Running head: Climatic niche of the Saker Falcon

Climatic niche of the Saker Falcon *Falco cherrug*: predicted new areas to direct population surveys in Central Asia.

LUKE J. SUTTON[1,2,3*] & ROBERT PUSCHENDORF[1]

[1]School of Biological and Marine Sciences, University of Plymouth, Drake Circus, Plymouth, PL4 8AA, United Kingdom

[2]The Peregrine Fund, 5668 West Flying Hawk Lane, Boise, Idaho 83709 USA

[3]Hierofalcon Research Group, Via Stazzone 235, 95124 Catania, Italy

*Corresponding author

Email: lukesutton696@gmail.com

Accurate species distribution data across remote and extensive geographic areas are difficult to obtain. Here, we use bioclimatic envelope models to determine climatic constraints on the distribution of the migratory Saker Falcon *Falco cherrug* to identify areas in data-deficient regions that may contain unidentified populations. Sakers live at low densities, across large ranges in remote regions, making distribution status difficult to assess. Using presence-background data and eight bioclimatic variables within a Species Distribution Modelling framework, we applied MAXENT to construct models for both breeding and wintering ranges. Occurrence data were spatially filtered, and climatic variables tested for multicollinearity, before selecting best fit models using Akaike Information Criteria by tuning MAXENT

parameters. Model predictive performance tested using Continuous Boyce index (B) was high for both breeding (B_{TEST} = 0.921) and wintering models (B_{TEST} = 0.735), with low omission rates and minimal overfitting. The Sakers' climatic niche was defined by precipitation in the warmest quarter in the breeding range model, and mean temperature in the wettest quarter in the wintering range model. Our models accurately predicted areas of highest climate suitability and defined the climatic constraints on a wide-ranging rare species, suggesting that climate is a key determinant of Saker distribution across macro-scales. We recommend targeted population surveys for the Saker based on model predictions to areas of highest climatic suitability in key regions with distribution knowledge gaps, in particular the Qinghai-Tibet plateau in western China. Further applications of our models could identify protected areas and reintroduction sites, inform development conflicts, and assess the impact of climate change on distributions.

Defining a species' geographic range is fundamental to conservation planning (Lawler *et al.* 2011), and for effective conservation management (Wu & Smeins 2000; Miller 2010). However, obtaining accurate species distribution data across remote and extensive geographic areas is difficult and time-consuming (Austin *et al.* 1996), especially for high-altitude steppe habitats (Osborne 2005). Species Distribution Models (SDMs) quantify correlations between species occurrence and environmental factors, generating maps of predicted species distributions (Franklin 2009; Peterson *et al.* 2011), helping land managers to set planning priorities for

species of conservation concern (Raxworthy *et al.* 2003; Elith & Leathwick 2009b; de Carvalho *et al.* 2017). Empirically-derived SDMs can highlight specific environmental factors influencing species' distributions and identify areas to locate populations in un-sampled areas within a species range (Guisan *et al.* 2006a; Miller 2010). Using presence-background data and interpolating to un-sampled areas within known ranges using predictive modelling (model-based interpolation, Elith & Leathwick 2009a), is a convenient and cost-effective approach to locate areas that are suitable for the species, and hence target ground surveys to these areas (Peterson 2001; Pearce & Boyce 2006; Rhoden *et al.* 2017).

Falcons (Falconiformes) are widespread diurnal raptors present on all continents apart from Antarctica. As tertiary-level consumers, they face multiple threats such as habitat loss and fragmentation, persecution, prey depletion and environmental contaminants (Newton 1979; Chancellor & Meyburg 2000). Conservation planning for threatened falcons based on SDMs has been effective for species such as the Red-footed Falcon *Falco vespertinus* in eastern Europe, guiding artificial nest site provision, directing population monitoring and designating protected areas (Fehévári *et al.* 2012). Booms *et al.* (2010), used SDMs to predict Gyrfalcon *Falco rusticolus* nest occurrence in remote areas of Arctic Alaska, with high prediction accuracy and successful ground-truthing of the predictive model. Both studies demonstrate how SDMs can guide best management practices, and direct future planning strategies for raptors of high conservation concern in remote landscapes.

The Saker Falcon *Falco cherrug* (hereafter 'Saker'), is a large falcon within the subgenus *Hierofalco,* along with other congeneric 'desert falcons': Gyrfalcon, Laggar

Falcon *Falco juggar* and Lanner Falcon *Falco biarmicus* (Nittinger *et al.* 2007). The Saker has a Palearctic breeding distribution extending from central Europe in the west, to north-eastern China and Mongolia in the east (Ferguson-Lees & Christie 2005), with a latitudinal breeding range of 56°N to 28°N. A migratory sub-population generally winters south of the breeding range in Afro-tropical (to latitude 3-4°S) and Indo-Malayan (to latitude 21°S) zoogeographic realms (Ferguson-Lees & Christie 2005), though many wintering Sakers also inhabit southerly regions of their breeding distribution (Dixon *et al.* 2016).

The Saker is a bioclimatic semi-specialist (Finlayson 2011), preferring open arid landscapes, semi-desert, and wooded steppe, from sea-level to 4700m (Kovács *et al.* 2014). However, Sakers also reside in open agricultural lowland landscapes, especially in eastern Europe (Ragyov *et al.* 2014). Saker diet is dominated by rodents such as ground-squirrels *Spermophilus sp.*, along with birds, and occasionally reptiles and insects (Watson & Clarke 2000; Wu 2011; Bondì *et al.* 2014; Nedyalkov *et al.* 2014). Currently, two biogeographic sub-species are recognised based on morphology: *F. c. cherrug* (east Europe to southern-central Siberia) and *F. c. milvipes* (from southern-central Siberia to north-eastern China) (Ferguson-Less & Christie 2005; Nittinger *et al.* 2007). Dispersing Sakers mainly follow a roughly north-south migratory pattern, with no evidence for breeding individuals dispersing into either sub-species population (Kovács *et al.* 2014), despite recent studies suggesting genetic similarity between individuals of the two sub-species (Nittinger *et al.* 2007).

Saker populations have declined significantly over the past 20 years due to multiple factors, including over-harvesting for falconry, habitat loss, electrocution, poisoning, persecution, prey loss and hybridisation with escaped falconry birds (Bagyura *et al.* 2004; Galushin 2004; Kovács *et al.* 2014; Dixon 2016). There has been a broad-scale conservation effort to reverse population declines, by regional conservation groups, research institutions, and non-governmental organisations (summarized in Kovács *et al.* 2014). However, addressing key knowledge gaps in Saker distribution is required to help achieve a sustainable global population into the future (Collar *et al.* 2013; Kovács *et al.* 2014). In 2015 the conservation status of the Saker was upgraded to Endangered on the IUCN Red List (BirdLife International 2017) after revised population estimates suggested a rapid decline of ~50% in the global population over the past 20 years, particularly in central Asia (Kovács *et al.* 2014). However, this classification is uncertain and may need revision based on new population estimates. Surveys need prioritisation to establish reliable population estimates, for important but remote breeding areas in China, Mongolia and central Asia (Dixon 2009; BirdLife International 2017).

Much of the Saker's Asian breeding range extends across remote geo-political regions with limited resources for biologists to conduct extensive surveys (Dixon 2005, 2009). Predicting areas with highest climatic suitability for breeding Sakers can help prioritise survey effort, saving time and resources (Nicholls 1989; Fielding & Bell 1997), thus informing revised population estimates, and directing conservation effort to key areas. Here, we report the results of the first SDMs covering the entire range of the Saker for both breeding and wintering distributions. Our approach aims to facilitate greater understanding of Saker distributions, and direct Saker recovery

projects in the long-term. Specifically, our objectives are: (1) predict whole range, breeding and wintering distributions for the Saker, (2) identify areas of highest climatic suitability, (3) determine the most important climatic characteristics that explain these respective distributions, (4) project models to high priority areas directing survey effort.

METHODS

Saker Falcon occurrence data

A total of 4907 Saker occurrence points were downloaded from the Global Biodiversity Information Facility database (GBIF 2017a,b; http://www.gbif.org/). The datasets were cleaned by removing duplicated occurrence points, records with no georeferenced location or over the sea, and any vagrant sightings appearing as outliers (Liu *et al.* 2017). We removed occurrences with < 2 decimal degree points to retain location accuracy. Geographic extent was defined by the subset of records within 10°-130° longitude and -5°-60° latitude following current known distribution (Kovács *et al.* 2014). After data cleaning, 854 geo-referenced records were compiled for the whole range distribution model. Georeferenced records with no date were excluded from analysis for seasonal breeding and wintering distributions. Breeding season records were defined as those occurrences recorded between March – August (Dixon *et al.* 2009; Kovács *et al.* 2014), and wintering records between September – February (Kovács *et al.* 2014; Dixon *et al.* 2015), resulting in 513 and 195 cleaned data records for breeding and wintering distributions respectively. We focused on model predictions and analysis for breeding and wintering ranges, but

also include a whole-range prediction map as a supplement to compare with current range maps which generally use the entire range of the Saker.

To reduce spatial auto-correlation and sampling bias in occurrence data, clustered occurrence points were removed using spatial filtering to reduce model over-fitting (Boria *et al.* 2014; Radosavljevic & Anderson 2014). Spatial filtering performs better than other methods of sampling bias correction (Kramer-Schadt *et al.* 2013; Fourcade *et al.* 2014), minimizing omission errors and improving model predictive performance (Aryal *et al.* 2016). A spatial filter distance of 40km between occurrence points was selected to minimise the effects of over-sampling in highly surveyed areas, rather than as a measure of home range size. Saker home range estimates vary widely, with wintering range sizes between 5-515 km² (Dixon *et al.* 2016), and breeding ranges between 78-215 km² (Potapov *et al.* 2000), thus making any estimate on range size for use in spatial filtering problematic. Filtering occurrence points for inclusion was determined by removing clustered points using the 'thin' algorithm function in the R package SPTHIN (Aiello-Lammens *et al.* 2015). Spatial auto-correlation was measured using Global Moran's *I* index on an inverse Euclidean distance matrix projected in Lambert Azimuth Equal Area (LAEA). Moran's *I* is an index ranging from -1 to +1, with values closer to zero indicating no spatial auto-correlation, and negative and positive values indicating negative and positive spatial auto-correlation respectively.

Inevitably, removing data points reduces sample size. Therefore, exponential spatial filters of 10, 20, 40, 80 and 160 km were compared using sample size and spatial auto-correlation as determinants of filter selection. Previous studies have used 10

km as an arbitrary distance for spatial filtering for species in high altitude heterogenous habitats (Pearson *et al.* 2007; Kramer-Schadt *et al.* 2013; Boria *et al.* 2014). Therefore, we considered a 40 km spatial filter as an intermediate distance to minimize spatial bias, whilst retaining a high number of data points for model accuracy and novel environmental conditions (Table S1, Franklin 2009; Boria *et al.* 2014). Cleaned whole-range occurrence data were spatially auto-correlated (Moran's I = 0.457, p = <0.001). After spatial filtering, spatial auto-correlation was reduced (Moran's I = 0.295, p = <0.001), resulting in 516 occurrence records for use in the whole-range distribution model. Spatial auto-correlation was apparent for both cleaned breeding range (Moran's I = 0.526, p = <0.001), and wintering range (Moran's I = 0.483, p = <0.001) occurrence data. After spatial filtering, spatial auto-correlation was reduced (breeding: Moran's I = 0.472, p = <0.001; wintering: Moran's I = 0.356, p = <0.001), resulting in 315 and 133 occurrence records for use in the breeding and wintering distribution models respectively. A common issue in SDM, especially with rare species, or those that are difficult to survey, such as the Saker, is sample size (Pearson *et al.* 2007; Gibson *et al.* 2007). We considered the respective occurrence records obtained after data cleaning and filtering as suitable for a machine-learning algorithm SDM (Stockwell & Peterson 2002; Wisz *et al.* 2008).

Environmental variables

Bioclimatic data were sourced from the WorldClim database (version 1.4, Hijmans *et al.* 2005). Bioclimatic variables were generated through interpolation of average monthly climate data from weather stations over the period 1950-2000. Raster layers were downloaded at a spatial resolution of 2.5 arc-minutes (~4.5km² at the equator)

and cropped to 1° beyond the extent of Saker occurrences using a delimited rectangle. This generates a defined envelope to capture the full range of suitable environmental conditions across the Sakers known range, thus improving model predictive power (Lawler *et al.* 2011). A limited raster extent also removes large regions not occupied by a species, thus reducing the area used for background selection points in model evaluation (Radosavljevic & Anderson 2014). We used solely climatic variables as a first evaluation of climatic constraints on geographic distribution for a rare species with little prior knowledge of biology (Collar *et al.* 2013; Galante *et al.* 2018).

Multi-collinearity between environmental variables can result in biased predictions through over-representing the biological relevance of correlated variables (Franklin 2009; Dormann *et al.* 2013; see Appendix 1 in Supporting Online Information for details). Therefore, before model construction, all 19 bioclimatic variables were tested for multi-collinearity using Variance Inflation Factor (VIF) analysis (Guisan *et al.* 2006b; Hair *et al.* 2006) in the R package USDM (Naimi *et al.* 2014). VIF is based on the square of multiple correlation coefficents, regressing a single predictor variable against all other predictors. VIF tests can detect hidden correlations in predictors not always apparent in pair-wise correlations. VIF > 10 indicates collinearity in the variables, thus we used a stepwise elimination of highly correlated variables retaining predictors with a more stringent VIF threshold of < 5 (Table S2), considered as suitable for multi-variable correlation (Dormann *et al.* 2013).

Eight variables were used as predictors in all models (mean diurnal temperature range; temperature seasonality; mean temperature wettest quarter; mean

temperature warmest quarter; precipitation wettest month; precipitation driest month; precipitation seasonality; precipitation coldest quarter), with predictor selection based on representing seasonal climatic trends, extremes and variability, which are strongly related theoretically and empirically to species distributions (Stockwell 2006; Reineking *et al.* 2016; Bradie & Leung 2017), and specifically to distributions of vagile bird species in arid environments (Reside *et al.* 2010). We selected climate predictors based on current knowledge and potential limiting environmental factors on Saker biology, preferring arid, temperate climatic conditions (Finlayson 2011), in predominantly high-altitude areas, or in regions of high daily and monthly temperature variability (Kovács *et al.* 2014; see Appendix 1 in Supporting Online Information for details).

Species Distribution Models

To predict areas of climate suitability a correlative SDM was developed using MAXENT (version 3.3.3k, Philips *et al.* 2006; Philips & Dudík 2008), a maximum entropy machine-learning algorithm. MAXENT uses presence-background data and is a robust algorithm that out-performs other SDM methods (Elith *et al.* 2006, 2011; Gibson *et al.* 2007; Duan *et al.* 2013), with good prediction accuracy across all sample sizes (Wisz *et al.* 2008). The MAXENT algorithm compares the environmental variables underlying the species occurrence points against the range of a random sample of background environmental conditions. A raster map is generated, with each cell ranked with an index of relative environmental suitability (Zeng *et al.* 2015). MAXENT uses L1-regularization (β) to reduce overfitting in complex models (Hastie *et al.* 2005), enabling improved discrimination of correlated predictors compared to other

SDM methods (Philips & Dudík 2008; Elith *et al.* 2011). Calibration and evaluation datasets were run using logistic output as a continuous index of climatic suitability, with 0 = low suitability and 1 = high climatic suitability. Iterations were increased from the default 500 to 5000 allowing for model convergence, and random seed selected to randomly sample occurrence test points for each replicate run. Default model parameters were used for background absences (10,000) and convergent threshold (0.00001).

Model evaluation

We used optimal-model selection based on Akikake Information Criterion corrected for small sample sizes (AIC$_c$; Warren & Seifert 2011), to determine the most parsimonious models from two key MAXENT parameters affecting model complexity: Regularization Multiplier (RM) and Feature Classes (FC). AIC$_c$ selects models based on a balance between goodness-of-fit and model complexity (Guisan & Thuiller 2005; Muscarella *et al.* 2014), and consistently selects less complex models robust to sampling bias (Galante *et al.* 2018; Gerstner *et al.* 2018). Species-specific tuning of RM and FC produces simpler and more realistic MAXENT models compared to using default settings (Warren & Seifert 2011; Muscarella *et al.* 2014), by smoothing response curves, limiting sampling bias, and reducing over-fitting in presence-only predictions (Merow *et al.* 2013; Radosavljevic & Anderson 2014). AIC$_c$ requires no external evaluation data, thus is a useful metric for presence-background SDMs (Muscarella *et al.* 2014; Galante *et al.* 2018). Lowest AIC$_c$ scores (i.e. ΔAIC$_c$ = 0) indicate the model that best predicts the training data, given data complexity and justified number of parameters (Gerstner *et al.* 2018), though all models with ΔAIC$_c$ <

2 are considered to have substantial support (Burnham & Anderson 2004). Forty-eight candidate models of varying complexity were built for each SDM by comparing a range of regularization multipliers from 0.5 to 4.0 in 0.5 increments, and five feature classes: linear (L); quadratic (Q); hinge (H); product (P); threshold (T), in all possible combinations using the ENMEVAL package in R (Muscarella *et al.* 2014).

Feature classes and regularization multiplier were selected based on the optimal-model output from all 48 candidate models with ΔAIC_c <2 using the 'random *k*-fold' cross-validation method in ENMEVAL (*k* = 10). The *k*-fold cross-validation method divides data into *k* number of mutually exclusive sub-sets, randomly selecting 90% of occurrence records for training, and 10% for testing. Model performance is determined by removing each successive dataset, then re-estimating the model based on the withheld data (Kohavi 1995; Hastie et al. 2005). The use of cross-validation is justified here in the context of a broad range single-species model applied with a spatial filter, with large distances between testing and training points reducing any spatial sorting bias (Hjimans 2012). We extracted a range of metrics from the ENMEVAL output to assess the best fit model, including AIC_c values, overfitting metrics (minimum omission rate (MOR); 10% omission rate (10%OR)) and Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC; see Appendix 2 in Supporting Online Information for details).

Omission rates measure the proportion of training points that are outside of the model prediction and evaluate discriminatory ability and over-fitting at specified thresholds. Low omission rates show improved discrimination between suitable and unsuitable habitats (indicating higher performance), whilst overfitted models show

higher omission rates than expected (Radosavljevic & Anderson 2014). For low overfitting models the expectation in MOR is a value close to zero and for 10%OR a value close to 0.10. Jackknife tests were used to estimate variable performance within the optimal calibration model by excluding each value, then developing the model with a sole variable to determine percentage contribution and regularized training gain of each environmental variable to model performance.

We used all occurrence points in final distribution maps to achieve the highest predictive accuracy for estimating distribution (Fielding & Bell 1997; Fehévári *et al.* 2012). Continuous Boyce index (*B*) was used as a threshold-independent metric to evaluate final output models (Hirzel *et al.* 2006; Ramírez-Albores *et al.* 2016). *B* measures how much climate suitability predictions differ from a random distribution of observed presences across the spatial predictions (Boyce *et al.* 2002), comparable to a Spearman correlation (R_s). Values of *B* range from -1 to +1, with positive values indicting climate suitability predictions consistent with observed presences, values closer to zero no different than a random model, and negative values indicating areas of poor predicted climate suitability but with a high number of observations. *B* evaluation was used on all presence data points split into 90% training (B_{TRAIN}) and 10% testing (B_{TEST}), calculated using the R package ECOSPAT (Di Cola *et al.* 2017), with a moving window for threshold-independence and ten defined bins.

Projected distribution maps were modelled for two important range countries within the Saker's known breeding and wintering ranges with key knowledge gaps: China and Mongolia (Dixon 2009; Collar *et al.* 2013). Both countries have the highest

spatial priority ranking to direct conservation efforts most likely to result in effective conservation outcomes in the United Nations Convention on Migratory Species (UN CMS) Saker Global Action Plan (Kovács *et al.* 2014). Both China and Mongolia have reasonable knowledge of distribution in some areas but limited knowledge elsewhere and are likely to hold significant unknown Saker populations (Dixon 2009). Within China we selected the Qinghai-Tibet plateau as a priority region lacking information on Saker distribution, ecology and population status (Dixon *et al.* 2015; Dixon *et al.* 2016). Clamping was applied when projecting to regions with limited distribution data to restrict model interpolation based on the extent of environmental predictors (Stohlgren *et al.* 2011). The process of clamping constrains environmental variable features to within the range of occurrence values given in the calibration data, identifying any uncertain predictions in projected distribution models (Elith *et al.* 2010). Model construction, analysis and predictive GIS maps were built in R (version 3.3.1; R Core Team 2016) using the DISMO (Hijmans 2017) and RASTER (Hijmans 2016) packages.

RESULTS

Species distribution models

From the 48 candidate models in all SDMs, we selected MAXENT settings based on the optimal-model output with lowest ΔAIC_c: breeding range Regularization Multiplier (RM) = 2, wintering range RM = 2.5, whole-range RM = 1.5; Feature Classes = all models = Linear, Quadratic, Hinge, Product. Predictive performance across all ten-fold cross-validated best fit models had moderate to high predictive accuracy, with all

models demonstrating robust predictive metrics with low omission rates (Table 1). Mean AUC_{TEST} and AUC_{TRAIN} values were all > 0.79 across the three range predictions, with low AUC_{DIFF} values indicating limited over-fitting (Table 1). Test omission rates were close to defined thresholds in all models, indicating effective discriminatory ability between suitable and unsuitable climatic areas, and a low number of localities falling outside of the model prediction (Table 1). Continuous Boyce index values in final model predictions showed high positive correlation between predicted climate suitability and presence training and test occurrence points. Both breeding (B_{TEST} = 0.921, n = 315) and wintering (B_{TEST} = 0.735, n = 133) range models had high predictive performance at identifying suitable habitat from test data. The whole-range model was consistent with current distribution maps (Figure S1), with high predictive accuracy (B_{TEST} = 0.882, n = 516). Continuous Boyce index plots showed positive correlations and increased predicted to expected ratios as the number of climate suitability bins increased (Fig. S2), demonstrating well-calibrated models.

New areas of climatic suitability

Final predictive maps using all occurrence points identified new areas of high climatic suitability in both breeding and wintering ranges for targeted population surveys (Figs. 1 & 2). Focal regions for potential Saker breeding populations where distribution data are limited include: central and eastern Turkey; western Caucasus from eastern Georgia south across Armenia and Azerbaijan, and into north-western Iran; western Uzbekistan, Kyrgyzstan; Qinghai-Tibetan plateau, and central-northern provinces of Gansu, Inner Mongolia, Ningxia and Xinjiang in China (Fig. 1). Focal regions for potential wintering populations where distribution data are limited include:

western Turkey; the Caucasus into northern Iran, the Levant and north African coast from Egypt to Tunisia; the Ethiopian highlands; central-southern Iran; north-east Afghanistan; the Yemen Highlands; eastern and southern Qinghai-Tibetan plateau, and extending across the provinces of western Sichuan, Gansu and Ningxia in China (Fig. 2).

Model predictions for regions within the Saker's known range, but with limited distribution data or low survey coverage, identified regions to prioritise future population surveys. Visual inspection of clamped projected distributions showed minimal or no prediction outside of the geographic extent of the calibration data. Predictive maps for the Qinghai-Tibet plateau identified highest climatic suitability for breeding Sakers in central and western Tibet, and eastern Qinghai (Fig. 3a). In Tibet, highest climatic suitability encompasses a large region west of Lhasa and extending north into the Nyenchen Tanglha and Tanggula mountain ranges, and further west into Ngari prefecture (Fig. 3a). In Qinghai, highest climatic suitability extends from the Bayan Har Shan range, north-east to the Qilian Shan range bordering Gansu province (Fig. 3b). Predicted wintering Saker climatic suitability covers a broad extent across southern and eastern Tibet, extending into the northern Himalayan foothills bordering Nepal, and further east into western Sichuan province (Fig. 3b). Predictive maps for Mongolia identified highest climatic suitability for breeding Sakers within the central steppe and forest-steppe ecoregions (Fig. S4a). With a large area of climatic suitability buffering this central core area, and to a lesser extent across eastern Mongolia. Predicted wintering climatic suitability identified a similar core region in central Mongolia, with further suitable climatic areas south close to the border with China (Fig. S4b).

Environmental predictors

Mean temperature of the warmest quarter (BIO10) contributed the highest percentage to the breeding range model prediction (37.24%; Table 2), followed by precipitation in the driest month (BIO14; 15.97%), precipitation in the wettest month (BIO13; 13.21%), and temperature seasonality (BIO4; 12.32%), with the remaining variables contributing ≤ 10%. Mean temperature in the wettest quarter (BIO8) contributed the highest percentage to model prediction in the wintering distribution model (53.03%; Table 2), followed by temperature seasonality (BIO4; 23.53%) and precipitation in the wettest month (BIO13; 14.72%), with the remaining variables all contributing < 5%. Jackknife tests of variable importance show that the environmental predictors with the highest percent contribution in both breeding and wintering models (mean temperature warmest quarter and mean temperature wettest quarter respectively), also had highest regularized training gain (Fig. S3). Overall, removing each variable whilst including all others demonstrated that no single predictor contained a significant amount of information not present in other predictors.

Response curves showing the probability of climate suitability relative to environmental variables, demonstrated pronounced climatic thresholds for Saker climate suitability (Fig. 4). In the breeding range model, suitable climatic conditions for mean temperature in the wettest quarter (BIO8) and warmest quarter (BIO10) both ranged from -10°C to ~35°C, peaking at the most suitable climatic conditions at 10°C (Fig. 4a). A similar curve was apparent for BIO8 in the wintering range model,

but with a higher tolerance of mean temperatures in the warmest quarter up to 40°C (Fig. 4b). Mean diurnal temperature range had highest climate suitability at ~15°C, ranging from 5°C to ~17°C in the breeding model, but decreasing rapidly after 5°C in the wintering range model. There were similar curves in precipitation of the driest month for both models, with a peak of most suitable climate conditions at ~12mm, then decreasing rapidly after (Figs. 4a & 4b). However, there were clear differences in suitable climate conditions in all other variables between the two models, demonstrating the importance of the same limiting climatic factors over others dependent on season.

DISCUSSION

Species distribution models are a powerful tool in conservation biogeography, particularly when used for predicting distribution knowledge gaps (Franklin, 2013). Our models for the Saker are consistent with current known distributions (Galushin 2004; Ferguson-Lees & Christie 2005, Karyakin 2012; Kovács *et al.* 2014; Figs. 1, 2, S1), suggesting that bioclimatic factors are a key determinant of Saker distributions at the macro-scale. Model predictive performance was consistently accurate using multiple metrics (see Appendix 3 in Supporting Online Information for details), demonstrating robust SDMs to identify areas with highest climatic suitability. Applying these models in targeted surveys should increase the chance of locating sparsely distributed Saker populations across remote areas, thus addressing knowledge gaps in Saker distribution. Climatic characteristics demonstrated pronounced thresholds for specific environmental conditions, quantifying which variables had the highest contribution to model prediction. Projected distributions in areas with limited

distribution data within the Saker's known range, indicated specific regions with highest climatic suitability, which should be prioritized for future surveys.

Environmental constraints

As an initial broad-scale assessment, only climatic variables were used as environmental predictors (Pearson & Dawson 2003), though incorporating land cover, topography and biotic interactions would likely improve model predictions, especially when used at smaller geographical scales (Elith & Leathwick 2009). Bioclimatic envelope models assume climate is the main limiting factor in species distributions (Pearson & Dawson 2003; Parmesan & Yohe 2003), and our models show how climate may drive Saker distributions at the macro-scale. Sakers generally prefer areas of low rainfall (Kovács *et al.* 2014), consistent with the response curves here, showing a marked decrease in climatic suitability with precipitation < 15mm in the wettest month (Fig. 4). However, from these models, Sakers can tolerate slightly higher levels of precipitation in the warmest quarter (25-30mm) in both breeding and wintering models (Fig. 4), and higher precipitation in the coldest quarter (50-80mm) in the breeding model (Fig. 4a), but with no effect in the wintering model (Fig. 4b).

The differences in response to temperature variables in both breeding and wintering predictions (Fig. 4), is likely due to the broad continental distribution of this migratory species. Wide variation in temperature would be expected between breeding areas mainly in colder, high altitude regions, compared to higher mean temperatures in some warmer, low-level wintering grounds. In the breeding model, Sakers can tolerate a ~15°C mean diurnal temperature range during the breeding season,

reflecting the wide latitudinal gradient in the species' breeding range (Ferguson-Lees & Christie 2005). However, wintering Sakers are less tolerant of diurnal temperature variation with a marked decrease in climatic suitability above 5°C diurnal temperature range (Fig. 4b). This reflects the seasonal movements of many northern breeding Sakers into more southerly wintering regions further south within their breeding range (Dixon *et al.* 2016).

Both mean temperature in the wettest quarter and mean temperature in the warmest quarter peaked at highest climate suitability of 10°C in the breeding model (Fig. 4a), with a similar response to mean temperature in the wettest quarter in the wintering model, but increased tolerance of temperatures up to 40°C for wintering Sakers measured with mean temperature in the warmest quarter (Fig. 4b). This suggests a strong relationship between temperature and precipitation for suitable breeding and wintering conditions for Sakers, reflecting the arid, temperate areas that Sakers mainly prefer across their range (Bondì *et al.* 2013). The response curves for temperature seasonality showed a greater range of tolerance during the breeding season consistent with the dominant high latitude and altitude areas used by breeding Sakers (Fig. 4a), compared to the peak of very low temperature seasonality for wintering Sakers (Fig. 4b) when moving further south in latitude or with decreasing altitude within their range.

Occurrence data and model limitations

Using online bioinformatic databases is increasingly common in biodiversity assessments (Maldonado *et al.* 2015) and is a cost-effective and convenient method.

However, using such records requires caution and careful appraisal before implementation into predictive models (Beck *et al.* 2013). The robust approach implemented here to clean datasets and remove spatial biases, whilst maintaining suitable sample sizes, shows the utility of using online database records for SDMs (See Appendix 4 in Supporting Online Information). Reliability of observer experience and accuracy of biodiversity database records can be difficult to quantify, as records are often compiled from varied sources with no systematic methodology. However, in these models, Saker occurrence records were sourced from reliable data sources such as raptor biologists, atlases, museum specimens, and citizen science data (i.e. eBird; Sullivan *et al.* 2009) and subsequently archived within the GBIF database.

Using presence-only data is now commonplace in biogeography, and its use should be encouraged when robust datasets can be generated from citizen-science data (Beck *et al.* 2013; Fournier *et al.* 2017a,b). For example, a comparison of model predictive performance from both satellite tracking and eBird data-derived SDMs for the Band-tailed Pigeon *Patagioenas fasciata*, demonstrated that an SDM using presence-only occurrence data was just as effective at predicting species distributions as one using satellite tracking data (Coxen *et al.* 2017). However, we recognise there are important methodological issues regarding spatial bias in presence-only occurrences from digital databases (Beck *et al.* 2014). Citizen-science data collection is often biased to areas near to human habitation or more accessible sites, resulting in gaps in sampling effort from hard-to-survey areas (Kramer-Schadt *et al.* 2013; Boria *et al.* 2014; Fourcade *et al.* 2014). However, as our occurrence dataset was compiled from both professional and volunteer surveyors, the sampling

area covers much of the Saker's known range. Inevitably there will always be variation in survey effort for species with such large ranges across remote areas, such as the Saker, and this is where we argue for using correlative SDMs to help fill in the gaps and direct future survey effort.

Predicted distributions for the Qinghai-Tibet plateau

China is an important range country for Sakers, with significant distribution knowledge gaps (Collar *et al.* 2013). The predicted distributions to a key region for breeding and wintering Sakers in the Qinghai-Tibet plateau is an initial step to improved understanding of Saker distribution across this vast country. Climatic suitability predictions for the Saker in this region are consistent with previous observations that Sakers mainly breed in the east and south of the plateau (Vaurie 1972; Potapov & Ming 2004). Cui *et al.* (2008) also report that Sakers are common in north-east Qinghai during the breeding season, and King & Tai (1991) state that Sakers are common in north-west Sichuan. However, even with this information surveying remote, inhospitable terrain, combined with large search areas, can often be a key factor constraining successful population surveys (Menon *et al.* 2010).

The predicted distributions into the Qinghai-Tibet plateau and Mongolia (Figure S4; Appendix 5 in supporting Online Information) illustrate this problem of scale, and how SDMs can identify the most suitable areas reducing the searchable area. Our models further support previous assessments for the importance of the Qinghai-Tibet plateau as an area of potentially high Saker population density (Dixon *et al.* 2015). Exploratory surveys in this vast biogeographic region (2.5 million km²), suggest the

plateau may hold Saker densities higher than for China as a whole, and further systematic surveys have been recommended (Dixon *et al.* 2015). Our models produced accurate distribution maps, and provided focused direction for future population surveys in remote regions of the Sakers' range. We recommend that population surveys for the Qinghai-Tibet plateau be prioritized to determine its potential global importance for breeding and wintering Saker populations.

Predictive spatial models are currently in high demand for conservation planning and to improve understanding of species distributions (Rodríguez *et al.* 2007; Wood *et al.* 2017). Broad-scale climatic assessments provide an initial step towards smaller-scale projects focused on planning reintroductions, mitigating development conflicts, defining protected areas and evaluating the impact of climate change on species distributions (Peterson *et al.* 2011). Here, we show a further application by identifying areas with highest climatic suitability to target population surveys for an endangered species, and potentially accelerating the discovery of new populations (Guisan *et al.* 2006a). Species geographical distributions are important units of information used to assess range size, population numbers, target management measures and determine conservation status (Elith *et al.* 2006). Species that live at low densities, across large ranges in remote regions, such as the Saker, are often the most difficult to assess (Collar *et al.* 2013). By using a presence-background, climate-based SDM, key knowledge gaps in Saker distribution can be addressed. Though there are multiple gaps in our understanding of Saker biogeography and ecology (Collar *et al.* 2013; Kovács *et al.* 2014), SDMs can start to address fundamental biogeographical questions, leading to improved understanding of Saker distribution, and informing effective conservation management.

ACKNOWLEDGEMENTS

REFERENCES

Aiello-Lammens, M. E., Boria, R. A., Radosavljevic, A., Vilela, B., & Anderson, R. P. (2015). spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography*, 38(5), 541-545.

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(6), 1223-1232.

Austin, G. E., Thomas, C. J., Houston, D. C., & Thompson, D. B. (1996). Predicting the spatial distribution of buzzard *Buteo buteo* nesting areas using a Geographical Information System and remote sensing. *Journal of Applied Ecology*, 33, 1541-1550.

Aryal, A., Shrestha, U. B., Ji, W., Ale, S. B., Shrestha, S., Ingty, T., ... & Raubenheimer, D. (2016). Predicting the distributions of predator (snow leopard) and prey (blue sheep) under climate change in the Himalaya. *Ecology and evolution*, 6(12), 4065-4075.

Bagyura, J., Szitta, T., Haraszthy, L., Demeter, I., Sándor, I., Dudás, M., ... & Viszló, L. (2004). Population trend of the Saker Falcon *Falco cherrug* in Hungary between 1980 and 2002. In R. D. Chancellor, & B. U. Meyburg (Eds.) *Raptors Worldwide* (pp. 663-772). Budapest, Hungary: Birdlife Hungary.

Beck, J., Ballesteros-Mejia, L., Nagel, P. & Kitching, I. J. (2013). Online solutions and the 'Wallacean shortfall': what does GBIF contribute to our knowledge of species' ranges? *Diversity and Distributions*, 19(8), 1043-1050.

Beck, J., Böller, M., Erhardt, A., & Schwanghart, W. (2014). Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. *Ecological Informatics*, 19, 10-15.

BirdLife International. (2017). *Falco cherrug*. The IUCN Red List of Threatened Species 2017: e.T22696495A110525916.

http://www.iucnredlist.org/details/22696495/0. Accessed 20 May 2017.

Booms, T. L., Huettmann, F. & Schempf, P. F. (2010). Gyrfalcon nest distribution in Alaska based on a predictive GIS model. *Polar Biology,* 33, 347-358.

Bondì S., Prommer M. & Sarà M. (2013). Habitat preferences and diet of the Saker Falcon *Falco cherrug* overwintering in Sicily. XVII Italian Ornithological Congress, 142-143 (abstract).

Bondì, S., Prommer, M, & Sarà, M. (2014). The diet of Saker Falcon *Falco cherrug* overwintering in the Mediterranean (Sicily). *Avocetta*, 38(2), 53-58.

Boria, R. A., Olson, L. E., Goodman, S. M. & Anderson, R. P. (2014). Spatial filtering to reduce sampling bias can improve the performance of ecological niche models. *Ecological Modelling*, 275, 73-77.

Boyce, M. S., Vernier, P. R., Nielsen, S. E. & Schmiegelow, F. K. (2002). Evaluating resource selection functions. *Ecological Modelling*, 157(2), 281-300.

Bradie, J. & Leung, B. (2017). A quantitative synthesis of the importance of variables used in MaxEnt species distribution models. *Journal of Biogeography*, 44(6),1344-1361.

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research*, 33(2), 261-304.

Chancellor, R.D. & Meyburg, B.-U. (2000). *Raptors at Risk.* World Working Group on Birds of Prey and Owls. Hancock House.

Collar, N., Symes, A., Dixon, A., Prommer, M., Shobrak, M. & Sielicki, J. (2013) A review of key knowledge gaps concerning the biology and ecology of the Saker Falcon *Falco cherrug* and the socio-economic factors affecting its use. In N. P. Williams, C. Galbraith & A. Kovács (Eds.) *Compilation Report on WorkPlan Objectives 4 – 8, including a modelling framework for sustainable use of the Saker Falcon Falco cherrug.* (pp. 54-83). Abu Dhabi: UNEP/CMS Raptors MoU Coordinating Unit, Saker Falcon Task Force.

Coxen, C. L., Frey, J. K., Carleton, S. A. & Collins, D. P. (2017). Species distribution models for a migratory bird based on citizen science and satellite tracking data. *Global Ecology and Conservation*, 11, 298-311.

Cui, Q., Su, J. & Jiang, Z. (2008). Summer diet of two sympatric species of raptors upland buzzard (*Buteo hemilasius*) and Eurasian eagle owl (*Bubo bubo*) in Alpine meadow: Problem of coexistence. *Polish Journal of Ecol*ogy, 56(1), 173-179.

de Carvalho, D. L., Sousa-Neves, T., Cerqueira, P. V., Gonsioroski, G., Silva, S. M., Silva, D. P. & Santos, M. P. D. (2017). Delimiting priority areas for the conservation of endemic and threatened Neotropical birds using a niche-based gap analysis. *PloS one*, 12(2), e0171838. DOI: 10.371/journal.pone.0171838.

Di Cola, V., Broennimann, O., Petitpierre, B., Breiner, F. T., D'Amen, M., Randin, C., ... & Pellissier, L. (2017). ecospat: An R package to support spatial analyses and modeling of species niches and distributions. *Ecography*, 40(6), 774-787.

Dixon, A. (2005). Falcon population estimates: how necessary and accurate are they? *Falco,* 25, 5-8.

Dixon, A. (2009). Saker Falcon breeding population estimates. Part 2: Asia. *Falco,* 33, 4-10.

Dixon, A. (2016). Commodification of the Saker Falcon *Falco cherrug*: Conservation Problem or Opportunity? In F. M. Angelici (Ed.) *Problematic Wildlife* (pp. 69-89). Switzerland: Springer International Publishing.

Dixon, A., Li, X., Rahman, M. L., Batbayar, N. & Zhan, X. (2016). Characteristics of home range areas used by Saker Falcons (*Falco cherrug*) wintering on the Qinghai-Tibetan Plateau. *Bird Conservation International*, pp.1-12.

Dixon, A., Ming, M. & Batbayar, N. (2015). Importance of the Qinghai-Tibetan plateau for the endangered Saker Falcon *Falco cherrug. Forktail,* 31, 37-42.

Dixon, A., Ragyov, D., Ayas, Z., Deli, M., Demerdzhiev, D., Angelov, I., Kmetova, E. & Nedyalkov, N. (2009). Population status of breeding Saker Falcons (*Falco cherrug*) in Turkey. *Avian Biology Research*, 2(4), 213-220.

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., ... Münkemüller, T. (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27-46.

Duan, R. Y., Kong, X. Q., Huang, M. Y., Fan, W. Y. & Wang, Z. G. (2014). The predictive performance and stability of six species distribution models. *PloS one*, 9(11), e112764. DOI: 10.1371/journal.pone.0112764.

Ellis, D. H., Kitowski, I. & Roundy, T. B. (2011). Nomadism in large falcons: Lessons from the Saker Falcon in Mongolia. In R. T. Watson, T. J. Cade, M. Fuller, G. Hunt, & E. Potapov (Eds.). *Gyrfalcons and Ptarmigan in a changing world, Volume II.* (pp. 291-306) Boise, Idaho, USA: The Peregrine Fund.

Elith, J., Graham, C. H., Anderson, R. P., Dudík, M., Ferrier, S., Guisan, A., … Zimmerman, E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29, 129–151. Elith, J., Kearney, M., & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in ecology and evolution*, 1(4), 330-342.

Elith, J. & Leathwick, J. R. (2009a). Species distribution models: ecological explanation and prediction across space and time. *Annual review of ecology, evolution, and systematics*, 40, 677-697.

Elith, J. & Leathwick, J. R. (2009b). The contribution of species distribution modelling to conservation prioritization. In A. Moilanen, K. A. Wilson & H. P. Possingham (Eds.) *Spatial Conservation Prioritization: Quantitative Methods and & Computational Tools.* (pp. 70-93). Oxford, UK: Oxford University Press.

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E. & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. *Diversity and Distributions*, 17, 43–57.

Engler, R., Guisan, A. & Rechsteiner, L. (2004). An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *Journal of Applied Ecology,* 41, 263–274.

Fehérvári, P., Solt, S., Palatitz, P., Barna, K., Ágoston, A., Gergely, J., Nagy, A., Nagy, K. and Harnos, A. (2012). Allocating active conservation measures using

species distribution models: a case study of red-footed falcon breeding site management in the Carpathian Basin. *Animal Conservation*, 15(6), 648-657.

Ferguson-Lees, J. & Christie, D. A. (2005). *Raptors of the World.* London: Christopher Helm.

Fielding, A. H. & Bell, J. F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, 24, 38–49.

Finlayson, C. (2011). *Avian Survivors. The History and Biogeography of Palearctic Birds.* London, UK: T. & A. D. Poyser.

Fourcade, Y., Engler, J. O., Rödder, D., & Secondi, J. (2014). Mapping species distributions with MAXENT using a geographically biased sample of presence data: a performance assessment of methods for correcting sampling bias. *PloS one*, 9(5), e97122. DOI: 10.1371/journal.pone.0097122.

Fournier, A., Drake, K. L., & Tozer, D. C. (2017a). Using citizen science monitoring data in species distribution models to inform isotopic assignment of migratory connectivity in wetland birds. *Journal of Avian Biology*. DOI: 10.1111/jav.01273.

Fournier, A., Sullivan, A. R., Bump, J. K., Perkins, M., Shieldcastle, M. C., & King, S. L. (2017b). Combining citizen science species distribution models and stable isotopes reveals migratory connectivity in the secretive Virginia rail. *Journal of Applied Ecology*, 54(2), 618-627.

Franklin, J. (2009). *Mapping Species Distributions.* Cambridge, UK: Cambridge University Press.

Franklin, J. (2013). Species distribution models in conservation biogeography: developments and challenges. *Diversity and Distributions,* 19, 1217-1223.

Galante, P.J., Alade, B., Muscarella, R., Jansa, S.A., Goodman, S.M. & Anderson, R.P. (2018). The challenge of modeling niches and distributions for data-poor species: a comprehensive approach to model complexity. *Ecography*, 41(5), 726-736.

Galushin, V. M. (2004). Status of Saker in Russia and Eastern Europe. *Falco,* 24, 3-8.

Gerstner, B. E., Kass, J. M., Kays, R., Helgen, K. M., & Anderson, R. P. (2018). Revised distributional estimates for the recently discovered olinguito (Bassaricyon neblina), with comments on natural and taxonomic history. *Journal of Mammalogy*, 99(2), 321-332.Gibson, L., Barrett, B. & Burbidge, A. (2007). Dealing with uncertain absences in habitat modelling: a case study of a rare ground-dwelling parrot. *Diversity and Distributions*, 13(6), 704-713.

Global Biodiversity Information Facility. (2017a). GBIF Occurrence Download *Falco cherrug.* Retrieved from https://www.gbif.org/ DOI: 10.15468/dl.thde9w. Accessed 8 March 2017.

Global Biodiversity Information Facility. (2017b). GBIF Occurrence Download *Falco cherrug.* Retrieved from https://www.gbif.org/ DOI: 10.15468/dl.pxrn6i. Accessed 28 December 2017.

Guisan, A., Broennimann, O., Engler, R., Vust, M., Yoccoz, N. G., Lehmann, A. & Zimmermann, N. E. (2006a). Using niche-based models to improve the sampling of rare species. *Conservation Biology*, 20(2), 501-511.

Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J.M.C., Aspinall, R. & Hastie, T. (2006b). Making better biogeographical predictions of species' distributions. *Journal of Applied Ecology*, 43(3), 386-392.

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology letters*, 8(9), 993-1009.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E. & Tatham, R. L. (2006). *Multivariate Data Analysis.* Upper Saddle River, NJ, USA: Pearson Prentice Hall.

Hanley, J. A. & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristics curve. *Radiology*, 143, 29–36.

Hastie, T., Tibshirani, R., Friedman, J. & Franklin, J. (2005). The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27, 83–85. DOI: 10.1016/j.neunet.2009.04.005 PMID: 19443179.

Hijmans, R. J. (2012). Cross-validation of species distribution models: removing spatial sorting bias and calibration with a null model. *Ecology*, 93(3), 679-688.

Hijmans, R. J. (2016). raster: Geographic Data Analysis and Modeling. R package version 2.5-8. https://CRAN.R-project.org/package=raster.

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. (2005). Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology,* 25, 1965-1978. http://www.worldclim.org.bioclim

Hijmans, R. J., Phillips, S., Leathwick, J. & Elith, J. (2017). dismo: Species Distribution Modeling. R package version 1.1-4. https://CRAN.R-project.org/package=dismo.

Hirzel, A. H., Le Lay, G., Helfer, V., Randin, C. & Guisan, A. (2006). Evaluating the ability of habitat suitability models to predict species presences. *Ecological Modelling*, 199(2), 142-152.

Huberty, C. J. (1994). *Applied Discriminant Analysis*. New York, NY, USA: Wiley Interscience.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence pp. 1137-1143. San Mateo, CA.

Kovács, A., Williams, N. P., & Galbraith, C. A. (2014). Saker Falcon *Falco cherrug* Global Action Plan (SakerGAP), including a management and monitoring system, to conserve the species. Raptors MOU technical publication no. 2. CMS technical series no. 31. Coordinating Unit—CMS Raptors MOU, Abu Dhabi.

Kramer-Schadt, S., Niedballa, J., Pilgrim, J. D., Schröder, B., Lindenborn, J., Reinfelder, V., ... & Cheyne, S. M. (2013). The importance of correcting for sampling bias in MaxEnt species distribution models. *Diversity and Distributions*, 19(11), 1366-1379.

Lawler, J. J., Wiersma, Y. F. & Huettman, F. (2011). Using species distribution models for conservation planning and ecological forecasting. In C. A. Drew, Y. F. Wiersma & F. Huettmann, F. (Eds.). *Predictive Species and Habitat Modeling in Landscape Ecology.* (pp. 271-290). New York, NY, USA: Springer.

Liu, C., White, M. & Newell, G. (2017). Detecting outliers in species distribution data. *Journal of Biogeography*, 2017; 00: 1-13. https://doi.org/10.1111/jbi.13122.

Maldonado, C., Molina, C. I., Zizka, A., Persson, C., Taylor, C. M., Albán, J., ... & Antonelli, A. (2015). Estimating species diversity and distribution in the era of Big Data: to what extent can we trust public databases? *Global Ecology and Biogeography*, 24(8), 973-984.

Menon, S., Choudhury, B. I., Khan, M. L., & Peterson, A. T. (2010). Ecological niche modeling and local knowledge predict new populations of Gymnocladus

assamicus a critically endangered tree species. *Endangered Species Research*, 11(2), 175-181.

Merow, C., Smith, M. J. & Silander, J. A. (2013). A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and settings matter. *Ecography*, 36(10), 1058-1069.

Miller, J. (2010). Species Distribution Modelling. *Geography Compass,* 4(6), 490-509.

Muscarella, R., Galante, P. J., Soley-Guardia, M., Boria, R. A., Kass, J. M., Uriarte, M., & Anderson, R. P. (2014). ENMeval: an R package for conducting spatially independent evaluations and estimating optimal model complexity for Maxent ecological niche models. *Methods in Ecology and Evolution*, 5(11), 1198-1205.

Naimi, B., Hamm, N.A., Groen, T.A., Skidmore, A.K. & Toxopeus, A.G. (2014). Where is positional uncertainty a problem for species distribution modelling? *Ecography*, 37(2), 191-203.

Nedyalkov, N., Levin, A., Dixon, A. & Boev, Z. (2014). Diet of Saker Falcon (*Falco cherrug*) and Eastern Imperial Eagle (*Aquila heliaca*) from Central Kazakhstan. *Ecologia Balkanica*, 6(1), 25-30.

Newton, I. (1979). *Population Ecology of Raptors.* London, UK: T. & A.D. Poyser.

Nicholls, A. O. (1989). How to make biological field surveys go further with generalized linear models. *Biological Conservation*, 50, 51–75.

Nittinger, F., Gamauf, A., Pinsker, W., Wink, M. & Haring, E. (2007). Phylogeography and population structure of the saker falcon (*Falco cherrug*) and the influence of hybridization: mitochondrial and microsatellite data. *Molecular Ecology*, 16(7), 1497-1517.

Osborne, P. E. (2005). Using GIS, remote sensing and modern statistics to study steppe birds at large spatial scales: A short review essay. In G. Bota, M.B. Morales, S. Mañosa & J. Camprodon (Eds.), *Ecology and conservation of steppe-land birds.* (pp. 169-184). Barcelona, Spain: Lynx Edicions & Centre Tecnològic Forestal de Cataluyna.

Parmesan, C. & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421(6918), 37.

Parviainen, M., Marmion, M., Luoto, M., Thuiller, W. & Heikkinen, R. K. (2009). Using summed individual species models and state-of-the-art modelling techniques to identify threatened plant species hotspots. *Biological Conservation,* 142, 2501–2509.

Pearce, J. L. & Boyce, M. S. (2006). Modelling distribution and abundance with presence-only data. *Journal of Applied Ecology*, 43, 405–412.

Pearson, R. G. & Dawson, T. P. (2003). Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global ecology and biogeography*, 12(5), 361-371.

Pearson, R. G., Raxworthy, C. J., Nakamura, M. & Peterson, A. T. (2007). Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, 34(1), 102-117.

Peterson, A. T. (2001). Predicting species' geographic distributions based on ecological niche modeling. *The Condor,* 103(3), 599–605.

Peterson, A. T., Soberón, J., Pearson, R. G., Anderson, R. P., Martinez-Meyer, E., Nakamura, M & Araújo, M. B. (2011). *Ecological Niches and Geographic*

Distributions. Monographs in Population Biology 49. Princeton, NJ, USA: Princeton University Press.

Phillips, S. J., Anderson, R. P. & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, 190, 231-259.

Phillips, S. J. & Dudík, M. (2008). Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, 31, 161–175.

Potapov, E., Fox, N., Shagdarsuren, O., Sumya, D. & Gombobaatar, S. (2000). Home ranges of Saker Falcons in Mongolia. *Falco*, 10, 10-11.

Potapov, E. & Ming, M. (2004). The highlander: the highest breeding Saker in the world. *Falco,* 23, 10-12.

Radosavljevic, A. & Anderson, R. P. (2014). Making better Maxent models of species distributions: complexity, overfitting and evaluation. *Journal of Biogeography*, 41(4), 629-643.

Ragyov, D., Biserkov, V., Gradev, G., Ivanov, I., Stoynov, E., Stoyanov, G., … Dixon, A. (2014). Past and present status of the Saker Falcon, *Falco cherrug* (Aves: Falconidae) in Bulgaria. *Acta Zoologica Bulgarica*, 66, 299-308.

Ramírez-Albores, J. E., Bustamante, R. O. & Badano, E. I. (2016). Improved predictions of the geographic distribution of invasive plants using climatic niche models. *PloS one*, 11(5), e0156029. DOI: 10.1371/journal.pone.0156029.

Raxworthy, C. J., Martinez-Meyer, E., Horning, N., Nussbaum, R. A., Schneider, G. E., Ortega-Huerta, M. A. & Peterson, A. T. (2003). Predicting distributions of known and unknown reptile species in Madagascar. *Nature*, 426, 837-841.

R Core Team. (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Reineking, B., Leutner, B. & Wegmann, M. (2016). Modelling species distributions. In M. Wegmann, B. Leutner & S. Dech (Eds.). *Remote sensing and GIS for ecologists: using open source software.* (pp. 258-293). Exeter, UK: Pelagic Publishing Ltd.

Reside, A.E., VanDerWal, J.J., Kutt, A.S. & Perkins, G.C. (2010). Weather, not climate, defines distributions of vagile bird species. *PLoS One*, 5(10), p.e13569.

Rhoden, C. M., Peterman, W. E., & Taylor, C. A. (2017). Maxent-directed field surveys identify new populations of narrowly endemic habitat specialists. *PeerJ*, 5, e3632. DOI: 10.7717/peerj.3632.

Rodríguez, J. P., Brotons, L., Bustamante, J. & Seoane, J. (2007). The application of predictive modelling of species distribution to biodiversity conservation. *Diversity and Distributions*, 13(3), 243-251.

Stockwell, D. R. (2006). Improving ecological niche models by data mining large environmental datasets for surrogate models. *Ecological Modelling*, 192, 188-196.

Stockwell, D. R. & Peterson, A. T. (2002). Effects of sample size on accuracy of species distribution models. *Ecological Modelling*, 148, 1–13.

Stohlgren, T. J., Jarnevich, C. S., Esaias, W. E., & Morisette, J. T. (2011). Bounding species distribution models. *Current Zoology*, 57(5), 642-647.

Sullivan, B. L., Wood, C. L., Iliff, M. J., Bonney, R. E., Fink, D. & Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. *Biological Conservation*, 142(10), 2282-2292.

Vaurie, C. (1972). *Tibet and its birds.* London, H. F. & G. Witherby.

Warren, D. L. & Seifert, S. N. (2011) Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection

criteria. *Ecological Applications* 21(2), 335-342.Watson, M. & Clarke, R. (2000). Saker Falcon diet. *British Birds*, 93, 136-143.

Wood, K. A., Stillman, R. A., & Hilton, G. M. (2017). Conservation in a changing world needs predictive models. *Animal Conservation*, 21(2), 87-88.

Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H. & Guisan, A. (2008).

  Effects of sample size on the performance of species distribution models. *Diversity and Distributions*, 14(5), 763-773.

Wu, Y. (2011). The diet of Saker Falcon (*Falco cherrug*) on the eastern fringe of Gurban Tunggut Desert, China. In *Remote Sensing, Environment and Transportation Engineering (RSETE), 2011 International Conference* (pp. 8647-8649). IEEE.

Wu, X. B. & Smeins, F. E. (2000). Multiple-scale habitat modelling approach for rare plant conservation. *Landscape and Urban Planning,* 51, 11-28.

Zeng, Q., Zhang, Y., Sun, G., Duo, H., Wen, L. & Lei, G. (2015). Using species distribution model to estimate the wintering population size of the endangered scaly-sided Merganser in China. *PloS one*, 10(2), e0117307.

Data Accessibility Statement

Due to the continuing threats facing Saker Falcons, location data are only available direct from the corresponding author. All GIS data and models are available from The Peregrine Fund Global Raptor Impact Network database.

Conflict of Interest

The authors have no conflicts of interest to declare.

TABLES

Table 1. Evaluation test statistics from ten-fold cross validation for three species distribution models for the Saker Falcon. AIC_c = Akaike Information Criterion corrected for small sample sizes, ΔAIC_c = the difference in AIC between the best model and a given model. AUC_{TRAIN}, AUC_{TEST} and AUC_{DIFF} are three different metrics representing Area Under the Curve (AUC) of the Receiver Operating Characteristic plot (ROC). MOR = minimum test omission rate & 10%OR = 10 percent test omission rate.

Distribution model	AIC_c	ΔAIC_c	AUC_{TRAIN}	AUC_{TEST}	AUC_{DIFF}	Test omission rates	
						MOR	10%OR
Breeding	8904.143	0.0	0.813	0.791	0.022	0.003	0.130
Wintering	3845.936	0.0	0.837	0.801	0.036	0.023	0.143
Whole-range	14746.47	0.0	0.851	0.840	0.011	0.002	0.120

Table 2. Percent contribution for environmental predictors in breeding and wintering distribution models for the Saker Falcon. Bioclimatic variables represent annual trends and extremes and limiting environmental factors from the period 1970-2000. Sourced from WorldClim (version 2, http://www.worldclim.org/bioclim/). All temperature values in °C and precipitation values in mm.

Bioclim code	Description	Breeding	Wintering
BIO2	Mean diurnal range[1]	5.92	4.97
BIO4	Temperature Seasonality[2]	12.32	23.53
BIO8	Mean temperature of wettest quarter	10.13	53.03
BIO10	Mean temperature of warmest quarter	37.24	1.44
BIO13	Precipitation of wettest month	13.21	14.72
BIO14	Precipitation driest month	15.97	0.98
BIO15	Precipitation seasonality[3]	3.60	1.34
BIO19	Precipitation of coldest quarter	1.72	0.00

[1] Mean of monthly temperature (maximum temperature – minimum temperature)

[2] Standard deviation *100

[3] Coefficient of variation. Standard deviation of monthly precipitation estimates expressed as a percentage of the mean estimate

FIGURE LEGENDS

Figure 1. Predicted breeding distribution model for the Saker Falcon. Map shows continuous logistic prediction with hotter red areas (values closer to 1) having high Saker climate suitability. Black filled circles define known Saker occurrence points. Boyce index $B_{TEST} = 0.921$ using 10% test data.

Figure 2. Predicted wintering distribution model for the Saker Falcon. Map shows continuous logistic prediction with hotter red areas (values closer to 1) having high Saker climate suitability. Black filled circles define known Saker occurrence points. Boyce index $B_{TEST} = 0.735$ using 10% test data.

Figure 3. Predicted distributions for the Saker Falcon projected into the Qinghai-Tibet plateau, China. Map **(a)** shows predicted breeding distribution, and **(b)** predicted wintering distribution. Hotter red areas (values closer to 1) have higher Saker climate suitability. Black filled circles show Saker occurrence points. Bold lines define Chinese border and administrative regions.

Figure 4. Response curves for all climatic variables used as predictors within distribution models for the Saker Falcon: **(a)** breeding range distribution, and **(b)** wintering range distribution. All temperature values in °C and precipitation values in mm.