

KNOWLEDGE DISCOVERY FROM MINING THE ASSOCIATION BETWEEN H5N1 OUTBREAKS AND ENVIRONMENTAL FACTORS

Y. L. Si ^{a,b,*}, T. J. Wang ^a, A. K. Skidmore ^a, H. H. T. Prins ^b

^a Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, 7500AA, Enschede, The Netherlands – (yali, tiejun, skidmore)@itc.nl

^b Resource Ecology Group, Wageningen University, 6708 PB Wageningen, The Netherlands – herbert.prins@wur.nl

Commission II

KEY WORDS: H5N1, wild birds, environmental factors, GIS

ABSTRACT:

The global spread of highly pathogenic avian influenza H5N1 in poultry, wild birds and humans, poses a significant zoonotic threat and a serious public health risk. An efficient surveillance and disease control system requires a deep understanding of their spread mechanisms, including environmental factors responsible for the outbreak of the disease. Previous studies suggested that H5N1 viruses occurred under specific environmental circumstances in Asia and Africa. These studies were mainly derived from poultry outbreaks. In Europe, a large number of wild bird outbreaks were reported in west Europe with few or no poultry infections nearby. This distinct outbreak pattern in relation to environmental characteristics, however, has not yet been explored. This research demonstrated the use of logistic regression analyses to examine quantitative associations between anthropogenic and physical environmental factors, and the wild bird H5N1 outbreaks in Europe. A geographic information system is used to visualize and analyze the data. Our results indicate that the H5N1 outbreaks occur in wild birds in Europe under predictable environmental conditions, which are highly correlated with increased NDVI in December, decreased aspect and slope, increased minimum temperature in October and decreased precipitation in January. It suggests that H5N1 outbreaks in wild birds are strongly influenced by food resource availability and facilitated by the increased temperature and the decreased precipitation. We therefore deduce that the H5N1 outbreaks in wild birds in Europe may be mainly caused by contact with wild birds. These findings are of great importance for global surveillance of H5N1 outbreaks in wild birds.

1. INTRODUCTION

The global spread of highly pathogenic avian influenza (HPAI) H5N1 in poultry, wild birds and humans, poses a significant zoonotic threat and a serious public health risk. Since July 2005, the H5N1 virus was detected outside of Asia, appeared in Russia, and then arrived in Romania in October (Gilbert et al. 2006a). In 2006, the disease became pandemic around the Black Sea region, the Mediterranean region, Western Europe and Africa. Recent studies (Kilpatrick et al. 2006; Si et al. 2009) suggested that both wild birds and domestic poultry have played roles in the global dispersion of the H5N1 virus. An efficient surveillance and disease control system requires more understanding of their spread mechanisms, including environmental factors responsible for the outbreak of the disease. In regard to the association with environment, the movement of wild birds generally depends on the availability of food and shelter, which is influenced by physical environmental factors. The movement of domestic poultry mainly relies on anthropogenic environmental factors, which is related to human activities.

Few efforts have been made to investigate the influence of environmental factors on H5N1 outbreaks. The transmission of the HPAI H5N1 in Nigeria and West Africa has been linked to differences in plant phenology and land-surface reflectance (Williams et al. 2008). In Southeast Asia the predictable factor of H5N1 activity was identified based on free-range duck farming and rice-paddy cultivation (Gilbert et al. 2008). And in mainland China, the minimal distance to the national highway, precipitation and the interaction between minimal distance to

the nearest lake and wetland are important predictive variables for the risk of H5N1 (Fang et al. 2008). While the land-use pattern, occurrence of seasonal wetlands, practices of backyard poultry and animal husbandry, and density of human population were identified as risk factors contributed to the transmission of the H5N1 virus in Indian subcontinent (Adhikari et al. 2009). These studies, mainly based on poultry outbreaks, revealed that both anthropogenic and physical environmental factors influence the disease transmission, suggesting that wild birds may facilitate the H5N1 spread in poultry. However, the environmental factors influencing the outbreak of H5N1 in wild birds are poorly understood. Wild bird H5N1 outbreaks may be caused by contact with infected wild birds if the disease pattern is highly correlated with physical environmental factors. Otherwise wild birds may be victims as a result of contact with infected domestic poultry if the disease pattern is more related to anthropogenic environmental factors. In this regard, the detection of key risk factors influencing disease outbreaks in wild birds may help to understand to what extent wild birds can be regarded as the major cause of the H5N1 outbreaks in wild birds.

Surveillance in wild birds tend to be more challenging than in poultry as the population and the movement of wild birds are not clearly known. Derivation of the key environmental indicators influencing the H5N1 spread in wild birds is of great importance for H5N1 surveillance and control. Different from Asia and Africa, in Europe large number of H5N1 outbreaks in wild birds were reported, and with few or no infections in

* Corresponding author

poultry nearby (Si et al. 2009). This distinct pattern provides an opportunity to explore the underlying spread mechanism and the key environmental factors for the disease outbreak.

This study, carried out at the European scale, aims to identify the key risk factors contributing to the outbreak of H5N1 in wild birds. The study will contribute to the knowledge about the spread mechanism of H5N1 virus in wild birds and help in taking necessary precautionary measures against any future disease outbreaks.

2. METHODS AND MATERIALS

2.1 Data

Data on avian influenza H5N1 outbreaks in wild birds in Europe, consisted of 808 confirmed events reported from July 18, 2005 to October 9, 2008, were provided by EMPRES-i: a global animal health information system of FAO's Emergency Prevention Programme for Transboundary Animal Diseases (<http://empres-i.fao.org/empres-i/home>). Environmental data were categorized into anthropogenic and physical environmental subsets, corresponding to two disease spread agents (i.e. poultry and wild birds). Table 1 shows the environmental data sets used for this study. The anthropogenic environmental data sets include urban areas, cities, metropolis, roads, major roads and railways, population density, poultry density. The physical environmental data sets involve DEM, Global Lakes and Wetlands Database, Ramsar sites, potential evapotranspiration, aridity index, precipitation, minimum and maximum temperature and remotely sensed data layers (i.e., monthly NDVI) from 2005 to 2008. Poultry density was classified into anthropogenic environmental factors as it is closely related to human activities. The topographic, wetlands, climatic and NDVI data were selected to stand for physical environmental factors because these conditions are closely related to the availability of food and shelters for wild birds.

2.2 Data pre-processing

The duplicate occurrences of the wild bird H5N1 outbreaks from same localities were discarded, resulting into 320 unique geographic coordinates. A total of 296 locations in 2005 and 2006 were assigned as H5N1 presence data for training the model, and additional 24 locations in 2007 and 2008 were utilized for the validation of the resulting output. To facilitate the use of the logistic regression model, 592 pseudo absent points (i.e. two times the H5N1 presence) were randomly generated within the maximum geographic ranges of H5N1 outbreaks from 2005 to 2008, with a minimum distance of 10 km which equals to the size of the surveillance area in Europe (Pittman and Laddomada 2008). All absence of H5N1 was located outside the 10 km radius buffer of the presence of H5N1 during this period. Figure 1 shows the distribution of presence and pseudo absence of H5N1 outbreaks in wild birds from July 2005 to December 2006.

GIS layers detailing the distance to the nearest urban area, city, metropolis, road, major road, railway, lake and wetland and Ramsar site were generated with a spatial resolution of 1 km. The minimum distances were then extracted from the distance layers for all presence or absence locations. An aspect and a slope layer were also generated based on DEM. Using a smooth algorithm in TIMESAT (Jonsson and Eklundh 2004), the monthly NDVI of 2006 were calculated based on monthly

NDVI layers from 2005 to 2008. Furthermore, using a zonal statistical calculation method, a 10 km mean buffer zone (equals to the size of the surveillance area in Europe) was calculated for the variables: population density, poultry density, DEM, aspect, slope, potential evapotranspiration, aridity index, precipitation, minimum temperature, maximum temperature and NDVI. A total of 63 environmental variables were generated for this study, comprising 10 anthropogenic environmental variables and 53 physical environmental variables respectively (Table 2).

2.3 Statistical analysis

Statistical analyses were performed using the Statistical Package (SPSS Inc, Chicago, IL, USA). Univariate analyses were conducted to examine the effect of each variable separately, reporting odds ratios (OR), 95% confidence intervals (CIs) and *P*-value in the output. Variables yielding nonsignificant changes in log-likelihood were excluded from the further analysis. The multicollinearity and autocorrelation were as well assessed by examining the Variance Inflation Factor (VIF) and Moran's I respectively. Collinearity is present when VIF for at least one independent variable is large. A smaller negative value of Moran's I indicate an increased negative autocorrelation and a smaller positive value reveals an decreased positive autocorrelation; no autocorrelation was detected when the value equals to zero. A subset of variables for identifying key environmental factors were then prepared based on the concern of odds ratios, collinearity and autocorrelation. A stepwise multiple logistic regression with forward entry mode was carried out using the subset of variables and entering each variable accounting for the highest change in the model log-likelihood. This procedure was repeated until no additional significant variable could be added, using a decision rule of $P < 0.01$ for entry and $P > 0.05$ for removal. And odds ratios, 95% confidence intervals (CIs) and *P*-value were reported in the output of the multivariate analysis. The performances of the models were assessed by the Hosmer-Lemeshow goodness of fit test.

Based on the predictive model derived from the multivariate logistic regression analysis, a risk map of wild bird H5N1 outbreaks was generated for a spatial resolution of 1 km. The risk map was then classified into four levels (i.e., very high, high, medium and low risk) for the validation using independent samples (i.e., H5N1 outbreaks in wild birds from 2007 to 2008). A 10 km buffer zone was generated for each validating occurrence and the mean risk value under each buffer zone was calculated. The percentage of buffer zones occurring on the high risk areas was then calculated to evaluate the accuracy of the predictive map.

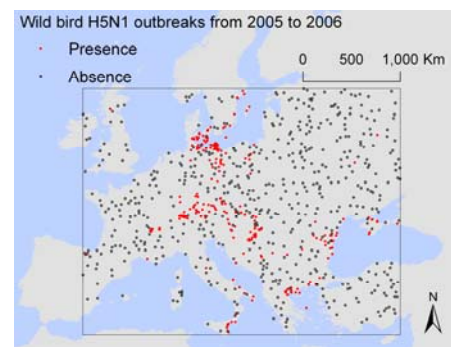


Figure 1. Distribution of presence and pseudo absence of H5N1 outbreaks in wild birds from 2005 to 2006 in Europe

Category	Environmental data sets	Format	Resolution	Data Producer & Source
Anthropogenic environmental data	Urban areas	Polygon	-	ESRI
	Cities	Polygon	-	ESRI
	Metropolis	Polygon	-	ESRI
	Roads	Polyline	-	ESRI
	Major roads	Polyline	-	ESRI
	Railways	Polyline	-	ESRI
	Population density in 2005	Raster	5 km	CIESIN, FAO, CIAT
Poultry density in 2005	Raster	6 km	FAO	
Natural environmental data	DEM	Raster	1 km	WORLDCLIM
	Global Lakes and Wetlands Database	Raster	0.5 km	WWF,ESRI, CESR
	Ramsar sites	Point	-	Wetlands International
	Mean annual potential evapotranspiration	Raster	1 km	CGIAR-CSI
	Mean annual aridity index	Raster	1 km	CGIAR-CSI
	Mean monthly precipitation	Raster	1 km	WORLDCLIM
	Mean monthly maximum temperature	Raster	1 km	WORLDCLIM
	Mean monthly maximum temperature	Raster	1 km	WORLDCLIM
	Monthly MODIS NDVI	Raster	1 km	NASA

Table 1. Environmental data sets used in generating variables for the analysis of H5N1 outbreaks in wild birds in Europe

Category	Environmental variables	Abbreviation	Unit
Anthropogenic environmental variables	Distance to the nearest urban area	Urban	km
	Distance to the nearest city	City	km
	Distance to the nearest metropolis	Metro	km
	Distance to the nearest road	Road	km
	Distance to the nearest major road	Mjroad	km
	Distance to the nearest railway	Railway	km
	Population density in 2005	Popden	p/km ²
	Poultry density in 2005	Poultryden	p/km ²
Natural environmental Variables	Distance to the nearest lake and wetland	GLWD	km
	Distance to the nearest Ramsar site	Ramsar	km
	DEM	DEM	m
	Aspect	Aspect	degree
	Slope	Slope	degree
	Mean annual potential evapotranspiration	Mapet	mm/km ² /year
	Mean annual aridity index	Maaridity	-
	Mean monthly precipitation	PrecJan to Dec	mm
	Mean monthly minimum temperature	TminJan to Dec	degree*10
	Mean monthly maximum temperature	TmaxJan to Dec	degree*10
Monthly MODIS NDVI	NDVIJan to Dec	-	

Table 2. Summary of environmental variables used in the analysis of H5N1 outbreaks in wild birds in Europe

3. RESULTS

The univariate analysis demonstrated that part of physical environmental variables significantly affect the spread of H5N1 in wild birds (Table 3), while no anthropogenic variables were significantly associated with H5N1 outbreaks in wild birds. Positive association were found with the minimum temperature variables, cold-season maximum temperature variables and winter NDVI variables. Negative association were detected

with DEM, aspect, slope and most precipitation variables. Among the associated environmental variables for the H5N1 outbreaks in wild birds, NDVI in December revealed the strongest influence, with an odds ratio of 12.915 reported.

Seven physical environmental variables (i.e., aspect, slope, mean monthly precipitation in January, mean monthly minimum temperature in February and October, monthly MODIS NDVI in March and December) were selected as the

inputs of the multivariate stepwise logistic regression. All selected variables showed a low to medium degree of autocorrelation and the VIF values are lower than 10 (Table 4). Multivariate logistic regression demonstrated that five variables, aspect, slope, mean monthly minimum temperature in October, mean monthly precipitation in January and monthly NDVI in December, were significantly associated with H5N1 outbreaks in wild birds (Table 4). H5N1 outbreaks in wild birds tend to occur in areas with increased NDVI in December, decreased aspect and slope, increased minimum temperature in October and decreased precipitation in January. The regression model adequately fits the data, evaluated by the Hosmer and Lemeshow goodness test ($X^2 = 8.596$, $P = 0.378$).

Variables	Univariate analysis		
	OR	95% CIs	P-value
DEM	0.998	0.998 ~ 0.999	<0.001
Aspect	0.99	0.985 ~ 0.995	<0.001
Slope	0.843	0.778 ~ 0.913	<0.001
PrecJan	0.988	0.982 ~ 0.995	0.001
PrecFeb	0.988	0.980 ~ 0.996	0.002
PrecMar	0.99	0.982 ~ 0.998	0.016
PrecApr	0.987	0.978 ~ 0.997	0.009
PrecAug	1.006	1.001 ~ 1.011	0.031
PrecOct	0.992	0.985 ~ 0.998	0.014
PrecDec	0.989	0.983 ~ 0.996	0.001
TminJan	1.012	1.009 ~ 1.016	<0.001
TminFeb	1.012	1.008 ~ 1.016	<0.001
TminMar	1.016	1.011 ~ 1.021	<0.001
TminApr	1.016	1.010 ~ 1.023	<0.001
TminMay	1.015	1.009 ~ 1.022	<0.001
TminJun	1.018	1.011 ~ 1.024	<0.001
TminJul	1.016	1.010 ~ 1.022	<0.001
TminAug	1.017	1.011 ~ 1.023	<0.001
TminSep	1.019	1.013 ~ 1.025	<0.001
TminOct	1.019	1.013 ~ 1.025	<0.001
TminNov	1.016	1.011 ~ 1.021	<0.001
TminDec	1.014	1.010 ~ 1.019	<0.001
TmaxJan	1.009	1.005 ~ 1.012	<0.001
TmaxFeb	1.009	1.005 ~ 1.012	<0.001
TmaxMar	1.01	1.006 ~ 1.014	<0.001
TmaxApr	1.006	1.001 ~ 1.011	0.013
TmaxSep	1.006	1.002 ~ 1.011	0.003
TmaxOct	1.008	1.004 ~ 1.012	<0.001
TmaxNov	1.008	1.004 ~ 1.012	<0.001
TmaxDec	1.009	1.005 ~ 1.012	<0.001
NDVIJan	5.599	1.958 ~ 16.010	0.001
NDVIFeb	5.437	1.897 ~ 15.585	0.002
NDVIMar	6.415	1.962 ~ 21.210	0.002
NDVINov	10.243	2.595 ~ 40.430	<0.001
NDVIDec	12.915	3.576 ~ 46.646	<0.001

Table 3 Significant variables associated with H5N1 outbreaks in wild birds reported by the univariate logistic regression analysis

A predictive risk map of H5N1 presence in wild birds in Europe was generated based on the predictive model derived from the logistic regression analysis. The validating samples were overlapped on the risk map, showing 79% of H5N1 outbreaks in wild birds (19/24) appeared in the predictive high and very high risk areas (i.e., predictive risk > 0.4) (Figure 2).

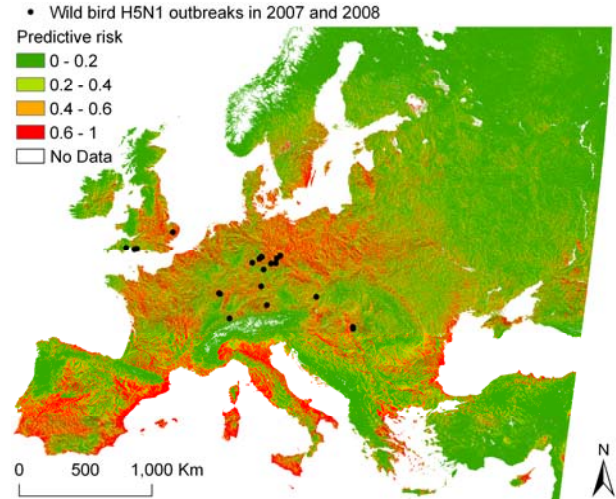


Figure 2. Predictive risk map of H5N1 outbreaks in wild birds in Europe

4. DISCUSSION

The results presented in this paper highlight two main findings: (i) H5N1 outbreaks in wild birds in Europe occur under consistent and predictable environmental circumstances. The key environmental factors influencing the presence of H5N1 outbreaks in wild birds are an increased NDVI in December, decreased aspect and slope (i.e., flat areas with low relief), increased minimum temperature in October and decreased precipitation in January. (ii) H5N1 outbreaks in wild birds in Europe may be mainly caused by contact with infected wild birds as the spatial distribution of wild bird H5N1 outbreaks strongly correspond to most physical environmental factors, but wild bird H5N1 outbreaks are not related to any anthropogenic environmental factors.

The slope and aspect showed consistently negative association with wild bird H5N1 outbreaks in Europe. And consistently positive association with NDVI and temperature during the winter was found, when large number of wild birds overwintering and staging in Europe. As wetlands, rivers, canals, ponds and irrigated networks are concentrated in lowlands, flat plains, deltas, and coastal areas (Gilbert et al. 2006b; Gilbert et al. 2008), the combination of increased winter NDVI and increased winter temperature with flat areas indicates a circumstance of increased food resources for wild birds. The availability of food may influence the movement of wild birds and hence influence the H5N1 outbreaks in wild birds. However, the risk of H5N1 infections is not increased in areas closer to wetlands. One explanation for this result may be the extensive distribution of wetlands in Europe.

This study also highlights that climatic factors significantly contribute to H5N1 outbreaks in wild birds. In agreement with a recent finding in mainland China (Fang et al. 2008), precipitation was found here to be negatively associated with

the risk of H5N1 outbreaks in wild birds, possibly because the lower precipitation leads to a higher concentration of birds in limited suitable habitats, and therefore increased opportunities for contact. Areas with higher temperatures tend to have higher risk of disease outbreaks because temperature can stimulate the viral activity. This explains why both monthly minimum and maximum temperature showed positive associations with H5N1 outbreaks.

The univariate logistic analysis in this study demonstrated that the risk of H5N1 infections in wild birds in Europe is not influenced by anthropogenic environmental factors but only by physical environmental factors, indicating that these outbreaks may be mainly caused by wild birds contact with infected wild birds. This is supported by the previous findings that wild birds are capable of excreting abundant viruses (e.g., in their faeces)

before and after the onset of clinical signs (Keawcharoen et al. 2008), or even asymptotically (Chen et al. 2006), and they are suspected to spread the virus over either long and short distances (Si et al. 2009).

Based on the five principal environmental factors identified by multivariate logistic analysis, a predictive risk map was generated, targeting important surveillance areas in Europe. Independent validation of the risk map showed a moderate accuracy of the predictive model, suggesting the key environmental factors identified are consistently affecting the outbreak of H5N1 in wild birds (Peterson and Williams 2008). As temperature and precipitation data are available and of good quality, improved food availability data could be used to enhance the accuracy of risk prediction on H5N1 outbreaks in wild birds in Europe.

Selected	Multivariate analysis				Moran I	VIF
	B	Odds ratio	95% CIs	P-value	P<0.01	
Aspect	-0.010	0.991	0.985 ~ 0.996	0.001	0.090	1.307
Slope	-0.140	0.869	0.780 ~ 0.969	0.011	0.350	1.502
PrecJan	-0.036	0.965	0.954 ~ 0.976	<0.001	0.360	1.822
TminOct	0.031	1.032	1.024 ~ 1.039	<0.001	0.470	4.272
NDVIDec	5.098	163.765	30.264 ~ 886.171	<0.001	0.510	1.799
Constant	-1.683	0.186	-	0.006	-	-

Table 4 Selected variables by multivariate stepwise logistic regression model for H5N1 outbreaks in wild birds in Europe from 2005 to 2006. Moran's I and VIF indicate lower level of autocorrelations and collinearity among variables

5. CONCLUSIONS

The key environmental factors influencing the presence of H5N1 outbreaks in wild birds in Europe are increased NDVI in December, decreased aspect and slope, increased minimum temperature in October and decreased precipitation in January. It suggests that H5N1 outbreaks in wild birds are strongly influenced by the availability of food resources, and facilitated by the increased temperature and the decreased precipitation. We therefore deduce that H5N1 outbreaks in wild birds in Europe may be mainly caused by contact with infected wild birds. These findings are of great importance for global surveillance of H5N1 outbreaks in wild birds.

REFERENCE

Adhikari, D., Chettri, A. and Barik, S. K. 2009. Modelling the ecology and distribution of highly pathogenic avian influenza (H5N1) in the Indian subcontinent. *CURRENT SCIENCE*, 97(1), pp. 73-78.

Chen, H., Li, Y., Li, Z., Shi, J., Shinya, K., Deng, G., Qi, Q., Tian, G., Fan, S., Zhao, H., Sun, Y. and Kawaoka, Y. 2006. Properties and Dissemination of H5N1 Viruses Isolated during an Influenza Outbreak in Migratory Waterfowl in Western China. *J Virol*, 80(12), pp. 5976-5983.

Fang, L.-Q., de Vlas, S. J., Liang, S., Looman, C. W. N., Gong, P., Xu, B., Yan, L., Yang, H., Richardus, J. H. and Cao, W.-C. 2008. Environmental Factors Contributing to the Spread of H5N1 Avian Influenza in Mainland China. *PLoS ONE*, 3(5), pp. e2268.

Gilbert, M., Chaitaweesub, P., Parakamawongsa, T., Premasithira, S., Tiensin, T., Kalpravidh, W., Wagner, H. and Slingenbergh, J. 2006b. Free-grazing Ducks and Highly Pathogenic Avian Influenza, Thailand. *Emerg Infect Dis*, 12(2), pp. 227-234.

Gilbert, M., Xiao, X., Domenech, J., Lubroth, J., Martin, V. and Slingenbergh, J. 2006a. Anatidae Migration in the Western Palearctic and Spread of Highly Pathogenic Avian Influenza H5N1 Virus. *Emerg Infect Dis* 12(11), pp. 1650-1656.

Gilbert, M., Xiao, X., Pfeiffer, D. U., Epprecht, M., Boles, S., Czarnecki, C., Chaitaweesub, P., Kalpravidh, W., Minh, P. Q., Otte, M. J., Martin, V. and Slingenbergh, J. 2008. Mapping H5N1 highly pathogenic avian influenza risk in Southeast Asia. *PNAS*, 105(12), pp. 4769-4774.

Jonsson, P. and Eklundh, L. 2004. TIMESAT – a program for analyzing time-series of satellite sensor data. *COMPUT GEOSCI*, 30, pp. 833-845.

Keawcharoen, J., van Riel, D., Amerongen, G. v., Bestebroer, T., Beyer, W. E., van Lavieren, R., Osterhaus, A. D. M. E., Fouchier, R. A. M. and Kuiken, T. 2008. Wild Ducks as Long-Distance Vectors of Highly Pathogenic Avian Influenza Virus (H5N1). *Emerg Infect Dis*, 14(4), pp. 600-607.

Kilpatrick, A. M., Chmura, A. A., Gibbons, D. W., Fleischer, R. C., Marra, P. P. and Daszak, P. 2006. Predicting the global spread of H5N1 avian influenza. *PNAS*, 103(51), pp. 19368-19373.

Peterson, A. T. and Williams, R. A. J. 2008. Risk Mapping of Highly Pathogenic Avian Influenza Distribution and Spread. *Conserv Ecol*, 13(2), pp. 15.

Pittman, M. and Laddomada, A. 2008. Legislation for the Control of Avian Influenza in the European Union. *Zoonoses and Public Health*, 55, pp. 29-36.

Si, Y., Skidmore, A. K., Wang, T., de Boer, W. F., Debba, P., Toxopeus, A. G., Li, L. and Prins, H. H. T. 2009. Spatio-temporal dynamics of global H5N1 outbreaks match bird migration patterns. *Geospat Health*, 4(1), pp. 65-78.

Williams, R. A. J., Fasina, F. O. and Peterson, A. T. 2008. Predictable ecology and geography of avian influenza(H5N1) transmission in Nigeria and West Africa. *Trans R Soc Trop Med Hyg* 102, pp. 471-479.

ACKNOWLEDGEMENTS

This research is sponsored by the China Scholarship Council (CSC) and the International Institute for Geo-Information Science and Earth Observation (ITC), the Netherlands. The authors thank FAO's Emergency Prevention Programme for Transboundary Animal Diseases (EMPRES-i) for providing the H5N1 outbreak data.