1	Benchmarking novel approaches for modelling species range
2	dynamics
3	
4	Damaris Zurell ¹ *, Wilfried Thuiller ^{2,3} , Jörn Pagel ⁴ , Juliano S. Cabral ^{5,6} , Tamara
5	Münkemüller ^{2,3} , Dominique Gravel ⁷ , Stefan Dullinger ⁸ , Signe Normand ⁹ , Katja H.
6	Schiffers ^{2,3,10} , Kara A. Moore ¹¹ and Niklaus E. Zimmermann ^{1,12}
7	¹ Dynamic Macroecology, Swiss Federal Research Institute WSL, Zürcherstrasse 111, CH-
8	8903 Birmensdorf, Switzerland
9	² Univ. Grenoble Alpes, Laboratoire d'Écologie Alpine (LECA), UMR-CNRS 5553
10	Université J. Fourier BP 53, F-38000 Grenoble, France
11	³ CNRS, Laboratoire d'Écologie Alpine (LECA), UMR-CNRS 5553 Université J. Fourier BP
12	53, F-38000 Grenoble, France
13	⁴ Institute of Landscape and Plant Ecology, University of Hohenheim, August-v.Hartmann-
14	Str. 3, D-70599 Stuttgart, Germany
15	⁵ Biodiversity, Macroecology and Conservation Biogeography, University Göttingen,
16	Büsgenweg 2, D-37077, Goettingen, Germany
17	⁶ Synthesis Centre of the German Centre for Integrative Biodiversity Research (iDiv),
18	Deutscher Platz 5e, D-04103 Leipzig, Germany
19	⁷ Université de Québec à Rimouski, 300 Allée des Ursulines, Rimouski, Canada. G5L 3A1
20	⁸ Department of Botany and Biodiversity Research, University of Vienna, Rennweg 14, 1030
21	Vienna, Austria

- ⁹ Section for Ecoinformatics and Biodiversity, Department of Bioscience, Aarhus University,
- 23 Ny Munkegade 114, DK-8000 Aarhus C, Denmark
- ¹⁰ Senckenberg Biodiversity and Climate Research Centre (BiK-F), Georg Voigt-Straße 14-
- 25 16, D-60325 Frankfurt (Main), Germany
- 26 ¹¹ Center for Population Biology, University of California, Davis, One Shields Avenue,
- 27 Davis, CA, 95616, USA
- 28 ¹² Department of Environmental Systems Science, Swiss Federal Institute of Technology
- 29 ETH, CH-8092 Zurich, Switzerland
- 30
- 31 Email addresses: DZ (damaris.zurell@wsl.ch), WT (wilfried.thuiller@ujf-grenoble.fr), JP
- 32 (joern.pagel@uni-hohenheim.de), JSC (jscabral@gmx.de), TM (tamara.muenkemueller@ujf-
- 33 grenoble.fr), DG (Dominique_Gravel@uqar.ca), SD (stefan.dullinger@univie.ac.at), SN
- 34 (signe.normand@bios.au.dk), KHS (katja.schiffers@senckenberg.de), KAM
- 35 (kmoore@ucdavis.edu), NEZ (niklaus.zimmermann@wsl.ch)
- 36 * Corresponding author: Damaris Zurell, Dynamic Macroecology, Swiss Federal Research
- 37 Institute WSL, Zürcherstr. 111, CH-8903 Birmensdorf, Switzerland, tel: +41 44 739 25 71,
- 38 fax: +41 44 7392 215, email: damaris.zurell@wsl.ch
- 39 Running head: Benchmarking range dynamic models
- 40 Keywords: climate change, species distribution models, demographic models, dispersal,
- 41 virtual ecologist approach, simulated data, prediction, population viability
- 42 Article type: Primary research article
- 43

45 Abstract

46	Increasing biodiversity loss due to climate change is one of the most vital challenges of the
47	21 st century. To anticipate and mitigate biodiversity loss, models are needed that reliably
48	project species' range dynamics and extinction risks. Recently, several new approaches to
49	model range dynamics have been developed to supplement correlative species distribution
50	models (SDMs), but applications clearly lag behind model development. Indeed, no
51	comparative analysis has been performed to evaluate their performance.
52	Here, we build on process-based, simulated data for benchmarking five range (dynamic)
53	models of varying complexity including classical SDMs, SDMs coupled with simple dispersal
54	or more complex population dynamic models (SDM hybrids), and a hierarchical Bayesian
55	process-based dynamic range model (DRM). We specifically test the effects of demographic
56	and community processes on model predictive performance. Under current climate, DRMs
57	performed best, although only marginally. Under climate change, predictive performance
58	varied considerably, with no clear winners. Yet, all range dynamic models improved
59	predictions under climate change substantially compared to purely correlative SDMs, and the
60	population dynamic models also predicted reasonable extinction risks for most scenarios.
61	When benchmarking data were simulated with more complex demographic and community
62	processes, simple SDM hybrids including only dispersal often proved most reliable. Finally,
63	we found that structural decisions during model building can have great impact on model
64	accuracy, but prior system knowledge on important processes can reduce these uncertainties
65	considerably.

66 Our results reassure the clear merit in using dynamic approaches for modelling species'
67 response to climate change but also emphasise several needs for further model and data
68 improvement. We propose and discuss perspectives for improving range projections through

- 69 combination of multiple models and for making these approaches operational for large
- 70 numbers of species.

72 Introduction

As climate change advances in its threat to biodiversity worldwide, reliable predictions of
range dynamics are needed to anticipate and mitigate potential impacts (Bellard *et al.*, 2012,
Pereira *et al.*, 2010), and we have seen an upsurge of related methods and applications in
recent years (Lurgi *et al.*, 2015, Normand *et al.*, 2014).

77 Range shifts are complex ecological processes driven by population dynamics and dispersal. 78 These processes are co-determined by a plethora of other factors including changes to the 79 abiotic and biotic environment (Sexton et al., 2009). Adequately representing these 80 interacting processes in an operational model and collecting data for reliably estimating the 81 many parameters of such complex models is a formidable challenge even for a single species 82 (Ehrlén & Morris, 2015) let alone for complex ecosystems. Older models relied on highly 83 simplified conceptualizations where the abiotic environment is the essential driver of climate-84 induced range shifts ignoring any demographic processes involved in range shifts. These so-85 called species distribution models (SDM; Guisan & Thuiller, 2005, Guisan & Zimmermann, 86 2000) have reached high popularity for providing biodiversity scenarios under climate 87 change, owing to the strong simplification of the represented processes and their relative ease 88 to use. However, their use for climate change projections has been discussed controversially 89 (Dormann et al., 2012, Thuiller et al., 2013) because SDMs assume that observed species' 90 distributions are not affected by dispersal limitations (Svenning et al., 2008) or source-sink 91 dynamics (Holt *et al.*, 2005), and ignore complex transient dynamics during range shifting 92 (Lawler et al., 2013, Zurell et al., 2009).

To address these issues, more mechanistic approaches of modelling range dynamics have
been advocated (Gallien *et al.*, 2010, Thuiller *et al.*, 2008) and several frameworks have been
developed (or revived) (Ehrlén & Morris, 2015, Lurgi *et al.*, 2015, Pereira *et al.*, 2010).

96	While all of them attempt to overcome the limitations associated with SDMs, their ability to
97	improve projections for species' range dynamics has never been compared systematically.
98	The lack of such evaluation is likely due to the difficulty to get appropriate benchmark data,
99	consisting of information on distribution, abundance and demography. Such complex datasets
100	are rare and benchmarking may be hampered if not all processes occurring in these
101	ecosystems are fully understood (Cheaib et al., 2012). We propose to conduct benchmarking
102	of new methods for modelling range dynamics using simulated community data, which allows
103	controlling all relevant ecological processes driving species distribution and track transient
104	dynamics in space and time ("virtual ecologist approach", Zurell et al., 2010).
105	We compare five generic modelling frameworks for predicting range dynamics capable of
106	(fast) calibration for any single species (Fig. 1). The choice of models was guided by our
107	objectives to include frameworks that are representative of current approaches for predicting
108	actual abundance (Ehrlén & Morris, 2015), that differ in their underlying assumptions and in
109	the complexity of data and process detail needed to parameterise them, and for which we
110	could find proficient users to run the simulations for our study to assure a fair comparison.
111	Due to a lack of common terminology, we refer to all five models (including SDMs) as range
112	models and to those models that explicitly consider dynamic processes of range shifts
113	(dispersal, population dynamics) as range dynamic models (Fig. 1). A subset of these models
114	relies on SDMs to predict habitat suitability and infer demographic rates ('SDM hybrids').
115	One approach infers demographic rates directly from the data and models habitat suitability as
116	an outcome of demographic processes (Fig. 1).
117	We explicitly focus our comparison of models on predicting range dynamics of single species.
118	Virtual species, however, were simulated within a virtual community to imitate constraints on

119 species distribution and abundance resulting from both abiotic factors and biotic interactions.

- 120 We first simulated virtual communities using a dynamic, individual-based, multi-species

121 simulation model, and then tracked these communities under climate change. Range models 122 were calibrated using data from the output of the virtual community model. We tested the 123 effects of different demographic (dispersal, source-sink dynamics) and community processes 124 (single species, species sorting, neutral dynamics) as well as of environmental stochasticity on 125 model predictive performance. 126 By comparing the performance of the different range models before and during climate 127 change we aim at answering the following questions: (1) Do SDMs and different range 128 dynamic models predict current species' distribution and abundance equally well? (2) Do 129 more complex, demography-based approaches consistently outperform SDMs under climate 130 change? (3) How is predictive model performance affected by prevailing demographic and 131 community processes? (4) What are the effects of structural decisions during model building? 132 We found that more complex range dynamic models improved current range projections 133 considerably compared to purely correlative SDMs. Under climate change, simpler range 134 dynamic models often outperformed more complex models, especially when benchmarking 135 data were simulated with more complex demographic and community processes. We discuss 136 guidelines and perspectives for increasing the reliability of climate change-induced range 137 predictions and for applying range dynamic models more widely in climate impact 138 assessments.

139 Material and methods

140 Overview of range dynamic models and hypotheses

141 We compare a classical SDM and four different range dynamic models, three of which are

- 142 SDM hybrids (Fig. 1, Appendix S1 in Supporting Information). The simplest SDM hybrid,
- 143 MigClim, supplements SDM predictions with distance-dependent colonisation probabilities
- 144 (Engler & Guisan, 2009, Normand et al., 2013). However, local demographic processes
- 145 including regeneration and mortality are not explicitly accounted for although these are

146 crucial for predicting population viability and spread rates. More complex SDM hybrids 147 couple SDM-derived habitat suitability maps and population models (Dullinger et al., 2012, 148 Keith et al., 2008, Zurell et al., 2012b). These models can be calibrated with simple 149 demographic information as we demonstrate with DemoNiche (Nenzén et al., 2012) or can be 150 inversely fitted to abundance data as demonstrated by the application of LoLiPop (Cabral & 151 Schurr, 2010). 152 The value of such SDM hybrids is debated, mainly because of potential circularity problems 153 (Gallien et al., 2010) and because the relation of SDM-derived habitat suitability to species 154 demographic parameters remains unclear (Thuiller et al., 2014). Dynamic range models 155 (DRM) have been introduced to overcome these issues. They do not rely on SDM output and 156 directly relate demographic rates to environmental factors and simultaneously estimate 157 parameters of the population model and the demography-environment relationship from 158 abundance and distribution data (Pagel & Schurr, 2012). We hypothesise that predictive 159 performance under climate change will increase from left to right in Fig. 1 because (i) range 160 dynamic models explicitly model the dispersal process and should hence outperform simple 161 SDMs, (ii) population dynamic models (DemoNiche, LoLiPop, DRM) additionally model 162 abundance dynamics from differently resolved demographic processes and should thus 163 outperform MigClim that only simulates potential colonisation, and (iii) DRMs jointly 164 estimate the effects of dispersal and demography on distribution and abundance dynamics and

- 166 Virtual species/communities

167 Simulation model

165

- 168 An individual-based, spatially-explicit, stochastic model (IBM; Gravel et al., 2006,
- 169 Münkemüller et al., 2012) was adapted to simulate the dynamics of a focal species and its co-
- 170 occurring competitors in heterogeneous environments. The IBM is a cellular automaton in

should thus outperform SDM hybrids which may suffer from circularity problems.

171	which each sub-cell is characterised by unique environmental conditions (temperature and
172	soil moisture) and can support one sessile individual. In the following, we provide a simple
173	overview over the main characteristics of the simulation model, while more details are given
174	in Appendix S2.
175	The IBM's spatial resolution only allows individual counts, yet the range models work at the
176	population level. We therefore implemented a hierarchy of two spatial scales so that the
177	sessile individuals are interacting locally but that aggregations at larger scale (10×10 sub-
178	cells) provide information on community composition and species abundance, which serve as
179	input for the range model comparison. One time step corresponds to one year and the four
180	main processes within one time step are large-scale, contagious disturbances at the coarse
181	resolution, and then the local (sub-cell) processes of adult mortality, propagule supply, and
182	recruitment of propagules to adult size (Fig. S1).
183	(1) Large-scale contagious disturbances like fire or windthrow act at the coarse scale with
184	an overall probability of 0.05, resulting in stochastic population dynamics and
185	incomplete range filling. Disturbances were omitted for some scenarios (see
186	Simulations), and are initiated in randomly selected cells at the coarse scale and spread
187	to the eight nearest neighbours. In response to disturbance, individuals are killed with
188	a probability of 0.9.
189	(2) Within each time step, adults die with a probability of 0.1 and can be replaced by
190	recruits from the local community or by immigrants.
191	(3) Propagule supply is determined by offspring production and by propagule dispersal.
192	Only adults that are at least one time step old can produce propagules. Offspring
193	production rate is determined by the local (temperature and moisture) environment,
194	and is described by a two-dimensional Gaussian function for the reproductive niche,

Page 10 of 38

195	with each species having a unique optimum. The Gaussian function is cut off at a
196	threshold of 0.001 to obtain finite niche breadth. Dispersal is simulated using a
197	negative exponential dispersal kernel where the rate parameter α determines the mean
198	dispersal distance $(1/\alpha)$.
199	(4) Recruitment into empty space or by competitive replacement follows a lottery function
200	of dispersal-driven propagule supply and interspecific differences in competitive
201	ability. Competitive performance depends on the species' niche, in particular the
202	probability λ of propagules to survive in the understory of adults prior to recruitment,
203	thus mimicking simple resource competition. An individual's competitive
204	performance (the probability of a propagule to establish and replace the resident) is
205	proportional to the ratio between its survival probability λ and that of the resident (or
206	between λ and a threshold of 0.1 in empty cells). The survival probability λ is a
207	function of the local environment, and we assumed a two-dimensional Gaussian
208	function for the survival niche (driven by soil moisture and temperature, see section
209	'Simulations and sampling' for more detail), which is equivalent to the reproductive
210	niche, if not mentioned otherwise. The cells can stay empty, if the overall propagule
211	rain is too small.

The species' fundamental niche is equivalent to the reproductive niche while the realised 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.

216 Simulations and sampling

The virtual communities were simulated in artificial landscapes of 20x125 (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

220	was represented as a linear latitudinal gradient with added spatially auto-correlated noise.
221	Sub-scale environmental heterogeneity at the scale of the 10×10 sub-cells was added as
222	normally distributed noise to the coarse-scale cells' mean temperature and moisture. We
223	simulated temporal variability in temperature by adding for each time step a temporally auto-
224	correlated random deviate to temperature.
225	Simulations were initiated by randomly distributing the virtual species in their respective
226	suitable habitats. First, simulations were run for 900 spin-up years under current
227	environmental conditions and variability to ensure that species/communities were in
228	(dynamic) equilibrium with the environment. After the end of the spin-up period (hereafter
229	referred to as year 0), climate change was initiated with a linear increase in temperature of 3°
230	Celsius over a period of 100 years. Soil moisture patterns were assumed to remain constant
231	over the simulation period.
232	We ran scenarios for three different community types: (1) neutral dynamics, (2) species
233	sorting, and (3) single species without any competitors. The first two community simulations
234	consisted of nine interacting species. For the species-sorting scenarios, the niche optima of the
235	eight co-occurring species in environmental space were symmetrically arranged around the
236	focal species niche with a fixed radius (Table S1, Fig. S4). In the neutral scenarios, all species
237	had equivalent niche optima and widths. All species within a community were assumed to
238	have equal dispersal ability.
239	We ran four scenarios for each type of community dynamics:
240	(1) SR: short-range dispersal without large-scale disturbances (with α =1/mean dispersal
241	distance=0.05).
242	(2) LR: like SR but with long-range dispersal (LR; α =0.1).

243 (3) SR+sinks: source-sink dynamics where the reproductive niche was narrower (by 1/3)

than the survival niche resulting in the realised niche being larger than the

fundamental niche (Fig. S4).

- 246 (4) SR+cont: with large-scale contagious disturbances.
- 247 Overall, we ran 12 different scenarios as input for the subsequent range model
- 248 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).
- 250 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).

255 Calibration of range models

256 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-maximising threshold (true

skill statistic; Allouche et al. 2006).

261 MigClim combines habitat suitability or presence/absence predictions from SDMs with a

dispersal kernel to predict colonisation probabilities (Engler & Guisan 2009). Here, habitat

- suitability was rescaled to range between 0 and 1 and a sigmoidal relationship was used to
- 264 relate habitat suitability to colonisation probability. Other structural relationships were tested
- and are discussed in Appendices S1 and S4. As dispersal kernel we took the known dispersalkernel from the IBM.

Global Change Biology

267	DemoNiche is a stage-structured matrix population model that constrains demographic rates
268	or carrying capacity by habitat suitability (given by an SDM) and connects local populations
269	by a dispersal kernel (Nenzén et al. 2012). Thereby, choices have to be made regarding the
270	demographic property (vital rates or carrying capacity) that should be constrained and the
271	relationship to habitat suitability (linear, sigmoidal, threshold). Here, we evaluated predictive
272	performance of different model configuration for year 0, and averaged predictions of the 5
273	best model configurations (see Appendix S1). Appendix S4 provides a more detailed
274	discussion on effects of structural uncertainty. We used a box-constraint variable metric
275	algorithm to calibrate the transition matrix by minimising differences between observed and
276	predicted mean demographic rates (Appendix S1). Calibrating the transition matrix was
277	necessary because the induced functional relationship between habitat suitability and
278	demographic transition probabilities implicitly assumes that demography is known for
279	optimal conditions (habitat suitability = 1) while demographic rates are usually averaged
280	across the population. As dispersal kernel we took the known dispersal kernel from the IBM.
281	LoLiPop simulates local population dynamics with different populations connected by two-
282	dimensional dispersal kernels (Cabral & Schurr 2010). Population dynamics can only take
283	place on suitable cells (given by SDM). Here, local population dynamics followed a
284	Beverton-Holt model extended with Allee effects (Cabral & Schurr 2010). Demographic
285	parameters were estimated from spatial abundance data from year 0 using maximum
286	likelihood estimation. As dispersal kernel we took the known dispersal kernel from the IBM.
287	The DRM is based on a simple stochastic model of local population dynamics (Ricker model)
288	coupled by a negative exponential dispersal kernel. The demography-environment
289	relationship is formulated as a regression of intrinsic population growth rate against the
290	environmental variables and the carrying capacity is assumed to likewise vary proportional to
291	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

295 (Pagel & Schurr 2012).

296 Assessing range model performance

297 Models were calibrated under equilibrium conditions and were then used to project species 298 distribution and abundance under climate change. Predictive performance was evaluated 299 against full known truth for each time step within the 100 years of climate change. We 300 assessed accuracy of spatial predictions by first converting all predictions to binary maps and 301 then calculating TSS (Allouche et al., 2006). Correlations between observed and predicted 302 abundance as well as predicted occurrence probabilities were assessed using Spearman's rank 303 correlation coefficient Rho. Further, we compared positions of range front, centre, and rear 304 edge by calculating quantiles (95%, 50%, and 5%, respectively) of the observed and predicted 305 latitudinal positions. Last, we calculated relative deviation in total abundance (summed over 306 all cells) as difference between predicted and observed abundance divided by observed 307 abundance (except for SDMs and MigClim), and differences in relative abundance estimates, 308 which is the relative decrease in observed and predicted abundance since year 0.

309

310 **Results**

311 Stochastic community IBM

312 The twelve different scenarios led to distinct spatial and temporal distribution of presence-

313 absence and abundance of the focal species (Appendix S2). Generally, short-range dispersal

- 314 resulted in stronger spatial clustering of populations. Differences in spatial clustering for
- 315 long- and short-range dispersal were particularly pronounced in neutral communities and

316 large-scale contagious disturbances resulted in even patchier spatial distributions. Under 317 climate change, the focal species exhibited range shifts accompanied with distinct population 318 decreases. Also, for all scenarios the focal species showed distinct time lags in range filling 319 following climate change with dispersal limitations at the range front and persistence in 320 unsuitable conditions at the rear range edge (Fig. 2, Fig. S6). Dispersal limitations were more 321 pronounced in short-range dispersal scenarios and under biotic interactions. In the species-322 sorting scenario, competing species were blocking the range front, whereas in neutral 323 communities, dispersal success became more random due to strong spatial clustering. Overall, 324 these diverse distributions provided a solid basis for comparing the predictive performance of 325 SDMs and range dynamic models.

326 Range model performance under current conditions

327 For the observation period (year 0), DRMs best predicted the focal species' mean and 328 maximum abundances along the temperature gradient (Figs. 2, S7-S8). SDMs often slightly 329 overestimated occurrence probability at range margins, which propagated differently to the 330 SDM hybrids. For example, LoLiPop predicted local abundances near range margins quite 331 successfully but underestimated abundances at range centres, whereas DemoNiche 332 overpredicted local abundances across nearly the whole range. Correspondingly, DRMs 333 obtained highest TSS and Rho in year 0 for most scenarios although differences between 334 range models were minor except that DemoNiche achieved much lower TSS scores while 335 Rho scores were consistently high (Fig. 3). By contrast, LoLiPop and SDMs predicted range 336 positions under equilibrium conditions best while DemoNiche and DRMs predicted slightly 337 too large ranges. DRMs and LoLiPop approximated total abundance best while DemoNiche 338 generally overestimated abundances (Fig. 4).

339 Range model performance under a warming climate

340 Under climate change, prediction accuracy of SDMs decreased significantly because the 341 range shifting potential was greatly overestimated by the full-dispersal SDMs (Fig. 3), 342 especially under short-range dispersal (Figs. 5-6). All range dynamic models were able to 343 overcome these limitations to some extent, achieving higher TSS and Rho values than SDMs 344 and generally showing less deviation between observed and predicted range margins (Figs. 3, 345 5, Figs. S11-S27). Here, DRMs were outperformed by SDM hybrids though often only 346 marginally, with MigClim and DemoNiche showing considerably higher TSS scores, slightly 347 higher Rho scores and smallest average deviations from range centre and rear edge. LoLiPop 348 predicted lowest average deviations from range front. Again, total abundance was best 349 predicted by DRMs and LoLiPop, and overestimated by DemoNiche. However, in terms of 350 relative change in abundance, all population models produced more similar projections with a 351 tendency towards underestimating the relative abundance and, thus, overestimating extinction 352 risks (Fig. 4, Figs. S28-S29). On average, DemoNiche best predicted relative change in 353 abundance.

354 Effects of demographic and community processes

355 Prediction accuracy (TSS and Rho) of all range models decreased most under short-range 356 dispersal and for scenarios including biotic interactions (Fig. 6). Under neutral dynamics, 357 which produced the most complex range-shifting dynamics, the simple MigClim generally 358 performed best. Overestimation of future total abundance by DemoNiche was particularly 359 strong under neutral dynamics. Here, LoLiPop predicted total abundance best, while there 360 were no clear winners among the population models for the other cases. However, in some 361 scenarios, LoLiPop estimated quite extreme demographic rates that resulted in over-362 compensatory local population dynamics, most severely for the long-range dispersal, species 363 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.

366 Effects of structural uncertainties

367 Effects of structural uncertainty were very pronounced in all range dynamic models

368 (Appendix S4). For example, assuming a linear relationship between SDM-derived habitat

369 suitability and colonisation probability in MigClim led to misleading predictions of near-

370 complete colonisation of the entire grid in long-range dispersal scenarios (Figs. S30-S31). In

371 DemoNiche, prior knowledge of the demography-environmental relationships helped

372 reducing uncertainty considerably. Also, model configurations that achieved highest

accuracies under equilibrium conditions usually remained among the best for climate change

374 predictions. We found no clear differences between constraining demographic rates or

375 carrying capacities by habitat suitability in DemoNiche, although the latter showed reduced

accuracy in long-range dispersal scenarios (Figs. S32-S34).

377 **Discussion**

378 We took the challenge of benchmarking state-of-the-art range models of varying complexity 379 using a comprehensive set of simulated data that account for demographic and community 380 processes. Based on the comparison across these benchmark data and diverse model outputs, 381 our initial questions can be answered as follows: (1) Under current climate, more complex 382 range dynamic models like DRMs better fit the observed species distributions and 383 abundances, although differences are small. (2) Under climate change, all dynamic 384 approaches improved predictions over simple SDMs. We could not, however, identify a 385 single, best approach for making predictions. Predictions of absolute abundance differed 386 markedly between population models while predictions of relative abundance were similarly 387 accurate. (3) Differences in dispersal ability and complex biotic interactions may introduce 388 high uncertainty in range predictions, while the effects of source-sink dynamics and increased

389	disturbance were minor. (4) In all range dynamic models, structural decisions during model
390	building can have great impact on model accuracy, but prior system knowledge on important
391	processes can reduce these uncertainties considerably. Our results reassure the clear merit in
392	using dynamic approaches for range predictions. But they also emphasise several needs,
393	namely: (i) to compare and combine multiple model outcomes for better capturing the
394	uncertainty associated with range predictions under climate change; (ii) to gather more and
395	higher quality data on species' demography; (iii) to run preliminary tests with each
396	demographic model in order to optimise the structural decisions and settings; and (iv) to
397	incorporate assembly processes for better capturing the within-community dynamic processes
398	and their constraints on range dynamics. In summary, implementing these insights will greatly
399	help advancing our ability of predicting future range dynamics and making these approaches
400	operational for larger numbers of species.

401 Range dynamic models on a par for current climates

402 Under current climates, DRMs provided the best fit in most cases, although their advantages 403 in prediction accuracies were generally small compared to other range dynamic models and 404 also compared to SDMs, which were not consistently outperformed by the more dynamic 405 approaches. DRMs jointly estimate the demography-environment relationship, dispersal and 406 other demographic parameters and, thus, avoid possible circularities that might arise from the 407 fitting steps for SDM hybrids (Gallien et al., 2010, Pagel & Schurr, 2012). Yet, for the 408 source-sink scenarios that could be assumed to be problematic for SDM hybrids while not for 409 DRMs (Pagel & Schurr, 2012, Schurr et al., 2012), we did not find evidence that DRMs 410 generally outperform hybrids. 411 Separate fitting steps in SDM hybrids of first fitting SDMs and then fitting the population

412 model may lead to bias when species are not in equilibrium with their environment. However,

413 if species are dispersal limited in some parts of their range but the entire niche in

414 environmental space is well-covered by data, as is the case for our source-sink and contagious 415 disturbance scenarios, circularity in SDM hybrids appear not to be a limiting problem. If, on 416 the other hand, observed species' distributions are biased by dispersal limitations such that 417 parts of the environmental niche are not filled, then DRMs may be the preferred choice, as 418 this approach directly accounts for the (limited) dispersal in the estimation of the species' 419 niche and can thereby reduce the bias that may otherwise arise from the disequilibrium (Pagel 420 & Schurr, 2012). Still, more research is needed to (i) develop robust approaches for assessing 421 limitations in range filling a priori (Svenning & Skov, 2004) as an important model 422 assumption of SDM hybrids, (ii) to assess how prevalent the phenomenon of incomplete 423 range filling is in extant species, and (iii) to evaluate DRM predictive performance and their 424 ability to accurately distinguish between environmental filtering and dispersal limitations in 425 such situations.

426 Large variation in predictive performance under future climates

427 All range dynamic models tested here considerably improved climate change projections 428 compared to SDMs, although relative model performance differed from those under current 429 climates. Surprisingly, advantages of DRMs in predicting current ranges did not result in 430 better predictions of future ranges compared to SDM hybrids. This may result from different 431 calibration approaches, calibration data and process detail covered by the models. For 432 example, the relatively poor performance of DRMs might arise to some extent from 433 uncertainty in estimated dispersal rates while dispersal kernels were known for SDM hybrids. 434 Also, DRMs overestimated the position of future trailing range edges (Figs. 3,6). This likely 435 originates from the fact that DRMs did not explicitly describe adult survival, which is 436 independent from the environment in the IBM, but summarised all demographic processes in 437 an environment-dependent population growth rate. Therefore, the transient persistence of 438 populations when temperatures have become unfavourable at the rear edge (Fig. 2) was not 439 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.

443 The main advantage of the more complex approaches is that they also predict population 444 dynamics and associated extinction risks. Both before and after climate change simulations, 445 DemoNiche largely overestimated abundance, while LoLiPop and DRMs produced more 446 reasonable estimates. Nevertheless, the relative decrease in abundance and associated 447 extinction risks were often better approximated by DemoNiche than by the other two 448 population dynamic models. This is likely due to the fact that DemoNiche uses demographic 449 instead of abundance data for calibration, and that DemoNiche does not restrict abundance 450 except when habitat suitability was related to carrying capacity. Calibrating demographic 451 parameters directly allows DemoNiche to more accurately predict population trajectories in 452 many cases. In fact, calibration on abundance data led to estimation of partially unrealistic 453 demographic rates in LoLiPop resulting in over-compensatory local population dynamics 454 (although this could be avoided by choosing an alternative underlying population model; 455 Cabral & Schurr 2010). Although also calibrating on abundance data, this effect was not 456 found in DRMs, as DRMs use abundance time series for calibration that inherently include 457 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 indispensible 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

465 abundance data may exhibit high uncertainty in future predictions and require careful testing 466 of model structure. A constraining point is that accurate data on demography and spatial 467 distribution and abundance are not available for many species and in high quality. Also, 468 density is not easy to measure in many plant communities, where biomass or relative cover is 469 preferentially recorded. Time series of relative cover may show strong inter-annual 470 variability, which is not necessarily related to population processes (Boulangeat *et al.*, 2012). 471 Thus, we do not only need to increase our efforts into data collection but also in defining 472 (more) meaningful response variables for population dynamic studies.

473 Effects of demographic and community processes

474 Generally, spatial prediction accuracies of all range models decreased when the focal species

475 was interacting with other competitor species, especially under short-range dispersal (Fig. 6).

476 Contrary, the effects of large-scale disturbances and source-sink dynamics on model accuracy

477 were comparably low. None of the tested range dynamics models account for biotic

478 interactions and hence they all experienced difficulties in these particular cases. More work is

479 thus needed to incorporate assembly processes in such dynamic models (Boulangeat et al.,

480 2012, Cabral & Kreft, 2012, Mokany et al., 2012). Therefore, caution is advised with these

481 models when biotic interactions are highly stochastic as was shown in the neutral scenarios.

482 Although the assumption of neutral community dynamics is much debated for temperate

483 ecosystems, observed spatial distributions are often astoundingly consistent with neutrality

484 even if driven by different mechanisms (Bell, 2005). Our results indicate that whenever the

485 ecosystem under study is strongly affected by biotic or environmental stochasticity, simpler

486 range models such as MigClim may be preferred over more complex population models

487 although prediction of population dynamics and abundances cannot be retrieved from such

488 models.

489 We additionally stress that although predictive performances of all models decreased stronger 490 in the short-range compared to long-range dispersal scenarios, this does not imply that 491 predictions are generally more reliable for long-distance dispersers. Our results need to be 492 judged with some caution and with respect to the virtual simulation model setup, in which all 493 species of the community had the same dispersal ability and recruitment was proportional to 494 the amount of propagule rain. This is, of course, a simplifying assumption and we will likely 495 observe even more complex community dynamics if species vary in dispersal ability (Cabral 496 & Kreft, 2012). It is, thus, reasonable to assume that long-distance dispersers may also 497 experience substantial migration limitation from competitors.

498 Uncertainty through structural decisions in range dynamic models

499 Most applications of range dynamic models ignore uncertainty in model structure and their 500 effect on prediction uncertainty (but see Cabral & Schurr, 2010). Our results clearly show 501 that structural decisions in model building are crucial and may strongly affect prediction 502 accuracies. Range dynamic models necessarily simplify the species-environment relationship 503 as well as the colonization and extinction processes and these simplifying assumptions may 504 lead to large uncertainties (Appendix S4). Important structural decisions in range dynamic 505 models relate to differences in the relationship between habitat suitability and demographic 506 rates, which has hitherto little empirical support (McGill, 2012, Thuiller et al., 2010, Thuiller 507 et al., 2014) and should receive more attention in the future. The different range dynamic 508 models may strongly differ in how variation in the environment-dependent demographic rate 509 drives variation in (i) local abundance or carrying capacity, (ii) persistence, and (iii) propagule 510 dispersal and associated colonisation success. This aspect becomes even more important, 511 when such models are applied in a climate change context, where different sources of 512 uncertainty need to be considered in order to make meaningful projections with sufficient 513 attention given to the sources of uncertainties.

514 Such structural aspects are rarely considered in dynamic range predictions, but we strongly 515 advise to do so and to assess to what degree prediction accuracies vary under different model 516 structures. In DemoNiche, for example, using habitat suitability to constrain carrying capacity 517 led to more realistic abundance estimates in some scenarios, but caused higher uncertainty 518 and erroneous predictions of extinction in other scenarios. Constraining vital rates seemed 519 more robust in that respect, but only if appropriate vital rates were selected according to prior 520 knowledge of the species' ecology. Overall, we found that the DemoNiche configurations that 521 achieved highest accuracy under equilibrium conditions usually remained among the best 522 during climate change as well. Accordingly, the best model configurations under current 523 climate could be used for making consensus forecasts. Alternatively, approximate Bayesian 524 computation might be employed to optimise the structural link between the habitat suitability 525 and demography in SDM hybrids given the data (Hartig et al., 2011). This idea is similar to 526 the information criterion based approach used by Cabral and Schurr (2010) in LoLiPop in 527 order to select among different underlying population models including or not Allee effects 528 and overlapping generations. Notably, the environment-demography relationships in LoLiPop 529 and DRMs, in particular more differentiated environmental responses of different 530 demographic processes, should also be explored more thoroughly in the future, although such 531 assessment will require larger computational efforts. Nevertheless, we want to stress that the 532 underlying idea of DRMs of jointly estimating the different constraints on the niche, namely 533 environmental limitation, population dynamics and dispersal, is better integrated with 534 ecological theory than hybrid approaches (for discussion see Schurr et al., 2012).

535 Limitations and extensions

536 Clearly, even a model intercomparison using simulated data can by no means be exhaustive

537 (Zurell et al., 2010). Our choice of assembly processes, dispersal and other demographic

- 538 processes, and landscape structure represents only one possible implementation and is still
- 539 highly simplified in many aspects. Nevertheless, such approach allowed for generating

540	consistent benchmarking data, producing an array of important demographic and community
541	processes that are known to affect species' range limits. Therefore our results provide insight
542	into comparative model performance in a wide subset of potential cases. Clearly, subsequent
543	comparison on field data and evaluation of other complicating factors such as sampling effort
544	or bias, unequal dispersal abilities and unequal competitive ability of heterospecifics among
545	others will be additionally informative. Also, the robustness of range dynamic models under
546	novel environments (Williams & Jackson, 2007, Zurell et al., 2012a), changing collinearity
547	structures in environmental predictors (Dormann et al., 2013) or arbitrary scale decisions
548	(Thuiller et al., 2015) should be tested in the future.
549	Moreover, although we aimed for a representative selection of current modelling frameworks
550	for predicting range dynamics and actual abundance (Ehrlén & Morris, 2015), our study
551	could only include a subset of available software applications. Lurgi et al. (2015) provide an
552	extensive review on available computer platforms for predicting population- or individual-
553	based range dynamics, which vary in the accommodated detail of demographic processes and
554	complexity in species' lifecycles. In comparison to some other platforms, for example
555	RAMAS (Akçakaya, 2000) that has been used rather widely in conservation context (cf.
556	Fordham et al., 2013) and more recently also for predicting range dynamics (e.g., Anderson et
557	al., 2009, Keith et al., 2008, Pearson et al., 2014, Swab et al., 2015), DemoNiche, as our
558	example of a stage-structured matrix population model, allows only a relatively low level of
559	complexity in the species' modelled life history. As the life cycle of the simulated species was
560	likewise simple, this is unlikely to have limited the performance in our model. However, for
561	other applications that demand a higher level of detail in the species' life cycle or in
562	environmental drivers, other approaches, such as RAMAS, might be preferable (Lurgi et al.,
563	2015).

564 Summary

565 Our range dynamic model intercomparison yielded no clear winners or losers. While all range 566 dynamic models show clear benefits over simple SDMs, we cannot provide simple 567 suggestions which model framework to choose for any single application. Currently, model 568 choice will depend to a large extent on data availability and on prior knowledge on species' 569 ecology (Lurgi et al., 2015). For example, we currently lack spatiotemporal abundance data or 570 solid demographic information for many taxa, which clearly limits model choice. As far as 571 possible given data limitation, we generally advise a comparison of predictions from multiple 572 models for improved understanding of model behaviour and prediction uncertainty (Cheaib et 573 al., 2012). Thereby, great discrepancies between model predictions may indicate that we 574 missed important ecological mechanisms. More efforts are needed to better understand the 575 underlying mechanisms and its calibration in range dynamic models. This is specifically true 576 for the interplay between demographic rates and biotic interactions in communities. Also, 577 uncertainty through structural decisions should be assessed more routinely, and important 578 model assumptions of range models should be verified *a priori*, for example the degree of 579 range filling. Clearly, broader application of range dynamic models is limited by data and by 580 computational efforts. Computation times are still quite high for DRMs compared to SDM 581 hybrids. However, data availability is more crucial. We have shown that different kinds of 582 calibration data (abundance, demographic rates) can be utilised, and that also SDM hybrids 583 can be calibrated in a (semi-) automated way. Still, more efforts should be given to collecting 584 longer-term and large-scale data on abundance and demography. Only such consistent data 585 basis can ensure wide application of range dynamic models for climate impact assessment. 586 Future studies should further focus on evaluating the effects of sampling effort and sampling 587 bias, and of other complicating processes such as asymmetric competition or niche evolution 588 on prediction accuracy of range dynamic models.

589

590 Acknowledgements

591 We thank Laure Gallien, Frank Schurr, Boris Schröder, Jane Elith and Antoine Guisan for 592 helpful discussions. This study arose from two workshops entitled 'Advancing concepts and 593 models of species range dynamics: understanding and disentangling processes across scales' 594 in 2012-2013. Funding was provided by the Danish Council for Independent Research – 595 Natural Sciences (grant no. 10-085056 to SN). DZ acknowledges funding from the People 596 Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme 597 (FP7/2007-2013) under REA grant agreement no. 624958. WT, TM and KS received funding 598 from the European Research Council under the European Union's Seventh Framework 599 Programme (FP7/2007-2013) Grant Agreement no. 281422 (TEEMBIO). The LECA is part 600 of Labex OSUG@2020 (ANR10 LABX56). JP acknowledges financial support from the 601 German Research Foundation (DFG, grant SCHU 2259/5-1). JSC acknowledges financial 602 support from German Research Foundation (DFG, via SA-21331 and the German Initiative of 603 Excellence). DG was supported by a NSERC-Strategic grant. SD acknowledges support from 604 the Austrian Panel on Climate Change (project B175127, SPEC-ADAPT). KS acknowledges 605 support from the EU COST Action 'HarmBio' (ES1101). NEZ acknowledges support form the 606 Swiss National Science Foundation SNSF (grant: 31003A 149508/1).

607

609	
610	References
611	Akçakaya HR (2000) Viability analyses with habitat-based metapopulation models.
612	Population Ecology, 42 , 45-53.
613	Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution
614	models: prevalence, kappa and the true skill statistic (TSS). Journal of Applied
615	Ecology, 43 , 1223-1232.
616	Anderson BJ, Akçakaya HR, Araújo MB, Fordham DA, Martinez-Meyer E, Thuiller W,
617	Brook BW (2009) Dynamics of range margins for metapopulations under climate
618	change. Proceedings of the Royal Society B, 276, 1415-1420.
619	Beissinger SR, Westphal MI (1998) On the Use of Demographic Models of Population
620	Viability in Endangered Species Management. Journal of Wildlife Management, 62,
621	821-841.
622	Bell G (2005) The co-distribution of species in relation to the neutral theory of community
623	ecology. Ecology, 86, 1757-1770.
624	Bellard C, Bertelsmeier C, Leadley P, Thuiller W, Courchamp F (2012) Impacts of climate
625	change on the future of biodiversity. Ecology Letters, 15, 365-377.
626	Boulangeat I, Gravel D, Thuiller W (2012) Accounting for dispersal and biotic interactions to
627	disentangle the drivers of species distributions and their abundances. Ecology Letters,
628	15 , 584-593.
629	Cabral JS, Kreft H (2012) Linking ecological niche, community ecology and biogeography:
630	insights from a mechanistic niche model. Journal of Biogeography, 39 , 2212-2224.
631	Cabral JS, Schurr FM (2010) Estimating demographic models for the range dynamics of plant
632	species. Global Ecology and Biogeography, 19, 85-97.

633	Cheaib A, Badeau V, Boe J et al. (2012) Climate change impacts on tree ranges: model
634	intercomparison facilitates understanding and quantification of uncertainty. Ecology
635	Letters, 15 , 533-544.
636	Dormann CF, Elith J, Bacher S et al. (2013) Collinearity: a review of methods to deal with it
637	and a simulation study evaluating their performance. Ecography, 36 , 27-46.
638	Dormann CF, Schymanski SJ, Cabral J et al. (2012) Correlation and process in species
639	distribution models: bridging a dichotomy. Journal of Biogeography, 39 , 2119-2131.
640	Dullinger S, Gattringer A, Thuiller W et al. (2012) Extinction debt of high-mountain plants
641	under twenty-first-century climate change. Nature Climate Change, 2, 619-622.
642	Ehrlén J, Morris WF (2015) Predicting changes in the distribution and abundance of species
643	under environmental change. Ecology Letters, 18, 303-314.
644	Engler R, Guisan A (2009) MigClim: Predicting plant distribution and dispersal in a changing
645	climate. Diversity and Distributions, 15, 590-601.
646	Fordham DA, Akçakaya HR, Araújo MB, Keith DA, Brook BW (2013) Tools for integrating
647	range change, extinction risk and climate change information into conservation
648	management. Ecography, 36, 956-964.
649	Gallien L, Münkemüller T, Albert CH, Boulangeat I, Thuiller W (2010) Predicting potential
650	distributions of invasive species: where to go from here? Diversity and Distributions,
651	16 , 331-342.
652	Gravel D, Canham CD, Beaudet M, Messier C (2006) Reconciling niche and neutrality: the
653	continuum hypothesis. Ecology Letters, 9, 399-409.
654	Guisan A, Thuiller W (2005) Predicting species distribution: offering more than simple
655	habitat models. Ecology Letters, 8, 993-1009.
656	Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology.
657	Ecological Modelling, 135, 147-186.

658	Hartig F, Calabrese JM, Reineking B, Wiegand T, Huth A (2011) Statistical inference for
659	stochastic simulation models - theory and application. Ecology Letters, 14, 816-827.
660	Holt RD, Keitt TH, Lewis MA, Maurer BA, Taper ML (2005) Theoretical models of species'
661	borders: single species approaches. Oikos, 108 , 18-27.
662	Keith DA, Akçakaya HR, Thuiller W et al. (2008) Predicting extinction risks under climate
663	change: coupling stochastic population models with dynamic bioclimatic habitat
664	models. Biology Letters, 4, 560-563.
665	Lawler JJ, Ruesch AS, Olden JD, Mcrae BH (2013) Projected climate-driven faunal
666	movement routes. Ecology Letters, 16, 1014-1022.
667	Lurgi M, Brook BW, Saltré F, Fordham DA (2015) Modelling range dynamics under global
668	change: which framework and why? Methods in Ecology and Evolution, 6, 247-256.
669	Mcgill BJ (2012) Trees are rarely most abundant where they grow best. Plant Ecology, 5, 46-
670	51.
670 671	51. Mokany K, Harwood TD, Williams KJ, Ferrier S (2012) Dynamic macroecology and the
670 671 672	51.Mokany K, Harwood TD, Williams KJ, Ferrier S (2012) Dynamic macroecology and the future for biodiversity. Global Change Biology, 18, 3149-3159.
670671672673	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 <i>et al.</i> (2012) From diversity indices to community
 670 671 672 673 674 	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 <i>et al.</i> (2012) From diversity indices to community assembly processes: a test with simulated data. Ecography, 35 , 468-480.
 670 671 672 673 674 675 	 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 <i>et al.</i> (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
 670 671 672 673 674 675 676 	 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 <i>et al.</i> (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.
 670 671 672 673 674 675 676 677 	 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 <i>et al.</i> (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 <i>et al.</i> (2013) A greener Greenland? Climatic potential
 670 671 672 673 674 675 676 677 678 	 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 <i>et al.</i> (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 <i>et al.</i> (2013) A greener Greenland? Climatic potential and long-term constraints on future expansions of trees and shrubs. Phil Trans R Soc
 670 671 672 673 674 675 676 677 678 679 	 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 <i>et al.</i> (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 <i>et al.</i> (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.
 670 671 672 673 674 675 676 677 678 679 680 	 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 <i>et al.</i> (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 <i>et al.</i> (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. Normand S, Zimmermann NE, Schurr FM, Lischke H (2014) Demography as the basis for

- Pagel J, Schurr FM (2012) Forecasting species ranges by statistical estimation of ecological
 niches and spatial population dynamics. Global Ecology and Biogeography, 21, 293304.
- 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.
- 687 Pereira HM, Leadley PW, Proença V *et al.* (2010) Scenarios for Global Biodiversity in the
- 688 21st Century. Science, **330**, 1496-1501.
- 689 Schurr FM, Pagel J, Cabral JS et al. (2012) How to understand species' niches and range

690 dynamics: a demographic research agenda for biogeography. Journal of

- 691 Biogeography, **39**, 2146-2162.
- 692 Sexton JP, Mcintyre PJ, Angert AL, Rice KJ (2009) Evolution and ecology of species range
 693 limits. Annual Review of Ecology, Evolution and Systematics, 40, 415-436.
- Svenning J-C, Normand S, Skov F (2008) Postglacial dispersal limitation of widespread forest
 plant species in nemoral Europe. Ecography, **31**, 316-326.
- 696 Svenning J-C, Skov F (2004) Limited filling of the potential range in European tree species.
- 697 Ecology Letters, 7, 565-573.
- 698 Swab RM, Regan HM, Matthies D, Becker U, Bruun HH (2015) The role of demography,
- 699 intra-species variation, and species distribution models in species' projections under
 700 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
- 703 Systematics, 9, 137-152.
- 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.

707	Thuiller W, Münkemüller	T, Lavergne S, Mouil	llot D, Mouquet N, S	Schiffers K,	Gravel D

- 708 (2013) A road map for integrating eco-evolutionary processes into biodiversity
 709 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.
- 714 Williams JW, Jackson ST (2007) Novel climates, no-analog communities, and ecological

surprises. Frontiers in Ecology and the Environment, **5**, 475-482.

- Zurell D, Berger U, Cabral JS *et al.* (2010) The virtual ecologist approach: simulating data
 and observers. Oikos, **119**, 622-635.
- 718 Zurell D, Elith J, Schröder B (2012a) Predicting to new environments: tools for visualising
- 719 model behaviour and impacts on mapped distributions. Diversity and Distributions,

720 **18**, 628-634.

- 721 Zurell D, Grimm V, Rossmanith E, Zbinden N, Zimmermann NE, Schröder B (2012b)
- Uncertainty in predictions of range dynamics: black grouse climbing the Swiss Alps.
 Ecography, 35, 590-603.
- 724 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**,
- 726 733-744.

729 Supporting information

- 730 Appendix S1: Range models descriptions
- 731 Appendix S2: Details and analysis of virtual community model
- 732 Appendix S3: Model comparison under climate change
- 733 Appendix S4: Structural uncertainty in range dynamic models



Figure 1. Main characteristics of implemented range models used for model inter-comparison. Habitat suitability maps (darker grey 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 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).

245x227mm (200 x 200 DPI)



Figure 2. Realised temperature range of focal species for the long- (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, realised 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.

153x110mm (200 x 200 DPI)



Figure 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 models. 81x152mm (200 x 200 DPI)



Figure 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 models. 81x60mm (200 x 200 DPI)



Figure 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. 214x101mm (200 x 200 DPI)



Figure 6. Prediction accuracies (grey) 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/N₀: absolute difference between observed and predicted relative abundance. Orange indicates overestimation and blue underestimation. Maximum circle sizes correspond to TSS=0.87, Rho=0.89, M=37 cells, N= 22, and N/N₀=57%.

225x246mm (200 x 200 DPI)