Articles

A temporally explicit species distribution model for a long distance avian migrant, the common cuckoo

Heather M. Williams, Mikkel Willemoes and Kasper Thorup

H. M. Williams (http://orcid.org/0000-0003-0755-3866) (hw49@buffalo.edu), Mikkel Willemoes and Kasper Thorup, Center for Macroecology, Evolution and Climate, Natural History Museum of Denmark, Univ. of Copenhagen, Copenhagen, Denmark. HMW also at: Graduate Program in Evolution, Ecology, and Behavior, Univ. at Buffalo, Cooke Hall, Buffalo, NY.

Modelling the distribution of migratory species has rarely been extended beyond breeding and wintering ranges despite many species showing much more complex movement patterns with multiple stopovers. We aimed to create a temporally explicit species distribution model describing the full annual distribution cycle, and use it to model the complex seasonal shifts in distribution of the common cuckoo *Cuculus canorus*, a declining long-distance migrant. To do this we used full-year satellite telemetry occurrence data, with their associated temporal information, to inform a temporally explicit species distribution model using MaxEnt. The resulting full-year distribution model was highly predictive (AUC = 0.894) and appeared to have generality at the species-level despite being informed by data from a single breeding population. Comparison of our methodology with seasonal distribution models describing the breeding, winter and migration ranges separately showed that our full-year method provided more general and extensive predictions and performed better when tested with an independent dataset. When species distribution models based on a single season exclude environmental conditions experienced by birds in other parts of the annual cycle they risk underestimating niche breadth and neglecting the importance of stopover habitat. Conversely, models which simply average conditions across a season may miss the significance of finer scale within-season movements and overestimate niche breadth. In contrast, our framework for a full-year migrant distribution model successfully captures the finer-scale changes expected in seasonal environments and can be used to inform conservation management at every stage of migration. The full-year model framework appears to produce temporal distribution models generalised to the species-level from occurrence data limited to few individuals of a single population and may have particular utility when aiming to describe the distribution of species with complex migration patterns from telemetry data.

Widespread population declines in long distance avian migrants have been observed over the past 25 years (Sanderson et al. 2006, Heldbjerg and Fox 2010, Runge et al. 2014, Vickery et al. 2014). Migrant species are considered especially vulnerable to environmental change due to their dependence on multiple locations for survival (Newton 2004, Klaassen et al. 2012, Runge et al. 2014, Runge et al. 2015). However, addressing these declines is complicated by incomplete knowledge of migrant wintering zones and migratory routes (Bridge et al. 2011). Although relatively precise migration tracks are now published for many migrant species (Klaassen et al. 2010, Willemoes et al. 2014) the emerging challenge is how to extract species-level distributions from these data representing relatively few individuals from relatively few breeding populations. Species distribution models (SDMs) offer a framework for such general inference from limited observational data. SDMs have many practical applications for conservation such as highlighting parts of the annual range in which species could be habitat limited, or predicting future changes in range with contemporary climate change (Barbet-Massin et al. 2009).

For migrant species, there are several complications when trying to fit SDMs because of migrants' dependence on seasonally changing landscapes and associated movements. Existing methodology for migrant SDMs typically do not include fine-scale temporal information, and instead create separate models to describe different parts of the annual cycle, based on seasonally averaged conditions (Peterson et al. 2005, Batalden et al. 2007, Gschweng et al. 2012, Cardador et al. 2014, Hayes et al. 2015). This seasonal approach may be appropriate where research is addressing a regional or landscape-scale issue in a subset of the range (Gschweng et al. 2012), or when there is strong evidence that a species is a niche switcher, changing its ecological requirements substantially between seasons (Martinez-Meyer et al. 2004). In many cases, however, these seasonal models may only reflect a small portion of the niche as they draw their occurrence data from just a small portion of the annual cycle (Colwell and Rangel 2009, Phillips et al. 2009). As such, their output may be over-fitted to these seasonal data and lack the generality needed to describe distributions at the species level, or to project distributions to new regions or

climates. Creating multiple models for a single species also presents a challenge when attempting to describe the complete distribution of complex migrants with multiple stopover areas, itinerant species or nomads (Jonzén et al. 2011). Stopover conditions have been shown to affect population sizes (Runge et al. 2014), but are essentially neglected in seasonal SDMs for migrants. To effectively model a migrant species' distribution, therefore, we should not only strive to reduce spatial bias in sample data, but also to capture the full range of annual conditions experienced by the species to more accurately describe the entire niche (Heikkinen et al. 2006, Kearney et al. 2012, Laube et al. 2015).

Fink et al. (2010) developed a spatiotemporal exploratory model (STEM) to deal with the spatiotemporal biases inherent in citizen science data (Hochachka et al. 2012). They created a range-wide ensemble model by scaling up from a set of models of lesser extents (in both space and time) and showed that pooling data over a larger region in the STEM provided a more realistic model for the changing spatiotemporal distribution of a migrant *Tachycineta bicolor* than a static SDM or smaller range STEM models (Fink et al. 2010). The temporal variation in distributions is not modelled explicitly in the STEM framework where habitat choice is only modelled based on static habitat assignments and does not take into account the seasonal states of the habitat when used by the birds.

Satellite telemetry provides temporally-explicit occurrence information for migratory species with quantifiable accuracies. Telemetry datasets avoid many of the spatial and temporal biases inherent in traditional occurrence data. For example, telemetry datasets allow for regularised temporal and spatial sampling schemes thus avoiding the often seen discrepancy between high and low latitude datasets (Boakes et al. 2010). Satellite telemetry data are also unaffected by anthropogenic landscape features such as road proximity or by human population size (Reddy and Davalos 2003). Furthermore, when considering migrant species, the high potential for an individual to move significantly between telemetry records implies that each telemetry occurrence is spatially independent of the location of the other records. However, as satellite telemetry datasets are often based on information from relatively few individuals, they can be difficult to generalise. Despite this drawback, satellite data have been successfully used in SDMs highlighting certain seasons or regions of migrant species' ranges (Martinez-Meyer et al. 2004, Edrén et al. 2010, Gschweng et al. 2012, Limaiñana et al. 2015) and it constitutes a promising resource for building SDMs (Hebblewhite and Haydon 2010).

One of the many migrant populations to be studied using satellite telemetry in recent years is the South Scandinavian population of common cuckoos *Cuculus canorus* (hereafter 'cuckoo'). The cuckoo is one of the many sub-Saharan avian migrant species in decline (Vickery et al. 2014), with the population in Denmark declining by more than 20% between 1975 and 2015 (Fenger et al. 2016) and even more dramatic declines of almost 50% being reported in the United Kingdom over a similar period (Gregory et al. 2004). The South Scandinavian population exhibits a complex loop migration between its European breeding grounds and its main wintering grounds in south-west-central Africa, with six distinct non-breeding sedentary periods (stopovers). As the tracked cuckoos spent an average of 141 days of the year neither actively migrating, nor at the breeding or winter grounds (Willemoes et al. 2014), the non-breeding stopover sites are likely important habitat with an impact on survival. We need to understand the full seasonal distribution of these declining migrants (including their stopovers) to enable their effective conservation and management.

Here, we use satellite telemetry data from cuckoos in a temporally explicit SDM to capture the full annual spatiotemporal cycle of a migrant species in a single niche model. Unlike previous temporally-specific models (Peterson et al. 2005, Batalden et al. 2007, Fink et al. 2010), this is accomplished with a single underlying model describing the full year, rather than separate models for separate seasons, by averaging the output of multiple models or by building an ensemble model. By allowing conditions experienced by the cuckoos throughout the whole year to influence overall model fit we make the assumption that cuckoos are tracking their niche throughout migration, and take a broad view of the niche as the full annual set of conditions capable of supporting the species, rather than splitting the niche into separate sets of seasonal requirements. However, as the cuckoo is a brood parasite the occurrence of host species likely affects the breeding distribution of the species and could result in apparent niche switching between the breeding season and the rest of the year. To account for this possibility, we additionally build a model for the non-breeding season only and compare its output with the full-year model. Finally, for comparison with the full-year model, we present a set of seasonal models (each built using only occurrence data from within the relevant season). We compare the predictive performance of these models and the correspondence of their predicted distributions with published distribution maps and an independent data source to test whether models built with occurence data from a very limited set of individuals can effectively scale up to the population or species level.

Material and methods

Satellite tracking

Seven (four male, three female) adult cuckoos from south Scandinavia were caught in mistnets, fitted with satellite transmitters (Solar PTT-100s, Microwave Telemetry Inc.) and later initiated migration. Transmitters operated on a 10 h on and 48 h off duty cycle, with the potential for multiple points to be logged within each duty cycle. Data were obtained between May 2010 and June 2013. Full details of this procedure are described by Willemoes et al. (2014).

Presence data preparation

Occurrence points were filtered to include only points accurate to within 1.5 km (Argos location quality classes 1, 2 and 3) (Argos 2011). SDMs work under the assumption of low spatial and temporal autocorrelation in the occurrence data. This is a reasonable assumption for our data, due to the high potential for cuckoo movement between points and the relatively coarse temporal grain (monthly) which we aim

to model. To further reduce spatial autocorrelation in the dataset we removed points generated by an individual that were located within 1 km of each other or were recorded within a two-hour period. We filtered the remaining set to only include one point/individual/day between sunset and sunrise to avoid nightly roosting locations being over-represented in the dataset. Finally, to prevent any one month or season from influencing the modelling process disproportionately, the dataset was randomly subsampled to create a training dataset with an equal sample size for each month (n = 50). Where high quality data were sparse in April (n = 45) and August (n = 17), data selection criteria were relaxed to allow inclusion of multiple points between sunset and sunrise. High quality points discarded during this process were saved for model testing. This left a training dataset of 600 points and a testing dataset of 542 points (Supplementary material Appendix 1 Table A1).

Whilst we consider most of our data to be representative of the cuckoo's niche and have spatial independence, the same is not necessarily true for points collected during the breeding season. Firstly, as a brood parasite, the cuckoo's breeding season distribution may be equally affected by the distribution of their host species as by their own ecological niche. This makes their resource needs more complex in the breeding season compared to the rest of the year (Williams et al. 2016) and means the inclusion of these data may jeopardise the ability of the model to generalise to the full year. Secondly, unlike the rest of the dataset, the breeding season data do not have spatial independence, as the original trapping locations of the satellite tracked birds were selected by the researchers and multiple birds were sampled within relatively limited areas. An alternate dataset was therefore maintained which excluded the breeding season data (hereafter called the 'full-year without breeding' model) to allow for comparison of both model projections.

MaxEnt

We used the MaxEnt software package (ver. 3.3.3k) (Phillips et al. 2006, Phillips and Dudik 2007) for modelling, as one of the leading algorithms for presence-only data in contemporary climate based SDMs (Elith et al. 2006).

As MaxEnt predictions are affected by the extent of the background (VanDerWal et al. 2009), the background extent should encompass the full geographical range of known species occurrence, but not exceed it (Elith et al. 2011). Given the wide distribution of the cuckoo (BirdLife International and NatureServe 2014) this study defined a two continent background (Europe and Africa) delimited by –40°–75° latitude and –20°–60° longitude. The extent included areas from the western coast of Africa, as far east as the Ural Mountains, and from northern Norway to the southern tip of Africa. Offshore islands (including Greenland, Iceland, Yuzhny Island, Madagascar, Cape Verde and the Canary Is.) were removed from the background as they are unlikely to be part of the cuckoo's distribution.

Using the R package 'dismo' (Hijmans et al. 2016), we randomly sampled 100 000 points within the terrestrial background area. Given our large background area and the temporal aspect of our model, we used ten times the default number of background points (10 000). Background points were split evenly into 12 monthly categories. We verified that the background points for each month span the range (at least 90%) of global values likely to be encountered by the cuckoo for each environmental variable (Supplementary material Appendix 1 Table A2).

Environmental variables

Simple models, with expertly selected predictors are preferable for studies, such as this, aimed at niche description (Austin 2002, Merow et al. 2014). To this end, we informed the model with a limited set of four climatic variables, each with a strong ecological rationale. The cuckoo is an insectivore, specialising in a caterpillar-based diet throughout the whole year (Payne 2005). This leads to the assumption that the cuckoo migrates to track invertebrate (principally caterpillar) abundance through space. As forest caterpillar survival is optimised by hatching at the time of budburst (van Asch and Visser 2007) we expect that caterpillar abundance will also peak at this time. As budburst advances with maximum monthly temperatures (Buse and Good 1996, Karlsson et al. 2003), and temperature has been shown to define the distributions of many species (Root 1988) this was included as the first model variable. The average precipitation of the month prior to cuckoo arrival was included as the second predictor, because of the limiting effect of precipitation on budburst in some tree species and regions (Pinto et al. 2011). As birds are sometimes thought to maximise primary productivity on their migration routes (Bridge et al. 2016) and a previous study found that cuckoos track high NDVI values through space (Thorup et al. 2017), we included mean monthly NDVI as our third predictor. Finally, the timing of budburst can be described by a sharp increase in local NDVI through time (Soudani et al. 2012), so we included derived data showing the difference in each month's mean NDVI from that of the previous month, such that positive values indicate increased NDVI since the previous month and negative values indicate decreasing NDVI.

Precipitation and temperature data for each month were obtained from the WorldClim database ver. 1.4 (Hijmans et al. 2005) at a 30-seconds resolution to make an approximate match with the spatial grain of our sample data. We extracted NDVI data at a 1 km resolution for each month (NASA 2015). We used the R package 'raster' (Hijmans 2015), to create the NDVI difference layers and to extract the temporally relevant data to each presence point and each background point.

MaxEnt modelling

As MaxEnt model output can be highly sensitive to model settings (Merow et al. 2013) we underwent a model selection process using ENMTools ver. 1.4.3 (Warren et al. 2008). The monthly outputs of models using each feature type (and high ranking combinations thereof) and different regularisation values (0.5, 1, 2, 3, 4) were ranked according to their fit to the data using AICc (Supplementary material Appendix 1 Table A3). Feature types allow different transformations of the data and regularisation values determine how closely the model is fit to the data (with low values being highly fit and higher values being more smoothed). The best regularisation

values for each month were averaged for the final output and the most appropriate features were selected by a consensus approach (Supplementary material Appendix 1 Table A3). This process led to our selection of quadratic and linear features with a regularisation value of 4.

MaxEnt produces a 'raw' output, which reflects the relative probability that an individual originated in each cell; and a 'logistic' output which is a rank conserved transformation of the raw output that defines the probability of species presence in each cell and allows for comparison between models of different extents and background sampling regimes. There has been recent debate over whether it is preferable to interpret only the raw MaxEnt output and limit inference to relative probability in order to avoid the need to estimate (unknown) prevalence (i.e. the proportion of occupied sites in the landscape) (Royle et al. 2012, Hastie and Fithian 2013, Merow et al. 2013, Phillips and Elith 2013). As our analysis focuses mainly on thresholded output (in which the logistic output is transformed to a binary presence/absence grid), which is not sensitive to prevalence (Supplementary material Appendix 1 Fig. A1), and to allow easier comparison with other models, we present the logistic output modelled with the default prevalence setting of 0.5.

The objective threshold rule 'maximum training sensitivity and specificity' was applied to convert the probabilistic output to a binary prediction showing presence and absence. This approach is the statistical equivalent of maximising the true skill statistic (TSS) for presence-only data (Allouche et al. 2006) and has been shown to have low variance irrespective of whether presence-only or presence-absence data were used (Liu et al. 2013). Furthermore, unlike other popular measures of threshold-dependent fit such as kappa, TSS is insensitive to prevalence and it gives equal weighting to both sensitivity and specificity (Allouche et al. 2006, Liu et al. 2013).

Environmental layers were supplied to MaxEnt in the 'samples with data' format without location information. This approach keeps all the temporal information that is needed to effectively describe the full distribution of a migrant species at a monthly scale and allowed for the same geographic location to be associated with multiple environmental conditions at different times of year. The MaxEnt model was then projected to geographic space in raster layers depicting climatic conditions for each calendar month. We then underwent the same process for the 'full-year without breeding data' model.

Crossvalidation

Crossvalidation checks were performed on the predictions to test for individual, annual or monthly idiosyncrasies in the model (Gschweng et al. 2012). To achieve this, the training data were partitioned to exclude each individual, sex, year and month in turn, and model performance was then tested with the excluded group. No substantial drop in model performance was detected after excluding any month, individual, sex or year. This indicates that, in common with other recent studies using satellite tracking data in SDMs (Gschweng et al. 2012) there is little signal of bias due to the low number of individuals in our dataset (Supplementary material Appendix 1 Table A4).

Seasonal models

In order to provide a comparison between our modelling framework and previous techniques for migrant SDMs, we constructed a set of four seasonal models representing the breeding season, autumn migration, winter and spring migration. We trained each model with data from only the corresponding season. Occurrence points from the full-year dataset (Supplementary material Appendix 1 Table A1) were spatially divided into the four seasons, largely following Willemoes et al. (2014) (Supplementary material Appendix 1 Fig. A2). The same predictors and model settings were used as for the full-year model. Seasonal precipitation, NDVI and NDVI difference layers were made by averaging raster layers for each month included in that season (May–July for the breeding season, July–November for autumn migration, October–April for the winter, March–May for spring migration). Maximum monthly temperature was calculated as the maximum value for the season. Each season was then represented in a single model projection. To test whether the background extent impacted model predictions we repeated the process with reduced background latitudinal extents relevant to each season (north of 35°N for the breeding season; between 0°N and 55°N for the migration seasons; south of 15°N for the winter) (Supplementary material Appendix 1 Fig A3). The resulting projections were qualitatively similar, but covered a somewhat reduced total extent. To maintain a conservative comparison between the full-year model and the seasonal models we only continued analysis with seasonal models built using the full-year background extent.

Model analysis

Model performance was assessed using AUC (area under the receiver operator curve). Although we are aware of potential shortcomings of the AUC statistic as a measure of model fit (Lobo et al. 2008), we include it due to its high prevalence in the literature and the lack of a threshold-independent alternative for presence-only data (Merow et al. 2013). As a measure of threshold-dependent model fit we used TSS (Allouche et al. 2006). The area of predicted occurrence (area with a logistic score above the threshold value) was calculated for each month of the full-year model and for the seasonal models using the 'raster' package (Hijmans 2015) and plotted using 'ggplot2' (Wickham 2009).

The full-year model performance was compared to the seasonal models in terms of its fit to the data with AUC and TSS statistics.

Each monthly projection from the full-year model was combined into a single binary map showing the full predicted annual cuckoo distribution, using the logistic threshold value which maximised TSS (0.12). The same process was followed to combine the seasonal models into a single binary distribution where thresholds were chosen for each season to maximise TSS (0.11 for breeding, 0.18 for autumn migration, 0.27 for winter, 0.11 for spring migration). Maps showing the estimated annual cuckoo distribution (BirdLife International and NatureServe 2014) were overlaid with each map and the correspondence visually assessed.

Cuckoo occurrence data were also obtained from the GBIF database population (global biodiversity information

facility 2015) to provide an independent dataset with which to compare the two models' abilities to predict the species-level distribution beyond the population-level data. The GBIF records for *Cuculus canorus* were downloaded and then grouped by observation month to match the time-frames of each seasonal model. These seasonal GBIF datasets were cleaned to remove apparently erroneous records (defined as those falling outside of the reduced background extents for each season in Supplementary material Appendix 1 Fig A3). This left a total GBIF dataset of 177 714 points. However, as the data were so heavily weighted towards records from the breeding season (with 5844 points available for June and only 43 points available for December) we produced five replicate datasets (n = 516) in which 43 points were randomly sampled without replacement from each month. Correspondence between these GBIF records and the model projections was assessed with the test AUC statistic and by considering the percentage of GBIF records covered by the predictions.

All analyses (except for the MaxEnt model itself) were performed in R Studio ver. 1.0.44 (RStudio Team 2015).

Results

Full-year model

Our modelling framework resulted in an excellent fit for the full-year model with a high AUC value and a low difference between scores for training and test data (training AUC = 0.894, test AUC = 0.893) (Swets 1988). The true skill statistic also indicated a good model fit under the threshold 'maximum training sensitivity plus specificity' (TSS = 0.648). Examination of the MaxEnt λ output showed that all parameters were maintained in the final model, with both linear and quadratic features with the exception of maximum monthly temperature which was only retained with quadratic features.

Monthly projections predict the highest probability winter (November–February) occurrence in southwest central Africa (mainly in DR Congo, Congo, Gabon, Equatorial Guinea, Cameroon and the Central African Republic) and along the south-eastern coast of Africa (largely in Tanzania and Mozambique). However, predicted winter presence is extensive and encompasses all of sub-Saharan Africa with the exception of the western halves of South Africa and Namibia and the Horn of Africa. In spring (March–April), this wide extent is largely maintained, but the highest probability areas move to the northwest, focused around the Gulf of Guinea from Sierra Leone to Nigeria. The predicted area of occurrence in Africa contracts northwards through the summer (May–August), but still leaves substantial suitable areas south of the equator. The projection spreads through Europe during the summer, and by June Europe is largely predicted suitable, with the exception of parts of Spain and northern Scandinavia. The prediction includes much of Russia but excludes countries East of the Caspian Sea. In autumn (September–October), predicted occupancy in Europe contracts southwards, and the highest probability of occurrence is predicted in a latitudinal band across central Africa, south of the Sahara (Fig. 1). The total area of predicted occurrence

for the cuckoo is at its lowest in Africa between May and September. However, as this time coincides with the peak availability in Europe, global availability is actually at its highest in the summer months (Fig. 2).

Mean monthly NDVI was the greatest single contributor to AUC for the full-year model (Table 1). Highest modelled probabilities of cuckoo presence occur at high NDVI (0.8 and above); moderately high (175 mm) precipitation in the previous month; a moderately high (25°C) maximum monthly temperature, and a high difference from the previous month's NDVI (Fig. 3). However, inspection of response curves highlighted a differential response to the predictors between the breeding season and the rest of the year, with a tight clustering of breeding season points away from the rest of the data (Fig. 3). This effect is visible in all four predictors, but is most evident in maximum monthly temperature. This is indicative of niche switching where one temperature is preferred during the breeding season, but another is selected in the winter and could argue against the inclusion of breeding data in the full-year model. However, excluding the breeding season data from the full-year model made very little difference to model projections for Africa throughout the year, both in terms of the geographic spread of suitable areas and the total area of occurrence (Supplementary material Appendix 1 Fig. A4, A5). Conversely, excluding the breeding season data, unsurprisingly, seems to compromise model fit during the breeding season and results in a less close match for what is known of the European cuckoo distribution (BirdLife International and NatureServe 2014) (Fig. 1, Supplementary material Appendix 1 Fig. A4).

Seasonal models

All four seasonal models generated very high AUC scores and a low difference between the training AUC score and the test score (breeding season: training AUC = 0.987, test AUC = 0.982; autumn migration: training AUC = 0.912, test AUC = 0.950; winter: training AUC = 0.973, test AUC = 0.961; spring migration: training AUC = 0.973, test AUC = 0.985) (Swets 1988). The true skill statistic also indicated a very good model fit under the threshold 'maximum test sensitivity plus specificity' for autumn migration, winter and spring migration (TSS = 0.670, 0.878 and 0.880 respectively) and an adequate fit for the breeding season (TSS = 0.584). All four parameters were maintained in each of the seasonal models, but combinations of linear and quadratic features varied between models.

The breeding season model produces a very limited and patchy area of predicted occurrence, with the highest probabilities centered on the northern parts of the known breeding range, especially in the United Kingdom, south and central Scandinavia and far northeastern Europe (Fig. 4a). The output for autumn migration shows a relatively extensive area of predicted presence across a wide longitudinal band of sub-Saharan Africa (Fig. 4b). The predicted winter distribution is focused on the Congo Basin and Angola (Fig. 4c). Finally, the spring migration stopovers are predicted to occur in a narrow latitudinal band of coastal West Africa and northern parts of the Congo Basin (Fig. 4d). According to the seasonal models, the lowest area of occurrence is during the breeding season (Fig. 4).
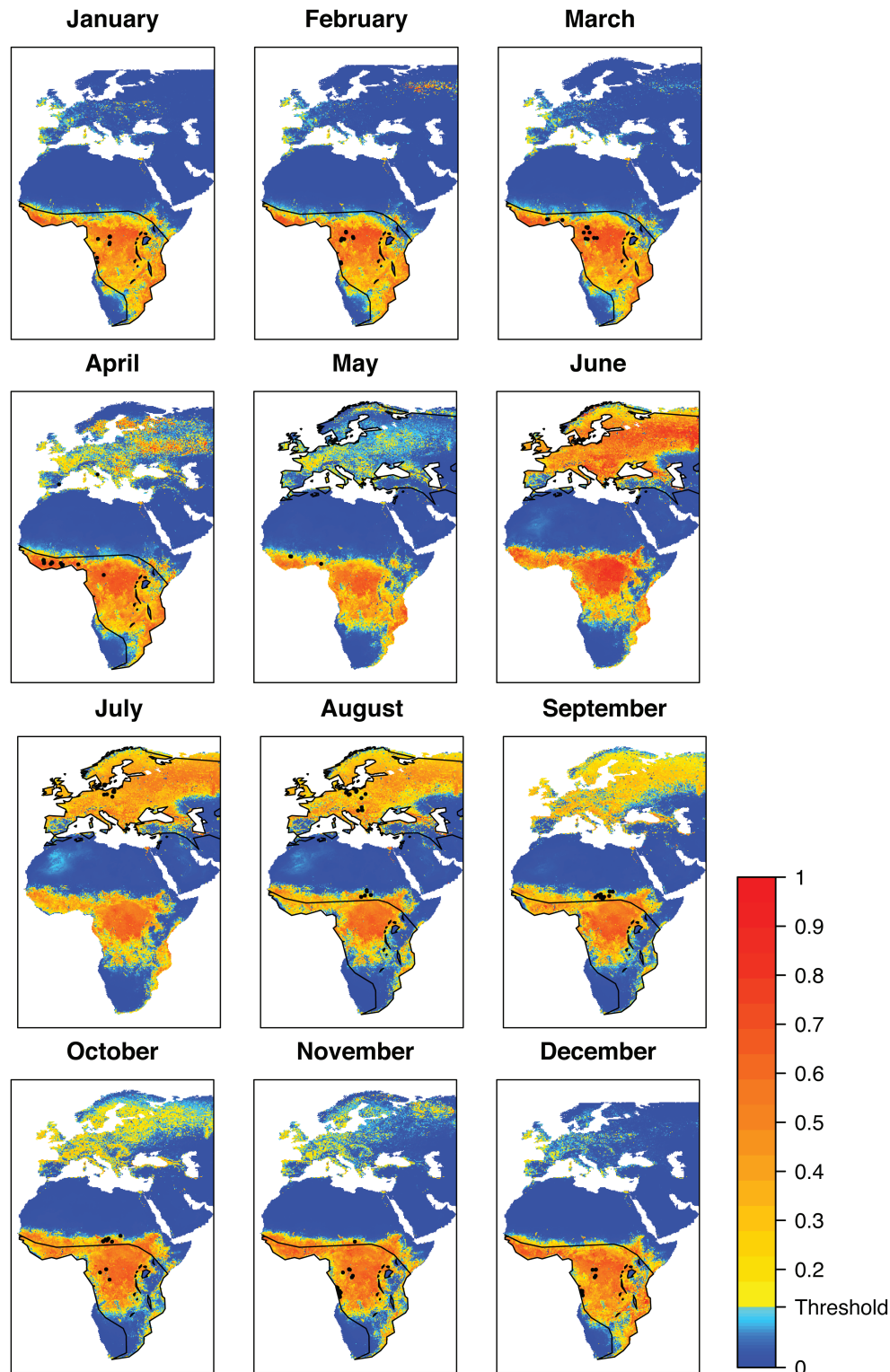
Figure 1. Logistic full-year model output showing probability of cuckoo presence throughout the year. Black points show cuckoo occurrence points used in model building from each month. Black outlines in Europe and Africa show the estimated area of cuckoo occurrence in the breeding and winter seasons respectively according to Birdlife International and NatureServe (2014). Areas of predicted presence according to the threshold rule are shown in yellows, oranges and reds. Probabilities below the threshold are shown in shades of blue.

NDVI was the greatest single contributor to AUC for each seasonal model (except the breeding season model where it was the second most important factor) (Table 1).

Full-year vs seasonal models

Combination of the model projections (12 months for the full-year model and 4 seasons for the seasonal model)
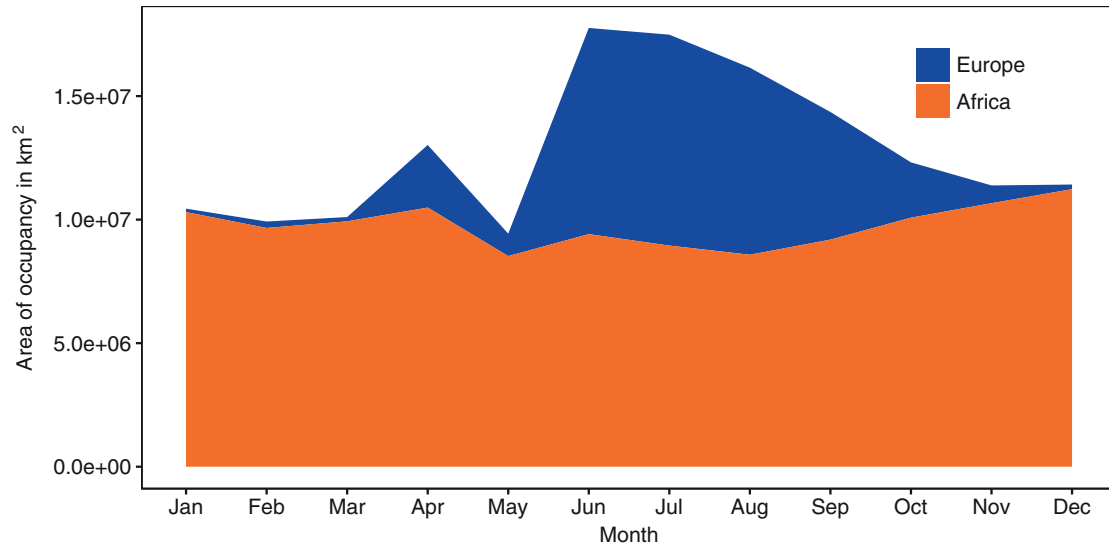
Figure 2. Predicted area of occupancy for the cuckoo through the full year. Areas above 30° latitude as shown as 'Europe' and areas south of 30° latitude are denoted as 'Africa'. Model built using training occurrence data that included breeding season locations. Data are shown cumulatively with available European area placed on top of the available African area.

produced quite different pictures of the full annual distribution of the cuckoo. The full-year model projection was very closely matched to the estimated distribution from BirdLife International and NatureServe (2014), but the combined seasonal models encompassed only a small portion of that range with large areas of under-prediction in the Eurasian breeding grounds and the South-East African part of their winter range (Fig. 5). The complete annual distribution from the combined seasonal models covered just 45% of the area encompassed by the full-year model, with the largest discrepancy in Europe where the seasonal models cover just 13% of the area of the full-year model.

In terms of AUC score the full-year and seasonal models performed similarly at predicting the occurrence of the GBIF data (full-year AUC test = 0.690; breeding AUC test = 0.693; autumn migration AUC test = 0.780; winter AUC test = 0.609; spring migration AUC test = 0.559, mean seasonal AUC test = 0.660). However, only an average of 28% (25.5–29%) of the GBIF occurrence points fell within the predicted distribution from the seasonal models compared with 90% (88.6–91%) with the full-year model (Fig. 5), implying that even though the models achieve similar AUC, that of the seasonal models may be driven by increased specificity without achieving adequate sensitivity.

Discussion

Our modelling framework produces an annual species distribution model for the common cuckoo with a high discriminatory power. SDMs for migrant species traditionally focus on distinct breeding-season and winter-season models but we have incorporated conditions experienced by the cuckoos in their multiple stopovers. As conditions in migratory stopovers can influence population size (Runge et al. 2014), this is a critical consideration for conservation purposes. By including temporal information describing the seasonally changing conditions, our SDM captures the full annual range of a migrant within a single framework. The framework may be generally applicable when aiming to extract species-level generality from population-level data for species with complex migrations.

Our SDM was informed by just four climatic variables: 1) maximum monthly temperature, 2) NDVI, 3) mean precipitation of the previous month, 4) difference in monthly NDVI from that of the previous month. Whilst our study was not primarily focused on optimising the selection of the environmental variables, the high model fit suggests that these are indeed useful parameters to predict the cuckoo distribution. Similar variables (mean precipitation, mean monthly temperature and NDVI) were also used in a recent study describing the distribution of 26 *Sylvia*

Table 1. Contribution of each variable to full-year model AUC.

| Variable	Full-year model permutation importance (%)	Breeding season model permutation importance (%)	Autumn migration model permutation importance (%)	Winter model permutation importance (%)	Spring migration model permutation importance (%)
Mean precipitation of previous month	1.3	12.2	39.6	4.9	0.1
Maximum monthly temperature	9.0	59.9	1.7	32.0	21.5
Mean monthly NDVI	76.9	26.6	43.9	57.1	71.9
Difference in monthly NDVI from previous month	12.8	1.3	14.8	6.0	6.5
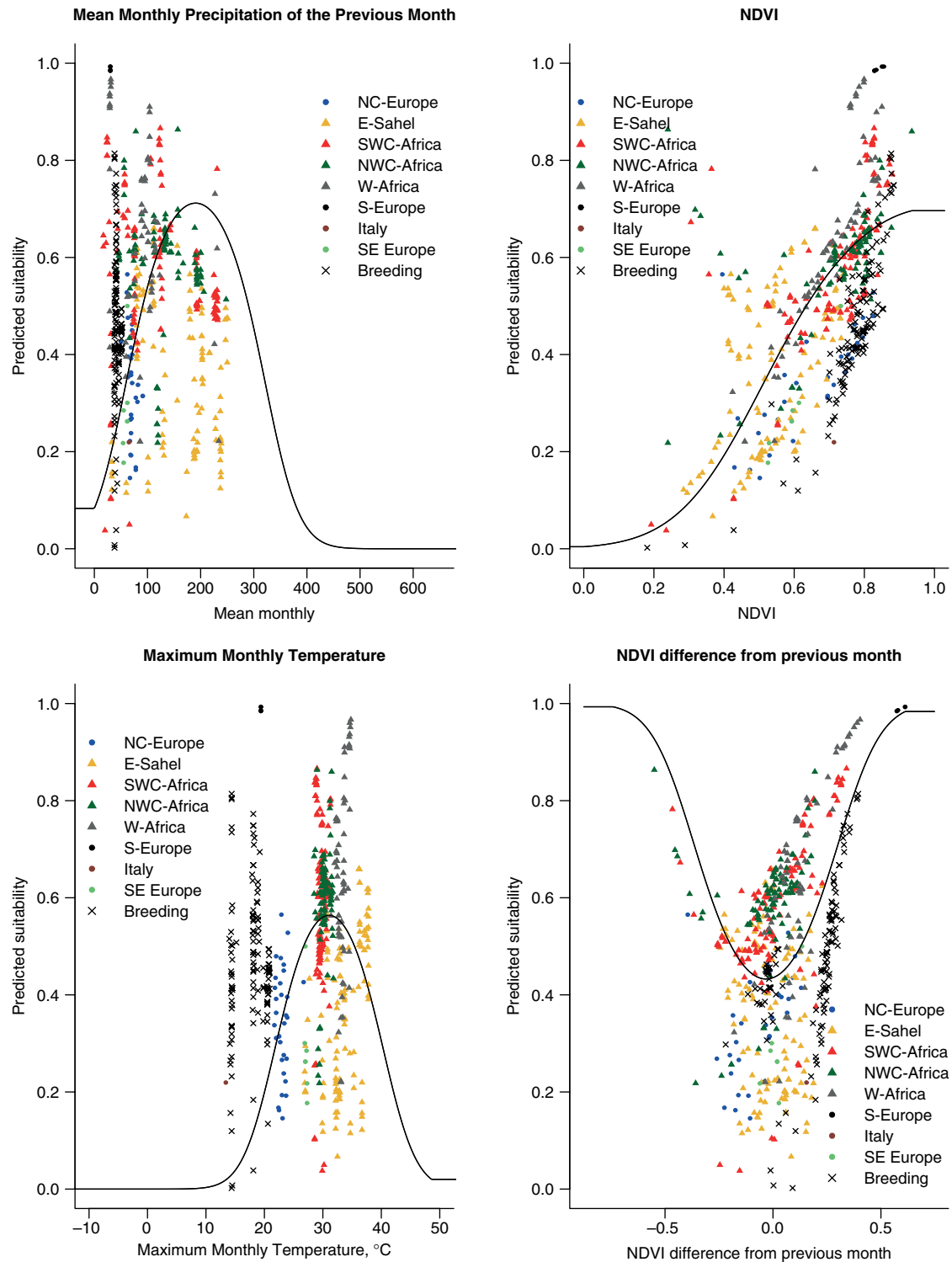
Figure 3. Response curves for each variable in the full-year model. Black lines show the mean values across five crossvalidations. Points show occurrences from each stopover location.

species (Laube et al. 2015) hinting that these factors may be generally predictive of the distribution of insectivorous (or at least partially insectivorous) avian migrants. As the cuckoo is a specialist insectivore, the climate variables used in our model may actually only indirectly determine cuckoo distribution, but have a greater direct impact on caterpillar abundance. That NDVI is predictive of cuckoo distribution is in

agreement with a recent study which found that cuckoos track high average 'greenness' (Thorup et al. 2017) and more general theories that migrant birds 'surf the green wave' of seasonal budburst (Drent et al. 1978).

We compared the performance of our full-year model with a set of four seasonal models at predicting an independent dataset of cuckoo occurrences from GBIF. The GBIF dataset
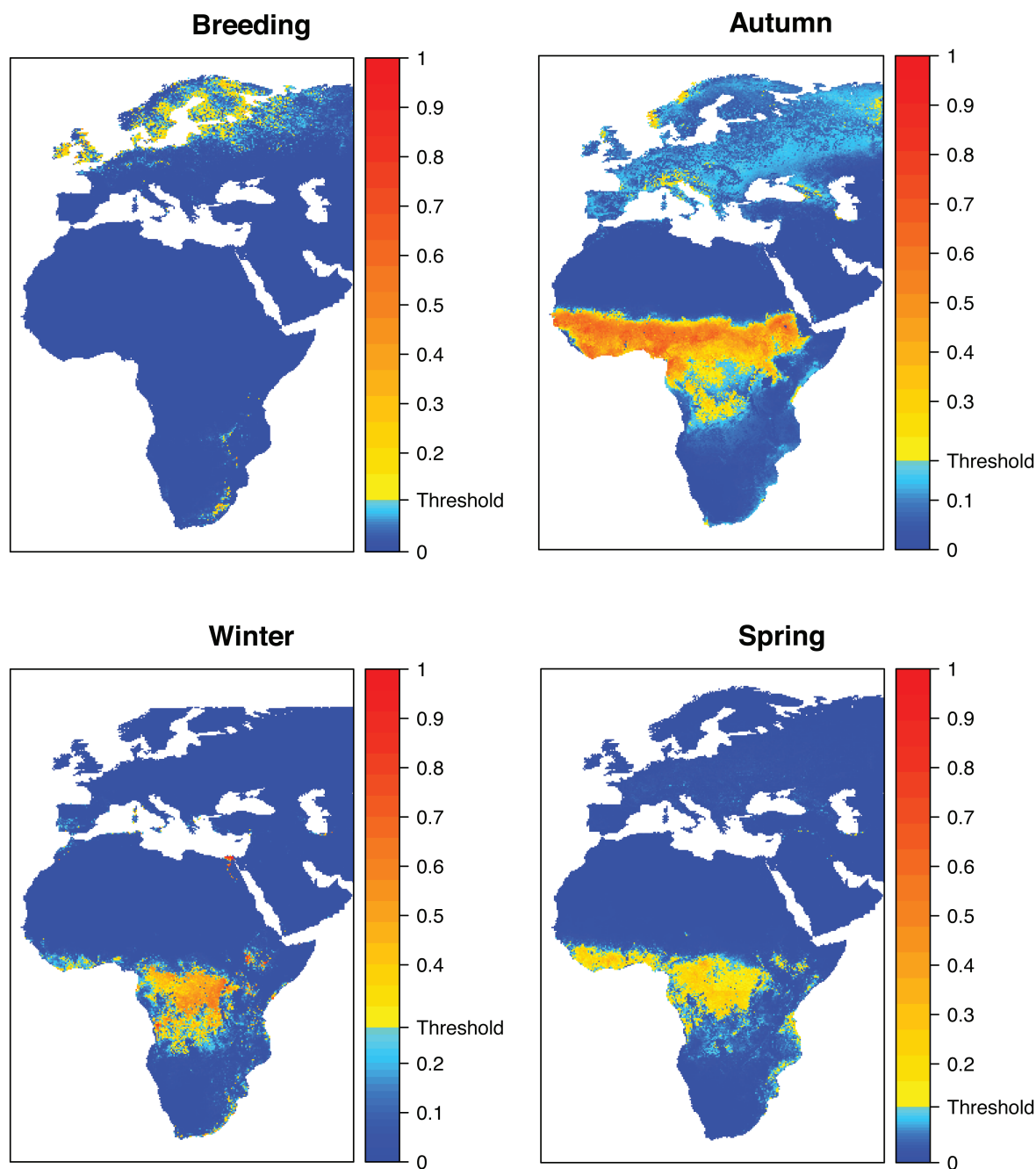
Figure 4. Logistic seasonal model outputs showing probability of cuckoo presence during the breeding season; autumn migration; winter and spring migration, respectively. Areas of predicted presence according to the threshold rule for each model are shown in yellows, oranges and reds. Probabilities below the thresholds are shown in shades of blue.

showed a strong signal of spatial bias, with a large proportion of winter records in South Africa and Tanzania (although other populations of cuckoos are known to winter closer to the South Scandinavian cuckoos in Central Africa (Hewson et al. 2016)) and the vast majority of European records coming from the northern parts of the cuckoo's breeding range (BirdLife International and NatureServe 2014). Even though not all of these regions were featured in our training occurrence data, the full-year model, was successful at predicting their suitability. By comparison, the seasonal models produced a much more spatially restricted overall prediction; were a relatively poor match for what is generally known

of cuckoo winter distribution (BirdLife International and NatureServe 2014) and failed to incorporate the majority of the GBIF occurrences (despite attaining similar test AUC scores as for the full-year model). As the seasonal models are only informed by occurrence data from a small part of the year, they appear to over-fit their projections to these limited data. This could imply that single season models generally underestimate suitable habitat. Furthermore, in situations where birds move between multiple locations within one season, single season models using averaged seasonal conditions may inform their model with false presences and end up overestimating the niche. In contrast, our full-year model

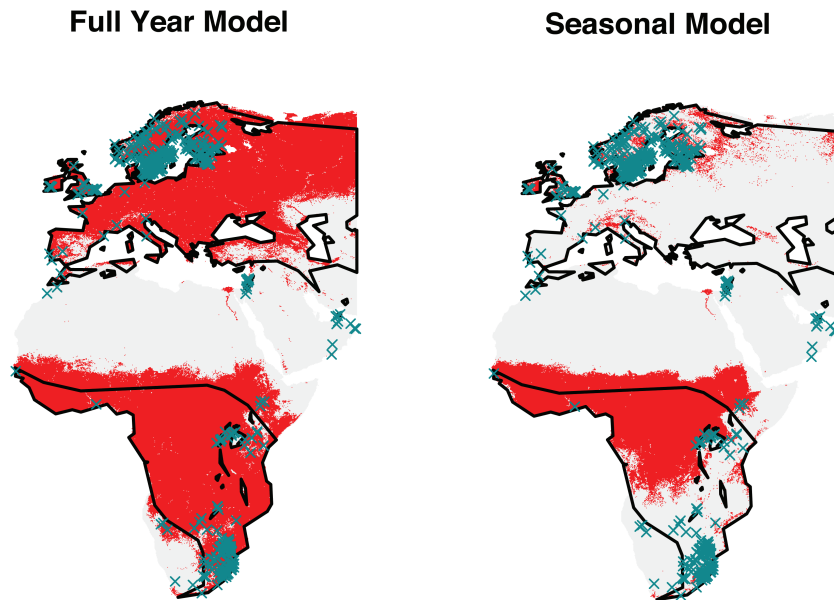| Full Year Model	Seasonal Model

Figure 5. Comparison of full-year (a) and seasonal (b) models in terms of model fit and ability to predict GBIF data from other cuckoo populations. Areas above the predicted occurrence thresholds have been summed for each month and season, respectively, to produce a total annual estimate for each model framework. Turquoise crosses show a seasonally balanced sample of occurrences from the GBIF dataset. Black outlines show the estimated area of cuckoo occurrence according to Birdlife International (2014).

both uses full-year data and incorporates these finer scale movements. Therefore, compared to an approach based only on a single season, full-year modelling may result in more general models for studies aiming at extracting whole-year generality at the species level. Conversely, however, if the research question is focused on delimiting suitable areas during a specific season, or for a specific population, a seasonal modelling approach may be preferred as the generality of the full-year model may create unhelpful over-prediction.

The cuckoo is an obligate trans-Saharan migrant, but our full-year SDM indicated that large areas of Africa remain climatically suitable throughout the entire year. Given the high potential costs of migration (Sillett and Holmes 2002), we therefore need to consider why the cuckoos do not stay year-round in Africa. The first possibility is that this could be simply an erroneous over-prediction of the model, and could imply that there are some un-measured habitat or climatic features making this region unsuitable for cuckoos in the boreal summer months. However, as these equatorial regions are relatively aseasonal and the cuckoo's ecologically similar sister species (the African cuckoo, *Cuculus gularis*) has a distribution overlapping the predicted areas of summer common cuckoo occurrence in Africa, it is also possible that there is a biotic reason for this apparent over-prediction. Whilst the full-year model shows that the global area of climatically suitable cuckoo habitat appears fairly constant through the year, there is a clear constriction of African habitat between May and September. As expansion of European climatic suitability occurs concurrently with the African habitat constriction it appears possible that the cuckoo migrates to the northern hemisphere to exploit the relatively untapped resource pulses (Cox 1968) and to limit potential competition with the African cuckoo.

Whereas the 'full-year with breeding' model (Fig. 1) appears more representative of the species' summer extent in

Europe (BirdLife International and NatureServe 2014) than the equivalent model with breeding data excluded (Supplementary material Appendix 1 Fig. A4) there was a strong signal of niche switching between the breeding season and the rest of the year. Niche switching between the breeding and non-breeding seasons in long distance migrants has been detected previously (Nakazawa et al. 2004, Batalden et al. 2007, Marini et al. 2012, Boucher-Lalonde et al. 2014) and may indicate a change in physiological requirements between life stages (Alerstam et al. 2003), that an abundant food resource becomes temporarily available in sub-optimal temperature conditions further north in the boreal summer, or that conditions in the non-breeding area deteriorate because of other unmeasured environmental factors. Indeed, the cuckoo is known to have different resource needs during the breeding season compared with the rest of the year (Williams et al. 2016), principally due to its brood parasitic life history. However, as excluding the breeding data had very little effect on predictions outside of the breeding season, and both the 'full-year' and 'full-year without breeding data' models performed better at predicting an independent dataset than the seasonal models, our conclusions remain the same whichever of the two full-year models are used. In that sense, the potential niche switching displayed by the cuckoo does not seem to have affected the generality of our model.

The full-year modelling framework includes temporally relevant covariate data from the full annual cycle of a complex long distance migrant and avoids many traditional spatial and temporal biases. This framework should be especially relevant for migrants with complex routes, itinerant species and nomadic species, which spend substantial parts of the year at stopovers other than the breeding or main wintering grounds. The increasing availability of satellite telemetry, geolocator and GPS data should make this modelling

framework feasible for many more migrant species in the near future.

References

Alerstam, T., Hedenström, A. and Åkesson, S. 2003. Long-distance migration: evolution and determinants. – Oikos 103: 247–260.

Allouche, O., Tsoar, A. and Kadmon, R. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). – J. Appl. Ecol. 43: 1223–1232.

Argos 2011. Argos user's manual. – <http://www.argos-system.org/web/en/76-user-s-manual.php>.

Austin, M. 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. – Ecol. Model. 157: 101–118.

Barbet-Massin, M., Walther, B. A., Thuiller, W., Rahbek, C. and Jiguet, F. 2009. Potential impacts of climate change on the winter distribution of Afro-Palaearctic migrant passerines. – Biol. Lett. 5: 248–251.

Batalden, R. V., Oberhauser, K. and Peterson, A. T. 2007. Ecological niches in sequential generations of eastern North American monarch butterflies (Lepidoptera: Danaidae): the ecology of migration and likely climate change implications. – Environ. Entomol. 36: 1365–1373.

Birdlife International and NatureServe 2014. Bird species distribution maps of the world 2012, *Cuculus canorus*. – The IUCN red list of threatened species, ver. 2014.3.

Boakes, E. H., Mcgowan, P. J., Fuller, R. A., Chang-Qing, D., Clark, N. E., O'connor, K. and Mace, G. M. 2010. Distorted views of biodiversity: spatial and temporal bias in species occurrence data. – PLoS Biol. 8: e1000385.

Boucher-Lalonde, V., Kerr, J. T. and Currie, D. J. 2014. Does climate limit species richness by limiting individual species' ranges? – Proc. R. Soc. B 281: 20132695.

Bridge, E. S., Thorup, K., Bowlin, M. S., Chilson, P. B., Diehl, R. H., Fléron, R. W., Hartl, P., Kays, R., Kelly, J. F., Robinson, W. D. and Wikelski, M. 2011. Technology on the move: recent and forthcoming innovations for tracking migratory birds. – BioScience 61: 689–698.

Bridge, E. S., Ross, J. D., Contina, A. J. and Kelly, J. F. 2016. Do molt-migrant songbirds optimize migration routes based on primary productivity? – Behav. Ecol. 27: 784–792.

Buse, A. and Good, J. 1996. Synchronization of larval emergence in winter moth (*Operophtera brumata* L.) and budburst in pedunculate oak (*Quercus robur* L.) under simulated climate change. – Ecol. Entomol. 21: 335–343.

Cardador, L., Sardà-Palomera, F., Carrete, M. and Mañosa, S. 2014. Incorporating spatial constraints in different periods of the annual cycle improves species distribution model performance for a highly mobile bird species. – Divers. Distrib. 20: 515–528.

Colwell, R. K. and Rangel, T. F. 2009. Hutchinson's duality: the once and future niche. – Proc. Natl Acad. Sci. USA 106: 19651–19658.

Cox, G. W. 1968. The role of competition in the evolution of migration. – Evolution 22: 180–192.

Drent, R., Ebbinge, B. and Weijand, B. 1978. Balancing the energy budgets of arctic-breeding geese throughout the annual cycle: a progress report. – Verh. Ornithol. Ges. Bayern 23: 239–264.

Edrén, S. M. C., Wisz, M. S., Teilmann, J., Dietz, R. and Söderkvist, J. 2010. Modelling spatial patterns in harbour porpoise satellite telemetry data using maximum entropy. – Ecography 33: 698–708.

Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Mcc. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M. and E. Zimmermann, N. 2006. Novel methods improve prediction of species' distributions from occurrence data. – Ecography 29: 129–151.

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E. and Yates, C. J. 2011. A statistical explanation of MaxEnt for ecologists. – Divers. Distrib. 17: 43–57.

Fenger, M., Nyegaard, T. and Jørgensen, M. F. 2016. Overvågning af de almindelige fuglearter i Danmark 1975–2015, årsrapport for punkttællingsprogrammet. – Dansk Ornitologisk Forening.

Fink, D., Hochachka, W. M., Zuckerberg, B., Winkler, D. W., Shaby, B., Munson, M. A., Hooker, G., Riedewald, M., Sheldon, D. and Kelling, S. 2010. Spatiotemporal exploratory models for broad-scale survey data. – Ecol. Appl. 20: 2131–2147.

Global biodiversity information facility 2015. *Cuculus canorus*, 276 010 records. – <http://www.gbif.org/>.

Gregory, R. D., Noble, D. G. and Custance, J. 2004. The state of play of farmland birds: population trends and conservation status of lowland farmland birds in the United Kingdom. – Ibis 146: 1–13.

Gschweng, M., Kalko, E. K. V., Berthold, P., Fiedler, W. and Fahr, J. 2012. Multi-temporal distribution modelling with satellite tracking data: predicting responses of a long-distance migrant to changing environmental conditions. – J. Appl. Ecol. 49: 803–813.

Hastie, T. and Fithian, W. 2013. Inference from presence-only data; the ongoing controversy. – Ecography 36: 864–867.

Hayes, M. A., Cryan, P. M. and Wunder, M. B. 2015. Seasonally-dynamic presence-only species distribution models for a cryptic migratory bat impacted by wind energy development. – PloS One 10: e0132599.

Hebblewhite, M. and Haydon, D. T. 2010. Distinguishing technology from biology: a critical review of the use of GPS telemetry data in ecology. – Phil. Trans. R. Soc. B 365: 2303–2312.

Heikkinen, R. K., Luoto, M. and Virkkala, R. 2006. Does seasonal fine-tuning of climatic variables improve the performance of bioclimatic envelope models for migratory birds? – Divers. Distrib. 12: 502–510.

Heldbjerg, H. and Fox, T. 2010. Long-term population declines in Danish trans-Saharan migrant birds. – Bird Study 55: 267–279.

Hewson, C. M., Thorup, K., Pearce-Higgins, J. W. and Atkinson, P. W. 2016. Population decline is linked to migration route in the Common Cuckoo. – Nat. Commun. 7: 12296.

Hijmans, R. J. 2015. raster: geographic data analysis and modeling. – R package ver. 2.4-18, <http://cran.r-project.org/package=raster>.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. and Jarvis, A. 2005. Very high resolution interpolated climate surfaces for global land areas. – Int. J. Climatol. 25: 1965–1978.

Hijmans, R. J., Phillips, S., Leathwick, J. and Elith, J. 2016. dismo: species distribution modeling. – R package ver. 1.1-1, <https://cran.r-project.org/package=dismo>.

Hochachka, W. M., Fink, D., Hutchinson, R. A., Sheldon, D., Wong, W. K. and Kelling, S. 2012. Data-intensive science applied to broad-scale citizen science. – Trends Ecol. Evol. 27: 130–137.

Jonzén, N., Knudsen, E., Holt, R. D. and Sæther, B.-E. 2011. Uncertainty and predictability: the niches of migrants and nomads. – In: Milner-Gulland, E. J., Fryxell, J. M. and Sinclair, A. R. E. (eds), Animal migration: a synthesis. Oxford Univ. Press, pp. 91–109.

Karlsson, P., Bylund, H., Neuvonen, S., Heino, S. and Tjus, M. 2003. Climatic response of budburst in the mountain birch at two areas in northern Fennoscandia and possible responses to global change. – Ecography 26: 617–625.

Kearney, M. R., Matzelle, A. and Helmuth, B. 2012. Biomechanics meets the ecological niche: the importance of temporal data resolution. – J. Exp. Biol. 215: 922–933.

Klaassen, R. H. G., Strandberg, R., Hake, M., Olofsson, P., Tøttrup, A. P. and Alerstam, T. 2010. Loop migration in adult marsh harriers Circus aeruginosus, as revealed by satellite telemetry. – J. Avian Biol. 41: 200–207.

Klaassen, M., Hoye, B. J., Nolet, B. A. and Buttemer, W. A. 2012. Ecophysiology of avian migration in the face of current global hazards. – Phil. Trans. R. Soc. B 367: 1719–1732.

Laube, I., Graham, C. H. and Böhning-Gaese, K. 2015. Niche availability in space and time: migration in Sylvia warblers. – J. Biogeogr. 42: 1896–1906.

Limaiñana, R., Arroyo, B., Terraaube, J., Mcgrady, M. and Mougeot, F. 2015. Using satellite telemetry and environmental niche modelling to inform conseravtion targets for a long-distance migratory raptor in its wintering grounds. – Oryx 49: 329–227.

Liu, C., White, M., Newell, G. and Pearson, R. 2013. Selecting thresholds for the prediction of species occurrence with presence-only data. – J. Biogeogr. 40: 778–789.

Lobo, J. M., Jiménez-Valverde, A. and Real, R. 2008. AUC: a misleading measure of the performance of predictive distribution models. – Global Ecol. Biogeogr. 17: 145–151.

Marini, M. Â., Barbet-Massin, M., Lopes, L. E. and Jiguet, F. 2012. Geographic and seasonal distribution of the Cock-tailed Tyrant (Alectrurus tricolor) inferred from niche modeling. – J. Ornithol. 154: 393–402.

Martinez-Meyer, E., Peterson, A. T. and Navarro-Siguenza, A. G. 2004. Evolution of seasonal ecological niches in the Passerina buntings (Aves: Cardinalidae). – Proc. R. Soc. B 271: 1151–1157.

Merow, C., Smith, M. J. and Silander, J. A. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. – Ecography 36: 1058–1069.

Merow, C., Smith, M. J., Edwards, T. C., Guisan, A., Mcmahon, S. M., Normand, S., Thuiller, W., Wüest, R. O., Zimmermann, N. E. and Elith, J. 2014. What do we gain from simplicity versus complexity in species distribution models? – Ecography 37: 1267–1281.

Nakazawa, Y., Peterson, A. T., Martínez-Meyer, E. and Navarro-Sigüenza, A. G. 2004. Seasonal niches of nearctic-neotropical migratory birds: implications for the evolution of migration. – Auk 121: 610–618.

Nasa 2015. NASA Data Portal. – <https://data.nasa.gov/>.

Newton, I. 2004. Population limitation in migrants. – Ibis 146: 197–226.

Payne, R. D. 2005. The cuckoos. – Oxford Univ. Press.

Peterson, A. T., Martinez-Campos, C., Nakazawa, Y. and Martinez-Meyer, E. 2005. Time-specific ecological niche modeling predicts spatial dynamics of vector insects and human dengue cases. – Trans. R. Soc. Trop. Med. Hyg. 99: 647–55.

Phillips, S. J. and Dudik, M. 2007. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. – Ecography 31: 161–175.

Phillips, S. J. and Elith, J. 2013. On estimating probability of presence from use–availability or presence–background data. – Ecology 94: 1409–1419.

Phillips, S. J., Anderson, R. P. and Schapire, R. E. 2006. Maximum entropy modeling of species geographic distributions. – Ecol. Model. 190: 231–259.

Phillips, S. J., Dudík, M., Elith, J., Graham, C. H., Lehmann, A., Leathwick, J. and Ferrier, S. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. – Ecol. Appl. 19: 181–197.

Pinto, C. A., Henriques, M. O., Figueiredo, J. P., David, J. S., Abreu, F. G., Pereira, J. S., Correia, I. and David, T. S. 2011. Phenology and growth dynamics in Mediterranean evergreen oaks: Effects of environmental conditions and water relations. – For. Ecol. Manage. 262: 500–508.

Reddy, S. and Davalos, L. M. 2003. Geographic sampling bias and its implications for conseravtion priorities in Africa. – J. Biogeogr. 30: 1719–1727.

Root, T. 1988. Energy constraints on avian distributions and abundances. – Ecology 69: 330–339.

Royle, J. A., Chandler, R. B., Yackulic, C. and Nichols, J. D. 2012. Likelihood analysis of species occurrence probability from presence-only data for modelling species distributions. – Methods Ecol. Evol. 3: 545–554.

Rstudio Team 2015. RStudio: Integrated Development for R. – RStudio, Boston, MA, <www.rstudio.com/>.

Runge, C. A., Martin, T. G., Possingham, H. P., Willis, S. G. and Fuller, R. A. 2014. Conserving mobile species. – Front. Ecol. Environ. 12: 395–402.

Runge, C. A., Watson, J. E., Butchart, S. H., Hanson, J. O., Possingham, H. P. and Fuller, R. A. 2015. Protected areas and global conservation of migratory birds. – Science 350: 1255–1258.

Sanderson, F. J., Donald, P. F., Pain, D. J., Burfield, I. J. and Van Bommel, F. P. J. 2006. Long-term population declines in Afro-Palearctic migrant birds. – Biol. Conserv. 131: 93–105.

Sillett, T. S. and Holmes, R. T. 2002. Variation in survivorship of a migratory songbird throughout its annual cycle. – J. Anim. Ecol. 71: 296–308.

Soudani, K., Hmimina, G., Delpierre, N., Pontailler, J. Y., Aubinet, M., Bonal, D., Caquet, B., De Grandcourt, A., Burban, B., Flechard, C., Guyon, D., Granier, A., Gross, P., Heinesh, B., Longdoz, B., Loustau, D., Moureaux, C., Ourcival, J. M., Rambal, S., Saint André, L. and Dufrêne, E. 2012. Ground-based network of NDVI measurements for tracking temporal dynamics of canopy structure and vegetation phenology in different biomes. – Remote Sens. Environ. 123: 234–245.

Swets, J. A. 1988. Measuring the accuracy of diagnostic systems. – Science 240: 1285–1293.

Thorup, K., Tøttrup, A. P., Willemoes, M., Klaassen, R. H. G., Strandberg, R., Lomas Vega, M., Dasari, H. P., Araújo, M. B., Wikelski, M. and Rahbek, C. 2017. Resource tracking within and across continents in long-distance bird migrants. – Sci. Adv. 3: 1–10.

van Asch, M. and Visser, M. E. 2007. Phenology of forest caterpillars and their host trees: the importance of synchrony. – Annu. Rev. Entomol. 52: 37–55.

VanDerWal, J., Shoo, L. P., Graham, C. and Williams, S. E. 2009. Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? – Ecol. Model. 220: 589–594.

Vickery, J. A., Ewing, S. R., Pain, D. J., Bairlein, F., Skorpilova, J. and Gregory, R. D. 2014. The decline of Afro-Palaearctic migrants and an assessment of potential causes. – Ibis 156: 1–22.

Warren, D. L., Glor, R. E. and Turelli, M. 2008. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. – Evolution 62: 2868–2883.

Wickham, H. 2009. ggplot2: elegant graphics for data analysis. – Springer.

Willemoes, M., Strandberg, R., Klaassen, R. H. G., Tøttrup, A. P., Vardanis, Y., Howey, P. W., Thorup, K., Wikelski, M. and Alerstam, T. 2014. Narrow-front loop migration in a population of the common cuckoo, *Cuculus canorus*, as revealed by satellite telemetry. – PLoS One 9: e83515.

Williams, H. W., Willemoes, M., Klaassen, R. H. G., Strandberg, R. and Thorup, K. 2016. Common cuckoo home ranges are larger in the breeding season than in the non-breeding season and in regions of sparse forest cover. – J. Ornithol. 157: 461–469.

Supplementary material (Appendix JAV-01476 at <www.avianbiology.org/appendix/jav-01476>). Appendix 1.