How complex should models be? Comparing correlative and mechanistic range dynamics models

Running head: Testing distribution model forecasts

Authors: Damien A. Fordham[1*], Cleo Bertelsmeier[1,2], Barry W. Brook[3], Dora Neto[4], Regan Early[5], Stuart C. Brown[1], Sébastien Ollier[6], Miguel B. Araújo[4,7,8]

[1] The Environment Institute, School of Biological Sciences, The University of Adelaide, South Australia 5005, Australia

[2] Univ Lausanne, Department of Ecololy & Evolution, CH-1015 Lausanne, Switzerland

[3] School of Biological Sciences, Private Bag 55, University of Tasmania, Hobart 7001, Australia

[4] CIBIO-InBIO, University of Évora, Largo dos Colegiais, 7000 Évora, Portugal

[5] Centre for Ecology and Conservation, University of Exeter Cornwall Campus, Penryn, Cornwall, TR10 9FE, UK

[6] Univ. Paris-Sud, UMR CNRS 8079, Bât 362, 91405 Orsay, France.

[7] National Museum of Natural Sciences, CSIC, C/ Jose Gutierrez Abascal, 2, 28006 Madrid, Spain

[8] Center for Macroecology, Evolution and Climate, Natural History Museum of Denmark, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark

*Corresponding author

ph:+61 8 83136711

fax: +61 8 83134347

Email: damien.fordham@adelaide.edu.au

Key words: climate change; hybrid ecological niche model; independent model validation**;** land use**;** mechanistic model; metapopulation and dispersal dynamics; species distribution model; transferability

Abstract

Criticism has been levelled at climate-change induced forecasts of species range shifts that do not account explicitly for complex population dynamics. The relative importance of such dynamics under climate change is, however, undetermined because direct tests comparing the performance of demographic models versus simpler ecological niche models are still lacking owing to difficulties in evaluating forecasts using real-world data. We provide the first comparison of the skill of coupled ecological-niche-population models and ecological niche models in predicting documented shifts in the ranges of 20 British breeding bird species across a 40-year period. Forecasts from models calibrated with data centred on 1970 were evaluated using data centred on 2010. We found that more complex coupled ecological-niche-population models (that account for dispersal and metapopulation dynamics) tend to have higher predictive accuracy in forecasting species range shifts than structurally simpler models that only account for variation in climate. However, these better forecasts are achieved only if ecological responses to climate change are simulated without static snapshots of historic land use, taken at a single point in time. In contrast, including both static land use and dynamic climate variables in simpler ecological niche models improves forecasts of observed range shifts. Despite being less skilful at predicting range changes at the grid cell level, ecological niche models do as well, or better, than more complex models at predicting the magnitude of relative change in range size. Therefore, ecological niche models can provide a reasonable first approximation of the magnitude of species' potential range shifts, especially when more detailed data are lacking on dispersal dynamics, demographic processes underpinning population performance, and change in land cover.

Introduction

There is unprecedented demand for forecasts of biodiversity change owing to the multiple human threatening processes affecting species and ecosystems worldwide (Mouquet *et al.*, 2015, Pereira *et al.*, 2010). The unparalleled access to large quantities of ecological data, coupled with increasingly sophisticated statistical and modelling tools, offers great promise for improving ecological forecasts. However, model usefulness is contingent on them being able to transfer predictions in space and/or time (Stewart *et al.*, 2015, Willis *et al.*, 2007). But how can one assess a model's capacity to anticipate global change impacts on species if climate change scenarios have not yet occurred?

Pragmatically, performance is usually assessed by testing the model predictions against records from the regions or time periods used to train the models (Araújo & Rahbek, 2006). Previous studies based on hindcasts (backwards projections) of past range shifts have used a variety of time horizons and taxonomic groups, and their results on model transferability are heterogeneous. For example, predictive ability across time was generally low in studies on cetaceans (hindcasts for the 1970s, Lambert *et al.*, 2014) and plants (late Pleistocene; Maguire *et al.*, 2016). In a study on birds, observed changes in abundance were significantly positively correlated to the model predictions in only 59% of cases (Illán *et al.*, 2014) — little better than an even bet. Even for models that were reasonably accurate in predicting range area for the training period, predictive skill tended to deteriorate substantially when used to forecast range shifts (Araujo *et al.*, 2005 [birds], Roberts & Hamann, 2012 [plants], Smith *et al.*, 2013[mammals]). Possible explanations for poor model transferability include the appearance of non-analogue climates (Pearson *et al.*, 2006), the lack of conservatism in species environmental tolerances (Pearman *et al.*, 2008a) and novel species interactions (Smith *et al.*, 2013). Where studies have found fair to good predictive accuracy over time, model transferability typically varied across species (Pearman *et al.*, 2008b), the level of management (Macias-Fauria & Willis, 2013), or the degree of stability in species-environment correlations (Rubidge *et al.*, 2011).

Importantly, all of these studies have used comparisons of different flavours of correlative ecological niche models (ENMs) — the most frequent type of modelling approach used to predict species range shifts (Araujo & Peterson, 2012). These models statistically correlate species ranges (occurrence or abundance data at known locations) with information on the environmental characteristics of those locations thought to delimit the species' tolerances for those environmental conditions (Elith & Leathwick, 2009). Once fitted to historical or present-day data, the models can be used to predict the species ranges across a landscape under future change. Ecological niche models have a set of inherent limitations (Guisan & Thuiller, 2005), but are generally considered to deliver a useful approximation of suitable areas, contingent on their appropriate use (Araujo & Peterson, 2012). However, one of the major shortcomings of ENMs for climate change applications is that they do not explicitly incorporate a species' propensity to colonise new locations, nor do they model the rate at which species will disappear from old locations (Elith *et al.*, 2010).

It has been argued that a more detailed understanding of species responses to environmental change, and an improvement in forecasts of species range shifts, requires more mechanistic models of range dynamics (Fordham *et al.*, 2014, Singer *et al.*, 2016). One approach is to combine species demography with climate suitability (typically derived from ENMs) and dispersal across a landscape, to simulate population dynamics within an integrative framework (Fordham *et al.*, 2013a). These models are potentially more realistic and less prone to bias than correlative ENMs alone (Zurell *et al.*, 2016), because they account for potentially important metapopulation processes, and multiple human impacts (Ehrlén & Morris, 2015, Fordham *et al.*, 2013c). Furthermore, they can directly measure extinction risk (population declines and other measures of stochastic viability), as well as change in habitat area, when assessing climate change impacts on biodiversity (Fordham *et al.*, 2012). Despite their increased popularity, demographic models linked to ENMs require more detailed field data for parameterisation, and have never been validated against real-world independent data. Such testing is critical for determining whether the addition of key information on species' vital rates (e.g., growth rate, density dependence) improves estimates of range shifts and extinction risk under climate change scenarios.

Here, we predict historical range shifts in response to past climate change using coupled demographic ENMs, and assess model performance using observed changes in species' ranges. In addition, we test if the performance of these complex models is superior to simpler correlative ENMs. To realise this aim, we made use of a unique dataset: a breeding bird population census of 20 species in the United Kingdom (UK) in 1970 and 2010. Because the 2010 dataset has only recently become available, results from previous model-validation studies have been based on observed range movements over 20 years (1970 – 1990) rather than 40 years period (e.g., Araujo *et al.*, 2005, Pearce-Higgins *et al.*, 2015, Rapacciuolo *et al.*, 2012)

We trained models of different complexity, using the species' occurrence in the 1970s, species' dispersal constraints, and demographic parameters estimated from time series abundance data centred on 1970, to forecast species ranges in 2010, where we had (independent) test distribution data. In total, we fitted 8 model types (Fig. 1) that represented a gradient from simpler correlative ENMs, to dispersal-linked ENMs, through to complex spatially explicit population models. Our general aim was to examine whether theoretically more realistic models (by virtue of accounting for a greater level of ecological processes) would deliver improved approximations of observed species range shifts. More specifically, we (i) compared models with and without the inclusion of land use and demographic processes (i.e., population growth and colonisation and extinction dynamics), (ii) evaluated models with different dispersal hypotheses (no dispersal, a distance-decay dispersal function, unlimited dispersal), and (iii) examined which models are more likely to result in 'false negative' or 'false positive' errors. Our results improve knowledge of whether simpler models are more transferable.

Material and Methods

Data

Bird occurrence, climate and land use data

British birds are one of the most extensively surveyed faunas in the world (Baillie *et al.*, 2014). We compiled occurrence data for 20 British bird species using the Breeding Atlas 10km grid-cell data for two time periods: 1968–72 (t_1), 2007–2011 (t_2). Species were chosen on the basis that (i) there were sufficient data available to parameterise demographic models; and (ii) they were not seasonal migrants or shore birds because of difficulties in defining explicit areas of occupancy for these groups (Bradshaw *et al.*, 2014). We also excluded species with fewer than 20 occurrence records in the first recording period, because of problems related to modelling data with small sample sizes (Stockwell & Peterson, 2002). There is a risk of biases in extrapolations associated with incomplete characterizations of climatic niches (Pearson *et al.*, 2004), potentially inflating forecast risks from climate change (Araújo *et al.*, 2011). To avoid this problem, data for birds for the baseline period (1968-1972) from the British Bird Atlas were amalgamated with European wide distributions at 50km grid-cell resolution from a digitized version of the Atlas of European Breeding Birds (Hagemeijer & Blair, 1997). Ecological niche models were thus initially trained with British and European wide distributions data, and then projected in Britain alone (Pearson *et al.*, 2004).

Annual mean values (1970–2000) for mean minimum temperature in February (°C), mean maximum temperature in July (°C), and mean total annual precipitation (millimetres) were derived from the Climate Research Unit (CRU) monthly climate data (New *et al.*, 2000). The data provide monthly values for 1901–2000 in a 10' (c.a. 16 x 16 km grid) resolution that overlapped with the terrestrial area of Europe, including England, Wales, and Scotland. The baseline climate record was extended to 2010 using the ALARM business-as-might-be-usual (BAMBU) storyline (Fronzek *et al.*, 2012). These combined data provided a homogenous 40-year annual-step climate data series. The difference between ALARM storylines were negligible for the UK in 2010 (see Supporting Information, Fig. S1); and are similar to more recent CRU TS 3.2 data (< ± 0.5 °C for temperature variables and < ± 0.25 mm/day for annual precipitation; Harris *et al.*, 2014). Our choice of climate variables reflects those known to impose constraints on bird ranges as a result of widely shared physiological limitations (Crick, 2004, Lennon *et al.*, 2000), having been used in several modelling studies of birds in the UK (Araujo *et al.*, 2005, Bradshaw *et al.*, 2014, Pearson *et al.*, 2004). All climate data were projected onto the British Breeding Bird Atlas occurrence 10km-grid and the European Breeding Birds Atlas 50km-grid using cubic spline interpolation.

We generated species-specific land-cover suitability maps using CORINE vector datasets for 1990, 2000, and 2006. We obtained these maps after: (i) classifying land cover classes as suitable or unsuitable at a 25m grid-cell resolution; and (ii) calculating the proportion of suitable land-cover within a 10km grid-cell. We used five land use categories (wet habitats, crop, pastures, forests, shrub lands), corresponding to the main habitats used by the birds in our study. The output was grid-cell habitat suitability values varying continuously between 0 and 1, representing the proportion of suitable habitat per grid cell (i.e., 0 = no suitable habitat; 1 = entire grid-cell consisted of suitable habitat). Habitat classifications were based on information from the British Bird Atlas (Baillie *et al.*, 2014) and expert advice. Spatial variation in land-cover suitability was similar between the three available time periods (Table S1). Therefore, to avoid the risk of uncertainty propagating through to results due to classification errors in the temporal sequence of land use layers (Mouquet *et al.*, 2015), and because we did not have land use data for the initial 20 years of the simulation, we treated land-cover suitability (centred on 1990, the midpoint of the simulation) as a static variable in the models focused on the mid-point of the study period. For species specific maps of land use suitability see Supporting Information, Fig. S2 and Table S1

Modelling

Using ecological niche models as the simplest modelling unit, we generated seven alternative model types with differing levels of complexity (Fig. 1). These models fell into three families (described in more detail below): simple Ecological Niche Models (ENMs); dispersal-linked niche models (Dispersal); and niche-population models (Metapop). More, specifically, these models were 1) ENMs affected by climate change and assuming unlimited dispersal (thereafter referred to as *ENM_UD*), 2) ENMs affected by climate change and land use and assuming unlimited dispersal (*ENM_UD_LU*), 3) ENMs affected by climate change and assuming no dispersal (*ENM_ND*), 4) ENMs affected by climate change and land use and assuming no dispersal (*ENM_ND_LU*), 5) ENMs affected by climate change and species' specific dispersal constraints (*DISPERSAL*), 6) ENM affected by climate change, land use and species' specific dispersal constraints (*DISPERSAL_LU*), 7) ENMs affected by climate change and species' specific extinction and colonization dynamics (*METAPOP*), and 8) ENMs affected by climate change, land use and species' specific extinction and colonization dynamics (*METAPOP*_LU).

Ecological Niche Models

We used 12 different ecological niche modelling techniques fitted with climate and land use predictors, using BIOENSEMBLES (Diniz-Filho *et al.*, 2009). An ensemble of ENMs was generated for each one of the 20 species considered. Ensemble forecasting approaches account for inter-model variation in predictions (Araújo & New, 2007), and there is empirical evidence that consensus predictions derived from multiple models within ensembles can improve projections of individual models in contexts of transferability under climate change (Araujo *et al.*, 2005). We fitted ensembles of forecasts using the following techniques: BioClim; Euclidian Distance (EUC); Gower Distance (GOW); Mahalanobis Distance (MAH); Generalized Linear Models (GLM); Generalized Additive Models (GAM); Random Forests (RF); Genetic Algorithm for Rule-set Production (GARP); Ecological Niche Factor Analysis (ENFA); MaxEnt, Neural Networks (NN); and Multivariate Adaptive Regression Splines (MARs). BIOCLIM, MAH, EUC and GOW were fitted to species occurrence records (presence only), while MaxEnt, ENFA and GARP use background information, describing a random sample of non-occurrences from the region of interest. GLM, GAM, RF, NN and MARs were parameterized assuming that absences represent true absence of the species. By varying the assumptions regarding absence data we captured the variability in projections accrued from such assumption in the models.

Models were calibrated using European-wide occurrence data for t_1 matched to average climate data for 1968–72. Models were trained using 80% random sample of the initial data and tested against the remaining 20% of data (Fielding & Haworth, 1995). Accuracy of predicted distributions in the training set were measured for every model using the area under the curve (AUC) of the receiver operation characteristic (ROC) and the true skill statistic (TSS) (Liu *et al.*, 2005). Models with low performance (TSS<0.3) were discarded from the ensemble (Garcia *et al.*, 2012). The remaining ENMs were used to predict probability of occurrence or climate suitability and presence and absence at annual time-steps from 1970 to 2010. Modelled probabilities or climate suitabilities were transformed into predictions of presence and absence of species in the grid cells, using thresholds defined by AUC for presence-absence models and fixed cut-offs for presence only models. Consensus about the predicted distribution of the species was obtained by recording the areas where at least 40% of the models agreed that the species would occur there (Araújo *et al.*, 2011).

In total, we generated four different types of ENMs (see above and Fig. 1). The ‘unlimited dispersal’ scenario (ENM_UD) assumes that the species can completely migrate to future suitable areas (in 2010), that no individuals remain in unsuitable grid cells, and that all suitable areas are occupied. In other words, species are presumed to be constantly in equilibrium with climate (Araujo & Peterson, 2012). The ‘no dispersal’ scenario (ENM_ND) assumes that the species cannot migrate beyond its observed range for the training period (1970). Therefore, only grid cells that were suitable in 1970 and 2010 were assumed to be occupiable between these time periods. The ENMs with climate, land use and unlimited dispersal (*ENM_LU*) were generated by classifying grid cells as unsuitable if land use suitability was below a minimum area threshold (even if the ENM classified those grid-cells as climatically suitable) of 0.0025 (i.e., a grid-cell needed 250m x 250m of suitable habitat to be considered habitable) needed to sustain a breeding pair of birds. This is likely to be a conservative threshold for some bird species in our analysis. The ENMs with climate, land use and no dispersal (*ENM_LU_ND*) used ENM_LU predictions but assumed that only grid cells that were occupied in both 1970 and 2010 were occupiable between these time periods.

Dispersal Models

Projections by ENMs of future grid cells suitable for colonization were linked to a stochastic dispersal model, using a cellular/lattice spatial structure consisting of 2 665 grid-cells (10 km x 10 km longitude/latitude grid-cell resolution). Cells were classed as either suitable or not suitable at each time step. Natal dispersal was modelled using published estimates from Paradis *et al.* (1998). More specifically, we used species’ specific dispersal kernels to model the probability of dispersal between grid cells of suitable habitat during each time step as an exponential function: $P = a.\exp(D^{1/b})$, where D is the distance between grid cell centroids, a is the proportion of individuals that disperse in all radial directions, b is the mean dispersal distance of the species (Akcakaya & Root, 2005). When D exceeds a specified maximum distance (D_{max}) that a species is expected to be able to disperse P is set to zero. See Table S2 for species-specific dispersal parameters. We modelled a high level of stochasticity in dispersal rates (co-efficient of variation = 1; Paradis et al. 1998) and assumed that colonised cells stabilise at a species-specific maximum density (set at maximum K between 1970 and 2010; see below) within a three year period using an exponential population growth function. The approach we used is similar to MigClim (Engler *et al.*, 2012), in that the model's basic unit is a cell that is occupiable or not, dispersal is defined by a dispersal kernel and propagule pressure is a function of the time since colonisation. It differs from MigClim in that it models the probability of the proportion of individuals that move between cells, not the probability of a dispersal event. This has both advantages and limitations (Engler & Guisan, 2009), however, we chose this method because outputs are directly comparable to those from the coupled niche-population models described below. The dispersal-only models were implemented in RAMAS Metapop (Akcakaya & Root, 2005). The model was initialised using the approach described for coupled niche-population models (see below), allowing us to directly compare model output with and without stochastic population growth and its interaction with dispersal.

Coupled niche-population models

ENMs with climate and with and without land use were coupled with a stochastic population model that captures extinction as well as colonization dynamics (*Metapop*) by simulating landscape-level population processes and dispersal with source and sink dynamics (Fordham *et al.*, 2013a). The demographic models for British birds used an identical cellular/lattice spatial structure to the dispersal-only models and were implemented in RAMAS Metapop. Each grid cell was modelled with a scalar-type stochastic model, which simulates the finite rate of population increase “R”, its variance and the population carrying capacity (Dunham *et al.*, 2006). The carrying capacity of birds in each grid cell for simulations with land use was calculated as:

$$K = thr\,(maximum\ abundance \times land\ cover \times climate\ suitability,\ minimum\ abundance)$$

Where, *maximum abundance* was the highest density of birds expected in a 10 km grid-cell when land cover = 1 and climate = 1. If abundance at any time was less than a *minimum abundance*, then a threshold function (*thr*) set abundance to zero, simulating a simple Allee effect. *Land cover* was the proportion of the grid cell that is potentially habitable based on land use type. *Climate suitability* was the output of the ecological niche model (with no threshold for prevalence), scaled between 0 and 1. The minimum abundance value was set iteratively by maximising the kappa score (Monserud & Leemans, 1992) between simulated and observed range in t_1 (i.e., 1970) using 10-fold cross validation for minimum abundance values ranging from 1 to 1,000 females per 10 x 10 km grid cell.

The carrying capacity of birds in each grid cell for simulations without land use was calculated as:

$$K = thr\,(maximum\ abundance \times climate\ suitability \times max\ land\ cover,\ minimum\ abundance)$$

Where, *max land cover* is the maximum area of suitable land use in any given 10km x 10km cell divided by the area of that cell. This scaling parameter prevents superabundant populations that can arise as a result of the relatively coarse spatial resolution of the model (10 km grids) (Fordham *et al.*, 2013b). The minimum abundance value was set using an identical technique to K with land use. Climate suitability was the same for *K* with and without land use.

We used long-term population dynamics time-series data to calculate finite rates of population increase and their variance (Brook & Bradshaw, 2006). The minimum length of these time series were 12 year-to-year transitions with a mean duration of 27 year-to-year transitions. The time series overlapped closely with the study period (mean focal year = 1965), which is close to the year used to calibrate the ENMs (1970), and, therefore, reasonable as the basis to estimate demographic parameters in the *Metapop* models. Estimates of maximum finite rate of population increase (R_{max}) and standard deviation around the intrinsic rate of population growth were calculated following Brook and Bradshaw (2006) and are reported in Table S2. The standard deviation value was used to model population fluctuations driven by environmental stochasticity (Fordham *et al.*, 2013b). We used multi-model inference (Burnham & Anderson, 2002) to assign strengths of evidence for two population dynamics models commonly used to describe phenomenological time series data: a density independent model (random walk) and a density dependent model (Ricker-logistic population growth). On this basis, each species was assigned either a density independent or density dependent model of population growth (Table S2). Density dependence was modelled using the "scramble competition" function in RAMAS, whereby as population abundance in a grid-cell increases, the amount of resources per individual decreases, as dictated by *K*. Density independence was modelled independent of *K,* by allowing *K* to affect grid-cell abundance only when climate and /or land use suitability = 0. The proportion of dispersers moving between grid cells of suitable habitat during each time step was modelled using a species-specific dispersal kernel and a CV = 1 (see Dispersal-only Model). Stochasticity in dispersal was driven by temporal variability in population growth rate as well as variability in the natal dispersal kernel.

Initial abundance in the first time step (*t*) was firstly modelled as being equal to 80% of *K*. A burn-in period of 10 years (1 000 iterations) was used to generate a stable initial equilibrium abundance and patch (occupied grid-cell) structure under the assumption of constant 1970 climate conditions (Fordham *et al.*, 2012). All simulations were based on 1 000 stochastic replicates and run over a 41 year period (i.e., 1970–2010).

Independent Model Testing

We compared observations and predictions for all 20 species using the three types of models with varying levels of realism and complexity (ENM, Dispersal, Metapop). Specifically, we compared observed and predicted spatial patterns of species ranges for t_2, and changes in range size between t_1 and t_2. This allowed us to identify models that give similar spatial projections and make generalisations across species regarding which model types best describe observed range dynamics (Garcia *et al.*, 2012). Change in range area between t_1 and t_2 was calculated as the difference between the number of 10 km grid-cells gained by the species (i.e., sites where the species was present in t_2 but absent in t_1) and the number of sites lost (i.e., sites where the species was absent in t_2 but present in t_1) relative to the total number of sites occupied in t_1 and t_2 (i.e., the stable range) (Delean *et al.*, 2013).

We used the True Skill Statistic (TSS) to compare predicted with observed patterns of presences and absences in 2010 for each species. This metric has been shown to be a simple and intuitive measure for discerning the accuracy of predictions when they are expressed as presence-absence maps (Allouche *et al.*, 2006). Because choice of evaluation metric can influence estimates of predictive accuracy (Allouche *et al.*, 2006) we also show results for area under the Receiver Operating Characteristic curve (AUC) (Swets, 1988). We used Generalized Linear Models (GLM, Gaussian-identity distribution-link) to explore the relative importance of different predictor variables on overall prediction accuracy (TSS and AUC), omission (falsely predicted absences) and commission errors (falsely predicted presences) i.e., based on results from 20 species x 8 different model configurations (n = 160). The predictor variables were '*species*', '*model type', 'land use'* and *'dispersal type'*. In each case, we inspected model residuals for normality and then chose an exponential transformation for TSS and omission error; and a log transformation for commission error. These transformations achieved normality for the response variable. We compared these models to a null model, which assumes a single rate across '*species*', '*model type'*, *'land use'* and *'dispersal type'*. For each GLM, we calculated the log-likelihood (LL), percentage of deviance explained, change in AIC compared to the best-ranked model (ΔAIC), model weights (ωAIC). To avoid over parameterising GLMs we tested single term models for omission and commission errors and models with simple two-way interaction terms (*model type* : *land use*; *dispersal type* : *land use*) for predictive accuracy.

Results

Predicting changes in patterns of occurrence

Projections of geographic patterns of range contraction and expansion varied considerably across models and species (Fig. S3), as illustrated in detail for two selected species (Fig. 2 & 3). We show that both the choice of dispersal type (no dispersal, unlimited dispersal, dispersal function) and how to model land use (and their interaction) influenced model skill in predicting observed patterns of occurrence in 2010 (Fig. 4 & Fig S4).

A multi-termed model with explanatory variables *dispersal* and *land use* (and their interaction) had the largest effect on predictive accuracy based on TSS (TSS ~ *species*; ω_i = 0.78), explaining 22% of the variance when compared to the null model (Table 1). There was also some support for the next two best-ranked models that modelled TSS as a function of *species* (TSS ~ *model*; ω_i = 0.13, ΔAIC_c =3.58, DEV = 36%); and as a function of type of *model*, *land use* and their interaction (TSS ~ *model : land use*; ω_i = 0.01, ΔAIC_c =0.01, DEV = 24%). Likewise, the choice of method of dispersal, and whether to consider land use, had the largest influence on AUC predictive accuracy (AUC ~ *dispersal : land use*; ω_i = 0.89), explaining 24% of structural deviance (Table 1). Including land use in ENM models tended to improve predictions of occurrence patterns in 2010 (Figures 4 & 5). In strong contrast, including land use in DISPERSAL and METAPOP models tended to provide less accurate predictions of occurrence patterns (Figures 4 & 5). In general, DISPERSAL and METAPOPmodels were most skilful in predicting changes in occurrence patterns, but only if land use was not considered in the model.

Predicting changes in range area

Models tended to do a fair-to-good job at predicting observed proportional changes in range area, regardless of model type (Figure 6). The difference from observed values was low (< 10%) for ≥ 50% of the birds modelled (with and without land-use) using ENMs (n = 10-11), ≥ 45% of birds with DISPERSAL models (n = 9-11), and ≥ 30 % of birds with METAPOP models (n = 6-9). The median difference between observed and predicted change in range area was 8.6 – 9.8 % for ENMs, 10.1 – 11.2 % for DISPERSAL and 9.9 – 16.3 % for METAPOP. Although skill in predicting changes in range area for a given modelling approach varied across species (Figure 6), all models did poorly at predicting observed range increases and decreases for some species. For example, no models were able to accurately predict the large range expansion undergone by *Alectoris rufa* or the large contraction experienced by *Perdix perdix* (Fig. 6). There were only three species (*Carduelis cannabina*, *Corvus corone* and *Pyrrhula pyrrhula*) for which all eight models predicted < ± 10% difference between observed and predicted net losses or gains of habitat (Figure 6). None of these species were in the upper or lower quartiles for observed range movement (i.e., they did not undergo relatively large levels of range expansion or contraction during the observation period). On average, models consistently predicted the correct direction of observed change (i.e., expansion or contraction) in range shifts in about 50% of cases (Figure S5), ranging from 25 - 35% (ENM_ND_LU and ENM_ND) to 60 to 70% (METAPOP_LU, METAPOP) depending on the type of model (10 ± 1.03 species, mean ± standard error).

In general more complex models without land use tended to better predict range size in 2010 (Table S3). The median difference between observed and predicted range size in

2010 was ± 9% for METAPOP, ± 10% for DISPERSAL, ± 13% for ENM_ND, ± 22% for ENM_UD. Accounting for land use greatly improved predictions of 2010 range size for ENM_UD_LU (± 13%), but reduced predictive accuracy for DISPERSAL and METAPOP models (± 20% and 23%, respectively). Masking unsuitable land-use types had no noticeable effect on predictions of 2010 range size for ENM_ND_LU (± 13%).

Commission and omission errors

There was greatest AIC support for modelling commission errors (falsely predicted presences) as a function of '*model type*' (ω_i = 0.63, DEV = 25.3%). There was slightly less support for the alternative hypothesis that '*dispersal type*' affects commission errors (ω_i = 0.37, ΔAIC_c =1.1, DEV = 23.8%). There was much less support for modelling commission errors as either a function of '*land use*', or '*species*' (ω_i = 0, ΔAIC_c =34.7, DEV = 4.8%; ω_i = 0, ΔAIC_c =41.8, DEV = 23.4%), compared to '*model type*'. Using ENMs to predict occurrence patterns in 2010 generally resulted in larger commission errors (Figure S6), particular when land use was not used to mask out unsuitable areas for occupancy because of non-climatic factors.

The explanatory variable '*land use*' had the largest effect on omission errors (falsely predicted absences), explaining 22.5% of model structural deviance (ω_i=0.68). Including land use resulted in higher omission errors. There was much less support for modelling omission errors as a function of '*model type*', '*dispersal*' and '*species*' (ω_i = 0, ΔAIC_c =15.1, DEV = 17.1%, ω_i = 0, ΔAIC_c =17.4, DEV = 14.7%, ω_i = 0, ΔAIC_c =38.8, DEV = 24.1%). Including land use in model predictions resulted in greater omission errors (Figure S6).

Discussion

Using independent validation data over a 40-year period, we found support for the view that more realistic and complex coupled niche-population models are likely to have higher predictive accuracy in forecasting species range shifts than structurally simpler models that only account for variation in climate (Ehrlén & Morris, 2015, Fordham *et al.*, 2013c). However, these better predictions of observed presence-absence patterns were only achieved when the effects of a single static snapshot of land use (focused on the midpoint of the simulation) on dispersal and other demographic processes were not considered in model simulations. In strong contrast, a mixture of static land use and dynamic climate variables improved ecological niche model forecasts of observed range shifts. These results reinforce the need for using statistically independent data to validate model outputs prior to making firm conclusions about the relative value of alternative modelling options (Araújo & Rahbek, 2006).

Comparison of models with and without land use

Approaches for combining dynamic and static environmental variables in range dynamics models for forecasting range shifts under climate projections remain poorly understood (Stanton *et al.*, 2012). Using maps of land use in 1990 to mask out areas that are unsuitable because of non-climatic factors in 2010, tended to improve predictions of observed range shifts from ecological niche models, by reducing false positive predictions (commission errors; Figure S6), with two rare exceptions. For *Cygnos olor* and *Gallinula chloropus* the addition of land use in ecological niche models largely reduced predictive performance (Figure 5). It might be that our classification of unsuitable habitat for these two species was incorrect and they were more flexible in their habitat requirements than assumed or that land use preferences changed markedly for these two species between 1990 and 2010. Alternatively, they might today be 'committed' to extirpation in these areas with unsuitable land use, and these delayed local extinction events have not yet been realised (Fordham *et al.*, 2016).

In strong contrast, masking out areas that are unsuitable for occupancy because of non-climatic factors in models with species' specific dispersal constraints (*DISPERSAL_LU*) and extinction and colonization dynamics (*METAPOP*_LU) resulted in a decrease in predictive accuracy, brought about by higher false negative predictions (Omission errors; Figure S6). This is because these more mechanistic approaches model dispersal and metapopulation processes as dynamic functions of land use (as well as climate suitability) continuously (usually at annual time steps) for the entire simulation period. By using a static snapshot of land use, focused on the midpoint of the simulation, the interactions between land use and ecological processes were simulated under the unlikely assumption that land use in the UK did not change between 1970 and 2010, and that the snapshot is a reliable projection of land use 20 years before and after 1990. Since land use and land cover in the UK has changed over short timescales since 1970 (Rounsevell & Reay, 2009), mismatches between simulated and actual land use prior to 1990 is likely to have resulted in the propagation of incorrect trajectories of species range movement early in the simulations, leading to inaccurate maps of presences and absences in 2010.

Although land use has been assumed to be important in models of range dynamics (Triviño *et al.*, 2011) and its capacity to improve range predictions has been tested previously using virtual species ranges (Stanton *et al.*, 2012), our new analysis provides an important test of these findings using a real-world independent validation dataset. We show that the common practice of using static land use predictors in coupled niche-population models to continuously mask out areas from forecasts that are unsuitable because of non-climate factors (e.g., Fordham *et al.*, 2013b, Harris *et al.*, 2012) should be abandoned if there is a high probability that land use will change over the simulation period. However, applying restrictive masks to ecological niche model predictions of habitat suitability will improve forecasts if land use closely resembles the future landscape. This is because the end point is what matters for the predictive accuracy of statistical-based models, not the conditions leading up to this end point (i.e., the road travelled). Models fitted with land use tended to have higher omission rates then models without land use regardless of model complexity.

Comparison of models with different dispersal hypotheses

In our models, we used a gradient of different dispersal hypotheses from unlimited to no dispersal and, unsurprisingly, the more restrictive assumptions (no dispersal or a dispersal function) generated predictions of smaller range sizes than unlimited dispersal. Different model types with different dispersal hypotheses explained > 20 % of the variance in the predictive accuracy (TSS, AUC), when choice of whether or not to model land use was also considered. Models with species' specific dispersal constraints and no land use tended to provide the most accurate presence/absence maps in 2010. This result supports the view that models, which explicitly simulate dispersal, should provide improved predictions of range shifts (Bocedi *et al.*, 2014), but only when there are reliable enough projections of land use change to effectively simulate the dynamic interaction between land use and dispersal. Not constraining dispersal in ecological niche models (ENM_UD) always resulted in lower TSS values then ecological niche models that assumed no dispersal at all (ENM_ND). This result provides a cautionary note for the common practice of using ecological niche models with an unlimited dispersal simplification to forecast species range movement under climate change, and differs from conclusions based on model convergence (Engler *et al.*, 2009), as opposed to our independent model validation.

Complex vs. simple models

Simpler models might theoretically be expected to outperform more complex demographic models because they are arguably more transferable, due to their generality (Bell & Schlaepfer, 2016, Randin *et al.*, 2006). Although, more complex (and potentially more ecologically realistic) models that included dispersal and other demographic processes as explicit parameters tended to improve predictions of recent range changes for our sample of British breeding birds, large levels of variation in predictive performance (TSS and AUC) were found among species. For example, the simplest model for *Alectoris rufa* − fitted with only climate variables and assuming unlimited dispersal − had as high, or higher, TSS than any of the alternative models that accounted for land use, dispersal or demography (Fig. 5). In contrast, species like *Accipiter nisus* were better modelled by the most complex METAPOP and DISPERSAL models without land use (Fig. 2). Similar results were found in a recent study that systematically examined model performance against complexity for families of ecological niche models (García-Callejas & Araújo, 2016), whereby properties of species ranges strongly influenced model performance (even more than model complexity).

The critical question is whether it is possible to classify (and predict) the circumstances in which different species are best predicted by different models. This is still an open question, but our results suggest that good estimates of dispersal dynamics and close approximations between future land use and species' occurrence will improve forecasts of species distributions. When there is scarce knowledge on a species dispersal dynamics, but their relationship between land use and occupancy is well understood for the model calibration period, and this relationship is unlikely to change greatly in space in the future, forecasts of species distributions will be maximised using simple ecological niche models with static land use masks. Conversely if species' dispersal dynamics are well documented, but the effect of land use on spatial colonisation patterns is unlikely to be static, forecasts of species' distributions will be maximised using a DISPERSAL model without land use. If robust estimates of population growth as well as dispersal constraints are available, and the model is to be used to estimate extinction risk as well as range movement, a METAPOP model should be used since the relationship between change in range area and extinction risk is often weak (Fordham *et al.*, 2012). The METAPOP model should be simulated with land use, only if land use is not expected to vary, or if spatiotemporal change in land use can be accurately projected. Furthermore, by modelling spatiotemporal abundance, coupled niche-population models not only allow extinction risk to be directly quantified, but the cost effectiveness of regional conservation alternatives and demographically oriented management interventions to be tested (Fordham *et al.*, 2013a).

Our research shows that for many species of British breeding birds, ecological niche models can provide a good approximation of the magnitude (but not necessarily the direction) of climate driven changes in geographic extent. Therefore if the objective of the study is to identify species that are likely to experience large range contractions and expansions in the future (regardless of where these play out in space and time), simple ecological niche models can routinely provide as good if not better predictions then more complex models. Similarly, Rapacciuolo *et al.* (2012) used temporally independent records to show that ecological niche models did well at predicting observed changes in total range area despite failing to predict correctly specific range changes at the grid cell level.

Previous findings and limitations

Previous studies have advocated the use of more complex range dynamics models that overcome some of the limitations of correlative ecological niche models by integrating demographic and physiological responses so that range shifts emerge from the interplay of relevant abiotic and biotic processes (Singer *et al.*, 2016). The use of more complex models to explore how changes in large-scale abundance distributions arise is leading to a more mechanistic understanding of the underlying processes of range dynamics (Lurgi *et al.*, 2015). Although the methodological frameworks of dynamic range models have been developed, empirical tests and applications of these models are rare because demographic data and time series of local abundances remain scarce (Urban *et al.*, 2016).

Accordingly, Zurell *et al.* (2016) recently compared model types using simulated data. They concluded that under present-day climatic conditions, complex demographic models are only marginally better than simple correlative models. However, in rapidly changing climates, complex range dynamic models that account for dispersal and/or demography, are likely to provide better forecasts. When community processes were included in simulated benchmarking data, and models were tested under conditions that better approximate real world conditions, DISPERSAL type-models often proved most reliable. In our study, biotic interactions were not directly considered in model forecasts, yet real world benchmarking data also revealed good evidence for using more complex models to predict where occupancy status changed due to observed climate change. Our study inevitably focused on low levels of observed climate change in the UK over the last 40 years, which are small compared to what is forecast for the future (Fordham *et al.*, 2016). Therefore, caution must be shown when using our results to make generalisations regarding how well models of species range dynamics will do at predicting range movement for the twenty-first century and beyond, because a models ability to predict (limited) 20th century climate-driven range movement does not necessarily equate to better predictions in response to forecast (larger) climate exposure (Fordham *et al.*, 2016 , Rapacciuolo *et al.*, 2012). Nevertheless, the in silico findings by Zurell *et al.* (2016) that models with dispersal and/or demography provide better predictions as climate change intensifies, gives us some confidence that our results will hold true, even under more extreme climate change.

The most complex models in our study were scalar-based demography models, which can be useful for ecological assessments but can overestimate risk of extinction (Dunham *et al.*, 2006). Further analysis should focus on testing more complex demographic models (e.g., stage/age structured demographic models (Caswell, 2001); Bayesian models of source-sink dynamics (Schurr *et al.*, 2012)) and improving model parameterisation using Approximate

Bayesian Computing (ABC) techniques to calibrate metapopulation models (Rougier *et al.*, 2015). These techniques offer the prospect of accumulative fine tuning of model parameters via the iterative re-casting of updated information in the prior distribution (van der Vaart *et al.*, 2015, Wells *et al.*, 2015). We suspect that our most complex (METAPOP) models would have provided even better predictions of species range dynamics if they captured life-history traits that permit population density to vary in different ways in response to key local spatial drivers (Csergő *et al.*, 2017), including dynamic land use change and recent conservation intervention.

Acknowledgements

The Australian Research Council supported D.A.F, C.B. and B.W.B (FT140101192, DP1096427, FT100100200, respectively). M.B.A and D.N. acknowledge support from the Foundation for Science and Technology (PTDC/AAG-GLO/0463/2014 and CIBIO - UID/BIA/50027/2013 (POCI-01-0145-FEDER-006821) and the Danish NSF. We thank T. Blackburn and C. Şekercioğlu for expert advice and the regional organisers, validators and volunteers responsible for the occurrence data used in this study.

References

Akcakaya HR, Root WT (2005) *RAMAS GIS: linking landscape data with population viability analysis (version 5.0),* Setaukey, N.Y., Applied Biomathematics.

Allouche O, Tsoar A, Kadmon R (2006) Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology,* **43**, 1223-1232.

Araujo M, Peterson AT (2012) Uses and misuses of bioclimatic envelope modelling. *Ecology,* **93**, 1527-1539.

Araújo MB, Alagador D, Cabeza M, Nogués-Bravo D, Thuiller W (2011) Climate change threatens European conservation areas. *Ecology Letters,* **14**, 484-492.

Araújo MB, New M (2007) Ensemble forecasting of species distributions. *Trends in Ecology & Evolution,* **22**, 42-47.

Araujo MB, Pearson RG, Thuiller W, Erhard M (2005) Validation of species-climate impact models under climate change. *Global Change Biology,* **11**, 1504-1513.

Araújo MB, Rahbek C (2006) How does climate change affect biodiversity? *Science,* **313**, 1396-1397.

Baillie S, Marchant J, Leech D *et al.* (2014) BirdTrends 2014: trends in numbers, breeding success and survival for UK breeding birds. In: *Research Report 662.* Thetford.

Bell DM, Schlaepfer DR (2016) On the dangers of model complexity without ecological justification in species distribution modeling. *Ecological Modelling,* **330**, 50-59.

Bocedi G, Palmer SCF, Pe'er G, Heikkinen RK, Matsinos YG, Watts K, Travis JMJ (2014) RangeShifter: a platform for modelling spatial eco-evolutionary dynamics and species' responses to environmental changes. *Methods in Ecology and Evolution,* **5**, 388-396.

Bradshaw CJA, Brook BW, Delean S *et al.* (2014) Predictors of contraction and expansion of area of occupancy for British birds. *Proceedings of the Royal Society B: Biological Sciences,* **281**.

Brook BW, Bradshaw CJA (2006) Strength of evidence for density dependence in abundance time series of 1198 species. *Ecology,* **87**, 1448-1451.

Burnham KP, Anderson DR (2002) *Model Selection and Multimodel Inference, 2nd edn.,* New York, Springer.

Caswell H (2001) *Matrix Population Models: Construction Anlysis and Interpretation, Second Edition,* Sunderland, USA, Sinauer Associates.

Crick HQP (2004) The impact of climate change on birds. *Ibis,* **146**, 48-56.

Csergő AM, Salguero-Gómez R, Broennimann O *et al.* (2017) Less favourable climates constrain demographic strategies in plants. *Ecology Letters,* **doi:10.1111/ele.12794**.

Delean S, Bull CM, Brook BW, Heard LMB, Fordham DA (2013) Using plant distributions to predict the current and future range of a rare lizard. *Diversity and Distributions,* **19**, 1472-4642.

Diniz-Filho JA, Bini LM, Rangel TFLB, Loyola RD, Hof C, Nogués-Bravo D, Araújo MB (2009) Partitioning and mapping uncertainties in ensembles of forecasts of species turnover under climate changes. *Ecography,* **32**, 1-10.

Dunham AE, Akcakaya HR, Bridges TS (2006) Using scalar models for precautionary assessments of threatened species. *Conservation Biology,* **20**, 1499-1506.

Ehrlén J, Morris WF (2015) Predicting changes in the distribution and abundance of species under environmental change. *Ecology Letters,* **18**, 303-314.

Elith J, Kearney M, Phillips S (2010) The art of modelling range-shifting species. *Methods in Ecology and Evolution,* **1**, 330-342.

Elith J, Leathwick JR (2009) Species distribution models: ecological explanation and prediction across space and time. *Annual Review of Ecology Evolution and Systematics,* **40**, 677-697.

Engler R, Guisan A (2009) MIGCLIM: Predicting plant distribution and dispersal in a changing climate. *Diversity and Distributions,* **15**, 590-601.

Engler R, Hordijk W, Guisan A (2012) The MIGCLIM R package – seamless integration of dispersal constraints into projections of species distribution models. *Ecography,* **35**, 872-878.

Engler R, Randin CF, Vittoz P, Czaka T, Beniston M, Zimmermann NE, Guisan A (2009) Predicting future distributions of mountain plants under climate change: does dispersal capacity matter? *Ecography,* **32**, 34-45.

Fielding AH, Haworth PF (1995) Testing the Generality of Bird-Habitat Models. *Conservation Biology,* **9**, 1466-1481.

Fordham DA, Akcakaya HR, Alroy J, Saltré F, Wigley TM, Brook BW (2016) Predicting and mitigating future biodiversity loss using long-term ecological proxies *Nature Climate Change,* **6**, 909–916.

Fordham DA, Akçakaya HR, Araújo MB *et al.* (2012) Plant extinction risk under climate change: are forecast range shifts alone a good indicator of species vulnerability to global warming? *Global Change Biology,* **18**, 1357–1371.

Fordham DA, Akçakaya HR, Araújo MB, Keith DA, Brook BW (2013a) Tools for integrating range change, extinction risk and climate change information into conservation management. *Ecography,* **36**, 956-964.

Fordham DA, Akcakaya HR, Brook BW *et al.* (2013b) Adapted conservation measures are required to save the Iberian lynx in a changing climate. *Nature Clim. Change,* **3**, 899-903.

Fordham DA, Brook BW, Hoskin CJ, Pressey RL, Vanderwal J, Williams SE (2016) Extinction debt from climate change for frogs in the wet tropics. *Biology Letters,* **12**.

Fordham DA, Brook BW, Moritz C, Nogués-Bravo D (2014) Better forecasts of range dynamics using genetic data. *Trends in Ecology & Evolution,* **29**, 436-443.

Fordham DA, Mellin C, Russell BD *et al.* (2013c) Population dynamics can be more important than physiological limits for determining range shifts under climate change. *Global Change Biology,* **19**, 3224-3237.

Fronzek S, Carter TR, Jylhä K (2012) Representing two centuries of past and future climate for assessing risks to biodiversity in Europe. *Global Ecology and Biogeography,* **21**, 19-35.

García-Callejas D, Araújo MB (2016) The effects of model and data complexity on predictions from species distributions models. *Ecological Modelling,* **326**, 4-12.

Garcia RA, Burgess ND, Cabeza M, Rahbek C, Araújo MB (2012) Exploring consensus in 21st century projections of climatically suitable areas for African vertebrates. *Global Change Biology,* **18**, 1253-1269.

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

Hagemeijer W, Blair J (1997) *The EBCC Atlas of European Breeding Birds, their Distribution and Abundance,* London.

Harris I, Jones PD, Osborn TJ, Lister DH (2014) Updated high-resolution grids of monthly climatic observations – the CRU TS3.10 Dataset. *International Journal of Climatology,* **34**, 623-642.

Harris JBC, Fordham DA, Mooney PA *et al.* (2012) Managing the long-term persistence of a rare cockatoo under climate change. *Journal of Applied Ecology,* **49**, 785-794.

Illán JG, Thomas CD, Jones JA, Wong W-K, Shirley SM, Betts MG (2014) Precipitation and winter temperature predict long-term range-scale abundance changes in Western North American birds. *Global Change Biology,* **20**, 3351-3364.

Lambert E, Pierce GJ, Hall K *et al.* (2014) Cetacean range and climate in the eastern North Atlantic: future predictions and implications for conservation. *Global Change Biology,* **20**, 1782-1793.

Lennon JJ, Greenwood JJD, Turner JRG (2000) Bird diversity and environmental gradients in Britain: a test of the species–energy hypothesis. *Journal of Animal Ecology,* **69**, 581-598.

Liu C, Berry PM, Dawson TP, Pearson RG (2005) Selecting thresholds of occurrence in the prediction of species distributions. *Ecography,* **28**, 385-393.

Lurgi M, Brook BW, Saltré F, Fordham DA (2015) Modelling range dynamics under global change: which framework and why? *Methods in Ecology and Evolution,* **6**, 247-256.

Macias-Fauria M, Willis KJ (2013) Landscape planning for the future: using fossil records to independently validate bioclimatic envelope models for economically valuable tree species in Europe. *Global Ecology and Biogeography,* **22**, 318-333.

Maguire KC, Nieto-Lugilde D, Blois JL, Fitzpatrick MC, Williams JW, Ferrier S, Lorenz DJ (2016) Controlled comparison of species- and community-level models across novel climates and communities. *Proceedings of the Royal Society B: Biological Sciences,* **283**.

Monserud RA, Leemans R (1992) Comparing global vegetation maps with the Kappa statistic. *Ecological Modelling,* **62**, 275-293.

Mouquet N, Lagadeuc Y, Devictor V *et al.* (2015) REVIEW: Predictive ecology in a changing world. *Journal of Applied Ecology,* **52**, 1293-1310.

Paradis E, Baillie SR, Sutherland WJ, Gregory RD (1998) Patterns of natal and breeding dispersal in birds. *Journal of Animal Ecology,* **67**, 518-536.

Pearce-Higgins JW, Eglington SM, Martay B, Chamberlain DE (2015) Drivers of climate change impacts on bird communities. *Journal of Animal Ecology,* **84**, 943-954.

Pearman PB, Guisan A, Broennimann O, Randin CF (2008a) Niche dynamics in space and time. *Trends in Ecology & Evolution,* **23**, 149-158.

Pearman PB, Randin CF, Broennimann O *et al.* (2008b) Prediction of plant species distributions across six millennia. *Ecology Letters,* **11**, 357-369.

Pearson RG, Dawson TP, Liu C (2004) Modelling species distributions in Britain: a hierarchical integration of climate and land-cover data. *Ecography,* **27**, 285-298.

Pearson RG, Thuiller W, Araújo MB *et al.* (2006) Model-based uncertainty in species range prediction. *Journal of Biogeography,* **33**, 1704-1711.

Pereira HM, Leadley PW, Proenca V *et al.* (2010) Scenarios for Global Biodiversity in the 21st Century. *Science,* **330**, 1496-1501.

Randin CF, Dirnböck T, Dullinger S, Zimmermann NE, Zappa M, Guisan A (2006) Are niche-based species distribution models transferable in space? *Journal of Biogeography,* **33**, 1689-1703.

Rapacciuolo G, Roy DB, Gillings S, Fox R, Walker K, Purvis A (2012) Climatic Associations of British Species Distributions Show Good Transferability in Time but Low Predictive Accuracy for Range Change. *PLoS one,* **7**, e40212.

Roberts DR, Hamann A (2012) Predicting potential climate change impacts with bioclimate envelope models: a palaeoecological perspective. *Global Ecology and Biogeography,* **21**, 121-133.

Rougier T, Lassalle G, Drouineau H *et al.* (2015) The Combined Use of Correlative and Mechanistic Species Distribution Models Benefits Low Conservation Status Species. *PLoS one,* **10**, e0139194.

Rounsevell MDA, Reay DS (2009) Land use and climate change in the UK. *Land Use Policy,* **26**, S160-S169.

Rubidge EM, Monahan WB, Parra JL, Cameron SE, Brashares JS (2011) The role of climate, habitat, and species co-occurrence as drivers of change in small mammal distributions over the past century. *Global Change Biology,* **17**, 696-708.

Schurr FM, Pagel J, Cabral JS *et al.* (2012) How to understand species' niches and range dynamics: a demographic research agenda for biogeography. *Journal of Biogeography,* **39**, 2146–2162.

Singer A, Johst K, Banitz T *et al.* (2016) Community dynamics under environmental change: How can next generation mechanistic models improve projections of species distributions? *Ecological Modelling,* **326**, 63-74.

Smith AB, Santos MJ, Koo MS *et al.* (2013) Evaluation of species distribution models by resampling of sites surveyed a century ago by Joseph Grinnell. *Ecography,* **36**, 1017-1031.

Stanton JC, Pearson RG, Horning N, Ersts P, Reşit Akçakaya H (2012) Combining static and dynamic variables in species distribution models under climate change. *Methods in Ecology and Evolution,* **3**, 349-357.

Stewart JaE, Perrine JD, Nichols LB *et al.* (2015) Revisiting the past to foretell the future: summer temperature and habitat area predict pika extirpations in California. *Journal of Biogeography,* **42**, 880-890.

Stockwell DRB, Peterson AT (2002) Effects of sample size on accuracy of species distribution models. *Ecological Modelling,* **148**, 1-13.

Swets JA (1988) Measuring the Accuracy of Diagnostic Systems. *Science,* **240**, 1285-1293.

Triviño M, Thuiller W, Cabeza M, Hickler T, Araújo MB (2011) The Contribution of Vegetation and Landscape Configuration for Predicting Environmental Change Impacts on Iberian Birds. *PLoS one,* **6**, e29373.

Urban MC, Bocedi G, Hendry AP *et al.* (2016) Improving the forecast for biodiversity under climate change. *Science,* **353**.

Van Der Vaart E, Beaumont MA, Johnston ASA, Sibly RM (2015) Calibration and evaluation of individual-based models using Approximate Bayesian Computation. *Ecological Modelling,* **312**, 182-190.

Wells K, Brook BW, Lacy RC *et al.* (2015) Timing and severity of immunizing diseases in rabbits is controlled by seasonal matching of host and pathogen dynamics. *Journal of The Royal Society Interface,* **12**.

Willis KJ, Araújo MB, Bennett KD, Figueroa-Rangel B, Froyd CA, Myers N (2007) How can a knowledge of the past help to conserve the future? Biodiversity conservation and the relevance of long-term ecological studies. *Philosophical Transactions of the Royal Society B: Biological Sciences,* **362**, 175.

Zurell D, Thuiller W, Pagel J *et al.* (2016) Benchmarking novel approaches for modelling species range dynamics. *Global Change Biology,* **22**, 2651-2664.

Table 1: Generalised linear model results for True Skill Score (TSS) and Area Under the Receiver Operating Characteristic Curve (AUC)

Metric	GLM	LL	k	AICc	ΔAIC_c	ω_i	DEV
TSS	*dispersal:landuse*	-190.59	6	395.73	0.00	0.78	22.45
	species	-175.64	20	399.32	3.58	0.13	35.68
	model:landuse	-190.50	8	399.96	4.22	0.09	22.55
	dispersal	-203.99	3	416.14	20.41	0.00	8.32
	model	-203.98	4	418.21	22.48	0.00	8.34
	landuse	-209.82	2	425.72	29.99	0.00	1.39
	null	-210.94	1	425.91	30.18	0.00	0.00
AUC	*dispersal:landuse*	151.46	6	-288.38	0.00	0.89	23.81
	model:landuse	151.61	8	-284.27	4.11	0.01	23.96
	species	160.68	20	-273.32	15.06	0.00	32.11
	dispersal	130.04	3	-268.63	19.75	0.00	10.28
	model	138.39	4	-266.52	21.86	0.00	10.29
	null	129.70	1	-255.37	33.01	0.00	0.00
	landuse	130.04	2	-254.02	34.37	0.00	0.43

Log likelihood (LL), Akaike's information criterion corrected (AICc), number of parameters (k), difference in AIC between the model with the lowest AIC (ΔAIC_c), AICc weights (ω_i), percentage deviance explained (DEV). GLM predictors were *species* being modelled (n=20), *model* (ENM_ND, ENM_UD, DISPERSAL, METAPOP), *dispersal* (no dispersal, unlimited dispersal, dispersal function) and *land use* (present or absent in the model).

Figure Captions

Fig. 1 Hierarchy of eight models used to predict species' ranges for British Breeding birds in 2010 using 1970 training data. (i) Bioensembles was used to generate ecological niche models (ENM) with or without a land-use, assuming either unlimited or no dispersal, (ii) RAMAS was used to mechanistically simulate annual dispersal over the 40 year period using species specific dispersal constraints (Dispersal), (iii) RAMAS was used to simulate metapopulations as well as dispersal dynamics (Metapop). See Methods for further details.

Fig. 2 Forecasts of range expansion vary between models. An example of where more complex models without land use are best at projecting range expansion. Maps are shown for observed and predicted range change between 1970 and 2010 for *Accipter nisus.* ENM_ND = climate with no dispersal; ENM_ND_LU = climate with no dispersal and land use; ENM_UD = climate with full dispersal; ENM_UD_LU = climate with full dispersal and land use; Dispersal= Climate with a dispersal function; Dispersal_LU= climate with a dispersal function and land use; Metapop = climate with a dispersal function and population model; Metapop_LU = climate with a dispersal function and population model and land use. True Skill Score values for predictions of range change between 1970 and 2010 for *Accipter nisus* are shown in Figure 5.

Fig. 3 Forecasts of range contractions vary between models. An example of where more complex models without land use are best at projecting range contraction. Maps are shown for observed and predicted range change between 1970 and 2010 for *Parus montanus.* ENM_ND = climate with no dispersal; ENM_ND_LU = climate with no dispersal and land use; ENM_UD = climate with full dispersal; ENM_UD_LU = climate with full dispersal and land use; Dispersal= Climate with a dispersal function; Dispersal_LU= climate with a dispersal function and land use; Metapop = climate with a dispersal function and population model; Metapop_LU = climate with a dispersal function and population model and land use. True Skill Score values for predictions of range change between 1970 and 2010 for *Parus montanus* are shown in Figure 5.

Fig. 4: Independent tests of model predictions for 20 species of British birds, showing the influence of model and land-use. Observed and predicted spatial patterns of species ranges are compared using True Skill Score (TSS) for three families of models: simple ecological niche models (ENMs), dispersal-linked niche models (Dispersal) and niche-population models (Metapop); each plotted with and without land-use (LU). Results for ENMs are shown assuming no dispersal (ND) and unlimited dispersal (UD).

Fig. 5: True Skill Score (TSS) for models independently validated using observed data on occurrence in 2010. Expanders represent the upper quartile for observed range movement (> 9 % increase in range area between 1970 to 2010). Contractors represent the lower quartile (> 13% decrease in range area between 1970 to 2010). ENM_ND = climate with no dispersal; ENM_ND_LU = climate with no dispersal and land use; ENM_UD = climate with full dispersal; ENM_UD_LU = climate with full dispersal and land use; Dispersal = Climate with a dispersal function; Dispersal_LU = climate with a dispersal function and land use; Metapop = climate with a dispersal function and population model; Metapop_LU = climate with a dispersal function and population model and land use.

Fig. 6: Absolute differences in observed and predicted percentage change in range area between 1970 and 2010 (%). Expanders represent the upper quartile for observed range movement (> 9 % increase in range area between 1970 to 2010). Contractors represent the

lower quartile (> 13% decrease in range area between 1970 to 2010). ENM_ND = climate with no dispersal; ENM_ND_LU = climate with no dispersal and land use; ENM_UD = climate with full dispersal; ENM_UD_LU = climate with full dispersal and land use; Dispersal = Climate with a dispersal function; Dispersal_LU = climate with a dispersal function and land use; Metapop = climate with a dispersal function and population model; Metapop_LU = climate with a dispersal function and population model and land use.

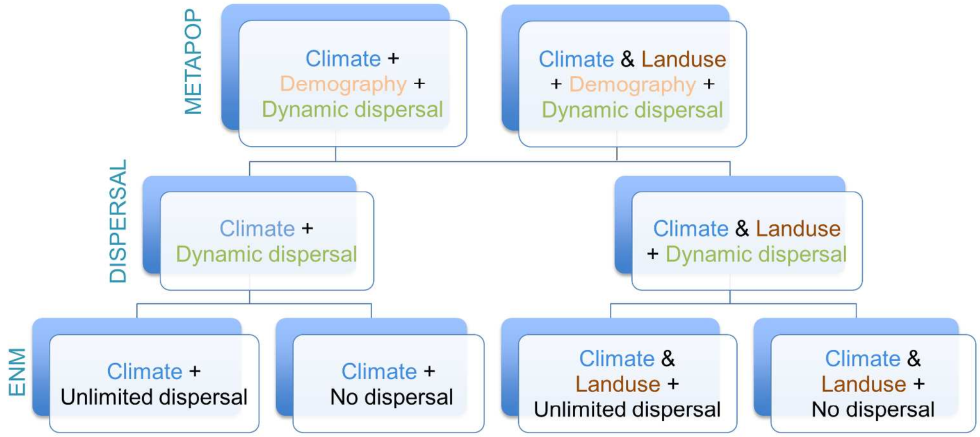
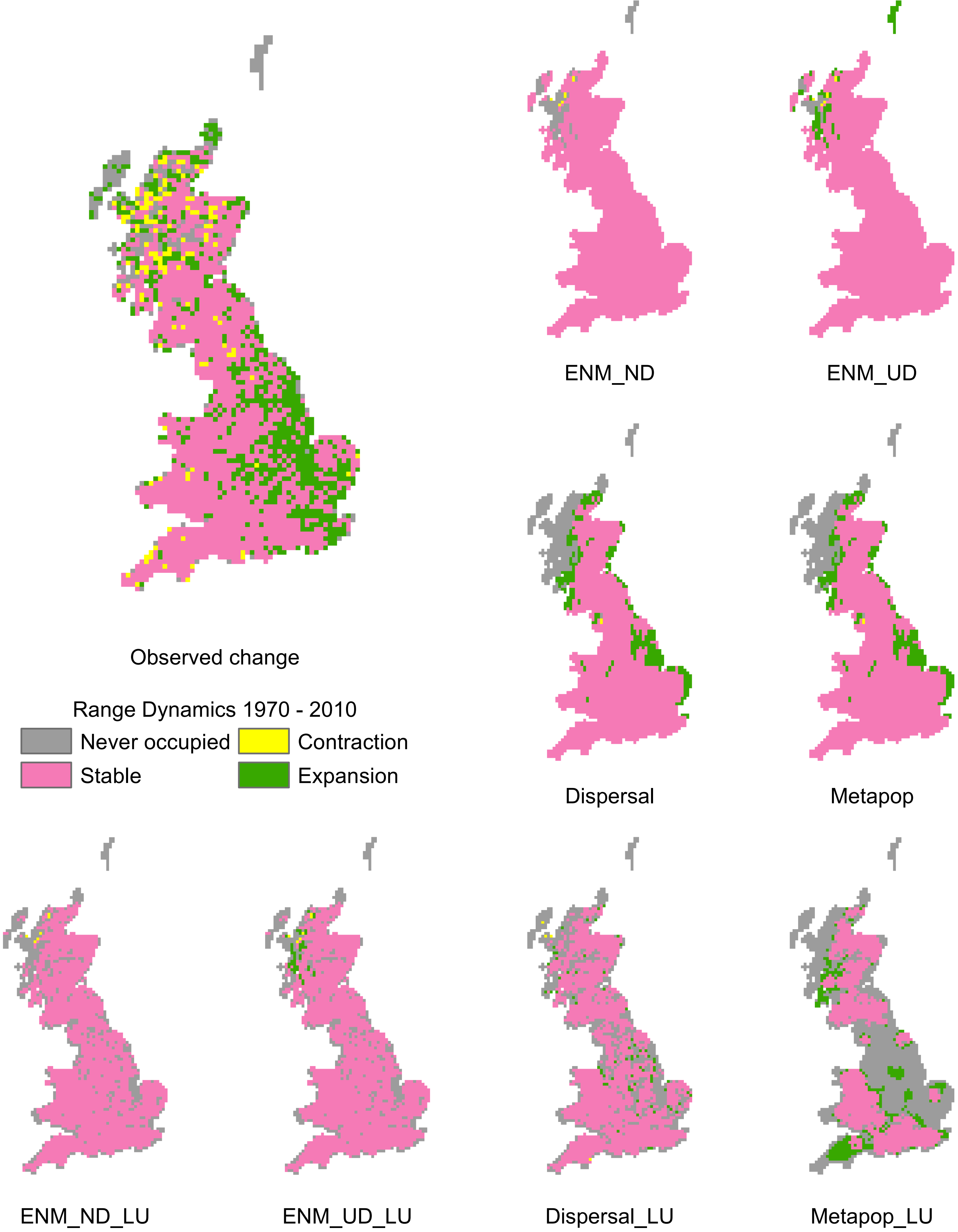
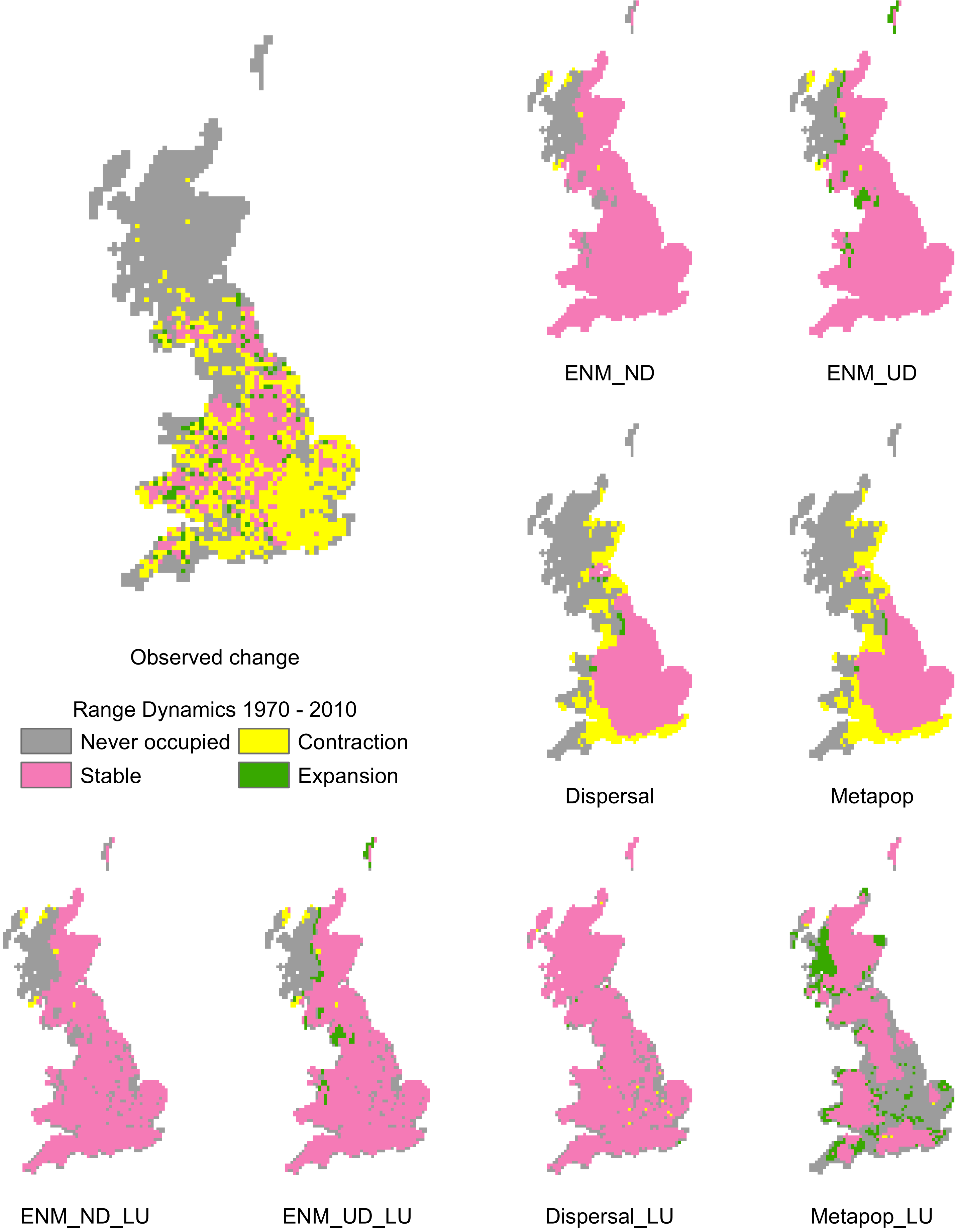


Fig. 1 Hierarchy of eight models used to predict species' ranges for British Breeding birds in 2010 using 1970 training data. (i) Bioensembles was used to generate ecological niche models (ENM) with or without a land-use, assuming either unlimited or no dispersal, (ii) RAMAS was used to mechanistically simulate annual dispersal over the 40 year period using species specific dispersal constraints (Dispersal), (iii) RAMAS was used to simulate metapopulations as well as dispersal dynamics (Metapop). See Methods for further details.

254x190mm (300 x 300 DPI)


Observed change
Range Dynamics 1970 - 2010
Never occupied
Contraction
Stable
Expansion
ENM_ND
ENM_UD
Dispersal
Metapop
ENM_ND_LU
ENM_UD_LU
Dispersal_LU
Metapop_LU


Observed change
Range Dynamics 1970 - 2010
Never occupied
Contraction
Stable
Expansion
ENM_ND
ENM_UD
Dispersal
Metapop
ENM_ND_LU
ENM_UD_LU
Dispersal_LU
Metapop_LU

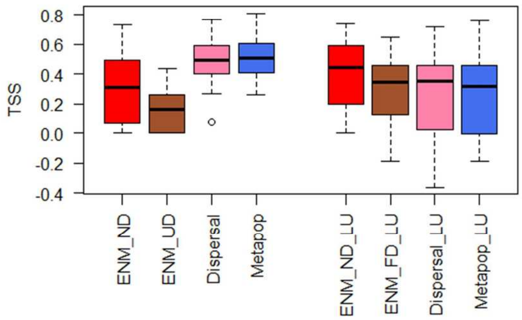
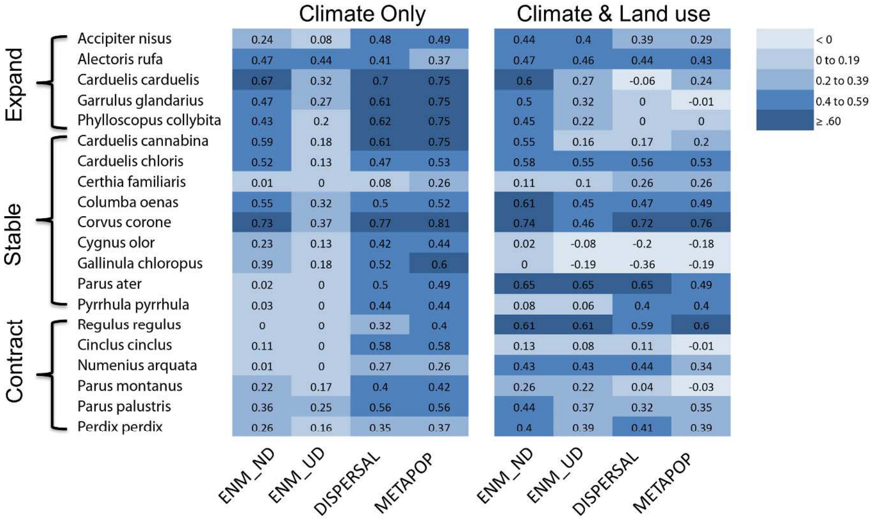


Fig. 4: Independent tests of model predictions for 20 species of British birds, showing the influence of model and land-use. Observed and predicted spatial patterns of species ranges are compared using True Skill Score (TSS) for three families of models: simple ecological niche models (ENMs), dispersal-linked niche models (Dispersal) and niche-population models (Metapop); each plotted with and without land-use (LU). Results for ENMs are shown assuming no dispersal (ND) and unlimited dispersal (UD).

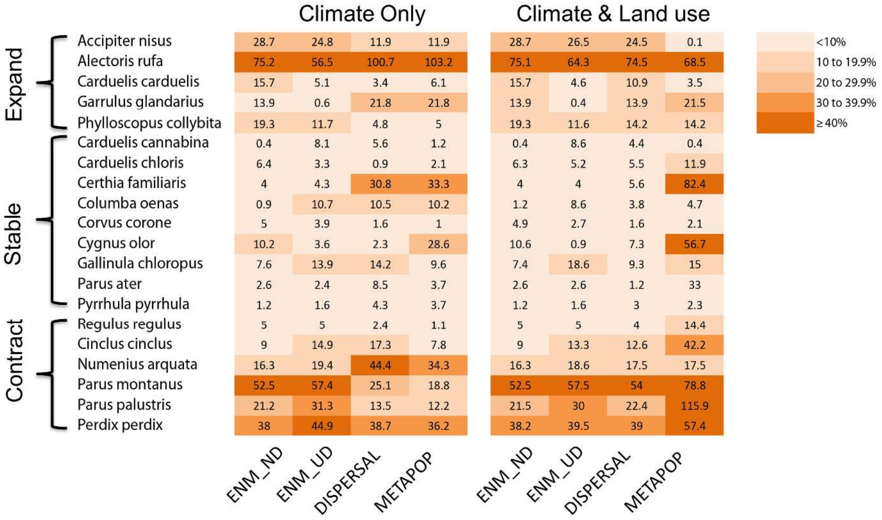
254x190mm (300 x 300 DPI)



		Climate Only				Climate & Land use			
		ENM_ND	ENM_UD	DISPERSAL	METAPOP	ENM_ND	ENM_UD	DISPERSAL	METAPOP
Expand	Accipiter nisus	0.24	0.08	0.48	0.49	0.44	0.4	0.39	0.29
	Alectoris rufa	0.47	0.44	0.41	0.37	0.47	0.46	0.44	0.43
	Carduelis carduelis	0.67	0.32	0.7	0.75	0.6	0.27	-0.06	0.24
	Garrulus glandarius	0.47	0.27	0.61	0.75	0.5	0.32	0	-0.01
	Phylloscopus collybita	0.43	0.2	0.62	0.75	0.45	0.22	0	0
Stable	Carduelis cannabina	0.59	0.18	0.61	0.75	0.55	0.16	0.17	0.2
	Carduelis chloris	0.52	0.13	0.47	0.53	0.58	0.55	0.56	0.53
	Certhia familiaris	0.01	0	0.08	0.26	0.11	0.1	0.26	0.26
	Columba oenas	0.55	0.32	0.5	0.52	0.61	0.45	0.47	0.49
	Corvus corone	0.73	0.37	0.77	0.81	0.74	0.46	0.72	0.76
	Cygnus olor	0.23	0.13	0.42	0.44	0.02	-0.08	-0.2	-0.18
	Gallinula chloropus	0.39	0.18	0.52	0.6	0	-0.19	-0.36	-0.19
	Parus ater	0.02	0	0.5	0.49	0.65	0.65	0.65	0.49
	Pyrrhula pyrrhula	0.03	0	0.44	0.44	0.08	0.06	0.4	0.4
Contract	Regulus regulus	0	0	0.32	0.4	0.61	0.61	0.59	0.6
	Cinclus cinclus	0.11	0	0.58	0.58	0.13	0.08	0.11	-0.01
	Numenius arquata	0.01	0	0.27	0.26	0.43	0.43	0.44	0.34
	Parus montanus	0.22	0.17	0.4	0.42	0.26	0.22	0.04	-0.03
	Parus palustris	0.36	0.25	0.56	0.56	0.44	0.37	0.32	0.35
	Perdix perdix	0.26	0.16	0.35	0.37	0.4	0.39	0.41	0.39

Fig. 5: True Skill Score (TSS) for models independently validated using observed data on occurrence in 2010. Expanders represent the upper quartile for observed range movement (> 9 % increase in range area between 1970 to 2010). Contractors represent the lower quartile (> 13% decrease in range area between 1970 to 2010). ENM_ND = climate with no dispersal; ENM_ND_LU = climate with no dispersal and land use; ENM_UD = climate with full dispersal; ENM_UD_LU = climate with full dispersal and land use; Dispersal = Climate with a dispersal function; Dispersal_LU = climate with a dispersal function and land use; Metapop = climate with a dispersal function and population model; Metapop_LU = climate with a dispersal function and population model and land use.

254x190mm (300 x 300 DPI)



		Climate Only				Climate & Land use			
		ENM_ND	ENM_UD	DISPERSAL	METAPOP	ENM_ND	ENM_UD	DISPERSAL	METAPOP
Expand	Accipiter nisus	28.7	24.8	11.9	11.9	28.7	26.5	24.5	0.1
	Alectoris rufa	75.2	56.5	100.7	103.2	75.1	64.3	74.5	68.5
	Carduelis carduelis	15.7	5.1	3.4	6.1	15.7	4.6	10.9	3.5
	Garrulus glandarius	13.9	0.6	21.8	21.8	13.9	0.4	13.9	21.5
	Phylloscopus collybita	19.3	11.7	4.8	5	19.3	11.6	14.2	14.2
Stable	Carduelis cannabina	0.4	8.1	5.6	1.2	0.4	8.6	4.4	0.4
	Carduelis chloris	6.4	3.3	0.9	2.1	6.3	5.2	5.5	11.9
	Certhia familiaris	4	4.3	30.8	33.3	4	4	5.6	82.4
	Columba oenas	0.9	10.7	10.5	10.2	1.2	8.6	3.8	4.7
	Corvus corone	5	3.9	1.6	1	4.9	2.7	1.6	2.1
	Cygnus olor	10.2	3.6	2.3	28.6	10.6	0.9	7.3	56.7
	Gallinula chloropus	7.6	13.9	14.2	9.6	7.4	18.6	9.3	15
	Parus ater	2.6	2.4	8.5	3.7	2.6	2.6	1.2	33
	Pyrrhula pyrrhula	1.2	1.6	4.3	3.7	1.2	1.6	3	2.3
Contract	Regulus regulus	5	5	2.4	1.1	5	5	4	14.4
	Cinclus cinclus	9	14.9	17.3	7.8	9	13.3	12.6	42.2
	Numenius arquata	16.3	19.4	44.4	34.3	16.3	18.6	17.5	17.5
	Parus montanus	52.5	57.4	25.1	18.8	52.5	57.5	54	78.8
	Parus palustris	21.2	31.3	13.5	12.2	21.5	30	22.4	115.9
	Perdix perdix	38	44.9	38.7	36.2	38.2	39.5	39	57.4

<10%
10 to 19.9%
20 to 29.9%
30 to 39.9%
≥40%

Fig. 6: Absolute differences in observed and predicted percentage change in range area between 1970 and 2010 (%). Expanders represent the upper quartile for observed range movement (> 9 % increase in range area between 1970 to 2010). Contractors represent the lower quartile (> 13% decrease in range area between 1970 to 2010). ENM_ND = climate with no dispersal; ENM_ND_LU = climate with no dispersal and land use; ENM_UD = climate with full dispersal; ENM_UD_LU = climate with full dispersal and land use; Dispersal = Climate with a dispersal function; Dispersal_LU = climate with a dispersal function and land use; Metapop = climate with a dispersal function and population model; Metapop_LU = climate with a dispersal function and population model and land use.

254x190mm (300 x 300 DPI)