

Review

The predictive performance of process-explicit range change models remains largely untested

David E. Uribe-Rivera^{1,2} , Gurutzeta Guillera-Arroita^{1,2,3} , Saras M. Windecker^{1,2},
Patricio Pliscoff^{4,5,6}  and Brendan A. Wintle^{1,2} 

RUNNER-UP

¹School of BioSciences, Univ. of Melbourne, Melbourne, Australia

²School of Ecosystem and Forest Science, Univ. of Melbourne, Melbourne, Australia

³Pyrenean Inst. of Ecology (CSIC), Zaragoza, Spain

⁴Depto de Ecología and Inst. de Geografía, Pontificia Univ. Católica de Chile, Santiago, Chile

⁵Center of Applied Ecology and Sustainability (CAPEs), Pontificia Univ. Católica de Chile, Santiago, Chile

⁶Inst. de Ecología y Biodiversidad (IEB), Santiago, Chile

Corresponding author: David E. Uribe-Rivera Email de.uribe.r@gmail.com

Ecography

2023: e06048

doi: [10.1111/ecog.06048](https://doi.org/10.1111/ecog.06048)

Subject Editor: Damaris Zurell

Editor-in-Chief:

Jens-Christian Svenning

Accepted 26 June 2022



Ecological models used to forecast range change (range change models; RCM) have recently diversified to account for a greater number of ecological and observational processes in pursuit of more accurate and realistic predictions. Theory suggests that process-explicit RCMs should generate more robust forecasts, particularly under novel environmental conditions. RCMs accounting for processes are generally more complex and data hungry, and so, require extra effort to build. Thus, it is necessary to understand when the effort of building a more realistic model is likely to generate more reliable forecasts. Here, we review the literature to explore whether process-explicit models have been tested through benchmarking their temporal predictive performance (i.e. their predictive performance when transferred in time) and model transferability (i.e. their ability to keep their predictive performance when transferred to generate predictions into a different time) against simpler models, and highlight the gaps between the rapid development of process-explicit RCMs and the testing of their potential improvements. We found that, out of five ecological processes (dispersal, demography, physiology, evolution, species interactions) and two observational processes (sampling bias, imperfect detection) that may influence reliability of forecasts, only the effects of dispersal, demography and imperfect detection have been benchmarked using temporally-independent datasets. Only nine out of twenty-nine process-explicit model types have been tested to assess whether accounting for processes improves temporal predictive performance. We found no benchmarks assessing model transferability. We discuss potential reasons for the lack of empirical validation of process-explicit models. Considering these findings, we propose an expanded research agenda to properly test the performance of process-explicit RCMs, and highlight some opportunities to fill the gaps by suggesting models to be benchmarked using existing historical datasets.

Keywords: ecological forecast, model transferability, predictive performance, process-explicit models, range shift, species distribution models

Introduction

Effective management of biodiversity amidst ongoing and accelerating environmental change requires accurate predictions of species range and community composition change (Evans 2012, Petchey et al. 2015, Dietze et al. 2018). Several factors may affect the capacity of ecological models used to forecast range change (range change models; RCMs) to generate accurate predictions over time, ranging from the amount of data fed into the model, to the inclusion of mechanistic descriptions in the modelling framework. Understanding the consequences of these factors on the accuracy of RCM predictions would help users to select modelling approaches that optimize the reliability of ecological forecasts, and consequently better inform biodiversity conservation applications (Clark et al. 2001, Grenouillet and Comte 2014).

Studies suggest that implementation of mechanistic understanding in ecological models is key to achieve more reliable biodiversity forecasts (Cuddington et al. 2013, Fordham et al. 2014, Evans et al. 2016, Singer et al. 2016). At least five ecological mechanisms govern range dynamics: demography, physiological tolerance to the abiotic environment, evolution, dispersal and species interactions (Urban et al. 2016). In addition to these ecological mechanisms, observational processes such as sampling bias (bias driven by an uneven distribution of sampling effort across the environmental space) and imperfect detection (bias induced by recording false absences, with detection probabilities potentially varying across the environmental space) can have substantial impacts on the predictive performance of distribution models that are fitted to observational data (Elith and Leathwick 2009, Lahoz-Monfort et al. 2014, Mesgaran et al. 2014). Consequently, the rate of development of novel modelling frameworks that incorporate one or more of these processes explicitly (so-called 'process-explicit' models; see Box 1 for more detailed definitions) has risen steadily during the last few decades (Briscoe et al. 2019).

Despite the growth of process-explicit models, few benchmarking experiments have been conducted to test whether these more complex RCMs improve the reliability of ecological forecasts compared to simpler models (Sequeira et al. 2018). In this paper, we review the literature to evaluate the existing evidence on the impact of incorporating processes in RCMs on model temporal predictive performance and transferability, and highlight the gaps between the rapid development of process-explicit RCMs and the testing of their potential improvements. We discuss the reasons for these gaps in empirical validation of process-explicit approaches. Considering this information, we propose an expanded research agenda to test the improvement in temporal predictive performance and transferability of process-explicit RCMs, and highlight opportunities to fill the identified gaps by suggesting models to be benchmarked using existing historical datasets. Due to inconsistencies in relevant definitions in the literature and to facilitate communication, we also clarify some of these concepts in Boxes 1 and 2.

Challenges to assessing temporal predictive performance and transferability of RCMs

The main constraint on assessing the temporal predictive performance and transferability of RCMs is that at least two temporally independent datasets with enough temporal separation to observe range changes are required: one to train the model, and another to assess its predictive performance (Fig. 1). There are two ways to assess temporal predictive performance of RCMs: to use retrospective forecasting of observed changes with a temporally-independent dataset for external model evaluations (Araújo et al. 2005, Araújo and Rahbek 2006) or to use simulated data (Zurell et al. 2016). Authors seeking to forecast species distributions often deal with the lack of temporally independent model evaluation data by reporting predictive performance based only on internal model evaluations conducted using a subsample of the dataset (e.g. using structured cross-validation protocols; Roberts et al. 2017). However, predictive performance assessed this way may not hold over time, because predictive performance usually decreases when models are used to generate predictions for a different time (Dobrowski et al. 2011, Rapacciuolo et al. 2012, Uribe-Rivera et al. 2017). For instance, Morán-Ordóñez et al. (2017) reported an inverse relationship between the discrimination power of correlative species distribution models (cSDMs) and the length of time in which the model had been transferred.

Reporting predictive performance only for training conditions (i.e. internal evaluation; Fig. 1) provides an unrealistic picture of both temporal predictive performance and model transferability, i.e. the degree to which the predictive performance is maintained from training to transferred conditions (Boxes 1 and 2). We expect model transferability to be inversely related to the degree of extrapolation under transferred conditions (Fitzpatrick and Hargrove 2009, Mesgaran et al. 2014), and directly related to how well the model represents the study system (Fielding and Haworth 1995, Randin et al. 2006). Temporal transferability might decrease if the model specification fails to account for non-stationary relationships between independent and dependent variables through time (Dormann 2007). In such cases, model parameters (e.g. describing relationships between occupancy and environmental features) will not properly characterize true relationships in future times and model predictions will not be accurate (Blois et al. 2013, Gharari et al. 2013). Assessing improvement of both temporal predictive performance and model transferability when processes are explicitly modelled therefore requires temporally-independent data for evaluation.

Knowledge gaps

In the last 20 years, an increasing number of models have been developed to incorporate ecological and observational processes in the pursuit of more reliable predictions.

Box 1. Glossary

Accuracy: Degree of proximity between the predicted and the true value.

Hindcasting, forecasting and retrospective forecasting: These three terms describe model transference to generate predictions to a temporal window beyond that used to fit the model. Clarifying these concepts is necessary to avoid confusing terminology identified during the literature review, especially when referring to predictions made for the past. We use hindcasting to refer to generating predictions for an earlier point in time; forecasting to refer to transferring a model forward in time, commonly to generate predictions for future scenarios; and retrospective forecasting to refer to a specific type of forecasting where a model fitted to past data is transferred forward in time to generate predictions for the present or the recent past (Fig. 1).

Model extrapolation: Projection into novel regions of covariate space (Eliot and Leathwick 2009). The extrapolation of a model occurs when predictions are generated for a combination of values of the predictive variables not observed in the data used for model training (novel conditions).

Model transferability: Represents the ability of a model developed for a specific site and/or time and/or taxon to keep its predictive performance when it is transferred to generate predictions into a different time or place or for a different taxon. It is a measure of the generality of the model and can be measured as the difference/ratio of predictive performance under the training and the transferred conditions (Randin et al. 2006, Dobrowski et al. 2011, Heikkinen et al. 2012). A model with good transferability is a model with similar predictive performance in the training and new conditions. Transferability may depend on the degree of novelty of the new conditions (i.e. how different are the new conditions to which the model is transferred from the original conditions, that is, the degree of extrapolation; Mesgaran et al. 2014).

Model transference: The exercise of using a model to generate predictions beyond the range of observed conditions in which it was trained. There are three types of transference: in time, in space or across taxonomic groups. It may result in model extrapolation.

Precision: Amount of uncertainty in the predictions. Also known as sharpness. Precise predictions convey low uncertainty (i.e. are closely grouped around the mean prediction). Imprecise predictions convey high uncertainty.

Predictive performance: A measure of how useful a prediction is, usually some function of accuracy and or precision. Also known as predictive skills (modified from Petchey et al. 2015).

Process-explicit models: Models that explicitly account for potentially relevant processes when predicting species' ranges. Definitions for this concept are not consistent in the literature, and usually refer to ecological processes only, including demography, physiology, dispersal, evolution and species interactions. We acknowledge that there is a range of possible ways to account for processes into an ecological model.

Range change models (RCM): Here we define range change models as a group of ecological models designed, used or potentially useful to predict species redistribution and community reassembly (as an emergent property of a single model or derived from stacking individual species predictions). This implies that we do not consider as RCM those models unable to make predictions of the distribution for every considered species individually (e.g. macroecological models, or community-level models with the 'assemble first predict later' strategy, sensu Ferrier and Guisan 2006).

Temporal predictive performance: The performance of predictions made beyond the temporal observation window (i.e. when the model is transferred in time beyond the training conditions). This concept differs from 'temporal transferability', which represents the relative decay in predictive performance with respect to the training conditions when models are transferred in time (Box 2).

Uncertainty: For predictive modelling, the spread of the distribution of predictions around the mean predicted value.

Here we review the literature to determine what evidence we have for the improvement in temporal predictive performance or transferability of RCMs that include ecological or observational processes.

Literature review methodology

We used combinations of three search terms on the Web of Science database ('forecast*' OR 'predict*' + 'chang*' OR

'range-shift*' + 'species-distribution*' OR 'species-composition*') and included publications from January 2006 to December 2020. We excluded studies that generated predictions backwards in time because the processes involved in species range change are directional: temporal predictive performance and transferability may not be equivalent for forecasting and hindcasting (Rubidge et al. 2011, Tapia et al. 2018). We focused on decadal range changes and did not incorporate evidence from studies with a broader temporal

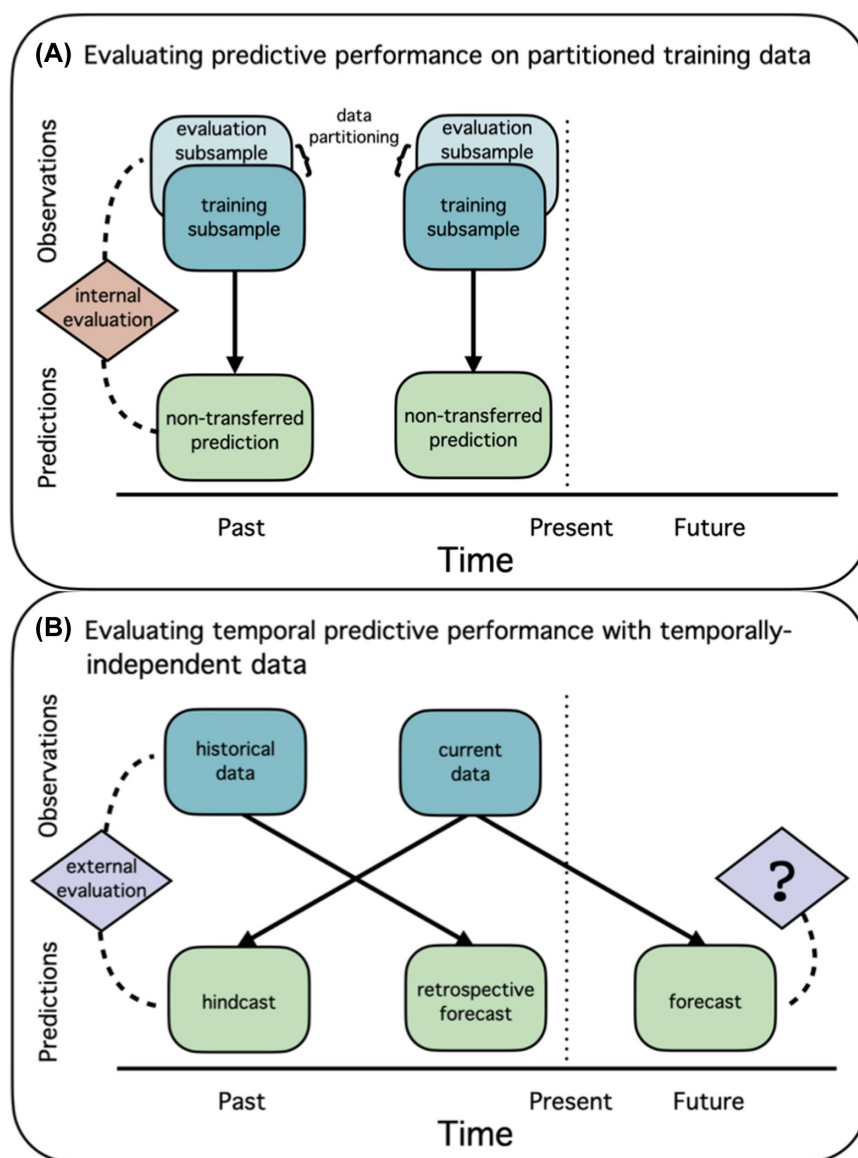


Figure 1. Schematic illustration of generating and evaluating predictions of species ranges across time (x-axis). Observations of the biological phenomena (i.e. observed occupancy or abundance patterns; blue boxes) can come from several times in the past. Model predictions (green boxes) are obtained from modelling the observed data (solid arrows) and can be generated with (B) or without (A) temporal transference. Dashed lines represent comparisons between model predictions and observed data that allow us to do model evaluation (diamonds). Most range change models are designed with the objective of producing forecasts into the future, but these forecasts cannot be evaluated as they predict to a time period for which there are no observed data (B). Given the constraint of temporally independent data for model evaluation of forecasts into the future, the predictive performance is often assessed through data partitioning of the training dataset: a subsample of the biological data not used for model training is used to evaluate non-transferred predictions (Internal evaluations; A). In contrast, hindcasts or retrospective forecasts (B) can be evaluated with external evaluations, using temporally independent observations to assess the performance of transferred predictions.

scale (i.e. geological timescales) in order to match the temporal scale of management measures that will benefit from ecological forecasts (Petchey et al. 2015).

We first classified studies as having 'testable' or 'non-testable' model predictions based on the feasibility of collecting the temporally-independent data necessary for an external model evaluation (in case it is not already available). Testable studies transferred models trained in the past forward in time

to generate predictions into a more recent time in which the range of species can be observed. These studies either test temporal predictive performance using a second temporally-independent dataset, or they do not test temporal predictive performance but could do so by generating such a dataset. Non-testable studies, on the other hand, have transferred the models into a future scenario for which observations required to contrast the predictions cannot be obtained (as depicted in

Box 2. Distinguishing model predictive performance from model transferability

Most assessments of transferred predictive performance use the term ‘transferability’ as a synonym for predictive performance under transferred conditions. Transferability has been defined as ‘the ability of a model developed for a specific site and/or time and/or taxon to predict biodiversity in a different time or place or for a different taxon’ (Sequeira et al. 2018) or as the ‘capacity of a model to produce accurate and precise predictions for a new set of predictors that differ from those on which the model was trained’ (Yates et al. 2018). Although these definitions capture the essence of what is sought from predictions beyond model training conditions, they do not explicitly separate processes that confer predictive performance at or near the training conditions to those that confer strong performance by maintaining predictive performance when transferred to new conditions. We propose to define ‘model transferability’ as the capacity of a model to maintain its predictive performance when transferred to new conditions. This definition therefore distinguishes whether a model has been properly specified (e.g. choosing the proper predictors) for the specific conditions in which it was trained, from whether it can maintain its predictive performance under transferred conditions. Hence, model transferability should measure the rate of change in, or loss of predictive performance (e.g. accuracy and precision) when moving from internal evaluation to external evaluation (Fig. 1). Its calculation should therefore be informed by the performance under both training and transfer conditions, and not only by the performance at the time of transfer.

Fig. 1). We designated all the retrospective forecasts as testable regardless of whether they used temporally-independent observations to assess the predictive performance.

In addition to needing a temporally independent dataset for assessing temporal predictive performance, understanding the effect of incorporating ecological and observational processes in RCMs requires comparison of models through systematic inclusions and exclusion of the focal process (hereafter, benchmarking experiments). This is important because more complex models (e.g. process-explicit models) should only be favored over simpler models (e.g. correlative models) if they improve predictive performance compared to simpler models (Zurell 2017). Of the studies classified as testable, we identified and examined those that benchmarked RCMs to assess the effects of explicitly modelling ecological or observational processes on either temporal predictive performance or transferability (Table 1). In order to match the processes modelled in the RCMs listed in Table 1, and the five ecological and two observational processes presented across the manuscript, we considered adaptation as an evolutionary process; community assembly as a result of species interactions; colonization-extinction dynamics as a result of demography and dispersal; and phenotypic plasticity as a result of evolutionary processes.

Model classification

Models were classified in families primarily by model complexity: the first level of classification depends on whether or not the models incorporate time as a model dimension (static versus dynamic models), and the second level of classification depends on the ecological level of the approach (individual-based, population or single-species models and multi-species models) (Table 1). Hybrid-SDMs refer to models that couple habitat suitability predictions across time, obtained from standard correlative species distribution models (cSDMs), with at least one extra sub-model that describes ecological processes. In the case of hybrid-SDMs, we decided to not group them

by model type and left them as individual models because each of them used different types of process-explicit submodels. For example, even though Demoniche (Nenzén et al. 2012) and Lolipop (Cabral and Schurr 2010) both model population dynamics and dispersal, the population dynamics submodel for the first is based on transition matrices, while Lolipop uses demographic models (e.g. Beverton–Holt or Ricker models; Cabral and Schurr 2010). Meta-models also couple two or more submodels, but do not rely on habitat suitability predictions obtained from cSDMs. Note that individual-based models could also be classified as hybrid-SDMs or meta-models, but due to the different ecological level at which they are trained (i.e. individual organisms), for clarity we decided to separate them as an independent family.

Results of the review

The results of the review showed that, out of 1769 individual studies, 649 transferred models to generate predictions forward in time. Only 36 (~ 6%) of them generated predictions that were potentially testable with either empirical or simulated temporally-independent data (Fig. 2). Of these, 28 used retrospective forecasting and nine used simulated data (for the complete list and classification see Supporting information). In the majority of the 36 studies (~ 87%) that generated testable predictions, predictions were created using cSDMs. Only nine out of the 29 process-explicit model types referred to in Table 1 were benchmarked against simpler models using temporally independent datasets, in six papers (Engler and Guisan 2009, Pagel and Schurr 2012, Zurell et al. 2016, Uribe-Rivera et al. 2017, Fordham et al. 2018, Briscoe et al. 2021). Of these nine benchmarked process-explicit model types, five were benchmarked through retrospective forecasting only (Uribe-Rivera et al. 2017, Fordham et al. 2018, Briscoe et al. 2021), one (MigClim) was benchmarked in three different studies using either simulated or empirical data (Engler and Guisan 2009, Zurell et al. 2016, Uribe-Rivera et al. 2017), and the remaining three were

Table 1. Classification of models used for predicting species range change under environmental change. Eight broadly defined model families subsume 30 model types. The processes explicitly modelled are shown in column three and coupled with key references. Whether or not the model temporal predictive performance has been benchmarked against simpler models using temporally independent (empirical or simulated) data and relevant references are indicated in the last column. Adapt=adaptation; CA-Co=community assembly through co-occurrence patterns; CA-S=community assembly through community saturation patterns; Dem=accounting for demography directly through modelling demographic rates; Col-Ext=accounting for demography and dispersal indirectly through modelling local colonization and extinction dynamics; Disp=dispersal capacity; ID=imperfect detection; Phy=physiology; Plas=phenotypic plasticity; SI=species interactions. E=using empirical data (i.e. using retrospective forecast); S=using simulated data. SDM=species distribution models; FIDIMO=fish dispersal model; dynamic-FOAM=dynamic framework of occurrence allocation in metacommunities; M-SET=metacom-
munity – space, environment, time; SESAM=spatially explicit species assemblage modelling. *Applications currently limited to plants or fish species.

Model family	Model type	Processes explicitly declared	Key references	Temporal predictive performance tested?
Static correlative species distribution models (static-cSDMs)	Standard cSDM	None	Guisan and Zimmermann 2000, Ferrier et al. 2002, Elith et al. 2010	Yes (E and S)
	Mechanistic niche models	Phy	Kearney and Porter 2004, Kearney and Porter 2009	No
	Static occupancy detection models (static-ODM)	ID	MacKenzie et al. 2002, Tyre et al. 2003, Kéry and Royle 2008	No
	Joint SDM (jSDM)	CA-Co	Pollock et al. 2014, Ovaskainen et al. 2017	No
	Detection-corrected jSDM	CA-Co, ID	Tobler et al. 2019	No
Dynamic correlative species distribution models (dynamic-cSDMs)	Integrated-SDM	SB, ID	Dorazio 2014, Fithian et al. 2015, Koshkina et al. 2017	No
	SESAM	CA-S	Guisan and Rahbek 2011	No
	The α -adjusted SDM	CA-S	Gavish et al. 2017	No
	Dynamic occupancy models (DOM)	Col-Ext	Royle and Kéry 2007, Kéry et al. 2013	No
	Dynamic occupancy detection models (dynamic-ODM)	Col-Ext, ID	Mackenzie et al. 2003, Kéry et al. 2013	Yes (E)
Hybrid-SDMs (h-SDMs)	Dynamic jSDM	CA-Co, Col-Ext	Thorson et al. 2016	No
	MigClim	Disp	Engler and Guisan 2009	Yes (E and S)
	KISSMig	Disp	Nobis and Normand 2014	No
	BioMove	Dem, Disp	Midgley et al. 2010	No
	Metapop	Dem, Disp	Fordham et al. 2013	Yes (E and S)
Individual-based Models (IBM)	Demoniche	Dem, Disp	Nenzén et al. 2012	Yes (S)
	SDM + FIDIMO*	Disp	Radinger et al. 2017	No
	Lolipop	Dem, Disp	Cabral and Schurr 2010	Yes (S)
	Adapt-R	Adapt, Disp, Phy, Plas	Bush et al. 2016	No
	Coupled stochastic population models with dynamic habitat suitability models	Dem, Disp	Keith et al. 2008, Thuiller et al. 2008, Dullinger et al. 2012	Yes (E)
Meta-models	RangeShifter	Adapt, Dem, Disp	Bocedi et al. 2014	No
	Dynamic ecoevolutionary model	Adapt, Dem, Disp	Cotto et al. 2017	No
	ALADYN	Adapt, Dem, Disp	Schiffers and Travis 2014	No
Meta-models	Dynamic-FOAM	CA-S	Mokany et al. 2011	No
	Coupled stochastic population models with mechanistic niche models	Phy, Dem, Disp, SI	Cabral and Kreft 2012	No
	Dynamic range models	Dem, Disp	Pagel and Schurr 2012	Yes (S)
	DynamicFOAM+M-SET	CA-S, Disp, Col-Ext	Mokany et al. 2012	No
	Spatially-explicit integrated population models (IPM)	Dem, ID	Chandler and Clark 2014, Riecke et al. 2019	No
	Demographic Distribution models	Dem	Diez et al. 2014, Merow et al. 2014	No
	Δtrait-SDMs*	Plas, Adapt + (Phy or Dem)	Benito-Garzón et al. 2019	No

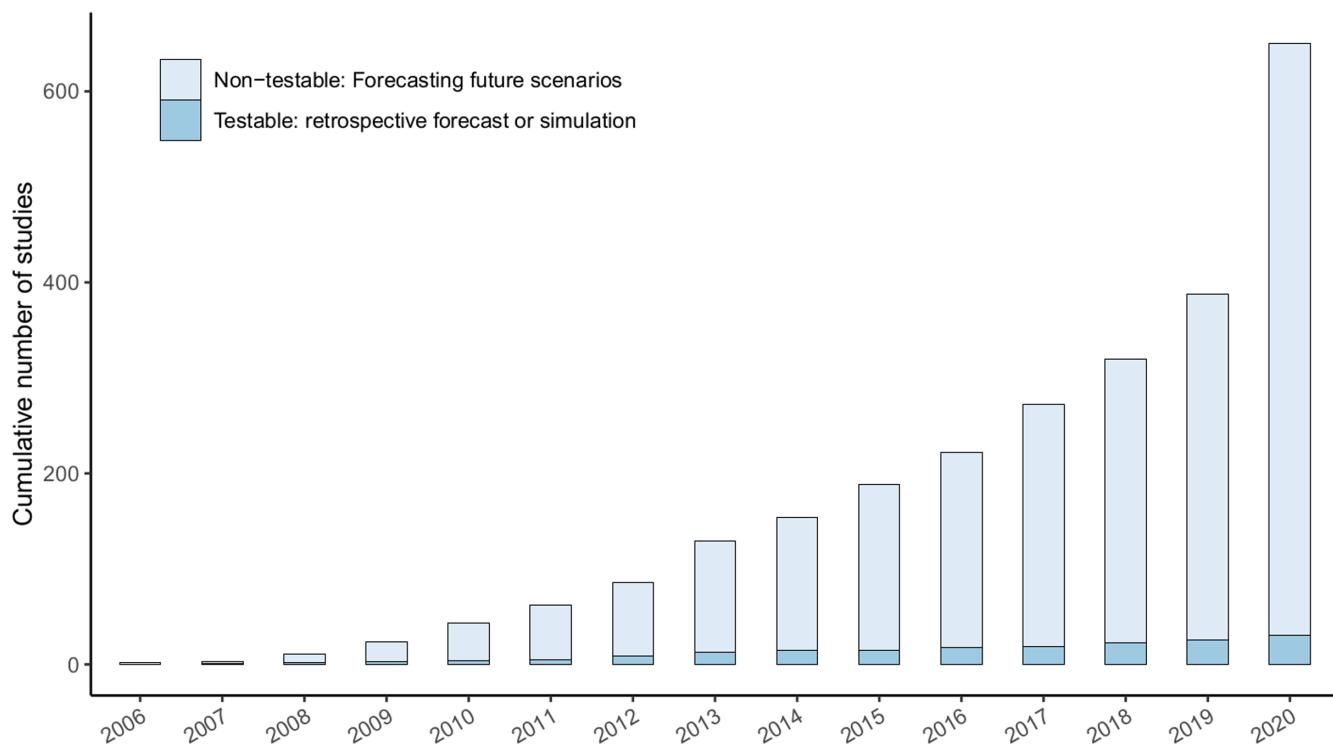


Figure 2. The emergence of RCM studies that make predictions forward in time, expressed as the cumulative number of papers published between January 2006 and December 2020 and available on Web of Knowledge database (accessed on 13 October 2019). Dark blue represents testable predictions (i.e. retrospective forecasting or simulated data) and light blue represent forecasting future scenarios (i.e. impossible to test using currently available observations).

tested with simulated datasets only (Pagel and Schurr 2012, Zurell et al. 2016). The bias towards non-testable future forecasts is expected because RCMs are ultimately designed to inform management under alternative scenarios of future global change. However, it is concerning that most of the RCMs that were explicitly designed to account for ecological and observational processes have not been compared against baseline simpler models.

Evidence from benchmarking studies

Benchmarking studies have been used to compare temporal predictive performance for only two of the five ecological processes we considered; dispersal and demography. Out of the observation processes, only imperfect detection has been benchmarked. None of the benchmarks, however, compared temporal transferability. Dispersal and demography are the most commonly included processes in RCMs (Fig. 3), so although this bias is expected, it implies a significant gap in our understanding. The evidence provided by the limited number of benchmarking studies supports the hypothesis that incorporating dispersal constraints may improve temporal predictive performance of RCMs (Engler and Guisan 2009, Pagel and Schurr 2012, Uribe-Rivera et al. 2017, Fordham et al. 2018). Improved performance may arise from incorporating dispersal constraints because the false presence predictive rate is likely to be reduced (Uribe-Rivera et al. 2017).

The evidence for improved predictive performance through incorporation of demographic processes is mixed (Zurell et al. 2016, Fordham et al. 2018). For example, Fordham et al. (2018) found that demographic hSDMs (accounting for metapopulation dynamics) tended to outperform cSDMs only when land use was incorporated as a predictor (see outstanding questions 1 and 2, Table 3). Evidence from simulated dynamic communities (Zurell et al. 2016) suggests that some inconsistency is due to differing performance of models that predict occupancy patterns compared to those that predict spatial distribution of abundances. Similarly, Briscoe et al. (2021) found that dynamic occupancy models (which indirectly account for demographic processes through modelling colonization and extinction dynamics) tend to outperform cSDMs when used to predict population trends, but tend to be less accurate in predicting spatial patterns of occurrence. In general, inconsistency in the improvement of RCMs with addition of demographic processes may be explained by the oversimplification of population dynamics in the evaluated RCMs, or of the virtual community used to test models through simulation studies. Difficulties in model selection with the increased number of parameters may contribute too (Briscoe et al. 2021).

The impacts of modelling species interactions in RCMs predictive performance or transferability has not been assessed. Modelling this process requires multi-species RCMs, which can model interactions directly (Cabral and Kreft 2012) or

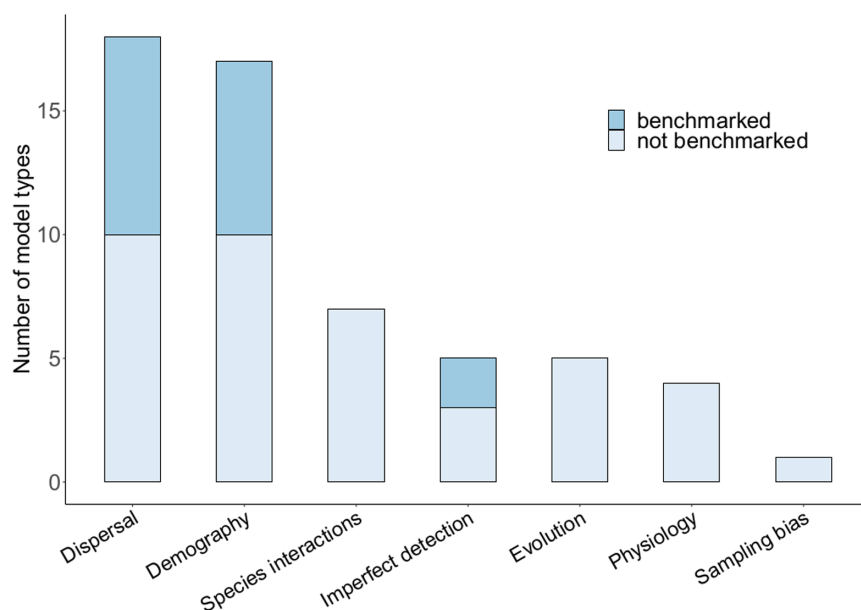


Figure 3. Number of RCM model types that implement ecological and observational processes and the proportion that are benchmarked. Of 29 model types that explicitly model ecological or observational processes only 9 were benchmarked. Light blue represents model types in which the process has been implemented but its effects on model temporal predictive performance or temporal transferability have not been benchmarked. Dark blue represents model types for which their temporal predictive performance has been benchmarked against a simpler model using temporally independent datasets. None of the benchmarking experiments compared model temporal transferability.

indirectly via modelling the resulting community-assembly processes (Guisan and Rahbek 2011, Mokany et al. 2012; Table 1). The consequences of incorporating these processes either directly or indirectly with multi-species RCMs remain unknown (outstanding questions 3 and 4, Table 3).

Rapid (generational) evolution in response to selective pressures is rarely accounted for in RCMs, and we found no studies that benchmarked the impact of including these processes on temporal predictive performance or transferability (Fig. 3). Rapid evolutionary processes such as evolution of dispersal or persistence related traits can generate eco-evolutionary dynamics with consequences for species range change that are poorly understood (Donelson et al. 2019, Nadeau and Urban 2019, Peterson et al. 2019). Implementations of these dynamics in RCMs are likely to improve ecological forecasts (Waldvogel et al. 2020) and represent another key knowledge gap (outstanding questions 5 and 6, Table 3).

Incorporation of imperfect detection was examined in only two model types in a single study (Fig. 3; Briscoe et al. 2021). Temporal predictive performance was not significantly improved over the baseline models across the 69 bird species studied, but there was improvement for species with lower detection probabilities (Briscoe et al. 2021). Imperfect detection is common in wildlife surveys, and the ODM framework allows estimating environment–occupancy relationships while accounting for imperfect detection (Lahoz-Monfort et al. 2014). More studies are needed to assess the benefits of accounting for imperfect detection on RCM predictions, including for multi-species and dynamic ODMs (outstanding question 7, Table 3).

The impacts of accounting for sampling bias on RCMs temporal predictive performance and transferability are still unknown. Recent simulation studies have shown that accounting for sampling bias through model-based data integration (the only model type accounting for sampling bias; Table 1) may generate more accurate predictions of species distributions than using a single data source (Simmonds et al. 2020, Ahmad Suhaimi et al. 2021). Testing whether these improvements are kept when projecting the models forward in time should, therefore, be considered another knowledge gap (outstanding question 8, Table 3).

Reasons for the knowledge gaps

The use of retrospective forecasts for benchmarking experiments in ecology lags other disciplines (Urban 2019, Xia et al. 2020). Several factors may discourage or prevent researchers from conducting benchmarking experiments. Historically, limiting factors were computational power, statistical methods and data availability (Luo et al. 2011). Arguably, improvements in computational capacity and development of statistical advances have cleared some of the main obstacles. The most likely reason for this lag is data limitation. Data limitation is not necessarily overcome by having two temporally-independent distributional datasets for training and testing RCMs. While contemporary distribution datasets are relatively easy to come by, the main constraint on benchmarking RCMs using retrospective forecasts is the lack of information-rich, accurately georeferenced data necessary for training process-explicit models in the past. If resources

Table 2. Examples of consolidated plant databases (i.e. compilations of datasets from several survey initiatives with predefined inclusion rules) readily available to inform retrospective forecasting exercises. *Accepted number of species (Plantae). **Published central tendency of individual measurements. Websites accessed on 19 October 2021.

Database	Information content	No. of plant species	No. of	Further information
Global Biodiversity Information Facility (GBIF)	Global compilation of geo-referenced occurrence records	> 537k*	~355.3M georeferenced records	www.gbif.org/what-is-gbif
sPlot	Global compilation of vegetation plots	> 43.1k	> 1.97M plots plant	www.idiv.de/en/sdiv/working_groups/wg_pool/splot/splot_database.html
TRY	Global compilation of plant traits	> 279.8k	> 11.8M trait records	www.try-db.org/
COMPADRE	Global compilation of demographic rates (from population matrices)	757	8708 matrix population models	www.compadre-db.org/Data/Compadre
GlobTherm	A global database of thermal tolerances	242	249 species level** measures of minimum or maximum thermal tolerances	https://doi.org/10.5061/dryad.1cv08
Global Inventory of Floras and Traits (GIFT)	A Global Inventory of Floras and Traits	> 355k	> 3.55M species-by-region occurrences; > 2.39M species-trait combinations; 109 traits	www.gift.uni-goettingen.de
forestREplot	Compilation of temperate forest plots revisited after at least 20 years with focus on forest herb layer	NA	4910 plots revisited	www.forestreplot.ugent.be/
EMBL-EBI	Molecular data resources	NA	NA Genes, proteins, genotypes among other resources	www.ebi.ac.uk/services

are available, high quality contemporary testing datasets can be collected. However, high quality historical data sets cannot generally be created. This issue is particularly acute for long-lived organisms. For example, little opportunity exists for testing retrospective forecasts produced by dynamic RCMs because most require very detailed spatio-temporally explicit distribution data to be trained (e.g. time-series of abundance/occurrence for many sites across the landscape or multi-temporal presence-absence maps; outstanding question 1, Table 3). Simulated data had been used to overcome this limitation. One of the model types that is likely restricted by this kind of data limitation is the dynamic range models (Pagel and Schurr 2012), which has only been tested using simulated data. Commonly, these simulations force an equilibrium between species and their environment at baseline time (i.e. the time in which the models are trained, or in some cases the center of the time-series used to train the models) for the virtual model-testing experiment (Pagel and Schurr 2012, Zurell et al. 2016). Establishing equilibrium baseline conditions during simulation testing may mask the improvements of accounting for non-equilibrium demographic, dispersal or evolutionary dynamics. Hence, it is also necessary to test whether the magnitude or direction of the effects of accounting for ecological processes on the temporal predictive performance of these models will change when simulating populations that are not in equilibrium when training data are collected (outstanding question 2, Table 3).

Other reasons that could explain the lack of benchmarking studies of process-explicit RCMs include lack of access to documentation and accessible tutorials to guide the parameterization of complex models (Briscoe et al. 2019), and the data-specificity of process-explicit models (Evans et al. 2016, Urban et al. 2022). The data required to train one particular process-explicit RCM may not be applicable to an alternative model, making it more difficult to collect the data required for benchmarking. Differences in the format of predictions may also induce some hesitation about embarking in benchmarking exercises. For example, individual-based models usually predict population densities and not probabilities of occurrence. This may limit the direct use of commonly used predictive performance metrics such as AUC-ROC and AUC-PR, which are better suited for probabilistic predictions (but we note that estimates of occurrence probability can be derived from mean density estimates). Finally, using theoretical knowledge of range change drivers to develop models that explicitly account for processes may increase the risk of confirmation bias, making researchers less inclined to interrogate models to reveal their weaknesses and limitations (Connolly et al. 2017).

Data and opportunities to improve model testing

Recent efforts to make consolidated databases more available are diminishing the barrier of historical data for benchmarking temporal predictive performance. We briefly reviewed

Table 3. Summary of outstanding questions; the process-explicit models that may be compared to a simpler reference model to test their potential improvements; and the consolidated datasets that could be used to train and test the models. ODM=occupancy detection models; DRM=dynamic range models; IPM=integrated population models; cSDM=correlative species distribution models; jSDM=joint species distribution models.

Topic	Outstanding questions	Candidate models for testing	Relevant datasets	Key references
Dynamism and non-equilibrium	1. Does accounting for colonization/extinction dynamics improve the temporal predictive performance and transferability compared to static RCMs?	Dynamic occupancy models versus cSDMs; dynamic ODM versus static ODM; dynamic jSDM versus stacked cSDM	sPlot/Forest-Replot/GBIF	Kéry and Royle 2008, Kéry et al. 2013, Thorson et al. 2016, Briscoe et al. 2021
	2. Can the decision of simulating communities in equilibrium or not influence the evaluation of the effects of processes on temporal predictive performance and transferability of RCM?	Dynamic and static RCMs in equilibrium-forced simulated community versus dynamic and static RCM in non-equilibrium simulated community	Simulated data	Pagel and Schurr 2012, Fordham et al. 2013, Merow et al. 2014, Thorson et al. 2016, Riecke et al. 2019
Community-level models or stacking single-species models	3. Does accounting for species interactions through modelling co-occupancy patterns improve the temporal predictive performance and transferability of RCM at species and community levels?	jSDM versus stacked single-species cSDM	sPlot/Forest-Replot/GBIF	Zurell et al. 2020, Poggiato et al. 2021
	4. Does accounting for community saturation through modelling species richness patterns improve the temporal predictive performance and transferability of RCM at community level?	SESAM, α -adjusted SDM or (DynamicFOAM+M-SET) versus stacked cSDMs	sPlot/Forest-Replot/GBIF	Guisan and Rahbek 2011, Mokany et al. 2012, Gavish et al. 2017, Zurell et al. 2020
Rapid evolutionary processes	5. Can accounting for local adaptation at the population level improve the temporal predictive performance and transferability of RCM?	Adapt-R versus KISSMig or MigClim	EMBL-EBI/GIFT/GBIF	Bush et al. 2016, Waldvogel et al. 2020
	6. Can the incorporation of spatio-temporal dynamics in the distribution of genotypes/phenotypes improve the temporal predictive performance and transferability of RCM?	Dynamic ecoevolutionary model or RangeShifter versus Demoniche or Metapop	EMBL-EBI/GIFT/GBIF/COMADRE	Cotto et al. 2017, Waldvogel et al. 2020
Observational processes	7. Does accounting for imperfect detection through modelling detection probabilities improve the temporal predictive performance and transferability of RCM?	Occupancy-detection models versus standard cSDM	sPlot/Forest-Replot	Lahoz-Monfort et al. 2014, Guillera-Arroita 2017, Briscoe et al. (2021)
	8. Does accounting for sampling bias/imperfect detection through data integration improve the temporal predictive performance and transferability of RCM?	Integrated-SDM versus cSDM	GBIF/sPlot/Forest-Replot	Koshkina et al. 2017, Miller et al. 2019, Isaac et al. 2020

consolidated datasets potentially useful for performing benchmarking experiments between alternative RCMs through retrospective forecasting of observed range-shifts. For this illustrative exercise, we limited the search to one of the taxa that has been more closely studied and so more data-rich, the plants (i.e. Plantae kingdom). However, most of the listed databases are not limited to plants (e.g. GBIF, GlobTherm, EMBL-EBI) or have analogue consolidation databases for other taxa (e.g. COMADRE is the equivalent to COMPADRE, but for animal species). Our findings highlight data of value for model development and testing, including opportunistic occurrence records (such as GBIF), occupancy, demographic and molecular and functional traits data for tree and other plant species (Table 2). Many of these consolidated databases have global coverage and include data collected as early as the 19th century. One of them, ForestREplot, was specifically designed to investigate temporal changes in species abundance, distribution, composition and other changes in biological diversity of understory plants.

Even though huge efforts have been recently directed to make data accessible on these platforms, the opportunities still come with challenges. Two of the main ones are 1) developments to verify the veracity and variety of the data are not keeping pace with the velocity of data accumulation (but see Sabatini et al. 2021); and 2) the velocity to assimilate data lags the velocity of data collection and model developments (Xia et al. 2020). The recent release of a new version of sPlot (Table 2; Bruelheide et al. 2019) is an indication that part of these challenges are being targeted. This new open-access dataset, implemented in an R package, subsampled the vegetation plots from sPlot to ensure a more environmentally balanced version for global comparisons (sPlotOpen; Sabatini et al. 2021). More efforts need to be redirected to coordinate not only filling the data gaps, but also to generate platforms and cyber-structures that will help identify issues and devote the resources in the necessary direction (Fer et al. 2020, Urban et al. 2022).

To stimulate future research on understanding the temporal predictive performance and transferability of RCMs, we collate the outstanding questions identified in previous sections, models requiring benchmarking and the data required to train and test them (Table 3). The proposed research would help fill the knowledge gaps identified in this review. We do not intend to answer the question of whether the data necessary for performing retrospective forecast, and particularly for training process-based RCMs in the past, is available for any of the proposed model benchmarking. However, we suggest a couple of databases in which researchers may find data potentially useful for benchmarking retrospective forecasts across RCMs. Generating simulated data to perform model benchmarking may remain the only plausible option in the short term for benchmarking some types of models. Ideally, benchmarking RCMs to better understand temporal predictive performance and transferability will be undertaken using a combination of empirical retrospective forecasting and simulation data studies.

Conclusions

Forecasting species distributions under changing environments remains a key challenge for biodiversity conservation policy and management. Our understanding of how explicitly modelled ecological and observational processes may influence the temporal predictive performance and transferability of RCMs is far from complete. Here we highlighted and described the gap between the diversification of RCMs that incorporate relevant processes and the assessment of the potential improvements to range change prediction that they may bring. We found that past assessments have been largely restricted to RCMs that included dispersal and demographic processes. Lack of rigorous assessments of models that incorporate other relevant processes such as local adaptation or sampling bias is concerning because there remains little basis on which to prioritize their inclusion in models that inform conservation management priorities (Román-Palacios and Wiens 2020).

Considering the identified knowledge gaps and range dynamics theory, we propose research to better understand the influence of ecological and observational processes on temporal predictive performance and transferability of RCMs. The use of historical databases and simulation exercises to explore key methodological questions, including relative predictive performance of process-explicit and process-implicit models, is an opportunity to fill knowledge gaps. Efforts to compile and improve the accessibility of biodiversity data, and guidelines to perform simulation exercises are crucial to making benchmarking experiments the standard protocol to test whether the modelling of processes actually improve range change forecasts.

Acknowledgements – The authors thank the reviewers and editors for their valuable comments which significantly improved the manuscript.

Funding – This study was funded by Beca Doctorado Becas Chile no. 72170498-2016, Proyecto Fondecyt no. 1181677-2018, and David Ashton Scholarship 2018 and 2019, Botany Foundation, University of Melbourne.

Author contributions

David E. Uribe-Rivera: Conceptualization (lead); Formal analysis (lead); Investigation (lead); Methodology (lead); Writing – original draft (lead); Writing – review and editing (equal). **Gurutzeta Guillera-Arroita:** Conceptualization (equal); Methodology (supporting); Supervision (lead); Writing – review and editing (equal). **Saras M. Windecker:** Writing – review and editing (equal). **Patricio Plischoff:** Resources (supporting); Supervision (supporting); Writing – review and editing (supporting). **Brendan A. Wintle:** Conceptualization (equal); Methodology (supporting); Supervision (lead); Writing – review and editing (equal).

Transparent peer review

The peer review history for this article is available at <https://publons.com/publon/10.1111/ecog.06048>.

Data availability statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Supporting information

The Supporting information associated with this article is available with the online version.

References

- Ahmad Suhaimi, S. S. et al. 2021. Integrated species distribution models: a comparison of approaches under different data quality scenarios – *Divers. Distrib.* 27: 1066–1075.
- Araújo, M. B. and Rahbek, C. 2006. How does climate change affect biodiversity? – *Science* 313: 1396–1397.
- Araújo, M. B. et al. 2005. Reducing uncertainty in projections of extinction risk from climate change: uncertainty in species' range shift projections. – *Global Ecol. Biogeogr.* 14: 529–538.
- Benito-Garzon, M. et al. 2019. Δ Trait SDMs: species distribution models that account for local adaptation and phenotypic plasticity. – *New Phytol.* 222: 1757–1765.
- Bennett, J. M. et al. 2018. GlobTherm, a global database on thermal tolerances for aquatic and terrestrial organisms. – *Sci. Data* 5: 1–7.
- Blois, J. L. et al. 2013. Modeling the climatic drivers of spatial patterns in vegetation composition since the last glacial maximum. – *Ecography* 36: 460–473.
- Bocedi, G. et al. 2014. Mechanistic modelling of animal dispersal offers new insights into range expansion dynamics across fragmented landscapes. – *Ecography* 37: 1240–1253.
- Briscoe, N. J. et al. 2019. Forecasting species range dynamics with process-explicit models: matching methods to applications. – *Ecol. Lett.* 22: 1940–1956.
- Briscoe, N. J. et al. 2021. Can dynamic occupancy models improve predictions of species' range dynamics? A test using Swiss birds. – *Global Change Biol.* 27: 4269–4282.
- Bruelheide, H. et al. 2019. sPlot – a new tool for global vegetation analyses. – *J. Veg. Sci.* 30: 161–186.
- Bush, A. et al. 2016. Incorporating evolutionary adaptation in species distribution modelling reduces projected vulnerability to climate change. – *Ecol. Lett.* 19: 1468–1478.
- Cabral, J. S. and Kreft, H. 2012. Linking ecological niche, community ecology and biogeography: insights from a mechanistic niche model. – *J. Biogeogr.* 39: 2212–2224.
- Cabral, J. S. and Schurr, F. M. 2010. Estimating demographic models for the range dynamics of plant species: demographic models for range dynamics. – *Global Ecol. Biogeogr.* 19: 85–97.
- Chandler, R. B. and Clark, J. D. 2014. Spatially explicit integrated population models. – *Methods Ecol. Evol.* 5: 1351–1360.
- Clark, J. S. et al. 2001. Ecological forecasts: an emerging imperative. – *Science* 293: 657–660.
- Connolly, S. R. et al. 2017. Process, mechanism and modeling in macroecology. – *Trends Ecol. Evol.* 32: 835–844.
- Cotto, O. et al. 2017. A dynamic eco-evolutionary model predicts slow response of alpine plants to climate warming. – *Nat. Commun.* 8: 15399.
- Cuddington, K. et al. 2013. Process-based models are required to manage ecological systems in a changing world. – *Ecosphere* 4: art20.
- Dietze, M. C. et al. 2018. Iterative near-term ecological forecasting: needs, opportunities and challenges. – *Proc. Natl Acad. Sci. USA* 115: 1424–1432.
- Diez, J. M. et al. 2014. Probabilistic and spatially variable niches inferred from demography. – *J. Ecol.* 102: 544–554.
- Dobrowski, S. Z. et al. 2011. Modeling plant ranges over 75 years of climate change in California, USA: temporal transferability and species traits. – *Ecol. Monogr.* 81: 241–257.
- Donelson, J. M. et al. 2019. Understanding interactions between plasticity, adaptation and range shifts in response to marine environmental change. – *Phil. Trans. R. Soc. B* 374: 20180186.
- Dorazio, R. M. 2014. Accounting for imperfect detection and survey bias in statistical analysis of presence-only data. – *Global Ecol. Biogeogr.* 23: 1472–1484.
- Dormann, C. F. 2007. Promising the future? Global change projections of species distributions. – *Basic Appl. Ecol.* 8: 387–397.
- Dullinger, S. et al. 2012. Extinction debt of high-mountain plants under twenty-first-century climate change. – *Nat. Clim. Change* 2: 619–622.
- Elith, J. and Leathwick, J. R. 2009. Species distribution models: ecological explanation and prediction across space and time. – *Annu. Rev. Ecol. Evol. Syst.* 40: 677–697.
- Elith, J. et al. 2010. The art of modelling range-shifting species. – *Methods Ecol. Evol.* 1: 330–342.
- Engler, R. and Guisan, A. 2009. MigClim: predicting plant distribution and dispersal in a changing climate. – *Divers. Distrib.* 15: 590–601.
- Evans, M. E. K. et al. 2016. Towards process-based range modeling of many species. – *Trends Ecol. Evol.* 31: 860–871.
- Evans, M. R. 2012. Modelling ecological systems in a changing world. – *Phil. Trans. R. Soc. B* 367: 181–190.
- Fer, I. et al. 2020. Beyond ecosystem modeling: a roadmap to community cyberinfrastructure for ecological data-model integration. – *Global Change Biol.* 27: 13–26.
- Ferrier, S. and Guisan, A. 2006. Spatial modelling of biodiversity at the community level. – *J. Appl. Ecol.* 43: 393–404.
- Ferrier, S. et al. 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II. Community-level modelling. – *Biodivers. Conserv.* 11: 2309–2338.
- Fielding, A. H. and Haworth, P. F. 1995. Testing the generality of bird-habitat models. – *Conserv. Biol.* 9: 1466–1481.
- Fithian, W. et al. 2015. Bias correction in species distribution models: pooling survey and collection data for multiple species. – *Methods Ecol. Evol.* 6: 424–438.
- Fitzpatrick, M. C. and Hargrove, W. W. 2009. The projection of species distribution models and the problem of non-analog climate. – *Biodivers. Conserv.* 18: 2255–2261.
- Fordham, D. A. et al. 2013. Tools for integrating range change, extinction risk and climate change information into conservation management. – *Ecography* 36: 956–964.
- Fordham, D. A. et al. 2014. Better forecasts of range dynamics using genetic data. – *Trends Ecol. Evol.* 29: 436–443.
- Fordham, D. A. et al. 2018. How complex should models be? Comparing correlative and mechanistic range dynamics models. – *Global Change Biol.* 24: 1357–1370.
- Gavish, Y. et al. 2017. Accounting for biotic interactions through alpha-diversity constraints in stacked species distribution models. – *Methods Ecol. Evol.* 8: 1092–1102.
- Gharari, S. et al. 2013. An approach to identify time consistent model parameters: sub-period calibration. – *Hydrol. Earth Syst. Sci.* 17: 149–161.

- Grenouillet, G. and Comte, L. 2014. Illuminating geographical patterns in species' range shifts. – *Global Change Biol.* 20: 3080–3091.
- Guillera-Arroita, G. 2017. Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. – *Ecography* 40: 281–295.
- Guisan, A. and Rahbek, C. 2011. SESAM a new framework integrating macroecological and species distribution models for predicting spatiotemporal patterns of species assemblages. – *J. Biogeogr.* 38: 1433–1444.
- Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. – *Ecol. Model.* 135: 147–186.
- Heikkinen, R. K. et al. 2012. Does the interpolation accuracy of species distribution models come at the expense of transferability? – *Ecography* 35: 276–288.
- Isaac, N. J. et al. 2020. Data integration for large-scale models of species distributions. – *Trends Ecol. Evol.* 35: 56–67.
- Kattge, J. et al. 2011. TRY – a global database of plant traits. – *Global Change Biol.* 17: 2905–2935.
- Kearney, M. and Porter, W. 2009. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. – *Ecol. Lett.* 12: 334–350.
- Kearney, M. and Porter, W. P. 2004. Mapping the fundamental niche: physiology, climate and the distribution of a nocturnal lizard. – *Ecology* 85: 3119–3131.
- Keith, D. A. et al. 2008. Predicting extinction risks under climate change: coupling stochastic population models with dynamic bioclimatic habitat models. – *Biol. Lett.* 4: 560–563.
- Kéry, M. and Royle, J. A. 2008. Hierarchical Bayes estimation of species richness and occupancy in spatially replicated surveys. – *J. Appl. Ecol.* 45: 589–598.
- Kéry, M. et al. 2013. Analysing and mapping species range dynamics using occupancy models. – *J. Biogeogr.* 40: 1463–1474.
- Koshkina, V. et al. 2017. Integrated species distribution models: combining presence-background data and site-occupancy data with imperfect detection. – *Methods Ecol. Evol.* 8: 420–430.
- Lahoz-Monfort, J. J. et al. 2014. Imperfect detection impacts the performance of species distribution models: imperfect detection impacts species distribution models. – *Global Ecol. Biogeogr.* 23: 504–515.
- Luo, Y. et al. 2011. Ecological forecasting and data assimilation in a data-rich era. – *Ecol. Appl.* 21: 1429–1442.
- MacKenzie, D. I. et al. 2002. Estimating site occupancy rates when detection probabilities are less than one. – *Ecology* 83: 2248–2255.
- MacKenzie, D. I. et al. 2003. Estimating site occupancy, colonization and local extinction when a species is detected imperfectly. – *Ecology* 84: 2200–2207.
- Merow, C. et al. 2014. On using integral projection models to generate demographically driven predictions of species' distributions: development and validation using sparse data. – *Ecography* 37: 1167–1183.
- Mesgaran, M. B. et al. 2014. Here be dragons: a tool for quantifying novelty due to covariate range and correlation change when projecting species distribution models. – *Divers. Distrib.* 20: 1147–1159.
- Midgley, G. F. et al. 2010. BioMove an integrated platform simulating the dynamic response of species to environmental change. – *Ecography* 33: 612–616.
- Miller, D. A. et al. 2019. The recent past and promising future for data integration methods to estimate species' distributions. – *Methods Ecol. Evol.* 10: 22–37.
- Mokany, K. et al. 2011. Combining α - and β -diversity models to fill gaps in our knowledge of biodiversity. – *Ecol. Lett.* 14: 1043–1051.
- Mokany, K. et al. 2012. Dynamic macroecology and the future for biodiversity. – *Global Change Biol.* 18: 3149–3159.
- Morán-Ordóñez, A. et al. 2017. Evaluating 318 continental-scale species distribution models over a 60-year prediction horizon: what factors influence the reliability of predictions? Temporal transferability of species distribution model predictions. – *Global Ecol. Biogeogr.* 26: 371–384.
- Nadeau, C. P. and Urban, M. C. 2019. Eco-evolution on the edge during climate change. – *Ecography* 42: 1280–1297.
- Nenzén, H. K. et al. 2012. Demoniche – an R-package for simulating spatially-explicit population dynamics. – *Ecography* 35: 577–580.
- Nobis, M. P. and Normand, S. 2014. KISSMig – a simple model for R to account for limited migration in analyses of species distributions. – *Ecography* 37: 1282–1287.
- Ovaskainen, O. et al. 2017. How to make more out of community data? A conceptual framework and its implementation as models and software. – *Ecol. Lett.* 20: 561–576.
- Pagel, J. and Schurr, F. M. 2012. Forecasting species ranges by statistical estimation of ecological niches and spatial population dynamics: statistical estimation of dynamic range models. – *Global Ecol. Biogeogr.* 21: 293–304.
- Petchey, O. L. et al. 2015. The ecological forecast horizon, and examples of its uses and determinants. – *Ecol. Lett.* 18: 597–611.
- Peterson, M. L. et al. 2019. Incorporating local adaptation into forecasts of species' distribution and abundance under climate change. – *Global Change Biol.* 25: 775–793.
- Poggiato, G. et al. 2021. On the interpretations of joint modeling in community ecology. – *Trends Ecol. Evol.* 36: 391–401.
- Pollock, L. J. et al. 2014. Understanding cooccurrence by modelling species simultaneously with a joint species distribution model (JSDM). – *Methods Ecol. Evol.* 5: 397–406.
- Radinger, J. et al. 2017. The future distribution of river fish: the complex interplay of climate and land use changes, species dispersal and movement barriers. – *Global Change Biol.* 23: 4970–4986.
- Randin, C. F. et al. 2006. Are niche-based species distribution models transferable in space? – *J. Biogeogr.* 33: 1689–1703.
- Rapacciuolo, G. et al. 2012. Climatic associations of british species distributions show good transferability in time but low predictive accuracy for range change. – *PLoS One* 7: e40212.
- Riecke, T. V. et al. 2019. Integrated population models: model assumptions and inference. – *Methods Ecol. Evol.* 10: 1072–1082.
- Roberts, D. R. et al. 2017. Cross-validation strategies for data with temporal, spatial, hierarchical or phylogenetic structure. – *Ecography* 40: 913–929.
- Román-Palacios, C. and Wiens, J. J. 2020. Recent responses to climate change reveal the drivers of species extinction and survival. – *Proc. Natl Acad. Sci. USA* 117: 4211–4217.
- Royle, J. A. and Kéry, M. 2007. A Bayesian state-space formulation of dynamic occupancy models. – *Ecology* 88: 1813–1823.
- Rubidge, E. M. et al. 2011. The role of climate, habitat and species co-occurrence as drivers of change in small mammal distributions over the past century: drivers of change in chipmunk distributions. – *Global Change Biol.* 17: 696–708.
- Sabatini, F. M. et al. 2021. sPlotOpen – an environmentally balanced, open-access, global dataset of vegetation plots. – *Global Ecol. Biogeogr.* 30: 1740–1764.

- Salguero-Gómez, R. et al. 2015. The COMPADRE Plant Matrix Database: an open online repository for plant demography. – *J. Ecol.* 103: 202–218.
- Schiffers, K. H. and Travis, J. M. J. 2014. ALADYN – a spatially explicit, allelic model for simulating adaptive dynamics. – *Ecography* 34: 1288–1291.
- Sequeira, A. M. M. et al. 2018. Transferring biodiversity models for conservation: opportunities and challenges. – *Methods Ecol. Evol.* 9: 1250–1264.
- Simmonds, E. G. et al. 2020. Is more data always better? A simulation study of benefits and limitations of integrated distribution models. – *Ecography* 43: 1413–1422.
- Singer, A. et al. 2016. Community dynamics under environmental change: how can next generation mechanistic models improve projections of species distributions? – *Ecol. Model.* 326: 63–74.
- Tapia, L. et al. 2018. Assessing the temporal transferability of raptor distribution models: Implications for conservation. – *Bird Conserv. Int.* 28: 375–389.
- Thorson, J. T. et al. 2016. Joint dynamic species distribution models: a tool for community ordination and spatio-temporal monitoring. – *Global Ecol. Biogeogr.* 25: 1144–1158.
- Thuiller, W. et al. 2008. Predicting global change impacts on plant species' distributions: future challenges. – *Perspect. Plant Ecol. Evol. Syst.* 9: 137–152.
- Tobler, M. W. et al. 2019. Joint species distribution models with species correlations and imperfect detection. – *Ecology* 100: e02754.
- Tyre, A. J. et al. 2003. Improving precision and reducing bias in biological surveys: estimating false negative error rates. – *Ecol. Appl.* 13: 1790–1801.
- Urban, M. C. 2019. Projecting biological impacts from climate change like a climate scientist. – *WIREs Clim. Change* 10: e585.
- Urban, M. C. et al. 2016. Improving the forecast for biodiversity under climate change. – *Science* 353: aad8466.
- Urban, M. C. et al. 2022. Coding for life: designing a platform for projecting and protecting global biodiversity. – *BioScience* 72: 91–104.
- Uribe-Rivera, D. E. et al. 2017. Dispersal and extrapolation on the accuracy of temporal predictions from distribution models for the Darwin's frog. – *Ecol. Appl.* 27: 1633–1645.
- Waldvogel, A. et al. 2020. Evolutionary genomics can improve prediction of species' responses to climate change. – *Evol. Lett.* 4: 4–18.
- Weigelt, P. et al. 2020. GIFT – a global inventory of floras and traits for macroecology and biogeography. – *J. Biogeogr.* 47: 16–43.
- Xia, J. et al. 2020. Research challenges and opportunities for using big data in global change biology. – *Global Change Biol.* 26: 6040–6061.
- Yates, K. L. et al. 2018. Outstanding challenges in the transferability of ecological models. – *Trends Ecol. Evol.* 33: 790–802.
- Zurell, D. 2017. Integrating demography, dispersal and interspecific interactions into bird distribution models. – *J. Avian Biol.* 48: 1505–1516.
- Zurell, D. et al. 2016. Benchmarking novel approaches for modeling species range dynamics. – *Global Change Biol.* 22: 2651–2664.
- Zurell, D. et al. 2020. Testing species assemblage predictions from stacked and joint species distribution models. – *J. Biogeogr.* 47: 101–113.