Agriculture, Ecosystems and Environment 141 (2011) 381–389

Contents lists available at ScienceDirect

Agriculture, Ecosystems and Environment

journal homepage: www.elsevier.com/locate/agee

Modelling the distribution of chickens, ducks, and geese in China

Diann J. Prosser[a,b,*], Junxi Wu[c], Erle C. Ellis[d], Fred Gale[e], Thomas P. Van Boeckel[f,g], William Wint[h], Tim Robinson[i], Xiangming Xiao[j], Marius Gilbert[f,g]

[a] United States Geological Survey, Patuxent Wildlife Research Center, Baltimore Avenue 10300, Beltsville, MD 20705, USA
[b] University of Maryland, College Park, MD 20740, USA
[c] Institute of Geographic Sciences & Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China
[d] Department of Geography & Environmental Systems, University of Maryland, Baltimore County, Baltimore, MD 21250, USA
[e] United States Department of Agriculture, Economic Research Service, Washington, D.C. 20250, USA
[f] Biological Control and Spatial Ecology, Université Libre de Bruxelles CP160/12, Av FD Roosevelt 50, B1050 Brussels, Belgium
[g] Fonds National de la Recherche Scientifique, rue d'Egmont 5, B-1000 Brussels, Belgium
[h] Environmental Research Group Oxford, P.O. Box 346, Oxford OX1 3QE, United Kingdom
[i] Food and Agriculture Organization of the United Nations (FAO), Viale delle Terme di Caracalla, 00100 Rome, Italy
[j] Department of Botany and Microbiology, Center for Spatial Analysis, University of Oklahoma, Norman, OK 73019, USA

ARTICLE INFO

Article history:
Received 2 December 2010
Received in revised form 1 April 2011
Accepted 4 April 2011
Available online 30 April 2011

Keywords:
Poultry
China
Distribution modelling
Population estimates
GIS
Epidemiology

ABSTRACT

Global concerns over the emergence of zoonotic pandemics emphasize the need for high-resolution population distribution mapping and spatial modelling. Ongoing efforts to model disease risk in China have been hindered by a lack of available species level distribution maps for poultry. The goal of this study was to develop 1 km resolution population density models for China's chickens, ducks, and geese. We used an information theoretic approach to predict poultry densities based on statistical relationships between poultry census data and high-resolution agro-ecological predictor variables. Model predictions were validated by comparing goodness of fit measures (root mean square error and correlation coefficient) for observed and predicted values for 1/4 of the sample data which were not used for model training. Final output included mean and coefficient of variation maps for each species. We tested the quality of models produced using three predictor datasets and 4 regional stratification methods. For predictor variables, a combination of traditional predictors for livestock mapping and land use predictors produced the best goodness of fit scores. Comparison of regional stratifications indicated that for chickens and ducks, a stratification based on livestock production systems produced the best results; for geese, an agro-ecological stratification produced best results. However, for all species, each method of regional stratification produced significantly better goodness of fit scores than the global model. Here we provide descriptive methods, analytical comparisons, and model output for China's first high resolution, species level poultry distribution maps. Output will be made available to the scientific and public community for use in a wide range of applications from epidemiological studies to livestock policy and management initiatives.

Published by Elsevier B.V.

1. Introduction

Globalization and a growing demand for meat products in developing regions in recent years have led to rapid expansion of the livestock sector, particularly pork and poultry meat in Asia. With these changes come an increased threat of emerging zoonotic diseases and a need for improved food safety and the implementation of appropriate biosecurity measures. Epidemiological efforts, livestock sector planning, and policy development all require knowledge of livestock distributions and abundance information that is often difficult to obtain in a consistent spatial format. For example, epidemiological modelling of highly pathogenic avian influenza (HPAI) type H5N1 (hereafter HPAI H5N1) in hot zones of re-emergence such as China is hampered by a lack of available data on spatial distributions of its main host, domestic poultry. HPAI H5N1 first emerged in 1996 in domestic geese of southeastern China (Xu et al., 1999). From 1997 to 2003, the virus continued to evolve and in early 2004, an extensive wave of outbreaks erupted across China and seven additional Asian countries (OIE, 2004). The virus showed varying degrees of pathogenicity and transmissibility among chickens, ducks, and geese, with ducks potentially serving as silent propagators of the virus (Li et al., 2004; Sturm-Ramirez

* Corresponding author at: USGS Patuxent Wildlife Research Center, Beltsville Lab, c/o BARC-East Building 308, 10300 Baltimore Ave., Beltsville, MD 20705, USA. Tel.: +1 301 497 5914; fax: +1 301 497 5624.
E-mail address: dprosser@usgs.gov (D.J. Prosser).

0167-8809/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.agee.2011.04.002

et al., 2005). Fourteen years later, HPAI H5N1 has spread from Asia to parts of Europe and Africa, and remains active in many regions, including China.

Since HPAI H5N1's first emergence in 1996, China has reported nearly 200 outbreaks in poultry and wild birds (primarily the former), and 39 cases in humans (OIE, 2010; World Health Organization, 2010). Strong government control efforts, including mass vaccination programs, a national active surveillance program, and culling of more than 35 million poultry, have led to a decrease in the number of outbreaks reported over the past year. The disease persists, however, with some human outbreaks occurring in regions without concurrent outbreak reports in poultry, raising questions as to whether underreporting of outbreaks or asymptomatic viral replication is occurring within the poultry population. High resolution distribution maps of individual poultry species would provide important input factors for disease risk modelling and vaccination strategies. To date, however, no such data have been available.

In 2007, the Food and Agriculture Organization of the United Nations (FAO) released the *Gridded Livestock of the World* (GLW): the first standardized, global, sub-national resolution population maps of livestock species, including poultry (FAO, 2007). An unprecedented accomplishment, these raster maps provide 3 arc-minute resolution livestock density estimates (approximately 5 km at the equator) based on disaggregation of agricultural census data (Robinson et al., 2007; Neumann et al., 2009). Until now, these were the only poultry distribution maps available that encompassed the whole of China. However, the temporal, spatial, and species resolutions available through GLW are not ideal for epidemiological modelling of HPAI H5N1 in China. The current version of the GLW uses poultry data from China in 1990s. Given that poultry production increased substantially from the 1990s to 2000s in China (http://kids.fao.org/glipha/), and HPAI H5N1 modelling efforts target this same timeframe, it is important to have distribution models based on updated poultry figures. In addition, the GLW dataset groups all poultry into one category. As chicken, duck and geese respond differently to HPAI H5N1 virus infection (Sturm-Ramirez et al., 2005), and their production systems have different spatial distributions, mapping poultry distributions at the species level is important for epidemiological modelling efforts.

In this study, we aimed to produce 1 km resolution population distribution maps for chickens, ducks, and geese across the extent of China. We hypothesized that strong statistical relationships exist between poultry populations and agro-ecological variables, which in turn could be used to spatially disaggregate census data. Building from previous work (FAO, 2007), we investigated quality of model output using remotely sensed predictors of meteorological data (Hay et al., 2006; Scharlemann et al., 2008) compared to ones that might offer more intuitive interpretation such as land cover variables. We also explored the effects of building predictive models within varying regional stratifications, and validated our data using a subset of the observed poultry data. Finally, in concert with related distribution modelling efforts for ducks across much of Monsoon Asia (see Van Boeckel et al., this issue), we compared the efficacy of using data solely from within China versus that from China and surrounding countries to determine whether the inclusion of outside data would improve model results.

The poultry distribution maps produced in this study are valuable for a variety of uses including epidemiological modelling, guiding policy decisions, livestock management, biosecurity and food safety, conflict resolution, and environmental impacts. We have made these data freely available through the FAO Gridded Livestock of the World (GLW) website: http://www.fao.org/AG/againfo/resources/en/glw/home.html.

2. Materials and methods

2.1. Poultry data

We aimed to obtain nationwide county level (administrative level 3) statistics for the 3 major types of poultry produced in China: chickens, ducks, and geese. Poultry statistics for China are published annually by the National Statistics Bureau (NSB) and the Ministry of Agriculture's Animal Husbandry Bureau (AHB). Both agencies report standard poultry statistics including: number of individuals sold per year (SOLD), number of individuals existing at the end of the calendar year (residual poultry; RESID), and meat and egg production by weight. Annual counts of each poultry type are collected from farms and households at the township level and are reported up through county, prefecture, and provincial administrative units with final submission to the national level. These data are publicly released as aggregated total poultry figures in provincial rural and statistical yearbooks (China National Bureau of Statistics, 2007). Differences between NSB and AHB statistics are attributed to the level of administrative unit for reporting and the type of poultry reported: NSB publishes aggregated estimates of total poultry (all species combined) at the county or prefecture level (levels 3 or 2, respectively) in provincial yearbooks; AHB publishes both aggregated (total poultry) and species level statistics (chickens, ducks, geese) at the coarser, provincial scale (level 1).

We extracted poultry census data from 96 rural and statistical yearbooks (printed in Chinese) for years 2003–2005 (reference list provided in Supplementary Table S1). Data were gathered for each of China's 22 provinces, 5 autonomous regions, and 4 municipalities (hereafter referred to as 31 provinces). We accessed yearbooks from the National Library of China in Beijing, the National Agricultural Information System of the Chinese Academy of Agricultural Sciences Agricultural Institute, the China National Knowledge Infrastructure (http://www.global.cnki.net/grid20/index.htm), and the United States Library of Congress in Washington, D.C.

Of the standard metrics reported, we used RESID poultry for the modelling process for 2 reasons: (a) RESID counts are conducted at the end of the calendar year at peak production prior to national Spring Festival holidays, and (b) RESID was the most comprehensive metric reported. In contrast to SOLD poultry, which comprised mainly poultry raised for meat consumption (broilers), RESID poultry provides a more complete representation of the poultry populations by including egg layers, meat poultry, and backyard poultry (poultry raised by households for personal consumption). As defined by the National Statistics Bureau, residual poultry is the number of poultry held in rural and urban areas at the end of the calendar year and includes "all size and breeds of poultry, etc. from rural cooperative economic organizations, state-operated farms, rural individuals, organizations, groups, schools, industrial/mining companies, government departments and units and raised by urban citizens" (China National Bureau of Statistics, 2007).

We employed a standardized protocol for filling gaps in available poultry statistics (See Fig. 1a and Section 3). In order of priority, 6 methods were used to create a complete set of poultry data for China: (1) county level RESID poultry; (2) prefecture level RESID poultry; (3) conversion of county level SOLD poultry to RESID poultry estimates; (4) conversion of prefecture level SOLD poultry to RESID poultry estimates; (5) provincial level RESID poultry; and (6) conversion of provincial level AHB RESID poultry to NSB RESID estimates (see Supplementary Fig. S1 for correlations between NSB and AHB provincial RESID poultry census data). We then divided total poultry figures into species estimates (chickens, ducks, and geese) using provincial species ratios from the 2006 Agricultural Census (China National Bureau of Statistics, 2008) which have not yet been released to the public. Poultry census estimates were

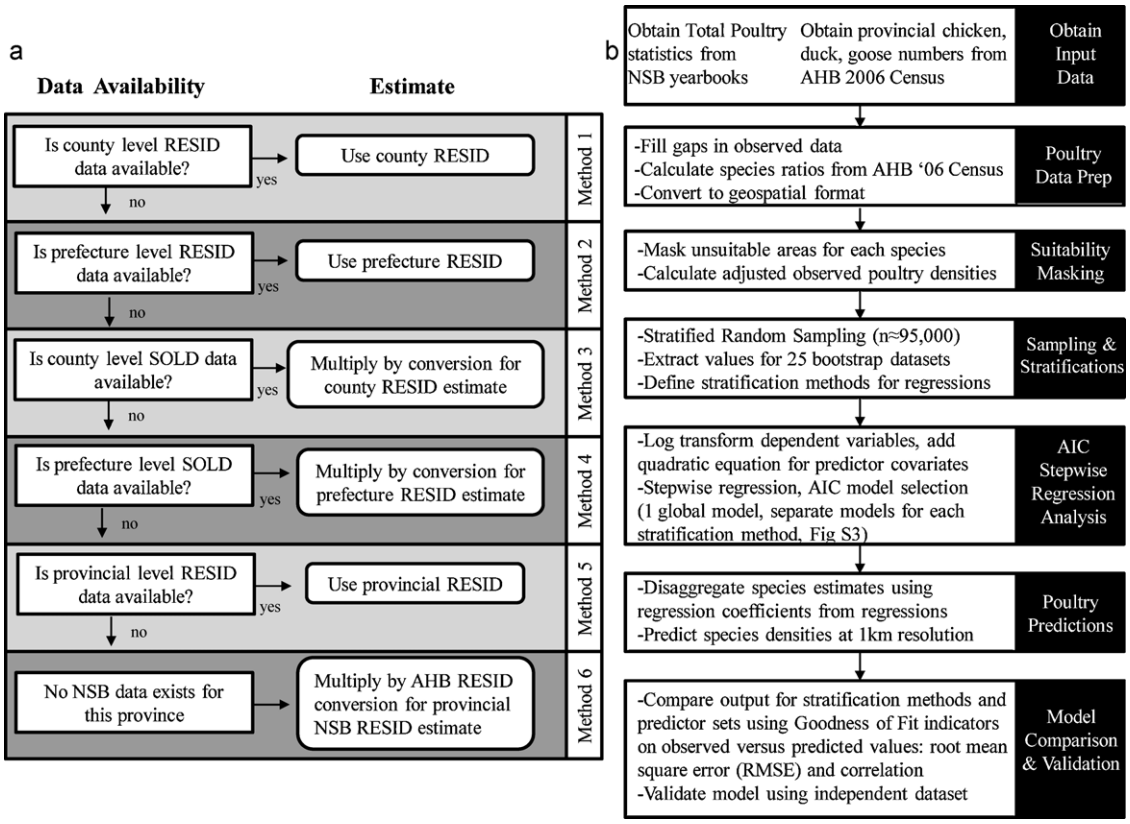

Fig. 1. (a) Methods used for filling data gaps in total poultry across China, (b) methodology for modelling chicken, duck, and goose distributions for China. RESID = residual poultry at end of year, SOLD = number poultry sold, NSB = National Statistics Bureau, AHB = Animal Husbandry Bureau (see Supplemental Fig. S1 for NSB and AHB relationships).

converted to geospatial format using ArcGIS 9.3 (Environmental Systems Research Institute, Inc., Redlands, CA, USA).

2.2. The modelling process

We modeled distributions of domestic chickens, ducks, and geese in China using the following steps modified from the GLW processing chain (FAO, 2007) (Fig. 1b): (1) obtain poultry census data; (2) fill data gaps, develop species level estimates, and convert to geospatial format; (3) mask unsuitable areas and calculate adjusted observed densities for each poultry species; (4) extract dependent (poultry) and independent (predictor) training and validation data using a stratified random sampling scheme; (5) establish statistical relationships between dependent poultry estimates and predictor covariates; (6) create predicted poultry distribution maps using equations from statistical relationships; and (7) assess model goodness of fit using sample points omitted from the training set.

After preparing the poultry census data for input into the modelling process, we calculated observed poultry densities for each administrative unit by correcting for the area of land unsuitable for poultry production. Suitability masks for chickens, ducks, and geese were modified from original GLW monogastric livestock (pigs and poultry) masks (FAO, 2007). Our suitability masks were restricted to exclude only the most environmentally unsuitable areas for production (e.g., extreme high elevations, tundra, ice, etc.; Table S2) but did not exclude heavily populated locations as certain phases of poultry production may occur in urban areas, such as chick hatcheries located within city limits.

We created a stratified random sampling frame that included one point per polygon (reporting administrative unit) and an average of 20 points per decimal degree across the extent of China. Sample points were bootstrapped to create 25 data sets to be used in assessing model variation. At each sample point, poultry estimates and predictor covariates were extracted. Seventy five percent of the points were used for training models and 25% were reserved for model validation.

We used an information theoretic approach to choose best models at iterative steps in a multivariate regression procedure (Burnham and Anderson, 2002; Whittingham et al., 2006). Dependent variables were log transformed for normality, and each independent variable was paired with its quadratic term to accommodate curvilinear relationships (Rawlings et al., 1998). The stepwise procedure began with a null model followed by inclusion of the predictor pair defined by the best Akaike Information Criterion (AIC). The process was successively repeated for each remaining pair of predictors until one of 2 conditions was met: (i) improvement in AIC score for 2 successive models was less than 1%, or (ii) a threshold minimum number of unique data values was not available for each predictor pair entered in the model (i.e., 15 data points per variable pair). Coefficients from the top regression models were then applied to the predictor imagery to create predicted maps of distributions for each species. Means and coefficients of variation (standard deviation divided by mean) were estimated from the 25 bootstrapped predictions. Two goodness of fit indicators were used to assess quality of model output: root mean square error (RMSE) and correlations (COR) between predicted and observed values. Lower RMSE and higher COR indicated better fits. Correction by country totals was applied to the final maps.

Environmental and demographic conditions relevant to poultry production vary widely across the extent of China. We therefore performed regression models within stratification zones chosen to reflect regional differences in association with poultry production. Model predictions for four stratification schemes were compared: (i) global livestock production systems (LPS), (ii) data driven ecozones (EZ) using unsupervised classification of Moder-

Table 1
Predictor variables used in China poultry distribution modelling. Three groups were compared: (1) Gridded Livestock of the World predictors (GLW; FAO, 2007), (2) a set of land use and anthropogenic predictors (LU), and (3) the GLW and LU predictors combined (GLW + LU).

GLW predictors	
MODIS channels	TFA processed channels 03, 07, 08, 14, 15, 35: mx, mn, d1, d2, d3, da, a1, a2, a3, p1, p2, p3, produced by SEEG, University of Oxford
1kgrumpdens	Alpha version kilometer resolution human population density for 2000 from GPW GRUMP, at Columbia University
1kgrumpdensb	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
1kaglgprc	Length of growing period, derived from FAO LGP layers using statistical modelling by ERGO
1kthlgprc	Length of growing period, derived from LGP layers produced by Thornton, using statistical modelling 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
1krdsdeg	Distance in decimal degrees to major roads – using landscan road layer, derived by ERGO
1kwatdeg	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
LU predictors	
Land cover	Forest, grassland, open water, vegetated wetland, rice paddy, cropland, developed, urban
Cropping intensity	Yan et al. (2010) and Yan et al. (in review)
Human population	Tian 2005
Elevation	Shuttle radar topography mission
Slope	GTOPO30

GPW GRUMP = gridded population of the world global rural urban mapping project; ERGO = environmental research group Oxford; SEEG = spatial epidemiology and ecology group.

ate Resolution Imaging Spectroradiometer (MODIS) remote sensing variables and Shuttle Radar Topography Mission (SRTM) digital elevation models, (iii) China Agro-ecological Regions (CAR), and (iv) a combination of the first three (All.BestRSE). The LPS regions, updated from those initially developed by Sere and Steinfeld (1996) and mapped by Thornton et al. (2002), represent 14 classes of livestock production based on grassland, mixed farming, and landless systems. The EZ regions consist of 4 hierarchical levels of clustering for Asia: EZ5, EZ12, EZ25, and EZ50 which represent 5, 12, 25, and 50 cluster classes using MODIS channels 3, 7, 8, 14, and 15, and SRTM data (see Van Boeckel et al., this issue for details). For the EZ stratifications, we built prediction maps at the pixel level, using regression coefficients of the EZ with the lowest residual squared error (hereafter referred to as EZ.BestRSE stratification). The CAR stratification, adapted from Verburg and Chen (2000), is a modification of the commonly used China agricultural regionalization by Crook (1993). CAR divides China into 8 regions based on agriculture, economics, environment, and provincial level administrative boundaries. Modifications from Crook (1993) consisted of removing the densely populated Sichuan province from sparsely populated Tibetan Plateau and including it with Yunnan and Guizhou provinces. The final stratification, All.BestRSE, chooses, pixel by pixel, the stratification with the lowest residual squared error from the stratifications described above. Examples of All.BestRSE, EZ.BestRSE, CAR, and LPS stratifications are displayed in Supplementary Fig. S3. We set model conditions to perform regressions within each stratification zone, however, if criteria of a minimum of 15 unique dependent estimates per variable pair were not met, coefficients from a single country level model were then used to create predictions within that zone.

GLW distribution models have traditionally been created using anthropogenic variables such as human density, distance to roads, etc., in combination with remotely sensed surrogates of meteorological data (e.g., middle infrared, land surface temperature, etc.) as predictors. We were interested in comparing capabilities of a predictor set using the GLW approach versus one that includes interpreted remote sensing variables such as land use (e.g., cropland, wetland, grassland, etc.). The incentive for using the latter group is the potential to draw more intuitive conclusions between significant predictor variables and poultry predictions. Thus, we ran models using 3 predictor datasets: GLW, LU, and the combined set GLW + LU (Table 1). The main difference between the GLW and LU sets was the inclusion of Fourier transformed MODIS data for GLW (Scharlemann et al., 2008) (see Van Boeckel et al., this issue for details) and land use variables for LU (Liu et al., 2002).

Goodness of fit indicators, RMSE and COR, were compared in an analysis of variance (ANOVA) to determine optimal predictor sets and regional stratification schemes. Data were reviewed for conformity to the assumptions of normality and homogeneity of variance. Histograms of RMSE and COR appeared normal for each of the predictor datasets and stratifications. Since sample sizes between levels were identical in the one-way ANOVA, we assumed the overall F test and multiple comparison tests were robust to departures from the unequal variance assumption (Neter et al., 1996).

Finally, to assess the value of including poultry and agro-ecological relationships from countries surrounding China, we compared goodness of fit scores for China versus those from a related study that models duck distributions across Monsoon Asia (see Van Boeckel et al., this issue). The modelling methodology in Van Boeckel et al. (this issue) is similar to that used in this study (although overall proposed hypotheses differ) and includes data from 14 countries: China, Cambodia, Bhutan, Thailand, Lao, Vietnam, Myanmar, Bangladesh, India, Nepal, Korea, Malaysia, Philippines, and Indonesia.

3. Results

We targeted NSB data for model development because of the finer scale at which they are reported (mainly county and prefecture versus provincial level for AHB). Of 3 years of data investigated, year 2004 was most complete (86% complete versus 82 and 78% for years 2003 and 2005, respectively), and thus was used for model input. We implemented a multi-level methodology for creating complete RESID estimates from the data available (Fig. 1a). We

Table 2
Data availability and method description for deriving 2004 residual poultry statistics for each of 31 provinces of China. Provinces denoted with asterisk indicate use of Ministry of Agriculture Animal Husbandry data (AHB); all others derived from National Statistics Bureau data (NSB).

Method	Data availability	Method description	Applicable provinces
1	RESID county level data	Use county RESID	Beijing, Jiangsu, Zhejiang, Anhui, Fujian, Henan, Hunan, Guangdong, Ningxia
2	RESID prefecture level data	Use prefecture RESID	Hebei, Heilongjiang, Jiangxi, Shandong, Shaanxi
3	SOLD county level data	Multiply by conversion for county RESID estimate	Tianjin, Hubei, Chongqing*
4	SOLD prefecture level data	Multiply by conversion for prefecture RESID estimate	Inner Mongolia, Shanghai*, Hainan*, Sichuan*, Qinghai*
5	RESID provincial level data	Use provincial RESID	Shanxi, Gansu
6	No NSB data at any level	Use AHB RESID data (provincial scale)	Liaoning, Jilin, Guangxi, Guizhou, Yunnan, Xizang, Xinjiang

applied methods 1–4 to approximately ¾ of the provinces (22 of 31) that had county and prefecture level data (Table 2). The remaining nine provinces had provincial level data; here we applied methods 5 and 6. Method 6 uses AHB data for those provinces lacking NSB data (based on high correlation between the 2 data sets: *r*-square value of 99.4%, see Supplemental Fig. S1).

Observed densities (census data), model predictions, and coefficient of variation are shown in Fig. 2a, b, and c, respectively. Observed densities were highest for chickens, and considerably lower for ducks and geese (111.2, 27.4, and 6.7 thousand per km^2 maximum, respectively). Geographically, maximum densities were higher in southern and eastern China than the remote northern and western regions (northern and western regions defined as CAR zones 5 and 6, see Supplemental S3c). Duck densities in particular were highest in southeastern China where lowland tropics and rice agriculture is prevalent. Chickens were most ubiquitous, with high densities across most of southern and eastern China, and moderate to low densities across remote regions of the north and west. Model uncertainty (COV) tended to be highest in the remote western regions of China where poultry numbers are lower.

Goodness of fit measures indicate that of the 3 predictor data sets, GLW + LU performed best (Fig. 3): one-way ANOVAs for RMSE and COR between predicted and observed values were both $P < 0.001$, and Tukey's pairwise comparisons were all $P < 0.005$. Goodness of fit measures for stratification methods were less distinct. We compared RMSE and correlation coefficients for each species, using the best predictor dataset only (GLW + LU). Of the 6 ANOVAs (RMSE and COR each for chickens, ducks, and geese) all but one (COR for ducks) were significant at $P < 0.05$, however, Tukey's pairwise comparisons did not indicate a single best stratification method for any of the species (Fig. 4). LPS and All.BestRSE tended to score better for chickens; LPS, All.BestRSE, and CAR for ducks; CAR and All.BestRSE for geese, however, we found that all stratifications chosen for analysis performed significantly better than the country model (Fig. S5): one-way ANOVA and Tukey's pairwise comparisons were all $P < 0.001$. Since each stratification method performed significantly better than the global model and without clear statistical difference among stratifications, we chose the stratification with the best mean goodness of fit scores for each species (see Fig. 4) to present our final output (Fig. 2b), which was LPS for chickens and ducks, and CAR for geese.

Predictor variables Elevation, Precipitation, and Evapotranspiration were consistently ranked among the top 5 predictors for each species (Table 3) based on mean Delta AIC score (the amount by which the AIC score of the best model was increased after removing the predictor). Other top predictors included Area Suitable for Monogastrics, Nighttime Land Surface Temperature, Enhanced Vegetation Index, Daytime Land Surface Temperature, and Middle Infra-red readings. The predicted poultry densities were generally positively associated with Precipitation, Evapotranspiration, Daytime Land Surface Temperature, Middle Infra-red, and Area Suitable for Monogastrics; they were generally negatively associated with Elevation, Nighttime Land Surface Temperature, and Enhanced Vegetation Index. The majority of predictors included in top ranked models by AIC were from the GLW set, however, important LU predictors included rice paddy for ducks and geese; and Elevation, open water, developed land, and cropland area for all three species.

We compared the effects of including training data from countries surrounding China (Cambodia, Bhutan, Thailand, Lao, Vietnam, Myanmar, Bangladesh, India, Nepal, Korea, Malaysia, Philippines, and Indonesia) versus restricting the analysis to using training data from within China. Goodness of fit indicators (RMSE and COR) were better for analyses restricted to China (Fig. 5) suggesting that the relationship between predictor variables and observed poultry densities within China are different from those of surrounding countries.

4. Discussion

The results of this work indicate that agro-environmental variables can be used to predict spatial poultry distributions in China. The process predicted density patterns that are consistent with known distribution patterns, for example high chicken densities across much of eastern China, particularly the Yellow River Basin and high duck densities in southeastern China and the Sichuan Basin. Geese were least abundant, but exhibited consistent patterns, with highest densities in Sichuan and parts of Guangdong. Validation measures between observed and predicted values indicated good fits based on RMSE and correlations. In comparison to goodness of fit values reported in the related Van Boeckel et al. (this issue) paper on duck distribution modelling in Monsoon Asia, goodness of fit scores for ducks within China ranked better than those produced for most other countries.

We observed statistically significant differences in goodness of fit scores among predictor data sets but not among regional stratifications. Each of the regional stratification methods we compared provided better goodness of fit scores than the country-wide model. However, because a clear best stratification scheme was not statistically evident, we chose the one with the best mean score for each species. This was the Livestock Production Systems (LPS) approach for chickens and ducks, and China agro-ecological (CAR) approach for geese. The combined approach (All.BestRSE) produced the second best mean scores across all species. Van Boeckel et al. (this issue) found similar results for their Monsoon Asia duck models with LPS and All.BestRSE showing highest fitness scores. The predicted density maps produced by models in this study and the Monsoon Asia study (Fig. 2b here and Fig. 4 in Van Boeckel et al., this issue) revealed similar output patterns. Here we conclude that for the China models, either stratification would be appropriate for use, however an advantage of LPS (and CAR, for geese) over the combined approach (All.BestRSE) is the more intuitive interpretation of a single stratification versus the combination of many.

Overall, uncertainty measures were low for each species (COV values ranged from <0.01 to 5). Areas with the highest uncertainties were located in northwestern China where poultry populations are scarce and environmental predictors are variable. In eastern and southern China, where poultry populations are high, uncertainty

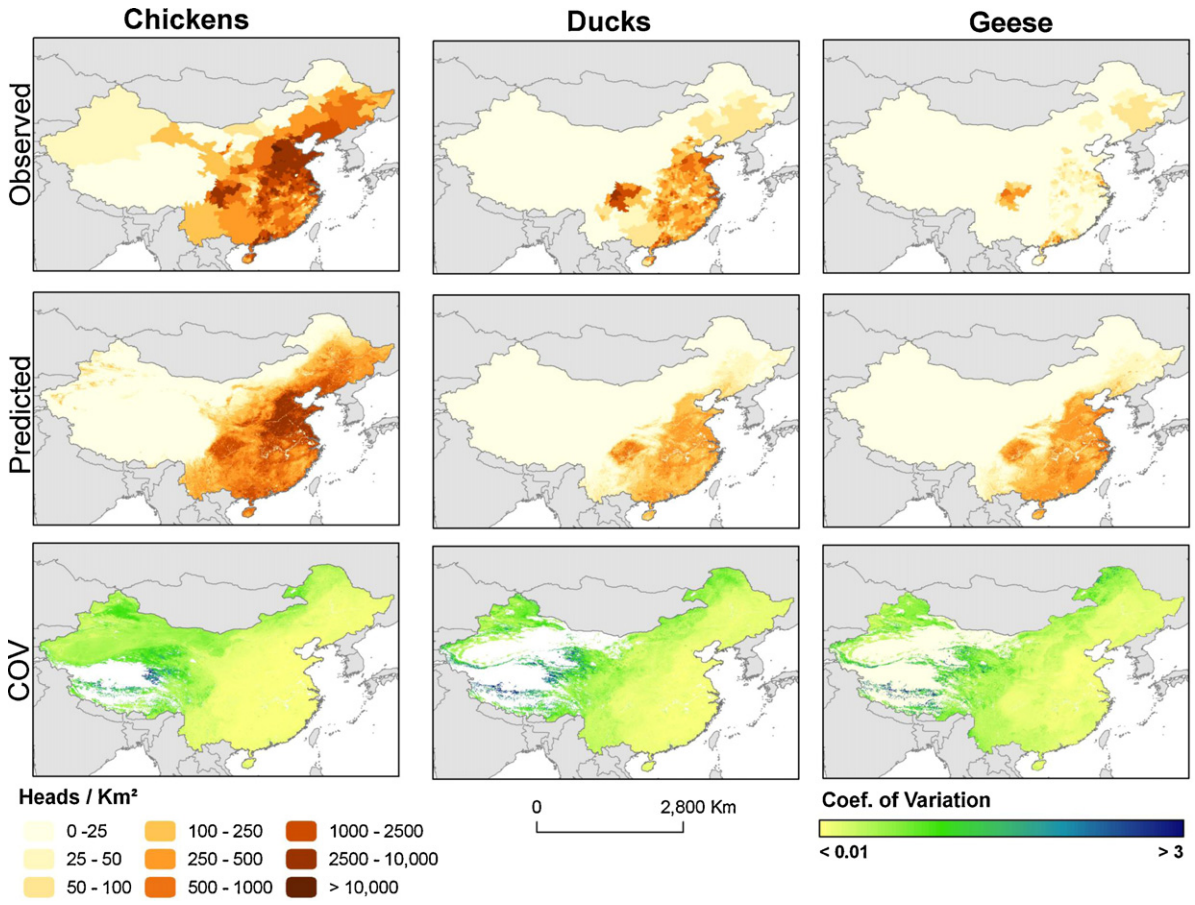

Fig. 2. (a) Observed densities, (b) model predictions, and (c) coefficient of variation, for chickens, ducks, and geese across China. Mean densities and coefficient of variation (standard deviation divided by mean) represent 25 bootstrapped models. Model output shown for the GLW + LU predictors and LPS (chickens, ducks) or CAR stratification (geese) method (defined by goodness of fit scores).

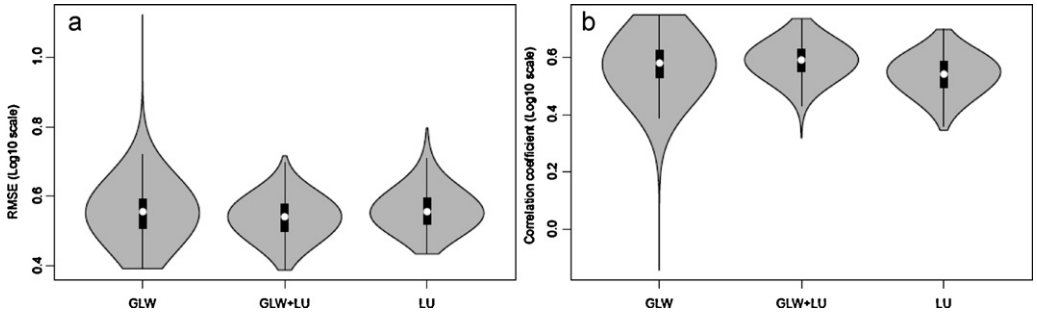

Fig. 3. Violin plots of (a) root mean square error (RMSE) and (b) correlation coefficient between predicted and observed chicken, duck, and goose densities (log transformed) for 3 predictor datasets: GLW (traditional Gridded Livestock of the World predictors), LU (land use and anthropogenic predictors), and GLW + LU (combination of GLW and LU predictors). ANOVA main effects ($P < 0.001$) and Tukey's pairwise comparisons (all $P < 0.005$) indicate significant differences among all 3 predictor sets with GLW + LU having lowest mean RMSE and highest mean Correlation between observed and predicted values.

Table 3
Top 5 predictor variables for chicken, duck, and goose distribution modelling regressions. Predictors are listed in decreasing order of mean Delta AIC (amount AIC score was increased after removing variable from the best model). A1 = amplitude of annual cycle, DA = combined variance in annual, bi-annual, and tri-annual cycles, D1 = variance in annual cycle (see Scharlemann et al., 2008).

Chicken	Ducks	Geese
Annual precipitation	Elevation	Elevation
Area suitable for Monogastrics	Annual precipitation	Annual precipitation
Elevation	Evapotranspiration (DA)	Daytime Land Surface Temperature (D1)
Evapotranspiration (A1)	EVI (mean)	Middle Infra-red (mean)
Nighttime Land Surface Temperature (max)	EVI (max)	Evapotranspiration (D1)

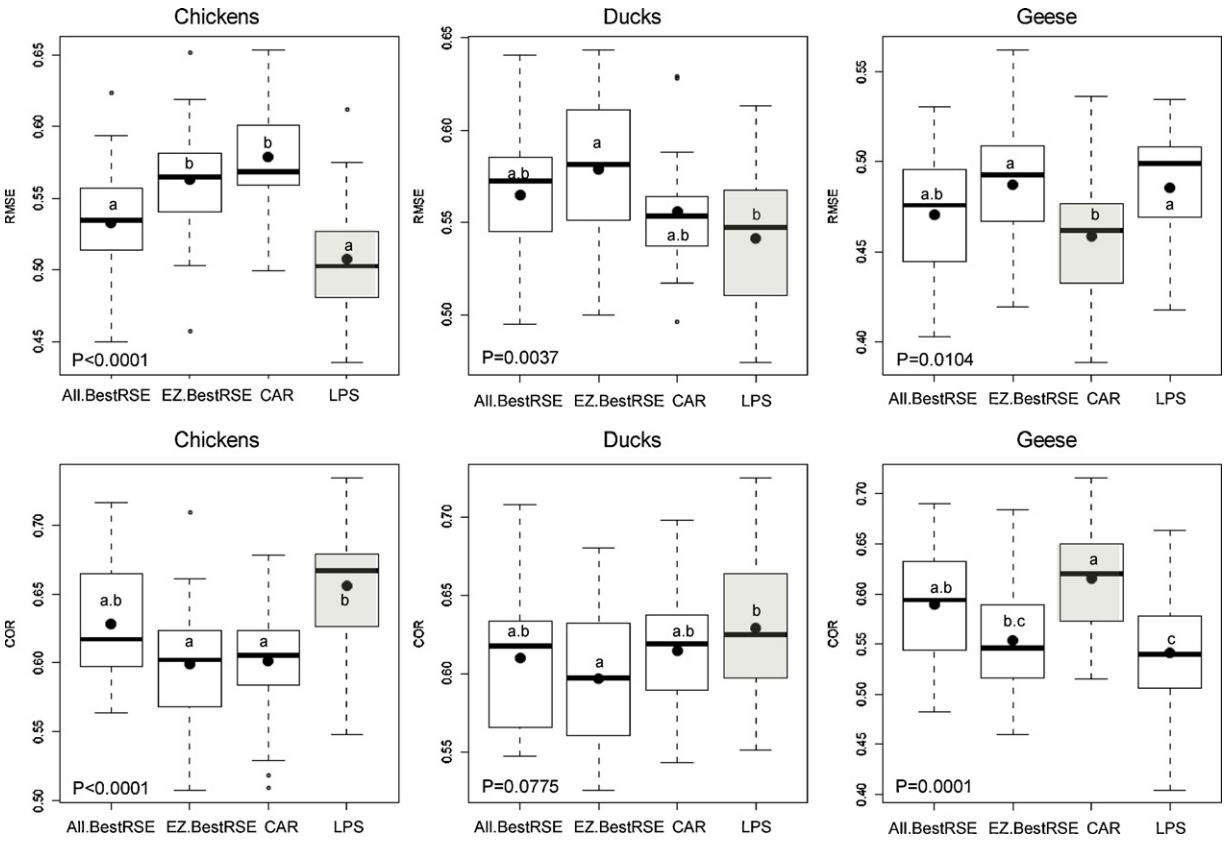

Fig. 4. Boxplots of root mean square error (RMSE) and correlation coefficient between predicted and observed chicken, duck, and goose densities (log transformed) for 4 stratification schemes: All.BestRSE (uses prediction from stratification (BestEZ, CAR, or LPS) with the best goodness of fit score on a pixel by pixel basis), EZ.BestRSE (uses prediction from data driven classifications (EZ5, EZ12, EZ25, EZ50) with best goodness of fit score on a pixel by pixel basis), CAR (China Agro-ecological Regions), and LPS (global livestock production systems). Main effects ANOVA significance values in lower left of each panel; means represented by black circles; Tukey's pairwise comparisons ($P < 0.05$) denoted by letters; grey boxplots represent statification with best mean GOF, LPS for chickens and ducks and CAR for geese. Although strong differences among stratifications were not evident, all stratifications examined performed better than the global model (i.e., no stratification; $P < 0.001$ see Fig. S4).

estimates were low (ranging from <0.01 to 0.08), indicating small standard deviations in relation to mean predicted densities. In general, uncertainty patterns across China were similar among species, and on average, COVs were lowest for chickens, then ducks, and geese.

The use of data external to China for training models produced inferior goodness of fit scores compared to those from models using training data entirely from within China. This exemplifies the fact that relationships between the predictor variables and poultry distributions differ for China in comparison to neighboring countries. The 13 countries included in this analysis were predominantly located to the south of China. These countries show greater similarity to China's tropical southeastern provinces than to the high-elevation drier provinces in western China and mixed grasslands of north central China, which could account for part of the differences in goodness of fit scores. In addition, China's poultry production system far exceeds those of its neighboring countries, ranking first in egg production and second in meat production (Qing, 2002; Wang, 2006) on a global scale. For example, in 2004, China's poultry production was more than an order of magnitude higher than those reported by its surrounding countries except Indonesia (5.1 billion versus 1.2 billion for China and Indonesia, respectively). Remaining countries ranged from 500 million (India) to 230 thousand (Bhutan); from UNFAO's Global Livestock Pro-

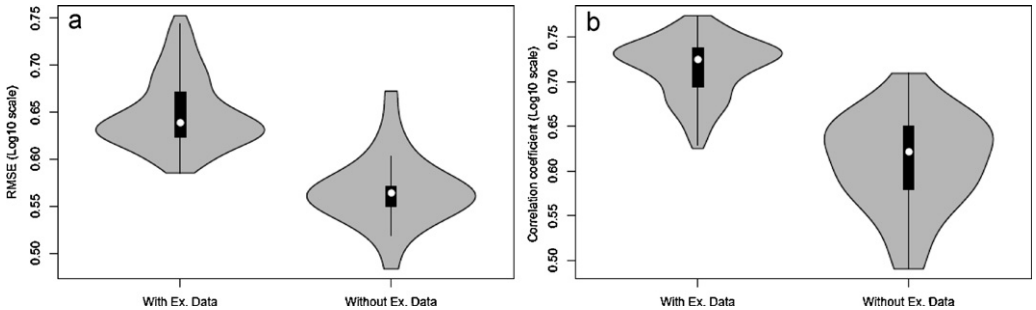

Fig. 5. (a) Root mean square error (RMSE) and (b) correlation coefficients for ducks (log densities) comparing predictions with and without data from surrounding countries. Data are presented as violin plots, a combination of box and kernel density plots (see Hintze and Nelson, 1998). Higher RMSE and lower correlation coefficients for analyses using data from surrounding countries suggest relationships between poultry densities and predictor variables within China are different from surrounding countries and such additional analyses do not improve predictions within China.

duction and Health Atlas (http://kids.fao.org/glipha/). Given the observed differences in goodness of fit scores, we do not recommend using external training data to create model predictions for China, nor should results from China be directly extrapolated to other regions in Asia.

The data fill methodology employed in this study (Fig. 1a) provides a consistent and repeatable method for assembling poultry statistics from multiple sources representing the diverse and expansive regions across China. Despite national efforts to report agricultural statistics in annual yearbooks for each province, the administrative level of reporting varies across regions, ranging from provincial to county level (administrative levels 1–3). Fig. S2 shows the spatial heterogeneity of input data used for our China models, the finest scale data being located in the poultry-rich regions of southeastern China. These differences are reflected in the uncertainty values (Fig. 2c) with higher COVs in regions in the western and northern regions of China. To accommodate the spatial heterogeneity of input data, we chose to use a mixed random and stratified sampling design that includes a minimum of one point per administrative unit as well as an average density across the country (20 points per decimal degree). Model predictions would likely be improved with finer scale input data for the remote regions of China, however, for the target time frame of our models, we have assembled the best data available to produce distribution predictions which have been qualified with estimates of uncertainty.

5. Conclusions

Our goal was to produce 1 km resolution population distribution maps each for chickens, ducks, and geese in China for use in HPAI H5N1 epidemiologic modelling. This research indicates that spatial distributions for these species can be modeled using agro-ecological predictors in a regression and disaggregation approach.

We found that a combination of traditional predictors (FAO Gridded Livestock of the World) and land use predictors produced output with the best goodness of fit scores between observed and predicted values. We also learned that of four stratification schemes used to build regression models within different regions of China, the livestock production systems (LPS), China Agro-ecological Regions (CAR), and combined approach (All.BestRSE) produced the best goodness of fit scores.

Obtaining observed population data across China for model training was challenging due to availability of data, however, using a multi-step approach to systematically incorporate the best data available for each region, we produced a complete and repeatable training set for model development. Should other datasets eventually be released to the public, the modelling process developed above can be used to create updated predictive spatial distribution maps for China.

Our poultry distribution models have been made available to the scientific and public community through the FAO Gridded Livestock of the World database for use in a multitude of applications from disease risk modelling to livestock and environmental management.

Acknowledgments

This work was funded in part by the USGS Wildlife Program and the U.S. National Institutes of Health (R01-TW007869) through the NIH/NSF Ecology of Infectious Disease program. Early exploration of poultry farming systems in China was supported by NSF East Asia Summer Institutes Program grant (OISE-0513222). Dr. Xiao is also supported by the Chinese Special Program for Prevention and Control of Infectious Diseases (No. 2008ZH1004-012) from the China Ministry of Health and China Ministry of Science and Technology. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. The authors would like to thank Dr. Huimin Yan of the Chinese Academy of Sciences, Institute of Geographical Sciences and Natural Resources Research for use of cropping intensity data for China. We thank R. Michael Erwin, Shane Heath, and anonymous reviewers for useful comments on earlier versions of this manuscript. The use of trade, product, or firm names in this publication is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.agee.2011.04.002.

References

Burnham, K.P., Anderson, D.R., 2002. Information and Likelihood Theory: A Basis for Model Selection and Inference. Model Selection and Multimodel Inference: A Practical Approach. Springer-Verlag, New York, NY, pp. 49–79.
China National Bureau of Statistics, 2007. Rural Statistical Yearbook. China Statistics Press, Beijing.
China National Bureau of Statistics, 2008. China 2006 National Agricultural Census Data Synthesis Summary. China Statistics Press, Beijing.
Crook, F.W., 1993. In: US Department of Agriculture, E.R.S. (Ed.), China's Grain Production Economy: A Review by Regions. China, International Agriculture and Trade Reports, Washington, DC.
FAO, 2007. In: Wint, G.R.W., Robinson, T.P. (Eds.), Gridded Livestock of the World 2007. , Rome, p. 131.
Hay, S.I., Tatem, A.J., Graham, A.J., Goetz, S.J., Rogers, D.J., 2006. Global environmental data for mapping infectious disease distribution. Adv. Parasitol. 62, 37–77.
Hintze, J.L., Nelson, R.D., 1998. Violin plots: A box plot-density trace synergism. Am. Stat. 52, 181–184.
Li, K.S., Guan, Y., Wang, J., Smith, G.J., Xu, K.M., Duan, L., Rahardjo, A.P., Puthavathana, P., Buranathai, C., Nguyen, T.D., Estoepangestie, A.T., Chaisingh, A., Auewarakul, P., Long, H.T., Hanh, N.T., Webby, R.J., Poon, L.L., Chen, H., Shortridge, K.F., Yuen, K.Y., Webster, R.G., Peiris, J.S., 2004. Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. Nature 430, 209–213.
Liu, J., Liu, M., Deng, X., Zhuang, D., Zhang, Z., Luo, D., 2002. The land use and land cover change database and its relative studies in China. J. Geogr. Sci. 12, 275–282.
Neter, J., Kutner, M.H., Nachtsheim, C.J., Wasserman, W., 1996. Applied Linear Statistical Models. McGraw-Hill, New York.
Neumann, K., Elbersen, B., Verburg, P., Staritsky, I., Pérez-Soba, M., de Vries, W., Rienks, W., 2009. Modelling the spatial distribution of livestock in Europe. Landscape Ecol. 24, 1207–1222.
OIE, 2004. Update on Highly Pathogenic Avian Influenza in Animals: Type H5 and H7., http://www.oie.int/downld/AVIAN%20INFLUENZA/A_AI-Asia.htm.
OIE, 2010. Update on Highly Pathogenic Avian Influenza in Animals: Type H5 and H7., http://www.oie.int/downld/AVIAN%20INFLUENZA/A_AI-Asia.htm.
Qing, F.S., 2002. China's Poultry Industry Affected by WTO. World Poultry, www.worldpoultry.net, www.worldpoultry.net, pp. 12–14.
Rawlings, J.O., Pantula, S.G., Dickey, D.A., 1998. Applied Regression Analysis: A Research Tool. Springer-Verlag, New York.
Robinson, T.P., Franceschini, G., Wint, W., 2007. The food and agriculture organization's gridded livestock of the world. Vet. Ital. 43, 745–751.
Scharlemann, J.P.W., Benz, D., Hay, S.I., Purse, B.V., Tatem, A.J., Wint, G.R.W., Rogers, D.J., 2008. Global data for ecology and epidemiology: a novel algorithm for temporal fourier processing MODIS data. PLoS ONE 3, e1408.
Sere, C., Steinfeld, H., 1996. World livestock production systems: current status, issues and trends. FAO Animal Production and Health Paper 127.
Sturm-Ramirez, K.M., Hulse-Post, D.J., Govorkova, E.A., Humberd, J., Seiler, P., Puthavathana, P., Buranathai, C., Nguyen, T.D., Chaisingh, A., Long, H.T., Naipospos, T.S., Chen, H., Ellis, T.M., Guan, Y., Peiris, J.S., Webster, R.G., 2005. Are ducks contributing to the endemicity of highly pathogenic H5N1 influenza virus in Asia? J. Virol. 79, 11269–11279.
Thornton, P.K., Kruska, R.L., Henninger, N., Kristjanson, P.M., Reid, R.S., Atieno, F., Odero, A., Ndeqwa, T., 2002. Mapping Poverty and Livestock in the Developing World. International Livestock Research Institute, Nairobi.
Van Boeckel, T.P., Prosser, D.J., Wint, G.R.W., Robinson, T., Gilbert, M. Modelling the distribution of domestic ducks in Monsoon Asia. Agric. Ecosyst. Environ., this issue.
Verburg, P.H., Chen, Y.Q., 2000. Multiscale characterization of land-use patterns in China. Ecosystems 3, 369–385.
Wang, H., 2006. The Chinese Poultry Industry at a Glance. World Poultry, www.worldpoultry.net, pp. 10–11.
Whittingham, M.J., Stephens, P.A., Bradbury, R.B., Freckleton, R.P., 2006. Why do we still use stepwise modeling in ecology and behaviour? J. Anim. Ecol. 75, 1182–1189.

World Health Organization, 2010. Cumulative Number of Confirmed Human Cases of Avian Influenza A/(H5N1) Reported to WHO as of 3 August 2010., http://www.who.int/csr/disease/avian_influenza/country/en/index.html.

Xu, X.Y., Subbarao, K., Cox, N.J., Guo, Y.J., 1999. Genetic characterization of the pathogenic influenza A/Goose/Guangdong/1/96 (H5N1) virus: similarity of its hemagglutinin gene to those of H5N1 viruses from the 1997 outbreaks in Hong Kong. Virology 261, 15–19.

Yan, H., Xiao, X., Huang, H.Q., Liu, J.Y., Wang, J.B., in review. Mapping multiple cropping index in China from agro-meteorological observations and time series MODIS imagery.

Yan, H., Xiao, X., Huang, H., 2010. Satellite observed crop calendar and its spatial-temporal characteristics in multiple cropping area of Huang-Huai-Hai Plain, China. Acta Ecologica Sinica 30 (9), 2416–2423, in Chinese with English abstract.