



Modeling habitat suitability for occurrence of highly pathogenic avian influenza virus H5N1 in domestic poultry in Asia: A spatial multicriteria decision analysis approach

Kim B. Stevens^{a,*}, Marius Gilbert^{b,c}, Dirk U. Pfeiffer^a

^a *Veterinary Epidemiology and Public Health Group, Department of Veterinary Clinical Sciences, Royal Veterinary College, Hawkshead Lane, North Mymms, Hatfield, Hertfordshire AL9 7TA, UK*

^b *Biological Control and Spatial Ecology, Université Libre de Bruxelles, Brussels, Belgium*

^c *Fonds National de la Recherche Scientifique, Brussels, Belgium*

ARTICLE INFO

Article history:

Received 10 October 2011

Revised 26 October 2012

Accepted 7 November 2012

Available online 24 November 2012

Keywords:

Emerging infectious diseases

Influenza in birds

Poultry diseases

Risk factors

Knowledge-driven spatial modeling

Model validation

ABSTRACT

Risk maps are one of several sources used to inform risk-based disease surveillance and control systems, but their production can be hampered by lack of access to suitable disease data. In such situations, knowledge-driven spatial modeling methods are an alternative to data-driven approaches. This study used multicriteria decision analysis (MCDA) to identify areas in Asia suitable for the occurrence of highly pathogenic avian influenza virus (HPAIV) H5N1 in domestic poultry. Areas most suitable for H5N1 occurrence included Bangladesh, the southern tip and eastern coast of Vietnam, parts of north-central Thailand and large parts of eastern China. The predictive accuracy of the final model, as determined by the area under the receiver operating characteristic curve (ROC AUC), was 0.670 (95% CI 0.667–0.673) suggesting that, in data-scarce environments, MCDA provides a reasonable alternative to the data-driven approaches usually used to inform risk-based disease surveillance and control strategies.

Crown Copyright © 2012 Published by Elsevier Ltd. All rights reserved.

1. Introduction

Although the first reported outbreak of highly pathogenic avian influenza virus (HPAIV) H5N1 in Hong Kong in 1997 was successfully controlled through mass slaughter of all chickens (Li et al., 2004; Lipatov et al., 2004), since its re-emergence in 2003 the disease has become

widespread in Eastern and Southeastern Asia (2003–04) followed by incursions into Southern Asia (2005–06) resulting in the cumulative deaths of millions of domestic poultry and over 250 humans in the three regions (WHO, June 2011). Not surprisingly, the disease has been highlighted as one of the main animal disease threats in Asia and its control described as ‘a formidable challenge’ (FAO, 2011).

As disease control resources are generally limited, targeting of disease surveillance and control using risk-based methods allows for the optimization of these finite resources. Risk-based methods include estimating the level of disease risk, mapping the spatial distribution of risk and identifying important risk factors for disease (Pfeiffer et al., 2008). While the majority of HPAIV H5N1 studies have focused on the latter (Gilbert et al., 2006, 2007; Martin et al., 2006; Henning et al., 2009; Loth et al., 2010; Desvaux et al., 2011), relatively few have mapped

Abbreviations: AUC, area under the curve; AHP, analytical hierarchy process; BPA, basic probability assignment; DST, Dempster-Shafer theory; EMPRES-i, Emergency Prevention System Global Animal Disease Information System; GARP, genetic algorithm for rule-set production; HPAIV, highly pathogenic avian influenza virus, Maxent maximum entropy; MCDA, multicriteria decision analysis; MPA, minimum predicted area; ROC, receiver operating characteristic; WLC, weighted linear combination.

* Corresponding author.

E-mail addresses: kstevens@rvc.ac.uk (K.B. Stevens), mgilbert@ulb.ac.be (M. Gilbert), pfeiffer@rvc.ac.uk (D.U. Pfeiffer).

the spatial distribution of disease risk (Pfeiffer et al., 2007; Fang et al., 2008; Gilbert et al., 2008; Hogerwerf et al., 2010; Martin et al., 2011).

All these studies have used data-driven methods which can be limited by their need for detailed locational information about disease events that have occurred in a defined animal population at risk. However, such methods become less applicable in situations where disease event details are unavailable or if a country has not yet detected a disease incursion. In such data-scarce situations, an alternative approach could be to use a knowledge-driven modeling method, such as multicriteria decision analysis (MCDA), which uses current knowledge regarding the factors associated with disease occurrence as model inputs in order to identify areas suitable for disease occurrence. For example, the knowledge that a certain disease vector or host species only occurs near water, within a specific temperature range, and its preferred habitat is grassland can be used to identify areas satisfying all those criteria and therefore potentially suitable for the presence of the disease vector or host, and hence the disease. Thus, it is possible to generate maps showing areas suitable for disease, in the absence of data, provided the necessary knowledge is available. However, as with data-driven models, knowledge-driven models are only as good as their inputs, and therefore models of new or emerging diseases based on meager knowledge are likely to be less useful than maps of established diseases based upon a large body of research.

Despite the fact that the method is not restricted by the need for data, MCDA has been used infrequently in the animal health field; the authors are aware of only three studies that have used this method for modeling the spatial distribution of animal diseases (Clements et al., 2006, 2007; Rakotomanana et al., 2007). One reason for its lack of use may be that none of these studies performed a quantitative assessment of their model's predictive accuracy relying rather on visual appraisal of the agreement between actual disease presence and predicted risk, which may inspire less confidence in the method than modeling approaches involving empirical quantitative assessment. We therefore felt that it was necessary to perform a quantitative evaluation of an MCDA model to allow for future comparisons of the method's predictive accuracy with those of the more frequently used data-driven approaches.

Traditional methods for quantitatively evaluating the predictive accuracy of maps, such as the area under the receiver operating characteristic curve (ROC AUC) (Greiner et al., 2000), require the use of both disease presence and absence data thus ostensibly precluding their use with knowledge-driven models. However, quantitative assessments of model validity modeling can also be conducted using disease presence data and a random sample of background data (Boyce et al., 2002; Franklin, 2009a) which provide a representative sample of the conditions available in the study area (Engler et al., 2004; Chefaoui and Lobo, 2008; VanDerWal et al., 2009; Wisz and Guisan, 2009) and discriminates used from available habitat. Suitability models based on disease presence and background data therefore model occupied versus available habitat. The

use of background data instead of disease absence data opens up possibilities for quantitatively evaluating MCDA models using disease-presence data from national or international databases such as the Food and Agricultural Organization of the United Nation's (FAO) Emergency Prevention System Global Animal Disease Information System (EMPRES-i) and, if the models prove to have a reasonable predictive accuracy, for the increased use of MCDA as a method for confidently modeling the spatial distribution of disease, especially in data-scarce situations.

Although MCDA is most useful in data-scarce situations, such circumstances do not allow for a robust estimation of the model's predictive accuracy; instead, such an objective calls for the use of a well-documented disease. As HPAIV H5N1 is one in the most well-documented animal diseases and, in Asia, occurs under widely differing socio-economic and agro-ecological conditions, it was chosen as an appropriate disease for assessing the predictive accuracy of the MCDA approach under a variety of conditions. Therefore, the objectives of this study were (i) to identify areas in Southern, Eastern and Southeastern Asia suitable for HPAIV H5N1 occurrence in domestic poultry using MCDA and (ii) to evaluate the predictive accuracy of the resulting model using quantitative methods of model validation in order to determine whether, in data-scarce situations, MCDA is able to provide a reasonable alternative to the data-driven approaches traditionally used to inform risk-based surveillance and disease control strategies.

2. Materials and methods

2.1. The MCDA method

The process of spatial MCDA is described in detail elsewhere (Malczewski, 1999, 2000; Pfeiffer et al., 2008; Hongoh et al., 2011). Briefly, the approach involves identifying risk factors associated with the outcome (in this instance, occurrence of HPAIV H5N1 in domestic poultry), based on existing or hypothesized knowledge of their relationship with the outcome, and which are available in georeferenced format. These spatial data layers are manipulated to produce appropriate spatial risk factor data layers for inclusion in the model and standardized on a common, continuous scale (IDRISI GIS and Image Processing software uses the scale 0–255). The outcome will always be scaled across the whole range of values (i.e. 0–255) and this can be done using either linear transformations, or preferably fuzzy logic (Malczewski, 1999, 2000), whereby uncertainty regarding the association between risk factor and outcome is modeled using a fuzzy membership function to describe the shape of the known or hypothesized relationship between each risk factor and the outcome. If only two risk factors of equal importance are being modeled they can simply be combined using the Boolean operators AND or OR. However, if the model includes two or more risk factors of unequal importance, weights can be applied to each risk factor, using a method such as weighted linear combination (WLC), so that the more important risk factors exert a greater influence on the

outcome. A final suitability estimate for each spatial unit (i.e. raster cell) is calculated using the equation:

$$S = \sum_{ij=1}^n (W_j RF_{ij}) \quad (1)$$

where S is the final suitability estimate for each raster cell, W is the weight for risk factor j and RF is the value of risk factor j for raster cell i . Suitability of a raster cell is expressed on a continuous scale (usually 0–255) with the lowest value indicating cells least suitable for the outcome and the highest value representing those cells most suitable. However, the final suitability of a raster cell will be relative to the range of suitability scores for that particular dataset.

2.2. Definitions

For the purpose of this study 'suitability' is defined as 'the ability of a habitat to support a species' (Franklin, 2009b); in this instance the species being the H5N1 virus. Disease 'occurrence' refers to the presence of HPAIV H5N1 without defining either incidence or prevalence.

2.3. Study area

The study area included all countries in Southern, Eastern and Southeastern Asia as defined by the United Nations Geoscheme (<http://millenniumindicators.un.org/unsd/methods/m49/m49regin.htm>; accessed March 2010).

2.4. Defining continental-level risk factors, constraints and their relationship with HPAIV H5N1 occurrence

The systematic review published by (Gilbert and Pfeiffer, 2012) which 'focused on studies that have studied explicitly the spatial distribution of HPAI H5N1 based on empirical data', was used to identify continental-level risk factors associated with occurrence of HPAIV H5N1 in domestic poultry in Asia, and to define the relationship between individual risk factors and disease occurrence. Briefly, a keyword-based search of the ISI Web of Science and CAB Abstracts databases from January 1996 to December 2010 identified papers with explicit reference to spatial data on avian influenza. These references were then grouped into six categories based on the type of study performed but only papers from the first category – those focusing primarily on the statistical analysis of HPAIV disease and infection data – were included in the review. In addition, selection of papers was restricted to those dealing with disease and infection caused by HPAIV subtype H5N1. Subsequently, the types of variables considered in the different studies were grouped into nine, broad categories namely, (i) farming practice and local biosecurity, (ii) poultry and livestock census data, (iii) anthropogenic variables, (iv) socio-economic variables, (v) variables indicative of the presence or abundance of wild birds, (vi) variables indicative of the presence or abundance of rivers, lakes or wetlands, (vii) eco-climatic variables, (viii) land-use and cropping variables, and (ix) topographic variables.

Variables were then selected from these categories for inclusion in the MCDA model provided they met the following criteria: (i) able to be mapped and (ii) the relevant spatial layer was available in the public domain at a sufficiently high resolution to differentiate within-country heterogeneity. In addition, the variables had to (iii) reflect broad causal relationships at a continental scale rather than be country-specific (i.e. have been identified to be of importance in more than one country), and to have been (iv) repeatedly identified to be (v) significantly ($p \leq 0.05$) associated with HPAIV H5N1 occurrence. The following six variables met all these criteria.

- (i) *Domestic waterfowl density (heads/km²; WfowlDen)*: Increased domestic waterfowl density is strongly associated with occurrence of HPAIV H5N1 (Gilbert et al., 2006, 2008; Pfeiffer et al., 2007; Minh et al., 2009; Martin et al., 2011) as the birds act as a reservoir for the disease. Owing to the highly right-skewed distribution of this variable together with its wide range of values we assumed a sigmoidal, monotonically increasing relationship between suitability for HPAIV H5N1 occurrence and domestic waterfowl density between 0 and 1000 heads/km² with constant risk thereafter.
- (ii) *Chicken density (heads/km²; ChickDen)*: Although chicken density has generally been shown to have only a weak, positive association with HPAIV H5N1 occurrence (Pfeiffer et al., 2007; Gilbert et al., 2008; Ward et al., 2008; Martin et al., 2011) it was included in the model as chicken production appears to drive the spatial expansion phase of HPAIV H5N1 (Slingenbergh and Gilbert, 2008). We assumed a quadratic relationship between chicken density and suitability for HPAIV H5N1 occurrence with highest risk associated with medium density of chickens (500–5000 heads/km²; semi-commercial/backyard producers (Tiensin et al., 2005, 2007)), and lowest risk associated with both low (0–500 heads/km²; subsistence farmers with few birds) and high (>5000 heads/km²; commercial producers with good biosecurity) chicken densities.
- (iii) *Human population density (heads/km²; PopDen)*: Increasing human population density has repeatedly been shown to be associated with HPAIV H5N1 (Pfeiffer et al., 2007; Gilbert et al., 2008; Martin et al., 2011) and it is hypothesized that population density is a proxy for factors such as trade-associated movement of domestic poultry and fomites. Different studies have shown the relationship between population density and HPAIV H5N1 occurrence to be positive linear (Paul et al., 2010; Martin et al., 2011) or quadratic (Loth et al., 2010). For the model we assumed a positive linear relationship (Paul et al., 2010; Martin et al., 2011) between population density and suitability for HPAIV H5N1 occurrence.
- (iv) *Proximity to open water (km; ProxWater)*: Proximity to open water and density of waterways have been shown to be associated with HPAIV H5N1 occurrence (Ward et al., 2008; Biswas et al., 2009b; Martin

et al., 2011) as open water and wetlands can be used as resting places by migratory birds resulting in H5N1 occurrence through the presence of infected wild waterfowl. As outbreaks of H5N1 in domestic poultry have been shown to generally occur within 5 km of open water (Ward et al., 2008) we hypothesized that suitability for HPAIV H5N1 occurrence in domestic poultry was highest closest to open water (0–5 km) and thereafter decreased in a sigmoidal, monotonic fashion with negligible risk after 10 km (Fang et al., 2008). The hypothesized mechanism behind this is that domestic poultry close to open water are more likely to come into contact with infected water or wild waterfowl, and that this risk decreases with decreasing proximity to open water.

- (v) *Proximity to areas suitable for rice-growing (km; Prox-Rice)*: The percentage of land used for rice has been shown to be associated with HPAIV H5N1 occurrence (Pfeiffer et al., 2007; Gilbert et al., 2008). We assumed that proximity to areas suitable for rice-growing would have a similar relationship with suitability for HPAIV H5N1 occurrence as proximity to open water and therefore hypothesized that risk was highest closest to areas suitable for rice-growing (0–5 km) (Ward et al., 2008) and thereafter decreased in a sigmoidal, monotonic fashion with negligible risk after 10 km (Fang et al., 2008).
- (vi) *Proximity to roads (km; ProxRoad)*: Density of roads (Loth et al., 2010; Yupiana et al., 2010) and distance to main roads (Biswas et al., 2009b) have been shown to increase the suitability for HPAIV H5N1 occurrence as roads act as major arteries for trade and movement of infected poultry or fomites. As outbreaks in domestic poultry have been shown to occur mainly within 5 km (Ward et al., 2008), but seldom further than 60 km, from a road we assumed that risk of HPAIV H5N1 occurrence followed a sigmoidal, monotonically decreasing relationship with greatest risk within 0–5 km of a road, decreasing risk thereafter and negligible risk after 60 km.

Two models were run; one incorporating all raster cells in the study area (Unconstrained Model) and one in which the area was constrained to exclude raster cells considered unsuitable for poultry production as defined by (Wint and Robinson, 2007) (Constrained Model) using a binary spatial data layer (suitable/unsuitable) derived from Gridded Livestock of the World (Wint and Robinson, 2007).

2.5. Data collection and generation of spatial risk factor data layers

The necessary digital spatial data layers were sourced primarily from the public domain. Human population density was obtained from Gridded Population of the World v3 (Center for International Earth Science Information Network (CIESIN), 2005). Chicken and domestic waterfowl densities for the study area were extracted from spatial data layers created by (Prosser et al., 2011) and (Van Boeckel et al., 2011). Location of open water was extracted from VMap0 Perennial Water Courses (Rivers) of the World

and the location of primary and secondary roads was extracted from VMap0 Roads of the World, both available from the Food and Agricultural Organization's (FAO) GeoNetwork website (<http://www.fao.org/geonetwork/srv/en/main.home>). Location of areas suitable for rice growing was extracted from Suitability for Rain-fed and Irrigated Rice (High Input) available from FAO's GeoNetwork website (<http://www.fao.org/geonetwork/srv/en/main.home>). Only those areas with suitability ratings of very high, high, good and medium were extracted as we assumed that these would be the areas used for intensive rice production, which has been shown to be associated with HPAIV H5N1 occurrence (Pfeiffer et al., 2007; Gilbert et al., 2008; Paul et al., 2010).

For the three 'proximity to' spatial data layers, raster layers showing Euclidean distance to the feature of interest were generated for inclusion in the model and, owing to computational limitations imposed by mapping at the continental level, all raster layers were resampled to a resolution of 5 km² with the attribute having the largest share of a cell's area assigned as the resampled cell's attribute value. Spatial data layers with clearly defined attributes (e.g. Euclidean distance to a feature) are known as *crisp* sets, as opposed to *fuzzy* sets which indicate the hypothesized strength of the association between different values of the risk factor and the outcome (Eastman, 2009). Crisp sets were converted to fuzzy sets using the fuzzy membership functions detailed in Section 2.4. The scale of all spatial risk factor data layers was continuous and positively correlated with suitability for HPAIV H5N1 occurrence. Crisp sets were produced in ArcGIS 10 (ESRI, Redlands, CA, USA) while IDRISI Andes (Clark Labs, Clark University, Worcester MA, USA) was used to create the fuzzy sets and run the MCDA model.

Locational point data on all confirmed HPAIV H5N1 outbreaks in domestic poultry in Asia between January 2004 and October 2010 ($n = 10,104$) were extracted from the EMPRES-i database. These data are compiled from numerous sources including reports from FAO and the World Organization for Animal Health (OIE). We assume that these reports are not subject to false-positives but recognize that there will be an element of under-reporting (i.e. false-negative locations) although this is likely to be spatially heterogeneous. All duplicate locations were removed from the dataset so that any disease location was represented only once, and when more than one disease point occurred in a 5 km² raster cell, only one was retained leaving a total of 3690 disease points.

2.6. Generation of weights and creation of the final suitability maps

Weights were generated for each factor using the pairwise comparison matrix of the analytical hierarchy process (AHP) (Saaty, 1987; Saaty, 1990; Malczewski, 1999). With this method, for each pair of risk factors, modelers are required firstly to specify whether risk factor A (for example) is more or less important than risk factor B (for example) with regards the outcome and secondly, to specify the degree of importance on a nine-point scale ranging from

extremely less important, through equal importance to extremely more important (Malczewski, 1999).

Pairs of risk factors were weighted based on (i) the number of times they had been reported in the literature as significant risk factors and (ii) the significance levels they achieved in the different studies (Table 1). Level of significance achieved by the risk factors in previous studies was used as an indicator of the importance of the risk factor to disease occurrence because the lower the *p*-value the stronger the statistical evidence required to prove that the result could not have happened by chance, suggesting that the lower the *p*-value the stronger the association between the risk factor and outcome. Thus, those which had been reported more frequently and which achieved a higher level of significance were considered to be more important with respect to disease occurrence than risk factors which were reported less often or achieved lower *p*-values. The principal eigenvector of the pairwise comparison matrix was then calculated to produce a best-fit set of weights (Malczewski, 1999).

The Pearson product-moment correlation coefficient was determined for each pair of factors and where this was greater than 0.4 the weights of both factors were decreased by 10%. In the absence of any statistical evidence upon which to base the amount by which weights of correlated factors were decreased, we experimented with a range of percentages (5–30%), finally selecting 10% as the amount which provided a sufficient change to the weights yet without allowing any of the altered risk factors to dominate the model. Once the necessary pairs of risk factors had been adjusted in this way, the weights of the unadjusted factors were increased, in proportion with their original ratio, so that all weights summed to one. The weights were combined with the constraint and spatial risk factor data layers using WLC to determine the

suitability of each 5 km² raster cell for HPAIV H5N1 occurrence. Suitability was expressed on a continuous scale ranging from 0 (lowest suitability) to 255 (highest suitability), although this is relative to the range of suitability scores for each dataset. IDRISI's RANK module, which applies a simple choice heuristic, was used to rank-order all raster cells and the resulting map was divided by the maximum rank thereby generating a map showing the probability of cells being suitable for HPAIV H5N1 occurrence, relative to the highest suitability ranking (Eastman, 1997; Malczewski, 2000).

2.7. Map validation

The suitability maps were validated against the EMPRES-i HPAIV H5N1 outbreak data specified in Section 2.5, using the ROC AUC. As this evaluation method requires data representing both disease presence and absence, the EMPRES-i data were used together with 10 000 background data points. These data were randomly generated within the constraints specified below, to represent the agro-ecological niche of areas without HPAIV H5N1 outbreaks in domestic poultry. Ten thousand points were considered sufficient to represent all environmental conditions in the study area and were generated within a constrained area based on three criteria: (i) being >5 km from a disease presence point (i.e. presence and background points could not occur in the same cell), (ii) being >5 km from another background point to ensure that there was never more than one background point per cell and (iii) having a PopDen ≥ 1 head/km² (i.e. to exclude unpopulated desert and mountainous areas which would be unsuitable for poultry production).

Incremental probability cut-off values were used to create multiple binary spatial data layers – cells ranked

Table 1

Selected risk factors associated with highly pathogenic avian influenza (HPAI) H5N1 outbreaks in domestic poultry and the significance level achieved by each in the respective studies. Plus (+) and minus (–) indicates positive and negative associations respectively between the risk factor and HPAI while +++/–––, ++/––, and +/- refer to significance of the risk factor at *p*-levels of <0.001, <0.01 and <0.05, respectively. Sign/significance, country and spatial unit are presented in the same sequence as the corresponding references.

Risk factor	Sign/Significance	Country	Spatial unit	References
Domestic waterfowl density	+++ , + , +++ , +++ , +	Thailand, South Asia, Thailand, Vietnam, Thailand	All AdminL3	Gilbert et al. (2006), Gilbert et al. (2008), Paul et al. (2010), Pfeiffer et al. (2007) and Tiensin et al. (2009)
Human population density	+++ , ++ , ++/ –– , +	Thailand, South Asia, Bangladesh, Thailand	All AdminL3	Gilbert et al. (2006, 2008), Loth et al. (2010) and Tiensin et al. (2009)
Chicken density	+ , +/-	Global, Vietnam	Country, AdminL3	Hogerwerf et al. (2010) and Pfeiffer et al. (2007)
<i>Roads</i>				
Road density, presence of a road, road length, minimum distance to highway	++ , ++ , – – – , +	Bangladesh, Thailand, Romania, China, Thailand	AdminL3, AdminL3, cases, AdminL3, AdminL3	Loth et al. (2010), Paul et al. (2010), Ward et al. (2008) and Fang et al. (2008)
<i>Water</i>				
Minimum distance to lake*minimum distance to wetland, number of rivers and streams, access to water	– – – , ++ , +	China, Romania, Bangladesh	AdminL3, cases, farm	Fang et al. (2008), Ward et al. (2008) and Biswas et al. (2009a,b)
<i>Crops</i>				
Cropping intensity, % land used for rice	+ , +++ , +	South Asia, Thailand, Vietnam	All AdminL3	Gilbert et al. (2008), Paul et al. (2010) and Pfeiffer et al. (2007)

suitable or unsuitable – from the continuously ranked suitability map. Each of these spatial data layers was overlaid with the HPAIV H5N1 occurrence and background data and their sensitivity and specificity calculated and plotted as a ROC curve (sensitivity versus 1-specificity), and the ROC AUC calculated.

2.8. Sensitivity analysis

A sensitivity analysis was conducted to determine how changes in the weights and membership functions applied to each risk factor affected the final suitability estimate. To determine the effect of a change in membership function the model was run assuming a linear relationship between all factors and risk of HPAIV H5N1 occurrence. Similarly, to determine the effect of a change in the weights applied to each risk factor, two new weights were calculated for each risk factor by adding and subtracting 25% from the original weight. In the absence of any statistical evidence upon which to base the amount by which the weights were altered, 25% was selected as it provided a wide range of uncertainty while simultaneously ensuring that the underlying model structure was maintained. Each of the newly calculated weights was individually incorporated into the MCDA model, while holding all other factor weights constant. In addition, equal weights were applied to each risk factor and the model re-run. Suitability estimates were extracted for all disease presence points from the original and new maps, and the mean change in the suitability estimate as a result of altering the weight or membership function was determined, together with the standard deviation.

2.9. Application of Dempster-Shafer theory (DST)

Dempster-Shafer theory (DST) is an extension of Bayesian probability theory and is explained in detail by Eastman (Eastman, 2009). Unlike Bayes, DST assumes the existence of ignorance in the body of knowledge and that absence of evidence in support of an hypothesis (in this instance, the suitability of a raster cell for HPAIV H5N1

occurrence) does not implicitly imply the presence of evidence in support of the alternative hypothesis (in this instance, the unsuitability of a raster cell for HPAIV H5N1 occurrence). DST recognizes six important concepts; basic probability assignment (BPA), ignorance, belief, disbelief, plausibility and belief interval. BPA represents the support that a piece of evidence provides for the hypothesis and can be obtained from knowledge or data. Belief represents the total support for the hypothesis while plausibility represents the degree to which the hypothesis cannot be disbelieved. In other words, belief signifies the degree of hard evidence in support of the hypothesis while plausibility denotes the degree to which the conditions appear to be right for it even though hard evidence is lacking. The difference between belief and plausibility is the belief interval and represents the level of uncertainty surrounding – in this instance – the suitability or unsuitability of a raster cell for HPAIV H5N1 occurrence.

As BPAs are fuzzy sets (Eastman, 2009), the spatial data layers resulting from the application of the membership functions described in Section 2.4, were used in the application of DST after first converting them to a scale of 0–1. The two risk factors that relate directly to the disease host and vector – chicken and domestic waterfowl densities – were assumed to provide hard evidence for the occurrence of HPAIV H5N1 while the remaining four risk factors were assumed to provide soft evidence for disease occurrence as they were considered proxies for transmission mechanisms.

3. Results

3.1. Weights

The pairwise comparison matrix for the AHP is detailed in Table 2, correlations between pairs of risk factors are shown in Table 3 and final adjusted weights presented in Table 4. Domestic waterfowl and population density were the highest weighted risk factors while chicken density

Table 2

Pairwise comparison matrix of the analytical hierarchy process (AHP) for risk factors associated with highly pathogenic avian influenza virus (HPAIV) H5N1 based on the significance level achieved by the different risk factors as detailed in Table 1.

	WfowlDen	PopDen	ProxRoad	ProxWater	ChickDen	ProxRice	Weight
WfowlDen	1						0.3768
PopDen	1/2	1					0.2472
ProxRoad	1/3	1/2	1				0.1574
ProxWater	1/4	1/3	1/2	1			0.1149
ProxRice	1/5	1/4	1/3	1/3	1		0.0652
ChickDen	1/6	1/5	1/4	1/4	1/3	1	0.0385

Table 3

Pearson product-moment correlation coefficient for pairs of risk factors associated with occurrence of highly pathogenic avian influenza virus H5N1.

	WfowlDen	PopDen	ProxRoad	ProxWater	ProxRice	ChickDen
WfowlDen	1	0.25	-0.08	-0.03	-0.20	0.34
PopDen		1	-0.33	0.26	-0.55	0.55
ProxRoad			1	-0.09	0.27	-0.24
ProxWater				1	-0.18	0.09
ProxRice					1	-0.47
ChickDen						1

Table 4

Risk-factor weights resulting from the analytical hierarchy process (AHP) and following adjustment for correlations between risk factors.

Risk factor	Weight		Adjustment
	Original	Adjusted for correlation	
WfowlDen	0.3768	0.4281	
PopDen	0.2472	0.1978	Adjusted ↓ for correlation with ProxRice (–10%) and ChickDen (–10%)
ProxRoad	0.1574	0.1767	
ProxRice	0.1149	0.0919	Adjusted ↓ for correlation with PopDen (–10%) and ChickDen (–10%)
ProxWater	0.0652	0.0748	
ChickDen	0.0385	0.0307	Adjusted ↓ for correlation with ProxRice (–10%) and PopDen (–10%)

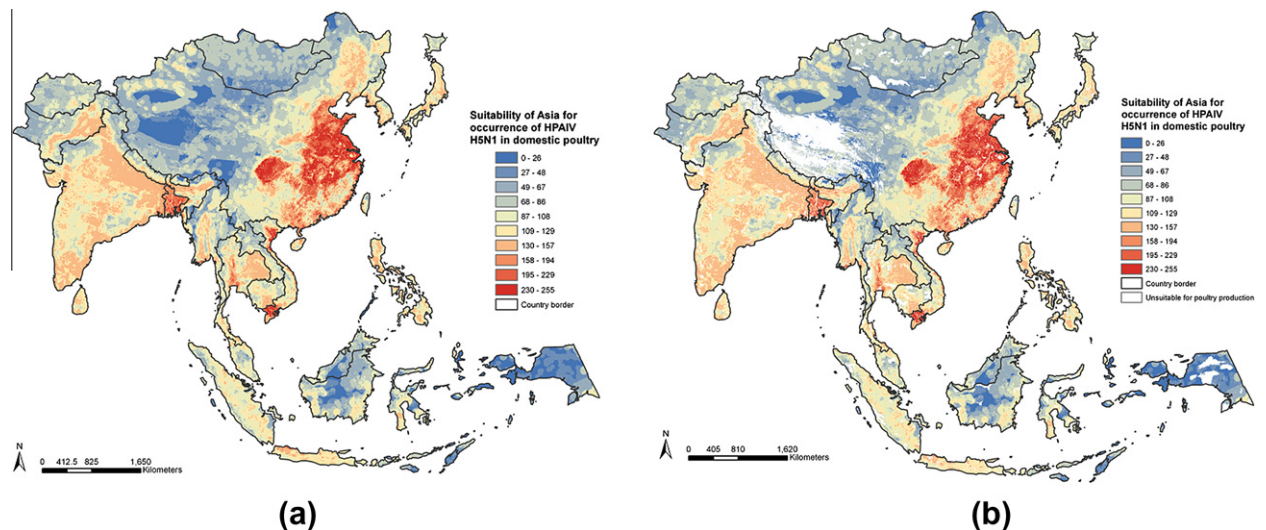


Fig. 1. Map illustrating the suitability of Asia for occurrence of highly pathogenic avian influenza virus (HPAIV) H5N1 in domestic poultry on a continuous scale from least to most suitable, as defined by multicriteria decision analysis. (a) Area unconstrained by suitability for poultry production and (b) constrained to areas considered suitable for poultry production by (Wint and Robinson, 2007).

received the lowest weighting. Chicken density was moderately correlated with population density ($r = 0.55$) while proximity to areas suitable for rice-growing was moderately correlated with population ($r = -0.55$) and chicken density ($r = -0.47$).

3.2. Suitability of Asia for occurrence of HPAIV H5N1 in domestic poultry

The suitability of Asia for occurrence of HPAIV H5N1 in domestic poultry was displayed on a continuous scale of

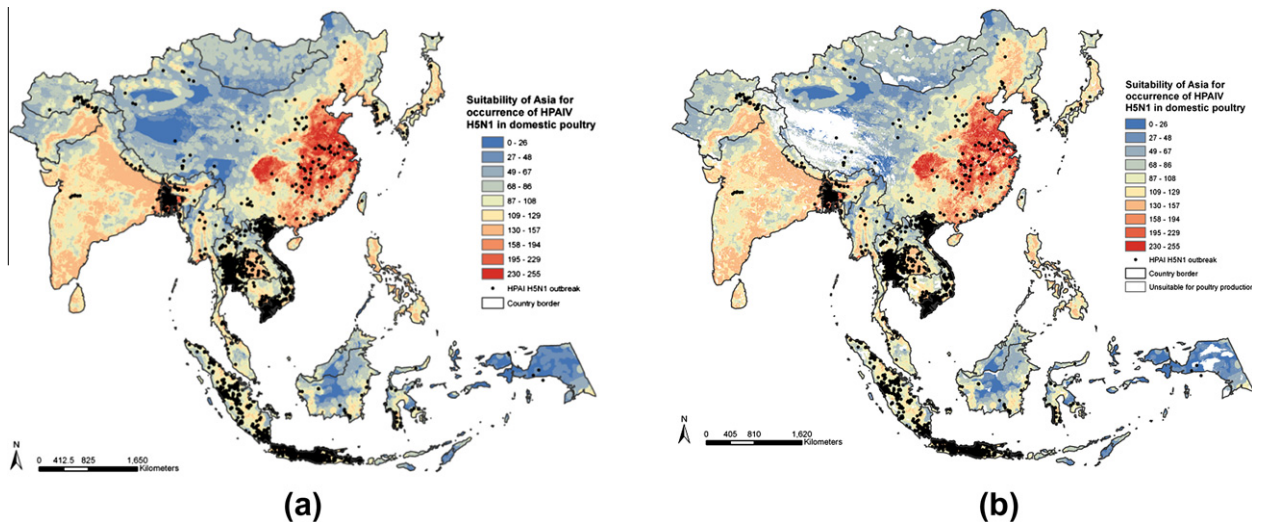


Fig. 2. Map illustrating the suitability of Asia for the occurrence of highly pathogenic avian influenza virus (HPAIV) H5N1 in domestic poultry overlaid with the location of confirmed HPAIV H5N1 outbreaks in domestic poultry between January 2004 and October 2010. (a) Area unconstrained by suitability for poultry production and (b) constrained to areas considered suitable by (Wint and Robinson, 2007) for poultry production.

least to most suitable represented by a blue–red color scale (Fig. 1). Dark-red areas were considered highly suitable for HPAIV H5N1 occurrence, light-red areas moderately suitable while light- and dark-blue indicated areas of low, and very low suitability, respectively. Areas most suitable for the virus included Bangladesh, the southern tip and eastern coast of Vietnam, parts of north-central Thailand and large parts of eastern China. Areas of moderate suitability included central Sumatra, Java, and northern India. In general, northern and central Asia, including Mongolia, Afghanistan, the majority of Pakistan, Nepal, Bhutan, western China, North Korea, northern and eastern Myanmar, northern Laos and Kalimantan were considered least suitable for HPAIV H5N1 occurrence. Areas considered unsuitable for poultry production (12% of study area) generally coincided with those predicted least suitable for HPAIV H5N1 occurrence (Fig. 1). Eighteen percent ($n/N = 675/3690$) of all outbreak locations occurred in areas deemed unsuitable for poultry production (Fig. 2).

3.3. Map validation

The ROC AUCs associated with the constrained and unconstrained models ranged from poor (0.563; 95% CI 0.559–0.566) to almost fair agreement (0.670; 95% CI 0.667–0.673) between areas predicted suitable for HPAIV H5N1 occurrence and the location of disease outbreaks in domestic poultry. However, the model in which the study area was unconstrained resulted in a higher ROC AUC than the model constrained to those areas suitable for poultry production (0.670 versus 0.563 respectively). As the unconstrained model had a higher ROC AUC than the constrained model, and those areas considered unsuitable for poultry production were generally modeled as being least suitable for HPAIV H5N1 occurrence, the unconstrained model was used in all further analyses.

Table 5

Sensitivity analysis of weights used to estimate the suitability of Asia for the occurrence of highly pathogenic avian influenza virus (HPAIV) H5N1. Average change in risk estimates was calculated using the point locations of 3690 outbreaks of HPAIV H5N1 in domestic poultry.

Factor	Mean change in suitability estimate \pm standard deviation	
	Factor weight increased by 25%	Factor weight decreased by 25%
WfowlDen	43.48 \pm 24.14	21.00 \pm 29.43
PopDen	33.24 \pm 29.86	33.87 \pm 23.31
ProxRoad	26.98 \pm 25.49	40.19 \pm 27.42
ProxRice	31.01 \pm 26.30	35.92 \pm 26.03
ProxWater	31.20 \pm 26.30	35.72 \pm 26.75
ChickDen	33.93 \pm 26.54	33.92 \pm 26.54

3.4. Sensitivity analysis

Altering the fuzzy membership functions of the risk factors resulted in a larger mean change in suitability estimates than altering the weights. Using the same weight for all factors (0.1667) resulted in a mean change in suitability estimates of 17 ± 30 units, while assuming a linear relationship between all factors and HPAIV H5N1 occurrence, instead of using appropriate membership functions, resulted in a mean change in suitability estimates of 99 ± 50 units. Increasing or decreasing individual factor weights while holding all other weights constant resulted in mean changes in suitability estimates ranging from 21 to 43 units (Table 5).

3.5. Application of DST

The map outputs resulting from the application of DST to the MCDA model are illustrated in Fig. 3. The evidence supported the belief that parts of eastern China, Bangladesh, the southern tip of Myanmar, and northern and

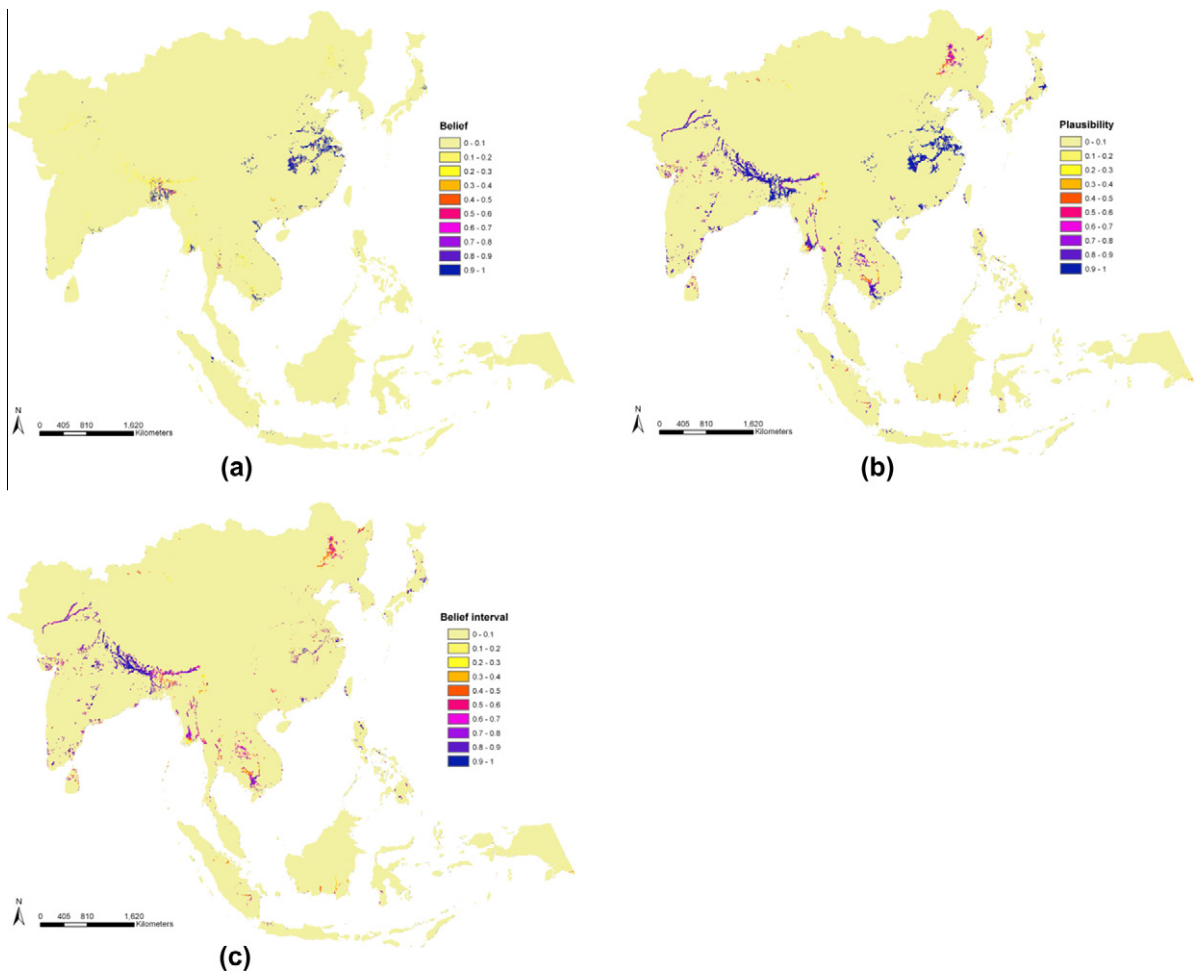


Fig. 3. Map outputs from the application of Dempster-Shafer theory (DST) to multicriteria decision analysis (MCDA) modeling of the suitability of Asia for the occurrence of highly pathogenic avian influenza virus (HPAIV) H5N1 in domestic poultry. (a) Belief, (b) plausibility and (c) belief interval.

southern Thailand were suitable for the occurrence of HPAIV H5N1. The model suggested that it was plausible that similar areas, although slightly more extensive than those indicated in the belief map, together with pockets in India – in particular along the north-eastern border of the country – parts of Thailand, Cambodia and Myanmar and along the Indus River, were suitable for HPAIV H5N1 occurrence, albeit generally at a lower intensity. As a result there were moderate levels of uncertainty surrounding the suitability estimates for Bangladesh, Myanmar, Thailand, Cambodia, Vietnam and eastern China but high uncertainty surrounding the suitability estimates along the north-eastern border of India and the Indus River.

4. Discussion

Only a few studies have attempted to map the risk or suitability of HPAIV H5N1 occurrence in Asia and those that have used a range of data-driven modeling methods. This study is unique for two reasons: firstly, it is the only study of which we know to use a knowledge-driven

approach to map the suitability of HPAIV H5N1 occurrence in domestic poultry and secondly, the only MCDA animal-health study to validate an MCDA model using quantitative methods rather than visual appraisal of the agreement between disease presence and predicted suitability.

Areas identified by this model as being most suitable for occurrence of the virus in domestic poultry included Bangladesh, the southern tip and eastern coast of Vietnam, parts of north-central Thailand and large parts of eastern China. These areas were generally characterized by high human population density, and to a lesser extent, high chicken density. Although high density of domestic waterfowl was generally associated with areas suitable for HPAIV H5N1 occurrence, domestic waterfowl density was low in Bangladesh –which was modeled as highly suitable for disease occurrence.

Most of India and large parts of eastern China were deemed to be moderately-to-highly suitable for the occurrence of HPAIV H5N1, yet have reported few outbreaks. This discrepancy between prediction and reality was supported by the high levels of uncertainty, as determined by DST,

surrounding the suitability estimates in these areas. In China the reason for this discrepancy may be due to a combination of mass vaccination and under-reporting while in India it may be the result of the prevailing semi-arid climate and relatively low domestic waterfowl density. In addition, WLC allows the low values of one risk factor to be compensated for by high values of another, and although India has a low domestic waterfowl density they have an extensive road network. The risk factor 'proximity to roads' therefore probably contributed largely to India receiving a moderately-high suitability rating for HPAIV H5N1 occurrence but the low domestic waterfowl density and semi-arid climate preclude a high disease incidence. Furthermore, this suggests that the occurrence of HPAIV H5N1 in this part of Asia may be under-reported as the extensive road network, together with its proximity to Bangladesh, implies that the area is likely to have been exposed to the virus yet has reported few outbreaks.

Our model showed good, broad visual agreement with most of the previous studies (Pfeiffer et al., 2007; Gilbert et al., 2008; Martin et al., 2011), in particular in the identification of disease 'hot spots'. However, there was poor visual agreement between this model and that of Fang et al. (Fang et al., 2008) whose predictions of high risk areas in mainland China were far more extensive than those of our model which identified only pockets in the east of the country to be suitable for HPAIV H5N1 occurrence. Martin et al. (Martin et al., 2011) and Hogerwerf et al. (Hogerwerf et al., 2010) limited their study areas to regions suitable for poultry production in order to exclude areas such as the desert regions of Inner Mongolia, Tibet and Xinjiang autonomous regions. These areas were all predicted to be high risk for HPAIV H5N1 occurrence by Fang et al. (Fang et al., 2008) even though they do not allow for the maintenance and transmission of HPAIVV H5N1. Our model was run twice; the first time the study area was constrained to those regions suitable for poultry production (Wint and Robinson, 2007) and the second time the study area remained unconstrained. In general, the unconstrained model categorized areas unsuitable for poultry production as least suitable for HPAIV H5N1 occurrence, suggesting that our model is a more accurate representation of the ecological niche of HPAIV H5N1 than that of Fang et al.'s (Fang et al., 2008).

MCDA has been used infrequently in the animal health field. One reason for the lack of use may be that none of the previous studies performed a quantitative assessment of their model's predictive accuracy, relying rather on visual appraisal of the spatial correlation between disease presence and predicted risk, which possibly inspires less confidence in the method than modeling approaches that include a quantitative assessment. This study is the first that we know of to use traditional quantitative methods of model evaluation to validate a MCDA model. A common assumption is that knowledge-driven models are inferior to data-driven ones; however, this assumption is not supported by any comparative studies of the two approaches as, until recently, it was not possible to quantitatively evaluate an MCDA model. However, validation of our unconstrained model using disease presence and background data produced a ROC AUC of 0.670 (95% CI 0.667–0.673)

which challenges, to some extent, the assumption of inferiority. In other words, a continental-level distribution map with reasonable predictive accuracy was produced without using data to fine-tune the model parameters. However, in order to allow for a direct comparison of the predictive accuracy of different modeling techniques, all models need to have dealt with the same species and have been modeled at the same extent and resolution (Lobo et al., 2008), which precludes a direct comparison of the accuracy of our model with those previously published (Pfeiffer et al., 2007; Fang et al., 2008; Gilbert et al., 2008; Hogerwerf et al., 2010; Martin et al., 2011).

The unconstrained model had a higher level of predictive accuracy than the constrained model (ROC AUC 0.670 versus 0.563 respectively), possibly due largely to disease presence points occurring in cells classified as open water and therefore unsuitable for poultry production. This was a particular problem on the island of Java. The reason for this is possibly twofold. Firstly, as the map showing areas suitable for poultry production was resampled to a resolution of 5 km², whole cells were classified as unsuitable for poultry production when only a portion of the cell may have contained open water. Thus, a disease outbreak may have, in reality, occurred near water but as a result of being resampled to a lower resolution the outbreak now appeared to occur in water. Secondly, disease outbreaks are frequently georeferenced to the centroid of an administrative area (Farnsworth et al., 2010; Gilbert and Pfeiffer, 2012) which may have inadvertently placed them in an area classified as unsuitable for poultry production due, for example, to the presence of open water. Thus, although biologically it seems reasonable to constrain a study area to only those regions capable of supporting the disease species, map resolution and how the data points have been georeferenced must be taken into account if such an approach is followed. This will be an important consideration when using ecological niche models, such as maximum entropy (Maxent) and genetic algorithm for rule-set production (GARP), which compare the ecological niche of disease point locations with that of background data. When almost 20% (as in this study) of those disease points occur in cells classified as unsuitable for poultry production, possibly as a result of map resolution or imprecise georeferencing, the accuracy of such models will be compromised.

It is important to bear in mind that the fair predictive accuracy of this model may be partially the result of the resolution used – 5 km². If we had used a higher resolution (e.g. 1 km²) the predictive accuracy might have been lower. Unfortunately, it was not possible to work with a higher resolution owing to computational limitations imposed by the size of the study area. However, we would argue that this limitation is mitigated by the fact that, owing to the potentially variable accuracy of the georeferencing of the disease points, a high resolution map might actually generate less meaningful output than a map of lower resolution. Thus, data- and knowledge-driven models both appear to be similarly constrained regarding the extent of the study area being modeled, albeit for different reasons. Conversely, our model may be under-predicting owing to the imprecise georeferencing frequently associated with

disease outbreaks. In addition, it is important to bear in mind that by validating against known outbreaks we assume that all areas have had the chance to become infected and, where conditions were suitable, outbreaks occurred and were reported. If some areas have not yet been exposed, our validation will be affected, in particular through occurrence of false-negatives.

Compared with data-driven modeling methods, especially those that require both disease presence and absence data, MCDA is relatively quick and inexpensive to implement, especially if the spatial risk factor data layers are sourced from the public domain. In addition, the method has a strong participatory element in that a range of stakeholders can be involved in the modeling process, thus bringing together different points of view. MCDA easily incorporates both qualitative and quantitative variables and, through the use of fuzzy membership sets, is able to accommodate non-linear relationships between risk factors and the outcome. Furthermore, unlike some of the data-driven methods, the incorporation of DST allows for the uncertainty surrounding the suitability estimates to be explored.

The application of DST to our model showed that, based on our choice of hard and soft evidence, disease has so far occurred primarily in areas with a low level of uncertainty (belief interval) while areas with a high level of uncertainty generally coincided with areas predicted suitable for disease yet which have so far reported few or no outbreaks. Such a spatial distribution of uncertainty may occur for one or more of the following reasons. Firstly, based on our classification of the risk factors as either hard or soft evidence for disease occurrence, the observed spatial distribution of uncertainty may suggest that the densities of the host (domestic chickens and waterfowl) and vector (domestic waterfowl) have so far been the main risk factors in the occurrence of HPAIV H5N1 in Asia and that the risk factors considered to represent soft evidence are indeed of secondary importance to the density of domestic chickens and waterfowl in disease occurrence. Secondly, the level of uncertainty may be a reflection of the varied degree of under-reporting in different countries or regions with higher levels of under-reporting resulting in higher levels of uncertainty being associated with the suitability estimates. Thirdly, risk factors used to reflect broad, continental-level causal relationships may not apply to all regions in the study area, reflected by high levels of uncertainty in those regions.

A potential reason why MCDA has been infrequently used to model the spatial distribution of disease is the subjectivity associated with the method, particularly in the selection of risk factors, definition of relationships between risk factors and disease occurrence and determination of weights, as opposed to the more objective data-driven methods. There is no doubt that, for each of these, people's perceptions will be influenced largely by their own experience with a disease, and therefore using an individual to build the model, based on their own experience with the disease, would introduce a considerable degree of subjectivity. One way to reduce this could be through identifying, defining and weighting risk factors based on group consensus. Another alternative would be to use the method

implemented initially by (Clements et al., 2006) and further developed in this study. By selecting and weighting risk factors based on the number of times each has been reported in the scientific literature to be significantly associated with a disease, together with the level of statistical significance achieved, the potentially biased perspective of an individual can be ameliorated. However, of greater concern is that this approach is heavily dependent on the results of previous studies for the identification and weighting of risk factors and thus, as our knowledge of the disease grows, the risk factors, relationships and weightings included in the model will inevitably be subject to change. In addition, as a disease evolves the focus of the studies regarding its epidemiology will inevitably change to account for our wider knowledge of the disease and the altered status of the disease in different locations resulting from the successful implementation of control measures.

On a similar note, in a study where the model inputs are based largely on published papers, of concern is the influence of publication bias on the results. According to (Ioannidis, 2005) publication bias is least likely to arise when (i) similar results have been identified (i.e. confirmed) by several, independent studies, (ii) the studies involved are large, (iii) the effect sizes are large, (iv) the dependent variable is “unequivocal and universally agreed” (Ioannidis, 2005; e124), (v) the studies use common, well-documented analytical methods, (vi) the research findings are not influenced by financial interests (e.g. pharmacokinetic trials) and (vii) there is no race between scientific teams working in the same field to be the first to publish significant findings on a ‘hot’ new topic. As the publications used to identify and define our model inputs spanned almost two decades, there is little likelihood that our choice of risk factors was wrongly influenced by competing teams hoping to be the first to publish significant findings on a new topic. In addition, the six risk factors used in our model have all been repeatedly identified, by several, independent studies as being significantly associated with HPAIV H5N1 occurrence. Furthermore, all studies used logistic regression together with a generally, well-defined outcome (i.e. laboratory confirmed cases of HPAIV H5N1) and, as a result of the disease characteristics, sample sizes were relatively large and the studies were unlikely to have been influenced by financial interests. All this suggests that the effect of publication bias on our model was likely to be minimal.

Despite concerns regarding the potential for error to be introduced through the subjective weighting of risk factors, it has been shown by both this study and Clements et al. (2006) that altering the weights has considerably less of an effect on the final suitability estimates than changing the shape of the relationships between risk factors and outcome. Assigning equal weights to all risk factors resulted in a mean change in suitability estimates of only 17 units, while modeling a linear relationship between all risk factors and the outcome resulted in a mean change in suitability estimate of 99 units, suggesting that choice of weights is reasonably robust. We suggest that weights have less of an effect than the relationships on the final suitability value of any raster cell, as individual weights

are always less than one (together they sum to one), while the relationship between risk factor and outcome determines the spread of values on the 0–255 scale. Thus, when calculating the final suitability estimate of a given pixel, the weight has less influence in the calculation than the value of the risk factor. This is particularly disquieting in light of the comment by Malczewski (Malczewski, 2000) that the most frequently modeled relationship between risk factor and outcome is a linear one. Another argument against using linear transformations, rather than fuzzy membership functions, to model the relationship between risk factors and the outcome is that suitability maps generated using linear transformations of datasets of different extent will not be comparable as the final suitability of a raster cell will be relative to the range of suitability scores for that particular dataset. However, if the same fuzzy membership functions and decision rules are applied to datasets of different extent then the resulting suitability maps will be comparable.

One of the main limitations of MCDA, and all spatial modeling techniques, is that risk factors must be able to be mapped. This means that farm-level risk factors known to be associated with HPAIV H5N1 occurrence, such as an individual farm's trading intensity (Desvaux et al., 2011; Paul et al., 2011) or biosecurity practices (Paul et al., 2011) cannot be included in the model. An extension of this problem is that of risk factors not being available at a sufficiently high resolution for incorporation in the model. For example, socio-economic factors such as purchasing power per capita were incorporated in their spatial model by Hogerwerf et al. (2010). However, such risk factors are generally only available in the public domain at a national level and therefore only serve to discriminate the level of risk between countries and cannot be used to differentiate between risk levels within a country. In addition, the difficulties associated with obtaining the relevant spatial data may mean that the contribution of certain risk factors to disease occurrence have not been studied and are therefore excluded from the MCDA model. For example, owing to the lack of relevant spatial data, few statistical studies have investigated the relationship between the presence of wild birds and HPAIV H5N1 occurrence, even though wild birds have been implicated in long-distance spread of the disease (Takekawa et al., 2010; Gilbert et al., 2011). However, as the entire study area (apart from a small corner in north-west Afghanistan, which had not experienced any HPAIV H5N1 outbreaks at the time of the study) was covered by either the East Asia/Australia or Central Asia flyways we assumed that any potential risk of HPAIV H5N1 occurrence associated with migratory birds would be reasonably constant across the study area and the omission of this risk factor from the model would not unduly influence the results. However, those implementing and using spatial models need to bear in mind that, by omitting potential risk factors for such reasons, the final model may not represent the whole picture. In addition, our knowledge of the epidemiology of HPAIV H5N1 is constantly growing and evolving and therefore modelers and decision-makers need to take into consideration that MCDA maps only reflect the current state of our knowledge and therefore the more we know about the epidemiology of

the disease the more accurate the suitability maps that can be produced.

As we were modeling the risk of HPAIV H5N1 occurrence for the whole of Asia rather than for a specific country risk factors had to reflect broad causal relationships at the continental-level rather than be country-specific. We therefore assumed that risk factors which had been identified as being significantly associated with HPAIV H5N1 occurrence in multiple countries were those that described general, continental-level causal relationships even if individual risk factor had not been found to be significant in certain countries. For example, although a range of studies have shown density of domestic wildfowl to be associated with HPAIV H5N1 persistence in Asia (Gilbert et al., 2006; Pfeiffer et al., 2007; Gilbert et al., 2008; Tiensin et al., 2009; Paul et al., 2010; Hogerwerf et al., 2010), Loth et al.'s high-resolution studies found no significant association between this variable and disease occurrence in either Java (Loth et al., 2010) or Indonesia (Loth et al., 2011) despite the fact that both countries have high densities of domestic waterfowl and are endemic for HPAIV H5N1. However, we assume this to be due to the variable reflecting different transmission mechanisms at different resolutions as the extent to which domestic waterfowl contribute to HPAIV H5N1 transmission may vary considerably depending on local farming and trading conditions, which themselves differ greatly between countries (Gilbert and Pfeiffer, 2012). As these mechanisms are seldom of the type that can be mapped, a model that reflects broad causal relationships at a crude scale is possibly more accurate and contains less uncertainty than one which attempts to model complex, fine-scale mechanisms.

Although the spatial unit of analysis for seven (70%) of the ten studies used to identify the variables for inclusion in the MCDA model (Table 1) was Administrative Level 3, for the remaining three studies spatial extent varied from individual cases (Ward et al., 2008) and farms (Biswas et al., 2009a) to country-level (Hogerwerf et al., 2010). Using studies performed at different spatial extents to identify variables for inclusion in the model and to define their relationship with disease occurrence can add an extra degree of uncertainty to the MCDA model as, variables modeled as significant at different spatial extents may reflect different processes in the causal relationship between variable and outcome (Gilbert and Pfeiffer, 2012). Similarly, studies performed in vastly differing countries yet which identify the same variable to be significantly associated with disease may again be modeling different processes. However, the coarser the spatial extent of the studies, the more likely they are to be modeling the same processes (Gilbert and Pfeiffer, 2012) and similarly, the more alike countries are with respect to climate, agro-ecological and socio-economic factors the more likely they are to be modeling the same causal relationship. In this study, the majority of papers used to identify model inputs were performed at a similar coarse spatial scale (i.e. Administrative Level 3) suggesting that the associations they identified between the different variables and HPAIV H5N1 occurrence were more likely to reflect the relatively straightforward relationships we modeled here as opposed to the complex and diverse mechanisms by which the

variables could contribute to disease occurrence at a finer scale. This assumption is also supported by the generally low levels of uncertainty presented in the Dempster-Shafer map (Fig. 3).

Our predictive risk map was validated using HPAI H5N1 outbreak data extracted from the EMPRES-i database, which contains data collected from both formal (e.g. reports from the OIE, World Health Organization, national authorities, FAO country or regional projects, field missions and field officers, non-governmental organizations, laboratories and reference centers) and informal sources (e.g. media reports and those disseminated by the Global Public Health Intelligence Network and ProMED) (Martin et al., 2007; Farnsworth et al., 2010). In addition, all outbreaks appearing in the EMPRES-i database are followed-up until either confirmed or denied (Farnsworth et al., 2010). An alternative to EMPRES-i data is the individual, national HPAIV databases which are generally more comprehensive than the international data sources with respect to number of outbreaks, level of detail included and are recorded at a finer spatial resolution (Farnsworth et al., 2010). However, these databases are not available for all countries and are generally not as easily accessible as their global counterparts. This suggests that EMPRES-i is an adequate source of data for the spatial modeling of HPAIV H5N1 at a continental level and for the inferences we wanted to draw, although users need to bear in mind the limitations of using data that have been spatially referenced to administrative centroids rather than exact outbreak locations, as discussed previously.

In conclusion, this work demonstrates the potential of MCDA for knowledge synthesis. In contrast to data-driven models, which require the availability of a single dataset containing all relevant information, MCDA can integrate data generated by a range of studies conducted together or independently from each other. Many risk management scenarios require action in the presence of scarce knowledge, and in such situations MCDA allows decision-makers to take advantage of existing qualitative expert knowledge while recognizing the uncertainty associated with their predictions. Furthermore, the resulting suitability models can be updated or modified once additional knowledge has been generated, and the quantitative relationships represented in MCDA models can be communicated and debated with an expert audience, thereby providing an opportunity for consensus building.

5. Conclusion

Using only knowledge regarding the relationship between key risk factors and HPAIV H5N1 occurrence in domestic poultry, this study showed that it is possible to model the continental-level risk of HPAIV H5N1 occurrence in domestic poultry with a fair degree of predictive accuracy, in particular the broad location of disease 'hot-spots'. However, by validating a knowledge-driven spatial model using a quantitative method, this study has paved the way for future research involving a comparison of knowledge- and data-driven spatial modeling methods. The work has also shown that MCDA is a practical tool for synthesizing knowledge about a disease and generating

reasonable spatial representations of variation in disease suitability which in turn has direct relevance for informing the design of risk-based animal health surveillance.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

KBS performed all data manipulation, analyses and spatial modeling and wrote the paper. MG conducted the literature search identifying factors associated with HPAIV H5N1, contributed the maps of chicken and domestic waterfowl density and edited the final draft of the manuscript. DUP conceived of the study, participated in its design and coordination and edited the final draft of the manuscript.

References

- Biswas PK, Christensen JP, Ahmed SSU, Barua H, Das A, Rahman MH, et al. Risk factors for infection with highly pathogenic influenza A virus (H5N1) in commercial chickens in Bangladesh. *Vet Res* 2009a;164:743–6.
- Biswas PK, Christensen JP, Ahmed SSU, Das A, Rahman MH, Barua H, et al. Risk for infection with highly pathogenic avian influenza virus (H5N1) in backyard chickens, Bangladesh. *Emerg Infect Dis* 2009b;15:1931–6.
- Boyce M, Vernier P, Nielsen S, Schmiegelow F. Evaluating resource selection functions. *Ecol Modell* 2002;157:281–300.
- Center for International Earth Science Information Network (CIESIN), C.U.a.C.I.d.A.T.C. Gridded population of the world Version 3 (GPWv3): population density grids. Palisades, NY: Socioeconomic Data and Applications Center (SEDAC), NY: Columbia University. Available at: <<http://sedac.ciesin.columbia.edu/gpw>>; 2005.
- Chefaoui R, Lobo J. Assessing the effects of pseudo-absences on predictive distribution model performance. *Ecol Modell* 2008;210:478–86.
- Clements ACA, Pfeiffer DU, Martin V. Application of knowledge-driven spatial modelling approaches and uncertainty management to a study of Rift Valley fever in Africa. *Int J Health Geogr* 2006;5:57–69.
- Clements ACA, Pfeiffer DU, Martin V, Pittiglio C, Best N, Thiongane Y. Spatial risk assessment of Rift Valley fever in Senegal. *Vector Borne Zoonotic Dis* 2007;7:203–16.
- Desvaux S, Grosbois V, Pham TTH, Fenwick S, Tollis S, Pham NH, et al. Risk factors of highly pathogenic avian influenza H5N1 occurrence at the village and farm levels in the red river delta region in Vietnam. *Transbound Emerg Dis* 2011. <http://dx.doi.org/10.1111/j.1865-1682.2011.01227.x>.
- Eastman JR. GIS and uncertainty management: new directions in software development. *Lurralde* 1997;20:53–66.
- Eastman JR. Decision support: uncertainty management. *IDRISI guide to GIS and image processing*. Accessed in IDRISI Andes., Worcester, MA: Clark University; 2009. p. 156–172.
- Engler R, Guisan A, Rechsteiner L. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. *J Appl Ecol* 2004;41:263–74.
- Fang L-Q, de Vlas SJ, Liang S, Looman CWN, Gong P, Xu B, et al. Environmental factors contributing to the spread of H5N1 avian influenza in mainland China. *PLoS One* 2008;3:e2268.
- FAO. EMPRES transboundary animal disease bulletin. p. 7–8. Available from: <<http://www.fao.org/docrep/014/j2249e/j2249e2200.pdf>>; 2011 [accessed June 2011].
- Farnsworth ML, Hamilton-West C, Fitchett S, Newman SH, de La Rocque S, De Simone L, et al. Comparing national and global data collection systems for reporting, outbreaks of H5N1 HPAI. *Prev Vet Med* 2010;95:175–85.
- Franklin J. Model evaluation. Mapping species distributions: spatial inference and prediction. New York: Cambridge University Press; 2009a. p. 209–234.

- Franklin J. Species distribution modelling. New York: Oxford University Press; 2009b. p. 3–20.
- Gilbert M, Chaitaweesub P, Parakamawongsa T, Premashthira S, Tiensin T, Kalpravidh W, et al. Free-grazing ducks and highly pathogenic avian influenza, Thailand. *Emerg Infect Dis* 2006;12:227–34.
- Gilbert M, Newman S, Takekawa J, Loth L, Biradar C, Prosser D, et al. Flying over an infected landscape: distribution of highly pathogenic avian influenza H5N1 risk in South Asia and satellite tracking of wild waterfowl. *EcoHealth* 2011;1–11.
- Gilbert M, Pfeiffer D. Risk factor modelling of the spatio-temporal patterns of highly pathogenic avian influenza (HPAIV) H5N1: a review. *Spat Spatio-Temporal Epidemiol* 2012;3:173–83.
- Gilbert M, Xiao X, Chaitaweesub P, Kalpravidh W, Premashthira S, Boles S, et al. Avian influenza, domestic ducks and rice agriculture in Thailand. *Agric Ecosyst Environ* 2007;119:409–15.
- Gilbert M, Xiao X, Pfeiffer DU, Epprecht M, Boles S, Czarnecki C, et al. Mapping H5N1 highly pathogenic avian influenza risk in Southeast Asia. *Proc Natl Acad Sci USA* 2008;105:4769–74.
- Greiner M, Pfeiffer D, Smith RD. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. *Prev Vet Med* 2000;45:23–41.
- Henning KA, Henning J, Morton J, Long NT, Ha NT, Meers J. Farm- and flock-level risk factors associated with highly pathogenic avian influenza outbreaks on small holder duck and chicken farms in the mekong delta of Viet Nam. *Prev Vet Med* 2009;91:179–88.
- Hogerwerf L, Wallace R, Ottaviani D, Slingenbergh J, Prosser D, Bergmann L, et al. Persistence of highly pathogenic avian influenza H5N1 virus defined by agro-ecological niche. *EcoHealth* 2010;2:13–25.
- Hongoh V, Hoen A, Aenishaenslin C, Waaub J-P, Belanger D, Michel P, et al. Spatially explicit multi-criteria decision analysis for managing vector-borne diseases. *Int J Health Geogr* 2011;10:70.
- Ioannidis JPA. Why most published research findings are false. *PLoS Med* 2005;2:e124.
- Li KS, Guan Y, Wang J, Smith GJD, Xu KM, Duan L, et al. Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. *Nature* 2004;430:209–13.
- Lipatov AS, Govorkova EA, Webby RJ, Ozaki H, Peiris M, Guan Y, et al. Influenza: emergence and control. *J Virol* 2004;78:8951–9.
- Lobo JM, Jiménez-Valverde A, Real R. AUC: a misleading measure of the performance of predictive distribution models. *Glob Ecol Biogeogr* 2008;17:145–51.
- Loth L, Gilbert M, Osmani MG, Kalam AM, Xiao X. Risk factors and clusters of highly pathogenic avian influenza H5N1 outbreaks in Bangladesh. *Prev Vet Med* 2010;96:104–13.
- Loth L, Gilbert M, Wu J, Czarnecki C, Hidayat M, Xiao X. Identifying risk factors of highly pathogenic avian influenza (H5N1 subtype) in Indonesia. *Prev Vet Med* 2011;102:50–8.
- Malczewski J. GIS and multicriteria decision analysis. Inc New York: John Wiley & Sons; 1999.
- Malczewski J. On the use of weighted linear combination method in GIS: common and best practice approaches. *Trans GIS* 2000;4:5–22.
- Martin V, De Simone L, Lubroth J. Geographic information systems applied to the international surveillance and control of transboundary animal diseases, a focus on highly pathogenic avian influenza. *Vet Italia* 2007;43:437–50.
- Martin V, Pfeiffer DU, Zhou X, Xiao X, Prosser DJ, Guo F, et al. Spatial distribution and risk factors of highly pathogenic avian influenza (HPAI) H5N1 in China. *PLoS Pathog* 2011;7:e1001308.
- Martin V, Sims L, Lubroth J, Pfeiffer D, Slingenbergh J, Domenech J. Epidemiology and ecology of highly pathogenic avian influenza with particular emphasis on South East Asia. *Dev Biol (Basel)* 2006;124:23–36.
- Minh PQ, Morris RS, Schauer B, Stevenson M, Benschop J, Nam HV, et al. Spatio-temporal epidemiology of highly pathogenic avian influenza outbreaks in the two deltas of Vietnam during 2003–2007. *Prev Vet Med* 2009;89:16–24.
- Paul M, Tavnornpanich S, Abrial D, Gasqui P, Charras-Garrido M, Thanapongtharm W, et al. Anthropogenic factors and the risk of highly pathogenic avian influenza H5N1: prospects from a spatial-based model. *Vet Res* 2010;41:28.
- Paul M, Wongnarkpet S, Gasqui P, Poolkhet C, Thongratsakul S, Ducrot C, et al. Risk factors for highly pathogenic avian influenza (HPAI) H5N1 infection in backyard chicken farms, Thailand. *Acta Trop* 2011;118:209–16.
- Pfeiffer D, Robinson T, Stevenson M, Stevens K, Rogers DAC. Spatial risk assessment and management of disease. Spatial analysis in epidemiology. Oxford University Press; 2008. p. 110–119.
- Pfeiffer DU, Minh PQ, Martin V, Epprecht M, Otte MJ. An analysis of the spatial and temporal patterns of highly pathogenic avian influenza occurrence in Vietnam using national surveillance data. *Vet J* 2007;174:302–9.
- Prosser D, Wu J, Ellis EC, Gale F, Van Boeckel TP, Wint W, et al. Modelling the distribution of chickens, ducks, and geese in China. *Agric Ecosyst Environ* 2011;141:381–9.
- Rakotomanana F, Randremanana R, Rabarijaona L, Duchemin J, Ratovonjato J, Ariey F, et al. Determining areas that require indoor insecticide spraying using multi criteria evaluation, a decision-support tool for malaria vector control programmes in the Central Highlands of Madagascar. *Int J Health Geogr* 2007;6:2.
- Saaty RW. The analytic hierarchy process – what it is and how it is used. *Math Modell* 1987;9:161–76.
- Saaty TL. Multicriteria decision making: the analytic hierarchy process. PA: RWS Publications Pittsburgh; 1990.
- Slingenbergh J, Gilbert M, 2008. Do old and new forms of poultry go together? In: Proceedings of the International Conference on Poultry in the 21st Century, Bangkok, Thailand.
- Takekawa JY, Newman SH, Xiao X, Prosser DJ, Spragens KA, Palm EC, et al. Migration of waterfowl in the East Asian flyway and spatial relationship to HPAI H5N1 outbreaks. *Avian Dis* 2010;54:466–76.
- Tiensin T, Ahmed SSU, Rojanasthien S, Songserm T, Ratanakorn P, Chaichoun K, et al. Ecologic risk factor investigation of clusters of avian influenza A (H5N1) virus infection in Thailand. *J Infect Dis* 2009;199:1735–43.
- Tiensin T, Chaitaweesub P, Songserm T, Chaisingh A. Highly pathogenic avian influenza H5N1, Thailand, 2004. *Emerg Infect Dis* 2005;11:1664–72.
- Tiensin T, Nielen M, Songserm T, Kalpravidh W, Chaitaweesub P. Geographic and temporal distribution of highly pathogenic avian influenza A virus (H5N1) in Thailand, 2004–2005: an overview. *Avian Dis* 2007;51: 182–118.
- Van Boeckel M, Prosser D, Wint W, Robinson T, Gilbert M. Modelling the distribution of domestic duck in Monsoon Asia. *Agric Ecosyst Environ* 2011;141:373–80.
- VanDerWal J, Shoo L, Graham C, William S. Selecting pseudo-absence data for presence-only distribution modeling: how far should you stray from what you know? *Ecol Modell* 2009;220:589–94.
- Ward M, Maftai D, Apostu C, Suru A. Environmental and anthropogenic risk factors for highly pathogenic avian influenza subtype H5N1 outbreaks in Romania, 2005–2006. *Vet Res Commun* 2008;32:627–34.
- WHO. Cumulative number of confirmed human cases of avian influenza A/(H5N1) reported to WHO. Available from: <http://www.who.int/csr/disease/avian_influenza/country/cases_table_2011_2006_2011/en/index.html>; 2011 [accessed June 2011].
- Wint W, Robinson T. Gridded livestock of the world. Food Agric Org Rome; 2007.
- Wisiz M, Guisan A. Do pseudo-absence selection strategies influence species distribution models and their predictions? An information-theoretic approach based on simulated data. *BMC Ecol* 2009;9:8.
- Yupiana Y, de Vlas SJ, Adnan NM, Richardus JH. Risk factors of poultry outbreaks and human cases of H5N1 avian influenza virus infection in west java province, Indonesia. *Int J Infect Dis* 2010;14:e800–5.