

## RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

### **H7N9 and H5N1 avian influenza suitability models for China: accounting for new poultry and live-poultry markets distribution data**

Artois, Jean; Lai, Shengjie; Feng, Luzhao; Jiang, Hui; Zhou, Hang; Li, Xiangping; Dhingra, Madhur S.; Linard, Catherine; Nicolas, Gaëlle; Xiao, Xiangming; Robinson, Timothy P.; Yu, Hongjie; Gilbert, Marius

*Published in:*

Stochastic Environmental Research and Risk Assessment

*DOI:*

[10.1007/s00477-016-1362-z](https://doi.org/10.1007/s00477-016-1362-z)

*Publication date:*

2017

*Document Version*

Publisher's PDF, also known as Version of record

[Link to publication](#)

*Citation for published version (HARVARD):*

Artois, J, Lai, S, Feng, L, Jiang, H, Zhou, H, Li, X, Dhingra, MS, Linard, C, Nicolas, G, Xiao, X, Robinson, TP, Yu, H & Gilbert, M 2017, 'H7N9 and H5N1 avian influenza suitability models for China: accounting for new poultry and live-poultry markets distribution data', *Stochastic Environmental Research and Risk Assessment*, vol. 31, no. 2, pp. 393-402. <https://doi.org/10.1007/s00477-016-1362-z>

#### **General rights**


Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# H7N9 and H5N1 avian influenza suitability models for China: accounting for new poultry and live-poultry markets distribution data

Jean Artois<sup>1</sup> · Shengjie Lai<sup>2,3,4</sup> · Luzhao Feng<sup>2</sup> · Hui Jiang<sup>2</sup> · Hang Zhou<sup>2</sup> · Xiangping Li<sup>5</sup> · Madhur S. Dhingra<sup>1,6</sup> · Catherine Linard<sup>1,7</sup> · Gaëlle Nicolas<sup>1</sup> · Xiangming Xiao<sup>8</sup> · Timothy P. Robinson<sup>9</sup> · Hongjie Yu<sup>4</sup> · Marius Gilbert<sup>1,10</sup> 

Published online: 5 December 2016

© The Author(s) 2016. This article is published with open access at Springerlink.com

**Abstract** In the last two decades, two important avian influenza viruses infecting humans emerged in China, the highly pathogenic avian influenza (HPAI) H5N1 virus in the late nineties, and the low pathogenic avian influenza (LPAI) H7N9 virus in 2013. China is home to the largest population of chickens (4.83 billion) and ducks (0.694 billion), representing, respectively 23.1 and 58.6% of the 2013 world stock, with a significant part of poultry sold through live-poultry markets potentially contributing to the spread of avian influenza viruses. Previous models have looked at factors associated with HPAI H5N1 in poultry and LPAI H7N9 in markets. However, these have not been studied and compared with a consistent set of predictor variables. Significant progress was recently made in the

collection of poultry census and live-poultry market data, which are key potential factors in the distribution of both diseases. Here we compiled and reprocessed a new set of poultry census data and used these to analyse HPAI H5N1 and LPAI H7N9 distributions with boosted regression trees models. We found a limited impact of the improved poultry layers compared to models based on previous poultry census data, and a positive and previously unreported association between HPAI H5N1 outbreaks and the density of live-poultry markets. In addition, the models fitted for the HPAI H5N1 and LPAI H7N9 viruses predict a high risk of disease presence for the area around Shanghai and Hong Kong. The main difference in prediction between the two viruses concerned the suitability of HPAI H5N1 in north-China around the Yellow sea (outlined with Tianjin, Beijing, and Shenyang city) where LPAI H7N9 has not spread intensely.

Jean Artois and Shengjie Lai have contributed equally to the work.

**Electronic supplementary material** The online version of this article (doi:[10.1007/s00477-016-1362-z](https://doi.org/10.1007/s00477-016-1362-z)) contains supplementary material, which is available to authorized users.

✉ Hongjie Yu  
cfetpyhj@vip.sina.com

✉ Marius Gilbert  
mgilbert@ulb.ac.be

<sup>1</sup> Spatial Epidemiology Lab. (SpELL), Université Libre de Bruxelles, Brussels, Belgium

<sup>2</sup> Division of Infectious Disease, Key Laboratory of Surveillance and Early-warning on Infectious Disease, Chinese Center for Disease Control and Prevention, 155 Changbai Road, Changping District, Beijing 102206, China

<sup>3</sup> WorldPop, Department of Geography and Environment, University of Southampton, Southampton SO17 1BJ, UK

<sup>4</sup> Key Laboratory of Public Health Safety, Ministry of Education, School of Public Health, Fudan University, Shanghai 200032, China

<sup>5</sup> Institute of Biodiversity Science, Fudan University, Shanghai 200433, China

<sup>6</sup> Department of Animal Husbandry & Dairying, Government of Haryana, Pashudhan Bhawan, Bays No. 9-12, Sector -2, Panchkula, Haryana 134109, India

<sup>7</sup> Department of Geography, Université de Namur, Namur, Belgium

<sup>8</sup> Department of Microbiology and Plant Biology, Center for Spatial Analysis, University of Oklahoma, 101 David L. Boren Blvd, Norman, OK 73019, USA

<sup>9</sup> Livestock Systems and Environment (LSE), International Livestock Research Institute (ILRI), Nairobi, Kenya

<sup>10</sup> Fonds National de la Recherche Scientifique, Brussels, Belgium

**Keywords** Poultry data · Avian influenza · Spatial epidemiology · HPAI H5N1 · LPAI H7N9

## 1 Introduction

In high-income countries, most of the intensification of poultry production took place in the second half of the 20<sup>th</sup> century and are not changing much anymore (FAO 2009). Hotspots of intensive poultry production can be found in several states of the USA and in north-western Europe (Robinson et al. 2014), where they have been present for decades and will probably remain so for a number of years. In contrast, transition economies such as China or Brazil are intensifying their animal production in response to rising demands from more urbanized and wealthy human populations that are increasing their per capita consumption of poultry meat and eggs (Robinson and Pozzi 2011). Therefore, both the number and geographical distribution of poultry is changing much faster than that in high-income economies. This has two main consequences. First, changes in densities and geographical distribution influence the conditions of spread and evolution of infectious diseases. Higher densities generally translate into higher contact rates between animals, which, alongside other mechanisms, may explain why several disease emergences were linked to recent intensification of livestock production systems (Jones et al. 2013). Secondly, the density of host is a key variable in any epidemiological investigations, and studies carried out in economies with a fast-changing agricultural sector need to account for those changes.

China is specifically in this situation. In the last 20 years, the stock of chickens and ducks was multiplied by a factor of 2.61 and 2.36, respectively, corresponding to a compounded annual growth rate of 3.91 and 2.36% (FAO 2009). The country now holds by far the largest population of chickens, and 70% of the world's ducks. It is in this context of intensification of poultry production and fast environmental and land-use changes (Wei and Ye 2014) that two major avian influenza viruses (AIVs) infecting humans emerged in the country. In 1996, the highly pathogenic avian influenza virus (HPAIV) H5N1 was first reported in southern China (Li et al. 2004) and for several years was only found in that country. In 2003–2004, it started to spread to other countries and reached a maximum geographical range in 2006, when a cumulative number of over 60 countries had reported the presence of the virus across Asia, Europe and Africa (Hogerwerf et al. 2010). In 2013, new infections caused by a low pathogenic avian influenza virus (LPAIV) H7N9 were notified in humans, and the infections were traced back to live-poultry markets (Cowling et al. 2013; Yu et al. 2013). The H5N1 and H7N9

viruses had very different pathogenicities in poultry and both showed the capacity to infect humans, though with somewhat different epidemiological characteristics (Cowling et al. 2013; Qin et al. 2015).

Surprisingly, although chicken and duck are so critical as potential driver of disease emergence or as separate epidemiological variables in risk-factor analyses, separate statistics on their distribution are not routinely produced with a high spatial level of details in China and there exists a great degree of heterogeneity in reporting and aggregation of census data (Prosser et al. 2011). Data are collected at a fine level (typically counties) through censuses or surveys, but when these data become centralized at the prefecture (administrative level 2) or province level (administrative level 3), details are lost, either because of spatial aggregation (e.g. provincial yearbook only reporting prefecture-level aggregated data, or national yearbook only reporting province-level data), or because of thematic aggregation with chicken and ducks being pooled together in a “poultry” category. So, many country-level data are currently not centralized in a high-resolution spatial database, although they may be collected on the ground and both the spatial and temporal resolution of database can be a limiting factor for epidemiological investigation in a fast-changing sector.

Previous investigations strongly demonstrated the need to separate domestic waterfowls (ducks and geese) from gallinaceous poultry as their association with HPAI H5N1 virus presence was consistently found to differ in several Asian countries (Gilbert et al. 2006, 2008; Gilbert and Pfeiffer 2012). In addition, different types of production systems, such as extensive backyard production on the one hand and more commercial modes of production on the other, may also have different types of influence on transmission risk. Recent work on HPAI H5N1 and on LPAI H7N9 found different patterns of association between extensively and intensively raised ducks and chickens in Thailand (Van Boeckel et al. 2012) and China (Gilbert et al. 2014). So, there is a strong need for analysing both HPAI H5N1 and LPAI H7N9 data in relation to detailed and up to date separate, duck and chicken data, in addition to making a distinction between extensive and intensive production systems.

Several studies have previously investigated the spatial distributions in China of both HPAI H5N1 (Li et al. 2015b; Martin et al. 2011a; Fang et al. 2008) and LPAI H7N9 (Gilbert et al. 2014; Fang et al. 2013). In a recent study, the spatial distribution of human infections with HPAI H5N1 and H7N9 was studied and compared, which yielded insights on the areas of co-circulation and potential infections (Li et al. 2015c), and in another study with a strong emphasis on climatic factors (Li et al. 2015a). However, due to the limited poultry data availability highlighted

above, few of these studies made a distinction between extensive and intensive systems, and several included a general “poultry” category that does not differentiate chicken from ducks.

In this paper, we aimed to revisit and update previous HPAI H5N1 and LPAI H7N9 suitability models using a novel and improved set of poultry data in China separating chicken and ducks and different production systems with state-of-the-art recent downscaling methodology (Nicolas et al. 2016). In addition, we also tested live poultry market density as predictor of HPAI H5N1 outbreaks distribution, which was not done in previous studies due to the lack of available census data at the time of previous studies. So, the specific objectives were to compare the result obtained with the updated poultry data sets, to test the effect of live-poultry market density on HPAI H5N1 distribution, and to compare the geographical distribution of high HPAI H5N1 and LPAI H7N9 suitability within China.

## 2 Materials and methods

### 2.1 Poultry statistics

Census data on chicken and duck numbers at the end of the calendar year 2011 and 2012, and the numbers of individuals sold per year were obtained from three sources: (a) published yearbooks, such as the China Animal Husbandry Yearbook, Statistic Yearbook of China or provincial yearbooks (e.g. <http://data.stats.gov.cn/>); (b) the official website of the Ministry of Agriculture of China (<http://english.agri.gov.cn/>) and the Agricultural Bureaus at province and prefecture level; (c) contact with provincial Bureaus of Animal Husbandry, provincial Departments of Commerce, Statistics Bureaus and Chinese Agricultural Universities to obtain any data not available from sources (a) or (b). These data were mostly available at the prefecture level (administrative level 2) but not consistently for all prefectures and provinces. For example, a prefecture could have a certain value in the 2011 data set, but not in the 2012 one. For a few provinces, namely Anhui, Jiangsu and Zhejiang in Eastern China, we were also able to find county-level data for the year 2010.

Despite the exploration of the available above data sources, we were not able to find 2010, 2011 or 2012 data for all provinces and prefectures. In a previous study, Prosser et al. (2011) already created a composite poultry dataset for chicken, duck and geese, based on census statistics from 2003 and 2005, and this composite data set remains to date the basis of the poultry data for China in the Gridded Livestock of the World (GLW) database. The term ‘composite’ refers to the fact that they already had to combine data from different spatial scale: province,

prefecture and counties. Therefore, we combined our recent data set with the one from Prosser et al. (2011) in case of complete absence of information in the concerned area. Namely, for each of the counties in China, we checked different available data sources and used the density data from the source with the following order of priority: (i) a 2010 value from the county-level data set (available only in Anhui, Jiangsu and Zhejiang), (ii) a 2011 or 2012 value from the prefecture-level data set, and (iii) a value from the Prosser et al. (Prosser et al. 2011) data set. In addition, when an administrative level 2 (prefecture) count was available in 2011 or 2012 in an area that had administrative level 3 (counties) data in the Prosser et al. (Prosser et al. 2011) data set, we used the county-level data corrected to match the 2011 or 2012 prefecture counts by multiplying them by a single scalar. In order to account for differences in reference years, we applied national-level growth correction factors from FAOSTAT (FAO 2009) to account for the differences in years, with 2010 as pivot year. A similar approach was used for both chicken and ducks. Due to a lack of data for geese, we did not consider that species in further processing.

### 2.2 Poultry downscaling

In order to avail the poultry data at the same spatial resolution as other risk factors, we used the downscaling methodology of the Gridded Livestock of the World (GLW), which was fully described in Robinson et al. (Robinson et al. 2014) and Nicolas et al. (Nicolas et al. 2016). This method relies on models being built based on the census counts and a set of covariates, and typically models species at the continental level (Robinson et al. 2014; Nicolas et al. 2016). Therefore, models can be trained using the data outside a specific country. In order to benefit from training chicken and duck data in countries near China, we extracted polygons with chicken and duck census counts from the GLW within an arbitrarily 500 km buffer from China’s border, to benefit from good quality input data from areas with similar agro-ecological conditions. The GLW methodology is only briefly summarized here. First, the density of animals per km<sup>2</sup> of suitable land is estimated in all polygons corresponding to the sub-national poultry census data and transformed to its logarithmic value (base 10). Second, a large set of sample points is built to cover the modelling extent, and values for the observed densities and predictor variables were extracted from their respective polygons (census data) or pixels (predictor variables). Third, the sample file was divided into  $n$  sub-samples for bootstrapping the analysis, and each sub-sample file was divided in two parts, one for building the model with 70% of the polygons, and one for evaluating the model goodness of fit with 30% of the polygons.

Fourth, each sub-sample was used to build a Random Forest model and the model was applied to the raster imagery to obtain a single predicted value for each pixel. Fifth, the predicted values were averaged over the  $n$  bootstraps. Finally, post-processing was carried out to correct pixels values by multiplying all pixels of a particular census polygon by a constant so that the sum of the grid cell values within the polygon was equal to the observed totals in the input subnational census data. Finally, the pixel values were also corrected so that the national total matches the FAOSTAT official total for a specified base year, in this case 2010. The spatial covariates used to make the predictions include Fourier-transformed remotely sensed variables (the normalized difference vegetation index and enhanced vegetation index, the day and night land surface temperature and the band 3 shortwave infrared band), eco-climatic variables (length of growing period and annual precipitation), topographic variables (elevation and slope) and anthropogenic variables (human population density and travel time to major cities). We used exactly the same set of predictor variables as described in Robinson et al. (Robinson et al. 2014) and Nicolas et al. (Nicolas et al. 2016), with the exception of the human population density, where the recent 2010 human population data published for China by the Worldpop project was used (Gaughan et al. 2016). The chicken density layer was finally broken down between extensively and intensively raised chickens, following the methodology outlined in Gilbert et al. (Gilbert et al. 2015).

### 2.3 Avian influenza data

For the HPAI H5N1 models, we used the outbreak locations from the epidemiological dataset described in Martin et al. (Martin et al. 2011a), complemented by outbreak locations extracted from the FAO Empres-i data base, including 76 recent records (from 2009 to 2016) of HPAI H5N1 infection in domestic poultry (Claes et al. 2014). Boosted regression trees (BRT) models require data on both presence and absence and pseudo-absences were generated throughout the country in locations: (1) where there was no evidence of previous HPAI H5N1 presence; (2) at a minimum distance of 0.0833 decimal degree of any positive (which correspond to the spatial resolution of the poultry density layer); and (3) in a location where human and poultry density was higher than five person or bird per  $\text{km}^2$  to exclude desert, unpopulated areas and areas with potentially very low surveillance from the analysis. There is no consensus on the optimal number of pseudo-absences to be used in niche modelling methods (Barbet-Massin et al. 2012). This number depends of the species under consideration, type of model used and the spatial extent of data. Pseudo-absences were generated in much greater

numbers than HPAI outbreaks (eight times more negatives than positives) to capture enough variability in data and optimizing the model performance.

For the LPAI H7N9 models, we used the set of positive and negative markets described in Gilbert et al. (Gilbert et al. 2014), complemented by H7N9 presence location recorded by China CDC up to 01/10/2015 (first three epidemic waves). As the epidemiological unit is the market and the data set included absence points (markets where H7N9 was never recorded), there was no need to distribute pseudo-absences. The spatial locations of markets described above was used to create a layer of market density ( $\text{market}/\text{km}^2$ ) on a grid of 0.0833 decimal degree of resolution.

### 2.4 Avian influenza suitability modelling

Boosted regression tree (BRT) models were used to model the probability of presence of HPAI H5N1 outbreaks at the pixel level and probability of LPAI H7N9 infection at markets level (Elith et al. 2008). The method is increasingly used in suitability modelling of infectious (Gilbert et al. 2014; Pigott et al. 2014) and vector-borne diseases (Bhatt et al. 2013) for its capacity to model interactions between variables as well as non-linear relationships between the outcome and predictor variables. Each model was evaluated with an eightfolds cross-validation procedure and the area under the receiver operating characteristic curve used as a measurement of the discriminatory capacity of models (Elith et al. 2008). The data set is split in 8 sub-data and a single sub-data is retained as the validation data and the remaining 7 sub-data are used as training data. The cross-validation is then repeated 8 times for each BRT model. Finally, in order to account for sources of uncertainty in the model (on the localisation of pseudo-absences and the data splitting of cross-validation), the analysis was bootstrapped with 15 independent BRT run for a total of 120 cross-validations (15 runs  $\times$  eightfolds).

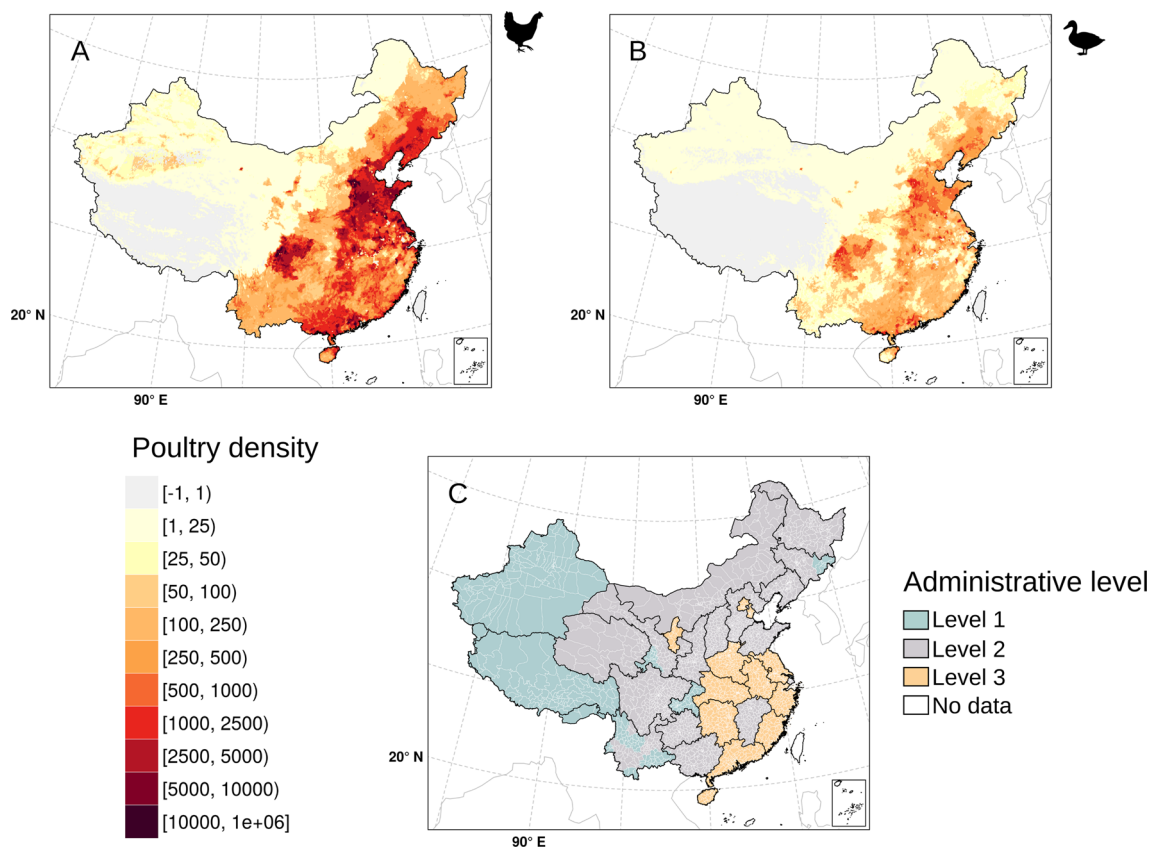
In order to ensure comparability with the previously published results, we used a similar set of predictor variables as in Martin et al. (Martin et al. 2011a) and Gilbert et al. (Gilbert et al. 2014), including the chicken and duck density data layers produced by the processing detailed in the previous section, the market density layer (Gilbert et al. 2014), human population density from the Worldpop project ([www.worldpop.org](http://www.worldpop.org)) (Gaughan et al. 2016), the proportion of land covered by water and the proportion of land covered by rice cropping from the GlobCover database (Bicheron et al. 2008) and the cropping intensity established through remote sensing (Xiao et al. 2005, 2006). So, the final set of predictor variables included: extensively raised chicken density (ChExtDn,  $\text{heads}/\text{km}^2$ ), intensively raised chicken density (ChIntDn,  $\text{heads}/\text{km}^2$ ), duck density

(heads/km<sup>2</sup>), live-poultry market density (MktDn, markets/km<sup>2</sup>), human population density (HpopDn, people/km<sup>2</sup>), the cropping intensity (CropInt, crop cycles/year), the proportion of area covered by rice paddy fields (RiceCov, %) and the proportion of area covered by permanent water (WatCov, %). Since a market’s epidemiological situation reflects the potential circulation of viruses in its catchment areas, a spatial filter was applied to each predictor variable, following the procedure outlined in Gilbert et al. (Gilbert et al. 2014). In this step, the covariate values were smoothed with weights determined by a Gaussian kernel and the parameter  $\sigma$  representing, the size of the catchment area. A range of values for the parameter  $\sigma$  was tested in Gilbert et al. (Gilbert et al. 2014) and the same figure was used in this study ( $\sigma = 0.7$ ). In addition, Gilbert et al. (2014) used modelled live-poultry market densities as predictor variables in their model, so that they could extrapolate their model to the rest of Asia. For the sake of comparability, we used the same layer, but the raw observed number of live-poultry market by pixel was also tested to evaluate the potential effect of the modelling procedure.

### 3 Results

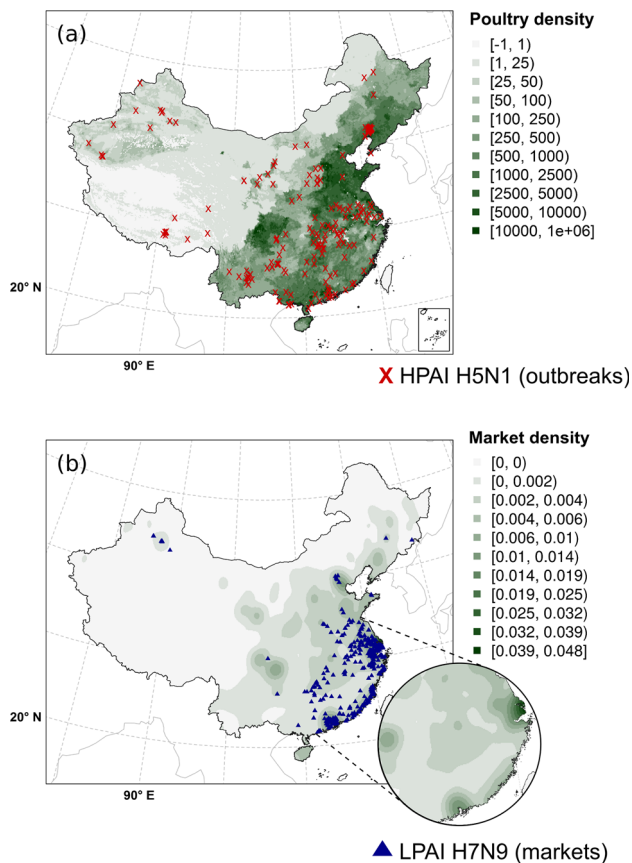
The new poultry distribution is displayed in Fig. 1, showing generally much higher densities of chicken over ducks, along with the final level at which input census data was available. The distribution of ducks in China largely follows that of chickens, with some of the highest densities observed in the northeastern provinces of Shanxi, Hebei, Shangdong, south of Beijing, and the southern provinces of Guangxi and Guangdong. In central China, the Sichuan province stands out as having particularly high densities too. Figure 2 displays the density of poultry and live bird markets with the distribution of HPAI H5N1 and LPAI H7N9 cases. The live bird market density is high on the east coast of China and inland, around some cities as Chongqing and Lanzhou.

The relative influences in the BRT models of the predictor variables as well as their profiles are displayed in Fig. 3. For HPAI H5N1 outbreaks, the variables with the highest relative contribution (RC) were the live-poultry market density (positive association, RC 34.5%), human population density (positive association, RC 23.4%), and



**Fig. 1** Overview of base poultry data resolution. The new data sets of chicken (a) and duck (b) density (heads/km<sup>2</sup>—on a logarithmic scale of base 10) are obtained by combining recent census data sets at different spatial levels (c; level 1: province; level 2: prefecture; level

3: county). This figure was built with the R-3.3.1 software (<https://cran.r-project.org/>). The graticule is composed of a 10-degree increments and the coordinate system is ‘SR-ORG:7564’



**Fig. 2** Distribution of HPAI H5N1 outbreaks (**a**; red cross) and LPAI H7N9 infected markets (**b**; blue triangles) in China included in this study. The density of poultry (the sum between the chicken and the duck density) and live bird markets (smoothed) are also displayed in the maps (**a**) and (**b**) respectively. The density of live bird markets was smoothed with weights determined by a Gaussian kernel and the parameter  $\sigma$  representing the size of the catchment area ( $\sigma = 0.7$ ). This figure was built with the R-3.3.1 software (<https://cran.r-project.org/>). The graticule is composed of a 10-degree increments and the coordinate system is ‘SR-ORG:7564’

cropping intensity (negative association, RC 17.5%). One should note that when the raw number of live-poultry markets is used as predictor instead of the one modeled in Gilbert et al. (2014), both the human population density and live-poultry market density remain the most important factors, but their respective relative contribution is inverted, i.e. human population density has the highest RC (SI 1 Fig A, B).

For LPAI H7N9 infected markets, the predictor variables with the highest relative contributions were the poultry market density (20.2%) followed by the density of intensively raised chicken (18.4%), the proportion of land covered by water (14.8%) and human population density (10.5%). Interestingly, when repeating the analysis by breaking down LPAI H7N9 records by seasonal epidemic waves, the relative contribution of live-poultry market density in the LPAI H7N9 market models tended to

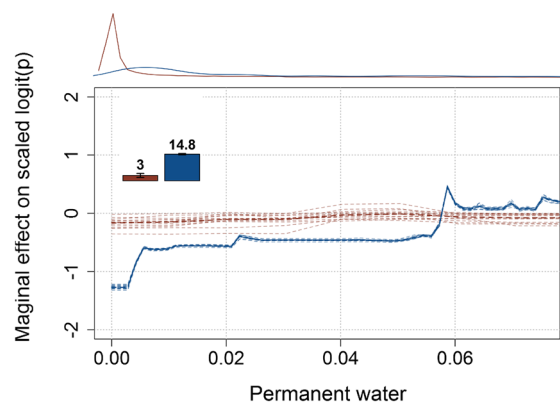
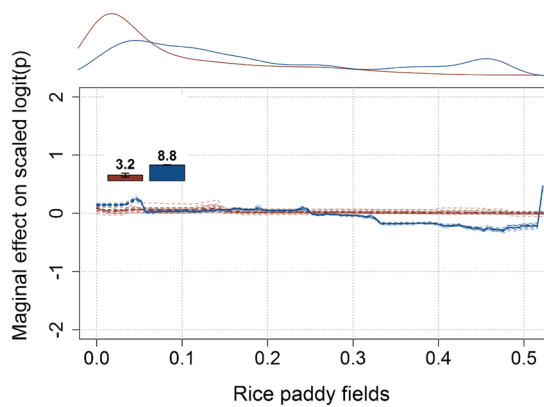
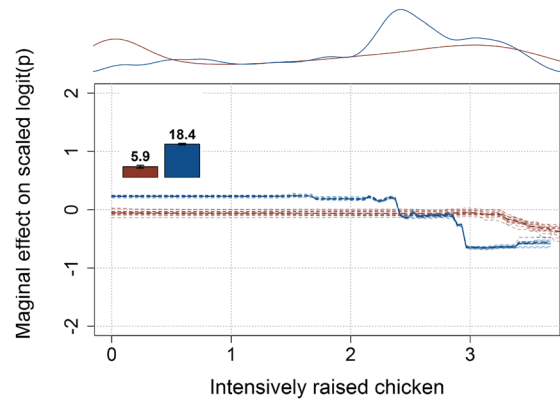
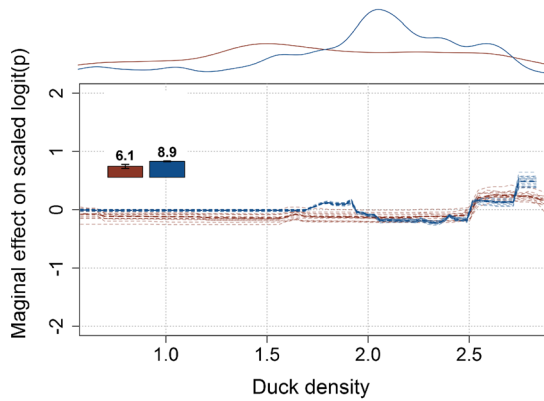
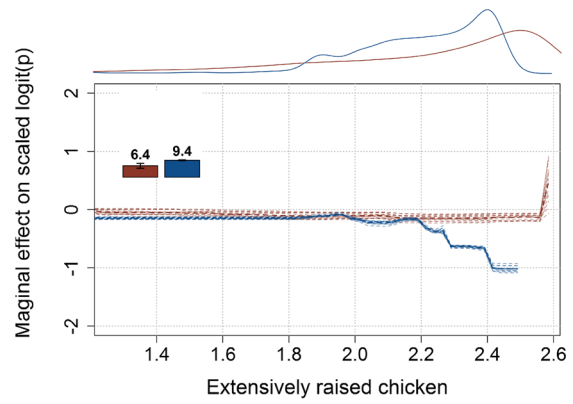
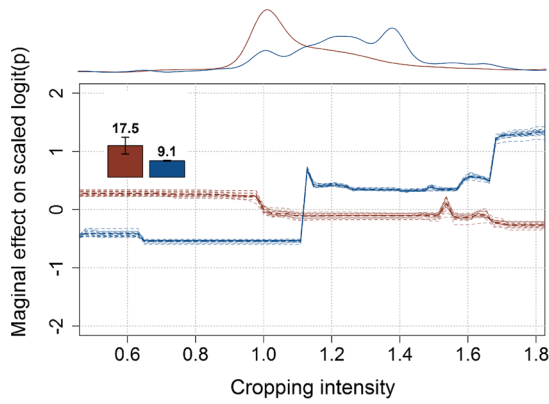
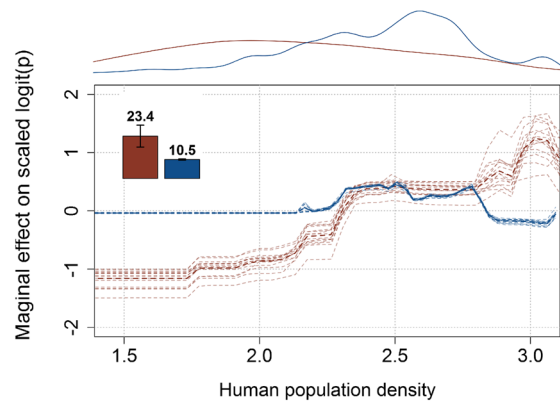
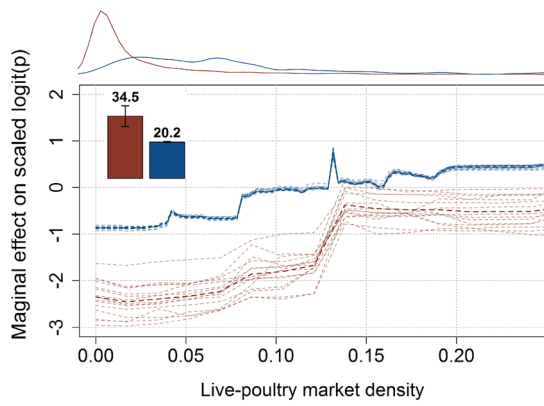
**Fig. 3** Relative contribution (*bar plots*) and partial dependent plot (*curves*) of each predictor of the BRT models of HPAI H5N1 outbreaks (*red*) and LPAI H7N9 infected markets (*blue*). The relative contribution of each predictor is scaled so that the sum of all predictor variables adds to 100%, and measures the number of times a predictor is selected for splitting the dataset over the trees. The partial dependent plot gives a graphical description of the marginal effect of a predictor on the predicted response. The opaque line represents the mean marginal effect, whilst transparent lines represent each bootstrap. On the top of each graph, the density function of the observed distribution of predictors is displayed for one bootstrap and for the two analyses (*red* HPAI H5N1; *blue* LPAI H7N9)

decrease over time, with values of 22.5, 17.7 and 11.7 for the 2012/2013, 2013/2014 and 2014/2015 winters epidemic waves, respectively (SI 2 Fig. A). Markets reporting LPAI H7N9 virus infections are associated positively with the proportion of area covered by water and the market density as observed in the BRT profiles of Fig. 3. The opposite trend is observed for the chicken density layers, which are negatively associated with LPAI H7N9 presence. The effect of using the raw number of live-poultry market per pixel instead of the modeled one was a reduction in its RC from 20.2 to 17.9, and it remained the top predictor in terms of RC.

HPAI H5N1 and LPAI H7N9 suitability maps are displayed in Fig. 4a and b, respectively. While LPAI H7N9 remained constrained to the two hotspots of Shanghai and Guangdong, coastal areas and a number of small and isolated pockets with higher suitability in and around inland cities, HPAI H5N1 probability of presence was found to be distributed over more widespread zones in inland China, with much higher probabilities of presence in large rural areas. Both models had comparable goodness of fit measurements, with AUC values of  $0.885 \pm 0.039$  and  $0.850 \pm 0.024$  for the HPAI H5N1 and LPAI H7N9 models respectively.

## 4 Discussion

Overall, the changes due to the effect of the improved poultry variables were relatively limited, and provided several results consistent with the previous analyses (Martin et al. 2011a; Gilbert et al. 2014). For HPAI H5N1 outbreaks, this involved a predominance of anthropogenic factors (human population density or live poultry market density) with a relatively limited influence of poultry variables. One can note, however, a slight increase in the marginal effect linked to the highest duck densities ( $>300$  ducks/km<sup>2</sup>;  $\sim 2.48$  on a logarithmic scale) and a decrease linked to the highest densities of intensively raised chickens ( $>1500$  chickens/km<sup>2</sup>;  $\sim 3.18$  on a logarithmic scale). For LPAI H7N9, the results are also fairly similar, and focusing on changes in the effect of the poultry factors, we



note a slightly higher RC of duck density, with a higher marginal effect linked to duck densities  $>300$  ducks/km<sup>2</sup>, and the confirmation of the negative association with intensively and extensively raised chicken. It is in fact quite surprising not to find a stronger association between HPAI H5N1 outbreaks and different poultry variable, as such association have consistently been found elsewhere. It is hard to say if this relates to a yet unsatisfying quality of the poultry data, differences in reporting, or whether this may be due to a true absence of statistical association with data adequately reflecting the situation on the ground.

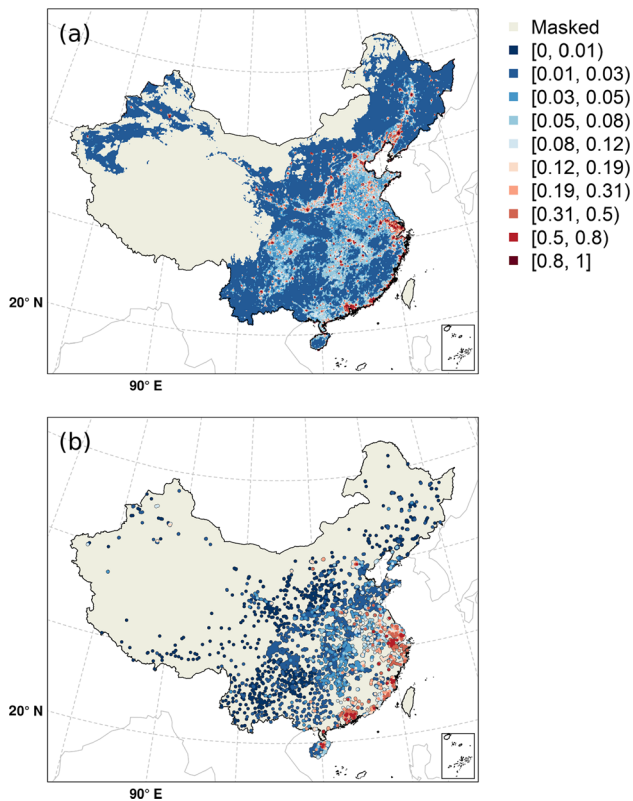
One noticeable change to the HPAI H5N1 model was the inclusion of the live-poultry market density, which was found to be a strong predictor of HPAI H5N1 presence, when used in its modelled (RC = 34.5%) or raw (RC = 17.9%) form. The distribution of live-poultry markets and human population are strongly correlated, so their respective effects are in fact quite difficult to separate on a purely statistical ground. However, the fact that both appeared as strongly significant suggests that they may exert simultaneous influences. This statistical association fits with many recent results on live-poultry market networks highlighting their possible key role in HPAI H5N1 persistence. This was suggested by simulation studies (Fournié et al. 2013) and social-network analysis of live-poultry market networks in China (Martin et al. 2011b) and Vietnam (Magalhães et al. 2010; Fournié et al. 2012) but was never quantitatively demonstrated on such a large spatial scale. One should note that this was made possible thanks to the important effort of collecting live-poultry market census data following the emergence of LPAI H7N9 (Gilbert et al. 2014). Many countries where HPAI H5N1 persisted over long periods of time have a large part of their poultry being traded through live-poultry markets, including China, Vietnam, Indonesia, Bangladesh or Egypt. By contrast, Thailand, for example, which presents all risk factors usually associated with high HPAI H5N1 risk such as high free-grazing duck density, dense irrigated areas, co-existing extensive and intensive poultry systems and human population density, has very few live-poultry markets, for cultural reasons, and this may contribute to explain the success of the country in eradicating the disease. Our results support the suggestions made recently by several authors that focusing surveillance and control in markets and adapting their management to include cleaning and/or closing day might be the key to preventing HPAI H5N1 persistence too (Fournié et al. 2011, 2013).

Interestingly, we also tentatively showed that the relative contribution of live-poultry market density to the LPAI H7N9 market models reduced over time, and this pattern is observed alongside an overall reduction in the number of human cases noted in the last years. The specific objective of this analysis was not to make a full assessment of the

model over time, which could be explored in future works. However, they may support the hypothesis that enhanced surveillance, control and management in markets after the first epidemic wave may have reduced their role in disease transmission. For example, previous papers have already showed a clear association between the timing of market closure and reductions in human cases (Yu et al. 2013; Wu et al. 2014). Our live-poultry market database does not take markets closures or opening status into account, and some markets may have changed their management practices, disappeared or have been closed over time. For example, much trading of live-poultry was banned in markets in the periphery of Beijing, but this was not accounted for in our model, which still highlights Beijing as being a potential local hotspot due to its high density of markets in our dataset.

In terms of geographical distributions, the resulting HPAI H5N1 suitability map reflects the higher contribution of live-poultry market density by producing a much more clustered distribution of suitability than that predicted by Martin et al. (2011a). Our HPAI H5N1 outbreak suitability map does not overlap so much with the recent H5N1 human infection risk map produced by Li et al. (2015c), who highlighted high risk regions for human infections as being mostly concentrated in southern China. This may be partly explained by the difference in outcome: they studied the distribution of human cases and we investigated the distribution of outbreaks in poultry. However, this does not entirely explain why there would be fewer human cases in Northern China if the landscape is suitable for H5N1 infections in poultry and if outbreaks were reported there. However recent work comparing the epidemiology of H7N9 and H5N1 viruses suggests that the susceptibility to HPAI H5N1 virus infections in humans could be more limited and family-based than for H7N9 (Qin et al. 2015). This could add an uncertainty in the link between human cases and the underlying circulation in poultry.

HPAI H5N1 suitability showed a much more widespread distribution than LPAI H7N9, which remains largely constrained to the southeastern coastal areas. There are many live-poultry markets in inland China (including Chengdu, Chongqing, Beijing and Shenyang cities) and these remain at relatively lower risk compared to the southeastern and coastal hotspot areas. When the new H7N9 virus emerged in China, there was an anticipation of a possible geographical expansion into other Asian countries that had many suitable areas for infection beyond China (Gilbert et al. 2014). Yet, after more than three years of seasonal and winter epidemic waves in humans, the disease has not spread much within China or internationally (or has not been observed). This may possibly relate to a yet unclear reservoir of the virus. Despite very large sampling for surveillance carried out in China, LPAI H7N9



**Fig. 4** Predicted maps of the probability of presence of HPAI H5N1 outbreaks (*top*) and the probability for a market of being infected by LPAI H7N9 (*bottom*). Note that infection risk is estimated as the probability that a pixel (HPAI H5N1) or market (LPAI H7N9) would be infected. The mask corresponds to the areas where human and poultry density was lower than five persons, heads/km<sup>2</sup>. This figure was built with the R-3.3.1 software (<https://cran.r-project.org/>). The graticule is composed of a 10-degree increments and the coordinate system is ‘SR-ORG:7564’

was only rarely found in poultry farms, and most positives were from human cases traced back to live-poultry markets or from sampling carried out within live-poultry markets themselves. LPAI H7N9 virus may show some specificity toward some particular chicken breeds, such as the yellow chicken, that are only raised and traded in those areas of suitability. However, this remains very speculative, and a clear understanding of the true poultry reservoir of LPAI H7N9 is still lacking. Another likely reason of the various observed distributions of LPAI H7N9 and HPAI H5N1 may be related to the low or highly pathogenic nature of the viruses which drive differences in transmission and spread which may translate into corresponding surveillance and intervention strategies for prevention and control.

In conclusion, this study found that improvements to the poultry census data had limited impact on the outputs of suitability models for HPAI H5N1 and LPAI H7N9, that a strong and positive association between HPAI H5N1 and

live-poultry markets was quantified for the first time, and that the distribution of both the HPAI H5N1 and LPAI H7N9 suitability show several areas in common, in particular in the Shanghai and Guangdong areas, which are both areas of rapid recent economic development (Yue et al. 2014).

**Acknowledgments** This study was supported by grants from the National Science Fund for Distinguished Young Scholars (81525023) by the US National Institutes of Health (1R01AI101028-02A1) by the FNRS project ‘Mapping people and livestock’ (PDR T.0073.13) and by the CGIAR Research Programmes on the Humid Tropics; Climate Change, Agriculture and Food Security (CCAFS) and Agriculture for Nutrition and Health (A4NH).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Barbet-Massin M, Jiguet F, Albert CH, Thuiller W (2012) Selecting pseudo-absences for species distribution models: how, where and how many? *Methods Ecol Evol* 3(2):327–338
- Bhatt S, Gething PW, Brady OJ et al (2013) The global distribution and burden of dengue. *Nature* 496:504–507. doi:10.1038/nature12060
- Bicheron P, Defourny P, Brockmann C et al (2008) Globcover: products description and validation report. MEDIAS, Toulouse
- Claes F, Kuznetsov D, Liechti R et al (2014) The EMPRES-i genetic module: a novel tool linking epidemiological outbreak information and genetic characteristics of influenza viruses. *Database*. doi:10.1093/database/bau008
- Cowling BJ, Jin L, Lau EHY et al (2013) Comparative epidemiology of human infections with avian influenza A H7N9 and H5N1 viruses in China: a population-based study of laboratory-confirmed cases. *Lancet* 382:129–137. doi:10.1016/S0140-6736(13)61171-X
- Elith J, Leathwick JR, Hastie T (2008) A working guide to boosted regression trees. *J Anim Ecol* 77:802–813. doi:10.1111/j.1365-2656.2008.01390.x
- Fang LQ, de Vlas SJ, Liang S et al (2008) Environmental factors contributing to the spread of H5N1 avian influenza in mainland China. *PLoS ONE* 3:e2268
- Fang L-Q, Li X-L, Liu K et al (2013) Mapping spread and risk of avian influenza A (H7N9) in China. *Sci Rep*. doi:10.1038/srep02722
- FAO (2009) FAO statistical database. <http://faostat.fao.org/>
- Fournié G, Guitian FJ, Mangtani P, Ghani AC (2011) Impact of the implementation of rest days in live bird markets on the dynamics of H5N1 highly pathogenic avian influenza. *J R Soc Interface* 8:1079–1089. doi:10.1098/rsif.2010.0510
- Fournié G, Guitian J, Desvaux S et al (2012) Identifying live bird markets with the potential to act as reservoirs of avian influenza A (H5N1) virus: a survey in northern Viet Nam and Cambodia. *PLoS ONE* 7:e37986. doi:10.1371/journal.pone.0037986
- Fournié G, Guitian J, Desvaux S et al (2013) Interventions for avian influenza A (H5N1) risk management in live bird market

- networks. *Proc Natl Acad Sci USA* 110:9177–9182. doi:[10.1073/pnas.1220815110](https://doi.org/10.1073/pnas.1220815110)
- Gaughan AE, Stevens FR, Huang Z et al (2016) Spatiotemporal patterns of population in mainland China, 1990 to 2010. *Sci Data*. doi:[10.1038/sdata.2016.5](https://doi.org/10.1038/sdata.2016.5)
- Gilbert M, Pfeiffer DU (2012) Risk factor modelling of the spatio-temporal patterns of highly pathogenic avian influenza (HPAIV) H5N1: a review. *Spat Spatiotemporal Epidemiol* 3:173–183. doi:[10.1016/j.sste.2012.01.002](https://doi.org/10.1016/j.sste.2012.01.002)
- Gilbert M, Chaitaweesub P, Parakarnawongsa T et al (2006) Free-grazing ducks and highly pathogenic avian influenza, Thailand. *Emerg Infect Dis* 12:227–234
- Gilbert M, Xiao X, Pfeiffer DU et al (2008) Mapping H5N1 highly pathogenic avian influenza risk in Southeast Asia. *Proc Natl Acad Sci USA* 105:4769–4774. doi:[10.1073/pnas.0710581105](https://doi.org/10.1073/pnas.0710581105)
- Gilbert M, Golding N, Zhou H et al (2014) Predicting the risk of avian influenza A H7N9 infection in live-poultry markets across Asia. *Nat Commun*. doi:[10.1038/ncomms5116](https://doi.org/10.1038/ncomms5116)
- Gilbert M, Conchedda G, Van Boeckel TP et al (2015) Income disparities and the global distribution of intensively farmed chicken and pigs. *PLoS ONE* 10:e0133381. doi:[10.1371/journal.pone.0133381](https://doi.org/10.1371/journal.pone.0133381)
- Hogerwerf L, Wallace RG, Ottaviani D et al (2010) Persistence of highly pathogenic avian influenza H5N1 virus defined by agro-ecological niche. *EcoHealth* 7:213–225. doi:[10.1007/s10393-010-0324-z](https://doi.org/10.1007/s10393-010-0324-z)
- Jones BA, Grace D, Kock R, Alonso S, Rushton J, Said MY, McKeever D, Mutuab F, Youngb J, McDermottb J, Pfeiffer DU (2013) Zoonosis emergence linked to agricultural intensification and environmental change. *Proc Natl Acad Sci* 110(21):8399–8404
- Li KS, Guan Y, Wang J et al (2004) Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. *Nature* 430:209–213
- Li J, Rao Y, Sun Q et al (2015a) Identification of climate factors related to human infection with avian influenza A H7N9 and H5N1 viruses in China. *Sci Rep* 5:18094. doi:[10.1038/srep18094](https://doi.org/10.1038/srep18094)
- Li X-L, Liu K, Yao H-W et al (2015b) Highly pathogenic avian influenza H5N1 in Mainland China. *Int J Environ Res Public Health* 12:5026–5045. doi:[10.3390/ijerph120505026](https://doi.org/10.3390/ijerph120505026)
- Li X-L, Yang Y, Sun Y et al (2015c) Risk distribution of human infections with avian influenza H7N9 and H5N1 virus in China. *Sci Rep* 5:18610. doi:[10.1038/srep18610](https://doi.org/10.1038/srep18610)
- Magalhães RJS, Ortiz-Pelaez A, Thi KL et al (2010) Associations between attributes of live poultry trade and HPAI H5N1 outbreaks: a descriptive and network analysis study in northern Vietnam. *BMC Vet Res* 6:10. doi:[10.1186/1746-6148-6-10](https://doi.org/10.1186/1746-6148-6-10)
- Martin V, Pfeiffer DU, Zhou X et al (2011a) Spatial distribution and risk factors of highly pathogenic avian influenza (HPAI) H5N1 in China. *PLoS Pathog* 7:e1001308. doi:[10.1371/journal.ppat.1001308](https://doi.org/10.1371/journal.ppat.1001308)
- Martin V, Zhou X, Marshall E et al (2011b) Risk-based surveillance for avian influenza control along poultry market chains in South China: the value of social network analysis. *Prev Vet Med* 102:196–205. doi:[10.1016/j.prevetmed.2011.07.007](https://doi.org/10.1016/j.prevetmed.2011.07.007)
- Nicolas G, Robinson TP, Wint GRW et al (2016) Using random forest to improve the downscaling of global livestock census data. *PLoS ONE* 11:e0150424. doi:[10.1371/journal.pone.0150424](https://doi.org/10.1371/journal.pone.0150424)
- Pigott DM, Golding N, Mylne A et al (2014) Mapping the zoonotic niche of Ebola virus disease in Africa. *eLife*. doi:[10.7554/eLife.04395](https://doi.org/10.7554/eLife.04395)
- Prosser DJ, Wu J, Ellis EC et al (2011) Modelling the distribution of chickens, ducks, and geese in China. *Agric Ecosyst Environ* 141(3–4):381–389
- Qin Y, Horby PW, Tsang TK et al (2015) Differences in the epidemiology of human cases of avian influenza A(H7N9) and A(H5N1) viruses infection. *Clin Infect Dis* 61:563–571. doi:[10.1093/cid/civ345](https://doi.org/10.1093/cid/civ345)
- Robinson T, Pozzi F (2011) Mapping supply and demand for animal-source food to 2030. *Animal Production and Health Working Paper, No. 2*. FAO, Rome
- Robinson TP, Wint GRW, Conchedda G et al (2014) Mapping the global distribution of livestock. *PLoS ONE* 9:e96084. doi:[10.1371/journal.pone.0096084](https://doi.org/10.1371/journal.pone.0096084)
- Van Boeckel TP, Thanopongtharm W, Robinson T et al (2012) Improving risk models for avian influenza: the role of intensive poultry farming and flooded land during the 2004 Thailand epidemic. *PLoS ONE* 7:e49528. doi:[10.1371/journal.pone.0049528](https://doi.org/10.1371/journal.pone.0049528)
- Wei YD, Ye X (2014) Urbanization, urban land expansion and environmental change in China. *Stoch Env Res Risk Assess* 28(4):757–765
- Wu P, Jiang H, Wu JT et al (2014) Poultry market closures and human infection with influenza A(H7N9) virus, China, 2013–14. *Emerg Infect Dis* 20:1891–1894. doi:[10.3201/eid2011.140556](https://doi.org/10.3201/eid2011.140556)
- Xiao X, Boles S, Liu J et al (2005) Mapping paddy rice agriculture in southern China using multi-temporal MODIS images. *Remote Sens Environ* 95:480–492
- Xiao X, Boles S, Froelking S et al (2006) Mapping paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images. *Remote Sens Environ* 100:95–113. doi:[10.1016/j.rse.2005.10.004](https://doi.org/10.1016/j.rse.2005.10.004)
- Yu H, Wu JT, Cowling BJ et al (2013) Effect of closure of live poultry markets on poultry-to-person transmission of avian influenza A H7N9 virus: an ecological study. *Lancet*. doi:[10.1016/S0140-6736\(13\)61904-2](https://doi.org/10.1016/S0140-6736(13)61904-2)
- Yue W, Fan P, Wei YD, Qi J (2014) Economic development, urban expansion, and sustainable development in Shanghai. *Stoch Env Res Risk Assess* 28(4):783–799