

RESEARCH

Open Access

A global model of avian influenza prediction in wild birds: the importance of northern regions

Keiko A Herrick¹, Falk Huettmann^{1,2*} and Michael A Lindgren³

Abstract

Avian influenza virus (AIV) is enzootic to wild birds, which are its natural reservoir. The virus exhibits a large degree of genetic diversity and most of the isolated strains are of low pathogenicity to poultry. Although AIV is nearly ubiquitous in wild bird populations, highly pathogenic H5N1 subtypes in poultry have been the focus of most modeling efforts. To better understand viral ecology of AIV, a predictive model should 1) include wild birds, 2) include all isolated subtypes, and 3) cover the host's natural range, unbounded by artificial country borders. As of this writing, there are few large-scale predictive models of AIV in wild birds. We used the Random Forests algorithm, an ensemble data-mining machine-learning method, to develop a global-scale predictive map of AIV, identify important predictors, and describe the environmental niche of AIV in wild bird populations. The model has an accuracy of 0.79 and identified northern areas as having the highest relative predicted risk of outbreak. The primary niche was described as regions of low annual rainfall and low temperatures. This study is the first global-scale model of low-pathogenicity avian influenza in wild birds and underscores the importance of largely unstudied northern regions in the persistence of AIV.

Introduction

The influenza viruses that caused the four deadliest human pandemics of the past century (1918, 1957, 1968, 2009) contained gene segments from avian influenza acquired through recent reassortment events (reviewed in [1]). Influenza is thought to have originated in wild birds, and waterfowl are considered the primary reservoir. Avian influenza virus (AIV) most commonly infects *Anseriformes*, *Passeriformes*, and *Charadriiformes* in wild populations, particularly family *Anatidae* [2]. The Asian strains of the highly pathogenic H5N1 AIV subtype in poultry have received the most attention because of economic losses caused by this subtype, the virus's transmissibility from chicken to human [3,4], and fears over a new human influenza pandemic [5]. However, HPAI H5N1 is not the only strain with pandemic potential: cases of human infection with interspecies H7 and H9 subtypes have been reported [6,7] and others, such as

H6, can be highly pathogenic to poultry. The vast majority of influenza strains are of low pathogenicity to poultry, but because AIV is a virus of great diversity with the potential for rapid evolution [8], the full range of its variation should be considered rather than just focusing on a single strain or subtype. A massive reservoir of genetic diversity for potential reassortment of AIV exists in wild bird populations, from which nearly all combinations of hemagglutinin (H) and neuraminidase (N) subtypes have been isolated [9-11].

A number of ongoing surveillance projects record the subtypes of AIV isolated from wild birds [9,10,12,13]. Large cooperative databases, such as the Influenza Research Database (IRD), curate the surveillance efforts of multiple institutions. IRD provides an opportunity to apply predictive modeling to AIV on a global scale. In the prediction and risk assessment of infectious diseases, geographic information systems (GIS) and predictive modeling techniques are important tools [14]. Predictive models of Chagas disease [15], malaria [16], leishmaniasis [17], and Lyme disease [18] have been used to map disease prevalence and identify important factors contributing to risk. Several models have assessed risk factors for H5N1 in domestic poultry and produced high resolution models for India [19], Vietnam [20], and

* Correspondence: fhuettmann@alaska.edu

¹Ecological Wildlife Habitat Analysis of the Land- and Seascape (EWHALE) Lab, Biology and Wildlife Department, University of Alaska Fairbanks, Fairbanks, AK 99775, USA

²Institute of Arctic Biology, University of Alaska Fairbanks, Fairbanks, AK 99775-7000, USA

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

China [21]. One predictive map of multiple species of wild birds and subtypes of AIV was developed for the continental United States [22], and one for flyways adjacent to the Pacific Rim [23].

As of this writing, there are no global-scale predictions of LPAI in wild birds. The development of a global model could be an important tool in the management and risk assessment of AIV in the interest of public and animal health. Our model extends the value of AIV surveillance efforts by using the data for predictive purposes, not simply for descriptive purposes. In addition, a global model encompasses the distribution range of important reservoir species, many of which travel vast distances [24] in cross-continental migratory journeys to and from breeding grounds each year. We used ensemble data-mining machine-learning methods to 1) identify important predictor variables, 2) quantitatively describe the environmental niche of AIV in wild bird populations, and 3) develop a near-global-scale predictive map (excluding Antarctica) of AIV based on this described niche.

Materials and methods

Wild bird data

Sample data points of AIV-negative and AIV-positive data for wild birds were obtained from the Influenza Research Database online [25]. This dataset spans five years (2005–2010) of surveillance data providing geo-referenced collection coordinates for each sample, species name, AIV-positive or –negative status (determined by the collecting institution), viral subtype (where available), and many other collection specifics. We did not distinguish between high- and low-pathogenicity AIV strains in our dataset. We groomed the database to remove samples from domestic species and samples from unidentified species (listed as “Unknown”). In addition, this version of the database contained many instances where the latitude and longitude values were inverted; we examined each point for a match between GPS coordinates and collection location, and corrected it if the error was obvious or removed it if uncertainty remained. We randomly divided the data points into two pools for training the model (47 898 points) and testing the model (12 080 points) using MS Excel and imported both sets as point layers into ArcMap v.10.0 (Esri, Redlands USA).

Environmental variable layers

Forty-one predictor variable layers for ArcGIS were acquired from open source projects and included bioclimatic, geographic, and anthropogenic variables (Table 1). The extent of this model is bounded by these data layers, which exclude Antarctica. Bioclimatic variables included mean temperature for each month, for quarters (e.g. wettest quarter), and annual means for

precipitation and temperature. A number of time-dependent variables were included (i.e. mean temperature in January - December) and were manipulated in order to maintain their relevance to collection locations in the Southern Hemisphere. For points with negative latitude values, time-dependent variables were shifted by 6 months, such that months were correctly associated with the austral seasons. Geographic variables included elevation, which has been identified as an important factor in other AIV models [19], and lakes, rivers, and wetlands, which are important to waterfowl. We calculated some layers from existing predictor variables using the Spatial Analyst Tool in ArcMap. The distances from fresh water features and coastline were calculated using the Euclidean Distance Tool. Slope was calculated from elevation and aspect was in turn calculated from slope. Anthropogenic variables included indices of human manipulation, infrastructure, and population density. Due to the importance of chickens and pigs in the transmission of AIV to humans, we included predicted poultry and pig densities [26,27]. Not all layers included Antarctica, so the entire continent was excluded from the study area (layers trimmed at -57° latitude) to prevent biases in calculation. We then used the Geospatial Modeling Environment (GME; [28]) to intersect, or extract the values of the predictor variables at the same geographic coordinates as the sample data points. GME adds the values of each predictor variable to the database as an additional column. The intersected database is then imported into ArcMap for visualization. Layers and metadata are stored at and can be obtained from the Ecological Wildlife Habitat Analysis of the Land- and Seascape (EWHALE) Lab at the University of Alaska Fairbanks (UAF).

Defining the outbreak niche

We used the Random Forests algorithm [34], an ensemble data-mining machine-learning method, to identify the variables that best predicted the AIV-positive niche. We chose this particular algorithm because it is a powerful method of data-mining that performs with equal or superior accuracy to other algorithms (such as TreeNet, MARS, and Regression Tree Analysis) when used in ecological prediction [35,36]. Random Forests is relatively immune to overfitting and noise [34], which is a valuable feature when many similar predictor variables are incorporated. In addition, Random Forests ranks predictor variables by their contribution to model accuracy and the Variable Importance Scores (VIS) are normalized to the highest scoring variable. Using the pool of training data, we ran the Random Forests analysis method for classification trees in Salford Predictive Miner (Salford Systems, San Diego USA) with the following settings: class weights were balanced to up-

Table 1 The predictor variables used by the Random Forests algorithm to create a global prediction map for avian influenza virus in wild birds.

| Predictor variable | Details | VIS | Project source |
|---|--|-------|---|
| Annual precipitation (mm) | ln mm; 30 arc-seconds, 1 km spatial resolution | 100.0 | WorldClim [29] |
| Mean temperature, June (°C) | ln °C; 30 arc-seconds, 1 km spatial resolution | 85.2 | WorldClim |
| Mean temperature, April (°C) | ln °C; 30 arc-seconds, 1 km spatial resolution | 76.1 | WorldClim |
| Precipitation of driest quarter (mm) | ln mm; 30 arc-seconds, 1 km spatial resolution | 68.4 | WorldClim |
| Mean temperature, November (°C) | ln °C; 30 arc-seconds, 1 km spatial resolution | 64.9 | WorldClim |
| Precipitation seasonality | ln mm; 30 arc-seconds, 1 km spatial resolution | 63.1 | WorldClim |
| Mean temperature of driest quarter (°C) | ln °C; 30 arc-seconds, 1 km spatial resolution | 62.5 | WorldClim |
| Annual mean temperature (°C) | ln °C; 30 arc-seconds, 1 km spatial resolution | 54.8 | WorldClim |
| Mean temperature, February (°C) | ln °C; 30 arc-seconds, 1 km spatial resolution | 49.7 | WorldClim |
| Mean temperature, January (°C) | ln °C; 30 arc-seconds, 1 km spatial resolution | 45.7 | WorldClim |
| Temperature seasonality | Standard deviation × 100 | 45.0 | WorldClim |
| Precipitation of wettest quarter (mm) | ln mm; 30 arc-seconds, 1 km spatial resolution | 42.9 | WorldClim |
| Mean temperature, December (°C) | ln °C; 30 arc-seconds, 1 km spatial resolution | 38.6 | WorldClim |
| Maximum temperature of warmest month (°C) | ln °C; 30 arc-seconds, 1 km spatial resolution | 38.0 | WorldClim |
| Precipitation of driest month (mm) | ln mm; 30 arc-seconds, 1 km spatial resolution | 37.5 | WorldClim |
| Mean temperature, October (°C) | ln °C; 30 arc-seconds, 1 km spatial resolution | 36.1 | WorldClim |
| Mean temperature, September (°C) | ln °C; 30 arc-seconds, 1 km spatial resolution | 32.6 | WorldClim |
| Precipitation of coldest quarter (mm) | ln mm; 30 arc-seconds, 1 km spatial resolution | 32.3 | WorldClim |
| Population density (persons/km ²) | Population density for 2010, 2.5' resolution, persons/km ² | 29.3 | Gridded Popn of the World, v.3 [30] |
| Mean temperature of coldest quarter (°C) | ln °C; 30 arc-seconds, 1 km spatial resolution | 28.4 | WorldClim |
| Mean temperature, July (°C) | ln °C; 30 arc-seconds, 1 km spatial resolution | 28.3 | WorldClim |
| Isothermality (°C) | (Mean Diurnal Range/Temperature Annual Range); ln °C; 30 arc-seconds, 1 km spatial resolution | 26.6 | WorldClim |
| Mean temperature of wettest quarter (°C) | ln °C; 30 arc-seconds, 1 km spatial resolution | 25.2 | WorldClim |
| Mean diurnal range (°C) | ln °C, (mean of monthly temperature(max – min)) | 24.3 | WorldClim |
| Mean temperature, August (°C) | ln °C; 30 arc-seconds, 1 km spatial resolution | 22.7 | WorldClim |
| Mean temperature, March (°C) | ln °C; 30 arc-seconds, 1 km spatial resolution | 21.0 | WorldClim |
| Temperature annual range (°C) | ln °C; 30 arc-seconds, 1 km spatial resolution | 18.4 | WorldClim |
| Elevation (m) | ln m; 30 arc-seconds, 1 km spatial resolution | 18.3 | WorldClim |
| Predicted pig density (per km ²) | Animal density/km ² ; 3 min of arc | 18.0 | Gridded livestock of the world [31] |
| Mean temperature of warmest quarter (°C) | ln °C; 30 arc-seconds, 1 km spatial resolution | 17.0 | WorldClim |
| Precipitation of wettest month (mm) | ln mm; 30 arc-seconds, 1 km spatial resolution | 15.0 | WorldClim |
| Distance from coast (m) | Calculated from coastline | 12.6 | Esri |
| Precipitation of warmest quarter (mm) | ln mm; 30 arc-seconds, 1 km spatial resolution | 12.1 | WorldClim |
| Minimum temperature of coldest month (°C) | ln °C; 30 arc-seconds, 1 km spatial resolution | 11.2 | WorldClim |
| Human footprint index | Percentage of relative human influence (0–100) | 10.4 | Last of the Wild [32] |
| Distance from rivers, lakes, or wetlands | Calculated from combined large and small lake polygons, and lakes and wetlands grid | 9.7 | Global Lakes and Wetlands Database [33] |
| Slope | Calculated from elevation | 9.0 | Esri |
| Human influence index | Summative index of human disturbance (0–72) | 8.8 | Last of the Wild [32] |

Table 1 The predictor variables used by the Random Forests algorithm to create a global prediction map for avian influenza virus in wild birds. (Continued)

| | | | |
|--|--|-----|-------------------------------------|
| Predicted poultry density (per km ²) | Animal density/km ² ; 3 min of arc | 7.9 | Gridded livestock of the world [31] |
| Aspect | positive degrees from 0 to 359.9, measured clockwise from north; calculated from slope | 6.7 | Esri |
| Mean temperature, May (°C) | In °C; 30 arc-seconds, 1 km spatial resolution | 5.4 | WorldClim |

Variables are listed in order of their Variable Importance Score (VIS) or relative contribution to model accuracy as calculated by Random Forests.

weight the smaller number of AIV-positive samples against AIV-negative samples; the number of trees was set to 500; and seven predictors were used at each node [34].

The top five variables with the highest VIS were chosen for further examination. To compare the number of AIV-positive and -negative samples taken across each variable's range of values we plotted density using Spotfire S+ (TIBCO, v.8.2, Palo Alto USA). The ranges within which peaks occur suggest underlying mechanisms, which may be driving AIV outbreaks. Partial dependence plots were produced using the "partialPlot()" command in the RandomForest package [37] in R statistical programming language [38]. Partial dependence can be thought of as an index summarizing the quantified relationship of a predictor with the response variable after averaging the noise of non-relevant predictors [39]. Partial dependence plots can be useful in illustrating general trends in model accuracy's dependence on predictors. The partial dependence of a variable's effect is best understood by examining general patterns in relation to the values of the predictor variable rather than the specific values of partial dependence.

As a negative control, we calculated AUC for the AIV-positive status of the training subset against three individual predictors: the predictor with the highest importance score, the lowest importance score, and Annual Mean Temperature. We examined partial dependence plots for the general relationship between the range of predictor values and AIV-positivity (e.g. Figure 1a for Annual Precipitation). Predictor values for each point in the subset of training data were normalized between 0 and 1. If the relationship was negative, then the values were inverted such that low values of the variable predicted high occurrence. All three sets of normalized values were then subjected to AUC calculations.

Predictive map

To predict the relative occurrence of AIV in unsampled areas, we applied the model to a lattice of points spaced 100 km apart and calculated a predicted value for each point. Random Forests expresses the predicted occurrence of AIV as a Relative Occurrence Index (ROI)

rather than a probability score [40]. In ArcMap, we applied the Inverse Distance Weighted Tool (IDW) to interpolate these ROI values between the points, and generated a map of predicted AIV outbreak locations. The final map was projected as Robinson (sphere) with the central meridian at 145° so that Africa and Europe are displayed intact.

To evaluate the performance of the model, we calculated the Receiver Operating Characteristic (ROC) curve by plotting true positive points (AIV-positive status) against false positives with the program ROC_AUC [41]. A ROC value of 0.5 means a model accuracy of 50% in predicting positives and is no better than the random assignment of positive or negative status. A ROC value of 1.0 shows the model accurately classified 100% of points. If the area under the resulting curve (AUC) exceeds the critical value of 0.7, the model has high predictive power [42]. To evaluate accuracy, the model was applied to the pool of testing points. A ROC curve was calculated for these points using their predicted ROI value against their experimental AIV-positive status.

Results

Important predictor variables

Annual precipitation, mean temperature in June, and mean temperature in April were the most important predictor variables with VIS of 100, 85.2, and 76.1, respectively. Predictor variables with VIS above 50 were split almost equally between precipitation measurements and the mean temperatures in November, the driest quarter (3 month period), and annual mean temperature (Table 1). In the density plots (Figure 1a-e) the relative frequency of sampling was approximated by the density of the AIV-negative group of samples (represented by the solid black line); the range of values over which sampling occurred was inferred from the AIV-negative group. In general, the lack of perfect correspondence between AIV-positive (dotted red line) and AIV-negative groups showed that there were unequal densities of AIV-positivity across the sampling range. Thus AIV-positive samples did not occur at the same relative frequency as sampling effort. The ranges where the density of AIV-positive samples exceeded those of AIV-negative samples imply conditions correlated with

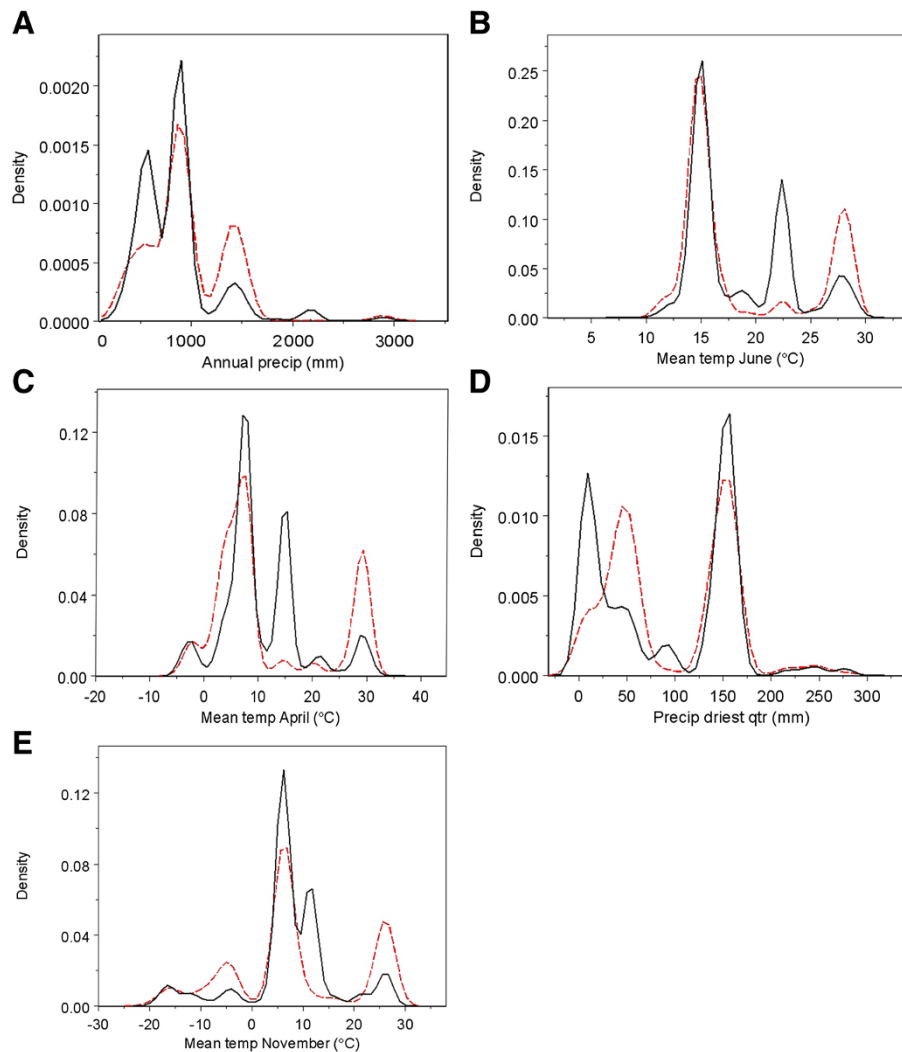


Figure 1 Density plots for important variables. Density plots for the variables with the five highest variable importance scores as calculated by Random Forests in the accuracy of the predictive model of avian influenza in wild birds (A-E). The density, or the likelihood of a variable to take on a value, of AIV-positive samples for each variable is represented by red dotted lines, the AIV-negative by black solid lines.

AIV-positivity. In the case of annual precipitation (Figure 1a), moderate (1400 mm) and very low (~0 mm) values were correlated with AIV-positivity. The partial dependence of AIV-positivity on annual precipitation exhibited a similar trend: very high dependence at 0 mm, a trough, and then moderately high dependence at values over 1000 mm (Figure 2a). These patterns imply that areas of low annual precipitation are most correlated with AIV-positivity, although areas of relatively high annual precipitation show some correlation as well. Areas of very low and high mean temperatures in June and April were correlated with AIV-positivity (Figures 1b,c), while areas of moderate temperature were not. June and April displayed similar patterns with a strong peak at the high range of sampling (~28°C and 30°C, respectively) and at the lowest

ranges (~10°C and 0°C, respectively). Examination of partial dependence revealed that AIV-positivity was high at the lowest temperatures, dropped sharply at moderate temperatures, and gradually increased at the higher end of the range (Figures 2b,c). Thus areas with low temperatures in June and April were correlated with AIV-positive samples. Precipitation of the driest quarter displayed one peak at 50 mm where AIV-positives had a higher density than AIV-negatives (Figure 1d). However, while the partial dependence was high at this value, it appears as a lone spike in an area of low partial dependence. Partial dependence on precipitation increases above 150 mm and reaches high levels above 250 mm (Figure 2d). While the highest density of AIV-positives occurred at relatively low annual precipitation, partial dependence was

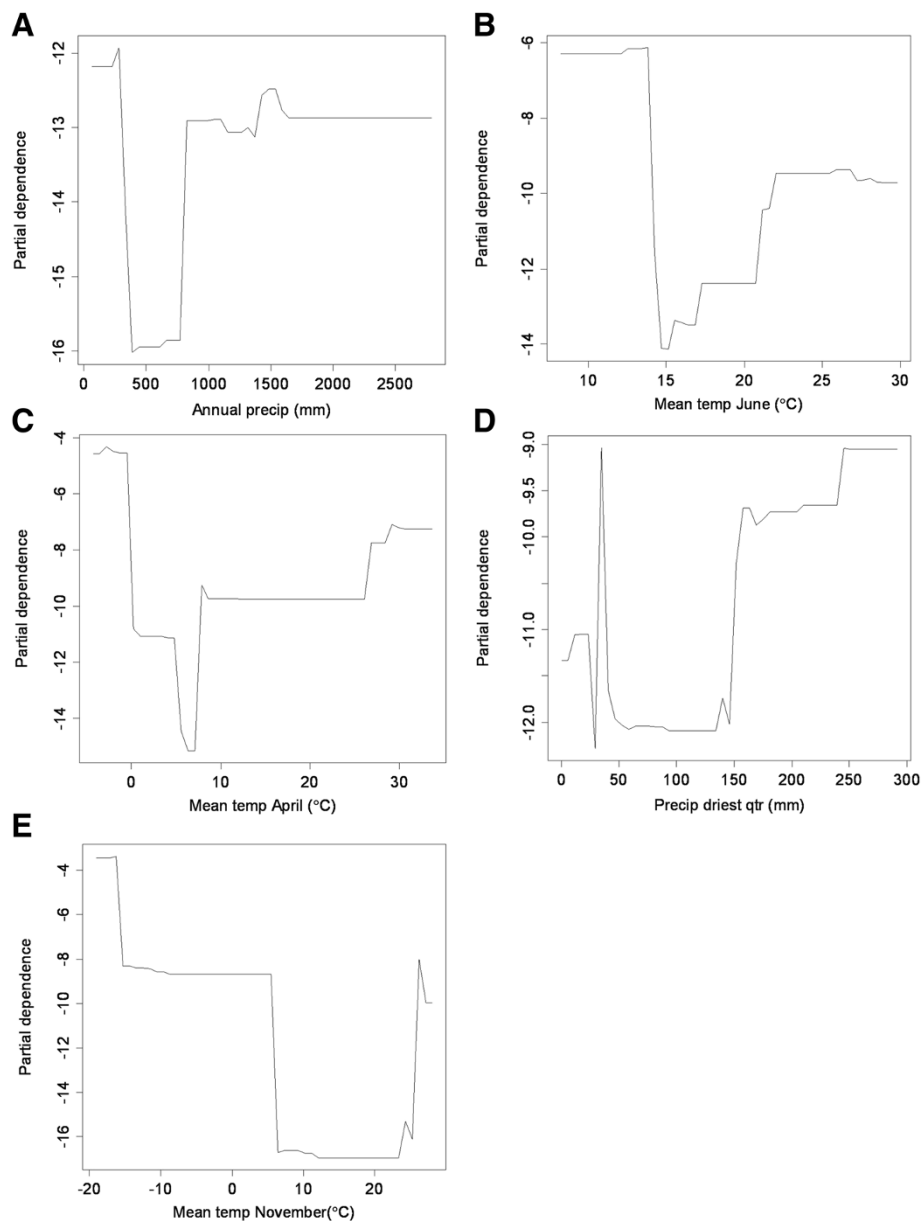


Figure 2 Partial dependence plots for important variables. Partial dependence plots for the variables with the five highest variable importance scores as calculated by Random Forests in the accuracy of the predictive model of avian influenza in wild birds (A-E). Plots show the partial dependence of a high Relative Occurrence Index value for avian influenza on each predictor variable.

highest at the highest range during the driest quarter, which may reflect a low seasonality or variation in rainfall during the year. Mean temperature in November was correlated with AIV-positivity at low and high values (Figures 1e and 2e). Highest partial dependence occurred at the lowest ranges ($< -20^{\circ}\text{C}$). Based on the important predictor variables, the niche of AIV-positive samples in this study was described as regions of low annual rainfall and low temperatures.

There appears to be a secondary niche that described regions of high precipitation and higher temperatures.

Ecological niche model

Random Forests produced a robust ecological niche model for AIV in wild birds and identified important predictor variables. The model had an ROC/AUC of 0.79 on the training points and 0.76 on the testing points, lending high confidence to its prediction of the

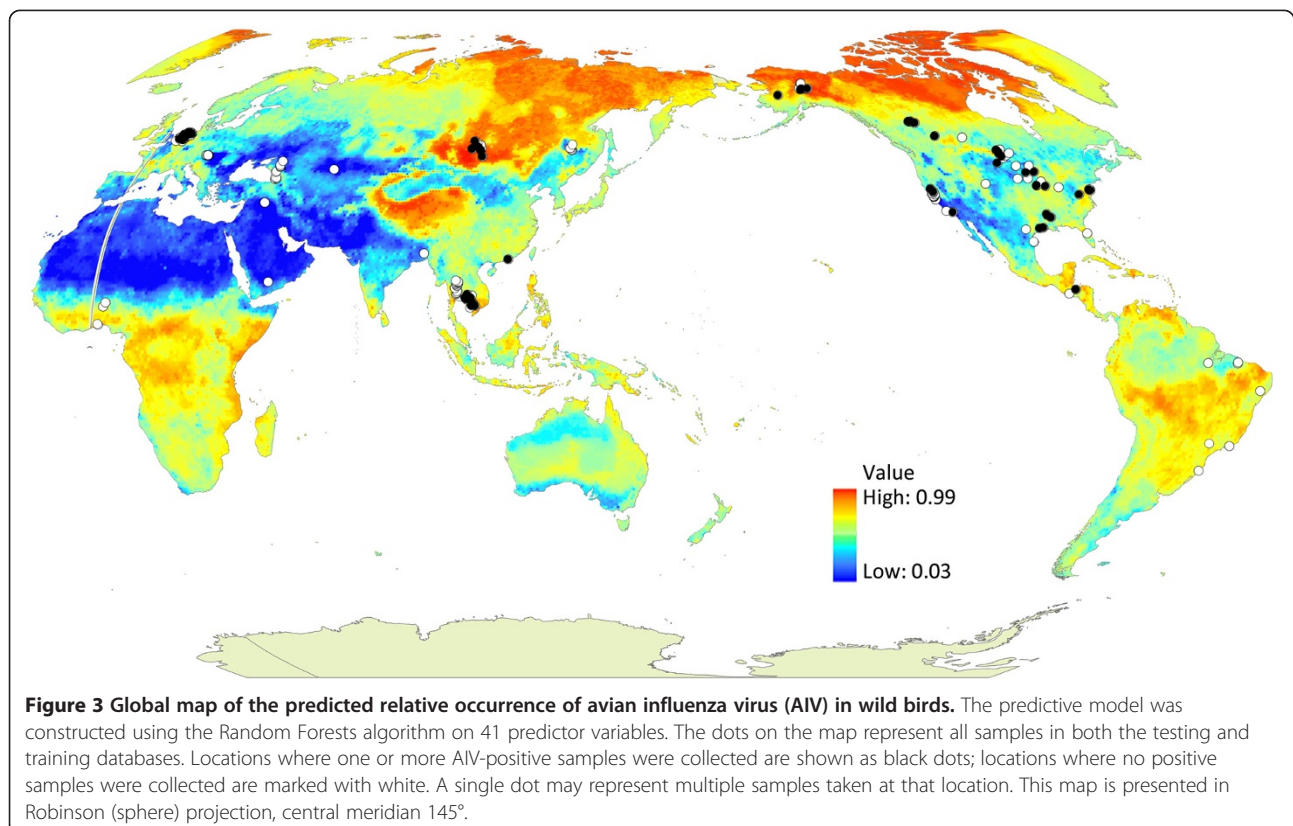
relative occurrence of AIV in wild birds on a global scale. The negative control test was performed by calculating AUC for the training subset against Annual Precipitation (the highest scoring predictor, AUC = 0.59), Mean Temperature in May (the lowest scoring predictor, AUC = 0.47), and Annual Mean Temperature (AUC = 0.47). Although Annual Precipitation received a higher AUC value than the other predictors, this AUC still did not reach the acceptability threshold of 0.7, demonstrating that individual variables were poor predictors of AIV. Northern areas had the highest values of Relative Occurrence Index and temperate regions had the lowest (Figure 3). Interestingly, an equatorial band of relatively high predicted occurrence was observed, which may reflect regions characterized by the secondary niche.

Discussion

While much of AIV modeling has focused on low-latitude regions and HPAI H5N1, we demonstrated that northern regions are important when all strains of AIV and wild reservoir species are taken into account. By creating a global-scale model, we identified important areas of high predicted occurrence that were missed by AIV models for temperate and sub-tropical regions. Small, local models are vital developing strategies for managing acute outbreaks

of specific diseases. However, a global scale perspective is necessary for AIV because, unlike other diseases, is carried by a host that is capable of migrating long distances and potentially infecting others along its path. Furthermore, a model that excludes wild birds, which are the natural reservoir for the virus, neglects the source of gene segments for future infections and potential pandemic strains.

Our model represents the first global-scale predictive map of AIV in wild birds. Using available global AIV data, we identified northern areas as having the highest relative predicted risk of outbreak. Important predictor variables included low temperatures and low annual precipitation. Cold winters and low rainfall may represent continental climates at high latitudes. Areas with these types of climatic conditions include landscapes in Siberia, the Russian Far East, Mongolia, and northern Canada, all of which had high indices of relative occurrence of AIV. Similar conditions at lower latitudes may be created by high elevation, such as the climate of the Tibetan Plateau, which also had a high score. The partial dependence of AIV-positivity on rainfall was bimodal and peaked at very low and high values. This apparently contradictory finding that extremes in rainfall were correlated with AIV-positivity may be explained through laboratory studies of transmission and persistence of the virus. Aerial, non-contact transmission of influenza between guinea pigs was most efficient below 35% relative



humidity [43]; thus, we expect dry climate to be conducive to the aerial spread of virus. At low relative humidity and temperature ($\sim 6^{\circ}\text{C}$, $< 46\%$ rh), virus persisted over two weeks on metal, glass, and in soil [44]. Wet conditions and low temperatures were also conducive to viral persistence: the virus remains viable nearly ten times longer in 17°C water than 28°C water [45]. At low temperatures and high relative humidity ($\sim 7^{\circ}\text{C}$, $\sim 88\%$ rh), the virus persisted over two weeks in chicken feces [44]. Low temperature is the common factor in these studies. While low relative humidity contributes to transmission and persistence on smooth surfaces, the virus also remains viable in water and damp materials such as bird feces. As the virus is transmitted efficiently in water, either through the fecal-oral route [46] or via tracheal shedding [47], dabbling ducks (such as *Anatidae*) in cool northern regions may be at increased risk of contracting AIV from the environment.

Our findings differed from other AIV models in the importance and range of anthropogenic variables. In our model, anthropogenic factors were represented by human population density as well as the Human Influence Index and the Human Footprint Index [32], which are indices calculated based on human population density, land transformation, transportation infrastructure, and electrical power infrastructure. All the anthropogenic variables received very low VIS with human population density scoring the highest at 29.3. Previous models identified high human population density and high farming intensity (especially rice cropping and aquaculture) as important predictors [19,20,48]. The niche they described is characterized as having a high human population, high level of anthropogenic disturbance, and the high annual temperature and humidity of the subtropical climates for which the models were designed (i.e. Bangladesh, Vietnam, and Thailand). However, these studies were specific to HPAI H5N1 in poultry. While the one North American model in wild birds identified low minimum temperatures, with which our model was consistent, they also identified the amount of cropland as an important factor [22]. In general, our model did not predict high occurrence of AIV in the continental United States when compared to northern regions, which have not been modeled previously.

Our model demonstrated a novel use of surveillance data that goes beyond the yearly reporting of infected species and viral subtypes isolated. The application of environmental data, GIS, and machine-learning extends the usefulness of surveillance results. However, the prediction of relative occurrence presented here is not a final, definitive map of avian influenza in wild birds, but rather an initial attempt that demonstrates that a useful signal can be gleaned from the noise found in a global dataset. Indeed, it serves to highlight shortcomings in available data. In particular, nearly all data were collected

in the Northern Hemisphere. In addition, this Northern Hemispheric niche could then be tested on southward-migrating birds to see if the same predictions are applicable. A predominance of *Anatidae* could create a spatial bias for northern regions and a temporal bias for summer months if most sampling is carried out during summer breeding season at high latitudes. However, if one uses the mean temperature in November as a proxy for latitude, there appears instead to be a strong temperate bias in collection with AIV-positive peaks occurring to either side. The bifurcate niche evident here is an interesting topic for future analysis. The mechanisms responsible for this niche require further investigation in order to clarify how the important bioclimatic variables contribute to AIV-positivity.

While ongoing surveillance is important to understanding the dynamics of AIV, efforts should include wilderness areas, such as Siberia, that have received less attention. Models such as this one could receive additional fine-tuning if these results were to guide future sampling efforts in regions of high predicted occurrence, much of which remains unsampled. As both AIV-positive and AIV-negative data are incorporated into this model, all results from prediction-guided sampling strengthen the prediction, even if only a small percentage of AIV-positive samples are isolated. Given the sheer quantity of data collected by long term surveillance efforts, an unprecedented opportunity exists to produce future models of greater accuracy. If data were curated and publically available, models could be treated as transparent, replicable science experiments. Improved global scale models could not only increase the understanding of viral ecology, but also serve to guide the management of influenza risk policy for the benefit of public health on a global scale. A global model of AIV must be a collaborative effort and we hope this initial attempt encourages greater cooperation and data-sharing among members of the AIV research community.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

KH carried out the research and the manuscript preparation. ML provided expertise in data analysis. FH was involved in critical review of the manuscript and gave final approval of the version to be published. All authors read and approved the final manuscript.

Acknowledgements

We thank IRD for providing the wild bird data used in this project. The Influenza Research Database Bioinformatics Resource Center has been wholly funded with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under Contract No. HHSN266200400041C. We thank the open source projects that provided the predictor variables used in this project. We thank Dr Abby Powell and BIOL 644 for helpful comments on the manuscript. KH was partially funded by the Biology & Wildlife Department (UAF). The funding body played no role in data collection or analysis; in the writing of the manuscript; or its decision to submit for application. FH and ML donated their time. This is EWHALE Lab Publication #111.

Author details

¹Ecological Wildlife Habitat Analysis of the Land- and Seascape (EWHALE) Lab, Biology and Wildlife Department, University of Alaska Fairbanks, Fairbanks, AK 99775, USA. ²Institute of Arctic Biology, University of Alaska Fairbanks, Fairbanks, AK 99775-7000, USA. ³Scenarios Network for Alaska and Arctic Planning, University of Alaska Fairbanks, Fairbanks, AK 99775, USA.

Received: 27 August 2012 Accepted: 28 May 2013

Published: 13 June 2013

References

- Garten RJ, Davis CT, Russell CA, Shu B, Lindstrom S, Balish A, Sessions WM, Xu X, Skepner E, Deyde V, Okomo-Adhiambo M, Gubareva L, Barnes J, Smith CB, Emery SL, Hillman MJ, Rivaller P, Smagala J, de Graaf M, Burke DF, Fouchier RA, Pappas C, Alpuche-Aranda CM, López-Gatell H, Olivera H, López I, Myers CA, Faix D, Blair PJ, Yu C, et al: **Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans.** *Science* 2009, **325**:197–201.
- Stallknecht DE, Brown JD: **Wild birds and the epidemiology of avian influenza.** *J Wildl Dis* 2007, **43**:S15–S20.
- Subbarao K, Klimov A, Katz J, Regnery H, Lim W, Hall H, Perdue M, Swayne D, Bender C, Huang J, Hemphill M, Rowe T, Shaw M, Xu X, Fukuda K, Cox N: **Characterization of an avian influenza A (H5N1) virus isolated from a child with a fatal respiratory illness.** *Science* 1998, **279**:393–396.
- Kandun IN, Wibisono H, Sedyaningsih ER, Yusharmen Hadisoedarsuno W, Purba W, Santoso H, Septiawati C, Tresnaningsih E, Heriyanto B, Yuwono D, Harun S, Soeroso S, Giriputra S, Blair PJ, Jeremijenko A, Kosasih H, Putnam SD, Samaan G, Silitonga M, Chan KH, Poon LL, Lim W, Klimov A, Lindstrom S, Guan Y, Donis R, Katz J, Cox N, Peiris M, et al: **Three Indonesian clusters of H5N1 virus infection in 2005.** *N Engl J Med* 2006, **355**:2186–2194.
- Ligon BL: **Avian influenza virus H5N1: a review of its history and information regarding its potential to cause the next pandemic.** *Semin Pediatr Infect Dis* 2005, **16**:236–335.
- Peiris JSM, de Jong MD, Guan Y: **Avian influenza virus (H5N1): a threat to human health.** *Clin Microbiol Rev* 2007, **20**:243–267.
- Centers for Disease Control: **Avian influenza A virus infections of humans.** <http://www.cdc.gov/flu/avian/gen-info/avian-flu-humans.htm>.
- Suarez DL: **Evolution of avian influenza viruses.** *Vet Microbiol* 2000, **74**:15–27.
- Munster VJ, Baas C, Lexmond P, Waldenström J, Wallensten A, Fransson T, Rimmelzwaan GF, Beyer WE, Schutten M, Olsen B, Osterhaus AD, Fouchier RA: **Spatial, temporal, and species variation in prevalence of influenza A viruses in wild migratory birds.** *PLoS Pathog* 2007, **3**:630–638.
- Krauss S, Walker D, Pryor SP, Niles L, Chenghong L, Hinshaw VS, Webster RG: **Influenza A viruses of migrating wild aquatic birds in North America.** *Vector Borne Zoonotic Dis* 2004, **4**:177–189.
- Webster R, Bean W, Gorman O, Chambers T, Kawaoka Y: **Evolution and ecology of influenza A viruses.** *Microbiol Rev* 1992, **56**:152–179.
- Parmley J, Lair S, Leighton FA: **Canada's inter-agency wild bird influenza survey.** *Integr Zool* 2009, **4**:409–417.
- CEIRS: **Centers of Excellence for Influenza Research and Surveillance (CEIRS).** National Institute of Allergy and Infectious Diseases. <http://www.niaid.nih.gov/labsandresources/resources/ceirs/Pages/default.aspx>.
- Peterson AT: **Ecologic niche modeling and spatial patterns of disease transmission.** *Emerg Infect Dis* 2006, **12**:1822–1826.
- Peterson AT, Sanchez-Cordero V, Beard CB, Ramsey JM: **Ecological niche modeling and potential reservoirs for Chagas disease.** *Mexico Emerg Infect Dis* 2002, **8**:662–667.
- Moffett A, Shackelford N, Sarkar S: **Malaria in Africa: vector species' niche models and relative risk maps.** *PLoS One* 2007, **2**:e824.
- Peterson AT, Vieglais DA, Andreasen JK: **Migratory birds modeled as critical transport agents for West Nile Virus in North America.** *Vector Borne Zoonotic Dis* 2003, **3**:27–37.
- Mak S, Morshed M, Henry B: **Ecological niche modeling of lyme disease in British Columbia, Canada.** *J Med Entomol* 2010, **47**:99–105.
- Adhikari D, Chettri A, Barik SK: **Modelling the ecology and distribution of highly pathogenic avian influenza (H5N1) in the Indian subcontinent.** *Curr Sci India* 2009, **97**:72–78.
- Pfeiffer DU, Minh PQ, Martin V, Epprecht M, Otte MJ: **An analysis of the spatial and temporal patterns of highly pathogenic avian influenza occurrence in Vietnam using national surveillance data.** *Vet J* 2007, **174**:302–309.
- Martin V, Pfeiffer DU, Zhou X, Xiao X, Prosser DJ, Guo F, Gilbert M: **Spatial distribution and risk factors of highly pathogenic avian influenza (HPAI) H5N1 in China.** *PLoS Pathog* 2011, **7**:e1001308.
- Fuller TL, Saatchi SS, Curd EE, Toffelmeier E, Thomassen HA, Buermann W, Smith TB: **Mapping the risk of avian influenza in wild birds in the U.S.** *BMC Infect Dis* 2010, **10**:187.
- Herrick KA, Huettmann F, Runstadler J, Chernetsov N, Antonov A, Valchuk O, Gerasimov Y, Matsyna E, Matsyna A, Markovets M, Druzyaka A, Saito K: **Predictive RISK modeling of avian influenza in the Pacific Rim and beyond.** In *Risk Models and Applications, 2010*. Edited by Kremers H, Susini A. Berlin: CODATA Germany: Lecture Notes in Information Sciences; 2010:135–148.
- Hedenström A: **Extreme endurance migration: what is the limit to non-stop flight.** *PLoS Biol* 2010, **8**:e1000362.
- Influenza research database.* <http://www.fludb.org>.
- FAO GeoNetwork: **Predicted global poultry density.** <http://www.fao.org/geonetwork/srv/en/resources.get?id=12720&name=glbpototcor.zip&access=private>.
- FAO GeoNetwork: **Predicted global pig density.** <http://www.fao.org/geonetwork/srv/en/metadata.show?id=12719&currTab=distribution>.
- Beyer HL: **Geospatial modelling environment.** <http://www.spatial ecology.com>.
- Hijmans RJ, Cameron SE, Para JL, Jones P, Jarvis A: **Very high resolution interpolated climate surfaces for global land areas.** *Int J Climatol* 2005, **25**:1965–1978.
- Center for International Earth Science Information Network (CIESIN), Columbia University, Centro Internacional de Agricultura Tropical (CIAT): **Gridded Population of the World Version 3 (GPWv3): Population Density Grids.** <http://sedac.ciesin.columbia.edu/gpw>.
- Robinson TP, Franceschini G, Wint W: **The food and agriculture Organization's gridded livestock of the world.** *Vet Ital* 2007, **43**:745–751.
- Sanderson EW, Jaiteh M, Levy MA, Redford KH, Wannebo AV, Woolmer G: **The human footprint and the last of the wild.** *Bioscience* 2002, **52**:891–904.
- Lehner B, Döll P: **Development and validation of a global database of lakes, reservoirs and wetlands.** *J Hydrol* 2004, **296**:1–22.
- Breiman L: **Random forests.** *Mach Learn* 2001, **45**:5–32.
- Prasad AM, Iverson LR, Liaw A: **Newer classification and regression tree techniques: bagging and random forests for ecological prediction.** *Ecosystems* 2006, **9**:181–199.
- Cutler DR, Edwards TC Jr, Beard KH, Cutler A, Hess KT, Gibson J, Lawler JJ: **Random forests for classification in ecology.** *Ecology* 2007, **88**:2783–2792.
- Liaw A, Wiener M: **Classification and regression by RandomForest R News** 2002, **2**:18–22.
- R Development Core Team: **A language and environment for statistical computing.** Vienna: R Foundation for Statistical Computing; 2010.
- Friedman JH: **Greedy function approximation: a gradient boosting machine.** *Ann Stat* 2001, **29**:1189–1232.
- Hegel TM, Cushman SA, Evans J, Huettmann F: **Current state of the art for statistical modelling of species distributions.** In *Spatial complexity, informatics, and wildlife conservation*. Edited by Cushman SA, Huettmann F. Tokyo: Springer; 2010:273.
- Schröder B: **ROC-Plotting and AUC Calculation Transferability Test v.1.3-7.** Potsdam: Universität Potsdam; 2006.
- Hosmer DW, Lemeshow S: **Applied logistic regression.** New York: Wiley; 2000.
- Lowen AC, Mubareka S, Steel J, Palese P: **Influenza virus transmission is dependent on relative humidity and temperature.** *PLoS Pathog* 2007, **3**:1470–1476.
- Wood J, Choi Y, Chappie D, Rogers J, Kaye I: **Environmental persistence of a highly pathogenic avian influenza (H5N1) virus.** *Environ Sci Technol* 2010, **44**:7515–7520.
- Brown JD, Goekjian G, Poulson R, Valeika S, Stallknecht DE: **Avian influenza virus in water: infectivity is dependent on pH, salinity and temperature.** *Vet Microbiol* 2007, **136**:20–26.
- Webster RG, Yakhno M, Hinshaw VS, Bean WJ, Murti KG: **Intestinal influenza: replication and characterization of influenza viruses in ducks.** *Virology* 1978, **84**:268–278.
- Löndt BZ, Nunez A, Banks J, Nili H, Johnson LK, Alexander DJ: **Pathogenesis of highly pathogenic avian influenza A/turkey/Turkey/1/2005 H5N1 in Pekin ducks (*Anas platyrhynchos*) infected experimentally.** *Avian Pathol* 2008, **37**:619–627.
- Gilbert M, Xiao X, Pfeiffer DU, Epprecht M, Boles S, Czarniecki C, Chaitaweessub P, Kalpravidh W, Minh PQ, Otte MJ, Martin V, Slingenbergh J: **Mapping H5N1 highly pathogenic avian influenza risk in Southeast Asia.** *Proc Natl Acad Sci USA* 2008, **105**:4769–4774.

doi:10.1186/1297-9716-44-42

Cite this article as: Herrick et al.: A global model of avian influenza prediction in wild birds: the importance of northern regions. *Veterinary Research* 2013 **44**:42.