Simulating the effects of migration on the uncertainty of species' range shifts

P. Holloway[1], J.A. Miller[1], S. Gillings[2]

[1]Department of Geography and the Environment, University of Texas at Austin, 305 E 23rd Street, CLA, Austin, TX 78712, USA
Telephone: (+1) 512 232 7336
Email: paul_holloway@utexas.edu
Email: jennifer.miller@austin.utexas.edu

[2]British Trust for Ornithology (BTO), The Nunnery, Thetford, Norfolk, IP24 2PU, UK
Telephone: (+44) 1842 750 050
Email: simon.gillings@bto.org

Abstract

Species distribution models (SDMs) are one of the most important GIScience research areas in biogeography and they are now the primary method used to study the effects of climate change on species distributions and ranges. We have identified 20 methods of incorporating migration alongside SDM to predict range shifts, and assess them for their accuracy in predicting future distributions, as well as the uncertainty between predictions. Ensemble models (generated by combining multiple projection outcomes) were created for each species, and a new index was developed to quantify uncertainty among the projections. Values ranged from 0.699 (large areas of disagreement and high uncertainty) to 0.999 (large areas of agreement and low uncertainty). This index will allow researchers to quickly identify high or low uncertainty amongst projection results from migration methods and discuss future range shifts.

Keywords: Species Distribution Modelling, Uncertainty, Migration, Climate Change.

1. Introduction

Species distribution models (SDMs) are one of the most important GIScience research areas in biogeography and they are now the primary method used to study the effects of climate change on species distributions and ranges. However, the extrapolation of species-environment relationships involves several conceptual and methodological issues that introduce uncertainty, such as novel environmental combinations, new species interactions, unpredictable feedbacks related to land cover change, as well as the migration capacity of the species.

Soberón and Peterson (2005) introduced the BAM framework to identify the three main factors that determine species distributions: Biotic, Abiotic and Movement. Biotic factors represent interactions with other species (i.e. competition, predation), abiotic factors represent the physiological tolerances of a species (i.e. temperature, precipitation) and movement refers to the area that has been or will be accessible to a species within a certain timeframe. While the importance of all three factors (abiotic, biotic and movement) is well recognized in the SDM discipline, the majority of SDM studies have not found a way to regularly incorporate movement in an SDM framework, and those that do often rely on an overly simplistic conceptualizations of movement (e.g. full dispersal or no dispersal) (Franklin 2010).

When migration has been implemented in SDM research, it has often taken one of two extreme approaches: unlimited or limited. These two approaches overly simplify movement, and as a result of their extreme variation, greatly increase the uncertainty associated with the effects of climate change, for example, Engler and Guisan (2009) found a 95% difference in projected distribution between their model of migration and these extreme approaches. In SDM research, many efforts have been directed towards identifying the uncertainty associated with the various model-building phases, such as the implementation of a statistical framework (Elith & Graham 2009) or the selection of absence or 'background' data (Barbet-Massin et al. 2012). With the variation in results due to migration methods currently used in SDM so substantial, research needs to be directed towards quantifying this (Miller & Holloway *in press*).

One of the challenges associated with comparing methods for projecting the effects of climate change on species' ranges has been the lack of appropriate data with which to validate the results. Datasets of species ranges that span temporal scales meaningful for assessing climate change impacts are becoming increasingly available. The British Trust for Ornithology (BTO) has led the production of bird atlases offering complete coverage of Great Britain at a 10km resolution in three periods spanning 40 years, in 1968–72, 1988–91 and 2007–11. These datasets can be used to test the accuracy of the migration methods implemented in SDM research, as we can take the earliest known range and project the species-environment relationships alongside migration to the latest time period, and compare the results with the current distribution. This will provide researchers with a framework to compare the differences between each individual migration method currently implemented, for a variety of species and spatiotemporal scales and will subsequently allow discussions about future range, extinction rates and migration patterns to be made with a better understanding of the inherent uncertainty.

2. Data and Methods

Distribution data for 50 breeding bird species were extracted from bird atlases for 1988–91 (Gibbons et al. 1993) and 2007–11 (Balmer et al. 2013). The species were selected to span a range of attributes such as their ubiquity, migration strategy and whether they were native or introduced. Corresponding climate data for the 12 UKCP09 annual variables was averaged to match the temporal range of the atlas data. Land cover data for the same periods (LCMGB 1990 and LCM2007) was obtained from the Centre of Ecology and Hydrology and elevation data from the USGS for bird analysis.

Three commonly used statistical methods known for generating accurate results were used to project species' ranges, Generalized Linear Models (GLM), MaxEnt, and Random Forests (RF). Each of these methods deals with extrapolation (predicting in future, and sometimes novel or non-analogue, climate conditions) differently. A total of 20 migration methods (see table 1) were coupled with the three SDM projections for both birds and plants (table 1). Unlimited migration is the 'best case' scenario, and assumes that all new suitable habitat is accessible. No migration is the 'worst case' scenario, assuming that only currently suitable habitat is available in future projections. All migration methods were implemented using original R functions written by the author in R 3.1.2 (R Development Core Team 2008), with the exception of the probabilistic dispersal kernels which use the kernel function in the MigClim package (Engler et al. 2013).

| Method	Description	Method #
No migration	Only suitable habitat that matches current range can become occupied	1
Limited migration (rook's)	Only overlapping suitable habitat can become occupied, using rook's connectivity	2
Limited migration (Queen's)	Only overlapping suitable habitat can become occupied, using Queens's connectivity	3
Fixed rate 1 step (arithmetic mean)	Fixed migration based on arithmetic mean recorded dispersal distance between earliest known distribution and latest known distribution	4
Fixed rate 1 step (geometric mean)	Fixed migration based on geometric mean recorded dispersal distance between earliest known distribution and latest known distribution	5
Fixed rate 2 steps (arithmetic mean)	Fixed migration based on arithmetic mean recorded dispersal distance between earliest known distribution, to an intermediately predicted distribution, and latest known distribution	6
Fixed rate 2 steps (arithmetic mean)	Fixed migration based on geometric mean recorded dispersal distance between earliest known distribution, to an intermediately predicted distribution, and latest known distribution	7
Discrete dispersal kernel (Gaussian arithmetic)	Probability of dispersal drawn from a Gaussian distribution and the maximum taken as the fixed rate. Based on arithmetic statistics	8
Discrete dispersal kernel (Gaussian geometric)	Probability of dispersal drawn from a Gaussian distribution and the maximum taken as the fixed rate. Based on geometric statistics	9
Discrete dispersal kernel (inverse exponential arithmetic)	Probability of dispersal drawn from an inverse exponential distribution and the maximum taken as the fixed rate. Based on arithmetic statistics	10
Discrete dispersal kernel (inverse exponential geometric)	Probability of dispersal drawn from an inverse exponential distribution and the maximum taken as the fixed rate. Based on geometric mean.	11
Discrete dispersal kernel (Cauchy arithmetic)	Probability of dispersal drawn from a Cauchy distribution and the maximum taken as the fixed rate. Based on arithmetic statistics	12
Discrete dispersal kernel (Cauchy geometric)	Probability of dispersal drawn from a Cauchy distribution and the maximum taken as the fixed rate. Based on geometric mean.	13
Probabilistic dispersal kernel (Gaussian arithmetic)	Assign each grid a conditional probability from a Gaussian distribution to be colonized and generate presences based on these values. Arithmetic statistics	14
Probabilistic dispersal kernel (Gaussian geometric)	Assign each grid a conditional probability from a Gaussian distribution to be colonized and generate presences based on these values. Geometric statistics	15
Probabilistic dispersal kernel (inverse exponential arithmetic)	Assign each grid a conditional probability from an inverse exponential distribution to be colonized and generate presences based on these values. Arithmetic statistics	16
Probabilistic dispersal kernel (inverse exponential geometric)	Assign each grid a conditional probability from an inverse exponential distribution to be colonized and generate presences based on these values. Geometric statistics	17
Probabilistic dispersal kernel (Cauchy arithmetic)	Assign each grid a conditional probability from a Cauchy distribution to be colonized and generate presences based on these values. Arithmetic statistics	18
Probabilistic dispersal kernel (Cauchy geometric)	Assign each grid a conditional probability from a Cauchy distribution to be colonized and generate presences based on these values. Geometric statistics	19
Unlimited	All suitable habitat can become occupied	20

Table 1. Summary of the migration methods used.
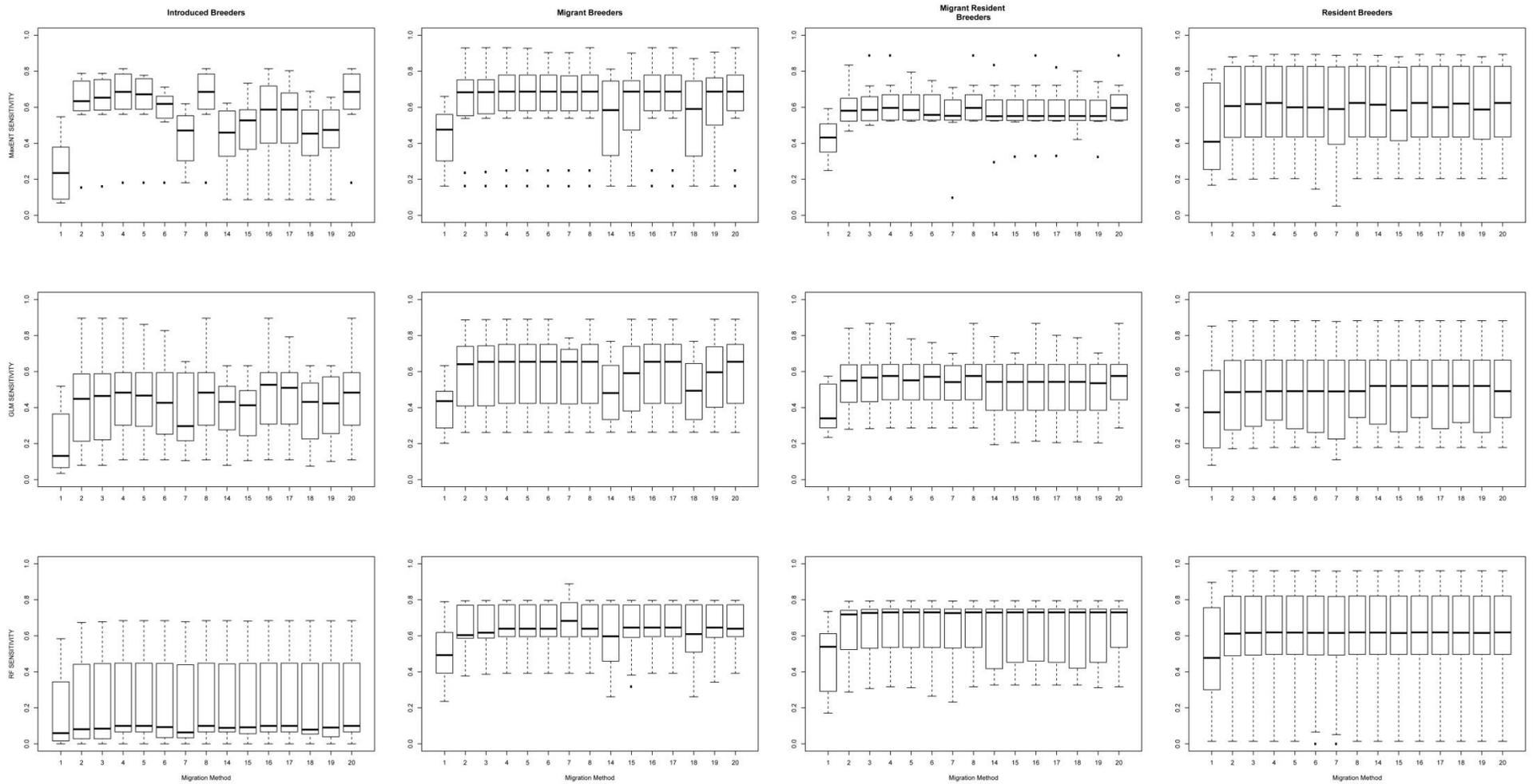
Figure 1. Sensitivity scores for the 20 migration methods (n.b. 8-13 are consolidated under 8 due to no variation), for the three SDM methods and four breeding groups.

Figure 2. Distribution maps of the Cetti's warbler with MaxEnt, a) 1990 observed distribution, b) 2010 observed distribution, c) unlimited migration, d) no migration, e) ensemble model, f) fixed rate 1 step (arithmetic), g) fixed rate 2 steps (arithmetic), h) Gaussian kernel (MigClim arithmetic), i) exponential (Migclim arithmetic) and j) Cauchy (MigClim arithmetic)
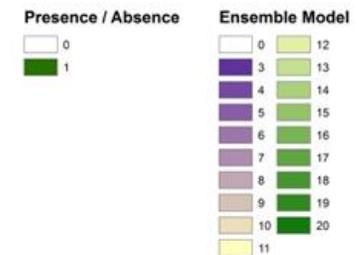
Species	Ensemble Richness	Ensemble Index	Species	Ensemble Richness	Ensemble Index
Blackcap	3	0.999	Kingfisher	6	0.985
Barn Owl	5	0.981	Lesser Black-backed Gull	4	0.985
Barnacle Goose	13	0.734	Long-eared Owl	6	0.956
Common Buzzard	7	0.971	Little Grebe	4	0.979
Cormorant	9	0.910	Little Ringed Plover	10	0.921
Canada Goose	4	0.990	Mandarin Duck	11	0.929
Common Tern	6	0.964	Nuthatch	5	0.990
Cetti's Warbler	18	0.770	Nightjar	8	0.912
Wood Duck	17	0.699	Great Skua	8	0.892
Dartford Warbler	15	0.705	Oystercatcher	5	0.992
Common Eider	5	0.985	Osprey	9	0.815
Egyptian Goose	15	0.773	Peregrine Falcon	4	0.982
Firecrest	15	0.787	Quail	4	0.970
Green Woodpecker	4	0.991	Ring-necked Parakeet	14	0.768
Gadwall	11	0.802	Red-legged Partridge	7	0.863
Goosander	4	0.982	Raven	5	0.988
Grasshopper Warbler	5	0.970	Reed Warbler	4	0.983
Goshawk	10	0.774	Stonechat	11	0.977
Greylag Goose	5	0.966	Siskin	8	0.984
Grey Wagtail	5	0.994	Sand Martin	4	0.985
Great Spotted Woodpecker	3	0.997	Shelduck	8	0.945
Garganey	8	0.864	Tree Sparrow	5	0.988
Grey Heron	3	0.996	Water Rail	9	0.947
Herring Gull	6	0.984	Woodlark	15	0.768
Hobby	6	0.944	Wigeon	9	0.938

Table 2: Ensemble statistics for each bird species across the whole study area

3. Results

The most notable difference in accuracy measures is for sensitivity, the proportion of correctly predicted presences (Figure 1), and occurs for MaxEnt models of the introduced breeders. The two most extreme migration methods (#1 and #20) bound the other methods, with the other methods changing quite substantially in their sensitivity scores. The two limited migration methods (#2 and #3) result in higher sensitivity than many of the other methods. The six iterations of the discrete dispersal kernel (#8 to #13) are consolidated under #8, as there were no significant differences in accuracy scores for these methods when all species were considered together. These discrete dispersal kernels produced results identical to unlimited migration due to large values of the maximum value drawn from fat-tailed kernels. Our results suggest that this method is possibly not the most realistic method of simulating migration of birds as it assumes that every location between the maximum distance drawn is equally accessible.

Figure 2 shows predictions from a number of the migration methods for the Cetti's warbler coupled with MaxEnt, as well as an ensemble model (the number of times an observation is predicted present by all 20 migration methods). Unlimited migration (Figure 2c) appears to over-predict the distribution of the Cetti's warbler into northern England and Scotland, as well as reduce the fragmentation in the center of the distribution. Conversely, no migration (Figure 4d) seriously under-predicts the distribution. This figure also highlights the differences between one- and two-step fixed rates (Figure 2f-g). The difference in area projected differs by 1940km^2 between these two methods. This is a large difference in projected future ranges for a relatively scarce species in Great Britain and could make up a considerable amount of its perceived range. This could result in contradicting interpretations of whether this species is considered as part of a stable population or a threatened population.

As it is not possible to show an ensemble map for all the species, in order to effectively interpret the uncertainty associated with the ensemble maps, we developed an ensemble index (*D)* which ranks full agreement across every observation as 1, and no agreement across every observation as 0:

$$D = \frac{\left(\frac{E-\min(E)}{\max(E)-\min(E)}\right)}{N} \tag{1}$$

where N is the total number of presence observations in the output, and E is defined as:

$$E = \left(\sum \frac{e}{m}\right) \tag{2}$$

where e is the ensemble score, and m is the total number of migration models used to generate the ensemble.

Table 2 identifies the ensemble index for all 50 bird species, as well as the total number of different ensemble scores (ensemble richness). The lowest score is 0.699 for the wood duck, an introduced breeder, and the second lowest is 0.705 for the Dartford warbler, a resident breeder. The lower the ensemble index, the higher the uncertainty of the projections. Several species have ensemble index scores over 0.99, suggesting very

high number of observations with full agreement. The higher total number of ensemble scores across an observation is obviously an indicator for a high ensemble index, but this is not always the case. Both Little ringed plover and Mandarin duck have an ensemble score of >0.9, but a total number of ensemble scores greater than 10. There are no really low index scores for any of the 50 birds. In part this could be due to a relatively low uncertainty surrounding these methods, although more likely this is due to the inclusion of the six discrete dispersal kernel methods which when combined suggest full agreement that may be a result of their specification. If these were removed, we may observe lower scores.

4. Acknowledgements

5. References and Citations

Balmer DE, Gillings S, Caffrey BJ, Swann RL, Downie IS, and Fuller FJ, 2013, *Bird Atlas 2007-11: the Breeding and Wintering Birds of Britain and Ireland.* BTO, Thetford, UK.

Barbet-Massin M, Jiguet F, Albert CH, and Thuiller W, 2012, Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution,* 3(2), 327-338.

Elith J, and Graham CH, 2009, Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography,* 32(1), 66-77.

Engler R, Hordijk W, and Pellissier L, 2013, MigClim: Implementing dispersal into species distribution models. R package version 1.6. http://CRAN.R-project.org/package=MigClim

Franklin J, 2010, Moving beyond static species distribution models in support of conservation biogeography. *Diversity and Distributions,* 16(3), 321-330.

Gibbons DW, Reid JB, and Chapman RA, 1993, *The new atlas of breeding birds in Britain and Ireland.* Poyser, London, UK.

Miller JA, and Holloway P, *in press,* Incorporating movement in species distribution models. *Progress in Physical Geography.*

R Devlopment Core Team 2008, R: A language and environment for statistical computer. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Soberón J and Peterson AT, 2005, Interpretation of models of fundamental ecological niches and species' distributional areas. *Biodiversity Informatics,* 2, 1-10.