

RESEARCH

Open Access



Species distribution modeling in regions of high need and limited data: waterfowl of China

Diann J. Prosser^{1,2*}, Changqing Ding³, R. Michael Erwin^{1,4}, Taej Mundkur⁵, Jeffery D. Sullivan⁶ and Erle C. Ellis⁷

Abstract

Background: A number of conservation and societal issues require understanding how species are distributed on the landscape, yet ecologists are often faced with a lack of data to develop models at the resolution and extent desired, resulting in inefficient use of conservation resources. Such a situation presented itself in our attempt to develop waterfowl distribution models as part of a multi-disciplinary team targeting the control of the highly pathogenic H5N1 avian influenza virus in China.

Methods: Faced with limited data, we built species distribution models using a habitat suitability approach for China's breeding and non-breeding (hereafter, wintering) waterfowl. An extensive review of the literature was used to determine model parameters for habitat modeling. Habitat relationships were implemented in GIS using land cover covariates. Wintering models were validated using waterfowl census data, while breeding models, though developed for many species, were only validated for the one species with sufficient telemetry data available.

Results: We developed suitability models for 42 waterfowl species (30 breeding and 39 wintering) at 1 km resolution for the extent of China, along with cumulative and genus level species richness maps. Breeding season models showed highest waterfowl suitability in wetlands of the high-elevation west-central plateau and northeastern China. Wintering waterfowl suitability was highest in the lowland regions of southeastern China. Validation measures indicated strong performance in predicting species presence. Comparing our model outputs to China's protected areas indicated that breeding habitat was generally better covered than wintering habitat, and identified locations for which additional research and protection should be prioritized.

Conclusions: These suitability models are the first available for many of China's waterfowl species, and have direct utility to conservation and habitat planning and prioritizing management of critically important areas, providing an example of how this approach may aid others faced with the challenge of addressing conservation issues with little data to inform decision making.

Keywords: Anatidae, Avian influenza, China, Habitat suitability, H5N1, Spatial analysis, Species distribution models, Waterfowl

Introduction

Environmental managers face numerous priority needs, ranging from the protection of critical habitat and mitigating effects of environmental stressors on wildlife to balancing the needs of our natural systems with the

impacts of an ever expanding anthropogenic footprint. Yet these challenges are linked by a very simple limiting factor: the need for reliable information regarding when and how the species we seek to manage and preserve are distributed across the landscape (Franklin and Miller 2010). While scientists have long been attempting to understand the relationships between wildlife populations and their spatio-temporal environment, modern technology has allowed for rapid advancements in using

*Correspondence: dprosser@usgs.gov

¹ Patuxent Wildlife Research Center, U.S. Geological Survey, Beltsville, MD 20705, USA

Full list of author information is available at the end of the article

modeling approaches to quantify and predict animal space use across time. Assessments based on species distribution maps to identify areas of importance during a species' main life cycle stages have been used globally and regionally (e.g. Williamson et al. 2013). Though the rapidly expanding field of species distribution modeling (SDM) offers new approaches for improved model development including Bayesian statistics, maximum entropy, artificial neural networks, genetic algorithms, and other machine learning techniques, most approaches require robust datasets for model development (Segurado and Araujo 2004; Franklin and Miller 2010; Guillerá-Arroita et al. 2015). Fine-grained distribution data may be locally available for some species; however, data are rarely available across large extents due to high costs of production. Some regions such as North America and parts of Europe have long-term monitoring efforts for target species from which consistent, quality data can be extracted (Root 1988; Sauer et al. 2003); whereas other, often developing regions, rarely have broad-scale programs despite having rich biological resources (Grenyer et al. 2006; Martin et al. 2012). In some cases, a lack of sufficient data may preempt the use of more advanced methods to address urgent societal needs, and here a balance of using the best approaches possible with available data and resources with clear reporting of shortcomings is warranted.

The frequent mismatch between the need for information regarding a species' spatio-temporal distribution and the data needed to make informed decisions is exemplified by avian influenza response efforts and wider conservation action in Asia. Wild waterfowl and shorebirds (orders Anseriformes and Charadriiformes) are known reservoirs for low-pathogenic forms of avian influenza viruses (LPAIV), which have the potential to mutate into lethal forms following entry into domestic poultry populations (Alexander 2007). Outbreaks of highly pathogenic avian influenza viruses (HPAIV) such as the Asian strain of H5N1 (Xu et al. 1999) have caused considerable damage to the health and economy of more than 60 countries from Asia to Africa since emergence in 1996 (OIE 2017), and the loss of thousands of waterfowl (Liu et al. 2005; OIE 2017). Waterfowl in the family Anatidae (ducks, geese, and swans; hereafter waterfowl) are of particular importance due to their migratory behavior, high abundance, propensity to congregate in high densities, and increased exposure to farmed ducks which can act as silent reservoirs of HPAI (Muzaffar et al. 2010). Due to the risk to both human and wildlife populations, appropriate understanding of the role wild birds' play in the epidemiology of these viruses is critical. Concurrent with the risk of avian influenza outbreaks leading to wild bird mortalities, the waterfowl of China face significant population-wide challenges from climate and land use

change (Yu et al. 2017). Unfortunately, few studies have incorporated wild birds into geographically explicit models (Gilbert and Pfeiffer 2012), largely because obtaining spatial inputs for these populations is difficult. Datasets within Asia have been driven by range maps (Williamson et al. 2013) or limited to small regions or few species (Zeng et al. 2015; Dai et al. 2016; Dronova et al. 2016). Therefore, a technique that makes the most of available data to address pressing conservation and human health concerns is of the utmost importance.

The objective of this project was to develop 1 km resolution (1 km × 1 km) binary grid maps of habitat suitability for all species of waterfowl known to spend the winter or breed in China. These base models would not only serve as inputs for modeling transmission risk of circulating avian influenza viruses at the poultry–waterfowl interface (Prosser et al. 2013), but also provide managers with data relevant to an array of conservation needs. As comprehensive nationwide waterfowl survey data were not available to apply newer techniques of SDM, we took the initial step of mapping potential distributions (habitat suitability) by linking habitat relationships and environmental predictors in a geographic information system (GIS). We validated the suitability models with available census or telemetry data and created a composite suitability map across all species for the breeding and wintering seasons. Here we present spatially explicit habitat suitability models for China's 30 breeding and 39 wintering waterfowl species. Although previous works with widely different methodologies have created SDM's for a few targeted species within this region (Moriguchi et al. 2013; Zeng et al. 2015; Dai et al. 2016), our models represent the first comprehensive set of models spanning the entirety of China. We hope that the assessment provided here will stimulate efforts for other priority areas faced with similar data challenges, and demonstrate that datasets built for a specific objective can have utility for a wide array of conservation oriented issues, particularly when the approach and assumptions are made clear.

Methods

Waterfowl data

Of the 51 waterfowl species listed in MacKinnon and Phillipps (2000) 42 were reported as utilizing China within either their breeding or wintering ranges. Thus, we conducted a review of the English and Chinese literature for China's 42 waterfowl species (Table 1) following the taxonomy provided by MacKinnon and Phillipps (2000). References included peer-reviewed journal articles, technical reports, as well as unpublished surveys from nature reserves, non-governmental organizations, etc. This literature review spanned several decades, with the first included text published in 1985. The database

Table 1 Thirty breeding and 39 wintering Anatidae waterfowl species of China

Common name	Scientific name	Models	Status ^a	% protected breeding ^b	% protected wintering ^c
Lesser Whistling Duck	<i>Dendrocygna javanica</i>	BW	LC	7.5	4.4
Mute Swan	<i>Cygnus olor</i>	B	LC	26.5	
Whooper Swan	<i>Cygnus cygnus</i>	BW	LC	18.6	5.8
Tundra Swan	<i>Cygnus columbianus</i>	W	LC		4.4
Swan Goose	<i>Anser cygnoides</i>	BW	V	18.8	13.9
Bean Goose	<i>Anser fabalis</i>	W	LC		6.3
Greater White-fronted Goose	<i>Anser albifrons</i>	W	LC		5.0
Lesser White-fronted Goose	<i>Anser erythropus</i>	W	V		5.6
Greylag Goose	<i>Anser anser</i>	BW	LC	17.3	6.0
Bar-headed Goose	<i>Anser indicus</i>	BW	LC	41.8	7.4
Snow Goose	<i>Anser caerulescens</i>	W	LC		2.0
Brent Goose	<i>Branta bernicla</i>	W	LC		5.3
Ruddy Shelduck	<i>Tadorna ferruginea</i>	BW	LC	11.7	16.6
Common Shelduck	<i>Tadorna tadorna</i>	BW	LC	18.0	7.3
Cotton Pygmy Goose	<i>Nettapus coromandelianus</i>	BW	LC	4.7	4.2
Mandarin Duck	<i>Aix galericulata</i>	BW	LC	14.9	4.9
Gadwall	<i>Anas strepera</i>	BW	LC	18.5	8.9
Falcated Duck	<i>Anas falcata</i>	BW	NT	21.9	4.8
Eurasian Wigeon	<i>Anas penelope</i>	BW	LC	17.1	5.5
Mallard	<i>Anas platyrhynchos</i>	BW	LC	20.6	5.8
Spot-billed Duck	<i>Anas poecilorhyncha</i>	BW	LC	8.1	5.6
Northern Shoveler	<i>Anas clypeata</i>	BW	LC	20.5	9.0
Northern Pintail	<i>Anas acuta</i>	BW	LC	15.8	8.8
Garganey	<i>Anas querquedula</i>	BW	LC	19.4	10.6
Baikal Teal	<i>Anas formosa</i>	W	LC		7.4
Common Teal	<i>Anas crecca</i>	BW	LC	17.0	11.3
Marbled Duck	<i>Marmaronetta angustirostris</i>	B	V	0.7	
Red-crested Pochard	<i>Rhodonessa rufina</i>	B	LC	13.2	
Common Pochard	<i>Aythya ferina</i>	BW	V	22.1	7.8
Ferruginous Pochard	<i>Aythya nyroca</i>	BW	NT	31.9	9.7
Baer's Pochard	<i>Aythya baeri</i>	BW	CE	19.6	7.3
Tufted Duck	<i>Aythya fuligula</i>	BW	LC	17.9	7.8
Greater Scaup	<i>Aythya marila</i>	W	LC		4.4
Steller's Eider	<i>Polysticta stelleri</i>	W	V		41.0
Long-tailed Duck	<i>Clangula hyemalis</i>	W	V		11.6
Black Scoter	<i>Melanitta nigra</i>	W	LC		14.3
White-winged Scoter	<i>Melanitta fusca</i>	W	LC		9.7
Common Goldeneye	<i>Bucephala clangula</i>	BW	LC	32.9	9.1
Smew	<i>Mergellus albellus</i>	BW	LC	1.8	8.7
Red-breasted Merganser	<i>Mergus serrator</i>	BW	LC	25.1	6.9
Scaly-sided Merganser	<i>Mergus squamatus</i>	BW	E	12.8	9.5
Common Merganser	<i>Mergus merganser</i>	BW	LC	22.7	10.0

Habitat suitability models were developed at 1 km resolution for the season in which a species occurs in China (B breeding, W winter, BW breeding and wintering). Taxonomy and seasonal presence follows MacKinnon and Phillips (2000)

^a A species conservation status according to the IUCN (2017) is denoted as LC least concern, V vulnerable, NT near threatened, E endangered, CE critically endangered

^b The percent of suitable breeding habitat within Chinas protected areas

^c The percent of suitable wintering habitat within Chinas protected areas

was then structured in three parts: (a) records outlining seasonal habitat requirements for individual species, (b) population and survey counts, and (c) habitat relationship matrices that we developed to relate habitat requirements to land cover predictors (see Model Development below). The database holds 9250 records drawn from more than 1000 references (China Anatidae Network 2012).

With permission from Wetlands International, we used the Asian Waterbird Census (AWC; Li et al. 2009; Wetlands International 2017) to validate the wintering waterfowl models. The AWC provides waterbird survey data collected at wintering sites throughout Asia during January of each year, making it ideal as a source of validation data for this time period. However, a similar consistent source of nationwide survey location data was not available for the spring and summer months to test the breeding models except for one focal species, the Bar-headed Goose (*Anser indicus*), for which we used satellite telemetry data from a related study (Prosser 2011).

Environmental variables

Remotely-sensed land cover data are readily available across large geographic extents and have been used successfully in modeling species distributions (Gottschalk et al. 2005). Land cover variables (e.g., rice-paddy, lake, river, marsh, grassland, forest, etc.; Additional file 1: Table S1) used in this study were derived from 30 m Landsat imagery and distributed by the Chinese Academy of Sciences (CAS) at 1 km spatial resolution (Liu et al. 2002). The land cover dataset is continuous fields, whereby each class is represented as the percent cover within the 30 m² pixel (e.g. 10% marsh, 64% forest, etc., summing to 100). We tested variables for correlation to avoid issues of multicollinearity (Graham 2003). Significant correlations were not observed (all below 0.67), however we reduced the data set from 25 variables to 18 (Additional file 1: Table S1) based on an a priori list of relevant cover classes. While climatic variables were available through platforms such as the Moderate Resolution Imaging Spectroradiometer (MODIS) program facilitated by the National Aeronautics and Space Administration (NASA), we decided against their inclusion as the literature did not provide adequate depictions of how they would influence habitat suitability.

Model development and validation

Using a habitat suitability approach (Fig. 1), we created presence-absence predictions for each of China's waterfowl species (Table 1). Development of habitat matrices included the extensive literature review (see above) and communication with local experts. Habitat relationships

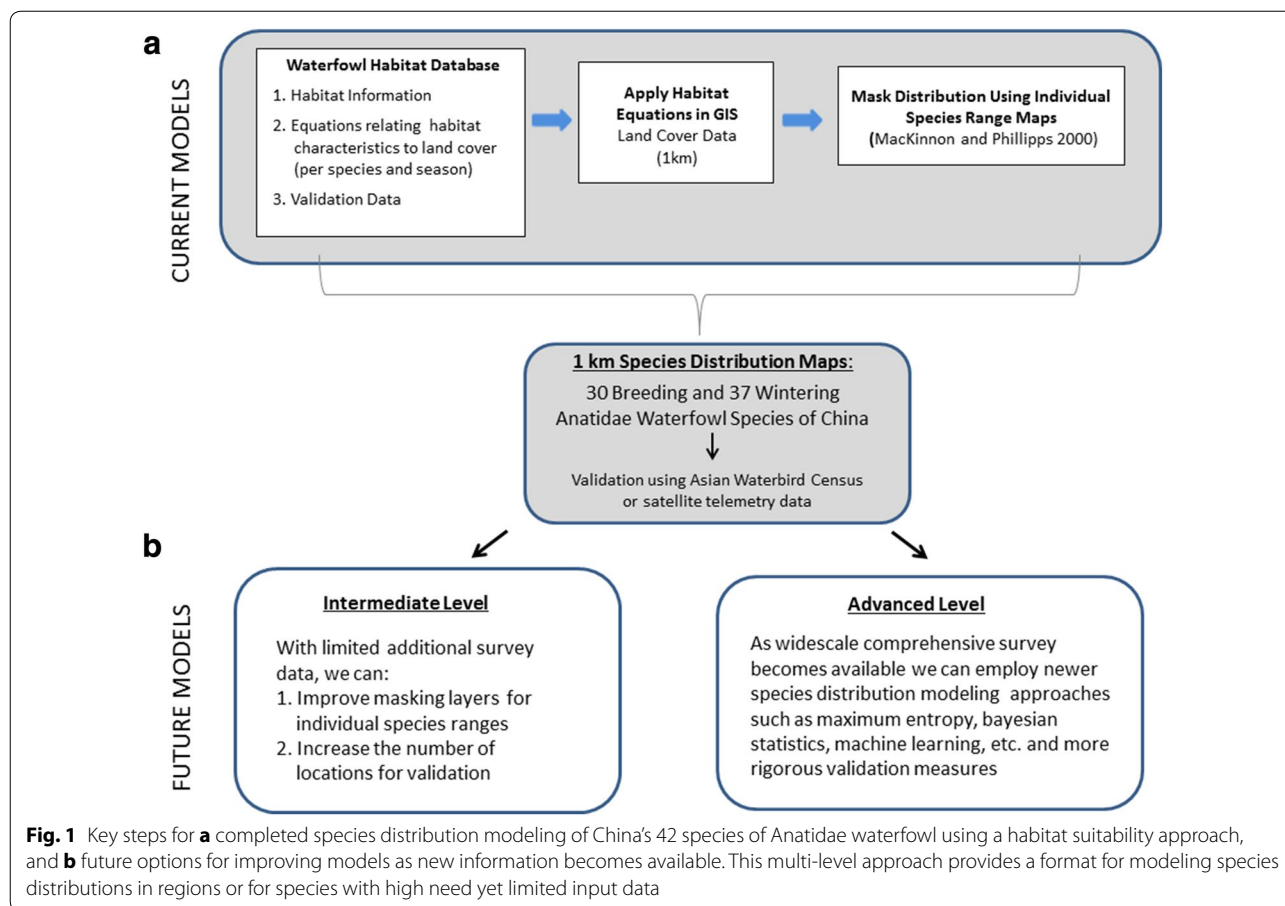
were developed by summarizing breeding or wintering habitats from the literature and linking them to appropriate land cover classifications of remotely-sensed Landsat imagery. For example, a summary of the literature indicated that the Bar-headed Goose breeds in habitats including shallow lakes, marshes, lake shores, highland moors, and salt lakes. We translated this summary into an equation that directly links land cover variables: marsh [Landsat category 64], rivers and irrigation channels [41], lakes [42], reservoir or pond [43], and river or lake shore [46] in combination greater than 0%. This descriptive equation was translated into the following equation:

$$\begin{aligned} &\text{Bar-headed Goose, Breeding} \\ &= (([64] + [41] + [42] + [43] + [46]) > 0) \end{aligned} \quad (1)$$

For the wintering season, the literature indicated that bar-headed geese use natural wetlands, agricultural fields, riverine wetlands, lacustrine wetlands, and freshwater lakes. The associated equation for land cover included: marsh [64], paddy [11], rainfed [12], rivers and irrigation channels [41], lakes [42], reservoir or pond [43] and river or lakeshore [46] in combination greater than 0:

$$\begin{aligned} &\text{Bar-headed Goose, Wintering} \\ &= (([64] + [11] + [12] + [41] + [42] + [43] + [46]) > 0) \end{aligned} \quad (2)$$

A complete list of species equations can be found in Additional file 2: Individual species models. The habitat equations were implemented in a geographic information system using Python coding (Python Software Foundation, Wilmington, Delaware) and ArcGIS 10.1 (ESRI, Redlands, California). The resulting suitability maps for each species were then masked (Fig. 1) using individual species range boundaries produced by MacKinnon and Phillipps (2000), one of the most comprehensive avian field references available for China. We used the mask to restrict suitable habitat to areas within the boundaries of known ranges for each species. This accounts for the natural distribution of each species and helps to avoid inclusion of regions that might contain suitable habitat but lie outside the range of a given species. While range maps present the broad distribution of a species, our approach identifies the areas within the species range that contain suitable habitat during the respective season. This approach reduces over-prediction of available space inherent in range maps (Graham and Hijmans 2006) and allows managers to focus only on relevant habitat for species of interest. One artifact of this approach is the appearance of a hard transition between predicted presence and absence cells along the outer boundary of each species range. While a soft transition could have been modeled, we chose to retain the



current boundaries so that results are explicit and easy to interpret.

The habitat suitability models were validated using AWC data for the winter season and satellite telemetry data for the breeding season (bar-headed goose only). Model validation included testing for errors of omission—identifying grid cells where the model predicts absence of a species but validation data shows that the species was present (Franklin and Miller 2010). Because coordinates for AWC sites represent the centroid of a larger census area that may range from less than 1 km to more than 10 km, we conducted validations at 3 spatial scales to appropriately reflect the possible spatial resolutions: (1) immediate (i.e., confirming model prediction as “suitable” within the 1 km pixel that encompasses the validation point), (2) within 5 km, and (3) within 10 km of known observation location. These values were chosen to reflect a progression towards the greatest possible locational error in the data. For the latter two, the test was used to confirm that within the specified distance of the validation point (5, 10 km), at least one pixel was predicted as “suitable” for the given species.

Species richness suitability maps

While species level outputs can provide important information, they can also serve as building blocks for a wide array of potential outputs. In order to depict regions which are suitable to the greatest number of species, and therefore likely to host large densities of waterfowl, we overlaid the model outputs for each species and summed the number of species within each cell for which the habitat was predicted to be suitable (Fig. 1). While an image containing all species gives insight into spatio-temporal trends in waterfowl distribution, disaggregation of the species included can allow examination of areas most important to relevant groupings. As an example, we also constructed species richness maps for birds within the genera *Anser* and *Aythya*. These genera were selected as they were well represented in our models and highly relevant to avian influenza and related conservation concerns. Species richness maps were derived separately for both winter and breeding seasons, with all calculations performed using Python coding (Python Software Foundation, Wilmington, Delaware) and ArcGIS 10.1 (ESRI, Redlands, California).

Protected areas

A dataset outlining the protected areas within China was obtained from the Chinese Academy of Sciences and their collaborators as outlined in Xu et al. (2017). This dataset represents the product of an extensive effort to identify the spatial boundaries for all protected areas within China, and contains 2412 terrestrial reserves which cover 15.1% of China's land surface (Xu et al. 2017). In this study we used the tabulate areas tool within ArcGIS to determine the percentage of suitable habitat per species located within protected areas, as determined by our breeding and wintering models. This is an approach similar to that of Zhang et al. (2017), which identified the number of important bird areas known to provide habitat for several waterbird species. Data were also visually examined to determine gaps in coverage at locations where habitat was suitable to multiple species within genera *Anser* or *Aythya*.

Results

Of the 42 waterfowl species reported in China, 39 are listed as winter residents and 30 as breeders (MacKinnon and Phillipps 2000). Based on this information, we produced suitability maps for 30 breeding and 39 wintering species (Table 1). Of these birds, 32 were ranked as species of least concern on the IUCN List of Threatened Species (IUCN 2017), while 6, 2, 1, and 1 were listed as Near threatened, Vulnerable, Endangered, and Critically endangered, respectively.

Breeding and wintering models as well as locations used for model validation are provided as an example in Fig. 2 for the Bar-headed Goose, a focal species due to its importance to HPAI transmission (Prosser et al. 2011) and its decreasing population (IUCN 2017). Suitable breeding habitat for the Bar-headed Goose included the high-elevation plateau of western China and a small section of Inner Mongolia in northeastern China. The pattern of suitable habitat for the breeding season was generally less dense (fewer 1 km cells within a given area) than for the wintering range, for the Bar-headed Goose (Fig. 2) as well as many of the waterfowl species (Fig. 3, Additional file 3: Fig. S1). Across species, there was variation in the extent and patchiness of suitable wintering habitat within each species' range. For example, the Greylag Goose (*Anser anser*) and Ruddy Shelduck (*Tadorna ferruginea*) models demonstrate a high density of suitable wintering habitat extending across much of their seasonal range, while suitable wintering habitat for the Common Teal (*Anas crecca*) was sparser and more patchy across a similar region (Fig. 3). In contrast, the winter model for the Cotton Pygmy-goose (*Nettapus coromandelianus*)

demonstrated a much more confined range of suitable wintering habitat, though it was comparable in patchiness to the Common Teal. Wintering and breeding maps for all species are included in Additional file 3: Fig. S1.

A total of 406 validation points from the winter Asian Waterbird Census were used to calculate omission rates for 14 of the 37 wintering species (23 species had no validation data). The number of validation points averaged 26 per species with a range of 1–166 (Table 2). Omission rates from validation procedures indicated a strong ability for the models to predict areas where a species might be found (zero errors of omission for 14 wintering species and three spatial scales; Table 2). For the Bar-headed Goose, 13 breeding and 21 wintering locations from a related satellite telemetry project (Prosser et al. 2011) were also used to validate the breeding and wintering models. Bar-headed Goose models had zero omission errors for the breeding season and an error rate of 0.095 for the wintering season (9.5%, or 2 of 21 breeding validation locations occurred within model cells that predicted species absence).

Cumulative species richness maps (Fig. 4) showed distinct spatial patterns for the breeding versus wintering seasons. Areas suitable for high species richness during the breeding season were centered in the northeast and the high-elevation western regions of China. In contrast, waterfowl richness during the wintering season was likely to be highest across much of the low-elevation southeastern part of China, particularly along the Yangtze River basin. Species richness potentials ranged from 0 to 20 species per grid cell for the breeding season and 0–31 for the wintering season, with similar patterns identified from the genus level maps (Fig. 5). Dalai Lake in the northeastern region of China was particularly likely to be important to both genera during the breeding season and was protected by the Dalai Lake National Nature Reserve, while regions of southeastern China such as the Yangtze River basin were of primary importance during the wintering period and had patchy protection provided by separated protected areas. Parsing out individual genera highlighted differences in important areas between groups. For instance, the Changtang Plateau, well protected by a network of large reserves, had far greater suitability to birds in the genus *Anser* than for those in *Aythya*. The percentage of suitable habitat within protected areas varied greatly by species and season, with breeding habitat generally better covered than wintering habitat (Table 1). The Ferruginous Pochard (*Aythya nyroca*), a near threatened species, demonstrates these trends with 31.9 and 9.7% of suitable breeding and wintering habitat within a protected area, respectively.

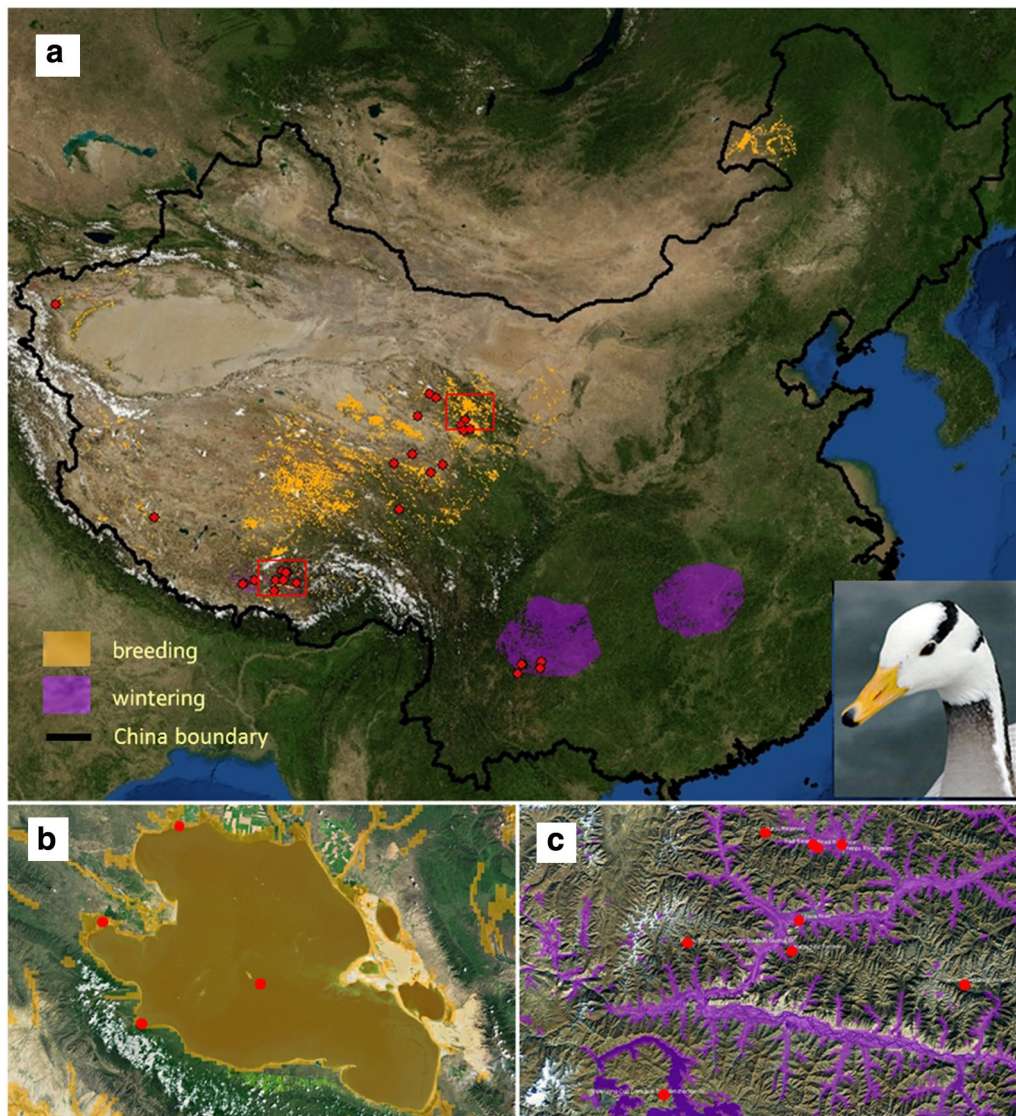


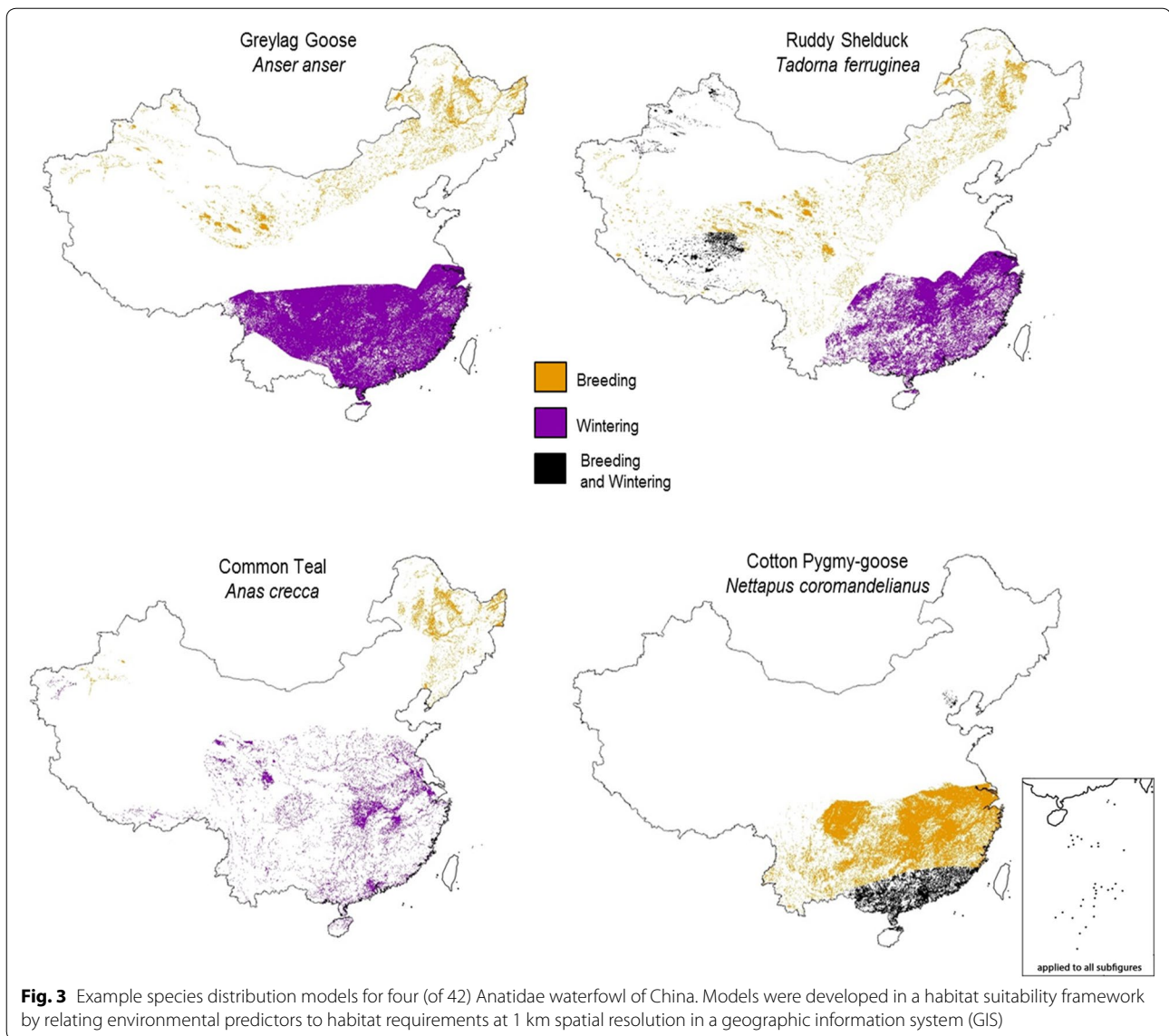
Fig. 2 Model results for example species, Bar-headed Goose, breeding (orange) and wintering (purple) seasons across China (**a**). Validation points from the waterfowl database (China Anatidae Network 2012) and our telemetry studies (Prosser et al. 2011) are depicted as red dots. Red frames delimit magnified insets **b** for breeding at Qinghai Lake, Qinghai and **c** wintering areas in southern Tibet

Discussion

The goal of this work was to develop high resolution spatial distribution maps for waterfowl species across seasons, providing a critical model input that would pave the way for informed management in a data-restricted region. Although there was a scientific deficit for producing data-driven models using newer SDM techniques (Guisan et al. 2013), the societal need to understand where these birds might occur both individually and cumulatively across species remained. Using the best data available, we took a conventional approach of mapping habitat suitability and combined it with model control measures (range masking) to develop and validate

habitat models for each species and combine these into species richness data layers. Such data layers, or any of the species specific layers, can now be directly incorporated as inputs into disease models (Prosser et al. 2016), critical habitat identification modeling, or any other need that requires estimation of a species most probable use of space, thereby resolving the challenge of incorporating information concerning wild birds into such efforts.

The key finding of this work was the seasonal changes in species distribution throughout the annual cycle. While there was variation between species, wintering distributions were predominately concentrated in the warmer and lower elevation regions of the southeast



while the breeding distributions were more evenly spaced across the northern latitudes and higher elevation regions of western China. These dynamic seasonal distributions have wide reaching implications for waterfowl management. For instance, the areas of greatest risk for interaction between poultry and wild birds for potential zoonotic disease transmission are likely to change throughout the year, and any attempts to limit such interactions must be responsive to these changes (Prosser et al. 2013). Therefore, future projects focused on modeling disease risk should incorporate the spatio-temporal behavior and abundance of wild birds, a step our approach makes possible even in regions of limited data availability. Seasonal distributions also indicate that any successful conservation attempts should include strategies to protect the birds throughout the year, which

requires attention to both wintering and breeding habitats (Cui et al. 2014; Dai et al. 2016).

While the spatio-temporal trends in avian influenza risk we observed have significant implications for how this disease should be modeled, they also have direct ecological implications. For instance, the Tufted Duck (*Aythya fuligula*) has been found to be especially prone to infection from highly pathogenic avian influenza, with infections often resulting in mortality (Keawcharoen et al. 2008). While the Tufted Duck is considered a species of least concern, other closely related species within the genus *Aythya* range from Vulnerable to Critically endangered (IUCN 2017, Table 1). Given the proclivity of some birds within the *Aythya* genus to suffer mortality from HPAI infections (Keawcharoen et al. 2008; Spackman et al. 2017) it is of great conservation importance

Table 2 Validation measures testing errors of omission for China waterfowl distribution maps at three scales: 1, 5, and 10 km

Common species name	Scientific name	Omission error rate ^a			Number of validation points ^b
		1 km	5 km	10 km	
Bar-headed Goose	<i>Anser indicus</i>				
Breeding season		0	0	0	13
Wintering season		0.095	0.095	0.095	21
Common Goldeneye	<i>Bucephala clangula</i>	0	0	0	74
Common Merganser	<i>Mergus merganser</i>	0	0	0	18
Common Pochard	<i>Aythya ferina</i>	0	0	0	1
Common Shelduck	<i>Tadorna tadorna</i>	0	0	0	2
Common Teal	<i>Anas crecca</i>	0	0	0	7
Greater Scaup	<i>Aythya marila</i>	0	0	0	5
Greylag Goose	<i>Anser anser</i>	0	0	0	1
Long-tailed Duck	<i>Clangula hyemalis</i>	0	0	0	40
Mallard	<i>Anas platyrhynchos</i>	0	0	0	166
Northern Pintail	<i>Anas acuta</i>	0	0	0	1
Red-breasted Merganser	<i>Mergus serrator</i>	0	0	0	64
Smew	<i>Mergellus albellus</i>	0	0	0	2
Tufted Duck	<i>Aythya fuligula</i>	0	0	0	8
Whooper Swan	<i>Cygnus cygnus</i>	0	0	0	17
Total number of validation points					406

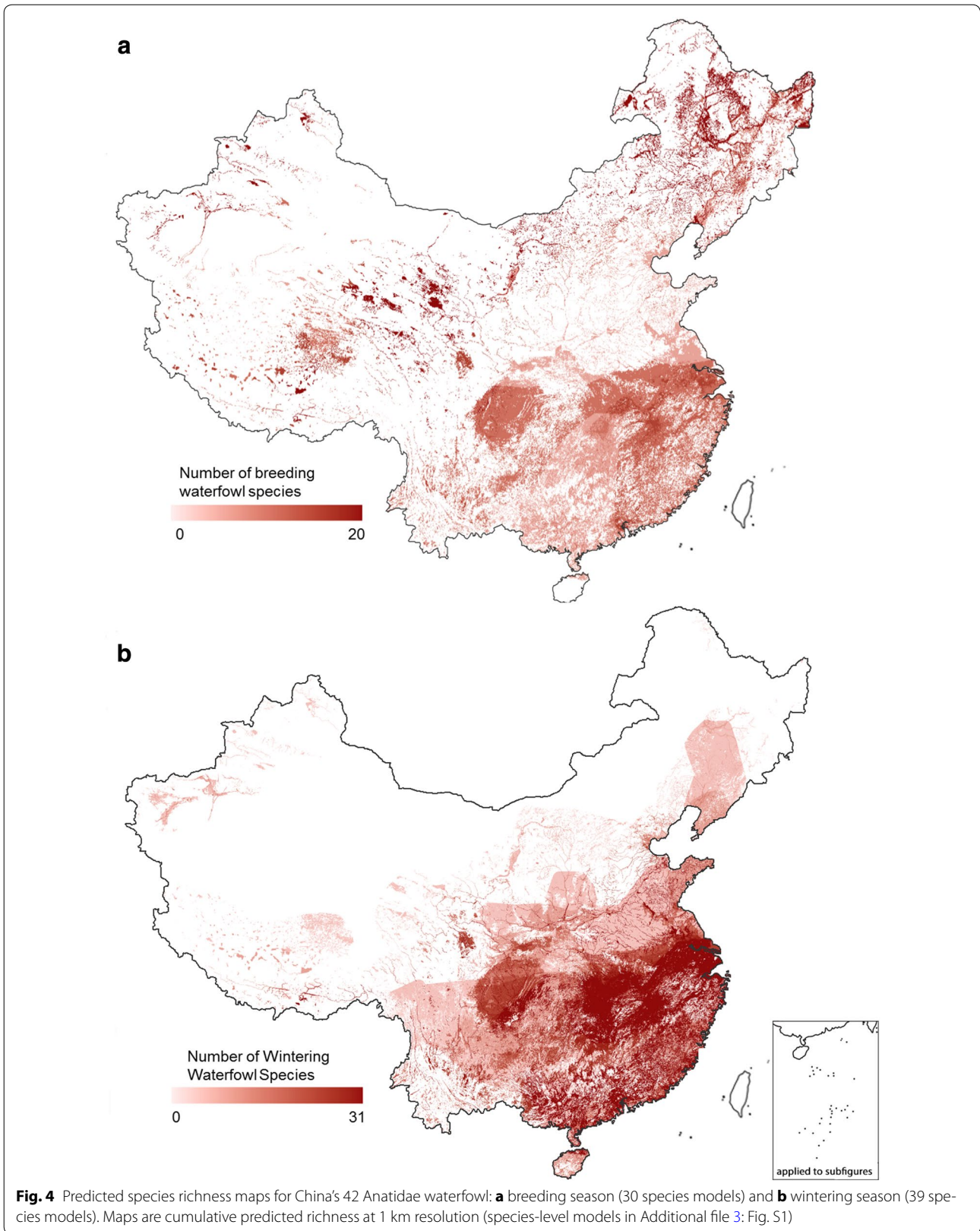
^a Omission rate is calculated by dividing the number of correctly predicted presence locations by the total number of validation (presence) points. For example, of the 21 validation locations where Bar-headed Geese were observed during the wintering season, two (or 9.5%) were incorrectly predicted as “absent” within the grid cell (within 1 km). In this example, increasing the number of neighboring cells to 5 or 10 km did not improve the error rate

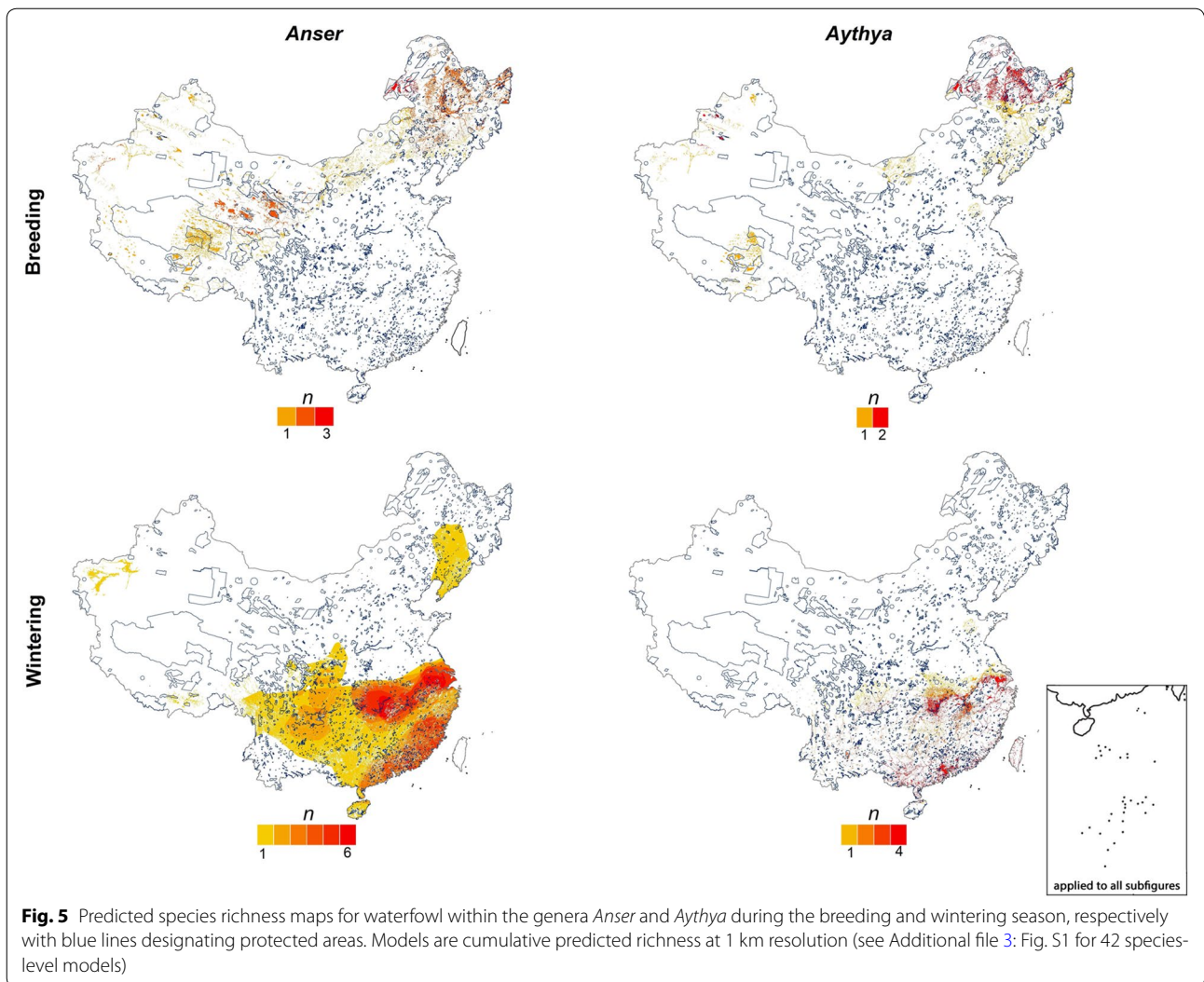
^b Asian Waterbird Census data were used as validation points for all species listed. In addition, for the Bar-headed Goose we tested for errors of omission using location data (both the breeding and wintering seasons) from our satellite telemetry work (Prosser et al. 2011)

to understand the risks such species face over time and space. Our findings suggest that the risk of these species to avian influenza virus is far from static, changing predictably based upon seasonal behavioral patterns. Informing our understanding of when and where vulnerable species are at greatest risk of exposure not only opens the door for more directed sampling efforts, but also enables the initiation of protective measures that could limit potential population losses. For instance, restrictions could be placed on poultry grazing and transportation through regions of seasonal importance so as to limit potential contact. Such practices could reflect the preventative measures applied in countries such as the United Kingdom (Gibbens 2017) and Thailand (Aengwanich et al. 2014) where risk to avian influenza is determined by geographic region and facility size, and biosecurity protocol are implemented accordingly.

The spatio-temporal changes in species richness observed between our wintering and breeding season models serve to highlight another key aspect of our approach: identifying areas of importance from a habitat conservation perspective. Though already established in the literature, our cumulative species richness maps confirm the importance of the Yangtze River (Cao et al. 2008,

2010; Cong et al. 2011; Williamson et al. 2013) and north-eastern China (Williamson et al. 2013) in a spatial context during the wintering and breeding seasons, respectively. Additionally, our genera-specific models allow for the identification of important locations at a much finer scale, such as identifying the Changtang Plateau in Tibet as being important to birds of the genus *Anser* during the breeding season (Zhang et al. 2015a, b). Our models also compared well to species distribution models prepared for this region. For instance, while there were some local areas of disagreement between our models and those produced by Zeng et al. (2015), both identified the Yangtze River Basin as the region within China of the greatest probability of occurrence for the Endangered Scaly-sided Merganser. Similarly, the models produced by Dai et al. (2016) report that key Bar-headed Goose summering habitat is found at Qinghai Lake and the Tibetan Plateau, findings which concur with our results. An oft noted benefit of SDM's is their usefulness in highlighting crucial habitats to target for conservation and intensive surveys (Moriguchi et al. 2013; Ochoa-Quintero et al. 2010). However, since all that is required to conduct our habitat suitability analysis is knowledge of a species' habitat needs and publicly available habitat data, this method





shows great promise to aid the conservation of species for which critical habitats may be less well understood, especially in these low data scenarios.

The ability of our approach to identify important habitats for waterfowl conservation is demonstrated when our model outputs are overlaid with the spatial boundaries of China’s protected areas. For instance, our outputs identify areas of suitable habitat for multiple species around Taihu Lake, near the mouth of the Yangtze River, where important wintering habitat is only sparsely protected. Our models also identify the need to focus on the protection of wintering habitats which are less protected than the breeding habitats (Cui et al. 2014). Both of these overarching trends are seen with the Ferruginous Pochard, a Near Threatened species. In China, this species breeds mostly throughout the Tibetan Plateau and the northwestern plain, with 31.9% of its suitable breeding habitat covered within expansive protected areas in

this region. Conversely, the Ferruginous Pochard winters in southeastern China with only 9.7% of its suitable wintering habitat protected via a patchwork of smaller protected areas. Therefore, this species would likely most benefit from increasing protected area connectivity within its suitable wintering habitat. The information provided by our models is especially important in the context of our study, as several of the species for which we created suitability maps are recognized as species of conservation concern by the IUCN (2017), and recent literature has highlighted the important role of protected areas in species conservation (Zhang et al. 2015a, b) within regions with effective governance (Amano et al. 2017), while advocating for the addition of new protected areas within China (Cao et al. 2010; Xia et al. 2016; Xu et al. 2017). Similarly, the outputs of our model can be used to monitor changes in available habitat over time. For instance, by utilizing imagery from different years,

researchers and managers can monitor changes in habitat availability for species of interest. The ability of conservationists to identify and respond to such changes is becoming increasingly important in the face of global climate change (Gillson et al. 2013; Yu et al. 2017), and models such as those created in this work provide the opportunity for responsive management even in these low data environments.

Though the comparability of our findings to other previous works provides some confidence in this approach, our formal validation provides critical additional support. The validation process demonstrated capacity to predict presence locations for individual species, however, because the validation data included presence-only records, we were not able to assess how well the models predicted the absence of a species, a common issue related to using presence-only data (Hernandez et al. 2006). Distribution models tend to be more successful in predicting presence locations than absence locations except for species with a very narrow niche (Brotons et al. 2004; Hernandez et al. 2006). We expected our models to be better at predicting presence than absence locations and recognize that this bias could translate into potential over-prediction of a species' distribution. However, with the intent of this model being to create an input layer for disease interface modeling we felt that a more inclusive approach was appropriate as less inclusive models would run the risk of excluding potentially important wild bird-domestic poultry interfaces. This inclusivity also has benefits when utilized in a conservation context. For instance, identifying the fundamental niche may be best suited when conserving areas around presently at-risk species as this provides buffers around the current areas of use and protects valuable habitat as populations rebound and expand. Yet it is still important to recognize that our approach identifies only suitable habitat for the given species, and thus depicts much less area than species range maps. It is through this enhanced degree of selectivity that important areas can be identified and targeted by managers and researchers as demonstrated above. In any modeling exercise, the implicit limitations should be identified and the model complexity should match the objectives of the study (Merow et al. 2014), as we did here.

One aspect of our approach that warrants special attention from those who wish to apply this method in future work is the impact the selected range map can have on the final model output. The benefit of masking model outputs to species range maps is that suitable habitat only within the species known distribution is considered. This provides obvious advantages for the utility of generated products. However, range maps often represent space use trends over many years, and may not

reflect current distributions. For instance, there have been recent reports of Mute Swan (*Cygnus olor*) wintering along the Yellow River delta (Brazil 2009), although the behavior is believed to be rare and consist of relatively few individuals. Unfortunately, observations of this behavior occurred after the publication of the range maps, chosen for their coverage of all relevant species, and as such no winter model was developed for this species. Given the rarity of this behavior, it is unlikely such an omission would have significant implications on the utility of model outputs, especially from a disease interface standpoint. Similarly, range maps make all areas of a species range appear equal in use, which can result in misleading results. For instance, White-winged Scoter (*Melanitta fusca*) winter predominantly along the eastern coast of China but have been known to utilize some inland habitats near Nanchong (MacKinnon and Philipps 2000). Therefore, while inland habitat use is very limited, our model output does not differentiate the likelihood of use between these two areas of suitable habitat. This issue could likely only be resolved via more complex modeling methods which would not be supported by the data available for this region. While these limitations do not remove the benefit of masking our model outputs with range maps, researchers should carefully consider the implications of potential shortcomings, and test for errors of omission whenever datasets are sufficient.

Our need for the information gained via modeling approaches, without having access to sufficient fine scale standardized datasets, forced us to use a habitat suitability approach. However, science is always changing, and as more data is collected additional analysis options become available. Therefore, we also outlined a framework for improving our current models as new data become available through the expansion and enhancement of waterfowl monitoring during the breeding season as well as the winter via the Asian Waterbird Census and other monitoring efforts (Fig. 1b). Improvements can be made iteratively based on availability of data. For example, an intermediate level model that revises species ranges and adds points for validation could be implemented with minimal additional effort and cost, and as more comprehensive data become available newer approaches to SDM can be employed (Fig. 1b). This multi-level approach, where models are meant to be improved as new information becomes available follows principles of Bayesian logic (Jewell et al. 2009) and Adaptive Resource Management (Allan and Stankey 2009), and can be applied towards a diverse range of distribution needs. Even if this method is not ideal when compared to approaches utilized under conditions of high data availability, it does not change for science-based management. Therefore, we hope that researchers from a wide array of disciplines will

recognize the opportunities this approach holds for conducting research and informing management in low data environments.

Conclusions

In this paper we present the first nationwide distribution models and predicted species richness maps for many of China's breeding and wintering waterfowl. We then use the information gained from these models to demonstrate their utility to conservation as it relates to avian influenza and the identification of critical habitats. We hope this example will encourage similar efforts in other regions with limited data but important needs for understanding the distribution of species across the landscape. We also hope this work will stimulate coordinated efforts to increase the level of input data and overall accuracy of these models. At this current stage, our high-resolution spatial models provide a unique and valuable resource to the research and planning communities across many disciplines from wildlife and habitat management to conservation medicine and beyond.

Additional files

Additional file 1: Table S1. China Land Cover dataset created by the Institute of Geographical Science and Natural Resources Research (IGSNRR), Chinese Academy of Sciences (CAS) from 30 m Landsat TM satellite imagery.

Additional file 2. Individual species models.

Additional file 3: Fig. S1. Species distribution models for 32 breeding and 37 wintering Anatidae waterfowl of China.

Authors' contributions

DJP, CD, RME and ECE conceived and designed the study, DJP and JDS analyzed the data, and DJP, JDS and TM wrote the paper. All authors read and approved the final manuscript.

Author details

¹ Patuxent Wildlife Research Center, U.S. Geological Survey, Beltsville, MD 20705, USA. ² University of Maryland, College Park, MD 20742, USA. ³ College of Nature Conservation, Beijing Forestry University, Beijing 100083, China. ⁴ University of Virginia, Charlottesville, VA 22904, USA. ⁵ Wetlands International, Horapark9, Ede 6717 LZ, The Netherlands. ⁶ Natural Systems Analysts, Winter Park, FL 32789, USA. ⁷ Department of Geography and Environmental Systems, University of Maryland Baltimore County, Baltimore, MD 21250, USA.

Acknowledgements

The authors thank Dingnan Lee (Beijing Forestry University) and Shane Heath (USGS) for collating Anatidae literature and developing the database; and Lei Cao (Chinese Academy of Sciences) for discussions on next steps for improving the models. We thank Ruth DeFries of Columbia University for her remote sensing and modeling expertise through all stages of this work. We thank Paul Marban (USGS) for formatting the species distribution maps, and Mike Haramis and two anonymous reviewers for improving earlier drafts of this manuscript. We also thank the many volunteers who contributed to the Asian Waterbird Census by conducting wintering waterbird surveys in China. The use of trade, product, or firm names in this publication is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

Data regarding waterfowl telemetry locations used during the current study are available from the corresponding author on reasonable request. Data regarding protected areas are available from Xu et al. (2017) but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of Xu et al. (2017). Final model outputs are available from the corresponding author on reasonable request.

Consent for publication

Not applicable.

Ethics approval and consent to participate

All telemetry data reported in this study were collected in accordance with protocol approved by the Patuxent Wildlife Research Center Animal Care and Use Committee.

Funding

This work was supported by the United States Geological Survey (Ecosystems Mission Area), the National Science Foundation Small Grants for Exploratory Research (No. 0713027), and Wetlands International.

Received: 23 August 2017 Accepted: 27 February 2018

Published online: 05 March 2018

References

- Aengwanich W, Boonsorn T, Srikot P. Intervention to improve biosecurity systems of poultry production clusters (PPCs) in Thailand. *Agriculture*. 2014;4:231–8.
- Alexander DJ. An overview of the epidemiology of avian influenza. *Vaccine*. 2007;25:5637–44.
- Allan C, Stankey GH. Adaptive environmental management: a practitioner's guide. Dordrecht: Springer; 2009.
- Amano TT, Szekeley B, Sandel S, Nagy T, Mundkur T, Langendoen T, Blanco D, Soykan CU, Sutherland WJ. Successful conservation of global waterbird populations depends on effective governance. *Nature*. 2017;2018(553):199–202.
- Brazil M. Birds of East Asia: Eastern China, Taiwan, Korea, Japan, and Eastern Russia. London: A&C Black; 2009.
- Brotans L, Thuiller W, Araujo MB, Hirzel AH. Presence-absence versus presence-only modelling methods for predicting bird habitat suitability. *Ecography*. 2004;27:437–48.
- Cao L, Barter M, Lei G. New Anatidae population estimates for eastern China: implications for current flyway estimates. *Biol Conserv*. 2008;141:2301–9.
- Cao L, Zhang Y, Barter M, Lei G. Anatidae in eastern China during the non-breeding season: geographical distributions and protection status. *Biol Conserv*. 2010;143:650–9.
- China Anatidae Network. Annual anatidae report to the State Forestry Administration. (2012);1–6 (in Chinese).
- Cong P, Cao L, Fox AD, Barter M, Rees EC, Jiang Y, Ji W, Zhu W, Song G. Changes in tundra swan *Cygnus columbianus bewickii* distribution and abundance in the Yangtze River floodplain. *Bird Conserv Int*. 2011;21:260–5.
- Cui P, Wu Y, Ding H, Wu J, Cao M, Chen L, Chen B, Lu X, Xu H. Status of wintering waterbirds at selected locations in China. *Waterbirds*. 2014;37:402–9.
- Dai S, Duole F, Bing X. Monitoring potential geographic distribution of four wild bird species in China. *Environ Earth Sci*. 2016;75:790.
- Dronova I, Beissinger SR, Burnham JW, Gong P. Landscape-level associations of wintering waterbird diversity and abundance from remotely sensed wetland characteristics of Poyang Lake. *Remote Sens*. 2016;8:462.
- Franklin J, Miller JA. Mapping species distributions: spatial inference and prediction. Cambridge: Cambridge University Press; 2010.
- Gibbins N. Declaration of an avian influenza prevention zone. London: Department for Environment, Food and Rural Affairs; 2017.

- Gilbert M, Pfeiffer DU. Risk factor modelling of the spatio-temporal patterns of highly pathogenic avian influenza (HPAIV) H5N1: a review. *Spat Spat Temp Epidemiol*. 2012;3:173–83.
- Gillson L, Dawson TP, Jack S, McGeoch MA. Accommodating climate change contingencies in conservation strategy. *Trends Ecol Evol*. 2013;28:135–42.
- Gottschalk TK, Huettmann F, Ehlers M. Thirty years of analysing and modelling avian habitat relationships using satellite imagery data: a review. *Int J Remote Sens*. 2005;26:2631–56.
- Graham CH, Hijmans RJ. A comparison of methods for mapping species ranges and species richness. *Global Ecol Biogeogr*. 2006;16:578–87.
- Graham MH. Confronting multicollinearity in ecological multiple regression. *Ecology*. 2003;84:2809–15.
- Grenyer R, Orme CDL, Jackson SF, Thomas GH, Davies RG, Davies TJ, Jones KE, Olson VA, Ridgely RS, Rasmussen PC, Ding TS, Bennett PM, Blackburn TM, Gaston KJ, Gittleman JL, Owens IPF. Global distribution and conservation of rare and threatened vertebrates. *Nature*. 2006;444:93–6.
- Guillera-Arroita G, Lahoz-Monfort JJ, Elith J, Gordon A, Kujala H, Lentini PE, McCarthy MA, Tingley R, Wintle BA. Is my species distribution model fit for purpose? Matching data and models to applications. *Global Ecol Biogeogr*. 2015;24:276–92.
- Guisan A, Tingley R, Baumgartner JB, Naujokaitis-Lewis I, Sutcliffe PR, Tulloch AIT, Regan TJ, Brotons L, McDonald-Madden E, Mantyka-Pringle C, Martin TG, Rhodes JR, Maggini R, Setterfield SA, Elith J, Schwartz MW, Wintle BA, Broennimann O, Austin M, Ferrier S, Kearney MR, Possingham HP, Buckley YM. Predicting species distributions for conservation decisions. *Ecol Lett*. 2013;16:1424–35.
- Hernandez A, Graham CH, Master LL, Albert DL. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*. 2006;29:773–85.
- IUCN. The IUCN Red List of threatened species. Version 2015–3. 2017. www.iucnredlist.org. Accessed 5 Dec 2017.
- Jewell CP, Kypraios T, Neal P, Roberts GO. Bayesian analysis for emerging infectious diseases. *Bayesian Anal*. 2009;4:465–96.
- Keawcharoen J, van Riel D, van Amerongen G, Bestebroer T, Beyer WE, van Lavieren R, Osterhaus ADME, Fouchier RA, Kuiken T. Wild ducks as long-distance vectors of highly pathogenic avian influenza virus (H5N1). *Emerg Infect Dis*. 2008;14:600–7.
- Li ZWD, Bloem A, Delany S, Martakis G, Quintero JO. Status of waterbirds in Asia: results of the Asian Waterbird Census 1987–2007. Kuala Lumpur: Wetlands International; 2009.
- Liu J, Liu M, Deng X, Zhuang D, Zhang Z, Luo D. The land use and land cover change database and its relative studies in China. *J Geogr Sci*. 2002;12:275–82.
- Liu J, Xiao H, Lei F, Zhu Q, Qin K, Zhang XW, Zhang XL, Zhao D, Wang G, Feng Y, Ma J, Liu W, Wang J, Gao GF. Highly pathogenic H5N1 influenza virus infection in migratory birds. *Science*. 2005;309:1206.
- Mackinnon J, Phillipps K. A field guide to the birds of China. New York: Oxford University Press Inc.; 2000.
- Martin LJ, Blossey B, Ellis E. Mapping where ecologists work: biases in the global distribution of terrestrial ecological observations. *Front Ecol Environ*. 2012;10:195–201.
- Merow C, Smith MJ, Edwards TC, Guisan A, McMahon S, Normand S, Thuiller W, Wuest R, Zimmermann N, Elith J. What do we gain from simplicity versus complexity in species distribution models? *Ecography*. 2014;37:1267–81.
- Moriguchi S, Amano T, Ushiyama K. Creating a potential distribution map for greater white-fronted geese wintering in Japan. *Ornithol Sci*. 2013;12:117–25.
- Muzaffar SB, Takekawa JY, Prosser DJ, Newman SH, Xiao X. Rice production systems and avian influenza: interactions between rice, poultry and wild birds. *Waterbirds*. 2010;33:219–30.
- Ochoa-Quintero JM, Szabolcs N, Flink S. Use of species distribution modelling based on data from the African waterbird census to predict waterbird distributions in Africa and identify gaps in knowledge of distribution. In: Anselin A (ed) *Bird Numbers 2010: Monitoring, indicators and targets*. Proceedings of the 18th Conference of the European Bird Census Council, Cáceres, Spain (partim). *Bird Census News*. 2010;23:29–40.
- OIE. Update on highly pathogenic avian influenza in animals: Type H5 and H7. 2017. <http://www.oie.int/animal-health-in-the-world/update-on-avian-influenza/2017/>. Accessed 5 Dec 2017.
- Prosser D, Cui P, Takekawa JY, Tang MJ, Hou YS, Collins BM, Yan BP, Hill NJ, Li TX, Li YD, Lei FM, Guo S, Xing Z, He YB, Zhou YC, Douglas DC, Perry WM, Newman SH. Wild bird migration across the Qinghai-Tibetan Plateau: a transmission route for highly pathogenic H5N1. *PLoS ONE*. 2011. <https://doi.org/10.1371/journal.pone.0017622>.
- Prosser DJ, Hungerford LL, Erwin RM, Ottinger MA, Takekawa JY, Ellis EC. Mapping risk of avian influenza transmission at the interface of domestic poultry and wild birds. *Front Public Health*. 2013;28:1–11.
- Prosser DJ, Hungerford LL, Erwin RM, Ottinger MA, Takekawa JY, Newman SH, Xiao X, Ellis EC. Spatial modeling of wild bird risk factors for highly pathogenic A(H5N1) avian influenza virus transmission. *Avian Dis*. 2016;60:329–36.
- Root T. Atlas of wintering North American birds: an analysis of Christmas Bird Count Data. Chicago: University of Chicago Press; 1988.
- Sauer JR, Fallon JE, Johnson R. Use of North American breeding bird survey data to estimate population change for bird conservation regions. *J Wildl Manage*. 2003;67:372–89.
- Spackman E, Prosser DJ, Pantin-Jackwood MJ, Berlin AM, Stephens CB. The pathogenesis of Clade 2.3. 4.4 H5 highly pathogenic avian influenza viruses in ruddy duck (*Oxyura jamaicensis*) and lesser scaup (*Aythya affinis*). *J Wildl Dis*. 2017;53:832–42.
- Segurado P, Araujo MB. An evaluation of methods for modelling species distributions. *J Biogeogr*. 2004;31:1555–68.
- Wetlands International. Asian waterbird census. 2017. <https://south-asia.wetlands.org/our-approach/healthy-wetland-nature/asian-waterbird-census/>. Accessed 28 May 2017.
- Williamson L, Hudson M, O'Connell M, Davidson N, Young R, Amano T, Szekely T. Areas of high diversity for the world's inland-breeding waterbirds. *Biodivers Conserv*. 2013;22:1501–12.
- Xia S, Yu X, Millington S, Liu Y, Jia Y, Wang L, Hou X, Jiang L. Identifying priority sites and gaps for the conservation of migratory waterbirds in China's coastal wetlands. *Biol Conserv*. 2016;210:72–82.
- Xu W, Xiao Y, Zhang J, Yang W, Zhang L, Hull V, Wang Z, Zheng H, Liu J, Polasky S, Jiang L, Xiao Y, Shi X, Rao E, Lu F, Wang X, Daily GC, Ouyang Z. Strengthening protected areas for biodiversity and ecosystem services in China. *Proc Natl Acad Sci USA*. 2017;114:1601–6.
- Xu X, Subbarao K, Cox NJ, Guo Y. Genetic characterization of the pathogenic influenza A/Goose/Guangdong/1/96 (H5N1) virus: similarity of its hemagglutinin gene to those of H5N1 viruses from the 1997 outbreaks in Hong Kong. *Virology*. 1999;261:15–9.
- Yu H, Wang X, Cao L, Zhang L, Jia Q, Lee H, Xu Z, Liu G, Xu W, Hu B, Fox AD. Are declining populations of wild geese in China 'prisoners' of their natural habitats? *Curr Biol*. 2017;27:365–77.
- Zeng Q, Zhang Y, Sun G, Duo H, Wen L, Lei G. Using species distribution model to estimate the wintering population size of the endangered scaly-sided merganser in China. *PLoS ONE*. 2015;10:e0117307.
- Zhang G, Liu D, Jiang H, Zhang K, Zhao H, Kang A, Liang H, Qian F. Abundance and conservation of waterbirds breeding on the Changtang Plateau, Tibet autonomous region, China. *Waterbirds*. 2015a;38:19–29.
- Zhang L, Wang X, Zhang J, Ouyang Z, Chan S, Crosby M, Watkins D, Martinez J, Su L, Yu Y, Szabo J, Cao L, Fox AD. Formulating a list of sites of waterbird conservation significance to contribute to China's ecological protection red line. *Bird Conserv Int*. 2017;27:153–66.
- Zhang Y, Jia Q, Prins HHT, Cao L, de Boer WF. Effect of conservation efforts and ecological variables on waterbird population sizes in wetlands of the Yangtze river. *Sci Rep*. 2015b;5:17136. <https://doi.org/10.1038/srep17136>.