Role of Urbanization, Land-Use Diversity, and Livestock Intensification in Zoonotic Emerging Infectious Diseases

Sumeet Saksena, Jefferson Fox, Michael Epprecht, Chinh Tran, Miguel Castrence, Duong Nong, James Spencer, Nguyen Lam, Melissa Finucane, Tran Duc Vien, and Bruce Wilcox

EAST-WEST CENTER
COLLABORATION · EXPERTISE · LEADERSHIP

WORKING PAPERS

Role of Urbanization, Land-Use Diversity, and Livestock Intensification in Zoonotic Emerging Infectious Diseases

Sumeet Saksena, Jefferson Fox, Michael Epprecht, Chinh Tran, Miguel Castrence, Duong Nong, James Spencer, Nguyen Lam, Melissa Finucane, Tran Duc Vien, and Bruce Wilcox

Sumeet Saksena, Jefferson Fox, and Melissa Finucane are Fellows in the Environment, Population, and Health group of the Research Program at the East-West Center in Honolulu, Hawaiʻi.

Michael Epprecht is a Research Scientist at the Center for Development and Environment, University of Bern, Bern.

Chinh Tran is a PhD candidate at the Department of Natural Resources and Environmental Management at the University of Hawaiʻi at Mānoa.

Miguel Castrence is a Remote Sensing and GIS scientist and currently working for Resource Mapping LLC based in Honolulu, Hawaiʻi.

Duong Nong is a PhD candidate at the Department of Natural Resources and Environmental Management at the University of Hawaiʻi at Mānoa.

James Spencer is Professor of City & Regional Planning, and Chair of Department of Planning, Development, and Preservation, Clemson University.

Nguyen Lam is a lecturer at the Hanoi University of Agriculture, Hanoi, Vietnam.

Tran Duc Vien is Chancellor of the Hanoi University of Agriculture, Hanoi, Vietnam.

Bruce Wilcox is with the Department of Pathology, Khon Kaen University, Khon Kaen.

The East-West Center promotes better relations and understanding among the people and nations of the United States, Asia, and the Pacific through cooperative study, research, and dialogue. Established by the US Congress in 1960, the Center serves as a resource for information and analysis on critical issues of common concern, bringing people together to exchange views, build expertise, and develop policy options.

The Center's 21-acre Honolulu campus, adjacent to the University of Hawaiʻi at Mānoa, is located midway between Asia and the US mainland and features research, residential, and international conference facilities. The Center's Washington, DC, office focuses on preparing the United States for an era of growing Asia Pacific prominence.

The Center is an independent, public, nonprofit organization with funding from the US government, and additional support provided by private agencies, individuals, foundations, corporations, and governments in the region.

EAST-WEST CENTER
COLLABORATION • EXPERTISE • LEADERSHIP

Role of Urbanization, Land-Use Diversity, and Livestock Intensification in Zoonotic Emerging Infectious Diseases: The Case of Highly Pathogenic Avian Influenza (H5N1) in Viet Nam

Saksena, Fox, Epprecht, Chinh, Castrence, Duong, Spencer, Lam, Finucane, Vien, Wilcox

Abstract
Emerging infectious diseases (EIDs) continue to significantly threaten human and animal health. While there has been some progress in identifying underlying proximal driving forces and causal mechanisms of disease emergence, the role of distal factors is most poorly understood. This article focuses on analyzing the statistical association between highly pathogenic avian influenza (HPAI) H5N1 and urbanization, land-use diversity and poultry intensification. A special form of the urban transition—peri-urbanization—was hypothesized as being associated with 'hot-spots' of disease emergence. Novel metrics were used to characterize these distal risk factors. Our models, which combined these newly proposed risk factors with previously known natural and human risk factors, had a far higher predictive performance compared to published models for the first two epidemiological waves in Viet Nam. We found that when relevant risk factors are taken into account, urbanization is generally not a significant independent risk factor. However, urbanization spatially combines other risk factors leading to peri-urban places being the most likely 'hot-spots'. The work highlights that peri-urban areas have highest levels of chicken density, duck and geese flock size diversity, fraction of land under rice, fraction of land under aquaculture compared to rural and urban areas. Land-use diversity, which has previously never been studied in the context of HPAI H5N1, was found to be a significant risk factor. Places where intensive and extensive forms of poultry production are collocated were found to be at greater risk.

Introduction
Unraveling mechanisms that underlie new and reemerging infectious diseases (EID) requires exploring complex interactions within and among coupled natural and human systems and poses one of the most difficult scientific problems facing society today (Wilcox and Colwell 2005). EID are diseases that have recently increased in incidence or in geographic or host range (e.g., tuberculosis, cholera, malaria, dengue fever), and diseases caused by new pathogens and new variants assigned to known pathogens (e.g., H5N1, SARS, Nipah virus, and avian influenza) (Morse 1995). Wilcox and Gubler (2005) and Wilcox and Colwell (2005) argue that transformations in ecological systems caused by multifaceted interactions with anthropogenic environmental changes such as urbanization, agricultural transformations, and natural habitat alterations produce feedbacks that affect natural communities and ultimately their pathogens, animal host, and human populations. These altered 'host-pathogen' relationships facilitate

pathogen spillover into "new" hosts, rapid adaptations by pathogens, more frequent generation of novel pathogen variants that result in new and reemerging infectious diseases, as well as range expansion and increasing epidemic intensity and frequency of existing diseases.

Kapan et al. (2007) hypothesized that the on-going process in Southeast Asia of replacing traditional farming methods such as multi-species livestock husbandry with industrial, mass-production-oriented operations poses significant environmental health risks (e.g., (Mallin and Cahoon 2003)) due to increases in livestock pools and thus opportunities for disease transmission. Simultaneously, rapid urban and peri-urban development in these countries has often been accompanied by more refuse, standing water, and animals in and around homes that have been correlated with environmental health risks (e.g., (Graham, Gurian et al. 2004)). With respect to HPAI, expansion of the urban fringe has placed a larger proportion of the human population in contact with formerly dispersed farm environments that include potentially infected poultry and swine populations. Such urban–rural interfaces have been hotspots of other infectious diseases such as leishmaniasis (Oliveira, Lacerda et al. 2004).

However, empirical studies on the link between urbanization and health are difficult to conduct (Vlahov and Galea 2002; Eckert and Kohler 2014) and scant evidence exists to support the theoretical argument. Importantly, such links have not been studied for EIDs, especially for zoonotic EIDs in developing countries. In addition, most studies and reviews of the relationship between urbanization and health are silent on the issue of whether the relationship is monotonic across the rural-urban gradient. This is mainly because of the prevailing belief that places are either rural or urban; in reality there is a rural-to-urban continuum, ranging from sparsely populated isolated settlements to small towns to secondary cities to megacities; and in any given country there is heterogeneity within areas that are classified as rural or urban (Aoyama and Horner 2010; Saksena, Fox et al. 2014).

Peri-urban areas are characterized by patchwork development and mixed land use, with large amounts of land still in agricultural use. McGee's concept of desakota (Indonesian for "village-town") is perhaps the best known model of the peri-urbanization process (McGee 1991) . McGee identifies six characteristics of a desakota region: 1) a large population of smallholder cultivators; 2) an increase in non-agricultural activities; 3) extreme fluidity and mobility of population; 4) a mixture of land uses, agriculture, cottage industries, suburban development; 5) increased participation of the female labor force; and 6) "grey-zones", where informal and illegal activities group (McGee 1991). In a similar vein, Webster (2002) lists four characteristics of the peri-urbanization process. These include: (1) a shift from an agriculturally based to a manufacturing-dominated economy; (2) a shift in employment from agriculture to manufacturing; (3) rapid population growth; and (4) changing spatial development patterns and rising land costs.

Spencer (2013) sought to establish whether bird deaths followed a Kuznets curve as settlement infrastructure patterns evolved. Vietnam's 1999 Census of Population and Housing

provides counts of households by housing construction materials (traditional/temporary or modern), water supply (stream, rain, well, piped), and sanitation infrastructure (none, pit, composting, flush). Spencer converted each of these 4-category, ranked urbanization measures into four distinct measures of settlement "coherence". He found evidence that increased diversity in water supply sources and sanitation infrastructure, evidence of places in rapid transition with neither typical rural nor urban sources of water and sanitation but a mixture of both, were associated with higher incidences of HPAI.

Previous studies have examined the effect of urbanization on H5N1 at different scales and for areas of varying sizes (Hogerwerf, Wallace et al. 2010; Loth, Gilbert et al. 2011). Some studies have used official administrative definitions, but this approach is limited in its bluntness. Others prioritize human population density as a satisfactory surrogate (Gilbert, Chaitaweesub et al. 2006; Gilbert, Xiao et al. 2008; Adhikari, Chettri et al. 2009; Tiensin, Ahmed et al. 2009; Loth, Gilbert et al. 2010; Yupiana, de Vlas et al. 2010; Loth, Gilbert et al. 2011; Martin, Pfeiffer et al. 2011), but this approach ignores the important fact that density is not a risk factor if it is accompanied by sufficient infrastructure to manage the density. These studies have used only a limited definition of urbanization, without a well-defined characterization of peri-urbanization, and without controls for other factors. Thus, it is not known whether the level of urbanization is truly an independent risk factor or whether there is an association between other risk factors and the level of urbanization.

One of the consequences of urbanization is the change in land-use diversity (which is also one of the consequences of other forces such as agricultural intensification, industrial development, deforestation,etc.). Changes in land- use and deforestation have been proposed as underlying drivers for a wide variety of emerging infectious diseases (Murray and Daszak 2013; Gottdenker, Streicker et al. 2014). Current conceptual models highlight two main mechanisms by which land- use change lead to disease emergence; a) land- use change perturbs disease dynamics in multi-host disease systems by disrupting the cross-species transmission rate and b) land- use change allows exposure of novels hosts to a rich pool of pathogen diversity, influencing the cross-species transmission rate (Murray and Daszak 2013) . It has been estimated that almost 75% of zoonotic diseases are associated with land- use changes (Taylor, Latham et al. 2001; Jones, Patel et al. 2008). Changes in land-use are often accompanied by landscape/habitat fragmentation. In general it has been theorized that landscape fragmentation can lead to EIDs, especially zoonotic diseases (Jones, Grace et al. 2013). The fragmentation mechanism has been tested empirically only in small areas using metrics such as ecotones, patchiness, etc. for diseases such as Lyme, Tsetse, etc. Such evidence is not available for H5N1. For all EIDs, empirical evidences are not available from country-wide studies. One reason for the limited data is that landscape fragmentation metrics (using software such as Fragstats, etc) rely on remote sensing data, which may not be easily accessible for large scale studies, especially in developing countries. Using census data is a novel way to study land-use diversity. Most

studies have shown only that fragmentation is associated with the abundance of the reservoir or vector related to a disease. Rarely has an attempt been made to relate land/habitat fragmentation to the final health endpoints in humans or livestock.

A handful of studies have examined the role of landscape fragmentation, which we describe here. In a small area study of Leishmaniasis in Southwest France, researchers used risk predictors reflecting both land-cover and composition (shape, size of forest and crop patches, distance between patches and a diversity index (Hartemink, Vanwambeke et al. 2011) using remote sensing. Mean shape index of crops was found to be the best performing predictor for one of the vectors of Leishmaniasis. In other small area studies in eastern Zambia (Ducheyne, Mweempwa et al. 2009) and Burkina Faso ((Guerrini, Bord et al. 2008), researchers found that indices of habitat fragmentation derived from remote sensing (area, mean patch size, number of patches, standard deviation of patch size, etc.) was associated with the abundance of tsetse flies – a vector for human and livestock trypanosomiasis. Hantavirus reservoirs in Panama were commonly found in fragmented habitats as measured by edges (Suzan, Marce et al. 2008). In Bangladesh Nipah Belt villages had forests that were more fragmented than elsewhere (Hahn, Gurley et al. 2014). In a larger regional study of the thirteen northeastern US states, Lyme disease incidence was found to be associated with greater fragmentation between forests and residential areas (Tran and Waller 2013). Based on 11 years of panel surveillance data in New York state, a study found that increasing points of contact between forested and developed land, as measured by their shared edges and by the perimeter length of forested patches, were associated with higher incidence of giardiasis cases (Walsh 2013). Thus greater ecological overlap leads to increased interspecies transmission in either direction. Ecotonal–level changes resulting from large-scale land-use conversion from wetlands to rice-cultivation or dryland forests to farming are also associated with disease emergence (Despommier, Ellis et al. 2006).

The recent rapid urbanization in developing countries is creating a huge demand from cities for vegetables, fruits, meat, etc. Much of this demand is being met by peri-urban farms (Birley and Lock 1998). New farms that have come up in peri-urban areas are mainly the consequence of Foreign Direct Investment (Burgos, Hong Hanh et al. 2007). Due to the globalization of poultry trade, intensive farms have expanded rapidly in Southeast Asia and compete with existing small backyard farmers. Truly enterprise-scale (15,000 – 100,000 birds) of operations are still rare in Viet Nam (only 33 communes have such a facility). On the other hand, farming based on contracts with regional multinational companies or national companies, as a mode of intensive farming, is very common. Such intensive farms typically have 2,000 - 15,000 birds. The relative role of extensive (backyard) systems and intensive systems has been the subject of recent studies (Gilbert, Chaitaweesub et al. 2006; Tiensin, Ahmed et al. 2009; Loth, Gilbert et al. 2010; Loth, Gilbert et al. 2011; Paul, Wongnarkpet et al. 2011). In much of Asia there is often a mix of commercial and backyard farming at any one location (Philip 2010). From a biosecurity perspective this has been suggested to be a matter of concern (Horby,

Pfeiffer et al. 2013). It has been theorized that the co-location of extensive and intensive systems is a potential risk factor. Intensive systems allow for virus evolution (e.g. LPAI to HPAI) and transformation, while extensive systems allow for environmental persistence and circulation (Jones, Grace et al. 2013). Empirical evidence is lacking so far. While studying chicken population as a risk factor many studies have made a distinction between production systems – native chickens, backyard chicken, flock density, commercial chicken, broilers and layers density, etc. (Gilbert, Chaitaweesub et al. 2006; Tiensin, Ahmed et al. 2009; Loth, Gilbert et al. 2010; Loth, Gilbert et al. 2011; Paul, Wongnarkpet et al. 2011; Jones, Grace et al. 2013). In isolation, however, none of these number and/or density based poultry metrics adequately measures the extent of colocation of intensive and extensive systems in any given place. Intensive and extensive systems in Viet Nam have their own fairly well defined flock sizes. A diversity index can capture better the effect of such co-location.

Aims

This study aimed to examine retrospectively the Viet Nam country-wide association between urbanization, land-use diversity, and livestock production structures and the risk of H5N1 in poultry at the lowest administrative level. To address confounding, we rely on an optimal number of known risk factors based on previous work and introduce a few new ones never before studied in the context of H5N1. We also aimed to understand the differences, if any, in country-wide risk dynamics compared to sub-national agro-ecological zones. For this purpose we chose to study the Red River Delta and Mekong River Delta that are well known hot spots of the disease. Given the highly complex dynamics of the epidemics and in keeping with recent methodological trends, we use multiple modeling approaches – parametric and non-parametric – with a focus on spatial analysis. We use both "place" oriented and "space" oriented approaches (Arcaya, Brewster et al. 2012).

Methods

Study area

The study population consisted of all the 10820 communes and wards of Viet Nam, excluding its islands. A digital map of the administrative boundaries at province, district and commune level was obtained from the Viet Nam General Statistics Office (GSO) for the year 2006. The point locations of all commune centroids in the study area were calculated using the geographical information system (GIS) software ArcGIS 10.2 (ESRI, Inc.)

HPAI outbreak information

Commune level data on outbreaks (absence or presence) were obtained from the publicly available database of the World Organisation of Animal Health (OIE). Viet Nam experienced its

first major epidemic waves between December 2003 to February 2006 (Pfeiffer, Minh et al. 2007). We chose to study the first wave (Wave 1) that ended in February 2004 and the second wave (Wave 2) that occurred between December 2004 and April 2005. In Wave 1, 21% of the communes and in Wave 2, 6% of the communes experienced an outbreak(s).

Risk predictors

We investigated a wide suite of socio-economic, agricultural, climatic and ecological variables relevant to poultry management and the transmission and persistence of the HPAI virus. Many of these variables were identified based on earlier studies of HPAI (Gilbert and Pfeiffer 2012). A couple of novel factors were included based our own hypothesis. All variables were measured or aggregated to the commune level. These risk factors are grouped below firstly by listing those directly related to our hypotheses and then according to their role in disease introduction, transmission and persistence, though some of these factors may have multiple roles.

- Degree of urbanization: We used a recently developed three-way classification of the rural-to-urban gradient in Viet Nam (Saksena, Fox et al. 2014). The classification framework is based on four characteristics: a) percentage of households whose main income is from agriculture, aquaculture and forestry, b) percentage of households with modern forms of toilets, c) percentage of land under agriculture, aquaculture and forestry and d) Normalized Differentiated Vegetation Index (NDVI). The three-way classification enabled testing for non-linear responses.
- Land use diversity: We measured land use diversity using the Gini-Simpson Diversity Index (Hill 1973). The Gini-Simpson Diversity Index is given by $1 - \lambda$, where $-\lambda$, where $\lambda = \sum_{i=1}^{n}(p_i)^2$. λ equals the probability that two entities taken at random from the dataset of interest represent the same type. In situations with only one class (complete homogeneity) the Gini-Simpson index would have a value equal to zero. Such diversity indices have been used to measure land use diversity (Forman 1995) We used the following five land use classes: annual crops, perennial crops, forests, aquaculture and built-up land (including miscellaneous uses). The area under the last class was calculated as the difference between the total area and the sum of the first four classes. These four classes are the main categories used in government census surveys in Viet Nam.
- Chicken flock size diversity: We used Gini-Simpson's Diversity index based on the following size classes: 1-50, 51-150, 151-2000 and > 2000. Previous studies have shown that in Viet Nam, typical backyard flock sizes are 1-50) and typical flock sizes in contract poultry operations are > 2000 (Burgos, Hong Hanh et al. 2007). The contract poultry owners are small commercial enterprises (intensive).
- Duck flock size diversity was also calculated using the Gini-Simpson Diversity index based on the following size classes: 1-50, 51-150, 151-2000 and > 2000
- Human population related transmission
  - Human population density
- Poultry trade and market

- o Towns and cities were assumed to be active trading places. So, the distance to the nearest town/city was used as indicator of poultry trade.
  - o Trade is facilitated by access to good transportation infrastructure. So, the shortest distance to the nearest a) national highway and b) provincial highway was used as indicator of transportation infrastructure.
- Disease introduction and amplification
  - o The density of non-industrial chicken and industrial chicken were estimated based on commune area.
- Intermediate hosts
  - o Duck and geese density were calculated using total commune area
  - o As previous studies have shown a link between scavenging in rice fields by ducks and outbreaks, we also calculated duck density using area under rice calculation.
- Agro-ecological and environmental risk factors
  - o Previous studies have shown that the extent of rice cultivation is a risk factor, mainly due its association with free ranging ducks acting as scavengers (Pfeiffer, Minh et al. 2007). We used percentage of land under rice cultivation as an indicator for this.
  - o Rice cropping intensity is also a known risk factor (Gilbert, Xiao et al. 2008; Paul, Tavornpanich et al. 2010; Loth, Gilbert et al. 2011) .
  - o The extent of aquaculture is a known risk factor (Pfeiffer, Minh et al. 2007), possibly because as a water body it offers routes for transmission and persistence of the virus. The percentage of land under aquaculture was used a metric.
  - o Proximity to water bodies increases the risk of outbreaks (Fang, de Vlas et al. 2008; Ward, Maftei et al. 2008; Biswas, Christensen et al. 2009; Cao, Xu et al. 2010), possibly by increasing the chance of contact between wild water birds and domestic poultry. We measured the shortest distance between the commune and the nearest a) lake and b) river.
  - o Bio-climatic variables –annual mean temperature and annual precipitation – have been associated with significant changes in risk (Fang, de Vlas et al. 2008; Tran, Yost et al. 2013).
  - o Elevation, which is associated with types of land cover and agriculture, has been shown to be a significant risk factor in Vietnam (Pfeiffer, Minh et al. 2007).
  - o Compound Topographical Index (also known as Topographical Wetness Index) is a measure of the tendency to pool water at a place. Studies in Thailand and elsewhere (Thanapongtharm, Van Boeckel et al. 2013) have shown that the extent of surface water is a strong risk factor, possibly due to the role of water in long-range transmission and persistence of the virus. In the absence of reliable and inexpensive data on the extent of surface water we used CTI as a proxy. CTI has

been used in Ecological Niche Models (ENM) of H5N1 (Williams, Fasina et al. 2008; Williams and Peterson 2009). However, given the nature of ENM studies, the effect of CTI as a risk factor has been unknown so far. CTI has been used a risk factor in the study of other infectious and non-infectious diseases (Sarkar, Strutz et al. 2010). Some studies have shown that at local scales, the slope of the terrain (a component of CTI) was significantly correlated with reservoir species dominance (Suzan, Marce et al. 2008). CTI is a function of both the slope and the upstream contributing area per unit width orthogonal to the flow direction. The implementation of CTI is: CTI = ln (As / (tan (beta)) where; As = Area Value calculated as (flow accumulation + 1) * (pixel area in m2) and beta is the slope expressed in radians (Gessler, Moore et al. 1995).

o Though previous studies have indicated that NDVI is a risk factor (Pfeiffer, Minh et al. 2007; Adhikari, Chettri et al. 2009; Henning, Pfeiffer et al. 2009; Si, Skidmore et al. 2009; Williams and Peterson 2009), we did not include it explicitly in our models, as the urban classification index we used includes NDVI.

Data sources

Human population data were obtained from the 1999 Population Census of Viet Nam. Agriculture Census is conducted every five years in Viet Nam. We noted that the first three epidemic waves occurred between the two censuses – 2001 and 2006. Most of these epidemic waves occurred closer in time to the 2006 agriculture census (Pfeiffer, Minh et al. 2007). So we chose to use data from the 2006 Agriculture Census as this would be more truly reflective of conditions around the outbreaks. However, for data on number of poultry we chose to use the 2001 Agriculture Census data set. The justification for this that between 1991 to 2003 poultry population grew at an average rate of 7% on a year over year basis. But due to the first wave of H5N1 epidemic the poultry population fell by 15% in 2004 (GSO). Only by mid-2008 did poultry population return close to pre-epidemic levels. Thus, poultry population data from the 2001 census were considered more representative. We aggregated the census household data to the commune level. Urban classification was based on a related study (Saksena, Fox et al. 2014). Raster data on annual mean temperature and precipitation were obtained from the WorldClim database and converted to commune level data. The bioclimatic variables were compiled from the monthly temperature and precipitation values and interpolated to surfaces at high spatial resolution (measured at 90m spatial resolution) (Hijmans, Cameron et al. 2005).

Elevation was generated from SRTM 90 meter Digital Elevation Models (DEM) acquired from the Consortium for Spatial Information (*CGIAR-CSI*). CTI data were generated using the Geomorphometry and Gradient Metrics Toolbox for ArcGIS 10.1.

Statistical and spatial analysis

In keeping with the principle of triangulation we used two statistical methods to assess the significance and direction of effect of the risk factors. Prior to risk factor analysis we cleaned the data by identifying illogical values for all variables and then either assigning a missing value to them or adjusting the values. Illogical values occurred mainly (less than 1% of the cases) for land-related variables such as percentage of commune land under a particular type of land use. Next we tested each variable for normality using the BestFit software (Palisade Corporation). Most of the variables were found to follow a log-normal distribution and a log-transform was used on them. We then examined the bi-variate correlations between all the risk factors (or their log-transform, as the case may be). Certain risk factors were then eliminated from consideration when │r│ ≥ 0.5 (r is the Pearson correlation coefficient). Notably, we excluded a) elevation (correlated with human population density, chicken density, duck density, percentage land under paddy, annual temperature and compound topographical index), b) human population density (correlated with elevation and CTI), c) duck and goose density (correlated with elevation, chicken density, percentage land under paddy, land use diversity index and CTI), d) annual temperature (correlated with elevation and CTI) and d) cropping intensity (correlated with percentage land under paddy).

Considering the importance of spatial autocorrelation in such epidemics, we used two modeling approaches: a) multi-level generalized linear mixed modeling (GLMM) and b) Boosted Regression Trees (BRT) (Elith, Graham et al. 2006; Elith, Leathwick et al. 2008) with an autoregressive term (Augustin, Mugglestone et al. 1996). We began by deriving an autoregressive term by averaging the presence/absence among a set of neighbors defined by the limit of autocorrelation, weighted by the inverse of the Euclidean distance (Augustin, Mugglestone et al. 1996). The limit of the autocorrelation of the response variable was obtained from the range of the spatial correlogram ρ (h) (Pebesma 2004). To determine which predictor variables to include in the two models, we conducted logistic regression modeling separately for each of them one by one but included the autoregressive term each time. We finally included only those variables whose coefficient had significance value p ≤0.2 (in at least one wave-region combination) and we noted the sign of the coefficient

We used a two-level GLMM (communes nested under districts) to take account of random effects for an area influenced by its neighbors, and thus, we studied the effect of spatial autocorrelation. We used robust standard errors for tests of fixed effects. Boosted regression trees, also known as stochastic gradient boosting, was performed to predict the probability of HPAI H5N1 occurrence and determine the relative influence of each physical environment factor to the HPAI H5N1 occurrence. This method has been recently developed and widely applied for distribution prediction in various fields of ecology (Elith, Graham et al. 2006; Elith, Leathwick et al. 2008). It is widely used for species distribution modeling where only the sites of occurrence of the species are known (Elith and Leathwick 2009). The method has been applied

in numerous studies for prediction distribution of HPAI H5N1 disease (Martin, Pfeiffer et al. 2011; Van Boeckel, Thanapongtharm et al. 2012; Fang, Li et al. 2013; Stevens, Gilbert et al. 2013; Gilbert, Golding et al. 2014). BRT utilizes regression trees and boosting algorithms to fit several models and combine them for improving prediction by performing iterative loop throughout the model (Elith, Graham et al. 2006; Elith, Leathwick et al. 2008).

The advantage of BRT is that it applies stochastic process which include probabilistic component to improve predictive performance. The use of regression trees is to select relevant predictor variables and boosting is to improve accuracy in a single tree. The sequential process allows trees to be fitted iteratively through a forward stage-wise procedure in the boosting model. Two important parameters specified in the BRT model are learning rate (lr) and tree complexity (tc) to determine number of trees for optimal prediction (Elith, Graham et al. 2006; Elith, Leathwick et al. 2008). In our model we used 10 sets of training and test points for cross-validation, a tree complexity of 5, a learning rate of 0.01 and a bag fraction of 0.5. Other advantages of BRT include its insensitivity to colinearity and non-linear responses. However, for the sake of consistency with the other two methods, we chose to eliminate predictors that were highly correlated with other predictors and to make log-transforms where needed.

The predictive performances of the models were assessed by the area under the curve (AUC) of the receiver operation characteristic (ROC) curve. AUC is a measure of the overall fit of the model that varies from 0.5 (chance event) to 1.0 (perfect fit) (Greiner, Pfeiffer et al. 2000). A comparison of AUC with other accuracy metrics concluded that it is the most robust measure of model performance because it remained constant over a wide range of prevalence rates (McPherson, Jetz et al. 2004). We used the Akaike Information Criteria (AIC) to compare each GLMM model with and without its respective suite of fixed predictors.

We used SPSS version 21 (IBM Corp., New York, 2012) for GLMM and R version 3.1.0 (The R Foundation for Statistical Computing, 2014) for the BRT. For calculating the spatial correlogram we used the spdep package of R.

Results

The range of spatial autocorrelation estimated from correlograms for Wave 1 was 30 km, 20 km and 100 km for Viet Nam, Red River Delta and Mekong River Delta respectively. For Wave 2 the range was 30 km, 40 km and 30 km respectively.

The unadjusted coefficients of the predictors based on logistic regression (but including the autocorrelation term) are shown in Table 1. All the fourteen predictors were found to be significantly associated with H5N1 outbreaks (P < 0.2) in at least one wave-region combination based on this univariate analysis. Land-use diversity, chicken density, poultry flock size diversity and distance to national highway were found to have significant associations across five of the six wave-region combinations. The final models for GLMM and BRT are shown in Tables 2-7.

The predictive power of the GLMM models, as measured by the AUCs is very good with AUC values ranging from 0.802 to 0.951 The predictive power of the national models was higher than that of the zonal models. The predictive power of the BRT models is good, with AUCs ranging from 0.737 to 0.914. The BRT models also had a better predictive power at the national level than the zonal level.

Our models have a far higher predictive power than that of the only other published model for Vietnam's Wave 1 (AUC = 0.69) and Wave 2 (AUC = 0.77) (Gilbert, Xiao et al. 2008). A common result between their model and our model is that, at the national level, the predictive performance for Wave 2 was higher than that for Wave 1.

We found very high spatial clustering effects as indicated by the fact that the in all waves and regions the BRT model found the autocorrelation term to have the highest rank of influence. As expected, the relative influence of the autocorrelation term at the national level was far higher (60-78%) than at the zonal levels (14– 35%).

In the GLMM models we found the AICs using the entire set of 14 variables to be much lower than the AICs of a GLMM model without fixed effects. This indicated that though clustering effects were significant, our theory driven predictor variables improved model performance.

For both waves the urbanization variable was not found to be significantly associated with H5N1 in any region according to the GLMM model except for the urban level in RRD for Wave 2 and in MRD for Wave 1. The BRT model ranked urbanization as one of the least influential variables. Land –use diversity was found to be significantly associated with H5N1 in both waves for Viet Nam according to the GLMM model, but at the zonal level the association was significant only for Wave 2 in MRD. The BRT model indicated that land-use diversity had a highly ranked influence in Viet Nam for Wave 2. For rest of the regions and waves land-use diversity had middle to below-middle rank of influence. Both the GLMM and BRT models indicated that chicken flock-size diversity had a very strong association with H5N1 for both waves at the national level. This was generally found to be true at the zonal levels with some exceptions. The duck and goose flock size diversity was also significantly associated with H5N1 in all regions, but the associations were much stronger in Wave 2 than in Wave 1. The GLMM model indicated that CTI had a very strong association with H5N1 at the national level in both waves both this was not true in the two deltas. The BRT model however indicated that CTI had a middle to low influence in all waves and regions.

Discussion

Based on univariate analysis, we found that urbanization, at the national level, was a significant risk factor for both waves. In addition, peri-urban areas were at higher risk than rural or urban areas. However, according to both multi-variate models the urbanization variable was not significant. This suggests that urbanization is a very broad surrogate for a group of risk factors. These risk factors combine in ways to make peri-urban areas the riskiest places. Indeed we found that chicken density, percentage of land under rice, percentage of land under aquaculture, flock size diversity for duck and geese and CTI to be highest in peri-urban areas (Figs 1 -5). We also found that land-use diversity was highest in rural areas, but peri-urban areas had diversity levels only marginally lower than that in rural areas (Fig 6).

Only one other study so far has explicitly examined peri-urban areas in the context of H5N1. A study in Indonesia found peri-urban areas to be significantly associated with H5N1 cases, even based on multivariate models (Loth, Gilbert et al. 2011). Our study has determined what features of peri-urban areas make them risky places. Therefore, after including those features in multi-variate models, the role of the urbanization variable per se diminishes.

At the delta level, generally urbanization had no significant association with H5N1. This may partly be due to the fact that the deltas are more homogenous, in terms of urbanization, when compared to the country as a whole.

This is the first time land use diversity has been studied as a risk factor for H5N1. Our results indicate a very strong association at the country level and in MRD. The results are similar to what has been observed for other EIDs using fragmentation metrics. Using census data for measuring land –use diversity is a novel approach that can be quickly implemented in large area (regional or national) studies. Our metric is constrained by the land-use categories used by GSO, which reflect priority policy issues for the government of Viet Nam. Future studies could examine the correlation between such a census-based metric with metrics derived from remote sensing.

This study has significantly advanced knowledge about the role of livestock production systems. Previous studies had examined the factors such as type of species, size of flocks and extent of commercialization. For the first time there is strong evidence that when intensive and extensive systems of production co-exist in the same place, such places are likely to experience higher risk. Future studies need to examine the biological causal mechanisms in this context.

For previously known risk factors, such as land under rice, land under aquaculture and poultry density, our models are agreement with associations reported in other studies, except that our coefficients are generally higher than what has been previously reported.

Both GLMM and BRT models found annual precipitation to be a significant factor. The GLMM model indicated a negative association, similar to what was found by another study in China (Fang, de Vlas et al. 2008) and in RRD (Tran, Yost et al. 2013). Generally, the role of precipitation was found to be far more significant in the deltas than for the country.

Acknowledgments

Table 1: Unadjusted coefficients (β) for the final set of predictors based on autologistic regression

Predictor	Wave 1			Wave 2		
	Viet Nam	Red River Delta	Mekong River Delta	Viet Nam	Red River Delta	Mekong River Delta
Urbanicity: rural	0 0.000#	0 0.027	0 0.462	0 0.000	0 0.027	0 0.683
Urbanicity: peri-urban	0.322 0.000	0.285 0.011	-0.105 0.507	0.591 0.000	0.656 0.007	0.027 0.871
Urbanicity: urban	0.231 0.112	-0.186 0.571	0.378 0.321	-0.077 0.792	0.085 0.909	-0.353 0.401
Percentage land under rice*	2.125 0.000	1.454 0.084	0.215 0.77	5.633 0.000	1.937 0.346	5.646 0.000
Percentage land under aquaculture*	1.535 0.086	0.912 0.739	4.630 0.000	-1.115 0.503	4.280 0.438	-4.314 0.019
Land-use diversity (Gini-Simpson index)	0.802 0.000	1.399 0.000	0.678 0.107	1.216 0.000	0.759 0.383	1.345 0.007
Chicken density*	0.399 0.000	0.495 0.000	0.030 0.747	0.536 0.000	1.158 0.015	0.489 0.000
Duck-rice area density	0.060 0.511	-0.288 0.743	0.247 0.223	0.105 0.059	-14.222 0.558	-0.880 0.341
Chicken flock size diversity (Gini-Simpson Index)	2.230 0.000	3.843 0.000	-0.211 0.770	1.295 0.032	3.523 0.012	1.741 0.046
Duck & goose flock size diversity (Gini-Simpson Index)	0.631 0.004	0.959 0.068	0.000 0.994	2.275 0.000	2.393 0.005	2.846 0.000
Annual precipitation*	1.287 0.001	6.699 0.172	3.08 0.015	0.161 0.823	13.743 0.184	0.234 0.834
Compound Topographical Index*	3.890 0.000	-1.561 0.660	-3.912 .619	6.366 0.000	16.959 0.116	-6.019 0.504
Shortest distance to nearest national highway*	-0.020 0.318	-0.041 0.161	0.063 0.061	-0.039 0.169	-0.184 0.006	-0.040 0.26
Shortest distance to nearest provincial highway*	-0.041 0.009	-0.020 0.436	0.000 0.919	-0.119 0.000	-0.114 0.140	-0.065 0.074
Shortest distance to nearest town*	-0.009 0.683	0.157 .002	0.060 0.181	-0.073 0.044	0.055 0.607	-0.073 0.120
Shortest distance to nearest lake*	-0.074 0.009	0.058 0.433	-0.010 0.591	0.069 0.330	-0.141 0.298	0.061 0.629

*Transform of the type $\log_{10}(1+x)$ was used, #p values

Table 2: Model results for Viet Nam, Wave 1

	GLMM			BRT		
	Coefficient	p	s.e	Relative Influence (%)	s.e	rank
Intercept	-5.961	0.160	4.245	n/a		
Urbanicity: rural	0			0.562	0.101	12
Urbanicity: peri-urban	0.050	0.693	0.126			
Urbanicity: urban	0.417	0.110	0.261			
Percentage land under rice*	2.014	0.038	0.970	1.09	0.04	8
Percentage land under aquaculture*	2.315	0.213	1.859	1.681	0.036	5
Land-use diversity (Gini-Simpson index)	0.683	0.065	0.370	1.715	0.103	4
Chicken density*	0.134	0.285	0.126	3.081	0.101	2
Duck-rice area density	0.085	0.007	0.032	1.268	0.101	7
Chicken flock size diversity (Gini-Simpson Index)	1.510	0.013	0.606	5.034	0.037	1
Duck & goose flock size diversity (Gini-Simpson Index)	0.481	0.135	0.322	1.649	0.037	6
Annual precipitation*	-1.908	0.100	1.159	3.005	0.04	3
Compound Topographical Index*	8.811	0.000	2.442	0.904	0.036	10
Shortest distance to nearest national highway*	0.008	0.763	0.026	0.268	0.103	14
Shortest distance to nearest provincial highway*	-0.014	0.534	0.023	0.341	0.04	13
Shortest distance to nearest town*	-0.055	0.198	0.043	0.793	0.037	11
Shortest distance to nearest lake*	-0.050	0.29	0.047	1.067	0.103	9
Autoregressive term	n/a			77.542	0.036	
AUC-ROC	0.907			Trg = 0.863, Eval = 0.841		

*Transform of the type $\log_{10}(1+x)$ was used

Rank = rank of relative influence excluding the rank of the autorgressive term

Trg = Training, Eval = Evaluation

Table 3: Model results for Viet Nam, Wave 2

	GLMM			BRT		
	Coefficient	p	s.e	Relative Influence (%)	s.e	Rank
Intercept	-8.511	0.206	6.729			
Urbanicity: rural	0			0.5	0.1	14
Urbanicity: peri-urban	0.273	0.124	0.178			
Urbanicity: urban	0.228	0.696	0.582			
Percentage land under rice*	6.046	0.000	1.400	5.22	0.04	1
Percentage land under aquaculture*	1.024	0.724	2.904	3.17	0.04	6
Land-use diversity (Gini-Simpson index)	2.212	0.001	0.646	2.22	0.1	9
Chicken density*	-0.525	0.003	0.178	4.21	0.1	4
Duck-rice area density	0.203	0.041	0.099	3.77	0.08	5
Chicken flock size diversity (Gini-Simpson Index)	1.837	0.020	0.788	4.6	0.08	3
Duck & goose flock size diversity (Gini-Simpson Index)	1.986	0.000	0.480	5.09	0.1	2
Annual precipitation*	-4.698	0.000	1.323	2.64	0.1	7
Compound Topographical Index*	14.627	0.000	3.661	2.56	0.08	8
Shortest distance to nearest national highway*	-0.040	0.271	0.036	1.96	0.08	10
Shortest distance to nearest provincial highway*	-0.119	0.001	0.035	0.75	0.08	13
Shortest distance to nearest town*	-0.127	0.053	0.066	1.52	0.1	12
Shortest distance to nearest lake*	0.940	0.021	0.408	1.75	0.04	11
Autoregressive term	n/a			60.04	0.08	
AUC-ROC	0.951			Trg = 0.944, Eval = 0.914		

*Transform of the type $\log_{10}(1+x)$ was used

Rank = rank of relative influence excluding the rank of the autorgressive term

Trg = Training, Eval = Evaluation

Table 4: Model results for Red River Delta, Wave 1

	GLMM			BRT		Rank
	Coefficient	p	s.e.	Relative Influence (%)	s.e	
Intercept	108.702	0.001	33.225	n/a		
Urbanicity: rural	0			0.911	0.042	13
Urbanicity: peri-urban	0.009	0.986	0.518			
Urbanicity: urban	0.179	0.383	0.205			
Percentage land under rice*	-1.643	0.362	1.802	5.028	0.058	7
Percentage land under aquaculture*	-0.974	0.799	3.832	6.3	0.042	3
Land-use diversity (Gini-Simpson index)	0.959	0.338	1.000	4.259	0.042	8
Chicken density*	0.970	0.012	0.385	8.275	0.058	2
Duck-rice area density	-6.879	0.129	4.525	6.067	0.042	5
Chicken flock size diversity (Gini-Simpson Index)	2.424	0.073	1.352	10.618	0.049	1
Duck & goose flock size diversity (Gini-Simpson Index)	0.051	0.920	0.511	5.611	0.049	6
Annual precipitation*	-32.626	0.003	11.114	6.208	0.088	4
Compound Topographical Index*	-7.193	0.355	0.119	3.363	0.042	10
Shortest distance to nearest national highway*	-0.079	0.082	0.045	1.594	0.058	12
Shortest distance to nearest provincial highway*	-0.027	0.472	0.038	0.493	0.088	14
Shortest distance to nearest town*	0.173	0.034	0.082	2.269	0.049	11
Shortest distance to nearest lake*	-0.041	0.730	0.119	3.57	0.088	9
Autoregressive term	n/a			35.434	0.042	
AUC-ROC			0.802	Trg = 0.827, Eval = 0.737		

*Transform of the type $\log_{10}(1+x)$ was used, #p values
Rank = rank of relative influence excluding the rank of the autorgressive term
Trg = Training, Eval = Evaluation

Table 5: Model results for Red River Delta, Wave 2

	GLMM			BRT		Rank
	Coefficient	p	s.e.	Relative Influence (%)	s.e	
Intercept	68.171	0.142	46.439			
Urbanicity: rural	0			2.13	0.04	13
Urbanicity: peri-urban	0.240	0.530	0.382			
Urbanicity: urban	0.041	0.033	1.245			
Percentage land under rice*	-7.892	0.046	3.944	6.81	0.09	6
Percentage land under aquaculture*	-2.552	0.688	6.362	6.31	0.07	7
Land-use diversity (Gini-Simpson index)	1.592	0.451	2.113	5.62	0.08	9
Chicken density*	1.017	0.033	0.476	10.34	0.07	2
Duck-rice area density	-70.094	0.030	32.181	7.95	0.08	4
Chicken flock size diversity (Gini-Simpson Index)	0.935	0.571	1.651	12.86	0.09	1
Duck & goose flock size diversity (Gini-Simpson Index)	1.934	0.087	1.130	8.07	0.04	3
Annual precipitation*	-32.390	0.038	15.625	7.85	0.28	5
Compound Topographical Index*	26.477	0.258	23.375	4.66	0.28	10
Shortest distance to nearest national highway*	-0.081	0.294	0.077	2.56	0.07	12
Shortest distance to nearest provincial highway*	-0.136	0.085	0.079	0.96	0.08	14
Shortest distance to nearest town*	0.093	0.619	0.188	2.72	0.09	11
Shortest distance to nearest lake*	-0.262	0.054	0.136	6.27	0.04	8
Autoregressive term	n/a			14.9	0.28	
AUC-ROC			0.902	Trg = 0.987, Eval = 0.755		

*Transform of the type $\log_{10}(1+x)$ was used, #p values
Rank = rank of relative influence excluding the rank of the autorgressive term
Trg = Training, Eval = Evaluation

Table 6: Model results for Mekong River Delta, Wave 1

	GLMM			BRT		Rank
	Coefficient	p	s.e.	Relative Influence (%)	s.e	
Intercept	35.702	0.129	23.528	n/a		
Urbanicity: rural	0			1.429	0.082	14
Urbanicity: peri-urban	0.149	0.592	0.277			
Urbanicity: urban	1.292	0.025	0.575			
Percentage land under rice*	3.632	0.021	1.571	4.317	0.059	8
Percentage land under aquaculture*	5.393	0.063	2.898	6.468	0.059	3
Land-use diversity (Gini-Simpson index)	-0.059	0.956	1.069	5.392	0.039	5
Chicken density*	-0.082	0.791	0.308	4.590	0.068	6
Duck-rice area density	0.297	0.454	0.396	4.457	0.072	7
Chicken flock size diversity (Gini-Simpson Index)	0.689	0.642	1.481	6.497	0.072	2
Duck & goose flock size diversity (Gini-Simpson Index)	-0.057	0.935	0.704	6.248	0.082	4
Annual precipitation*	-12.623	0.000	3.288	15.665	0.068	1
Compound Topographical Index*	3.183	0.863	18.482	3.516	0.082	9
Shortest distance to nearest national highway*	0.141	0.036	0.067	1.819	0.059	13
Shortest distance to nearest provincial highway*	0.012	0.825	0.054	2.058	0.068	11
Shortest distance to nearest town*	0.074	0.504	0.111	2.039	0.072	12
Shortest distance to nearest lake*	0.077	0.135	0.569	2.879	0.039	10
Autoregressive term	n/a			32.627	0.039	
AUC-ROC		0.891		Trg = 0.911, Eval = 0.811		

*Transform of the type $\log_{10}(1+x)$ was used, #p values
Rank = rank of relative influence excluding the rank of the autorgressive term
Trg = Training, Eval = Evaluation

Table 7: Model results for Mekong River Delta, Wave 2

	GLMM			BRT		Rank
	Coefficient	p	s.e.	Relative Influence (%)	s.e	
Intercept	-10.082	0.554	17.043			
Urbanicity: rural	0			0.72	0.07	14
Urbanicity: peri-urban	0.105	0.641	0.225			
Urbanicity: urban	0.176	0.758	0.569			
Percentage land under rice*	5.009	0.001	1.529	9.04	0.05	3
Percentage land under aquaculture*	0.936	0.781	3.360	6.75	0.07	6
Land-use diversity (Gini-Simpson index)	1.956	0.018	0.823	6.85	0.07	5
Chicken density*	0.364	0.145	0.250	7.2	0.07	4
Duck-rice area density	0.597	0.239	0.507	9.7	0.08	2
Chicken flock size diversity (Gini-Simpson Index)	1.674	0.147	1.153	5.91	0.08	7
Duck & goose flock size diversity (Gini-Simpson Index)	2.533	0.000	0.704	10.21	0.07	1
Annual precipitation*	-2.719	0.277	2.498	5.8	0.05	8
Compound Topographical Index*	8.874	0.529	14.076	4.13	0.07	19
Shortest distance to nearest national highway*	-0.019	0.716	0.052	3.84	0.07	10
Shortest distance to nearest provincial highway*	-0.149	0.005	0.053	2.14	0.05	13
Shortest distance to nearest town*	-0.074	0.433	0.094	3.36	0.07	11
Shortest distance to nearest lake*	1.006	0.005	0.382	3.2	0.08	12
Autoregressive term	n/a			21.14	0.07	
AUC-ROC		0.849		Trg = 0.926, Eval = 0.763		

*Transform of the type $\log_{10}(1+x)$ was used, #p values
Rank = rank of relative influence excluding the rank of the autorgressive term
Trg = Training, Eval = Evaluation

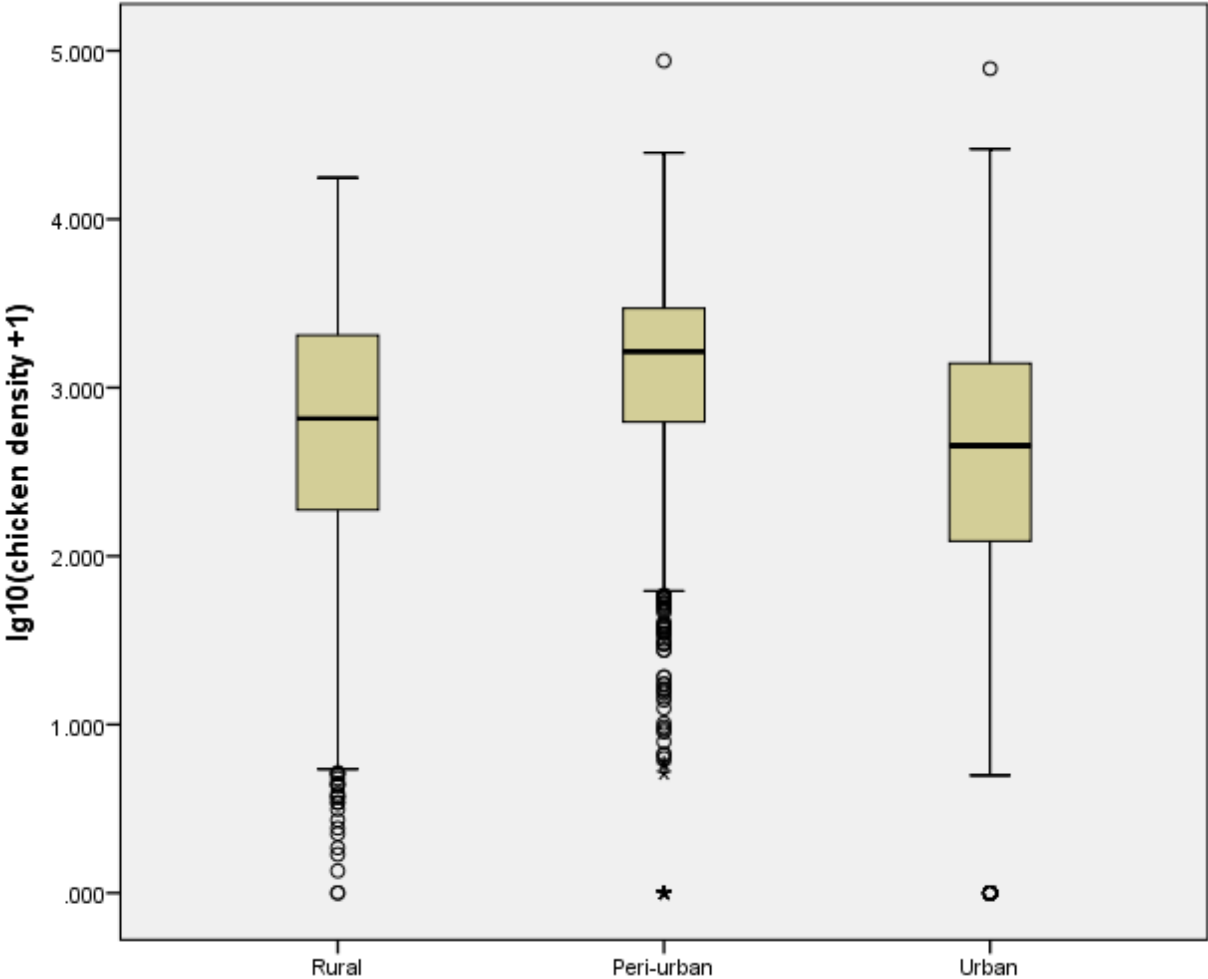**Figure 1: Variation of chicken density across urbanicity: Viet Nam**

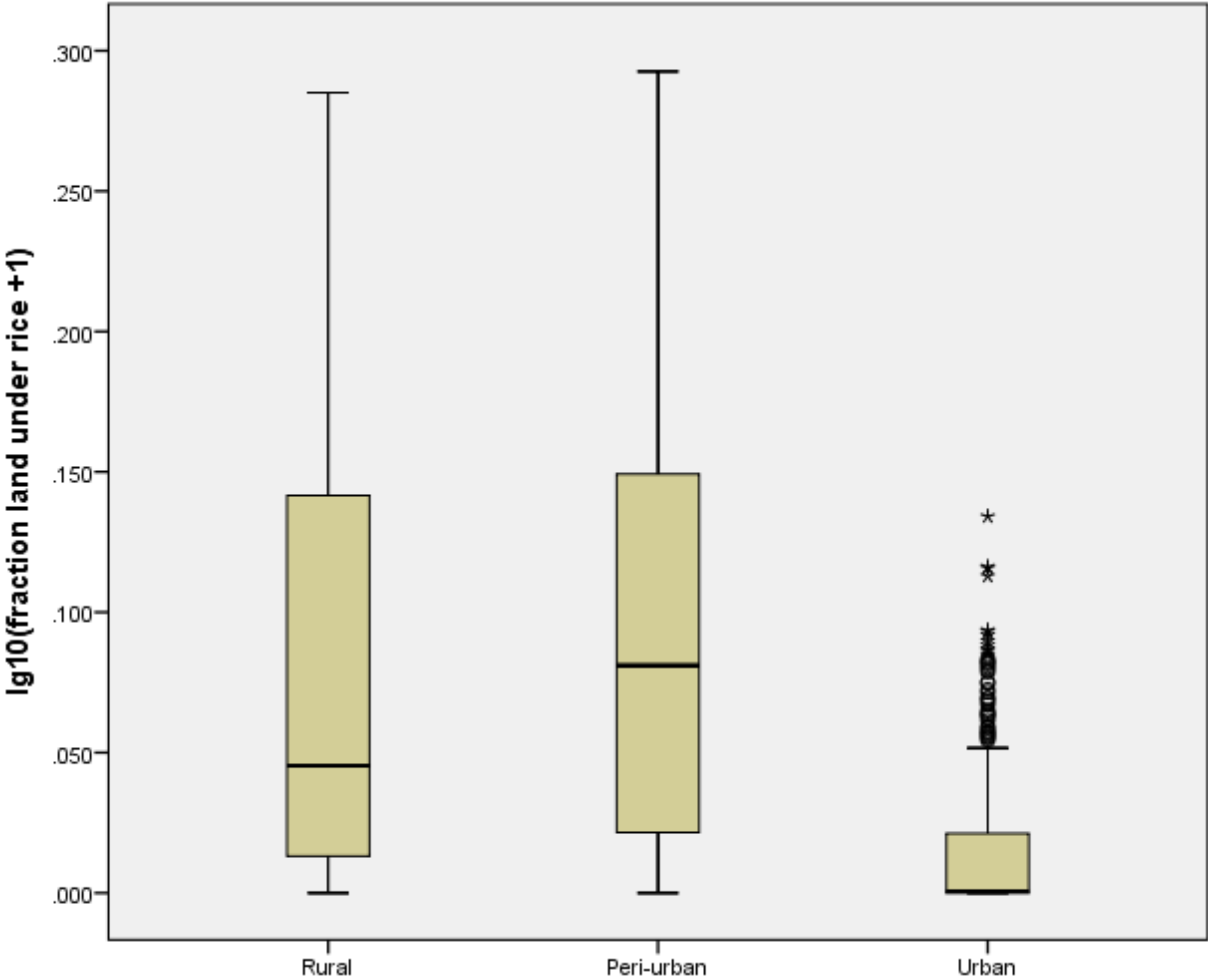**Figure 2: Variation of fraction of land under rice across urbanicity: Viet Nam**

Figure 3: Variation of fraction of land under aquaculture across urbanicity: Viet Nam
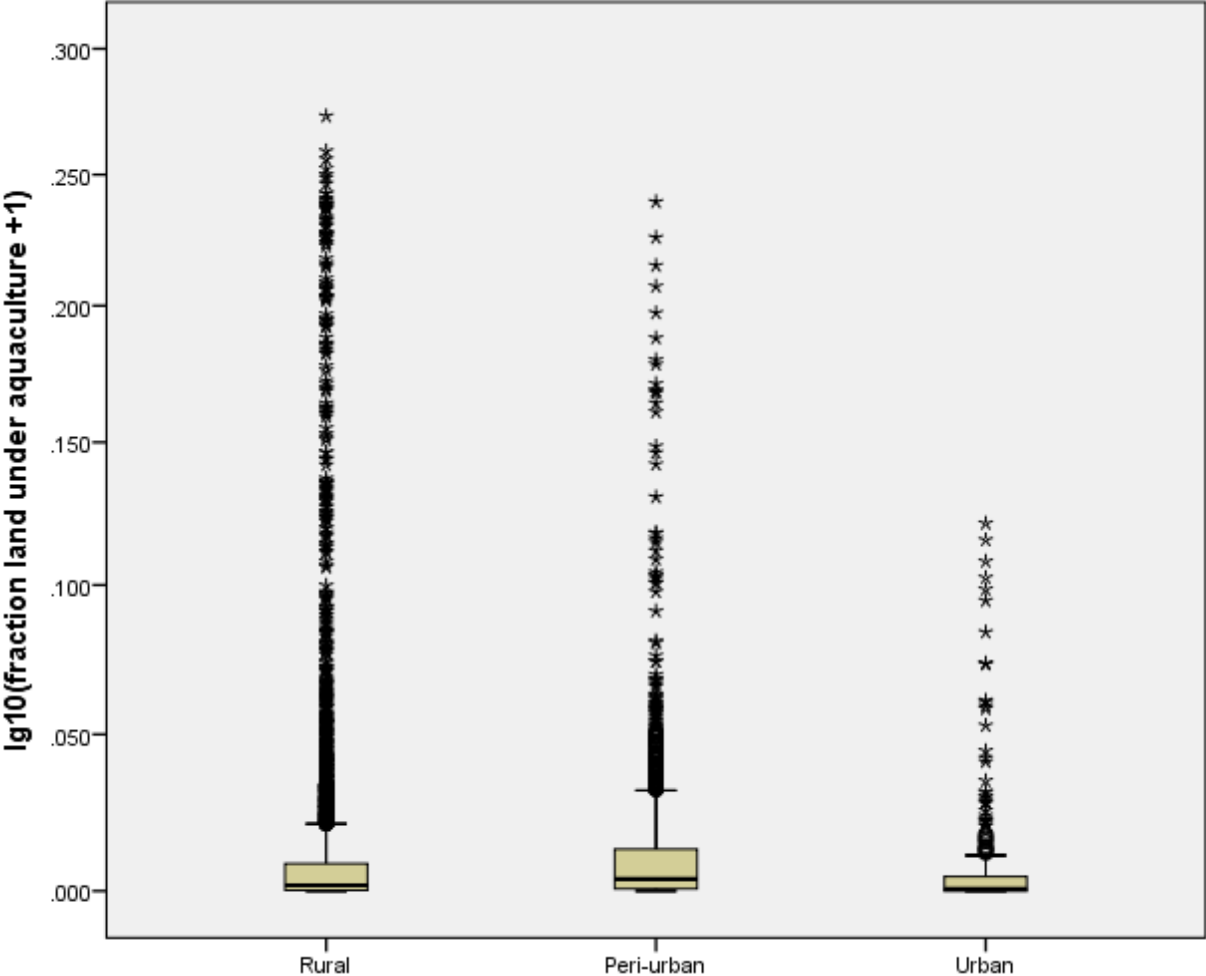
Fig 4: Variation of duck and geese flock size diversity across urbanicity: Viet Nam
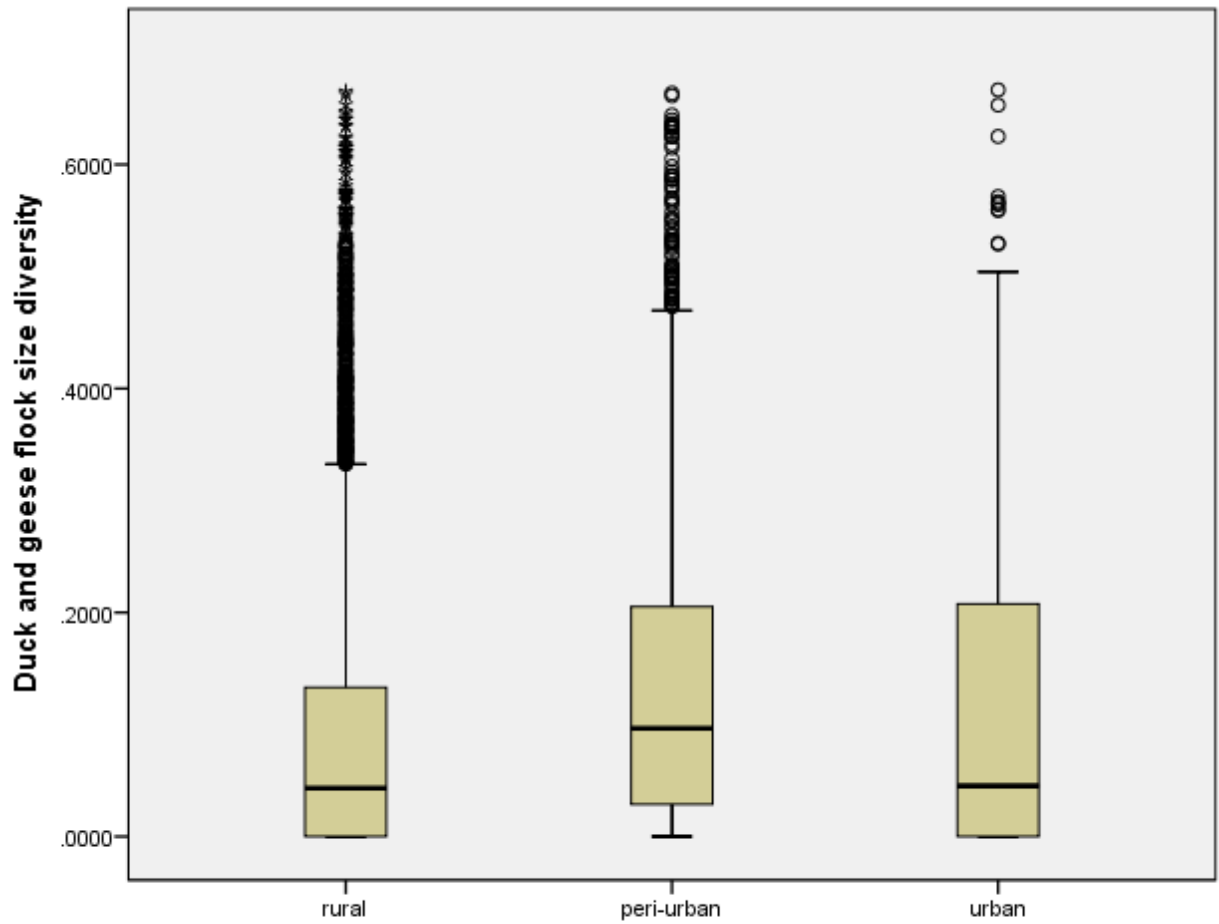
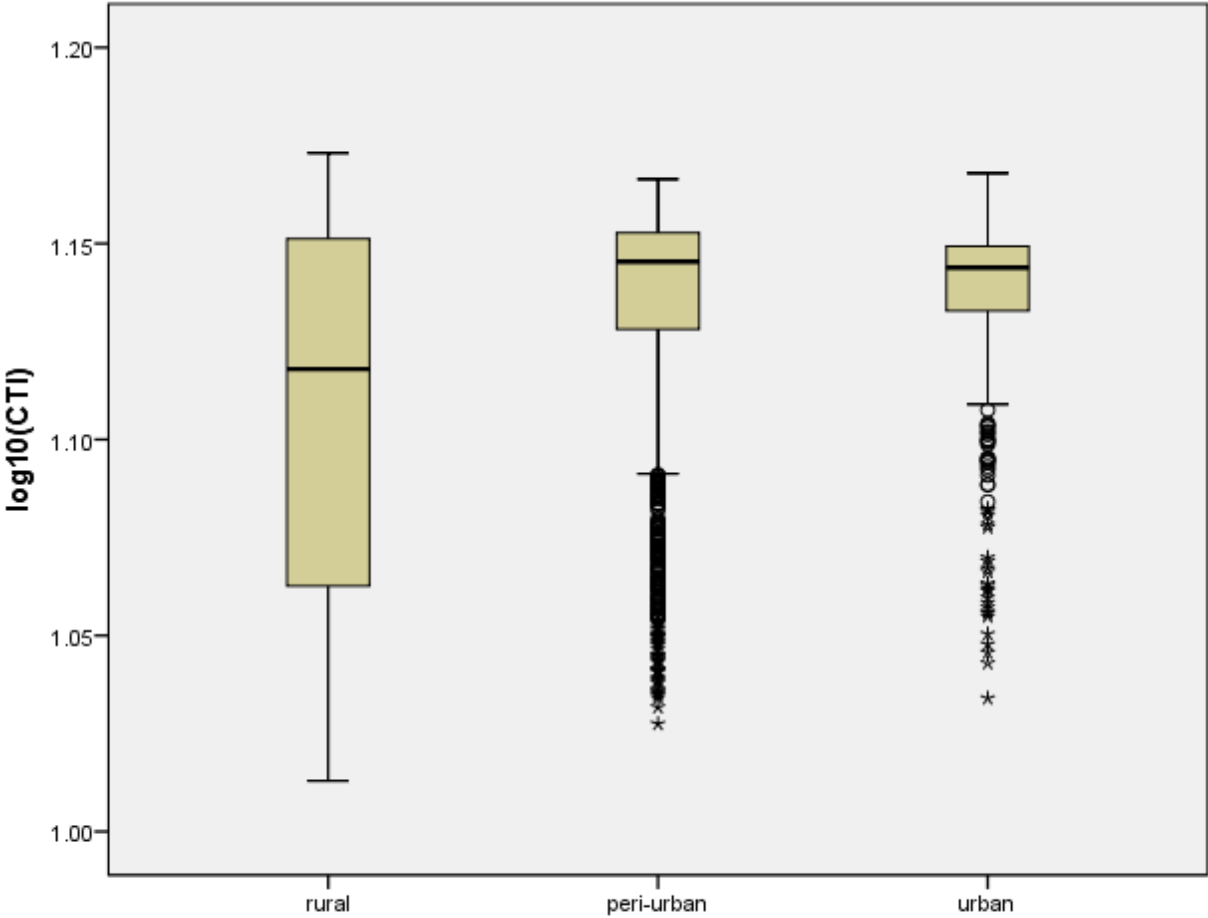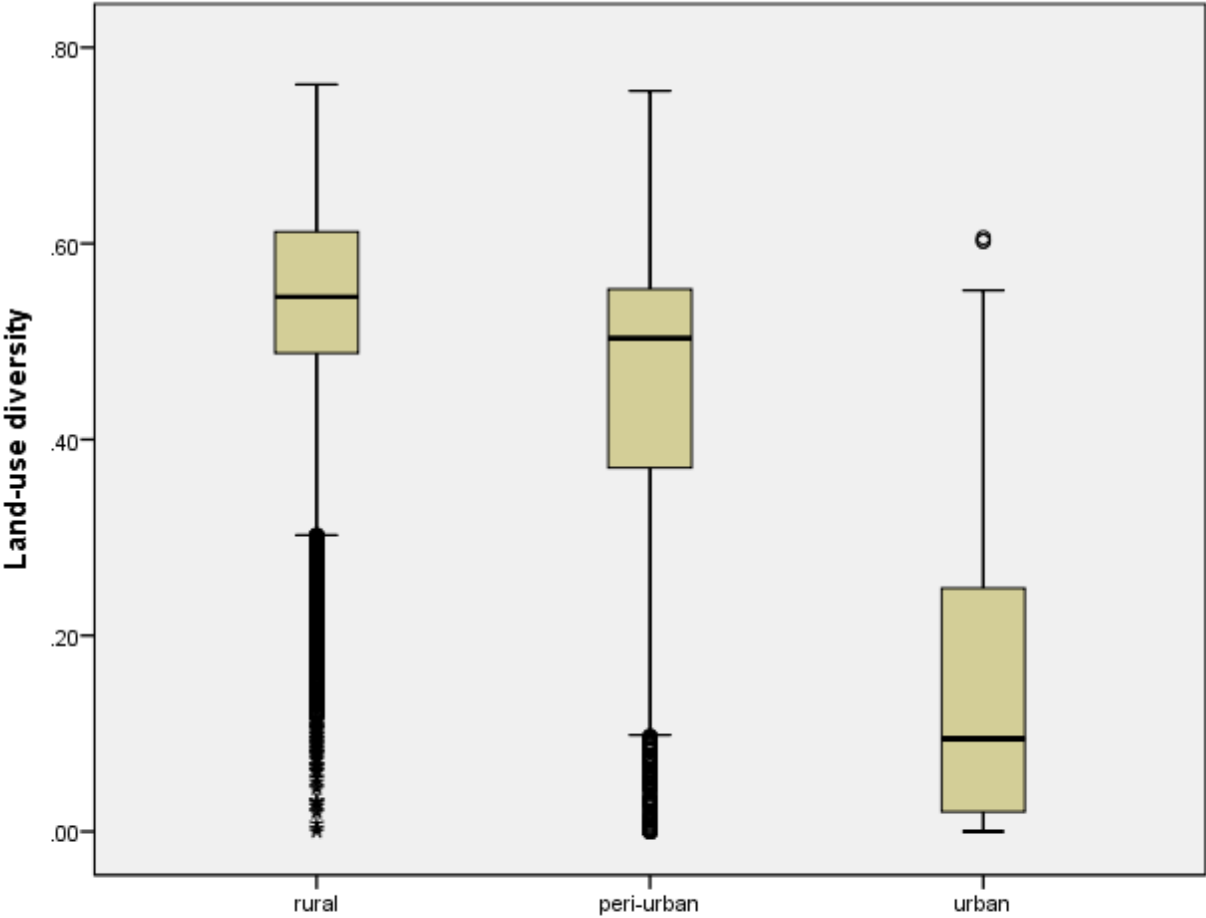**Fig 5: Variation of CTI across urbanicity: Viet Nam**

Fig 6: Variation of land-use diversity across urbanicity: Viet Nam

References

Adhikari, D., A. Chettri, et al. (2009). "Modelling the ecology and distribution of highly pathogenic avian influenza (H5N1) in the Indian subcontinent." <u>Current Science</u> **97**(1): 73-78.

Aoyama, Y. and R. Horner (2010). "World Development Report 2009: Reshaping Economic Geography." <u>Journal of Regional Science</u> **50**(5): 1001-1004.

Arcaya, M., M. Brewster, et al. (2012). "Area variations in health: a spatial multilevel modeling approach." <u>Health Place</u> **18**(4): 824-831.

Augustin, N. H., M. A. Mugglestone, et al. (1996). "An autologistic model for the spatial distribution of wildlife." <u>Journal of Applied Ecology</u> **33**(2): 339-347.

Birley, M. H. and K. Lock (1998). "Health and peri-urban natural resource production." <u>Environment and Urbanization</u> **10**(1): 89-106.

Biswas, P. K., J. P. Christensen, et al. (2009). "Risk for infection with highly pathogenic avian influenza virus (H5N1) in backyard chickens, Bangladesh." <u>Emerg Infect Dis</u> **15**(12): 1931-1936.

Burgos, S., P. T. Hong Hanh, et al. (2007). "Characterization of Poultry Production Systems in Vietnam " <u>International Journal of Poultry Science</u> **6**(10): 709-712.

Cao, C. X., M. Xu, et al. (2010). "Risk analysis for the highly pathogenic avian influenza in Mainland China using meta-modeling." <u>Chinese Science Bulletin</u> **55**(36): 4168-4178.

Despommier, D., B. R. Ellis, et al. (2006). "The role of ecotones in emerging infectious diseases." <u>Ecohealth</u> **3**(4): 281-289.

Ducheyne, E., C. Mweempwa, et al. (2009). "The impact of habitat fragmentation on tsetse abundance on the plateau of eastern Zambia." <u>Prev Vet Med</u> **91**(1): 11-18.

Eckert, S. and S. Kohler (2014). "Urbanization and health in developing countries: a systematic review." <u>World Health Popul</u> **15**(1): 7-20.

Elith, J., C. H. Graham, et al. (2006). "Novel methods improve prediction of species' distributions from occurrence data." <u>Ecography</u> **29**(2): 129-151.

Elith, J. and J. R. Leathwick (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. <u>Annual Review of Ecology Evolution and Systematics</u>. Palo Alto, Annual Reviews. **40:** 677-697.

Elith, J., J. R. Leathwick, et al. (2008). "A working guide to boosted regression trees." <u>J Anim Ecol</u> **77**(4): 802-813.

Fang, L. Q., S. J. de Vlas, et al. (2008). "Environmental factors contributing to the spread of H5N1 avian influenza in mainland China." <u>PLoS One</u> **3**(5): e2268.

Fang, L. Q., X. L. Li, et al. (2013). "Mapping spread and risk of avian influenza A (H7N9) in China." <u>Sci Rep</u> **3**: 2722.

Forman, R. T. T. (1995). <u>Land mosaics : the ecology of landscapes and regions</u>. Cambridge ; New York, Cambridge University Press.

Gessler, P. E., I. D. Moore, et al. (1995). "SOIL-LANDSCAPE MODELING AND SPATIAL PREDICTION OF SOIL ATTRIBUTES." <u>International Journal of Geographical Information Systems</u> **9**(4): 421-432.

Gilbert, M., P. Chaitaweesub, et al. (2006). "Free-grazing ducks and highly pathogenic avian influenza, Thailand." <u>Emerg Infect Dis</u> **12**(2): 227-234.

Gilbert, M., N. Golding, et al. (2014). "Predicting the risk of avian influenza A H7N9 infection in live-poultry markets across Asia." <u>Nat Commun</u> **5**: 4116.

Gilbert, M. and D. U. Pfeiffer (2012). "Risk factor modelling of the spatio-temporal patterns of highly pathogenic avian influenza (HPAIV) H5N1: a review." <u>Spat Spatiotemporal Epidemiol</u> **3**(3): 173-183.

Gilbert, M., X. Xiao, et al. (2008). "Mapping H5N1 highly pathogenic avian influenza risk in Southeast Asia." <u>Proc Natl Acad Sci U S A</u> **105**(12): 4769-4774.

Gottdenker, N. L., D. G. Streicker, et al. (2014). "Anthropogenic Land Use Change and Infectious Diseases: A Review of the Evidence." Ecohealth.

Graham, J., P. Gurian, et al. (2004). "Peri-urbanization and in-home environmental health risks: the side effects of planned and unplanned growth." Int J Hyg Environ Health **207**(5): 447-454.

Greiner, M., D. Pfeiffer, et al. (2000). "Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests." Prev Vet Med **45**(1-2): 23-41.

GSO. "Agriculture, Forestry and Fishery " Retrieved 1 October, 2014, 2014, from http://www.gso.gov.vn/default_en.aspx?tabid=469&idmid=3&ItemID=15369.

Guerrini, L., J. P. Bord, et al. (2008). "Fragmentation analysis for prediction of suitable habitat for vectors: example of riverine tsetse flies in Burkina Faso." J Med Entomol **45**(6): 1180-1186.

Hahn, M. B., E. S. Gurley, et al. (2014). "The role of landscape composition and configuration on Pteropus giganteus roosting ecology and Nipah virus spillover risk in Bangladesh." Am J Trop Med Hyg **90**(2): 247-255.

Hartemink, N., S. O. Vanwambeke, et al. (2011). "Integrated mapping of establishment risk for emerging vector-borne infections: a case study of canine leishmaniasis in southwest France." PLoS One **6**(8): e20817.

Henning, J., D. U. Pfeiffer, et al. (2009). "Risk factors and characteristics of H5N1 Highly Pathogenic Avian Influenza (HPAI) post-vaccination outbreaks." Vet Res **40**(3): 15.

Hijmans, R. J., S. E. Cameron, et al. (2005). "Very high resolution interpolated climate surfaces for global land areas." International Journal of Climatology **25**(15): 1965-1978.

Hill, M. O. (1973). "Diversity and evenness: A unifying notation and its consequences." Ecology **54**(2): 427.

Hogerwerf, L., R. G. Wallace, et al. (2010). "Persistence of highly pathogenic avian influenza H5N1 virus defined by agro-ecological niche." Ecohealth **7**(2): 213-225.

Horby, P. W., D. Pfeiffer, et al. (2013). "Prospects for emerging infections in East and southeast Asia 10 years after severe acute respiratory syndrome." Emerg Infect Dis **19**(6): 853-860.

Jones, B. A., D. Grace, et al. (2013). "Zoonosis emergence linked to agricultural intensification and environmental change." Proc Natl Acad Sci U S A **110**(21): 8399-8404.

Jones, K. E., N. G. Patel, et al. (2008). "Global trends in emerging infectious diseases." Nature **451**(7181): 990-993.

Kapan, D. D., S. N. Bennett, et al. (2007). "Avian influenza (H5N1) and the evolutionary and social ecology of infectious disease emergence (vol 3, pg 187, 2006)." Ecohealth **4**(2): 239-239.

Loth, L., M. Gilbert, et al. (2010). "Risk factors and clusters of Highly Pathogenic Avian Influenza H5N1 outbreaks in Bangladesh." Prev Vet Med **96**(1-2): 104-113.

Loth, L., M. Gilbert, et al. (2011). "Identifying risk factors of highly pathogenic avian influenza (H5N1 subtype) in Indonesia." Prev Vet Med **102**(1): 50-58.

Mallin, M. A. and L. B. Cahoon (2003). "Industrialized animal production - A major source of nutrient and microbial pollution to aquatic ecosystems." Population and Environment **24**(5): 369-385.

Martin, V., D. U. Pfeiffer, et al. (2011). "Spatial distribution and risk factors of highly pathogenic avian influenza (HPAI) H5N1 in China." PLoS Pathog **7**(3): e1001308.

McGee, T. G. (1991). The emergence of desakota in Asia: expanding a hypothesis. The extended metropolis: Settlement transition in Asia. N. Ginsburg, B. Koppel and T. G. McGee. Honolulu, University of Hawaii Press**:** pp:15-55.

McPherson, J. M., W. Jetz, et al. (2004). "The effects of species' range sizes on the accuracy of distribution models: ecological phenomenon or statistical artefact?" Journal of Applied Ecology **41**(5): 811-823.

Morse, S. S. (1995). "Factors in the emergence of infectious diseases." Emerg Infect Dis **1**(1): 7-15.

Murray, K. A. and P. Daszak (2013). "Human ecology in pathogenic landscapes: two hypotheses on how land use change drives viral emergence." Curr Opin Virol **3**(1): 79-83.

Oliveira, C. C., H. G. Lacerda, et al. (2004). "Changing epidemiology of American cutaneous leishmaniasis (ACL) in Brazil: a disease of the urban-rural interface." Acta Trop **90**(2): 155-162.

Paul, M., S. Tavornpanich, et al. (2010). "Anthropogenic factors and the risk of highly pathogenic avian influenza H5N1: prospects from a spatial-based model." Vet Res **41**(3): 28.

Paul, M., S. Wongnarkpet, et al. (2011). "Risk factors for highly pathogenic avian influenza (HPAI) H5N1 infection in backyard chicken farms, Thailand." Acta Trop **118**(3): 209-216.

Pebesma, E. J. (2004). "Multivariable geostatistics in S: the gstat package." Computers & Geosciences **30**(7): 683-691.

Pfeiffer, D. U., P. Q. Minh, et al. (2007). "An analysis of the spatial and temporal patterns of highly pathogenic avian influenza occurrence in Vietnam using national surveillance data." Vet J **174**(2): 302-309.

Philip, K. T. (2010). "Livestock production: recent trends, future prospects." Philosophical Transactions of the Royal Society B: Biological Sciences **365**(1554): 2853-2867.

Saksena, S., J. Fox, et al. (2014). "Classifying and mapping the urban transition in Vietnam." Applied Geography **50**: 80-89.

Sarkar, S., S. E. Strutz, et al. (2010). "Chagas disease risk in Texas." PLoS Negl Trop Dis **4**(10).

Si, Y., A. K. Skidmore, et al. (2009). "Spatio-temporal dynamics of global H5N1 outbreaks match bird migration patterns." Geospat Health **4**(1): 65-78.

Spencer, J. H. (2013). "The urban health transition hypothesis: empirical evidence of an avian influenza Kuznets curve in Vietnam?" J Urban Health **90**(2): 343-357.

Stevens, K. B., M. Gilbert, et al. (2013). "Modeling habitat suitability for occurrence of highly pathogenic avian influenza virus H5N1 in domestic poultry in Asia: a spatial multicriteria decision analysis approach." Spat Spatiotemporal Epidemiol **4**: 1-14.

Suzan, G., E. Marce, et al. (2008). "The effect of habitat fragmentation and species diversity loss on hantavirus prevalence in Panama." Ann N Y Acad Sci **1149**: 80-83.

Taylor, L. H., S. M. Latham, et al. (2001). "Risk factors for human disease emergence." Philos Trans R Soc Lond B Biol Sci **356**(1411): 983-989.

Thanapongtharm, W., T. P. Van Boeckel, et al. (2013). "Rivers and flooded areas identified by medium-resolution remote sensing improve risk prediction of the highly pathogenic avian influenza H5N1 in Thailand." Geospat Health **8**(1): 193-201.

Tiensin, T., S. S. Ahmed, et al. (2009). "Ecologic risk factor investigation of clusters of avian influenza A (H5N1) virus infection in Thailand." J Infect Dis **199**(12): 1735-1743.

Tran, C., R. Yost, et al. (2013). "Spatio-Temporal Occurrence Modeling of Highly Pathogenic Avian Influenza Subtype H5N1: A Case Study in the Red River Delta, Vietnam." ISPRS International Journal of Geo-Information **2**(4): 1106-1121.

Tran, P. M. and L. Waller (2013). "Effects of landscape fragmentation and climate on Lyme disease incidence in the northeastern United States." Ecohealth **10**(4): 394-404.

Van Boeckel, T. P., W. Thanapongtharm, 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**(11): e49528.

Vlahov, D. and S. Galea (2002). "Urbanization, urbanicity, and health." J Urban Health **79**(4 Suppl 1): S1-S12.

Walsh, M. G. (2013). "Forest fragmentation and risk of giardiasis in New York State." Ecohealth **10**(4): 405-414.

Ward, M. P., D. Maftei, et al. (2008). "Environmental and anthropogenic risk factors for highly pathogenic avian influenza subtype H5N1 outbreaks in Romania, 2005--2006." <u>Vet Res Commun</u> **32**(8): 627-634.

Webster, D. (2002). On the edge: Shaping the future of peri-urban East Asia. Stanford, CA, Asia-Pacific Research Center.

Wilcox, B. and R. Colwell (2005). "Emerging and Reemerging Infectious Diseases: Biocomplexity as an Interdisciplinary Paradigm." <u>Ecohealth</u> **2**(4): 244-257.

Wilcox, B. A. and D. J. Gubler (2005). "Disease ecology and the global emergence of zoonotic pathogens." <u>Environ Health Prev Med</u> **10**(5): 263-272.

Williams, R. A., F. O. Fasina, et al. (2008). "Predictable ecology and geography of avian influenza (H5N1) transmission in Nigeria and West Africa." <u>Trans R Soc Trop Med Hyg</u> **102**(5): 471-479.

Williams, R. A. and A. T. Peterson (2009). "Ecology and geography of avian influenza (HPAI H5N1) transmission in the Middle East and northeastern Africa." <u>Int J Health Geogr</u> **8**: 47.

Yupiana, Y., S. J. de Vlas, et al. (2010). "Risk factors of poultry outbreaks and human cases of H5N1 avian influenza virus infection in West Java Province, Indonesia." <u>Int J Infect Dis</u> **14**(9): e800-805.