

Original Contribution

Ecological Modeling of the Spatial Distribution of Wild Waterbirds to Identify the Main Areas Where Avian Influenza Viruses are Circulating in the Inner Niger Delta, Mali

Julien Cappelle,¹ Olivier Girard,² Bouba Fofana,^{3,4} Nicolas Gaidet,¹ and Marius Gilbert^{5,6}

¹CIRAD ES, UR Animal et Gestion intégrée des risques (AGIR), TA C 22/E, Campus International de Baillarguet, 34398 Montpellier Cedex 5, France

²ONCFS, Réserve de Chanteloup, 85340 l'Île d'Olonne, France

³Direction Nationale des Eaux et Forêts du Mali, BP 275 Bamako, Mali

⁴Wetlands International, BP 97 Mopti-Sevare, Mali

⁵Biological Control and Spatial Ecology, Université Libre de Bruxelles, av FD Roosevelt, 50, 1050 Brussels, Belgium

⁶Fonds National de la Recherche Scientifiques, rue d'Egmont 5, 1000 Brussels, Belgium

Abstract: Predicting areas of disease emergence when no epidemiological data is available is essential for the implementation of efficient surveillance programs. The Inner Niger Delta (IND) in Mali is a major African wetland where >1 million Palearctic and African waterbirds congregate. Waterbirds are the main reservoir of Avian Influenza Viruses (AIV). Our objective was to model their spatial distribution in order to predict where these viruses would be more likely to circulate. We developed a generalized linear model (GLM) and a boosted regression trees (BRT) model based on total aerial bird counts taken in winter over 6 years. We used remotely sensed environmental variables with a high temporal resolution (10 days) to predict the spatial distribution of four waterbird groups. The predicted waterbird abundances were weighted with an epidemiological indicator based on the prevalence of low pathogenic AIV reported in the literature. The BRT model had the best predictive power and allowed prediction of the high variability of waterbird distribution. Years with low flood levels showed areas with a higher risk of circulation and had better spatial distribution predictions. Each year, the model identified a few areas with a higher risk of AIV circulation. This model can be applied every 10 days to evaluate the risk of AIV emergence in wild waterbirds. By taking into account the IND's ecological variability, it allows better targeting of areas considered for surveillance. This could enhance the control of emerging diseases at a local and regional scale, especially when resources available for surveillance programs are scarce.

Keywords: emerging infectious diseases, surveillance, wild birds, distribution models, AIV, boosted regression trees

INTRODUCTION

Emerging infectious diseases are a growing concern because of their impact on public health, global economies, and

wildlife species (Morens et al., 2004). Although several ecological, environmental, and socioeconomic factors driving the emergence of infectious diseases have been identified at a global scale (Morse, 1995; Morens et al., 2008), it remains difficult to forecast and quickly detect emergences. A better allocation of surveillance efforts is needed, especially for wildlife zoonotic and vector-borne emerging infectious

Correspondence to: Julien Cappelle, e-mail: julien.cappelle@cirad.fr

diseases originating at lower latitudes (Jones et al., 2008). Compared to standard retrospective epidemiological studies, the main challenge in trying to predict disease emergence is a lack of epidemiological data. One consequently must turn to knowledge-based approaches combining environmental and socioeconomical data to derive predictions (Smith et al., 2005). These approaches then can be used to determine high-risk areas (Cumming et al., 2008; Brochet et al., 2009), prioritize surveillance programs, and plan interventions (Clements and Pfeiffer, 2009).

Avian influenza viruses (AIV) endemically circulating in wild waterbirds (Webster et al., 1992) are particularly prone to evolve and emerge in highly pathogenic forms, termed highly pathogenic AIV (HPAIV), that are responsible for high mortality in domestic poultry, and can occasionally transmit to humans (Koopmans et al., 2004). Both low and highly pathogenic AIV have been found circulating in wild waterbirds in West African wetlands (Gaidet et al., 2007, 2008), and the Asian H5N1 HPAIV spread to West Africa in 2006 where it since has persisted and reemerged in several countries (Fusaro et al., 2009).

The Inner Niger Delta (IND) in Mali is the largest continental wetland in West Africa, and the second largest in Africa. Stretching over 41,195 km² in the midst of the Sahelian zone, this low elevation floodplain area includes a number of seasonally inundated lakes, ponds, and river channels. It is a key West African wetland and supports many waterbird species, including up to one million migrating Palearctic ducks in the northern winter, 100,000 Afro-tropical ducks, and 300,000 waders (Girard, 2006). The ecology of the area is mainly driven by the flood level, which itself depends on rainfalls in the region (Zwarts and Grigoras, 2005). This flood level is quantified by the water level of the Niger River at a reference point (Akka, 15°24' N, 4°14' W) and correlates to the extent of the water cover that provides suitable habitats for waterbirds. Since 1970, the maximum flood level at Akka has ranged between 336 cm in 1984 and 534 cm in 1994, leading to a maximum flooded area ranging between 1500 and 20,000 km². The area is flooded between August and December, decreasing water levels in the weeks and months that follow leads to aggregations of waterbirds on the remnant water bodies. The mixing of wild waterbirds coming from the Palearctic region and Africa provides favorable conditions for AIV transmission, making the IND a potential hub for the spread of new AIV variants.

The flood level varies considerably within and between years (Zwarts et al., 2009), and this naturally influences the

spatial distribution of wild waterbirds and of the possible areas of increased AIV transmission. The number of birds present in the IND is also highly variable depending on the year. For example, aerial global censuses of wild waterbird species undertaken by one ornithologist have shown that the total number of Garganeys in the IND varied from about 815,000 to 225,000 between two successive years (Trollet et al., 2008). Therefore, the prediction of wild waterbird abundance is challenging, and a model with a good spatial and temporal resolution is needed to capture this high variability.

This article aimed to model the spatial distribution of the main wild reservoir of AIV in the IND in order to predict areas where AIV may be expected to circulate more intensively and where surveillance should be prioritized. We used a set of environmental variables and different modeling approaches to predict the spatial distribution of four waterbirds groups with different levels of AIV prevalence. Our models aimed to be both spatial and temporal to best capture the temporal variability of the IND's ecology with a good spatial resolution. We combined results from our spatial distribution model with measurements of AIV prevalence reported in the literature, to predict which areas would have a higher risk of AIV circulation for different 10-day periods.

METHODS

Ornithological Data

The ornithological data were collected from six aerial censuses performed in January 1999–2001, 2006–2008. Counts were carried out by the same observer from a small aircraft flying at 100–300 feet above the ground. Each year, an approximate total of 40 h was spent flying over the IND. The entire Delta was divided into 171 (10-minute-side) cells which, at this latitude, were equivalent to ca. 18 km × 18 km. Every year, a certain number cells surveyed—generally ones situated at the periphery of the area—were found to have dried out entirely. These either were overflown rapidly by the plane or ignored. For each year, the census was reported by cell and by species. More details on the aerial counts may be found in Girard et al. (2004). All of the most abundant waterbird species (ducks, herons, and large waders) were counted with the exception of rails and small waders, which could not be detected due to their smaller body size and inconspicuous behavior. In general, aerial censuses allow an estimation of the total

population of an area but lack accuracy, especially when counting large groups of gregarious birds (Zwarts et al., 2009).

We classified the species counted into four systematic groups according to the role they may play in AIV circulation. The species were grouped based on the AIV prevalence recorded in the literature on avian influenza in wild waterbirds (Olsen et al., 2006; Munster et al., 2007). Palearctic Anatids, mainly represented by Garganeys (*Anas querquedula*) and Pintails (*Anas acuta*), i.e., dabbling ducks recognized as major reservoirs of AIV, constituted the first group. The second group included Afro-tropical Anatids that have shown lower AIV prevalence than Palearctic Anatids (Gaidet et al., 2007). Waders and Ardeids constituted the two remaining groups. A principal component analysis (PCA) was run with the abundance data of all of the species counted over the 6 years to assess the validity of this systematic species grouping as a spatial distribution model dataset.

Environmental Variables

We compiled a total of eight environmental variables (Table 1) related to three main environmental indicators to predict the distribution of wild waterbirds: water indices indicative of favorable habitats, vegetation indices indicative of favorable feeding areas, and indicators of human disturbance. To deal with the highly variable habitats of the IND, we used environmental data with a high temporal resolution. We used the water level of the Niger River at a reference point (Akka) and remotely sensed environmental indicators compiled over 10 days matching the aerial census period of each year. The remote sensing products were acquired by the

vegetation instruments onboard SPOT satellites and were downloaded from the Vgt4Africa project website (Vgt4Africa, 2006). We used the Normalized Difference Vegetation Index (NDVI) (Reed et al., 1994) and the Dry Matter Production (DMP) derived from the net primary productivity (Chen et al., 1999) as vegetation indices. The Normalized Difference Water Index (NDWI) (Gao, 1996) and the Small Water Bodies (SWB) (Gond et al., 2004) product were used as indicators of humidity and water. These four remotely sensed environmental indicators were originally provided with a spatial resolution of 1 km and a temporal resolution of 10 days. We also used the human population density as an indicator of disturbance for birds based on a 1998 population census with a spatial resolution of 5 km (CIESIN, 2005), and a digital flooding model of the IND based on 30-m spatial resolution Landsat images (Zwarts and Grigoras, 2005). These two variables were not adjusted to match either the dates or the year of the census data. All of these environmental variables were aggregated at the cell level (ca. 18 km × 18 km). For NDVI, NDWI, DMP, population density, and elevation, we calculated the mean value of all pixels of the variable included in each of the cells for the 10-day period matching the aerial census. In addition, we calculated the total number of 1-km pixels of each class (Free Water, Humid area, and Free Water + Humid area) of the SWB indicator in each of the cells. For each census period, we obtained a dataset including an aggregated value of each environmental variable for the 171 cells of the IND.

Waterbird Distribution Modeling

Our aerial census data showed evidence of zero-inflated distribution with up to 75.8% of empty cells, while the

Table 1. Explanatory variables used in the different models

Variable	Indicator type	Source	Temporal resolution	Spatial resolution	Aggregation at cell level
NDVI	Vegetation density	Spot satellite	10 days	1 km	Mean
DMP	Vegetation production	Spot satellite	10 days	1 km	Mean
NDWI	Humidity	Spot satellite	10 days	1 km	Mean
SWB	Land use (water)	Spot satellite	10 days	1 km	Count
Water level	Water level	Field data	Day	Inner Niger Delta	–
Max. water level	Maximum water level	Field data	Year	Inner Niger Delta	–
Human density	Population density	GRUMP	–	5 km	Mean
DFM	Elevation	Zwarts and Grigoras (2005)	–	30 m	Mean

NDVI Normalized Difference Vegetation Index; DMP Dry Matter Production; NDWI Normalized Difference Water Index; SWB Small Water Bodies; DFM Digital Flooding Model (Zwarts and Grigoras, 2005); GRUMP Global Rural–Urban Mapping Project.

distribution of the remaining data was assumed to be log-normal. Modeling zero-inflated distribution can be achieved by building and combining two models (Barry and Welsh, 2002): first, a presence–absence model assuming a binomial distribution, and second, an abundance-if-presence model assuming a Gaussian distribution with the log-transformed data. The final abundance prediction then is calculated as the product of the probability of presence given by the first model and the predicted abundance given by the second model. Many methods have been developed to predict species distribution (Guisan and Zimmermann, 2000; Elith et al., 2006) and there is no agreement regarding which is the best statistical method for distribution modeling or which may be applicable to all ecological systems. As our main objective was to build robust predictions, we decided to build two different types of models and to select the one with the best predictive power (Pearce and Ferrier, 2000). We implemented a generalized linear model (GLM) (McCullagh and Nelder, 1989) and a boosted regression trees (BRT) model (Elith et al., 2008). The GLM has been applied widely in ecology and can be used as a reference (Guisan and Zimmermann, 2000). The BRT model was used because it reportedly provides better predictions than generalized models (Elith et al., 2006) and is said to better handle interactions and nonlinear relationships. We also chose the BRT model because inferring the role of the environmental variables was not our primary objective. The response variable was the presence of a bird group for the presence–absence model, and the \log_{10} -transformed abundance of a bird group for the abundance-if-presence model. We considered the impact of the interannual variability of bird abundances by standardizing them to the mean abundance estimated over the 6 census years.

The GLM included all of the environmental variables as explanatory variables and an autocovariate term to account for spatial autocorrelation (Augustin et al., 1996). The autocovariate term was a weighted average of the number of occupied cells or of the abundance of the cell in the range of spatial autocorrelation. The weight of each surrounding cell was estimated as the inverse of the Euclidian distance to the considered cell. This model was implemented in R software (R Development Core Team, 2009).

We implemented the BRT model using the functions developed by Elith et al. (2008). This type of model combines many regression trees using a boosting method to improve the predictive performance of the final model. Stochasticity can be introduced into BRT models by randomly selecting a fraction of the total data to fit each of the

regression trees. This fraction, called bag fraction, was fitted to 0.75 after showing no significant differences in the predictions for values ranging from 0.5 to 0.75. Optimization of BRT models requires fitting three main parameters: tree complexity, learning rate, and number of trees. Tree complexity is defined as the number of nodes in each regression tree; the learning rate is defined as the contribution of each regression tree in the final model; and the number of trees is defined as the total number of regression trees included in the boosting process. We used a tree complexity of five because we expected several interactions between our variables, especially between vegetation and water indicators. High values of vegetation variables could indicate a favorable habitat for waterbirds when associated with high values of water variable. On the other hand, they also could be indicating a nonfavorable habitat when associated with low values of water variables. The learning rate of the BRT model was fitted to 0.001 to have at least 1,000 trees fitted for each run of the model. The number of trees to be fitted for each model was identified using a cross-validation procedure described by Hastie et al. (2001) and implemented in the BRT functions in R. The BRT model was run 25 times for each prediction.

Model Selection and Validation

For each model (GLM and BRT), we ran six different data sets to predict the abundance of each of the four bird groups. Each data set was constituted by a combination of the results of five out of the six census years. The results from the remaining year were used to evaluate the predictions. To determine the best model, we calculated two different indicators. For the presence–absence model, we calculated the area under the receiver operating characteristic (ROC) curve (AUC) as implemented in Elith et al. (2008). The ROC curve is obtained by plotting sensitivity vs. $(1 - \text{specificity})$ for varying probability thresholds. High performance models are indicated by large areas under the curve (Manel et al., 2001). We also estimated the percentage of deviance explained D^2 for the presence–absence predictions and for the final abundance predictions. These performance indicators were calculated for each of the four bird groups and the six datasets, providing an overall mean and a standard deviation of the performance metrics. To improve the models, we applied a mask on the final results for each bird group. The mask was constituted with the cells where no bird of the group was recorded during the six census years.

Potential Virus Circulation

The statistical models provided a predicted spatial distribution for the four bird groups for each year. An indicator of AIV circulation risk was estimated by weighting the predicted abundance of each group with a coefficient corresponding to the global prevalence of AIV measured for these species. As little information was available on avian influenza prevalence in wild birds in Africa (Gaidet et al., 2007), we also used prevalence measured in other areas for Palearctic Anatids and waders (Olsen et al., 2006; Munster et al., 2007), and for Ardeids (Siembieda et al., 2010) to inform estimations of prevalence coefficients. We used a coefficient of 0.06 for the migratory Palearctic Anatids, 0.03 for the Afro-tropical Anatids, 0.01 for the waders, and 0.005 for the Ardeids. Finally, for each cell of the IND, we added the values of the risk indicator of the four bird groups to obtain maps of potential virus circulation.

As explained above, we trained the model with six different data sets constituted by different census year data combinations to assess the variability of the potential AIV circulation predictions. As our predictors were explicit in time, the models could be run in real time for January 2010. We made the prediction for 2010 with the environmental indicators recorded from January 11–20, 2010.

RESULTS

A total of 16 species were classified in four systematic groups according to the role they may play in AIV circulation (Appendix, Table 4). This grouping was validated as a spatial distribution model dataset by the results of the PCA. It showed that the species from each group constituted clusters (Fig. 1), and that 55.43% of the total variance was supported by the first two components of the PCA.

The BRT model showed the best predictive power for the presence–absence predictions; however, as the standard errors were high, the AUC values of the GLM and BRT models were of the same order (Table 2). The BRT model also provided the best predictive power for the final predictions of abundance with a mean D^2 of 0.372 (SE = 0.151) (Table 2). The high standard errors (coefficient of variation = 40.6% for the BRT model) are related to the high variability of the predictive power for each of the 24 distributions (one per group and per year). However, the BRT model had the best predictive power for 23 predictions out of 24. The application of the mask increased the

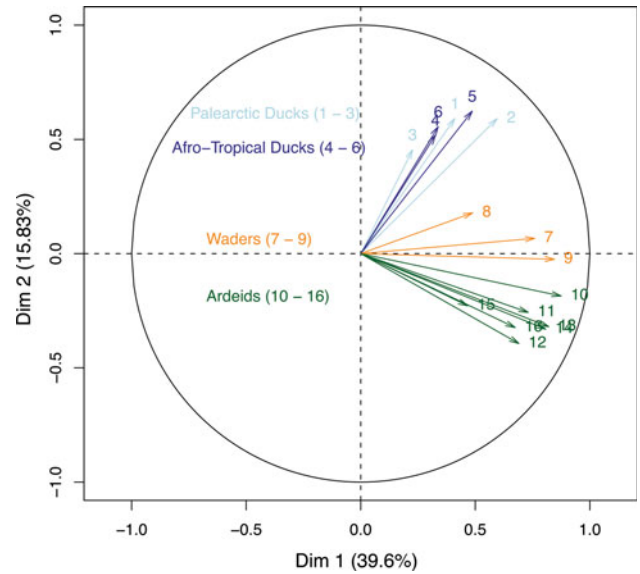


Figure 1. Plot of the first two components of the PCA run on the aerial census data of the wild waterbird species. Each arrow corresponds to a single species (see Appendix, Table 4, for the list of the species). The four colors indicate the four epidemiological bird groups (Color figure online).

mean D^2 of both models (Table 2). The effect of the mask was more important for the GLM (13.3%) than for the BRT model (5.1%), and it was more important for gregarious species (increase of the D^2 ranging from 7% to 60% for Palearctic Anatids with the BRT model) than for nongregarious species (increase of the D^2 ranging from 1% to 3% for Ardeids with the BRT model).

As shown for the BRT model (Table 3), the predictive power of the models was dependent on the bird group and the year. Overall the predictive power was better and less variable for the Ardeids group (mean D^2 = 0.509, SE = 0.132, coefficient of variation = 26%), than for the Palearctic Anatids group (mean D^2 = 0.345, SE = 0.155, coefficient of variation = 45%). Results were intermediate for Afro-tropical Anatids and waders. Predicted spatial distributions for the Ardeids were similar for the BRT model and for the observed data (Fig. 2). For the spatial distribution of the Palearctic Anatids, the model accurately predicted the general pattern of the distribution variability, with an aggregation of the birds in the central part of the Delta during years with low flood levels (Fig. 2). A map of the model residuals, i.e., the difference between observed and predicted abundances, is available in the Appendix, Figure 5. The predictions were significantly more accurate for the years with lower flood levels (Fig. 3).

Table 2. Predictive performance of the models

	AUC	D^2	D^2 with mask
GLM	0.838 (0.058)	0.255 (0.153)	0.289 (0.157)
BRT	0.853 (0.052)	0.354 (0.154)	0.372 (0.151)

AUC area under the ROC curve; D^2 percentage of deviance explained; GLM generalized linear model; BRT boosted regression trees.

The means and standard errors of the AUC are given for the presence-absence models. The means and standard errors of the D^2 are given for the final abundance prediction models. These two indicators were calculated from the 24 predictions made for the four bird groups and the 6 years. The deviance explained is given for the predictions before and after applying a mask (see text for more details).

The predictions of potential AIV circulation allowed the identification of several priority cells for surveillance each year. The years with a low water level showed cells with higher values for the AIV circulation indicator (Fig. 4). The variability of the AIV circulation risk predictions was low for years with a low water level, i.e., years that showed a higher risk of AIV circulation (Fig. 4). All the models ran for the period from January 11–20, 2010 identified the same three cells with a higher risk of AIV circulation. As environmental variables were available on January 23, 2010, the model could be run in real time, and surveillance could be prioritized according to relevant environmental data.

DISCUSSION

Our analyses provided two main results. First, despite the very high spatial-temporal variability of the IND ecosystem,

it was possible to model the distribution of several bird groups with a low-to-good predictability (predictability of the presence-absence model could be considered to be good, while the predictability of the abundance was low-to-good) explaining on average 37.2% (SE = 15.1%) of the deviance. Second, combining these predictions and weighting them according to AIV prevalence made it possible to identify at what points in time and space the risk of transmission or emergence of highly pathogenic strains was highest. We thereby could identify when and where surveillance should be prioritized.

There was substantial variation across years in the predictability of the different bird groups. For the BRT model, which showed the best predictive power, the deviance explained varied from 10.3% for the Palearctic ducks in 2000 to 63.7% for the Ardeids in 2008. Variation in the ability to predict was highest for the Palearctic ducks and lowest for the Ardeids, with intermediate levels for Afro-tropical ducks and waders. The model always had a lower predictive power for Palearctic ducks, highly gregarious species, than for Ardeids, the least gregarious species in this study. This is in accordance with a negative impact of gregariousness on the predictive power of the model, an observation already made by Seoane et al. (2005). One approach to improve this was to mask out areas where a bird group had never been observed. As expected, the mask increased the predictive power of the model, especially for gregarious species. This mask could reflect specific behavior of gregarious birds such as wintering site fidelity (Leyrer et al., 2006) that results in suitable habitats being left unoccupied. Despite this variation in the predictability, the model identified the main areas where wild waterbirds congregated in the IND for each year.

Table 3. Percentage of deviance explained by the BRT model

Year	Max. flood level (cm)	Afro-tropical ducks	Palearctic ducks	Ardeids	Waders
1999	486	0.154 (0.009)	0.227 (0.006)	0.327 (0.002)	0.112 (0.002)
2000	511	0.153 (0.002)	0.103 (0.005)	0.541 (0.001)	0.390 (0.002)
2001	465	0.255 (0.001)	0.335 (0.003)	0.387 (0.001)	0.354 (0.001)
2006	442	0.337 (0.005)	0.488 (0.004)	0.507 (0.001)	0.401 (0.002)
2007	462	0.281 (0.004)	0.483 (0.001)	0.645 (0.001)	0.499 (0.002)
2008	482	0.347 (0.002)	0.435 (0.002)	0.647 (0.001)	0.512 (0.002)

BRT boosted regression trees; D^2 percentage of deviance explained.

The means and standard errors of D^2 by the BRT model after applying a mask are given for each year and each bird group. These results were estimated from 25 runs of the model.

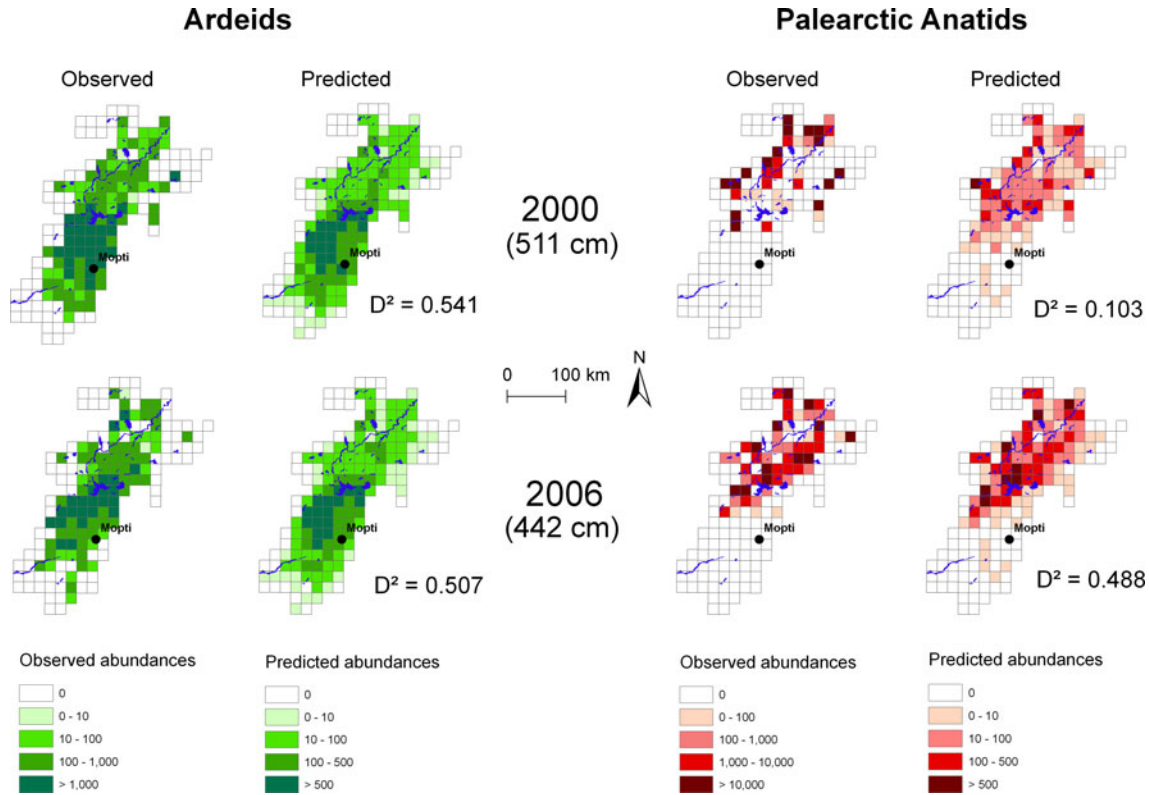


Figure 2. Spatial distributions of Ardeids and Palearctic Anatids in the IND. The observed data from the aerial census and the predicted distribution from the BRT model are shown for two different years. The observed data of the year considered were not included in the training dataset used for the predictions. Year 2006 was the one with the lowest flood level (442 cm), whereas 2000 had the highest flood

level (511 cm). The percentage of deviance explained by the model D^2 is indicated for each prediction. Different legend scales were applied to observed and predicted abundances in order to show the correspondence between the areas of highest bird density despite a general underestimation of the model.

The main factor affecting the distribution of the bird groups, and hence their prediction, was the flood level. Results showed a significant higher predictive power of the spatial distribution models of Palearctic ducks for years of low flood level, which can be explained by the high gregariousness of Palearctic ducks. In years with higher flood levels, more water bodies are available and many of them are unoccupied even though the model identified them as being suitable habitats. This results in the lower predictive power of the model. The results also showed that a lower flood level was associated with a higher risk of AIV circulation. This is in accordance with an increased circulation risk related to a higher concentration of wild birds on the remnant water bodies. The circulation of AIV thus is thought to be strongly influenced by the flood level, which itself can be extremely variable. For example, between December 1999 and February 2000, the size of the flooded area declined from

17,400 to 6,200 km², but in the same months in 1984 and 1985, it declined from 1680 to 480 km² (Zwarts et al., 2009). Overall, our model could take into account this flood level variability, and had a better predictability and a lower variability for years of low flood level showing a higher AIV circulation risk.

One limitation of our model is that it can only be applied during the wintering period because total census data were available only for January. Further, the flood level in the IND, the habitat use of waterbirds, and their associated dispersal behavior are highly variable between seasons. However, from an epidemiological point of view, this is the most important period for AIV circulation in the IND because it combines several factors: the annual peak in abundance of the Palearctic migrants, potential carriers of AIV (Artois et al., 2009), an increasing density of birds as the flooding level decreases, enhancing density-dependent and waterborne transmission processes (Roche et al., 2009),

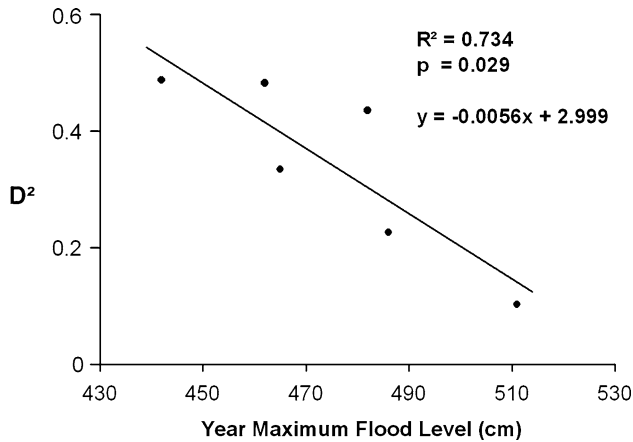


Figure 3. Influence of the maximum flood level of the year on the percentage of deviance explained D^2 by the BRT model for the Paelearctic duck distribution.

and cooler temperatures allowing better virus survival in the environment (Stallknecht et al., 1990).

Aside from the water level, other extrinsic factors also are involved in the variability of waterbird abundance in the IND. High interannual abundance variations have been observed in the IND, especially for Paelearctic migrants. This may be explained by annual variations in breeding

performances or by the selection of other West African wintering grounds. As our model is only based on local variables, it cannot take into account these variations. An improvement of the model would be to consider the West African scale and apply our model to the major wetlands of the region, i.e., the Senegal River Delta, IND, the Hadejia–Nguru wetlands, and Lake Chad, because a lower variability in overall abundance has been observed at this regional scale (Zwarts et al., 2009). Since the variables that we used are available for any region, our model could be applied to predict the distribution of the group species in these West African wetlands, and predictions could be validated using similar aerial censuses undertaken in the Senegal Delta and the Lake Chad area. This would allow quantification of the extrapolation capacity of our model, and would provide a method for regional-level prediction of potential AIV circulation hot spots in wild waterbirds.

As in other studies, the BRT models showed a better predictive power than GLM (Elith et al., 2006; Leathwick et al., 2009). One of their main advantages is the ability to deal with correlated variables. BRT models thus can include several variables related to the same ecological aspect (e.g., two vegetations indices related to food availability, NDVI

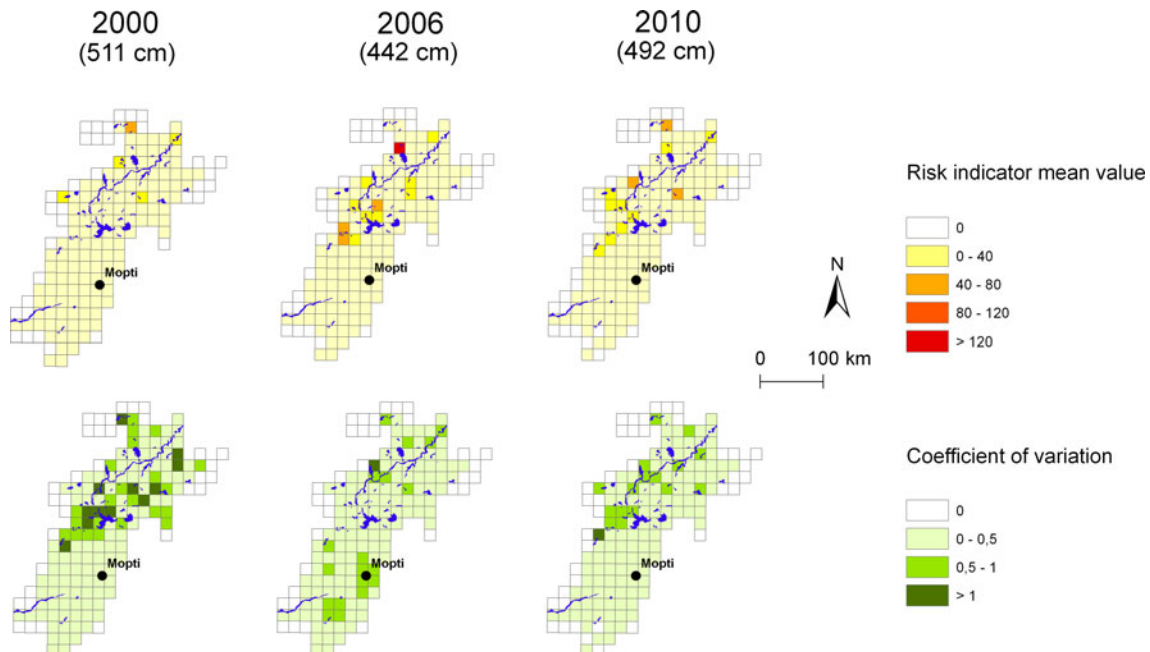


Figure 4. Predicted areas of AIV circulation. The maps on the first row show, for years with different flood levels (maximum flood level of the year in brackets), the predicted cells with a higher risk of AIV circulating in wild waterbirds of the IND. The mean value of the risk indicator was estimated with six runs of the BRT model trained on

six different datasets. Each training dataset included a different combination of five out of the six census years. The maps on the second row show the coefficient of variation (SE/mean) of these six runs of the model. AIV, Avian Influenza Viruses; BRT, boosted regression trees.

and DMP, could be included in the same model). Another difference between GLM and BRT models was shown by the effect of the mask we applied, which was to increase the capacity of the model to predict absence and, consequently, to increase the D^2 . The higher increase of the D^2 for the GLM due to this mask shows that the BRT model has a better capacity to predict absence than the GLM. Though variable interpretation is easier with GLM and computation time is shorter, BRT models are worth using when the main objective is predictive performance.

As little is known about the ecology and the epidemiology of HPAIV in wild birds, we used epidemiological data from low pathogenic AIV (LPAIV). Experimental infections with H5N1 HPAIV (Brown et al., 2006, 2008; Keawcharoen et al., 2008) showed that most Anatids species are able to replicate and shed H5N1 HPAIV, half of the time without any clinical sign (Gaidet et al., 2010). Furthermore, almost all of healthy, free-living wild birds that have tested positive for H5N1 HPAI with no apparent clinical signs were Anatids (Chen et al., 2006; Saad et al., 2007; Hesterberg et al., 2009; OIE, 2009). The highest circulation rates of LPAIV in the field also have been detected for Anatids species (Olsen et al., 2006). Our model thus is likely to point out the potential areas at risk for new strains of HPAIV despite the differences between HPAIV and LPAIV epidemiology, in particular, in terms of host range and host susceptibility. This model can be used to improve the surveillance of AIV in wild birds, but it cannot be used to directly predict the emergence of highly pathogenic AIV in poultry. Information on poultry distribution, their potential contacts with wild birds, and the transmission and survival rates of the virus should be added to the model to predict the emergence of highly pathogenic AIV in poultry.

CONCLUSIONS

Our model could predict the spatial distribution of several waterbirds groups and could identify areas with a higher AIV circulation risk in the IND. Based on environmental variables with a temporal resolution of 10 days, this model could take into account high seasonal ecological variability. This model therefore could be applied in other areas or for other emerging infectious diseases with strong seasonal drivers. When census data are available for a host population, the model can help prioritize where surveillance is needed. Applications of this methodology could enhance

the control of emerging diseases at a local and regional scale, especially when the resources available for surveillance programs are scarce.

ACKNOWLEDGMENTS

We thank Bourama Niagate from the Direction Nationale des Eaux et Forêts du Mali, Jean-Marie Boutin and Jean Thal from the Office National de la Chasse et de la Faune Sauvage (ONCFS), and all the people that helped collect the data. We also thank Leo Zwartz and Jan Van der Kamp from Altenburg and Wymenga (A&W) for sharing their knowledge about the Inner Niger Delta, and for providing helpful advice on the spatial modeling. Finally, we thank Olivier Gimenez from the Centre National de Recherche Scientifique (CNRS) for advice during the early development of the models.

APPENDIX

Table 4. List of the 16 species used for the analysis classified in four systemic groups according to the role they may play in Avian Influenza Virus circulation

<i>Palaearctic ducks</i>	
1. Northern pintail	<i>Anas acuta</i>
2. Garganey	<i>Anas querquedula</i>
3. Ferruginous duck	<i>Aythya nyroca</i>
<i>Afro-tropical ducks</i>	
4. Fulvous whistling duck	<i>Dendrocygna bicolor</i>
5. White-faced whistling duck	<i>Dendrocygna viduata</i>
6. Comb duck	<i>Sarkidiornis melanotos</i>
<i>Waders</i>	
7. Black-winged stilt	<i>Himantopus himantopus</i>
8. Black-tailed godwit	<i>Limosa limosa</i>
9. Ruff	<i>Philomachus pugnax</i>
<i>Ardeids</i>	
10. Grey heron	<i>Ardea cinerea</i>
11. Purple heron	<i>Ardea purpurea</i>
12. Squacco heron	<i>Ardeola ralloides</i>
13. Cattle egret	<i>Bubulcus ibis</i>
14. Great egret	<i>Ardea alba</i>
15. Unidentified white heron	<i>Ardeinae</i>
16. Little egret	<i>Egretta garzetta</i>

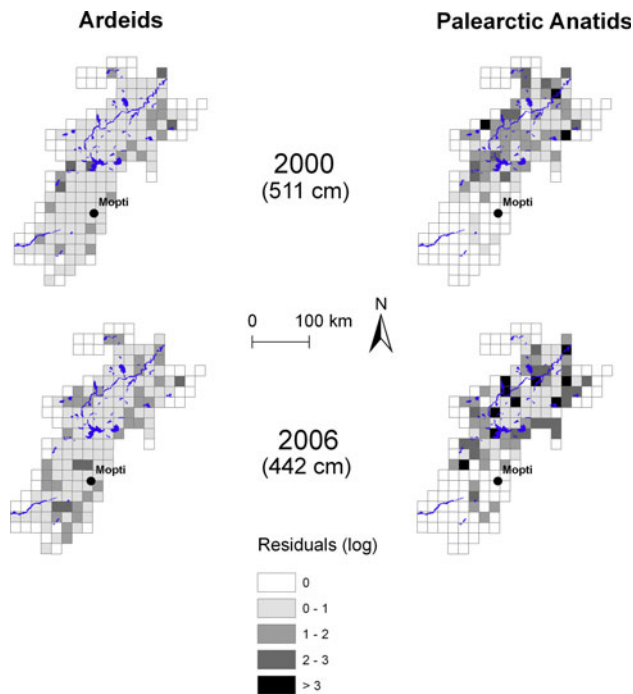


Figure 5. Residuals of the spatial distribution model of Ardeids and Palearctic Anatids. The residuals, i.e., the difference between the observed and predicted abundances, are displayed for two different years.

REFERENCES

- Artois M, Bicoût D, Doctrinal D, Fouchier R, Gavier-Widen D, Globig A, et al. (2009) Outbreaks of highly pathogenic avian influenza in Europe: the risks associated with wild birds. *Revue Scientifique Et Technique—Office International Des Epizooties* 28:69–92
- Augustin NH, Muggleston MA, Buckland ST (1996) An autologistic model for the spatial distribution of wildlife. *Journal of Applied Ecology* 33:339–347
- Barry SC, Welsh AH (2002) Generalized additive modelling and zero inflated count data. *Ecological Modelling* 157:179–188
- Brochet A-L, Guillemain M, Lebarbenchon C, Simon G, Fritz H, Green A, et al. (2009) The potential distance of highly pathogenic avian influenza virus dispersal by mallard, common teal and Eurasian pochard. *EcoHealth* 6:449–457
- Brown JD, Stallknecht DE, Beck JR, Suarez DL, Swayne DE (2006) Susceptibility of North American ducks and gulls to H5N1 highly pathogenic avian influenza viruses. *Emerging Infectious Diseases* 12:1663–1670
- Brown JD, Stallknecht DE, Swaynet DE (2008) Experimental infection of swans and geese with highly pathogenic avian influenza virus (H5N1) of Asian lineage. *Emerging Infectious Diseases* 14:136–142
- Chen H, Li Y, Li Z, Shi J, Shinya K, Deng G, et al. (2006) Properties and dissemination of H5N1 viruses isolated during an influenza outbreak in migratory waterfowl in western China. *Journal of Virology* 80:5976–5983
- Chen JM, Liu J, Cihlar J, Goulden ML (1999) Daily canopy photosynthesis model through temporal and spatial scaling for remote sensing applications. *Ecological Modelling* 124:99–119
- CIESIN (2005) Gridded Population of the World, Version 3 (GPWv3): Population Density Grids. Socioeconomic Data and Applications Center (SEDAC), Columbia University. <http://sedac.ciesin.columbia.edu/gpw>. Accessed 10 October 2009
- Clements ACA, Pfeiffer DU (2009) Emerging viral zoonoses: frameworks for spatial and spatiotemporal risk assessment and resource planning. *Veterinary Journal* 182:21–30
- Cumming G, Hockey PAR, Bruinzeel LW, Du Plessis MA (2008) Wild birds movements and avian influenza risk mapping in Southern Africa. *Ecology and Society* 13:26
- Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, et al. (2006) Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29:129–151
- Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. *Journal of Animal Ecology* 77:802–813
- Fusaro A, Joannis T, Monne I, Salvato A, Yakubu B, Meseko C, Oladokun T, et al. (2009) Introduction into Nigeria of a distinct genotype of avian influenza virus (H5N1). *Emerging Infectious Diseases* 15:445–447
- Gaidet N, Dodman T, Caron A, Balanca G, Desvaux S, Goutard F, et al. (2007) Avian influenza viruses in water birds, Africa. *Emerging Infectious Diseases* 13:626–629
- Gaidet N, Cattoli G, Hammoumi S, Newman SH, Hagemeijer W, Takekawa JY, et al. (2008) Evidence of infection by H5N2 highly pathogenic avian influenza viruses in healthy wild waterfowl. *PLoS Pathogens* 4:e1000127
- Gaidet N, Cappelle J, Takekawa JY, Prosser DJ, Iverson SA, Douglas DC, et al. (2010) Potential spread of highly pathogenic avian influenza H5N1 by wildfowl: dispersal ranges and rates determined from largescale satellite telemetry. *Journal of Applied Ecology* 47:1147–1157
- Gao BC (1996) NDWI—a normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment* 58:257–266
- Girard O (2006) Anatidae wintering in the Inner Niger Delta, Mali. In: *Waterbirds Around the World*, Boere GC, Galbraith CA, Stroud DA (editors), Edinburgh, UK: The Stationery Office, p 960
- Girard O, Thal J, Niagate B (2004) The Antadis (*Anatidae*) wintering in the Inner Niger Delta. *Game and Wildlife Science* 21:107–137
- Gond V, Bartholome E, Ouattara F, Nonguierma A, Bado L (2004) Monitoring and mapping of waters and wetlands in arid regions using the SPOT-4 VEGETATION imaging system. *International Journal of Remote Sensing* 25:987–1004
- Guisan A, Zimmermann NE (2000) Predictive habitat distribution models in ecology. *Ecological Modelling* 135:147–186
- Hastie T, Tibshirani R, Friedman JH (2001) *The Elements of Statistical Learning: Data Mining Inference, and Prediction*, New York: Springer-Verlag
- Hesterberg U, Harris K, Stroud DA, Guberti V, Busani L, Pittman M, et al. (2009) Avian influenza surveillance in wild birds in the European Union in 2006. *Influenza and Other Respiratory Viruses* 3:1–14
- Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. (2008) Global trends in emerging infectious diseases. *Nature* 451:990–993
- Keawcharoen J, van Riel D, van Amerongen G, Bestebroer T, Beyer WE, van Lavieren R, et al. (2008) Wild ducks as long-distance vectors of highly pathogenic avian influenza virus (H5N1). *Emerging Infectious Diseases* 14:600–607

- Koopmans M, Wilbrink B, Conyn M, Natrop G, van der Nat H, Vennema H, et al. (2004) Transmission of H7N7 avian influenza A virus to human beings during a large outbreak in commercial poultry farms in the Netherlands. *The Lancet* 363:587–593
- Leathwick JR, Elith J, Rowe D, Julian K (2009) Robust planning for restoring diadromous fish species in New Zealand's lowland rivers and streams. *New Zealand Journal of Marine and Freshwater Research* 43:659–671
- Leyrer J, Spaans B, Camara M, Piersma T (2006) Small home ranges and high site fidelity in red knots (*Calidris c. canutus*) wintering on the Banc d'Arguin, Mauritania. *Journal of Ornithology* 147:376–384
- Manel S, Williams HC, Ormerod SJ (2001) Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology* 38:921–931
- McCullagh P, Nelder JA (1989) *Generalized Linear Models*, London: Chapman & Hall
- Morens DM, Folkers GK, Fauci AS (2004) The challenge of emerging and re-emerging infectious diseases. *Nature* 430:242–249
- Morens DM, Folkers GK, Fauci AS (2008) Emerging infections: a perpetual challenge. *Lancet Infectious Diseases* 8:710–719
- Morse SS (1995) Factors in the emergence of infectious-diseases. *Emerging Infectious Diseases* 1:7–15
- Munster VJ, Baas C, Lexmond P, Waldenström J, Wallensten A, Fransson T, et al. (2007) Spatial, temporal, and species variation in prevalence of influenza A viruses in wild migratory birds. *PLoS Pathogens* 3:e61
- OIE (2009) Highly Pathogenic Avian Influenza, Germany. World Organization for Animal Health. www.oie.int/wahis/public.php?page=single_report&pop=1&reportid=7874. Accessed 12 July 2010
- Olsen B, Munster VJ, Wallensten A, Waldenstrom J, Osterhaus A, Fouchier RAM (2006) Global patterns of influenza A virus in wild birds. *Science* 312:384–388
- Pearce J, Ferrier S (2000) Evaluating the predictive performance of habitat models developed using logistic regression. *Ecological Modelling* 133:225–245
- R Development Core Team (2009) *R: A Language and Environment for Statistical Computing*, v. 2.9.0, Vienna, Austria: R Foundation for Statistical Computing
- Reed BC, Brown JF, VanderZee D, Loveland TR, Merchant JW, Ohlen DO (1994) Measuring phenological variability from satellite imagery. *Journal of Vegetation Science* 5:703–714
- Roche B, Lebarbenchon C, Gauthier-Clerc M, Chang C-M, Thomas F, Renaud F, et al. (2009) Water-borne transmission drives avian influenza dynamics in wild birds: the case of the 2005–2006 epidemics in the Camargue area. *Infection Genetics and Evolution* 9:800–805
- Saad MD, Ahmed LS, Gamal-Eldein MA, Fouda MK, Khalil FM, Yingsl SL (2007) Possible avian influenza (H5N1) from migratory bird, Egypt. *Emerging Infectious Diseases* 13:1120–1121
- Seoane J, Carrascal LM, Alonso CL, Palomino D (2005) Species-specific traits associated to prediction errors in bird habitat suitability modelling. *Ecological Modelling* 185:299–308
- Siembieda JL, Johnson CK, Cardona C, Anchell N, Dao N, Reisen W, et al. (2010) Influenza A viruses in wild birds of the Pacific flyway, 2005–2008. *Vector-Borne and Zoonotic Diseases*. doi: 10.1089/vbz.2009.0095. Online 8 January 2010
- Smith KF, Dobson AP, McKenzie FE, Real LA, Smith DL, Wilson ML (2005) Ecological theory to enhance infectious disease control and public health policy. *Frontiers in Ecology and the Environment* 3:29–37
- Stallknecht DE, Kearney MT, Shane SM, Zwank PJ (1990) Effects of pH, temperature, and salinity on persistence of avian influenza viruses in water. *Avian Disease* 34:412–418
- Trolliet B, Girard O, Benmergui M, Schricke V, Boutin JM, Fouquet M, Triplet P (2008) Oiseaux d'eau en Afrique subsaharienne: Bilan des dénombrements de janvier 2007. *Faune Sauvage* 279:4–11
- Vgt4Africa (2006) VGT4Africa project. VITO NV. <http://www.vgt4africa.org>. Accessed 8 March 2010
- Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y (1992) Evolution and ecology of influenza-A viruses. *Microbiological Reviews* 56:152–179
- Zwarts L, Grigoras I (2005) Flooding of the Inner Niger Delta. In: *The Niger, A Lifeline*, Zwarts L, Beukering Pvan, Kone B, Wymenga E (editors), Lelystad, The Netherlands: RIZA/Wetlands International/IVM/A&W, pp 43–77
- Zwarts L, Bijlsma RG, van der Kamp J, Wymenga E (2009) *Living on the Edge: Wetlands and Birds in a Changing Sahel*, Zeist, The Netherlands: KNNV Publishing