (library mgcv). No variable selection but variables with low influence have no importance in the model fit.
- Multiple adaptive regression spline (MARS). The maximum interaction degree, which determines whether interactions between variables are fitted or not, was set to 2.
- Classification Tree Analysis (CTA). The length of the tree was selected through a 5-fold cross-validation.
- Artificial Neural Networks: The learning rate and decay were selected using a 5-fold cross-validation.
- Flexible Discriminant Analysis (FDA): FDA is a more flexible approach of linear discriminant analysis in which the linear predictors are replaced by MARS.
- Random Forest (RF). We set the maximum number of trees at 500. The rest was the defaults of the random forest package.
- MAXENT: Default parameterisation (Phillips *et al.* 2006).

Each model was calibrated on a random sample of 80% of the initial data and evaluated on the 20% remaining using the True Skill Statistics (Allouche *et al.* 2006). The split-sampling procedure was repeated 10 times, generating 80 different predictions for scenario. Before

combining the prediction, we discounted models' run with a TSS lower than 0.5. Final models' predictions were then averaged.

RELATIVE COMPUTATION TIME FOR VIRTUAL BENCHMARK DATA:

Few minutes for each species and each scenario on super-computers.

RESPONSIBLE MODELLER/CONTACT:

Wilfried Thuiller, Email: wilfried.thuiller@ujf-grenoble.fr

SDM hybrids

MigClim

FULL NAME OF METHOD/ ABBREVIATION:

MigClim

SOFTWARE USED:

R package MigClim (Engler *et al.* 2013)

KEY REFERENCES:

Engler and Guisan (2009), Engler *et al.* (2012)

EXAMPLES OF IMPLEMENTATION FOR PREDICTION OF RANGE DYNAMICS:

Normand *et al.* (2013)

BRIEF DESCRIPTION:

MigClim is a cellular automaton to simulate plant migration for scenarios of environmental change, and predict dispersal-constrained projected species distribution. It couples predicted distribution maps obtained from SDMs with a dispersal kernel and simulates colonization of suitable cells. MigClim requires only a limited number of input parameters that describe the dispersal kernel, the frequency of long-distance dispersal and the propagule production potential (Engler & Guisan 2009; Engler *et al.* 2012). MigClim runs either with binary maps predicted by SDMs or probabilistic, habitat suitability maps from SDMs. In case of binary maps, all potentially suitable cells are interpreted to be equally susceptible for colonization. In case of probabilistic maps (and if reclassification threshold is zero), habitat suitability values are interpreted as the cells' susceptibility of getting colonized (invasion probability). For this, habitat suitability should be rescaled to range between 0 and 1 (respectively 0 and 1000 in BIOMOD annotation). The probability of an unoccupied cell x for becoming occupied is then the product of the dispersal probability from occupied cell y to cell x , the propagule production potential of cell y and the invasion probability of cell x .

DATA INPUT:

- Species initial distribution: we used the predicted presence/absence map from SDMs for the year 0.
- Habitat suitability maps: biomod2 output (predicted habitat suitability and predicted presence/absence) was used as input to MigClim.

OUTPUT:

MigClim outputs a presence/absence map of colonized cells. Here, we ran a number of replicate simulations and averaged the resulting maps. We, thus, obtained a probability map showing the number of times the cell was colonized in replicate simulations.

SETTINGS:

We used four different model configurations that differ in the way that habitat suitability is rescaled and related to invasion probability (see Fig. S5):

1. Configuration T (threshold): here, we used the TSS-maximising threshold (from biomod2) to convert SDM derived habitat suitability values to binary maps. Subsequently, all suitable cells had the same invasion probability (given dispersal constraints).
2. Configuration L (linear): here, the SDM derived habitat suitability values were rescaled to range between 0 and 1, and we then assumed a linear relationship between rescaled habitat suitability and a cell's invasion probability.
3. Configuration LT (linear-threshold): here, all cells with habitat suitability less than the TSS-maximising threshold (from biomod2) were set to 0 and all other habitat suitability values were rescaled to values between 0 and 1. We then assumed a linear relationship between thresholded and rescaled habitat suitability and a cell's invasion probability.
4. Configuration S (sigmoidal): here, the SDM derived habitat suitability values were rescaled to range between 0 and 1, and we then assumed a sigmoidal relationship between rescaled habitat suitability and a cell's invasion probability.

In the model intercomparison in the main article, we used the sigmoidal relationship between habitat suitability and invasion probability. More detailed results on structural uncertainty in MigClim are given in Appendix S4.

The parameter values were the following:

- Reclassification threshold: for T and LT configurations, the TSS-maximising threshold obtained from biomod2 was used to convert habitat suitability to presence/absence predictions.
- Environmental change steps: equivalent to the virtual simulation model, we simulated 100 years/time steps of environmental change.
- Dispersal time steps: within each year/time step, only one dispersal step occurred.
- One-dimensional dispersal kernel: we used the known exponential dispersal kernel that was used to simulate the virtual species
- Propagule production potential: initial maturity age was 1 year, and all age classes have the same probability of propagule production of 1.0.
- Barriers: we assumed no barriers to dispersal.
- Long-distance dispersal LDD: no LDD was included.
- Replicates: We ran 10 replicate simulations for each scenario and configuration.

RELATIVE COMPUTATION TIME FOR VIRTUAL BENCHMARK DATA:

3-6 minutes for 10 replicate simulations on a 2.6 GHz Intel Core i5 machine with 16 GB RAM.

RESPONSIBLE MODELLER/CONTACT:

Damaris Zurell, Email: damaris.zurell@wsl.ch

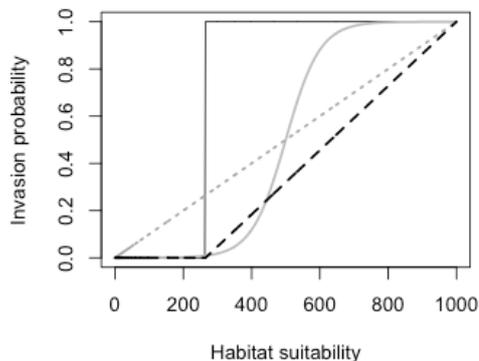


Figure S1. Relationship between habitat suitability and MigClim invasion probability for the single species short-distance dispersal scenario. The threshold relationship T is shown as black solid line, linear relationship L as grey dotted line, linear-threshold relationship LT as black dashed line, and sigmoidal relationship S as grey solid line.

DemoNiche

FULL NAME OF METHOD/ ABBREVIATION:

DemoNiche - coupled matrix population model

SOFTWARE USED:

R package *demoniche* (Nenzen 2013)

KEY REFERENCES:

Nenzén *et al.* (2012)

EXAMPLES OF IMPLEMENTATION FOR PREDICTION OF RANGE DYNAMICS:

Nenzén *et al.* (2012); other studies using matrix population models are Keith *et al.* (2008) and Anderson *et al.* (2009) who use RAMAS-GIS; and Dullinger *et al.* (2012) who use CATS

BRIEF DESCRIPTION:

DemoNiche is a spatially explicit demographic model in which local population dynamics are simulated by a matrix population model and constrained by SDM-derived habitat suitability, and local populations are connected by dispersal. DemoNiche requires the definition of a stage matrix that contains the mean survival rates and fecundities for each stage class. For this, the population of the focal species was structured into two stage classes, juveniles (here: seeds/seedlings) and adults (see Fig. 1 in main article). We assume a single adult stage class. The stage matrix thus contained 3 vital rates: the probability of juveniles surviving their first year to become adults P_{JA} , the survival probability of adults P_{AA} , and the fertility of adults F_{AJ} . Seeds were able to disperse according to the pre-defined dispersal kernel. Additionally, each cell is characterised by its carrying capacity K and a ceiling effect of K is assumed such that in local populations that exceed K , population size will be reduced to the level of K .

DemoNiche may be linked to habitat suitability in different ways. In the simplest case, habitat suitability together with a reclassification threshold can be used to derive patch-matrix maps for the metapopulation model. This corresponds to the original proposition of habitat-based, spatially explicit population models by Akçakaya (2000). Alternatively, habitat suitability can also be used to scale the carrying capacity K (Keith *et al.* 2008; Anderson *et al.* 2009; Zurell *et al.* 2012) or to scale the vital rates (Dullinger *et al.* 2012; Nenzén *et al.* 2012). Care should be taken when choosing between these options because, hitherto, there is no strong empirical evidence how habitat suitability relates to demographic rates or to carrying capacity (Thuiller *et al.* 2014b). Moreover, it is mostly not an easy task to obtain adequate demographic data for

constructing stage matrices (Dullinger *et al.* 2012). When using habitat suitability to scale the vital rates in the transition matrices, the definition of this transition matrix is even more complicated. Usually, vital rates are derived from long-term inventories or mark-recapture studies (Nenzén *et al.* 2012) and thus correspond to the long-term mean demographic rates. However, when scaling any of the stage matrix' transition probabilities with habitat suitability, the mean demographic rates get skewed because we make the implicit assumption that vital rates are at optimum levels at places with high habitat suitability although we usually only estimate mean instead of maximum demographic rates from the data. To overcome this problem, we here suggest to use the mean demographic rates derived from data not as a direct input to the stage matrix but to use it for calibrating the stage matrix such that the simulated mean demographic rates outputted from the matrix population model approximates the observed mean demographic rates. Here, we used a box-constraint quasi-Newton method to minimize the differences between simulated and observed mean demographic rates (using function *optim* with method 'L-BFGS-B' in base package *stats*; R Development Core Team 2014).

DATA INPUT:

- Habitat suitability maps: biomod2 output (predicted habitat suitability and predicted presence/absence) was used as input to DemoNiche.
- Mean demographic rates were estimated from 20-year time series sampled from the virtual simulation model before climate change set in. In every year, we recorded the probability of juveniles surviving to become adults $P_{J,A}$ as the ratio between the number of one-year old adults of the current year and the number of juveniles at the end of last year. The survival probability of adults $P_{A,A}$ was simply the fraction of last year adults that survived until the current year. The fertility of adults $F_{A,J}$ was the number of juveniles produced in the current year divided by the number of adults that survived the current year. From these time series, we estimated the means and the standard deviations of these vital rates.

OUTPUT:

Mean abundance over replicate simulations.

SETTINGS:

We used 13 different model configurations to test for effects of structural uncertainty:

1. SDM derived habitat suitability was converted to binary maps using the TSS-maximising threshold. This served as patch-matrix map for DemoNiche and local populations could only survive in suitable patches.
2. SDM derived habitat suitability indices (HSI) were used to scale the per cell carrying capacity K . Thereby we tested three different functional relationships between HSI and K : linear, linear-threshold, and sigmoidal. These relationships were defined in the same way as in MigClim (see above; Fig. S5).
3. HSI were used to scale vital rates of the stage matrix. Again, we used three different functional relationships between HSI and K (linear, linear-threshold, sigmoidal). Also we tested for differences in model performance depending on the choice of which vital rates are related to HSI.
 - a. Relate fertility of adults $F_{A,J}$ to HSI: Given the system knowledge that environment mainly determines the seed production and seedling recruitment phase in our virtual species, the ecologically most plausible choice is to relate fertility of adults $F_{A,J}$ to HSI.
 - b. Relate all vital rates to HSI: In this setting, we assume that environment determines also the survival rates of juveniles and adults, which is not the case in

the virtual simulation model. Hence, this setting should lead to underestimation of persistence rates and, thus, to overestimation of extinction risk.

- c. Relate survival rates $P_{J,A}$ and $P_{A,A}$ to HSI: This is the ecologically least realistic setting given prior system knowledge. As in the previous point (b), this setting should lead to underestimation of persistence rates. At the same time, reproduction is not constrained by HSI, which may lead to increased dispersal rates compared to (a-b).

In the model intercomparison presented in the main article, we used a consensus approach. For this, we first evaluated model performance for the 13 different configurations using Spearman's rank correlation coefficient ρ between observed and predicted abundances for the calibration/sampling data. Then, we averaged predictions of the best 5 models. More detailed results on the effect of structural uncertainty in DemoNiche are provided in Appendix S4.

Other important parameter values were set as follows:

- Carrying capacity K : if not scaled by HSI, K was set to 100 individuals per cell.
- Dispersal: the proportion of short-distance dispersal (in DemoNiche defined as spread to direct neighbours) was set to zero. For long-distance dispersal, we used the known exponential dispersal kernel that was used to simulate the virtual species.
- Environmental stochasticity: all stages of the transition matrix were affected by environmental stochasticity. Coefficient of variation in vital rates was set to 0.1 for all elements of the stage matrix.
- Replicates: We ran 10 replicate simulations for each scenario and configuration (effects of stochasticity were overall low).
- Number of years: we let the model spin up for 900 years using the habitat suitability map of year 0. Then, the model was run for the 100 years of climate change.

RELATIVE COMPUTATION TIME FOR VIRTUAL BENCHMARK DATA:

15-30 minutes for a standard model run with 10 replicate simulations on a 2.6 GHz Intel Core i5 machine with 16 GB RAM; with optimisation of transition matrix approx. 20-60 hours on Linux cluster 2.8GHz with 2 GB RAM.

RESPONSIBLE MODELLER/CONTACT:

Damaris Zurell, Email: damaris.zurell@wsl.ch

LoLiPop

FULL NAME OF METHOD/ ABBREVIATION:

Maximum log-likelihood population model for range dynamics (LoLiPop)

SOFTWARE USED:

Own R codes

KEY REFERENCES:

Cabral and Schurr (2010)

EXAMPLES OF IMPLEMENTATION FOR PREDICTION OF RANGE DYNAMICS:

Cabral *et al.* (2011); Cabral *et al.* (2013)

BRIEF DESCRIPTION:

A spatially-explicit stochastic demographic model for range dynamics of single plant species that uses habitat models to describe suitable cells. The approach is hybrid because it combines mechanistic (demographic models) and phenomenological (habitat models) niche components. The demographic parameters (maximum reproductive rate, individual mortality probability, local population extinction, Allee critical point and carrying capacity) can be combined with an

observation model (parameters: clumping index and per-individual observation probability) to be fitted to abundance distributional data via pseudo-maximum log-likelihood:

We estimated demographic and observation parameters by searching the parameter set β for which a vector of observed abundances \mathbf{A} has the lowest negative log-likelihood, calculated as

$$-\ln L(\mathbf{A}|\beta) = \sum_{i=1}^{N_{\text{cells}}} \ln L(A_i|\beta)$$

where N_{cells} is the total number of cells and $L(A_i|\beta)$ is the likelihood of observing A_i individuals at cell i . This likelihood is the probability density of the negative binomial distribution with mean $\rho \bar{N}_i$ and size parameter s (clumping index), where \bar{N}_i is the abundance at cell i averaged over 100 timesteps during the quasi-stationary state (assured after initial 900 timesteps) per replicate run and over five replicate runs. The parameter ρ is the per-individual observation probability and s describes the degree individuals are clumped in the cell (values around 1 indicate even distributed individuals, whereas lower values indicate stronger clumping). The abundance distribution \mathbf{N} at a given time step followed

$$\mathbf{N}(t+1) = S(\mathbf{N}(t)) + G(\mathbf{N}(t))$$

where $\mathbf{N}(t+1)$ and $\mathbf{N}(t)$ are vectors describing local abundances in all cells at time t and $t+1$, S is a function describing adult survival and G is a function describing dispersal and recruitment. The survival function S is a binomial random variable with binomial denominator $N_i(t)$ and success probability $1-M$, where $N_i(t)$ is local abundance in cell i and M is the per time step probability of adult mortality. The number of recruits, G , is a Poisson random variable whose mean is the expected number of offspring dispersed to each cell. For cell i , this expected number is

$$\sum_j D_{i,j} N_j(t) R(N_j(t))$$

where $D_{i,j}$ is the probability per individual of dispersal from cell j to cell i given by an input dispersal kernel \mathbf{D} and R is a function describing per capita reproduction. For R , we applied the Beverton-Holt reproductive function extended for inclusion of Allee effects (Cabral & Schurr 2010).

For a given set of demographic parameters and the simulated abundances $\bar{\mathbf{N}}$, we estimated the observation parameters by minimizing the negative log-likelihood $-\ln L$ with the Nelder-Mead simplex algorithm in R 3.0.1 (R Development Core Team, 2014). To estimate the demographic parameters, we followed the general strategy for direct parameter search (Bolker 2008): we varied demographic parameters in a multidimensional grid spanning the realistic range of values, and iteratively zoomed at least twice into the subregion of the grid with lowest negative log-likelihoods. The zooming procedure was stopped when the refined parameter grid did not produce lower negative log-likelihoods than the previous grid. Due to this direct search, we could not ascertain full maximum likelihood parameter estimates. However, the complete grid of demographic parameter values was broad and fine enough to cover ecologically important variation.

DATA INPUT:

- Habitat suitability maps: biomod2 output (binary 0 vs. 1, or unsuitable vs. suitable habitat, respectively).
- Abundance data from a subset of 500 cells for the year 0.
- 2-dimensional negative exponential dispersal kernels for both dispersal scenario. For the short-distance dispersal scenario ($\alpha=0.5$), the 2D kernel was 5 x 5 cells, whereas for the long-distance dispersal scenario ($\alpha=0.1$), the 2D kernel was 11 x 11 cells.

OUTPUT:

- Optimized parameter combination per scenario (maximum reproductive rate, individual mortality probability, Allee critical point and size parameter s)
- Abundance distribution N averaged over five time steps per replicate run over five replicate runs under no climate change.
- Abundance distribution N for every time step under climate change averaged over five replicate runs.

SETTINGS:

For all scenarios, we fixed carrying capacity at 100 individuals, local extinction probability at 0 and per-individual observation probability at 1 to resemble the simulated and observed data from base model used for benchmarking. We applied the respective dispersal kernel as input given the dispersal ability scenario (either short-distance or long-distance dispersal kernels). For the scenarios with auto-correlated local extinction (contagious disturbance), we extended the model to incorporate the same auto-correlation and probabilities as simulated in the base model. Therefore, the free parameters fitted to data were: maximum reproductive rate, individual mortality probability, Allee critical point and size parameter s . Initial values for reproductive rate were: 0.9, 1, 1.5, 2, 5, 10, 25, 50, 100, 200. Initial values for mortality were: 0.005, 0.01, 0.05, 0.1, 0.15, 0.2, 0.3, 0.5. Initial values for Allee critical point were: $-K$, $-0.8K$, $-0.6K$, $-0.4K$, $-0.2K$, 0, $0.04K$, $0.1K$ (where K was the carrying capacity per cell fixed at 100 individuals). The parameter s was optimized directly during maximum log-likelihood estimations using the s as the size parameter for the negative binomial distribution.

RELATIVE COMPUTATION TIME FOR VIRTUAL BENCHMARK DATA:

The computation time for a total of c. 5000000 iterations (e.g. 1000 parameter combinations, five replicate runs and 1000 time steps) ran on average for c. 15 hours (2.3 GHz base frequency/3.4 GHz Max Turbo frequency, 2nd Generation Intel Core i7 QuadCore). Maximum log-likelihood estimations for the same amount of parameter combinations would take c. 5 hours.

RESPONSIBLE MODELLER/CONTACT:

Juliano Sarmiento Cabral, Email: jscabral@gmx.de, juliano.cabral@idiv.de

Dynamic range models

FULL NAME OF METHOD/ABBREVIATION:

Dynamic range models (DRM)

SOFTWARE USED:

Own R codes

KEY REFERENCES:

Pagel and Schurr (2012)

EXAMPLES OF IMPLEMENTATION FOR PREDICTION OF RANGE DYNAMICS:

Pagel and Schurr (2012)

BRIEF DESCRIPTION:

A basic state-space model of spatial population dynamics embedded in a hierarchical Bayesian framework for the simultaneous estimation of all model parameters from observational data. Local population dynamics are described by a stochastic Ricker model with Poisson error:

$$N_t \sim \text{Poisson}(N_{t-1} \cdot \exp(r - h N_{t-1})).$$

Note that in this parameterization of the Ricker model the carrying capacity K is assumed to be proportional to r ($K = r/h$, where h describes a constant intensity of intraspecific competition in

all environments). Variation of r in response to environmental conditions is described by a probabilistic niche model:

$$r_{i,t} \sim \text{Normal}(\mu_{i,t}, \sigma_r)$$

$$\mu_{i,t} = \beta_0 \cdot \left(1 - \left(\frac{T_{i,t} - \beta_1}{\beta_2} \right)^2 - \left(\frac{X_i - \beta_3}{\beta_4} \right)^2 \right)$$

with parameters describing the growth rate under optimal conditions (β_0) and the niche optima (β_1, β_3) and relative niche widths (β_2, β_4) with respect to temperature T and the static landscape variable X .

Local populations are then linked by dispersal. The probability of dispersal from cell j to cell i in each time step is given by a spatially discrete kernel $P_{j \rightarrow i}(f_{LDD}, \alpha)$, where f_{LDD} is the fraction of individuals dispersed beyond cell j and α is the mean of a negative exponential distribution of dispersal distances for these individuals.

DATA INPUT:

- presence/absence data from a subset of 500 cells for the years -10 and 10
- time series of annual abundance data for the years -10 to 10 from a random subset of 50 cells

OUTPUT:

- Posterior samples of the model parameters and abundance states for all cells in the observation period (years -10 and 10).
- Predictions under climate change were generated by Monte Carlo simulation of the model based on the joint posterior distribution of parameter and abundance states in year 0. Predicted abundances at the end of the climate change period (year 100) were summarized as the median over 300 Monte Carlo simulations.

SETTINGS:

The Bayesian parameter estimation requires the specification of prior distribution for all parameters ($\beta_0, \beta_1, \beta_2, \beta_3, \beta_4, \sigma_r, h, f_{LDD}, \alpha$). For most parameters no available knowledge were assumed and weakly informative prior were specified:

$$\beta_0 \sim \text{Normal}(0, 100)$$

$$\beta_1 \sim \text{Normal}(0, 100)$$

$$\log(\beta_2) \sim \text{Normal}(0, 100)$$

$$\beta_3 \sim \text{Normal}(0, 100)$$

$$\log(\beta_4) \sim \text{Normal}(0, 100)$$

$$\sigma_r \sim \text{Uniform}(0, 100)$$

$$h \sim \text{InverseGamma}(10, 10)$$

For the mean dispersal distance α an informative lognormal prior was formulated with the mean given by the mean dispersal distance in the respective scenario simulation and a standard deviation of 0.1. In order to account for the discrepancy between the perennial life-cycle of the study species and the annual life-cycle as described by the Ricker model (i.e. no adult survival), the parameter f_{LDD} was fixed at a value of 0.1, thereby describing that only a fraction of population growth between time steps is realized by recruitment from dispersed propagules.

Samples from the parameter posterior distribution were generated by a Markov chain Monte Carlo algorithm. Three independent chains were run for 50000 iterations each and samples were stored during the final 10000 iterations with a thinning of 100.

RELATIVE COMPUTATION TIME FOR VIRTUAL BENCHMARK DATA:

The computation time for a total of 50000 iterations of a MCMC run was on average c. 240 hours (2.7 GHz Intel Xeon E5-2697 v2 CPU).

RESPONSIBLE MODELLER/CONTACT:

Jörn Pagel, Email: joern.pagel@uni-hohenheim.de

References

1.
Akçakaya, H.R. (2000). Viability analyses with habitat-based metapopulation models. *Population Ecology*, 42, 45-53.
2.
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.
3.
Anderson, B.J., Akçakaya, H.R., Araújo, M.B., Fordham, D.A., Martinez-Meyer, E., Thuiller, W. *et al.* (2009). Dynamics of range margins for metapopulations under climate change. *Proceedings of the Royal Society B*, 276, 1415-1420.
4.
Bolker, B.M. (2008). *Ecological Models and Data in R*. Princeton University Press.
5.
Cabral, J.S., Bond, W.J., Midgley, G.F., Rebelo, A.G., Thuiller, W. & Schurr, F.M. (2011). Effects of Harvesting Flowers from Shrubs on the Persistence and Abundance of Wild Shrub Populations at Multiple Spatial Extents. *Conservation Biology*, 25, 73-84.
6.
Cabral, J.S., Jeltsch, F., Thuiller, W., Higgins, S., Midgley, G.F., Rebelo, A.G. *et al.* (2013). Impacts of past habitat loss and future climate change on the range dynamics of South African Proteaceae. *Diversity and Distributions*, 19, 363-376.
7.
Cabral, J.S. & Schurr, F.M. (2010). Estimating demographic models for the range dynamics of plant species. *Global Ecology and Biogeography*, 19, 85-97.
8.
Dullinger, S., Gatttringer, A., Thuiller, W., Moser, D., Zimmermann, N.E., Guisan, A. *et al.* (2012). Extinction debt of high-mountain plants under twenty-first-century climate change. *Nature Climate Change*, 2, 619-622.
9.
Engler, R. & Guisan, A. (2009). MigClim: Predicting plant distribution and dispersal in a changing climate. *Diversity and Distributions*, 15, 590-601.
10.
Engler, R., Hordijk, W. & Guisan, A. (2012). The MIGCLIM R package – seamless integration of dispersal constraints into projections of species distribution models. *Ecography*, 35, 872-878.
11.
Engler, R., Hordijk, W. & Pellissier, L. (2013). MigClim: Implementing dispersal into species distribution models. R package version 1.6. <http://CRAN.R-project.org/package=MigClim>.
12.
Gravel, D., Canham, C.D., Beaudet, M. & Messier, C. (2006). Reconciling niche and neutrality: the continuum hypothesis. *Ecology Letters*, 9, 399-409.

13.

Guisan, A. & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8, 993-1009.

14.

Guisan, A. & Zimmermann, N.E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135, 147-186.

15.

Keith, D.A., Akçakaya, H.R., Thuiller, W., Midgley, G.F., Pearson, R.G., Phillips, S.J. *et al.* (2008). Predicting extinction risks under climate change: coupling stochastic population models with dynamic bioclimatic habitat models. *Biology Letters*, 4, 560-563.

16.

Münkemüller, T., de Bello, F., Meynard, C.N., Gravel, D., Lavergne, S., Mouillot, D. *et al.* (2012). From diversity indices to community assembly processes: a test with simulated data. *Ecography*, 35, 468-480.

17.

Nenzen, H. (2013). *demoniche*: *demoniche* - spatial population dynamics. R package version 1.0/r32. <http://R-Forge.R-project.org/projects/demoniche/>.

18.

Nenzén, H.K., Swab, R.M., Keith, D.A. & Araújo, M.B. (2012). *demoniche* – an R-package for simulating spatially-explicit population dynamics. *Ecography*, 35, 577-580.

19.

Normand, S., Randin, C., Ohlemüller, R., Bay, C., Høye, T.T., Kjær, E.D. *et al.* (2013). A greener Greenland? Climatic potential and long-term constraints on future expansions of trees and shrubs. *Phil Trans R Soc B*, 368, 20120479-20120479.

20.

Pagel, J. & Schurr, F.M. (2012). Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics. *Global Ecology and Biogeography*, 21, 293-304.

21.

Pearman, P.B., Randin, C.F., Broennimann, O., Vittoz, P., Knaap, W.O.v.d., Engler, R. *et al.* (2008). Prediction of plant species distributions across six millennia. *Ecology Letters*, 11, 357-369.

22.

Phillips, S.J., Anderson, R.P. & Schapire, R.E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231-259.

23.

R Development Core Team (2014). R: A Language and Environment for Statistical Computing. 24.

Thuiller, W., Georges, D. & Engler, R. (2014a). *biomod2*: Ensemble platform for species distribution modeling. R package version 3.1-64. <http://CRAN.R-project.org/package=biomod2>.

25.

Thuiller, W., Lafourcade, B., Engler, R. & Araújo, M.B. (2009). BIOMOD – a platform for ensemble forecasting of species distributions. *Ecography*, 32, 369-373.

26.

Thuiller, W., Lavergne, S., Roquet, C., Boulangéat, I., Lafourcade, B. & Miguel, B.A. (2011). Consequences of climate change on the tree of life in Europe. *Nature*, 470, 531-534.
27.

Thuiller, W., Münkemüller, T., Schiffrers, K.H., Georges, D., Dullinger, S., Eckhart, V.M. *et al.* (2014b). Does probability of occurrence relate to population dynamics? *Ecography*, 37, 1155-1166.
28.

Zimmermann, N.E., Yoccoz, N.G., Thomas C. Edwards, Jr., Meier, E.S., Thuiller, W., Guisan, A. *et al.* (2009). Climatic extremes improve predictions of spatial patterns of tree species. *PNAS*, 106, 19723-19728.
29.

Zurell, D., Grimm, V., Rossmannith, E., Zbinden, N., Zimmermann, N.E. & Schröder, B. (2012). Uncertainty in predictions of range dynamics: black grouse climbing the Swiss Alps. *Ecography*, 35, 590-603.

Appendix S2. Details and analysis of virtual community models

Here, we provide a more detailed description of the individual-based community model along with the specific IBM parameters used in the different scenarios, and an overview over the patterns produced by the IBM.

Tables:	Pages
Table S1. IBM parameters	19
Figures:	
Figure S2. Overview of IBM scheduling.	17
Figure S3. Observed abundance for different scenarios.	21
Figure S4. Observed abundance along niche axes for different scenarios.	22
Figure S5. Boxplots of observed local abundance distribution for focal species in different scenarios.	22
Figure S6. Observed temperature niche of focal species before and after climate change.	23

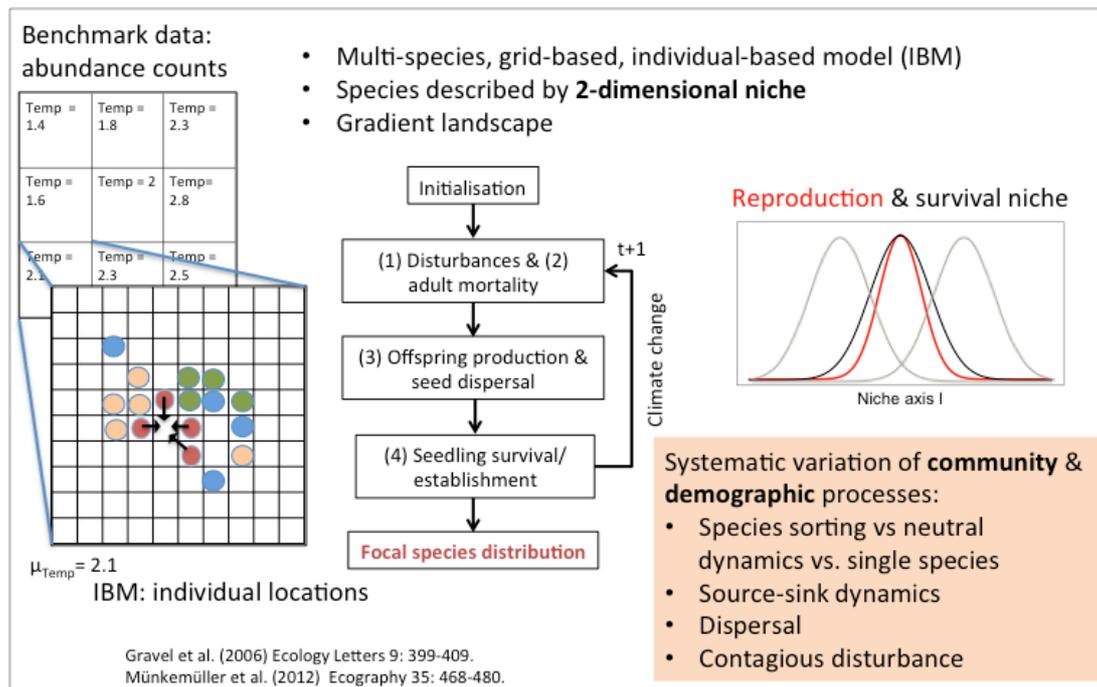


Figure S2. Overview of individual-based community model. For description see main article text.

A recently published, individual-based, spatially-explicit, stochastic model (Gravel *et al.* 2006; Münkemüller *et al.* 2012) was adapted to simulate the dynamics of individuals competing for resources in heterogeneous environments. The model is a cellular automaton and each cell was characterised by its environmental conditions (temperature and soil moisture) and could support one sessile individual from the species pool, or none. Because this spatial resolution only allows individual counts but the range models work on the population level, we implemented a hierarchy of two spatial scales so that the sessile individuals are interacting on a local scale but the model outputs community composition and species abundances at a coarser scale. To this end, a certain number of local cells (here, 10x10) that we henceforth call subcells, were aggregated to form one cell at the coarse, final resolution. One time step corresponded to one year and the main processes within one time step were (1) large-scale disturbances at the coarse resolution, and the local (subcell) processes of (2) adult mortality, (3) dispersal of propagules, and (4) recruitment of propagules to adult size (Fig. S1).

(1) Large-scale contagious disturbances like fire or wind throw acted at the coarse scale resulting in more stochastic population dynamics and incomplete range filling, and were only switched on for some scenarios. Disturbance was initiated in random cells at the coarse scale and spread to the eight nearest neighbours with a certain probability such that the overall contagious disturbance probability was 0.05. Within these coarse-scale cells, individuals (in the subcells) got killed with a probability of 0.9. Then,

(2) within each time step, adults died with a probability of 0.1 and could be replaced by recruits from the local community or by immigrants.

(3) Propagule supply of species i at a given site x was determined by the offspring rain O from parents:

$$O_{ix} = \sum f_{iy} w(d_{xy}) n_{iy} + I \quad \text{eq. 1}$$

with

$$w(d_{xy}) = \exp(-\alpha d_{xy}) \quad \text{eq. 2}$$

where f_{iy} was a scaling factor for offspring production of species i at site y , the dispersal kernel $w(d_{xy})$ was a function of Euclidian distance d between cells x and y , and n_{iy} was the incidence of species i in cell y . Only adults that were at least one year of age could produce seeds. Seed rain I (immigration) from outside the lattice was set to zero. Dispersal dynamics were simulated using a negative exponential dispersal kernel where the rate parameter α determines the inverse of the mean dispersal distance. The species-specific scaling factor for offspring production f_{iy} of species i at site y was determined by the local environment E_y , and was modelled by a Gaussian function for each axis j of the two-dimensional species niche:

$$f_{iy} = c_i \prod_{j=1}^2 \exp(-(E_{yj} - \mu_{f,ij})^2 / (2\sigma_{f,ij}^2)) \quad \text{eq. 3}$$

where c_i was the relative offspring production at the reproductive niche optimum $\mu_{f,ij}$ and $\sigma_{f,ij}$ determined the reproductive niche width. The Gaussian function was cut off at a threshold of 0.001 to obtain finite niche breadth.

(4) Recruitment follows a lottery function of dispersal-driven propagule supply and interspecific differences in competitive ability. The latter is determined by the ability of propagules to survive in the understory of adults prior to recruitment. The probability of a propagule to establish and replace the resident is proportional to the ratio between its survival probability λ_{ix} and that of the resident, or between λ_{ix} and a threshold of 0.1 in empty cells. The survival probability of species i at site x was determined by the local environment E_x , and was modelled by a Gaussian function for each axis j of the two-dimensional species niche:

$$\lambda_{ix} = h_i \prod_{j=1}^2 \exp(-(E_{xj} - \mu_{l,ij})^2 / (2\sigma_{l,ij}^2)) \quad \text{eq. 4}$$

where h_i was the species maximal performance at survival niche optimum $\mu_{l,ij}$ and $\sigma_{l,ij}$ was the niche width. There was lottery competition among the propagules that survived in the cell, and the chances of a species to establish were proportional to the relative amount of offspring rain. The cells could stay empty if the overall amount of offspring rain was too small which was simulated by taking as probability the ratio between actual number of offspring from all neighbours divided by the potential maximum number of offspring rain if all neighbours were occupied and produced offspring at optimum fecundity.

Reproductive niche optimum* μ_f of focal species for temperature	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
Reproductive niche optimum* μ_f of focal species for landscape variable X_1	0	0	0	0	0	0	0	0	0	0	0	0
Reproductive niche width* σ_f of focal species for temperature & X_1	0.4	0.4	0.4	0.32	0.6	0.6	0.6	0.48	0.6	0.6	0.6	0.48
Survival:												
Scaling factor h for performance at survival niche optimum	1	1	1	1	1	1	1	1	1	1	1	1
Survival niche optimum* μ_l of focal species for temperature	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7	0.7
Survival niche optimum* μ_l of focal species for landscape variable X_1	0	0	0	0	0	0	0	0	0	0	0	0
Survival niche width* σ_l of focal species for temperature & X_1	0.4	0.4	0.4	0.48	0.6	0.6	0.6	0.72	0.6	0.6	0.6	0.72
Distance* to niche optima of competing species	-	-	-	-	1.2	1.2	1.2	1.2	0	0	0	0

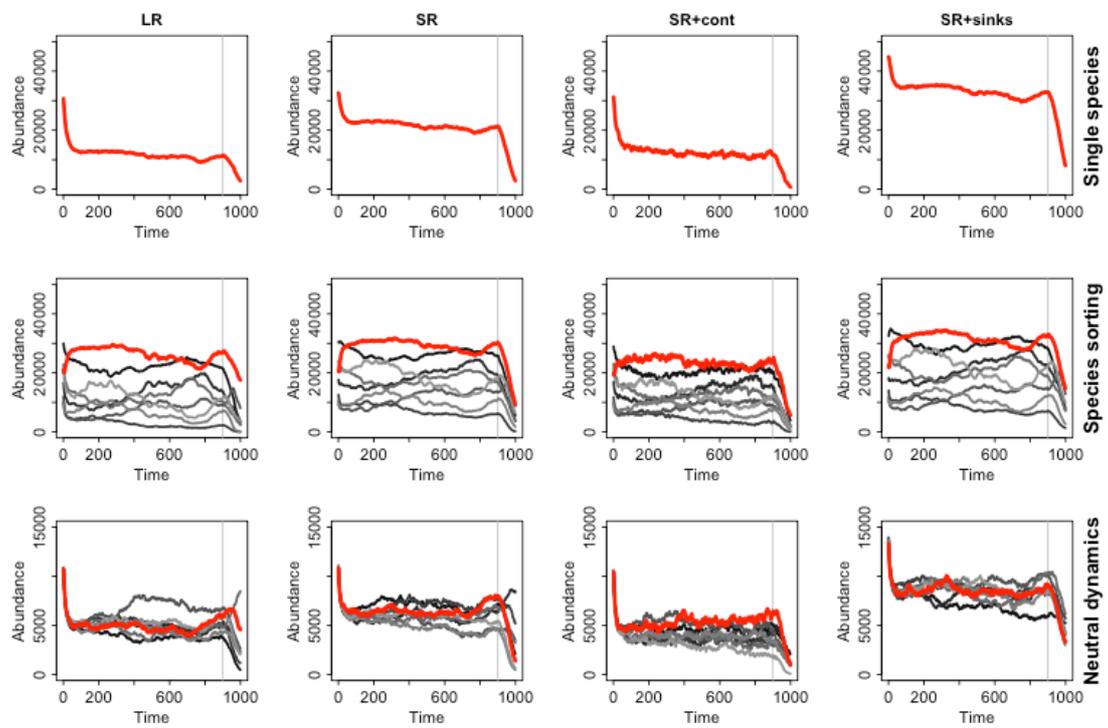


Figure S3. Observed abundance in virtual simulation model for different scenarios. Focal species is presented in red, competitor species (except in single species scenarios) in different grey shading. Model spins up for 900 years, then climate change is induced (marked by vertical grey line). LR: long range dispersal, SR: short range dispersal, cont: contagious disturbance, sinks: source sink dynamics.

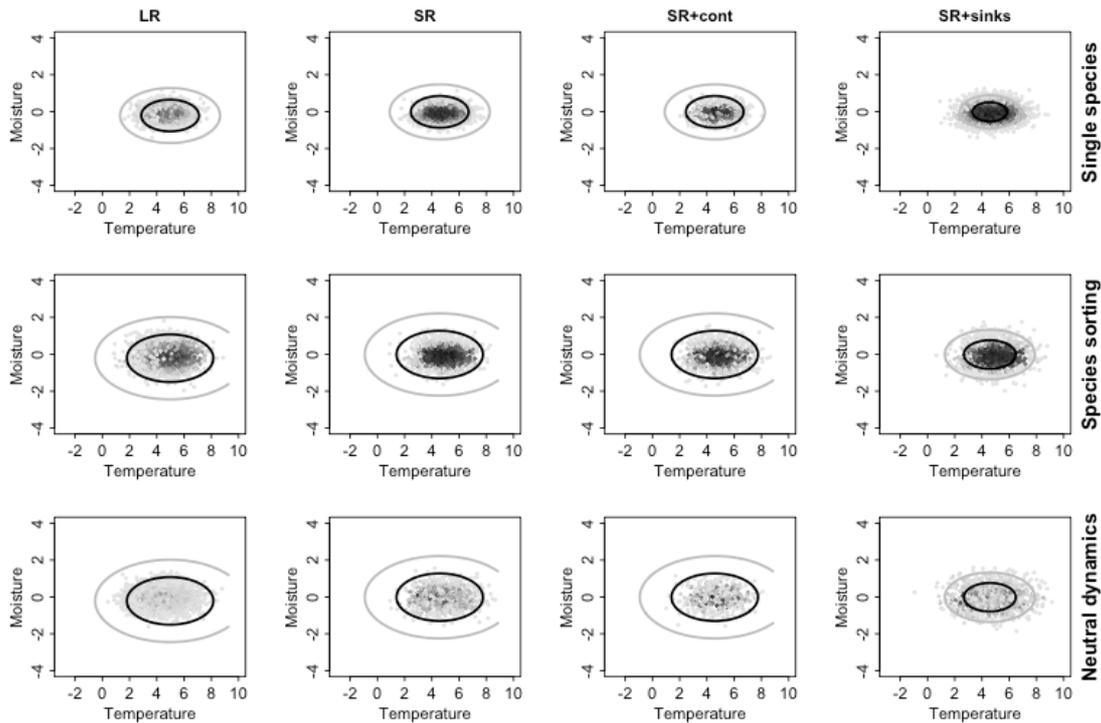


Figure S4. Observed abundance along niche axes in virtual simulation model for different scenarios. Black contour (inner circle) indicates survival probability of 0.1, which corresponds to competitive ability of empty cells. Grey contour (outer circle) corresponds to fundamental niche limit beyond which no reproduction is possible (Gaussian function cut off at reproduction rate of 0.001). LR: long-range dispersal, SR: short-range dispersal, cont: contagious disturbance, sinks: source-sink dynamics.

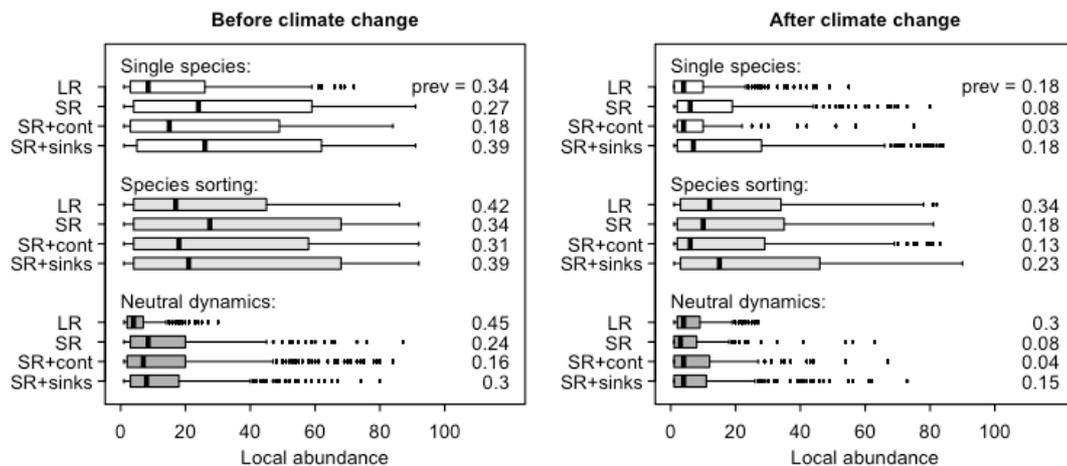


Figure S5. Focal species' abundance distribution observed from virtual simulation model for year 0 (left) and year 100 (right). Boxplots depict the distribution of abundance values over all cells and for all scenarios, together with the respective prevalence values (number of cells occupied). Maps exemplarily depict the spatial distribution of focal species' abundances for the long- and short-range variants of the species sorting and neutral dynamics scenarios. Darker colour indicates higher local abundance.

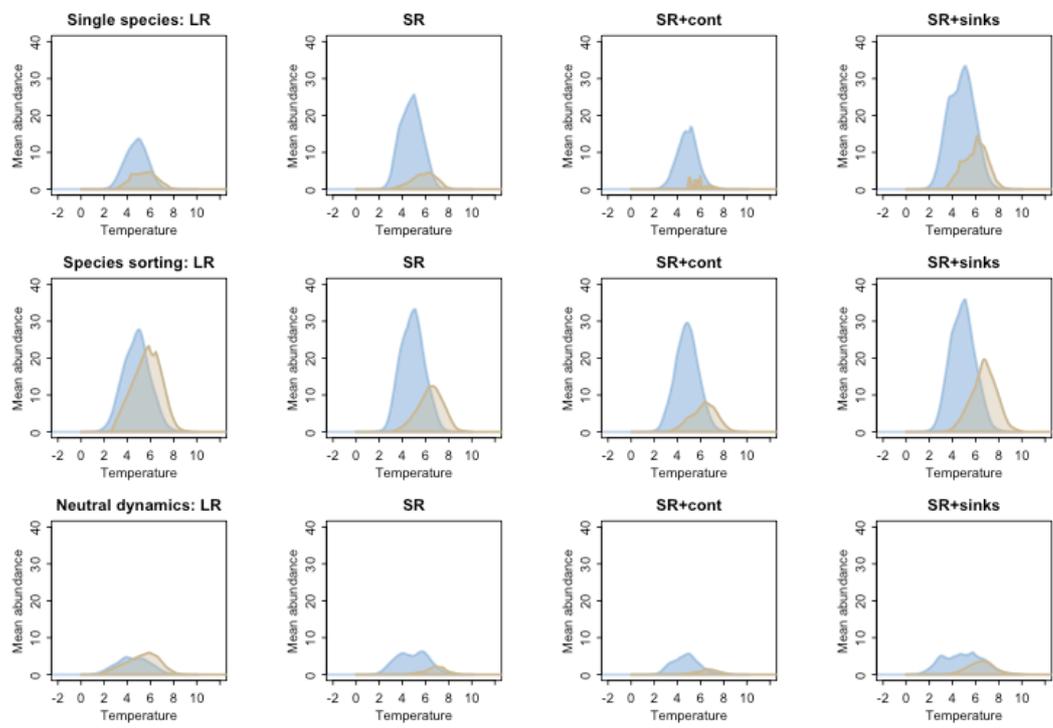


Figure S6. Temperature niche of focal species before (year 0; lightblue) and after climate change (year 100; wheat-coloured) as approximated by mean observed abundances along temperature gradient. For all scenarios, we see imprints of dispersal limitations at leading, cold edge and persistence at rear, warm edge.

Appendix S3. Model comparison under climate change

Here, we provide additional figures illustrating the comparative model performance.

Figures:	Pages
Figure S7. Predicted temperature niche of focal species approximated by maximum abundances along temperature gradient.	25
Figure S8. Predicted temperature niche of focal species approximated by mean abundances along temperature gradient.	26
Figure S9. Time series of sensitivity obtained for SDMs for different scenarios.	26
Figures S10. Time series of specificity obtained for SDMs for different scenarios.	27
Figures S11-22. Maps of observed and predicted abundances before and after climate change for different scenarios.	27-34
Figures S23-S29. Time series of prediction accuracy (TSS, Spearman's Rho), range margins, total and relative abundance.	39-45

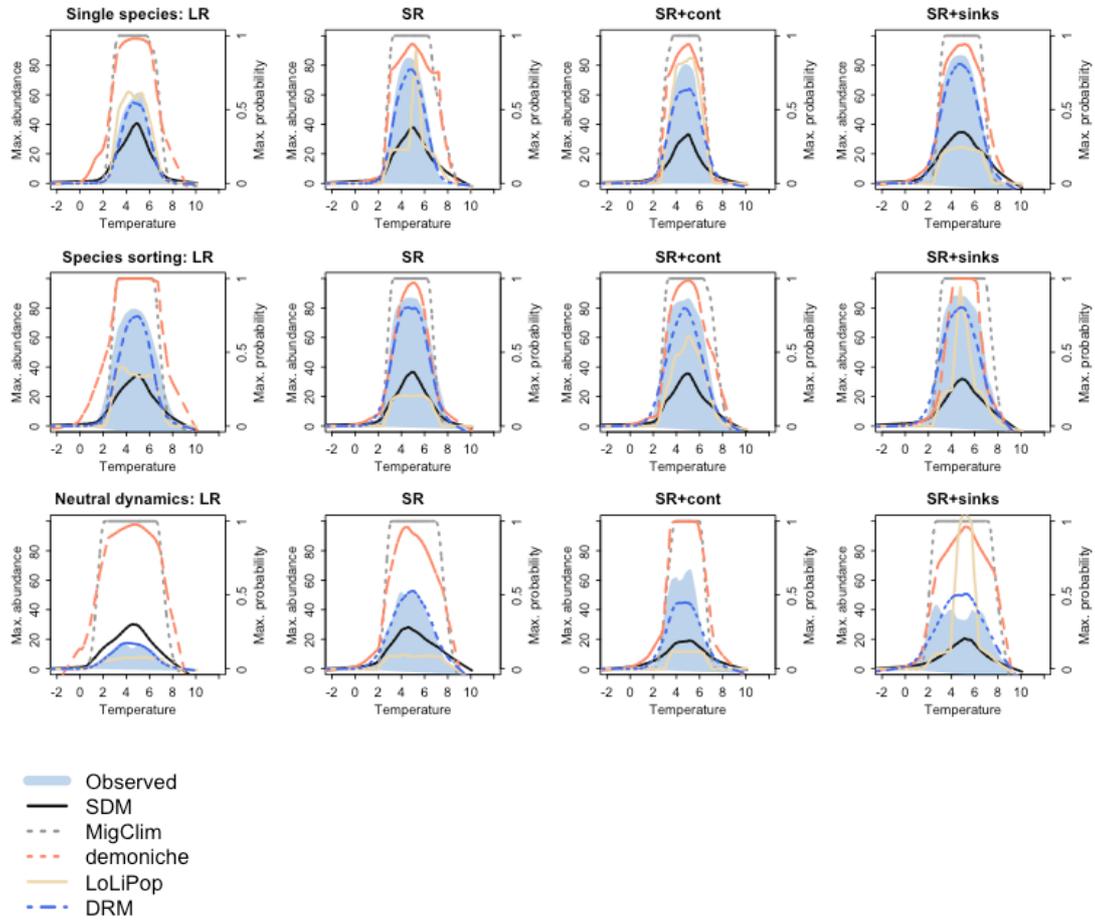


Figure S7. Observed and predicted temperature niche of focal species. We show the maximum observed abundances along temperature gradient as proxy for the realised niche and the respective predictions by the different range models.

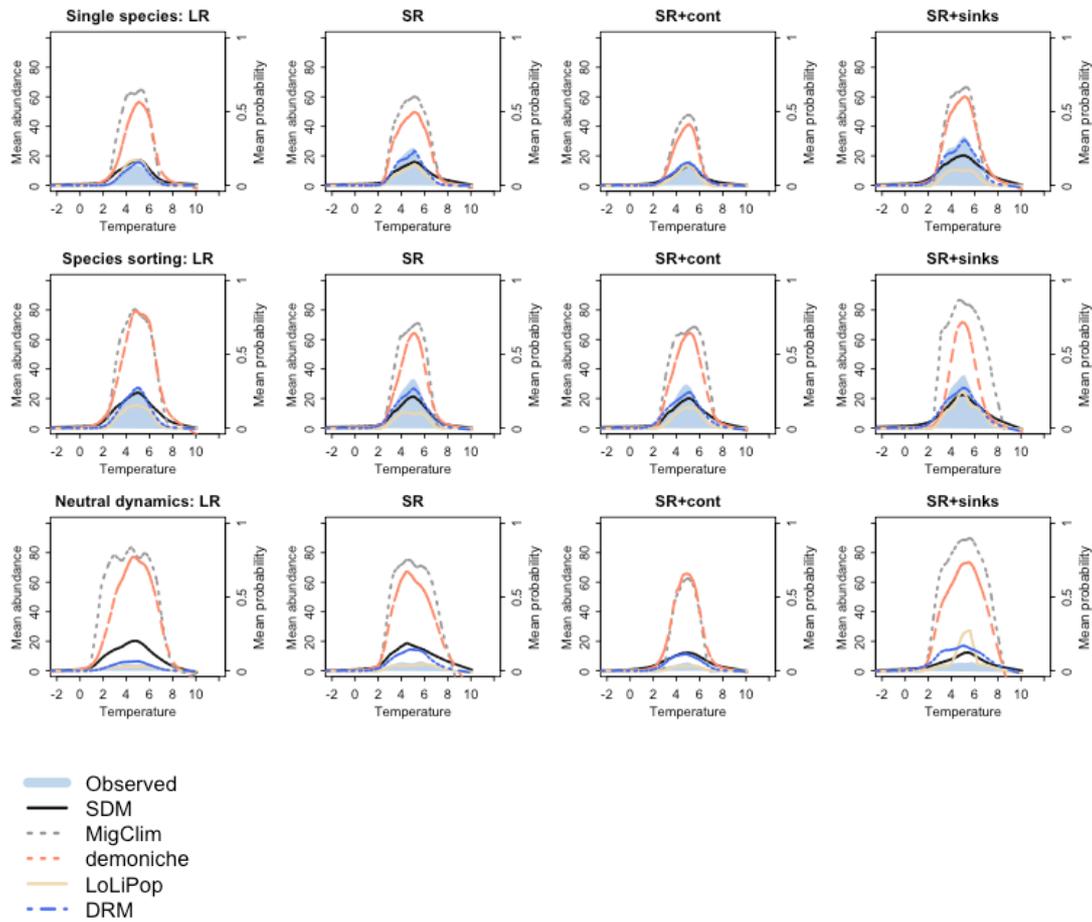


Figure S8. Observed and predicted temperature niche of focal species. We show the mean observed abundances along temperature gradient as proxy for the realised niche and the respective predictions by the different range models.

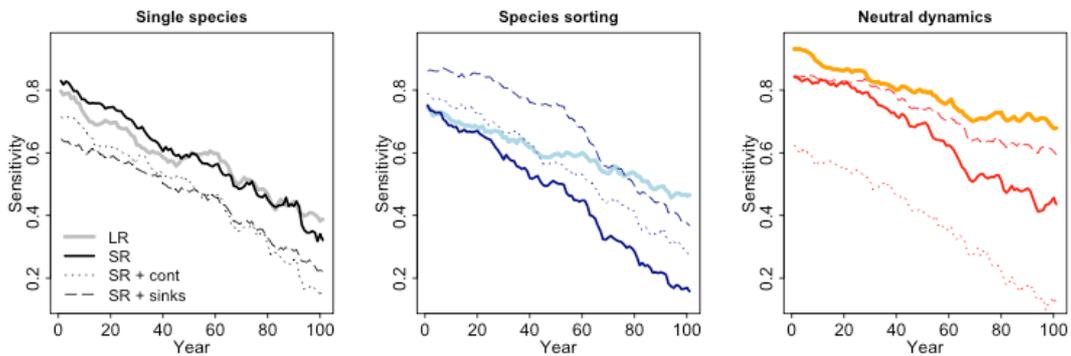


Figure S9. Time series of sensitivity obtained for SDMs for different scenarios.

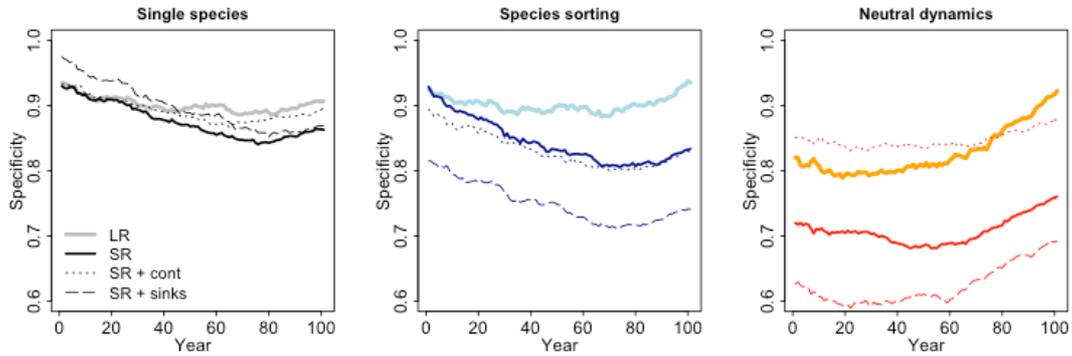


Figure S10. Time series of specificity obtained for SDMs for different scenarios.

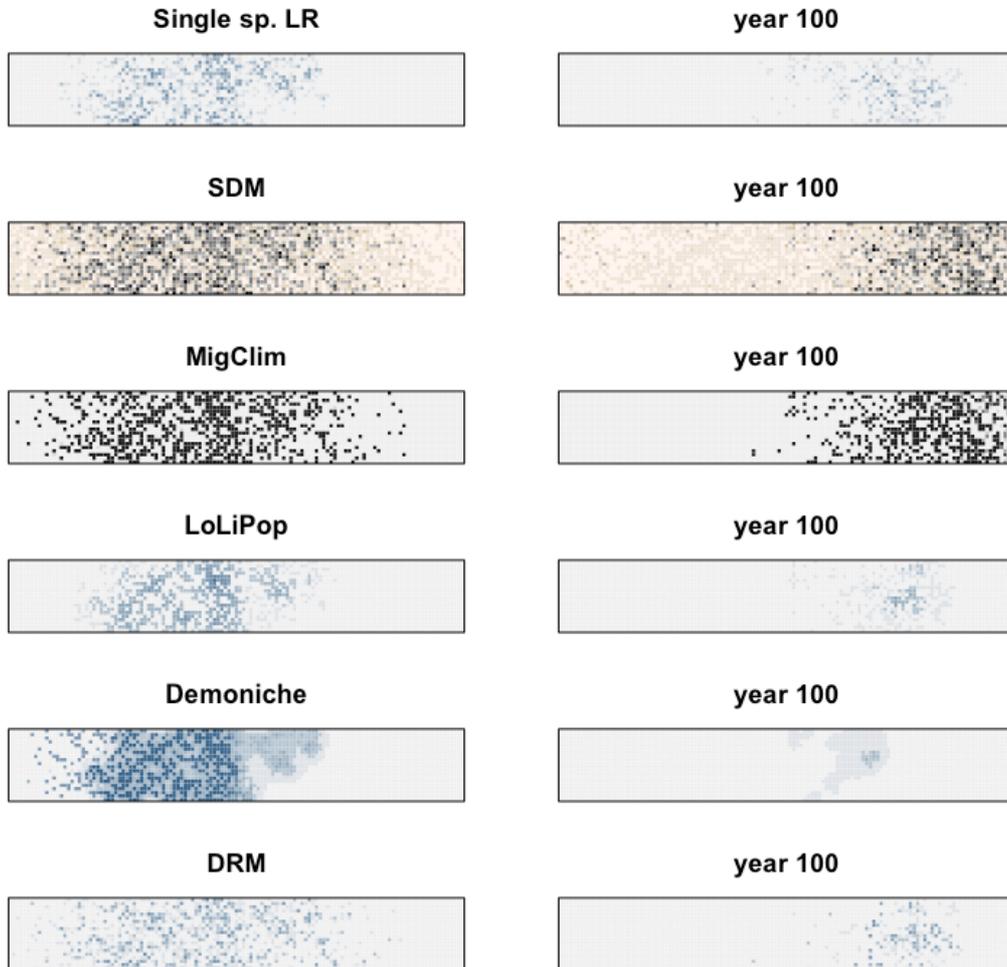


Figure S11. Observed abundances of virtual model (top row) and predictions of range models for year 0 before climate change (left) and year 100 after climate change (right) for the single species scenario with long range dispersal. Abundances are presented in blue shading with darker colour indicating 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.

presences, with darker colours indicating higher habitat suitability. MigClim predicts colonized (in black) vs. uncolonized cells. Grey shading in MigClim indicates that these cells were not colonized in all replicate runs.

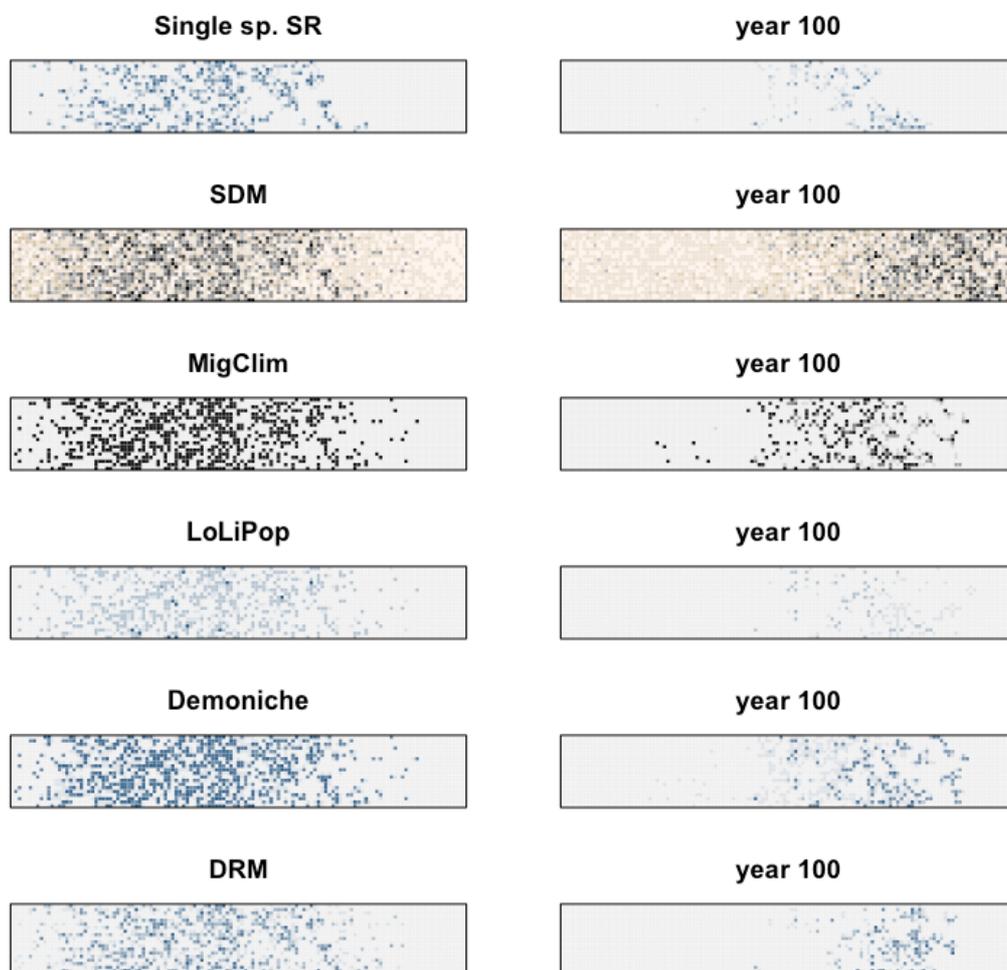


Figure S12. Observed abundances of virtual model (top row) and predictions of range models for year 0 before climate change (left) and year 100 after climate change (right) for the single species scenario with short range dispersal. Abundances are presented in blue shading with darker colour indicating 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 colonized (in black) vs. uncolonized cells. Grey shading in MigClim indicates that these cells were not colonized in all replicate runs.

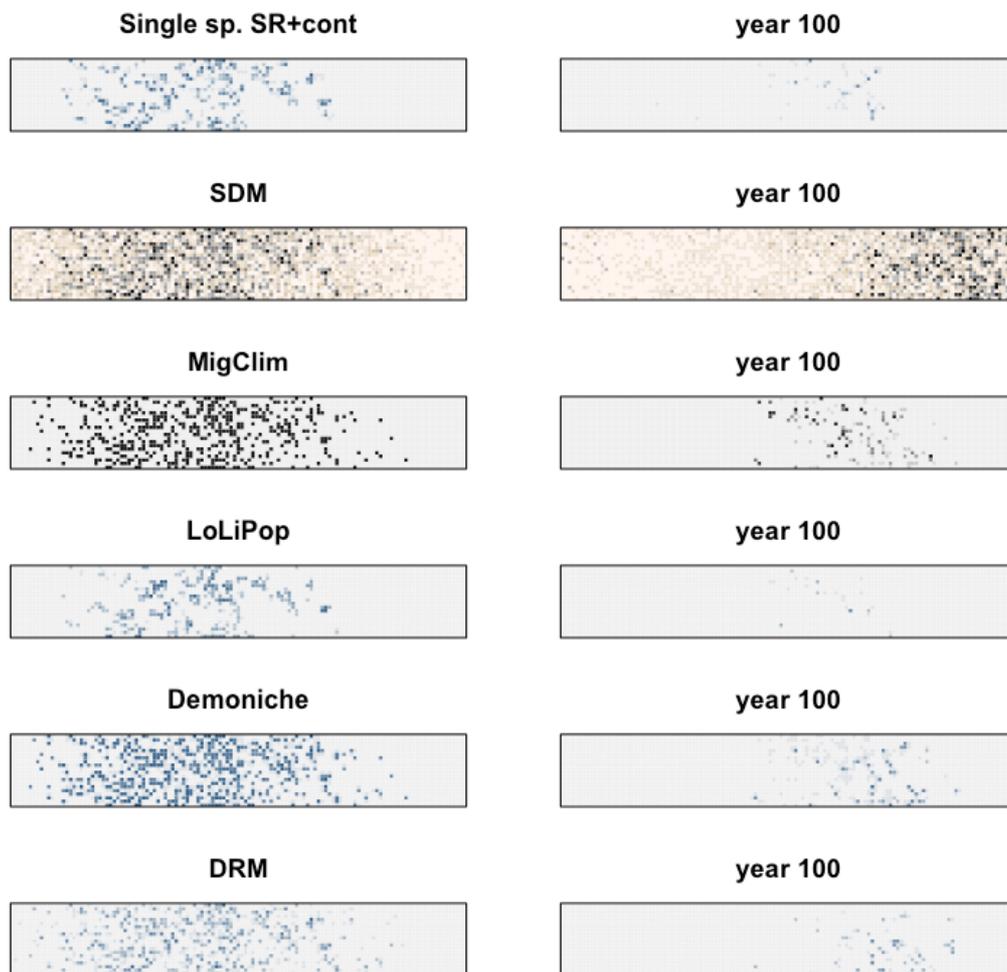


Figure S13. Observed abundances of virtual model (top row) and predictions of range models for year 0 before climate change (left) and year 100 after climate change (right) for the single species scenario with short range dispersal and contagious disturbances. Abundances are presented in blue shading 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 colonized (in black) vs. uncolonized cells. Grey shading in MigClim indicates that these cells were not colonized in all replicate runs.

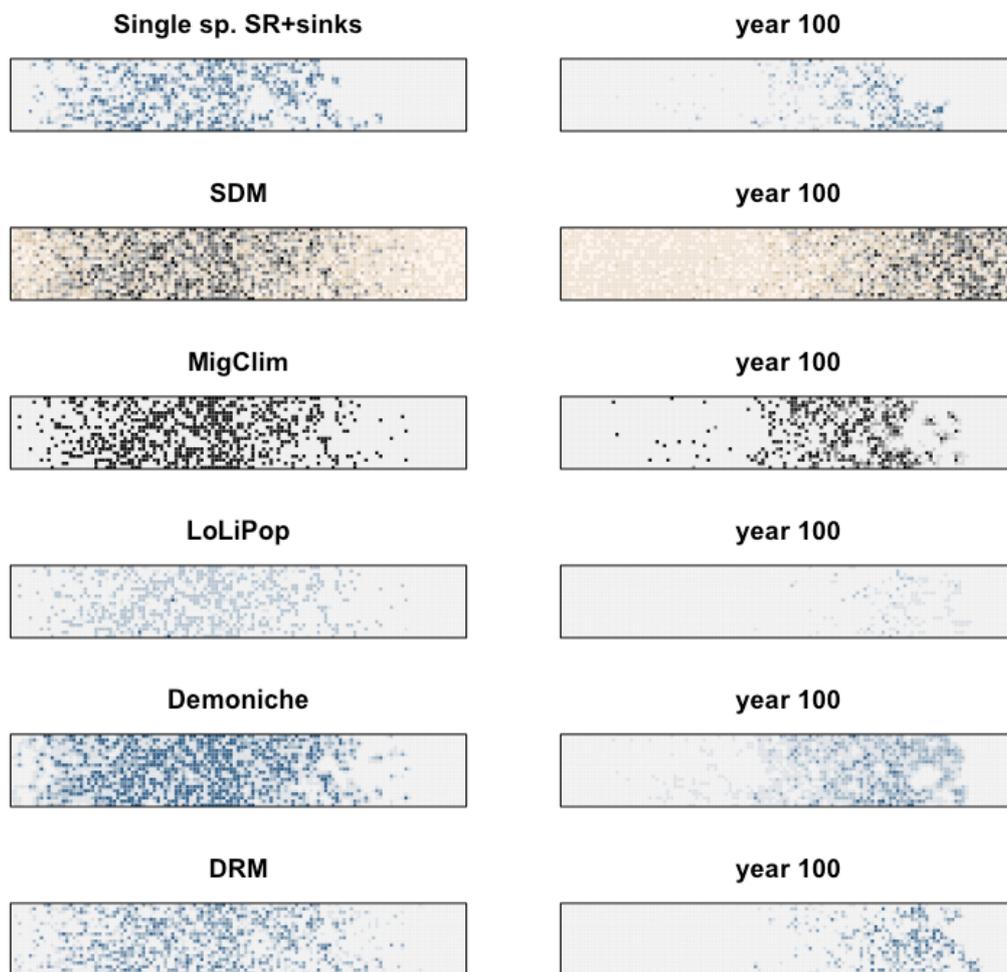


Figure S14. Observed abundances of virtual model (top row) and predictions of range models for year 0 before climate change (left) and year 100 after climate change (right) for the single species scenario with short range dispersal and source sink dynamics. Abundances are presented in blue shading with 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 colonized (in black) vs. uncolonized cells. Grey shading in MigClim indicates that these cells were not colonized in all replicate runs.

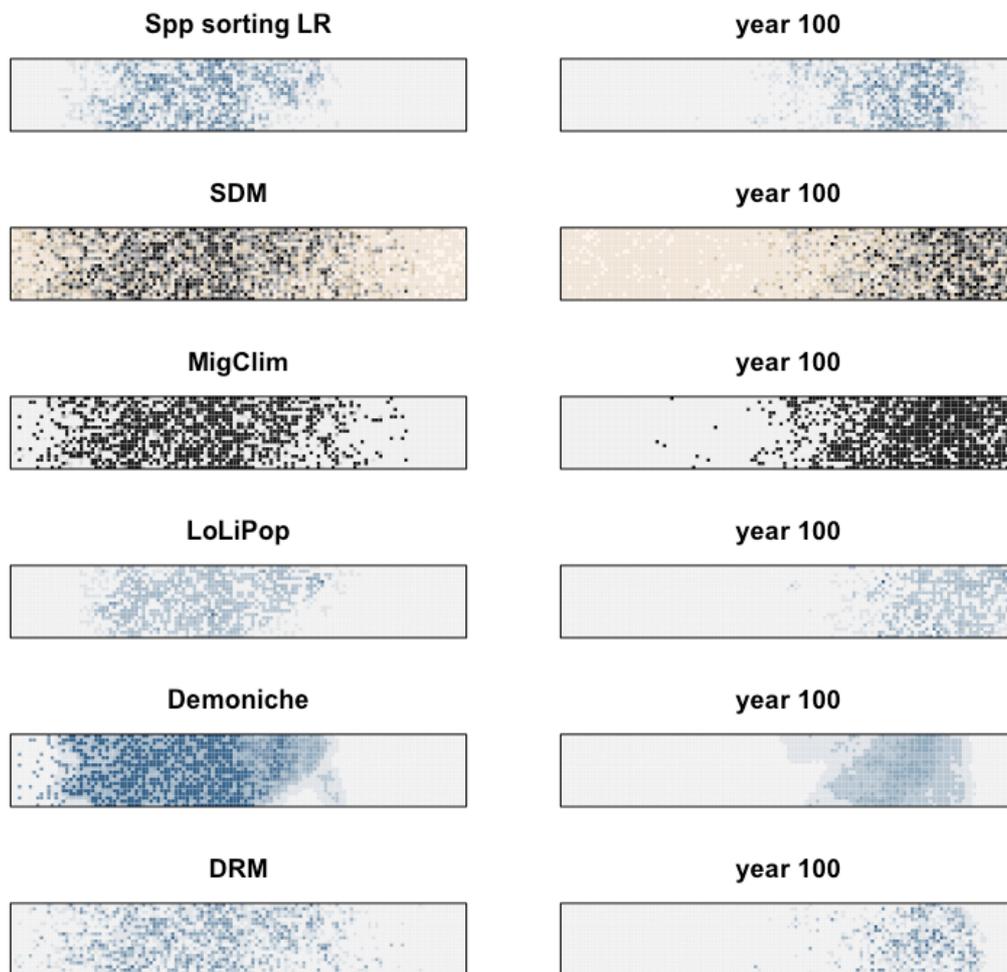


Figure S15. Observed abundances of virtual model (top row) and predictions of range models for year 0 before climate change (left) and year 100 after climate change (right) for the species sorting scenario with long range dispersal. Abundances are presented in blue shading with darker colour indicating 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 colonized (in black) vs. uncolonized cells. Grey shading in MigClim indicates that these cells were not colonized in all replicate runs.

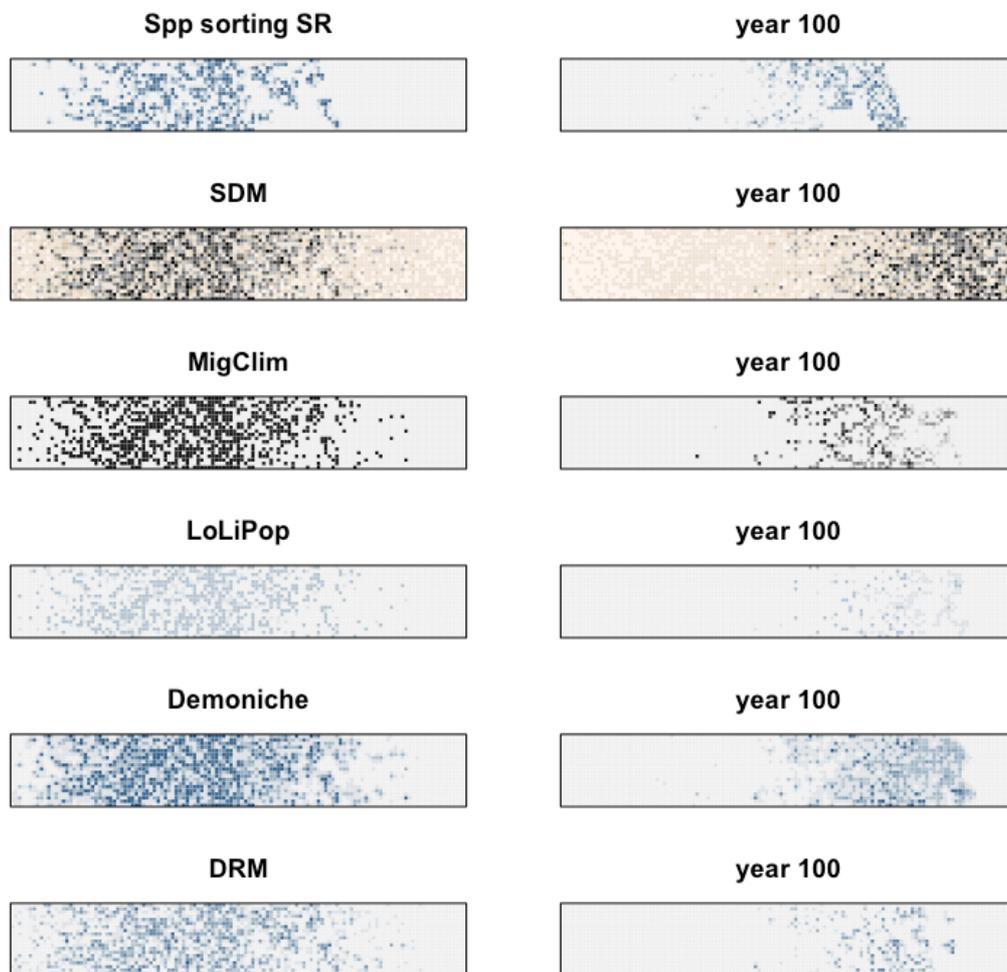


Figure S16. Observed abundances of virtual model (top row) and predictions of range models for year 0 before climate change (left) and year 100 after climate change (right) for the species sorting scenario with short range dispersal. Abundances are presented in blue shading with darker colour indicating 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 colonized (in black) vs. uncolonized cells. Grey shading in MigClim indicates that these cells were not colonized in all replicate runs.

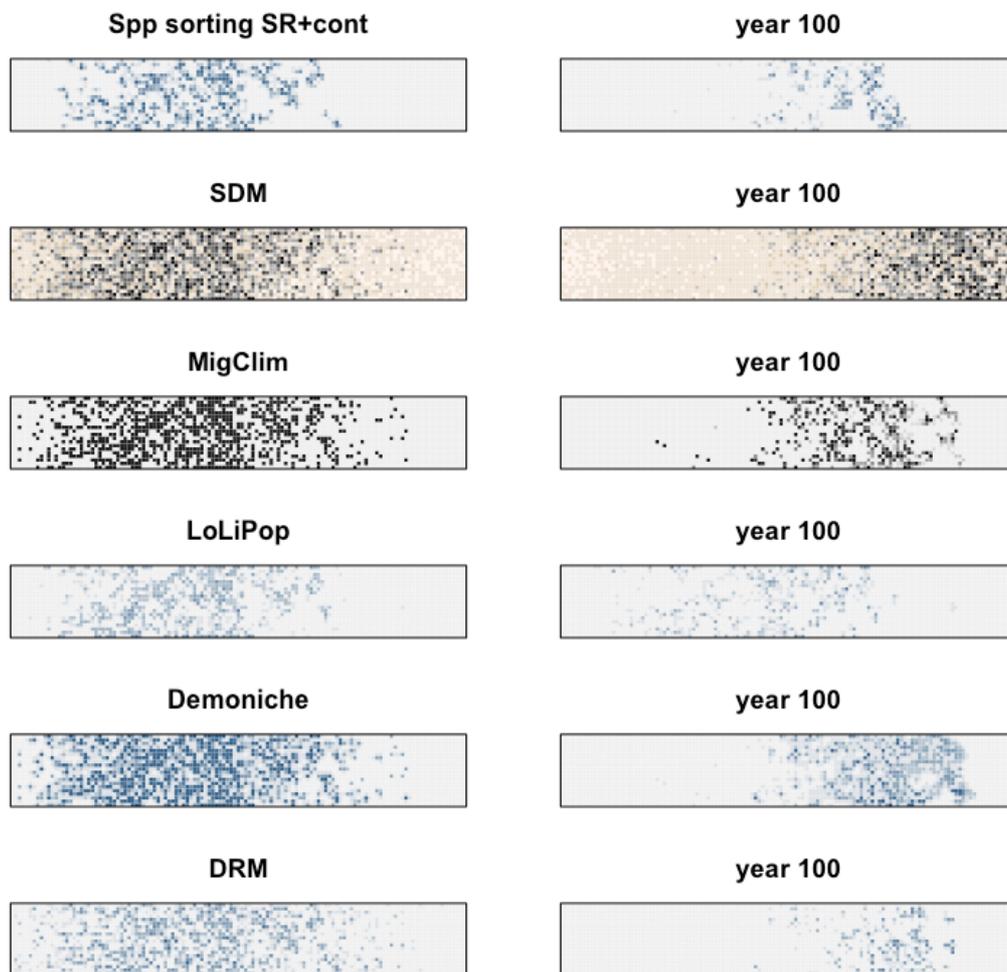


Figure S17. Simulated abundances of virtual model (top row) and predictions of range models for year 0 before climate change (left) and year 100 after climate change (right) for the species sorting scenario with short range dispersal and contagious disturbance. Abundances are presented in blue shading 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 colonized (in black) vs. uncolonized cells. Grey shading in MigClim indicates that these cells were not colonized in all replicate runs.

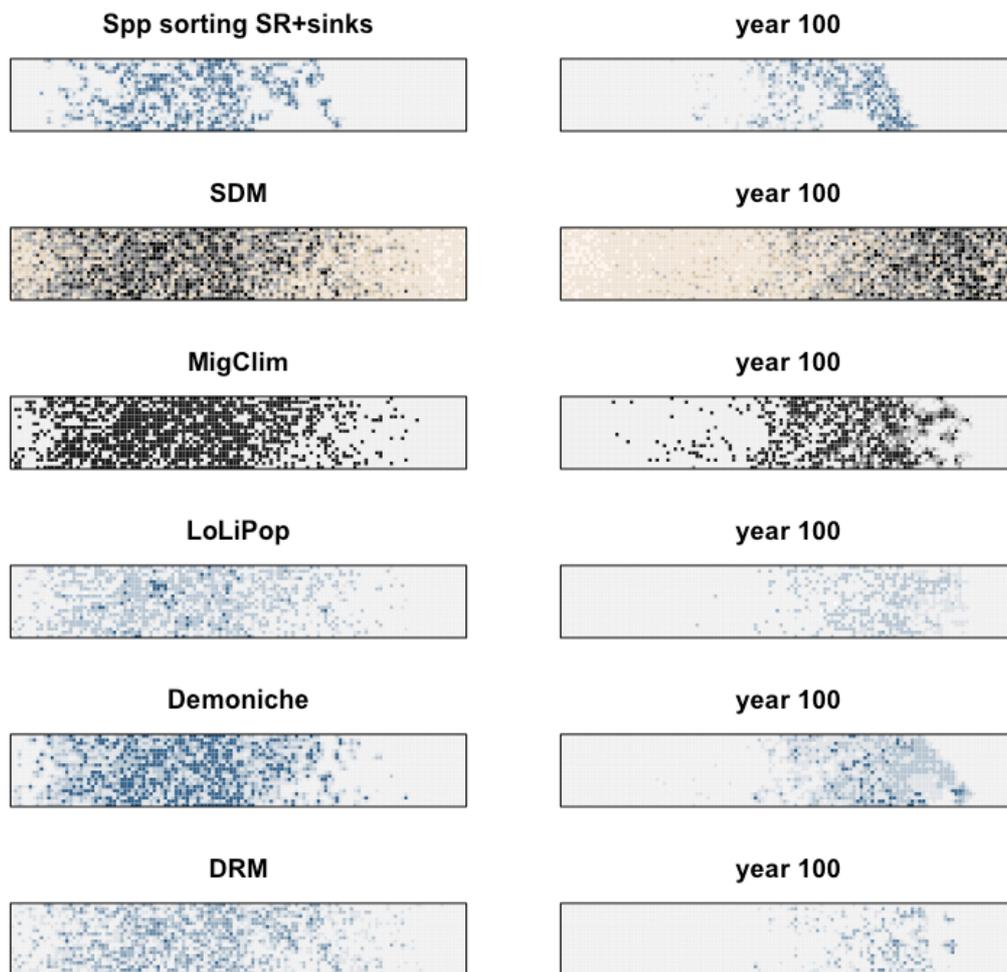


Figure S18. Observed abundances of virtual model (top row) and predictions of range models for year 0 before climate change (left) and year 100 after climate change (right) for the species sorting scenario with short range dispersal and source sink dynamics. Abundances are presented in blue shading with 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 colonized (in black) vs. uncolonized cells. Grey shading in MigClim indicates that these cells were not colonized in all replicate runs.

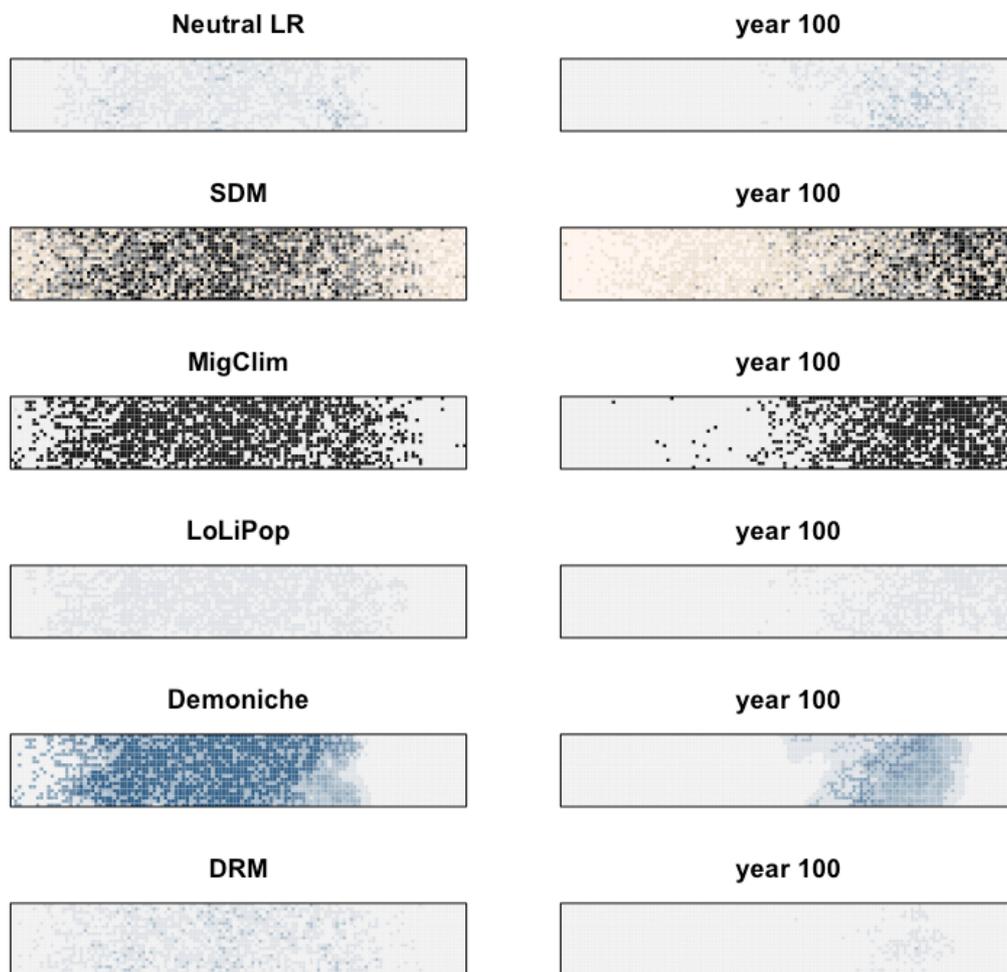


Figure S19. Observed abundances of virtual model (top row) and predictions of range models for year 0 before climate change (left) and year 100 after climate change (right) for the neutral dynamics scenario with long range dispersal. Abundances are presented in blue shading with darker colour indicating 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 colonized (in black) vs. uncolonized cells. Grey shading in MigClim indicates that these cells were not colonized in all replicate runs.

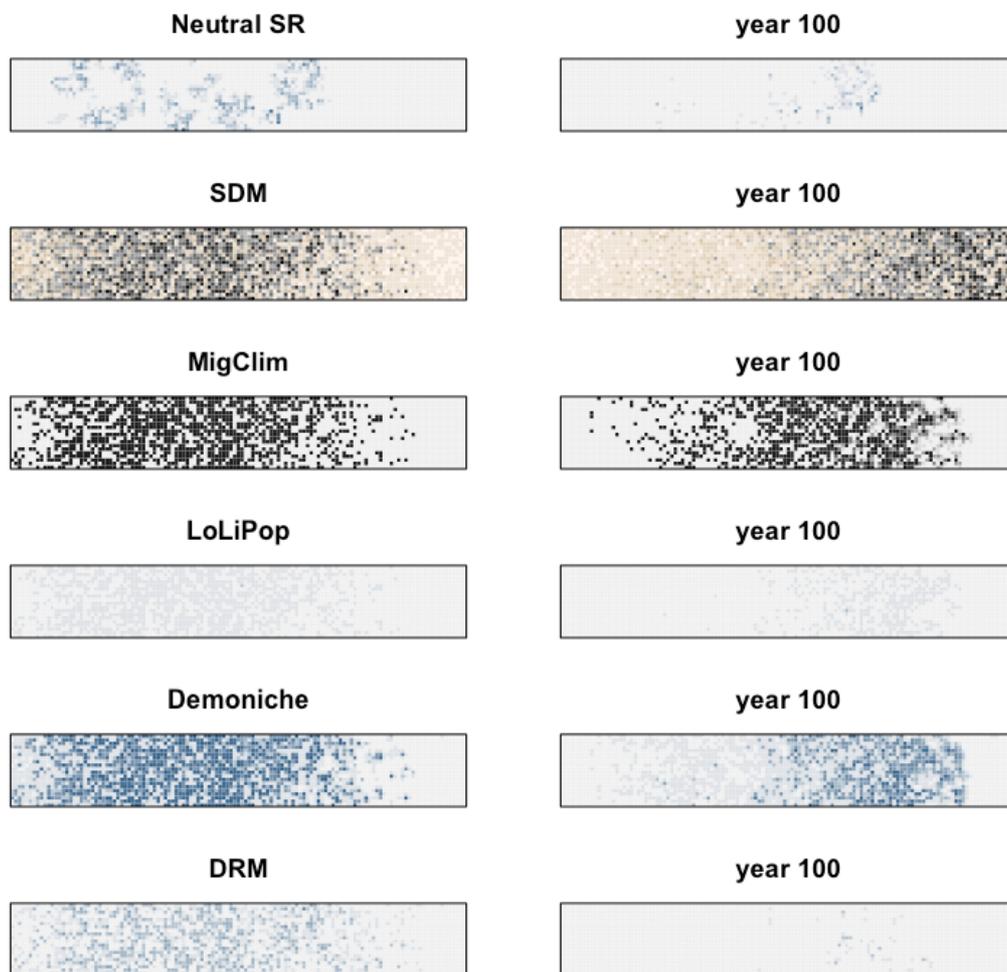


Figure S20. Observed abundances of virtual model (top row) and predictions of range models for year 0 before climate change (left) and year 100 after climate change (right) for the neutral dynamics scenario with short range dispersal. Abundances are presented in blue shading with darker colour indicating 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 colonized (in black) vs. uncolonized cells. Grey shading in MigClim indicates that these cells were not colonized in all replicate runs.

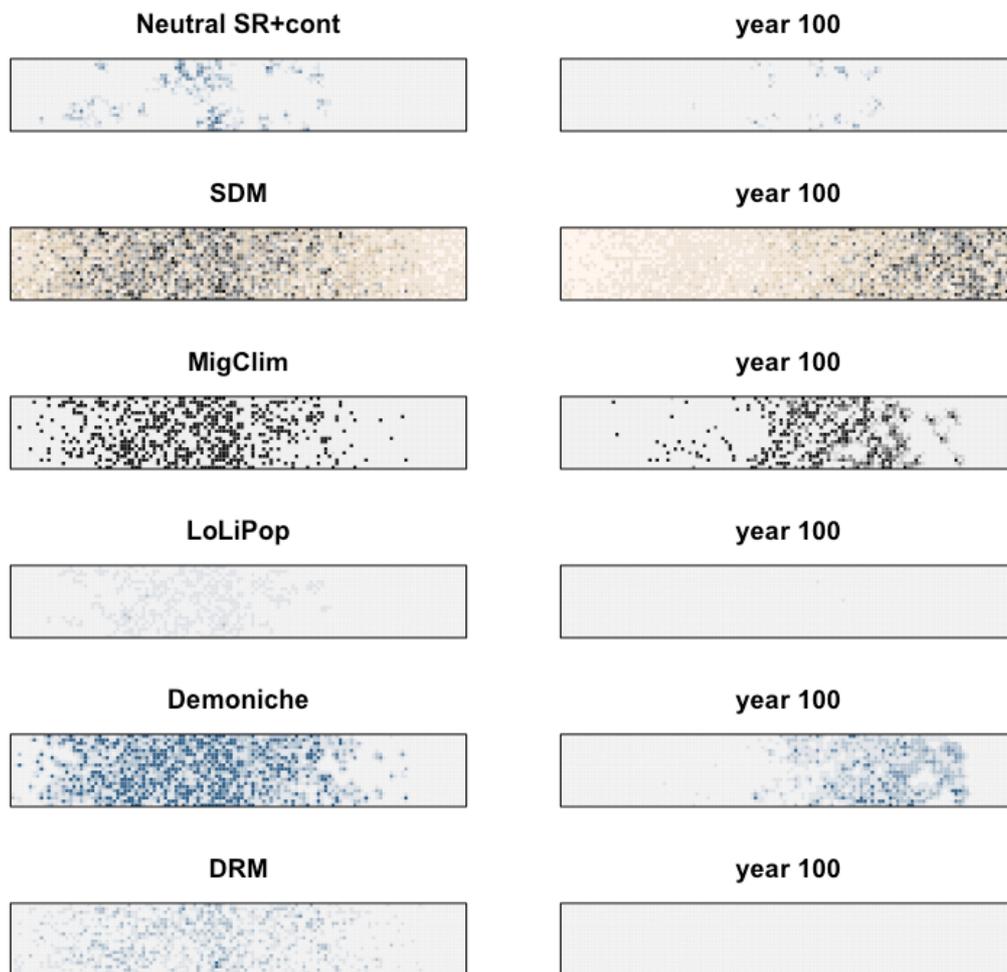


Figure S21. Observed abundances of virtual model (top row) and predictions of range models for year 0 before climate change (left) and year 100 after climate change (right) for the neutral dynamics scenario with short range dispersal and contagious disturbance. Abundances are presented in blue shading 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 colonized (in black) vs. uncolonized cells. Grey shading in MigClim indicates that these cells were not colonized in all replicate runs.

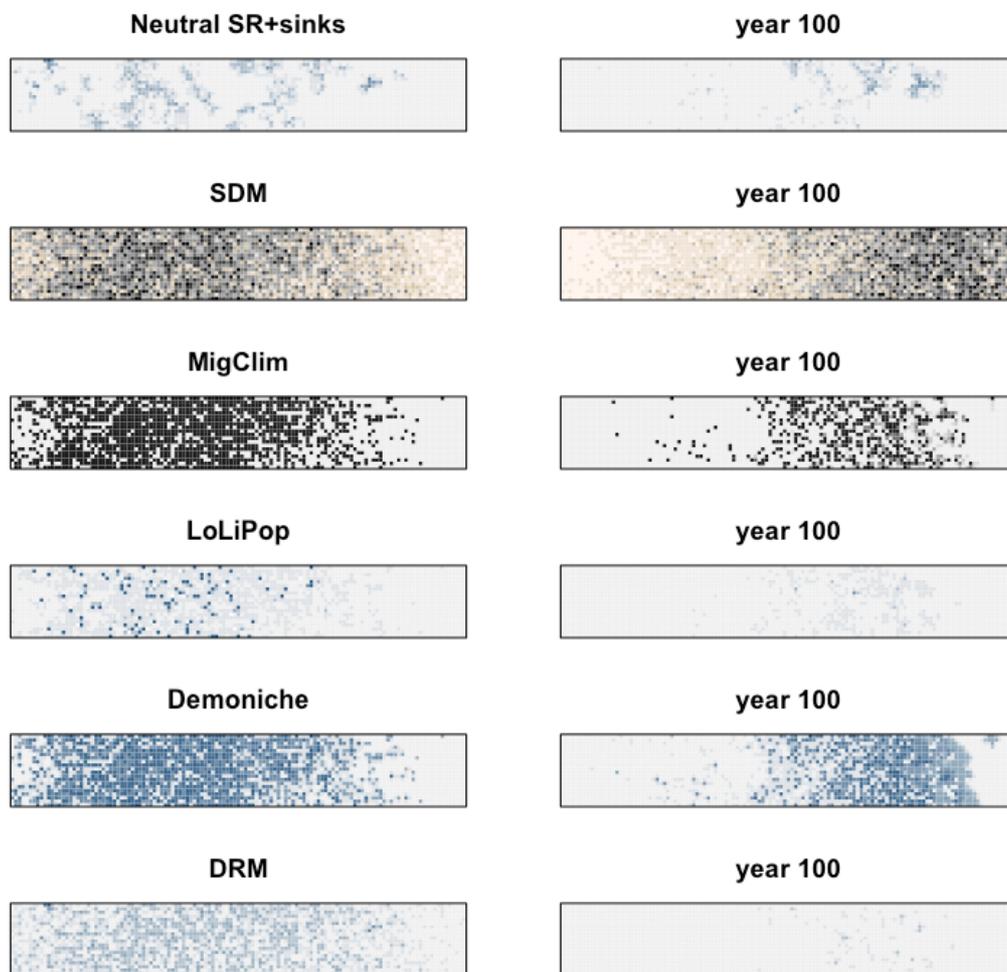


Figure S22. Observed abundances of virtual model (top row) and predictions of range models for year 0 before climate change (left) and year 100 after climate change (right) for the neutral dynamics scenario with short range dispersal and source sink dynamics. Abundances are presented in blue shading 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 colonized (in black) vs. uncolonized cells. Grey shading in MigClim indicates that these cells were not colonized in all replicate runs.

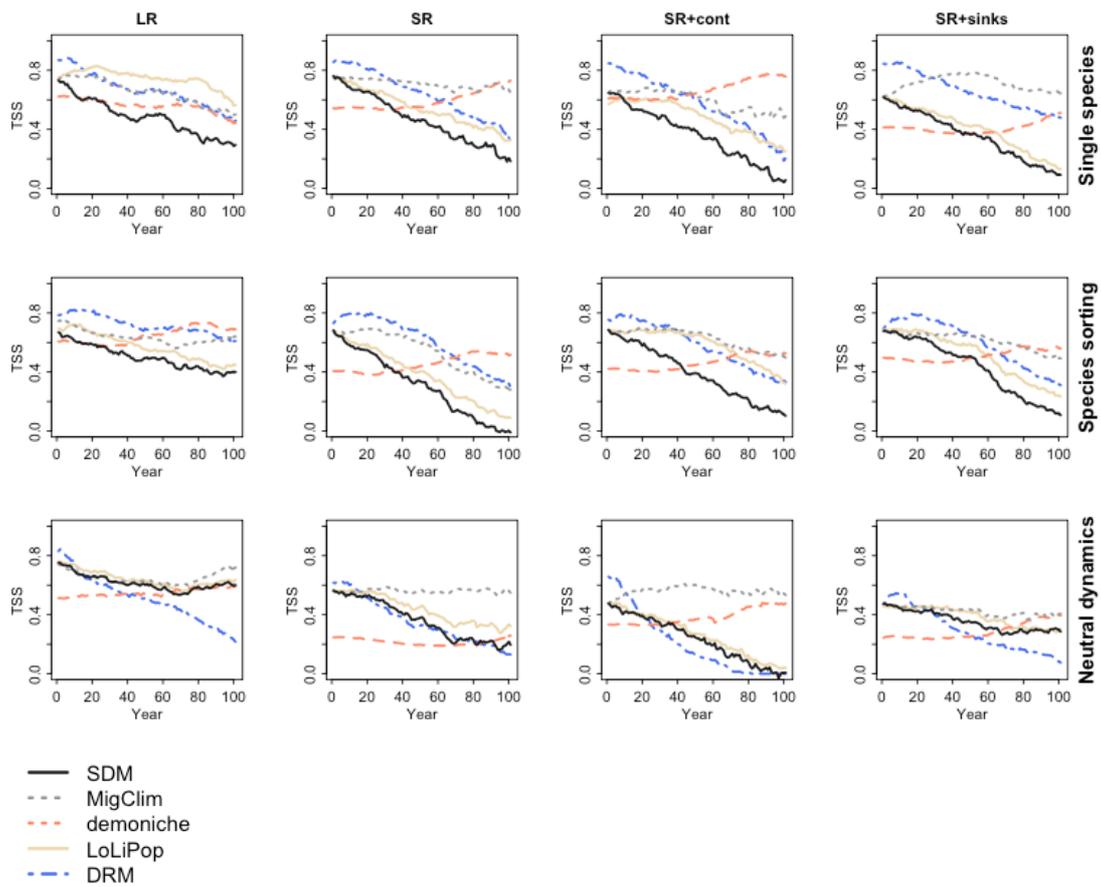


Figure S23. Time series of TSS (true skill statistic) obtained for different range models and scenarios.

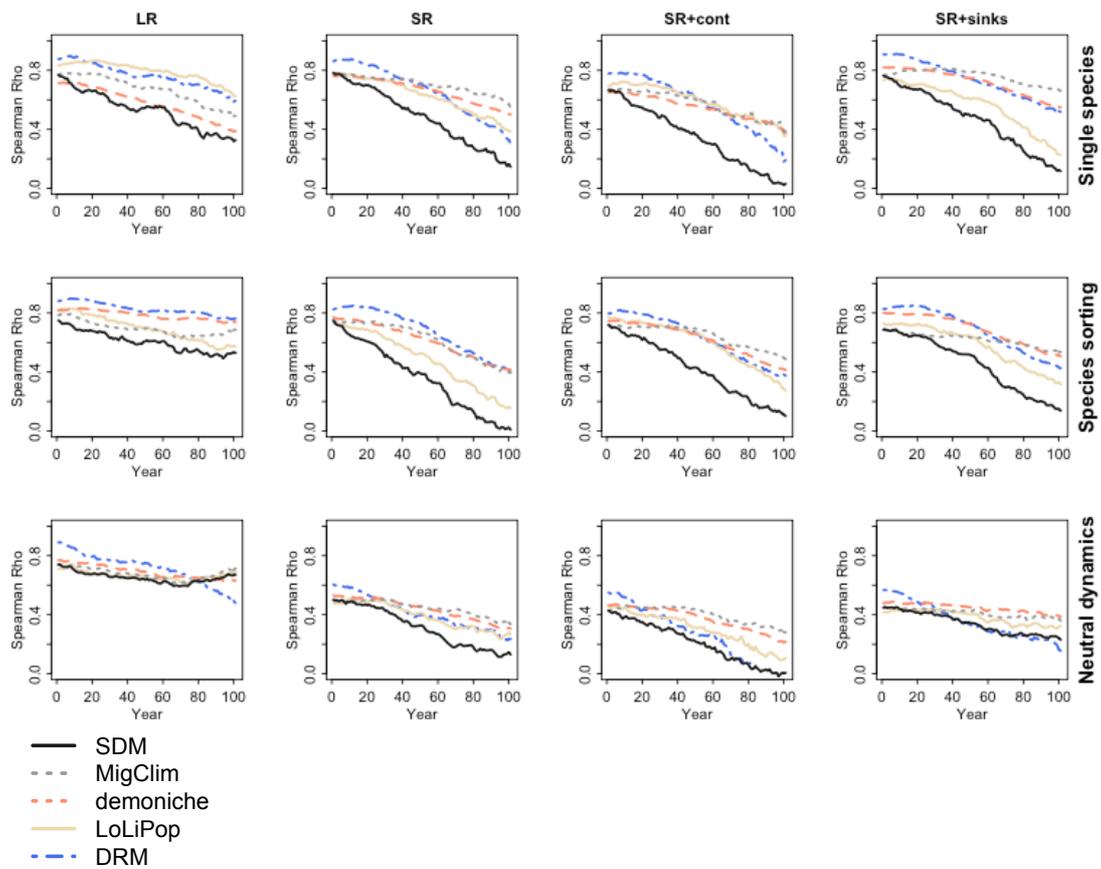


Figure S24. Time series of Spearman's rank correlation coefficient Rho for different range models and scenarios.

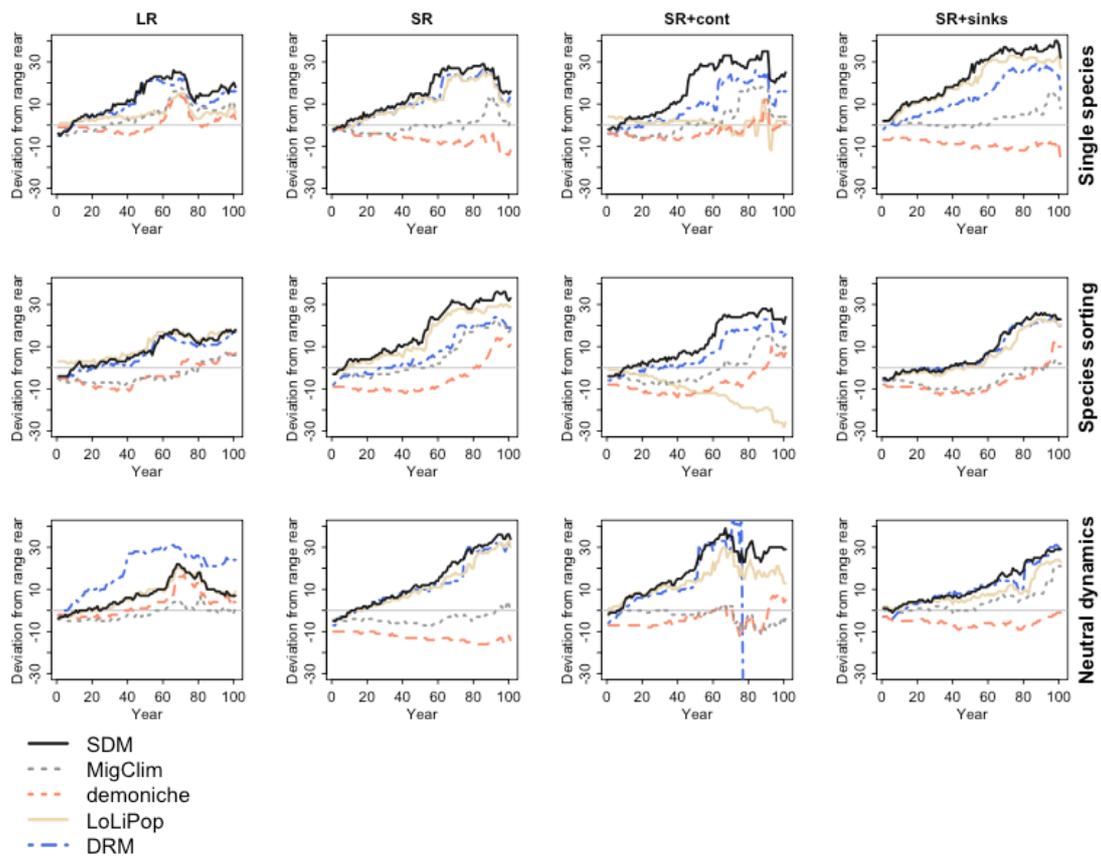


Figure S25. Time series of predicted rear range edge (5% quantile of latitudinal distribution) for different range models and scenarios.

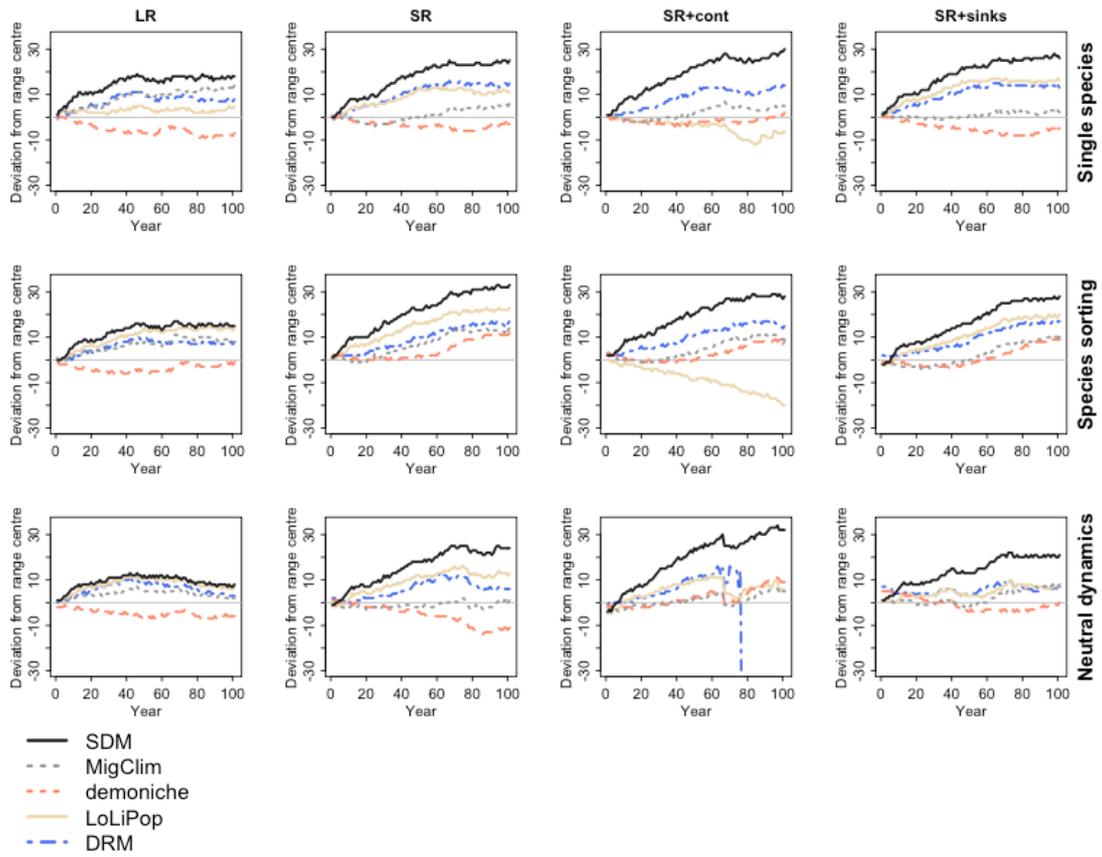


Figure S26. Time series of predicted range centre (Median of latitudinal distribution) for different range models and scenarios.

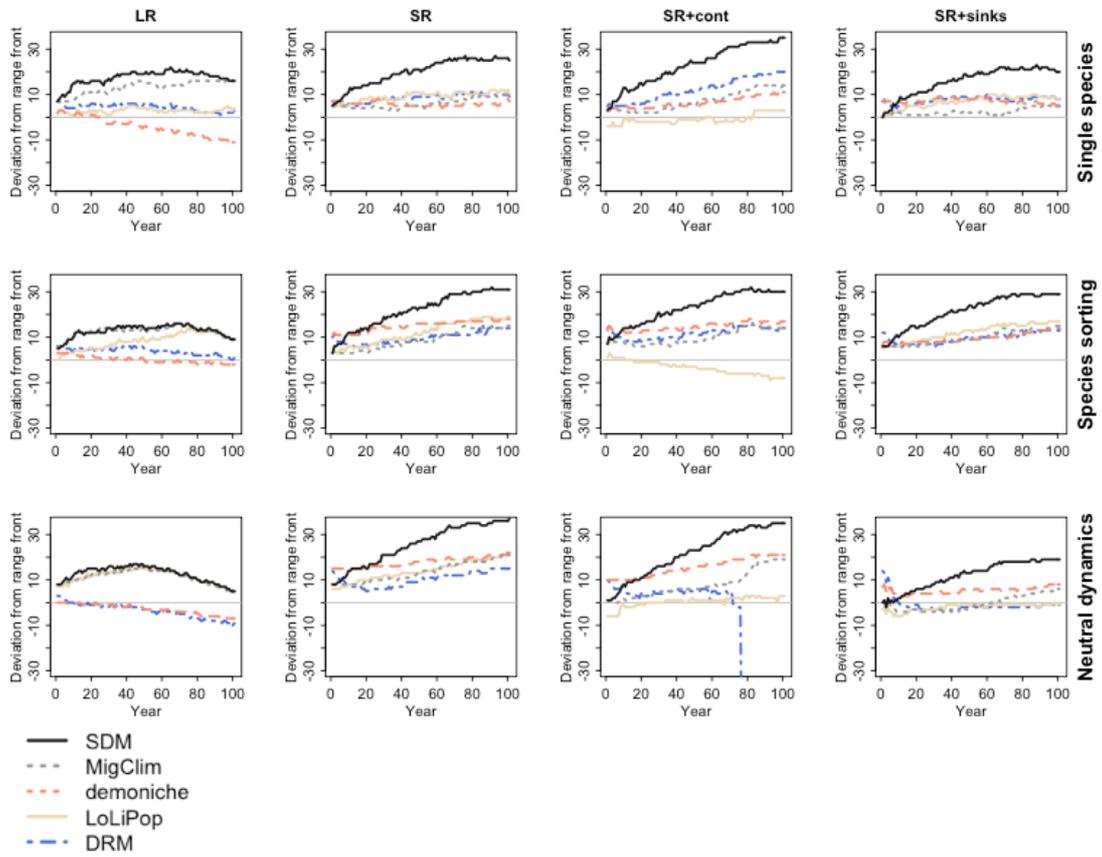


Figure S27. Time series of predicted range front (95% quantile of latitudinal distribution) for different range models and scenarios.

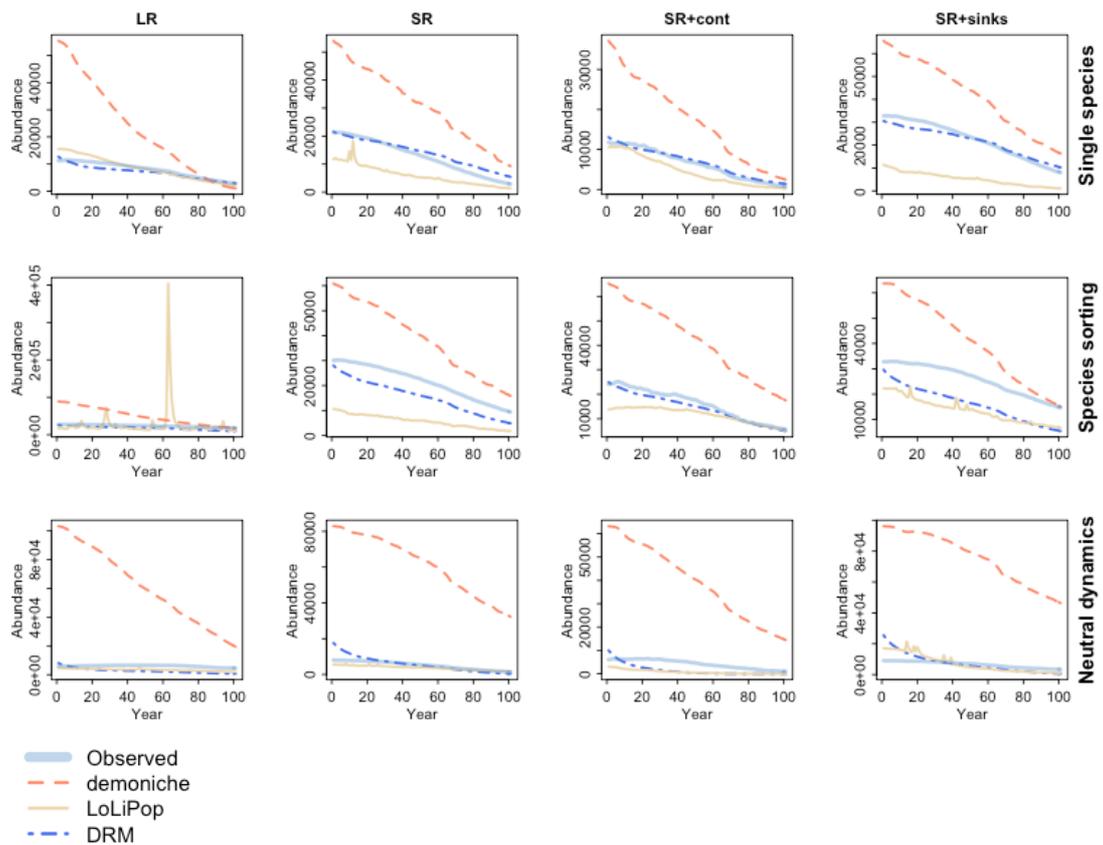


Figure S28. Time series of observed and predicted total abundance for different range models and scenarios.

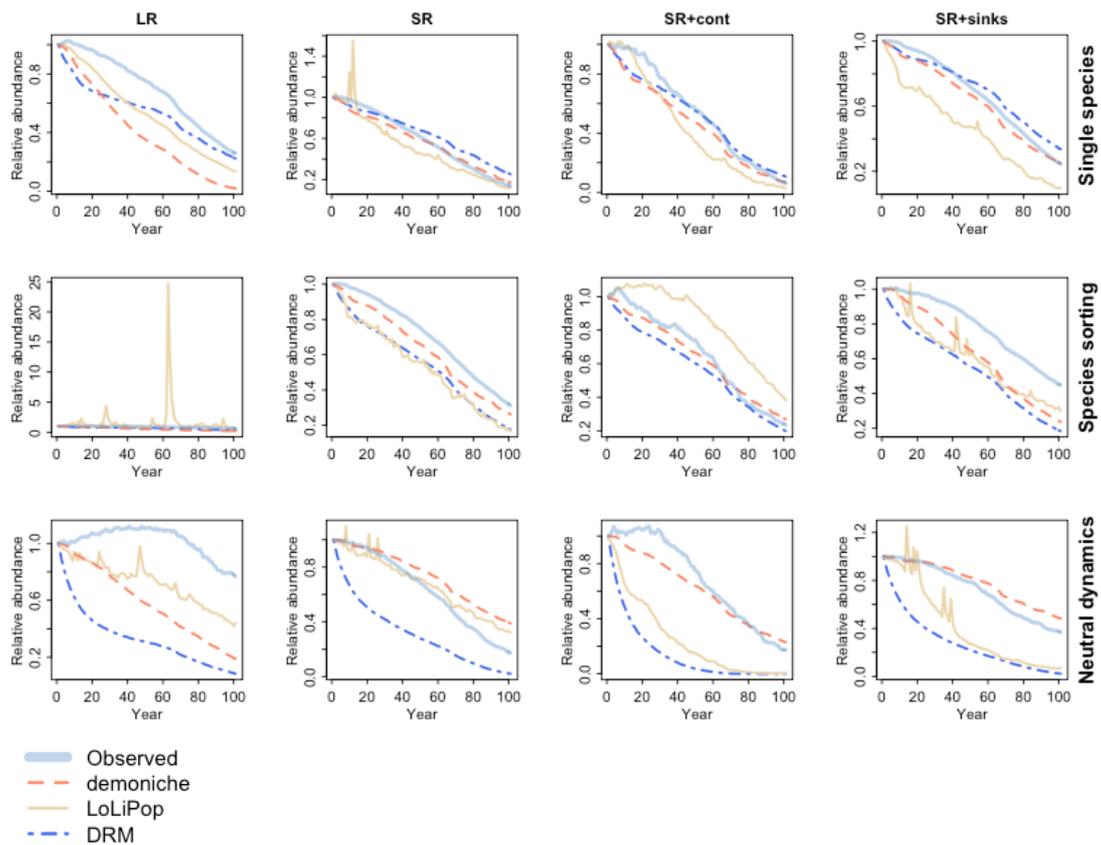


Figure S29. Time series of observed and predicted relative abundance for different range models and scenarios.

Appendix S4. Structural uncertainty in range models

Here, we present results from different model configurations in MigClim and DemoNiche, and discuss possible ways to deal with this structural uncertainty.

Appendix S4:	Pages
Structural uncertainty in MigClim.	47
Structural uncertainty in DemoNiche.	48
Structural uncertainty in DRMs.	52
Figures:	
Figure S30. Maps of predicted colonization by different MigClim configurations.	47
Figure S31. True skill statistic for different MigClim configurations.	48
Figure S32. Spearman's Rho for different DemoNiche configurations.	50
Figure S33. Total abundance predicted for different DemoNiche configurations.	51
Figure S34. Relative abundance predicted for different DemoNiche configurations.	52
Figure S35. Spearman's Rho for different DRM configurations.	53
Figure S36. Total abundance predicted for different DRM configurations.	54

Structural uncertainty in MigClim

The introductory articles on MigClim only compared the linear and the threshold relationship between habitat suitability and invasion probability for equilibrium conditions (Engler & Guisan 2009; Engler *et al.* 2012). Here, we extended the analyses to also test linear-threshold and sigmoidal relationships (cf. Appendix S1) and evaluated predictions under climate change. Our results show that all model configurations perform similarly well under equilibrium conditions while predictions differ considerably under climate change, especially for scenarios of long-range dispersal (Figs. S29-S30). Most strikingly, linear relationships may lead to completely misleading predictions under climate change with near-complete invasions of the entire grid. The threshold and linear-threshold relationships produced almost equivalent predictions. Under short-range dispersal, their performance decreased considerably throughout the climate change period while this performance decrease was not as strong in long-range dispersal scenarios. The sigmoidal relationship performed similarly well for all scenarios tested. In some cases, it was outperformed by the linear relationship but, overall, it can be judged the most robust configuration for MigClim. Here, more tests are needed for the choice of steepness in the sigmoidal function.

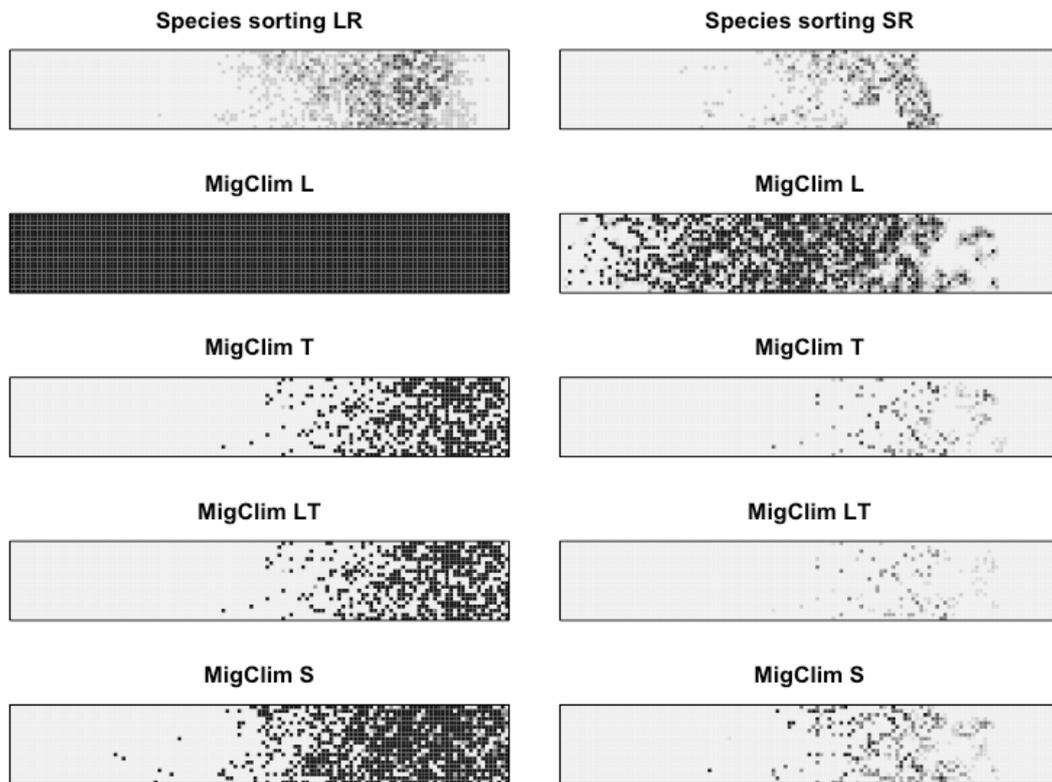


Figure S30. Observed abundances of virtual model (top row) and predictions of different MigClim configurations for year 100 after climate change for the species sorting scenario with long range (left) and short range dispersal (right). MigClim predicts colonized (in black) vs. uncolonized cells. Grey shading in indicates that these cells were not colonized in all replicate runs. L refers to linear scaling between habitat suitability and colonization probability; T refers to a threshold approach with colonization probability of 1 for predicted presences from SDMs and 0 for predicted absences from SDMs; LT uses the same threshold but linearly scales habitat suitability above that threshold; S refers to sigmoidal scaling.

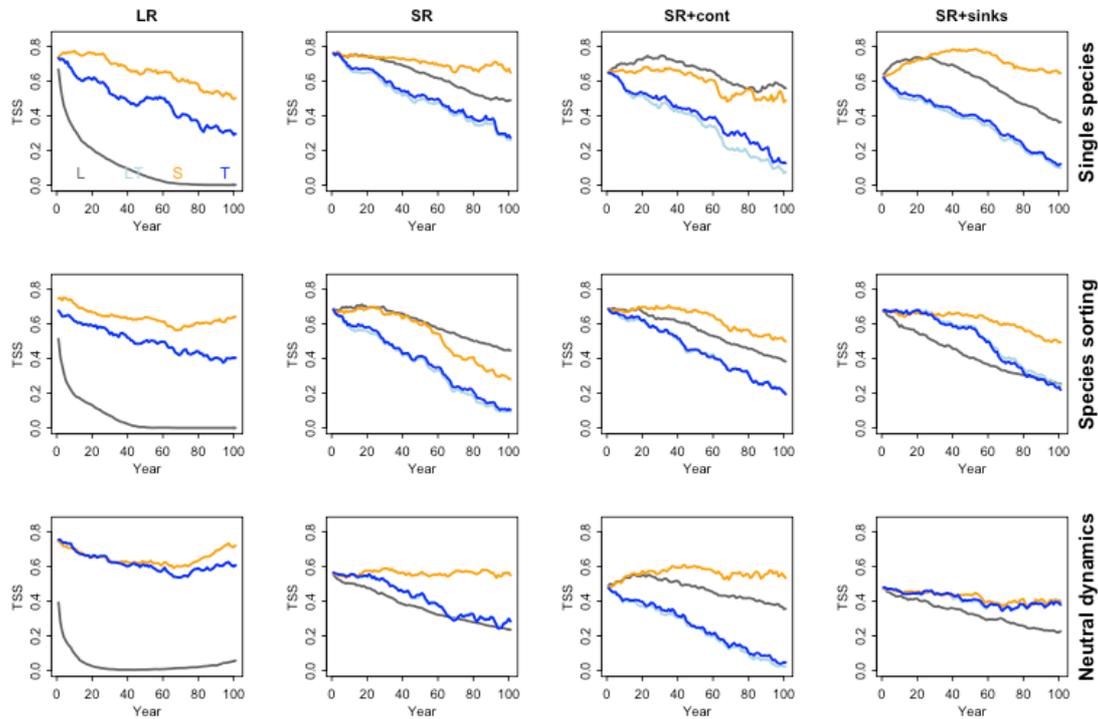


Figure S31. Time series of TSS (true skill statistic) values obtained for different MigClim configurations. **L** refers to linear scaling between habitat suitability and colonization probability; **T** refers to a threshold approach with colonization probability of 1 for predicted presences from SDMs and 0 for predicted absences from SDMs; **LT** uses the same threshold but linearly scales habitat suitability above that threshold; **S** refers to sigmoidal scaling. In many cases, **LT** and **T** are not distinguishable.

Structural uncertainty in DemoNiche

In DemoNiche, different functional relationships between habitat suitability and demographic property (threshold, linear, linear-threshold, sigmoidal) and the choice of demographic property that is constrained by habitat suitability (vital rates, carrying capacity; cf. Appendix S1) resulted in highly variable model predictions and evaluation scores. One first important notion from our virtual analyses is that when scaling vital rates by habitat suitability (except for threshold relationship), the stage matrix needs to be calibrated because such procedure implicitly assumes that optimal demographic rates for habitat suitability values near 1 are known. Otherwise, these model configurations will lead to strong underestimation of population viability (not shown). Here, we suggested calibrating the stage matrix by minimising differences between observed and predicted mean demographic rates (Appendix S1). Alternatively, abundance data could be used for calibration similar to the estimation approach in LoLiPop.

As expected, we found that scaling adult survival rates by habitat suitability led to overestimation of extinction risks (Fig. S33), because in our virtual simulation model only fecundity and recruitment were related to environment. Excluding such model configurations through prior knowledge of the system greatly reduces structural uncertainty.

Under short-range dispersal, differences in model performance when constraining (plausible) vital rates and carrying capacity by habitat suitability were relatively small. Not surprisingly, the most striking differences between these configurations was found in total abundance estimates (Fig. S32) while differences in, for example, Rho (Fig. S31) and relative abundance (Fig. S33) were much smaller. In contrast, under long-range dispersal, constraining carrying capacity by habitat suitability often led to stronger overestimation of extinction risk, and for the different community scenarios, extinction was falsely predicted in at least one of the carrying capacity settings.

Overall, we found no signs that one functional relationship between demographic property and habitat suitability performed superior to others. We did find, however, that those model configurations that achieved highest evaluation scores under equilibrium conditions usually remained among the best also under climate change. We, thus, suggest comparing multiple structural relationships and only using the best models for making predictions. Here, we chose a fixed measure of taking the best five model configurations (evaluated on sampling data using Spearman's rank correlation coefficient ρ between observed and predicted abundances). One could also choose to include only those model configurations that achieved evaluation scores above a certain threshold, e.g. $AUC > 0.7$ as is common practice in SDM ensembles. The uncertainty in predictions from these best model configurations may already serve as indication for overall reliability of predictions.

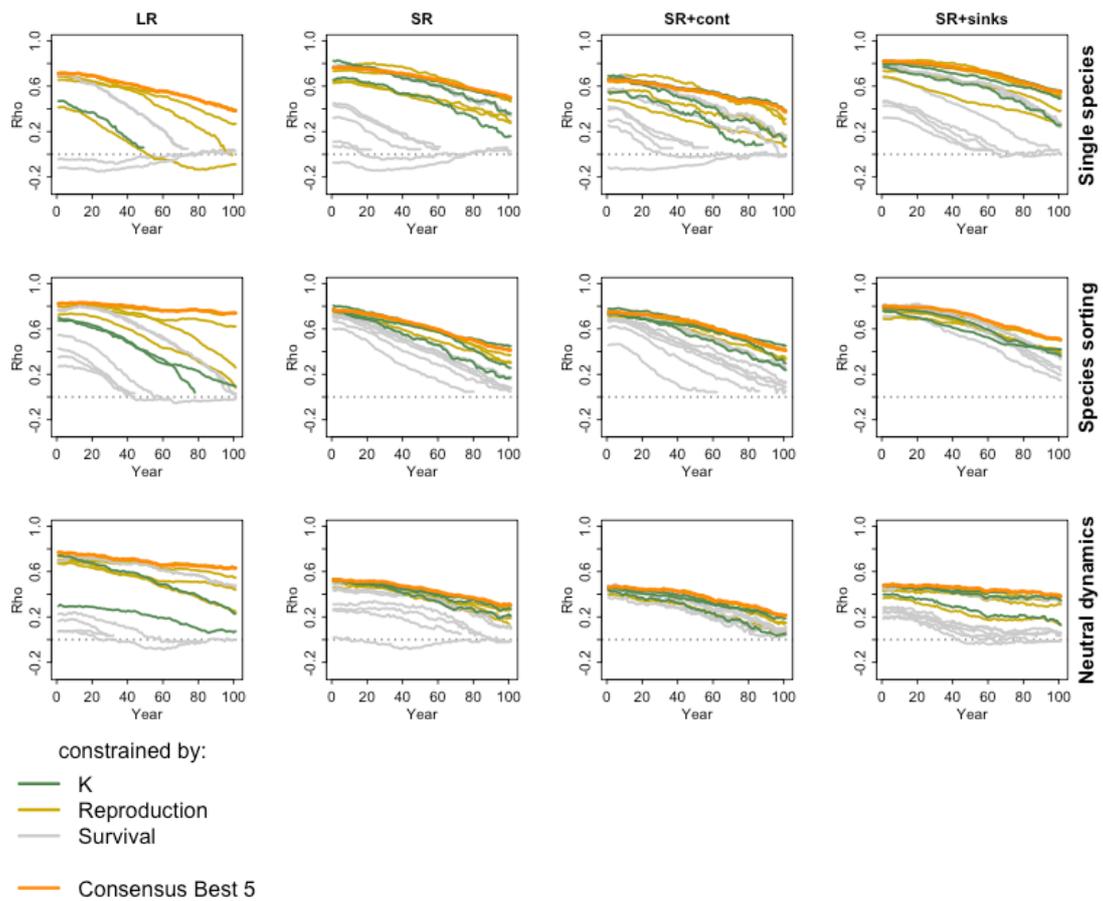


Figure S32. Time series of Spearman's rank correlation coefficient Rho for different DemoNiche configurations that differ in how demography is constrained by habitat suitability. Orange line represents prediction from committee averaging of best 5 models; green lines indicate models that constrain carrying capacity, golden lines indicate models that constrain reproduction only, and light grey lines indicate models that constrain survival.

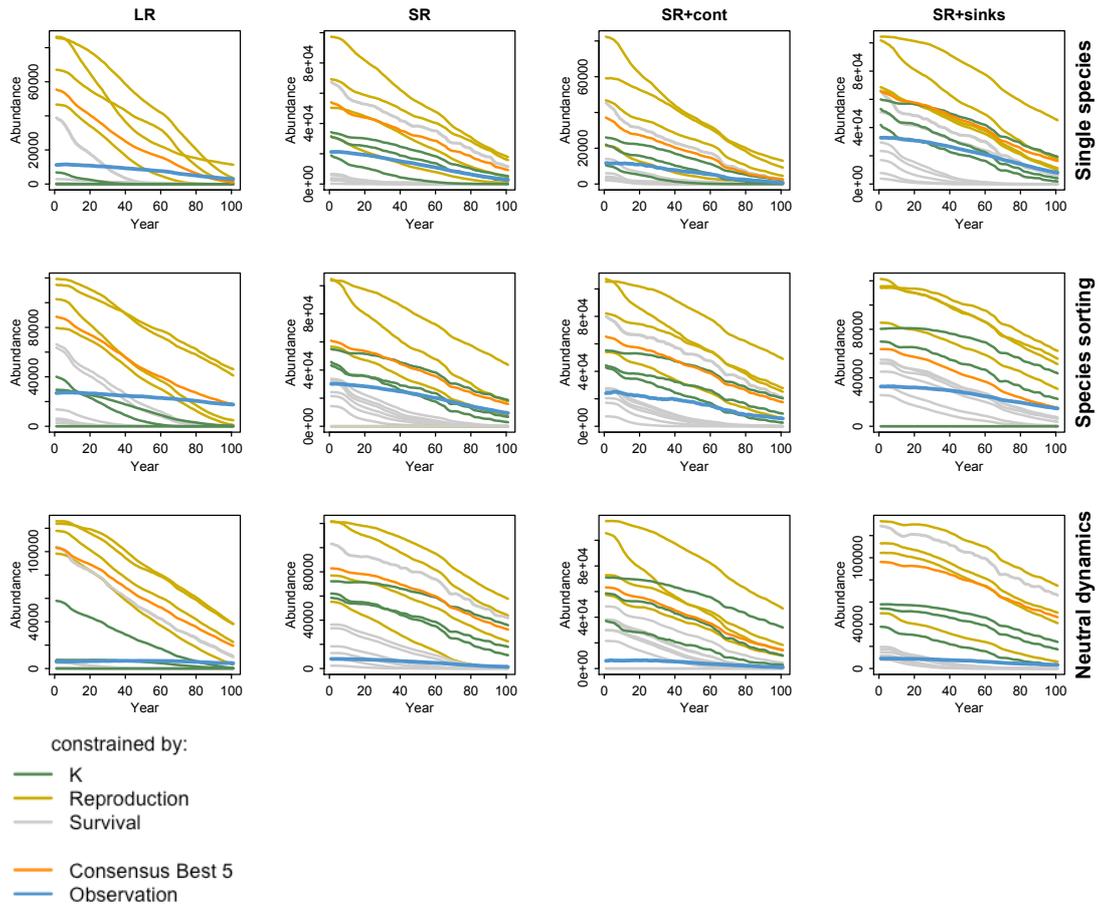


Figure S33. Time series of observed total abundance and total abundance predicted for different DemoNiche configurations. Blue line indicates observations; orange line represents prediction from committee averaging of best 5 models; green lines indicate models that constrain carrying capacity; golden lines indicate models that constrain reproduction only; and light grey lines indicate models that constrain survival.

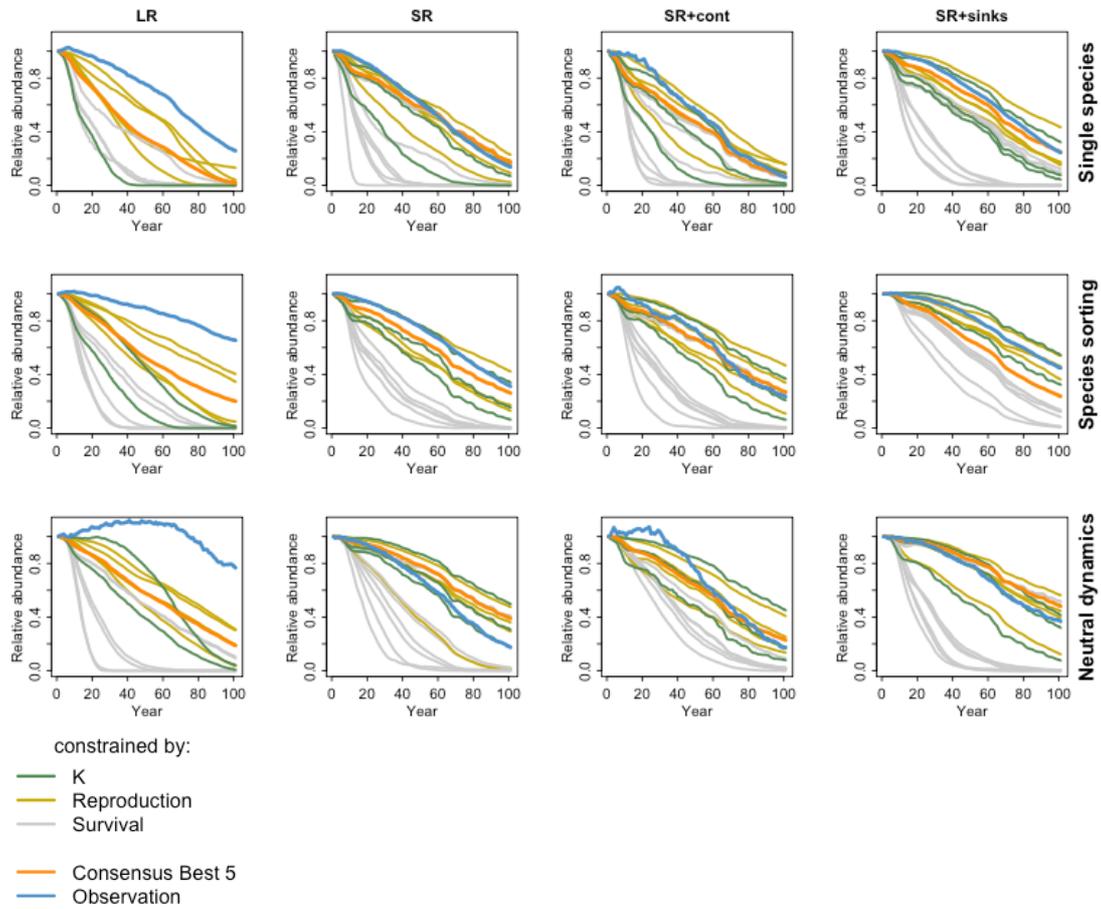


Figure S34. Time series of observed relative abundance and relative abundance predicted for different DemoNiche configurations. Blue line indicates observations; orange line represents prediction from committee averaging of best 5 models; green lines indicate models that constrain carrying capacity; golden lines indicate models that constrain reproduction only; and light grey lines indicate models that constrain survival.

Structural uncertainty in DRMs

We tested slight modifications to the DRM settings: (i) Instead of estimating the dispersal kernel from the data, dispersal rates were fixed to correspond to the dispersal parameter $2 \times \alpha$. (ii) The estimated abundance distribution in year 0 is corrected prior to making predictions by setting positive abundance counts in locations beyond the observed range margins to zero.

Overall, these different DRM configurations resulted in only slightly different predictions. In the future, different environmental-demography relationships should also be explored.

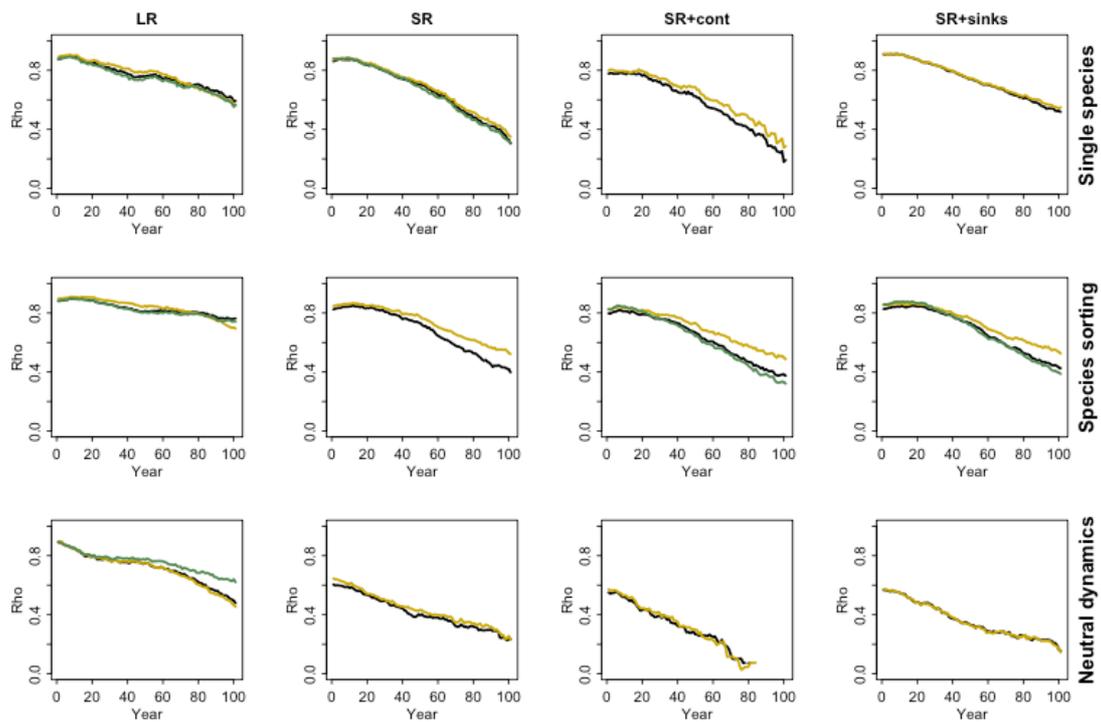


Figure S35. Time series of Spearman's rank correlation coefficient Rho for different DRM configurations. Blue line indicates observations; black line represents predictions as presented in main article; golden lines indicate models that correct abundances of year 0 prior to prediction; green lines indicate models with fixed dispersal parameters.

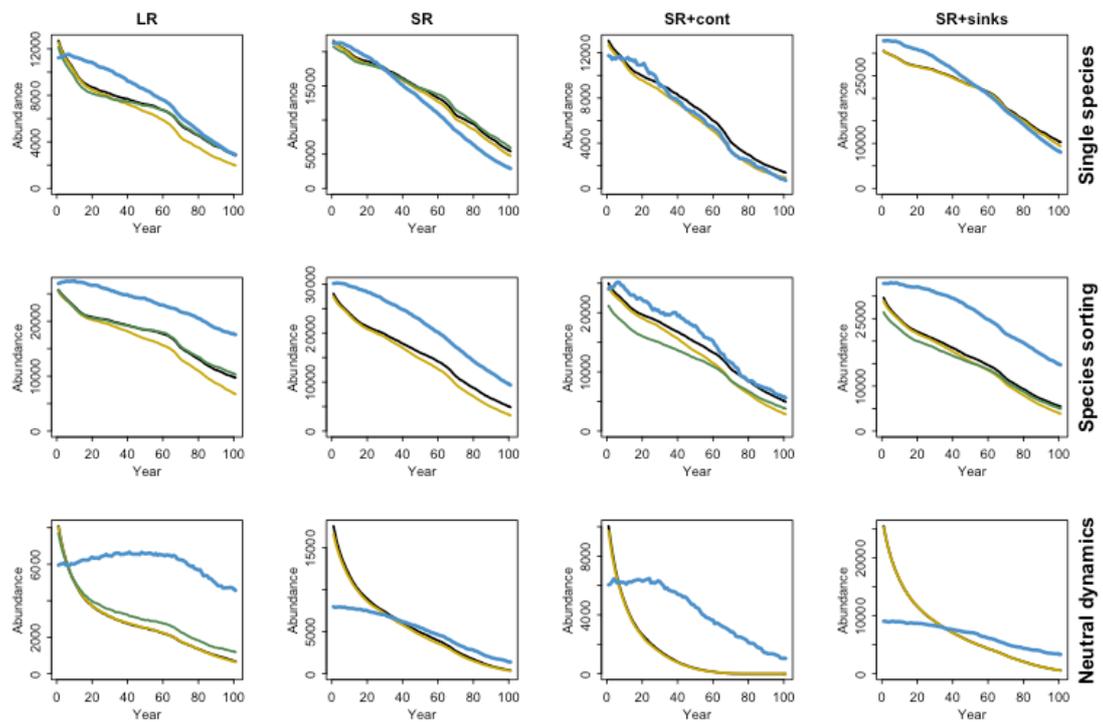


Figure S36. Time series of observed total abundance and total abundance predicted for different DRM configurations. Blue line indicates observations; black line represents predictions as presented in main article; golden lines indicate models that correct abundances of year 0 prior to prediction; green lines indicate models with fixex dispersal parameters.