Published in final edited form as:

Agric Ecosyst Environ. 2011 May 1; 141(3-4): 373–380. doi:10.1016/j.agee.2011.04.013.

Modelling the distribution of domestic ducks in Monsoon Asia

Thomas P Van Boeckel^{1,2}, Diann Prosser³, Gianluca Franceschini⁴, Chandra Biradar⁵, William Wint⁶, Tim Robinson⁴, and Marius Gilbert^{1,2}

¹ Biological Control and Spatial Ecology, Université Libre de Bruxelles CP160/12, Av FD Roosevelt 50, B1050, Brussels, Belgium

² Fonds National de la Recherche Scientifique, rue d'Egmont 5, B-1000 Brussels, Belgium

³ USGS Patuxent Wildlife Research Center, Baltimore Avenue 10300, MD 20705 Beltsville, and University of Maryland, College Park, USA

⁴ Food and Agriculture Organization of the United Nations (FAO), Viale delle Terme di Caracalla 00153, Rome, Italy

⁵ Department of Botany and Microbiology, Center for Spatial Analysis, University of Oklahoma, Norman, Oklahoma, 73019 USA

⁶ Environmental Research Group Oxford, P.O. Box 346, Oxford OX1 3QE, United Kingdom

Abstract

Domestic ducks are considered to be an important reservoir of highly pathogenic avian influenza (HPAI), as shown by a number of geospatial studies in which they have been identified as a significant risk factor associated with disease presence. Despite their importance in HPAI epidemiology, their large-scale distribution in monsoon Asia is poorly understood. In this study, we created a spatial database of domestic duck census data in Asia and used it to train statistical distribution models for domestic duck distributions at a spatial resolution of 1 km. The method was based on a modelling framework used by the Food and Agriculture Organisation to produce the Gridded Livestock of the World (GLW) database, and relies on stratified regression models between domestic duck densities and a set of agro-ecological explanatory variables. We evaluated different ways of stratifying the analysis and of combining the prediction to optimize the goodness of fit of the predictions. We found that domestic duck density could be predicted with reasonable accuracy (mean RMSE and correlation coefficient between log-transformed observed and predicted densities being 0.58 and 0.80, respectively), using a stratification based on livestock production systems. We tested the use of artificially degraded data on duck distributions in Thailand and Vietnam as training data, and compared the modelled outputs with the original high-resolution data. This showed, for these two countries at least, that these approaches could be used to accurately disaggregate provincial level (administrative level 1) statistical data to provide high resolution model distributions.

© 2011 Elsevier B.V. All rights reserved.

Corresponding author: Marius Gilbert, Biological Control and Spatial Ecology, Université Libre de Bruxelles CP160/12, Av FD Roosevelt 50, B1050, Brussels, Belgium, Tel +32 2 650 37 80, Fax + 32 2 650 24 45, mgilbert@ulb.ac.be.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

Livestock Mapping; Domestic ducks; Monsoon Asia; Regression models; Highly pathogenic avian influenza

1. Introduction

Maps of livestock distribution are a concise way to visualize and analyze large census datasets. They have a wide variety of applications such as assessing the risk of zoonotic disease, food safety management, determination of the potential daily protein production capacity, monitoring of the land-use changes, assessment of the environmental risk associated with animal production (Wint and Robinson, 2007).

The Highly Pathogenic Avian Influenza (HPAI) H5N1 virus that first appeared in southern China in the late 1990's (Li et al., 2004) is one of the most significant recent epizootics which has had dramatic consequences on smallholders' livelihoods and poultry production in many countries (Brown, 2010). To date, the human death toll of these events remains moderate despite the very high mortality rates observed in wild and domestic fowl (World Health Organization, October 2010: 507 cases reported, 302 deaths confirmed).

Domestic ducks play a significant role in the epidemiology of HPAI H5N1 virus. First, experimental studies have demonstrated that they can be apparently healthy carriers of the HPAI H5N1 virus and have even been referred to as the "Trojan horse of the avian flu" (Kim et al., 2009). Domestic ducks have been shown to survive HPAI H5N1 virus infections and excrete large quantities of the virus without showing clinical signs of disease (Hulse-Post et al., 2005). As a result, domestic ducks may play a determinant role in the transmission of the virus during the movement of flocks between different feeding locations. Studies have shown that the distribution of HPAI H5N1 virus in parts of Asia is heavily influenced by the distribution of duck farming systems (Gilbert et al., 2007). More specifically, the density of ducks has been found to be a key variable for predicting of the presence of HPAI H5N1 virus in Thailand (Gilbert et al., 2006), Vietnam (Pfeiffer et al., 2007), over the Mekong region (Gilbert et al., 2008), and even at the global scale (Hogerwerf et al., 2010).

However, despite the importance of ducks in HPAI H5N1 transmission, data on domestic duck populations derived from national censuses are often incomplete and vary considerably in resolution between countries. Some countries such as Thailand and Vietnam that produce large quantities of duck meat and eggs have undertaken massive survey campaigns (*e.g.* X-ray Survey in Thailand involving 400,000 inspectors) following the HPAI H5N1 virus epidemics of 2004 and have very high resolution data reported at administrative levels 3 and 4 (*e.g.* Thailand: county and village level). In some other countries where HPAI H5N1 is of concern, however, data are available at best only at administrative levels 1 and or 2. A striking example is China, which hosts more than 75% of the world's domestic duck population (FAOSTAT, 2010), yet species level duck data are available only at the provincial level (administrative level 1). In the Chinese province of Xinjiang, for example, there is only one figure provided for the entire 600,000 square kilometres. At this time modelers lack comprehensive and accurate information relative to the domestic duck population density and location. Ideally such information could be modeled at fine spatial resolution by applying an appropriate *disaggregation* methodology to the existing data. Various disaggregation techniques have been explored and applied with relative success to a wide range of livestock species (Wint and Robinson, 2007; Neumann et al., 2009; Gerber et al., 2005). To our knowledge, however, no previous attempts have focused on the

geographical distribution of domestic ducks specifically in Monsoon Asia, a region of particular concern for the persistence of HPAI H5N1 virus. In this study we present a statistical modelling procedure inspired by earlier efforts to map livestock distributions (Wint and Robinson, 2007) and specially adapted to the case of domestic ducks. The methodology was applied to disaggregate reported domestic duck data to a 1×1 km continuous population density surface across 14 countries.

The methodology relies on the use of agro-ecological predictors providing information on both bioclimatic and anthropogenic factors which are assumed to affect the geographical distribution of duck farming. The central hypothesis is that robust statistical relationships can be established between domestic duck population density and these agro-environmental predictors, and that these relationships can be used to disaggregate the administrative level distribution of domestic duck data across Monsoon Asia.

2. Materials & Methods

We aimed to obtain the most recent duck data at the finest possible administrative level within Monsoon Asia. This required data to be compiled at different administrative levels depending on sources available. Table 1 provides details of the reported statistics used. The absolute numbers of ducks per reporting unit were converted into densities (birds km⁻²) by dividing the number of birds by the area of land considered suitable for duck production. This step was important to prevent densities from being underestimated in administrative units containing vast areas of land unsuitable for duck production (e.g. Xinjiang Province, China). The suitability mask for ducks was modified from the original GLW monogastric livestock (pigs and poultry) mask (Wint and Robinson, 2007). We took a conservative approach in suitability masking by excluding only the most unsuitable areas for duck production (extreme high elevations, tundra, ice and deserts, for example, see Table S4 in Prosser *et al.* this volume). The resulting dependent (*y*) variable data are referred to as ‘training data’.

Two main groups of predictors were included. First, eco-climatic predictors derived from Fourier transformed MODIS Imagery (<http://modis.gsfc.nasa.gov/>) over the period 2001–2005 (Scharlemann *et al.*, 2008) and second, anthropogenic variables.

The Global MODIS Imagery is a compilation of images of the surface of the earth taken every 8 or 16 days for different channels of reflectance. These images are then composited over a period of time to generate a time series for each channel. The different channels are subject to atmospheric corrections and used in combination to define bioclimatic indices such as the Normalized Difference Vegetation Index (NDVI) or Land Surface Temperature (LST); variables that are expected to be of relevance in explaining the spatial distributions of livestock species (Wint and Robinson 2007) and associated animal diseases (Hay *et al.*, 2006; Sinka *et al.*, 2010).

The Fast Fourier Transform (FFT) is a widespread ordination method used in ecology to summarize the information contained in a time-series in a way that is directly subject to interpretation (which is not always the case of other ordination methods such as the principal component analysis). The resulting output of the FFT is an image with each pixel being attributed the value the Fourier Coefficients (phase or amplitude) of the time series for the different harmonics of the signal. In this study we used the values of the amplitude and phase of the first to third harmonic, describing respectively the amplitude and the delay of the annual, bi-annual and tri-annual biological cycles for the channels listed in Table 2.

Domestic livestock production is constrained by natural limitations such as food availability, temperature, aridity and access to water, which can be captured through statistical

relationships with the eco-climatic predictors. But production is also driven by human activities and related to factors such as population density, demand and access to production inputs (feeds, services etc.) and output markets. As these factors are suspected to explain a substantial share of the spatial variability of domestic duck distribution in specific areas, we also used a set of anthropogenic predictor variables including human population density (Balk et al., 2006), distance to the nearest road (Dobson et al., 2000) and travel time to major cities (Nelson 2008; Table 2.)

The first step of the modelling procedure consisted of the spatial sampling of duck density data. However, because of the high heterogeneity in spatial resolution of the reported training data from different countries, applying a sampling scheme of equal density was not ideal. Applying a density related sampling scheme to take advantage of high resolution training data in some countries would have resulted in a large proportion of data points being associated with the same, single duck density value reported within an administrative unit. Conversely, using a low density of sampling points would not take full advantage of the higher resolution data available in other areas.

We therefore used a stratified sampling approach that included at least one point per administrative unit with training data and, beyond that, an average of 20 points per decimal degree across our analysis extent (Supplementary Information, Fig. S3). Twenty five random sets of sample points were created in order to bootstrap the analysis - each containing 75% of points used to train the statistical models (model set) and 25% for model validation (test set).

A forward inclusion multiple regression approach was developed using log-transformed duck densities as the dependent variable (to normalize duck densities). This involved the iterative inclusion of each predictor, paired with its quadratic term to account for non-linear relationships. Starting from a null-model, the algorithm searched for and included the pair of predictors that resulted in the best reduction in Akaike Information Criterion (AIC) (Burnham et Anderson, 2002; Whittingham et al., 2006). This process continued until the improvement in AIC was less than 1% (Fig. 1). As a measure to prevent over-fitting, we limited the maximum number of pairs of predictors to one for every 15 sampling points from different administrative units. The coefficients of the model were then applied to the imagery to obtain the log-transformed predicted density of ducks for each pixel. A ceiling value 50% higher than the highest observed duck density in the census data was applied to the predictions to remove any anomalous predictions.

The analysis was also stratified to allow for different quantitative relationships linking duck densities to covariates in different agro-ecological systems. We considered two types of stratification. The first was based on the distribution of livestock production systems (LPS) proposed by Seré and Steinfeld (1996) and mapped by Kruska et al. (2003). The second was based on 4 different levels of a data-driven ecozonation (EZ) comprising respectively 5, 12, 25 and 50 clusters. These ecozones were obtained by an isodata unsupervised clustering of TFA-processed MODIS channels for Middle Infra Red, day time Land Surface Temperature, night time Land Surface Temperature, Normalized Difference Vegetation Index, Enhanced Vegetation Index and the Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM). For each stratum within each stratification system (LPS, EZ5, EZ12, EZ25 & EZ50), a minimum of 15 points per variable pair with different duck densities were required to train a model. When this condition was not met, an unstratified regional model based on all points was used.

Wint and Robinson (2007) developed a method to combine predictions made under different stratification schemes whereby the prediction from the stratification that resulted in a model

with the best R^2 was selected for each pixel. We evaluated this method against alternatives and compared different ways of combining stratifications. Three combination methods were based on the ecozone classifications at different levels (EZ5, EZ12, EZ25 and EZ50). In the first method (EZ.Finest), the algorithm selects for each pixel the value predicted by the finest (most detailed) ecozonation. For example, if for a given pixel, a prediction was available with EZ12 and EZ25 stratifications, but not EZ50 (because the EZ50 strata did not have enough unique data points), the prediction of the EZ25 stratification would be selected. In the second and third combination methods; EZ.BestRSE and EZ.BestR2, predictions were selected from the ecozone stratification scheme that resulted in a model with the lowest residual squared error (RSE), or highest R^2 value, respectively. Lastly, two methods used all stratifications (LPS, EZ5, EZ12, EZ25 and EZ50), and simply used for each pixel, the prediction from the stratification scheme which gave the lowest RSE (All.BestRSE) or highest R^2 (All.BestR2).

The goodness of fit of the resulting predictions were evaluated using the 25% of points that had been withheld from the analyses, using two indices: the root mean squared difference (RMSE) and the correlation coefficient (COR) between the predicted and observed densities (in Log10-scale). Both RMSE and COR were estimated for each bootstrap. The mean geographical distribution of the predicted densities was obtained by averaging the distributions predicted by each bootstrap, and the coefficient of variation of the prediction was also estimated as an index of the variability in predictions. In order to standardize the results to a single reference year the predicted values were adjusted such that the national totals of ducks matched those provided by FAOSTAT for the year 2007 (FAOSTAT, 2010). RMSE and COR were compared for different stratification schemes, combination methods and levels of detail in training data using one-way ANOVA followed by Tukey multiple comparison tests.

A restricted set of analyses was carried out for Vietnam and Thailand, in order to assess the sensitivity of the mapping procedure to the resolution of the input data. Vietnam and Thailand are the only countries where the reported duck data were fully available at the 3rd administrative level. These high-resolution data were aggregated to administrative levels 2 (district), 1 (province) and 0 (country) to produce artificially degraded training datasets from which to derive regression models. In addition, we also evaluated a scenario in which no data were available for Thailand and Vietnam so that predictions for these countries were based on relationships established using data from other countries. For this specific set of comparisons, we used the LPS stratification system and predictions were evaluated against the original, high-resolution sub-district data (administrative level 3).

3. Results

RMSE and correlation coefficient are illustrated in Fig. 2 for the five different stratification systems (LPS, EZ5, EZ12, EZ25 & EZ50) and five combination methods (EZ.Finest, EZ.BestRSE, EZ.BestR2, All.BestRSE, All.BestR2) tested in this study. The best goodness of fit scores were produced by the LPS stratification scheme (RMSE = 0.58, correlation coefficient = 0.80), and for the All.BestRSE combination method (RMSE = 0.57, correlation coefficient = 0.81), each indicating good agreement between observed and predicted values. No significant difference was observed between the LPS stratification and Best RSE combination method (Post-Hoc Tukey adj. p = 0.473), thus we chose the LPS system for the sensitivity analysis carried out on Vietnam and Thailand.

Figure 3 presents the map of observed duck density (top), the mean predicted density (middle) and its coefficient of variation (bottom) of the All.BestRSE method. The geographical distribution of the modelled duck density values correspond with the regions

known to be the main duck production areas in Monsoon Asia, (Fig. 3. top). These include Bangladesh; the deltas of the Mekong, Red, Irrawaddy, and Chao Phraya rivers; the Chinese province of Sichuan; the South and East coasts of China; and the Island of Java in Indonesia. Finer scale displays are provided as Supplementary Information Figure S4. The variability in the predictions quantified by the coefficient of variation of the predicted duck density (Fig 3, bottom) was moderate in the main duck production areas, and high in regions with low absolute duck densities, such as the western side of the Indian sub-continent, Nepal, Bhutan, the north-eastern deserts of China, Borneo Island, and the Ridge of Papua New Guinea.

The mean RMSE and correlation coefficients for each of the 14 countries included in the study are presented in Fig. 4. Bhutan, Malaysia, and South-Korea Republic show the highest variability in predicted goodness of fit; while China, Indonesia, Thailand and Vietnam are characterized by a relatively low variability of the error. The lowest RMSE values were observed for Laos, Myanmar, and Cambodia. The outputs of the predictions made for Vietnam and Thailand using various administrative levels of training data showed that the quality of output maps and the goodness of fit metrics were highly dependent on the level of detail in the training data (Figs. 5 & 6). The goodness of fit indices worsened with decreasing spatial resolution of training data: the correlation coefficient increased from 0.30 for administrative level 0 training data; through 0.64 and 0.69 for administrative levels 1 and 2, respectively, up to 0.71 when administrative level 3 training data were used. As expected, the best result for the predicted duck density was obtained with the finest spatial resolution of training data as illustrated in Fig. 6. This figure also highlights that relatively poor predictions were obtained when only one density for the entire country was used (admin. Level 0 input data). In those circumstances, improved outputs were obtained by basing the predictions only on sampling points from adjacent countries (NoData).

4. Discussion

The main objective of this study was to assess different disaggregation techniques based on spatial correlations between domestic duck densities and agro-ecological covariates. The broad-scale pattern of the predictions matches that of the observed densities, and highlights regions with favourable agro-ecological conditions for duck production. These include areas where flood plain agriculture is practiced and in low altitude areas such as the large plains of the South Asian river deltas, which are known to be major areas for extensive duck production. The high duck densities predicted in these regions are supported by the multiple rice cropping cycles that provide a sustained food source for duck production, as suggested in previous studies (Gilbert et al., 2007). Ducks are also raised in relatively high densities in North-Eastern areas of China, where market accessibility and demand for duck meat are high. The goodness of fit metrics RMSE and COR obtained using the best-performing methods (All.BestRSE and LPS stratification) indicate that spatial correlations between duck densities and agro-ecological covariates can be used to predict duck densities with reasonably good accuracy (Correlation coefficients are respectively 0.76 and 0.75 for All.BestRSE and LPS). The more detailed analyses carried out for Thailand and Vietnam demonstrate that the methodology developed in this study can be used to disaggregate artificially degraded training data to predict the pattern of duck densities that would be obtained from lower spatial resolution training data. Whilst improved predictive accuracy is evident using higher resolution training data, the analysis showed that reasonable disaggregation results were also obtainable using training data from administrative levels 1 or 2 (coefficients of correlation are respectively 0.64 and 0.69; Fig 6). This result is particularly useful because it suggests that good predictions can be obtained even in countries where data are only available at relatively low resolution. For example, we only have administrative level 1 data for Myanmar; a country that was struck by the HPAI H5N1 epidemic. A second interesting result of this analysis based on Thailand and Vietnam is that

when only one figure is available for the entire country (Admin. Level 0), it is advisable not to use it to train the model, rather to train the model using exclusively values from neighbouring countries (NoData) (provided that sufficient data are available there). This suggests that the techniques can be used to fill gaps - predicting duck densities for countries where we have little or no information - provided that the models can be trained using higher accuracy data from other regions sharing similar agro-ecological conditions. This is likely to be less reliable when data are missing for very large countries with a wide range of ecological conditions, and this indeed seems to be the case for China (see Prosser et al., this issue). One should note that the high spatial resolution of the output (1 km) does not necessarily reflect ground truth at the same resolution, because disaggregation was only evaluated at the administrative level 3. Finer-scale evaluation of the output could only be achieved by the collection of validation data at an equivalent scale of 1 km.

Model uncertainty varied spatially and in relation to observed duck density. Coefficients of variation (expressed as a proportion of the predicted value) (Fig 3. bottom) was lowest in regions with the highest duck densities, and higher in areas with low duck densities. This indicates that low densities of ducks in peripheral regions are more difficult to model reliably, but also that the main duck breeding areas are predicted with relatively low uncertainty. Isolated peaks in COV were mainly found in the western part of the Indian sub-continent, with extreme values located in Bhutan. The high model uncertainty may be due to extremely low densities in the reported data and the small statistical weight of the Bhutan's specific ecosystems in the overall dataset. Goodness of fit assessed by country showed significant variability between countries (Fig. 4). The variability in goodness of fit relates to the quality of the training data with countries like Thailand or Vietnam showing the lowest variability in goodness of fit, both countries having the most detailed input data. One could also note that countries with low intensity of duck production such as Myanmar, Cambodia or Laos had relatively good accuracy metrics.

One aim of this study was to explore ways by which model accuracy could be improved through the use of different stratification schemes and different ways of combining the results thereof. The LPS performed best among the stratification schemes evaluated in this study. The areas of high domestic duck densities can be associated with either extensive (e.g. Irrawaddy Delta) or intensive farming practices (e.g. Chinese peri-urban breeding). Naturally, these differences can potentially be more easily captured by a livestock-oriented stratification system than by unsupervised, data driven stratification systems. The LPS classes associated with the highest domestic duck densities were the mixed irrigated humid systems (MIH), with mean densities of 223 birds km⁻², followed by the mixed irrigated temperate systems (MIT), which averaged 200 birds km⁻², and the Urban systems, with 131 birds km⁻². Considering the lack of biological interpretability of the combined stratification All.BestRSE method and its more complex implementation, the use of the LPS layer as single stratification scheme for modelling domestic duck densities appears as the best approach. Nevertheless, in conditions where a thorough study has not demonstrated the value of a particular stratification system, for example modelling a different species, or in multiple areas with different environmental conditions, the appropriateness of alternative stratification schemes and combination methods should be evaluated.

In this study we showed that an improved version of previously developed methodology for livestock disaggregation (Wint and Robinson, 2007) could be applied successfully to model the distribution of domestic duck in Monsoon Asia. In the general framework developed for this purpose we used various stratifications and combination strategies. Among these, the Livestock Production System layer originally produced by Seré and Steinfeld (1996) comes out to be an essential mapping tool when applied to domestic ducks: first because it generates the best results in terms of statistical goodness of fit, second because of the

easiness of interpretation it offers in term of farming practices. All maps produced in this study are made available through the FAO Gridded Livestock of the World database.

Acknowledgments

This work was partly supported by the National Institutes of Health Fogarty International Center through the NSF/NIH Ecology of Infectious Diseases program (7R01TW007869-04).

References

- Balk DL, Deichmann U, Yetman G, Pozzi F, Hay SI, Nelson A. Determining global population distribution: methods, applications and data. *Adv Parasit*. 2006; 62:119–156.
- Brown IH. Summary of Avian Influenza Activity in Europe, Asia, and Africa, 2006–2009. *Avian Dis*. 2010; 54(s1):187–193. [PubMed: 20521631]
- Burnham, KP.; Anderson, DR. *Model Selection and Multimodel Inference: a Practical Information—Theoretic Approach*. Springer; New York: 2002.
- Dobson JE, Bright EA, Coleman PR, Durfee RC, Worley BA. LandScan: a global population database for estimating populations at risk. *Photogram Eng Rem S*. 2000; 66(7):849–857.
- Gerber P, Chilonda P, Franceschini G, Menzi H. Geographical determinants and environmental implications of livestock production intensification in Asia. *Bioresource Technol*. 2005; 96(2):263–276.
- FAOSTAT. [Accessed April 2011] FAOSTAT Production Data. 2006. Available online at: <http://faostat.fao.org/>
- Gilbert M, Chaitaweesub P, Parakamawongsa T, Premashtira S, Tiensin T, Kalpravidh W, Wagner H, Slingenbergh J. Free-grazing ducks and highly pathogenic avian influenza, Thailand. *Emerg Infect Dis*. 2006; 12(2):227–234. [PubMed: 16494747]
- Gilbert M, Xiao X, Chaitaweesub P, Kalpravidh W, Premashtira S, Boles S, Slingenbergh J. Avian influenza, domestic ducks and rice agriculture in Thailand. *Agr Ecosyst Environ*. 2007; 119(3–4): 409–415. [PubMed: 18418464]
- Gilbert M, Xiao X, Pfeiffer DU, Epprecht M, Boles S, Czarnecki C, Chaitaweesub P, Kalpravidh W, Minh PQ, Otte MJ, Martin V, Slingenbergh J. Mapping H5N1 highly pathogenic avian influenza risk in Southeast Asia. *P Natl Acad Sci USA*. 2008; 105(12):4769–4774.
- Hay SI, Tatem AJ, Graham AJ, Goetz SJ, Rogers DJ. Global environmental data for mapping infectious disease distribution. *Adv Parasit*. 2006; 62:37–77.
- Hogerwerf L, Wallace RG, Ottaviani D, Slingenbergh J, Prosser D, Bergmann L, Gilbert M. Persistence of Highly Pathogenic Avian Influenza H5N1 Virus Defined by Agro-Ecological Niche. *EcoHealth*. 2010; 7(2):213–225. [PubMed: 20585972]
- Hulse-Post DJ, Sturm-Ramirez KM, Humbird J, Seiler P, Govorkova EA, Krauss S, Scholtissek C, Puthavathana P, Buranathai C, Nguyen TD, Long HT, Naipospos TSP, Chen H, Ellis TM, Guan Y, Peiris JSM, Webster RG. Role of domestic ducks in the propagation and biological evolution of highly pathogenic H5N1 influenza viruses in Asia. *P Natl Acad Sci USA*. 2005; 102(30):10682–10687.
- Kim J, Negovetich NJ, Forrest HL, Webster RG. Ducks: The “Trojan Horses” of H5N1 influenza. *Influenza Other Respi Viruses*. 2009; 3(4):121–128. [PubMed: 19627369]
- Kruska RL, Reid RS, Thornton PK, Henninger N, Kristjanson PM. Mapping livestock oriented agricultural production systems for the developing world. *Agr Syst*. 2003; 77:39–63.
- Li KS, Guan Y, Wang J, Smith GJD, Xu KM, Duan L, Rahardjo AP, Puthavathana P, Buranathai C, Nguyen TD, Estoepangestie ATS, Chaisingh A, Auewarakul P, Long HT, Hanh NTH, Webby RJ, Poon LLM, Chen H, Shortridge KF, Yuen KY, Webster RG, Peiris JSM. Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. *Nature*. 2004; 430(6996):209–213. [PubMed: 15241415]
- Nelson, A. *Travel time to major cities: A global map of Accessibility*. Office for Official Publications of the European Communities; Luxembourg: 2008.

- Neumann K, Elbersen BS, Verburg PH, Staritsky I, Pérez-Soba M, Vries W, Rienks WA. Modelling the spatial distribution of livestock in Europe. *Landscape Ecol.* 2009; 24(9):1207–1222.
- 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(2):302–309. [PubMed: 17604193]
- Séré, C.; Steinfeld, H.; Groenewold, J. World livestock production systems: current status, issues and trends. International Livestock Research Institute; Nairobi: 1995.
- Scharlemann JPW, Benz D, Hay SI, Purse BV, Tatem AJ, Wint GRW, Rogers DJ. Global Data for Ecology and Epidemiology: A Novel Algorithm for Temporal Fourier Processing MODIS Data. *PLoS ONE.* 2008; 3(1):e1408. [PubMed: 18183289]
- Sinka M, Rubio-Palis Y, Manguin S, Patil A, Temperley W, Gething P, Van Boeckel T, Kabaria C, Harbach R, Hay SI. The dominant *Anopheles* vectors of human malaria in the Americas: occurrence data, distribution maps and bionomic précis. *Parasite Vector.* 2010; 3(1):72.
- Whittingham MJ, Stephens PA, Bradbury RB, Freckleton RP. Why do we still use stepwise modelling in ecology and behaviour? *J Anim Ecol.* 2006; 75(5):1182–1189. [PubMed: 16922854]
- World Health Organization. [Accessed April 2011] Global Alert and Response. Available online at: http://www.who.int/csr/disease/avian_influenza/country/en/
- Wint, W.; Robinson, T. Gridded Livestock of the World. Food and Agricultural Organisation; Rome: 2007.

Fig. 1.

Stepwise integration procedure of independent variables in regressions models: Each step of the model consists of integrating a pair of predictor variables (a variable and its quadratic term) that leads to the best AIC improvement of the model among the different predictors. This operation is successively repeated for each of the pairs of predictors that have not been selected previously. A new variable is thus added to the regression model after each iteration. The process stops under two conditions i) the improvement in the AIC criterion of two successive models is inferior to 1%, ii) the number of unique points within an Agro-Ecological Zone is insufficient given the number of unique points needed by variable (15 points, for linear and squared term integration).

Fig. 2.

RMSE and correlation coefficients obtained with different stratifications (top left and right) and with the different methods for combining their predictions (bottom left and right).

Different letters denote significantly different means at the $p = 0.05$ level (One-way ANOVA; RMSE: $F_{10,264} = 308.8$, $p < 0.001$; COR: $F_{10,264} = 213.71$, $p < 0.001$).

Fig. 3. Maps of observed duck density in data pooled from national statistics (top), predicted values using the Livestock Production Systems as stratification layer (middle), and variability of the predictions expressed as coefficient of variation (bottom)

Fig. 4. RMSE and coefficient of correlation of predictions vs. observed data grouped by country using the Livestock Production Systems as stratification layer.

Fig. 5. Predictions in the Mekong obtained with different levels of input data: administrative level 3, level 2, level 1, level 0, and no data.

Fig. 6.
RMSE and Correlation coefficient in the Mekong, by administrative level of training data.

Table 1

Original data source and resolution for domestic duck population density

Country	Date	Admin. Level	Original Source
China	2005	HYBRID 1–3	Animal husbandry yearbook data for Chinese province
Cambodia	2002	1	Department of Animal Health and Protection 2003 Census
Bhutan	-	1	Bhutan Ministry of Agriculture
Thailand	2004	3	Department of Livestock Development
Lao	1999	1	Ministry of Agriculture and Forestry
Vietnam	2001	3	2001 Agricultural Census Database, Ministry of Agriculture
Myanmar	2006	2	Myanmar Animal Census
Bangladesh	-	3	Bangladesh Department of Livestock Services
India	2003	2	2003 Agricultural Census
Nepal	-	3	Department of Livestock Services
Korea	2004	1	Ministry of Agriculture and Forestry
Malaysia	2004	1	Department of Veterinary Services
Philippines	2002	2	Philippines 2002 Agricultural Census
Indonesia	2005	HYBRID 1–2	Statistik Peternakan 2006

Table 2
Predictors variables used in the stepwise multiple regression modeling procedure

MODIS Imagery	
MODIS Channels	TFA Processed Channels 03,07,08,14,15,35: mx, mm, d1, d2, d3, da, a1, a2, a3, p1, p2, p3, produced by SEEG, University of Oxford
CONTINENTAL TILES OF MODELLING COVARIATES	
lkgrumpdens	Alpha version kilometer resolution human population density for 2000 from GPW GRUMP, at Columbia University
lkgrumpdensb	Beta version kilometer resolution human population density for 2000 from GPW GRUMP at Columbia University
green0301c1rc	MODIS Phenology datasets, Greenup band 1, January 2003, Boston University, Dept Geography (see text)
green0301c2rc	MODIS Phenology datasets, Greenup band 2, January 2003, Boston University, Dept Geography (see text)
senes0301c1rc	MODIS Phenology datasets, Senescence band 1, January 2003, Boston University, Dept Geography (see text)
wd1kslp	Slope, GTOPO30 dataset
lkaglgprc	Length of Growing Period, Derived from FAO LGP layers using statistical modeling by ERGO
lklhlgprc	Length of Growing Period, Derived from LGP layers produced by Thornton, using statistical modeling by ERGO
rmsuitdeg	Distance in Decimal Degrees to land suitable for Ruminants, derived by ERGO
mgsuitdeg	Distance in Decimal Degrees to land suitable for Monogastrics, derived by ERGO
lkrdsdeg	Distance in Decimal Degrees to Major Roads - using Landsat Roads layer, derived by ERGO
lkwaideg	Distance in Decimal Degrees to Sea, Major Lakes and Rivers, Derived by ERGO
glurdeg	Distance in Decimal Degrees to GRUMP alpha urban areas, Derived by ERGO
2kprecyr1k	Annual Precipitation, synoptic period to 2000, produced by Worldclim
acc50k	Travel time to major cities (>50,000) European Commission GEM
V590ELC	MODIS SRTM Elevation product, sea level corrected
V590EL	MODIS SRTM Elevation product
ZONES AND MASKS FOR MODELLING IMPLEMENTATION	
lkaream	land pixel area in meters, Produced by SDRN, FAO
lkmsuit	Area Suitable for monogastrics
lkad0cd	Country GAUL Administrative Level Zero codes
AGRO-ECOLOGICAL ZONES FOR MODELLING STRATIFICATION	
Stratification layers	
	378141514p1dem50 Ecological Zones, derived by unsupervised classification by ERGO from MODIS imagery
	378141514p1dem25 Ecological Zones, derived by unsupervised classification by ERGO from MODIS imagery

AGRO-ECOLOGICAL ZONES FOR MODELLING STRATIFICATION		
Stratification layers		
	378141514p1dem12	Ecological Zones, derived by unsupervised classification by ERGO from MODIS imagery
	378141514p1dem5	Ecological Zones, derived by unsupervised classification by ERGO from MODIS imagery
	SandS	Livestock Production Systems by Seré and Steinfeld
* GPW GRUMP = Gridded Population of the World Global Rural Urban Mapping Project		
* ERGO = Environmental Research Group, Oxford		
* SEEG = Spatial Epidemiology and Ecology Group		