Risk analysis for the highly pathogenic avian influenza in Mainland China using meta-modeling

CAO ChunXiang[1*], XU Min[1,3], CHANG ChaoYi[1,3], XUE Yong[1], ZHONG ShaoBo[1], FANG LiQun[2], CAO WuChun[2], ZHANG Hao[1], GAO MengXu[1,3], HE QiSheng[1,3], ZHAO Jian[1,3], CHEN Wei[1,3], ZHENG Sheng[1,3] & LI XiaoWen[1]

[1] *State Key Laboratory of Remote Sensing Science, the Institute of Remote Sensing Applications of the Chinese Academy of Sciences, Beijing 100101, China;*
[2] *Beijing Institute of Microbiology and Epidemiology, State Key Laboratory of Pathogen and Biosecurity, Beijing 100071, China;*
[3] *Graduate University of the Chinese Academy of Sciences, Beijing 100049, China*

A logistic model was employed to correlate the outbreak of highly pathogenic avian influenza (HPAI) with related environmental factors and the migration of birds. Based on MODIS data of the normalized difference vegetation index, environmental factors were considered in generating a probability map with the aid of logistic regression. A Bayesian maximum entropy model was employed to explore the spatial and temporal correlations of HPAI incidence. The results show that proximity to water bodies and national highways was statistically relevant to the occurrence of HPAI. Migratory birds, mainly waterfowl, were important infection sources in HPAI transmission. In addition, the HPAI outbreaks had high spatiotemporal autocorrelation. This epidemic spatial range fluctuated 45 km owing to different distribution patterns of cities and water bodies. Furthermore, two outbreaks were likely to occur with a period of 22 d. The potential risk of occurrence of HPAI in Mainland China for the period from January 23 to February 17, 2004 was simulated based on these findings, providing a useful meta-model framework for the application of environmental factors in the prediction of HPAI risk.

highly pathogenic avian influenza, meta-modeling, remote sensing, geographical information system, Bayesian maximum entropy, logistic regression, spatiotemporal autocorrelation

There has been much progress in the study of influenza viruses. Influenza viruses are important members of the family Orthomyxoviridae and are characterized as Types A, B, and C on the basis of the character of the internal nucleoprotein antigen. Influenza A viruses infect domestic birds, wild birds, humans, pigs, and ocean mammals. Until recently, the other types had not been found to infect birds. When referring to avian influenza, we mean a syndrome of avian diseases caused by influenza A viruses. Moreover, influenza viruses are further divided into subtypes determined by hemagglutinin (H) and neuraminidase (N) antigens. To date, 16 H subtypes and 9 N subtypes have been identified from influenza A. New viruses can be formed by any combination of subtypes. Thus, theoretically, there are 144 influenza A viruses, although there are fewer in reality. When humans are co-infected with different types of viruses, these viruses have the propensity to re-assort or mutate and produce new viruses that spread among humans rapidly. This leads to serious socio-economic and public health consequences. According to the pathogenicity of viruses to avian species, avian influenza is usually divided into highly pathogenic avian influenza (HPAI) and mildly pathogenic avian influenza (MPAI). While MPAI is non-virulent, does not have clinical signs, and causes only

*Corresponding author (email: cao413@irsa.ac.cn)

mild respiratory or reproductive disease, HPAI causes severe illness and results in mortality approaching 100% [1–3].

1 Introduction of the prevalence of HPAI

Since the end of 2003, avian influenza has frequently broken out in many countries. In 2004 and 2005, some areas in Mainland China were affected by the disease while more than 10 other Asian countries suffered from HPAI (mainly H5N1). In the short period from January 23 to February 26, 2004, there were a total of 49 occurrences in 16 provinces (autonomous regions and municipalities) of Mainland China (http://www.agri.gov.cn/). In April/May 2005, an outbreak of H5N1 was detected in bar-headed geese at Qinghai Lake (located in Qinghai Province, China). From June to November 2005, more avian influenza cases were subsequently reported throughout Mainland China. At the end of November 2005, the first human case of infection with avian influenza in Mainland China was reported [4]. Figure 1 illustrates the monthly HPAI occurrences from January 2004 to November 2005 in Mainland China.

Several preliminary studies showed that the occurrences of HPAI have high temporal and spatial clustering and are closely related to environmental factors [5]. Most of the avian influenza cases reported involved contact with large bodies of water and waterfowl. Chen et al. [6] justified that the H5N1 virus had already been transmitted between migratory birds at a lake. Waterfowl not only comprised wild birds but also migratory birds and domestic birds (such as ducks). Because of their stopover functions, it is strongly suspected that water bodies are significant locations for the transmission of avian influenza, especially between migratory birds to poultry. In addition, seasonal changes affected outbreaks, which can probably be attributed to temperature changes. Therefore, even though the clear relationship between environmental factors and avian influenza outbreaks was not explained and substantiated, we suspect that environmental factors affect the spread of influenza. In early research, remote sensing (RS) and geographic information systems (GISs) were successfully used in analyzing malaria

and schistosomiasis, demonstrating the tremendous potential of adapting GIS and RS techniques in explaining epidemics in specific environments [7,8]. RS can obtain abundant data and a GIS can provide spatial analysis tools to interpret the relationships among flyways, the environment, and HPAI. Early qualitative analyses using GIS and RS techniques showed that various factors (such as climate, bird migration, transportation, and water bodies) have significant correlation with the occurrence of HPAI [9,10]. Determining the exact environmental factors influencing outbreaks is now the key to understanding the spread of HPAI. Recent research indicates that rice agriculture positively corresponds to HPAI occurrence [11], which implies that vegetation and the spill-over function of poultry should not be neglected. Besides vegetation, the land surface temperature (LST) and circumstance of poultry have roused the interest of researchers to identify how HPAI spreads. Previous research has mainly employed spatial analysis functions and exploratory spatial data analysis techniques such as overlapping analysis, buffering analysis, histograms, scatterplots, and spatial clustering (in these studies, climate data such as water vapor, land surface temperature, and aerosol optical depth were obtained by R S) We selected essential environmental factors only and reduced the number of correlations between factors. Conventionally, the auto-logistic regression model has been applied in epidemiology, especially in binary outbreak modeling [12,13]. In addition, including modern spatial and temporal statistics has enhanced the accuracy of simulation. The Bayesian maximum entropy (BME) approach applies outcomes of the binary logistic model efficiently by tolerating uncertainties throughout assessment [14–16]. Employing the BME and binary logistic model reveals systematic associations between the environmental influence and spatiotemporal autocorrelations, and allows the development of a mathematically rigorous and scientifically meaningful modeling framework by accounting for environmental factors, space, time and epidemic determinants of HPAI diffusion [17]. These studies provide a valuable theoretical basis for applying RS and GIS technologies in analyzing HPAI diffusion. To further understand the mechanism of an outbreak and epidemic of HPAI, it is necessary to quantitatively analyze the correlation between
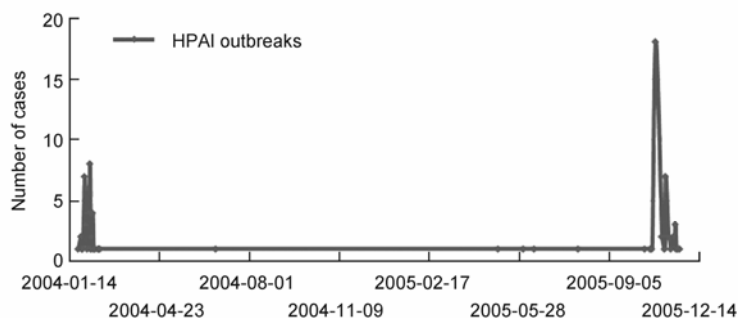

Figure 1   Number of HPAI occurrences in Mainland China in 2004. Shown are laboratory-confirmed H5N1 cases only.

the occurrence of HPAI and related factors.

2 Algorithm

2.1 Data and preprocessing

(i) Data and preprocessing for GIS analyses. Data of the occurrences of HPAI in 2004 were collected from monthly reports published by the Ministry of Agriculture, China (http://www.agri.gov.cn/). The data items include the positions of villages where disease broke out and the dates of the outbreaks. According to the 1:1000000 ArcGIS digital map of China provided by the State Bureau of Surveying and Mapping, the sites of the outbreaks (the centroids of the villages where HPAI occurred) are geocoded precisely onto the map. The dates when the disease broke out were defined as when the suspected cases were reported and confirmed finally, or when the confirmed cases were reported directly without suspected cases reported previously.

In this study, we used MODIS data. The National Aeronautics and Space Administration provides full and free land, atmosphere, and ocean data at their website for download. All remotely sensed data were preprocessed throughout the territory of Mainland China and its surrounding areas (72°–136°E, 16°–56°N) before they were used with data of the occurrences of HPAI. Refer to the ArcGIS online manual for details of the algorithms (http://www.esri.com/). Eight environmental factors were chosen for consideration: rivers, lakes, reservoirs, national expressways, railways, land temperature, vegetation, and moisture. Rivers, lakes, reservoirs, national highways, and railway data were obtained from layers of 1:1000000 digital maps. Corresponding MODIS standard products for all of China were downloaded from the website of the Goddard Earth Sciences Data and Information Services Center (http://disc.gsfc.nasa.gov/) to obtain land temperature and moisture data. SPOT vegetation products for all of China corresponding to the period of occurrences of HPAI were downloaded from a website providing free SPOT vegetation products (http://free.vgt.vito.be). Figure 2 illustrates the operational flow before modeling analysis.

(ii) Statistical analyses. In our study, a binary logistic regression model that considers spatial and temporal weights was used to analyze the correlation between environmental factors and HPAI occurrence. The LST, infrared water vapor (IWV), and normalized difference vegetation index (NDVI) were retrieved from RS data while the proximity to water bodies (rivers, lakes, and reservoirs) and proximity to transportation lines (national highways and railways) were calculated using GIS (Figure 3). Considering the relationships among spatial environmental data, flyways of migratory birds, and the occurrences of HPAI, we defined the spatial weights of the occurrences of HPAI.

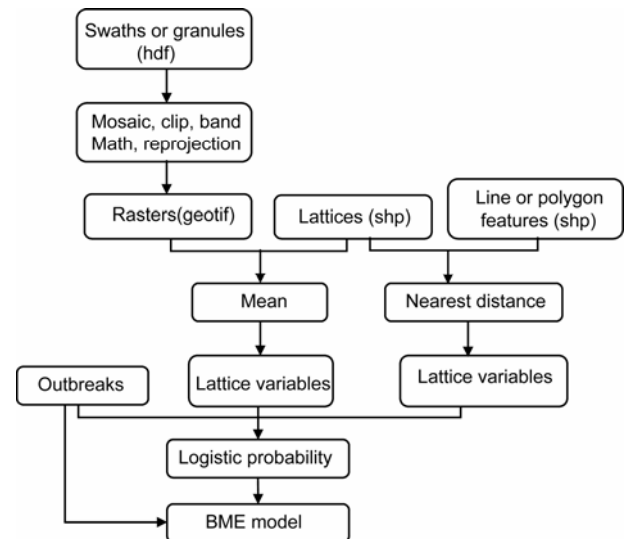(iii) Strategy for spatial analysis. In accordance with the

Figure 2 Procedures of meta-modeling. Logistic regression analysis with HPAI occurrence as the response variable and various factors derived from remotely sensed images or GIS vector data as covariates. The main steps of the processing of remotely sensed data and GIS vector data are also illustrated. In the BME model, the outbreaks are hard data and the logistic probability map is soft data.

spatial resolution of MODIS satellite imagery, the digital vector map was divided into a lattice with 0.5°×0.5° cells, on which the occurrences of HPAI were located and regarded as virtual objects for spatial analysis. Objects that are similar to those in case-control (or case-reference) studies in epidemiology [18] were selected. In this way, we reduce selective bias and avoid heavy yet inefficient data processing and computational tasks. For this purpose, we designed the following sampling strategies. First, we partitioned the study area (Mainland China) into rows and columns of small spatial cells with dimensions of 1°×1°. Each cell having an HPAI occurrence was selected as a case category, which is used in the following discussion to interpret the evolving pattern of H5N1 and establish a prediction program. In addition, twice as many cells were selected as a control group to verify the accuracy of the modeling result.

2.2 Methodology and modeling

(i) Formulation. It is recommended to use a nonlinear function in the ordinary linear regression of a binary variable. A logistic function has been widely used, and the corresponding regression model is usually called a logistic regression model [19]. Let the conditional probability of the occurrence of an event be $P(y_i=1|x_i)=p_i$. The logistic regression model assumes that this probability can be expressed as

$$p_i = \frac{1}{1+e^{-(\alpha+\beta x_i)}} = \frac{e^{(\alpha+\beta x_i)}}{1+e^{(\alpha+\beta x_i)}}, \tag{1}$$

where p_i is the conditional probability of the occurrence of the ith case. The function is a nonlinear function of the covariates and can be transformed into a linear function.
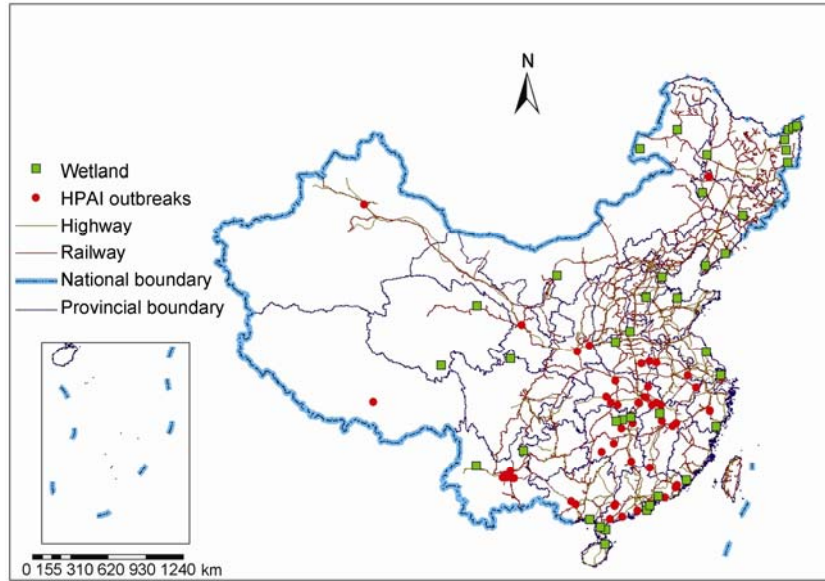
Figure 3 Geographic distributions of highways, railways, wetland, and HPAI occurrence. There is correlation between the transportation network and HPAI outbreaks that requires analysis.

According to eq. (1), the ratio of the occurrence of an event to non-occurrence is

$$\frac{p_i}{1-p_i} = e^{\alpha+\beta x_i}.$$
(2)

This ratio is referred to as "the odds". We obtain a linear function by taking a natural logarithm transform of the odds:

$$\ln\left(\frac{p_i}{1-p_i}\right) = \alpha + \beta x_i.$$
(3)

(ii) Binary logistic regression model. Owing to the spatial variability influenced by migratory birds, the classic logistic model is not suitable for the analysis of spatial data. Thus, the binary logistic regression model, which is especially designed for the analysis of spatial data, was put forward [20]. This model incorporates a spatial weighting function into the classic logistic regression model. It can deal with inherent spatial variability in spatial statistical analysis. The logistic regression model has been developed over several decades and is widely used.

In contrast to the classical logistic regression model, binary logistic regression is better able to represent the 0/1 value of outbreaks (i.e. the "yes" or "no" situation). The binary logistic regression model determines the conditional probability of y_i according to

$$p_i = P(y_i \mid y_j(j \neq i)) = \frac{e^{\eta_i}}{1+e^{\eta_i}},$$
(4)

where $\hat{y}_i = \sum_{j=1}^{n} y_j w_{i,j}$, $\beta=(\beta_1, \beta_2,\cdots,\beta_p)'$ is the parameter vec-

tor of the covariates, $X_i=(x_{i1}, x_{i2},\cdots, x_{ip})'$ is the covariate vector, and γ is the parameter corresponding to the spatial covariate $\hat{y}_i$.

However, for time-series data like these in our study, we have to consider both a spatial function and temporal function comprehensively. Here, we let

$$w_{ij} = \varpi_{ij}\omega_{ij}.$$
(5)

Thus, the formula is the same as for the logistic model. However, the meaning of the final term in the equation is quite different. Thus, the probability of occurrence in each lattice cell can be obtained to construct the BME soft data.

2.3 Modeling and parameter estimation

In our study, all variables initially incorporated are theoretically acceptable. In the phase of model building, we need to select all candidate variables that can predict the response variable well. In general, we first fit a preliminary model and then modify the results to produce a new model. Each model has a different mathematic meaning; the binary logistic model considering various environmental factors as well as the BME model focuses on spatial and temporal statistics. We then compare the results obtained by employing the models. Adjustments and comparisons are likely to be repeated many times in producing the final model. In the first step, when there are many covariates, the work is time and energy consuming. In this case, we may let a computer automatically select covariates. In practice, the variables for constituting the model should pass the statistical *f*-test.

Because incorporating the spatial and temporal function ruins the assumption of variable independence, the maximum

likelihood method is not suitable in determining parameters of the model. In principle, we should employ other approaches in determining parameters; e.g. the Monte Carlo Markov chain. However, for convenience, we used the binary logistic procedure of SPSS13 software (SPSS Inc., 1989–2004) to estimate parameters; i.e. we used the pseudolikelihood method [21].

2.4 Spatial weighting function for the flyways of migratory birds

In our study, migratory birds were considered to be a main factor influencing the outbreaks. Figure 3 portrays the four acknowledged migratory routes of birds in Mainland China [22]: (1) from Inner Mongolia, Gansu Province, and Qinghai Province to India and Pakistan; (2) from northeastern and northern China to Southeast Asia; (3) from northeastern and northern China to Japan, the Philippines, and Australia; and (4) from Mongolia and Siberia to southern China. On these migratory routes, 20 stopover sites were considered in our model. Figure 4 shows the occurrences of HPAI superimposed on the main flyways and important wetlands, and it is clear that the occurrences are quite near the flyways.

It is known that birds rarely fly incessantly while migrating. Because of the length of their journeys, they stop over at certain locations, and these locations are of particular importance in terms of the spread of disease. Stopovers are generally locations that provide abundant food, are hardly affected by human beings, and provide appropriate habitat. These places are mainly wetlands such as lakes, reservoirs, and rivers. When we superimpose major wetlands in Mainland China, the main flyways in China, and the occurrences of HPAI breakouts during 2004 and 2005,

we see that outbreaks were mostly around wetlands. However, there were few or no occurrences of HPAI around some wetlands.

There are two main approaches to determining the spatial weighting function: using prior knowledge and exploring the spatial patterns of existing data. From the correlation between flyways and the occurrences of HPAI, we deduced the following.

(1) There is high correlation between stopovers on the same flyway or areas around these stopovers and the occurrence of HPAI.

(2) The correlation between stopovers outside the flyways and the occurrence of HPAI depends on the distance from the nearest flyway.

(3) The correlation between stopovers on different flyways is relatively low and should be determined by their spatial connections.

According to the above conclusions, we present rules for determining the spatial weighting function.

(1) $\omega_{ij} = 0$.

(2) If two sites i, j are stopovers on the same flyway, then $\omega_{ij} = 1$.

(3) If at least one of two sites is not on a flyway, we calculate the distance to the nearest stopover for each site and determine the weight according to the sum of the distances (e.g. the inverse of the sum).

(4) In other situations, we calculate the spatial distance between two sites and determine the weight (e.g. the inverse of the Euclidian distance between the two sites).

We set the distance between two stopovers on the same flyway as zero; therefore, the spatial weight is expressed as

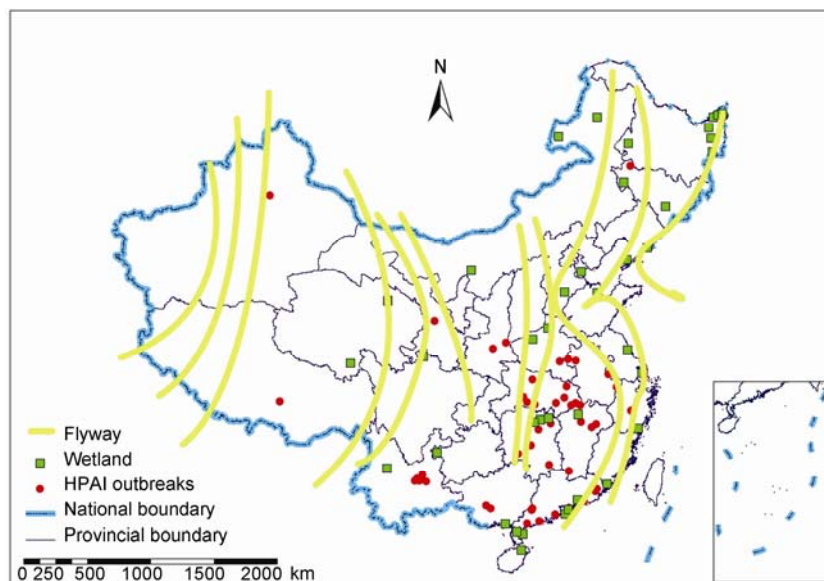$$w_{ij} = \varpi_{ji} = \frac{1}{1+S}. \tag{6}$$

Figure 4 Occurrences of HPAI superimposed on the main flyways and important wetlands (migratory flyways of wild birds are adapted from information collected and analyzed by Wetlands International). Taken from Global Patterns of Influenza A Virus in Wild Birds.

2.5 Temporal weighting function

The spatial pattern is sometimes relevant to an earlier spatial pattern, which is usually seen during the spread of infectious diseases. The reason for this is that the interaction between infective and susceptible individuals develops the spatial pattern in different ways. We can understand this as the historical occurrences of the disease affecting the current status of the disease. In addition, the relationship between two time points is relevant to the interval between them. Intuitionally, the shorter the interval is, the stronger the association is. Therefore, we use a monotonically decreasing function to express the temporal autocorrelation; e.g.

$$w_{ij} = \varpi_{ji} = \frac{1}{e^{|t_i - t_j|}}. \tag{7}$$

2.6 Meta-modeling employing BME

While the binary logistic model can represent and evaluate processes relevant to the epidemic, the BME method can analyze spatiotemporal variability, which equally determines the HPAI risk. Considering the different perspectives of parameters and research application, the optimal solution in this study is to combine the two models in an attempt to make the simulation more accurate than is possible using each model alone. The BME approach is a useful geostatistical method for estimating data with uncertainties by incorporating a multi-disciplinary knowledge base and high-order statistics. All events (outbreaks and potential outbreaks) are regarded as realizations (possibilities) in a random field that has a temporal dimension in addition to two spatial dimensions. All events are divided into two classes: hard data and soft data. Hard data correspond to accurate variables, such as observables and records, labeled $\chi(\rho)=\chi(s,t)$. Soft data correspond to links between hard data and potential realizations, and can be portrayed by a multi-variate probability density function (PDF); e.g.

$$F_\chi(\chi)d_\chi = \text{prob}\left[\chi^- < P < \chi^+ + d_\chi\right], \tag{8}$$

where prob[.] is a probability operator. Thus, the probability of soft data can be calculated according to the cumulative density function. The soft data are established according to multi-disciplinary knowledge, and they are combined with hard data to form a dataset referred to as the specificatory knowledge base. The stochastic expectation of a random field is then defined as

$$E[g(\chi)] = \int d_\chi g(\chi)\text{prob}F_\chi(\chi) = m_\chi, \tag{9}$$

With a known mean, all realizations can be separated into stochastic expectations and residuals; i.e.

$$\chi = \chi_d + m_\chi, \tag{10}$$

where χ_d represents the residuals of the different realiza-

tions. Based on the residuals, we can use covariance to quantify space-time variability. The distance comprises spatial lag and temporal lag. Spatial lag is the spatial distance between two spatial points while temporal lag is the time interval between two points. The spatiotemporal covariance is obtained through a calculation in which the parameters of sill and range have epidemic meanings. Sill refers to the decline in covariance to a constant value and the spatiotemporal range is the range of autocorrelation on spatial and temporal scales. A mathematic model should then be selected to fit the calculated spatiotemporal covariance. To maximize the Bayesian entropy, the BME method calculates the posterior PDF at each estimation realization χ_k on the basis of equations obtained through Bayesian conditionalization:

$$F(\chi_k) = A^{-1}\int d_\chi f(map)F_\chi(\chi), \tag{11}$$

where A is a normalization coefficient and *f(map)* is the weight deduced from the spatiotemporal covariance model.

Even though the binary logistic model passes the statistical *f*-test at a 95% confidence level, it brings plenty of uncertainty to the estimation. Therefore, the result is considered soft data. The HPAI outbreaks are hard data. In general, BME estimation significantly increases the accuracy of the binary logistic model.

3 Results and validations

In our study, eight covariates were initially analyzed. They were proximity to a national highway, proximity to a reservoir, proximity to a lake, proximity to a railway, proximity to a river, the LST, the NDVI and the IWV. The response was the status of occurrence of HPAI (presence or absence). The status was a binary variable; cells that contained an occurrence(s) had a value of 1, and 0 otherwise. Table 1 is a descriptive summary of the covariates and responses.

3.1 Analysis of individual factors

Through analysis of individual factors, six of the investigated eight factors were found to be significantly relevant to the occurrence of HPAI. Table 2 gives the detailed results of analysis. The *P* values of five factors are less than 0.05, which is statistically significant. These factors include all transportation and water body factors. This is consistent with the assumption that the trade of waterfowl and poultry, which is related to transportation, affects the occurrence of HPAI.

3.2 Analysis of results obtained using the binary logistic regression model

Table 3 presents the results of analysis using the binary

Table 1   Description of variables

Variable	Description	Unit	Type
Covariate			
National	Proximity to national highway	km	Continuous
Reservoir	Proximity to reservoir	km	Continuous
Lake	Proximity to lake	km	Continuous
Railway	Proximity to railway	km	Continuous
River	Proximity to river	km	Continuous
LST	Land surface temperature	°C	Continuous
Wv	Water vapor	10^{-1} cm	Continuous
NDVI	Normalized difference vegetation index	10^{-1}	Continuous
Response			
Present	Occurrence(s) or nonoccurrence		Binary

Table 2   Analysis of individual factors

Covariate	Odds ratio	Odds ratio (95% confidence interval)	P value
National	0.949	0.928–0.97	<0.001
Railway	0.995	0.992–0.998	<0.05
Reservoir	0.983	0.974–0.992	<0.001
Lake	0.927	0.900–0.956	<0.001
River	0.997	0.995–0.999	<0.05
LST	0.974	0.927–1.024	0.304
Wv	0.999	0.967–1.031	0.941
NDVI	0.957	0.806–1.137	0.616

logistic procedure of SPSS13. First, we set the P value of the variable sigma to less than 0.001, which strictly controls the number of model parameters. According to the results of the analysis, we can express the final model as

$$p_i = \frac{\exp(1.464 - 0.113\text{national} - 0.011\text{reservoir} - 0.039\text{lake} + 1.849\text{sigmaw})}{1 + \exp(1.464 - 0.113\text{national} - 0.001\text{reservoir} - 0.039\text{lake} + 1.849\text{sigmaw})}. \quad (12)$$

Using eq. (12), with input of the status of disease before the current date and the environment factors in the area of interest, we obtain the probability of occurrence of HPAI at the corresponding site. Dividing Mainland China into an array of cells and calculating the probability of occurrence for each cell, we can predict the occurrence of HPAI in Mainland China.

Figure 5 is a prediction map of the probability of occurrence of HPAI obtained employing the above model. Using the binary logistic model, a simulation closely related to HPAI outbreaks is available for BME approaches. It is clear that the binary logistic model provides prominent environmental determinants of the occurrence of bird flu but intro-

duces uncertainties in regions without disease. The main aim of BME estimation is to eliminate the uncertainty and increase the simulation accuracy.

3.3   Analysis of meta-modeling employing BME approaches

Points within a certain distance of hard data were taken as soft data. The soft dataset has strong correspondence with the environment which impact the HPAI outbreaks, as well as a range of uncertainty. As natural residuals, the uncertainty is distributed normally (Gaussian distribution of the PDF). For the period from January 23 to February 17, 2004, there were 50 hard data and 2238 soft data.

Figure 6 shows the spatial distribution of all data over the research duration. After the expectation was removed, the residuals of all data accord to a normal distribution and are correlated spatiotemporally. We calculate the covariance of the BME three-dimensional random field, and present the results in Figure 7.

After experimental covariance is calculated, we construct a theoretical covariance model that fits it suitably:

$$c_{\chi R}(r = |s - s'|, \tau = |t - t'|)$$
$$= c_1 \exp\left(\frac{-3r}{a_{r1}}\right)\exp\left(\frac{-3\tau^2}{at_1^2}\right) + c_2 \exp\left(\frac{-3r}{a_{r2}}\right)\exp\left(\frac{-3\tau}{at_2}\right), \quad (13)$$

where $c_1 = 0.08$, $c_2 = 0.005$, $a_{ri} = 234.5$ km, $a_{r2} = 45$ km, $at_1 = 22$ d, and $at_2 = 4.6$ d. As our covariance model indicates the epidemic spatiotemporal variability, the parameters have epidemic meaning. The spatiotemporal structure of HPAI indicates that the epidemic risk was influenced by two categories of spatiotemporal patterns: (1) the main spatial and temporal spill-over and (2) the disturbance. Thus, for mapping purposes, we extract a BME mean value for each estimation point chronologically and validate its accuracy.

3.4   Validation of BME meta-modeling

A dataset was selected randomly among HPAI outbreaks and blank spots to validate the model. A total of 55 validation points were picked randomly among the hard and soft data points (see Figure 8).

To measure the mapping accuracy, we use the simplest

Table 3   Analysis using the binary model and considering spatial and temporal functions

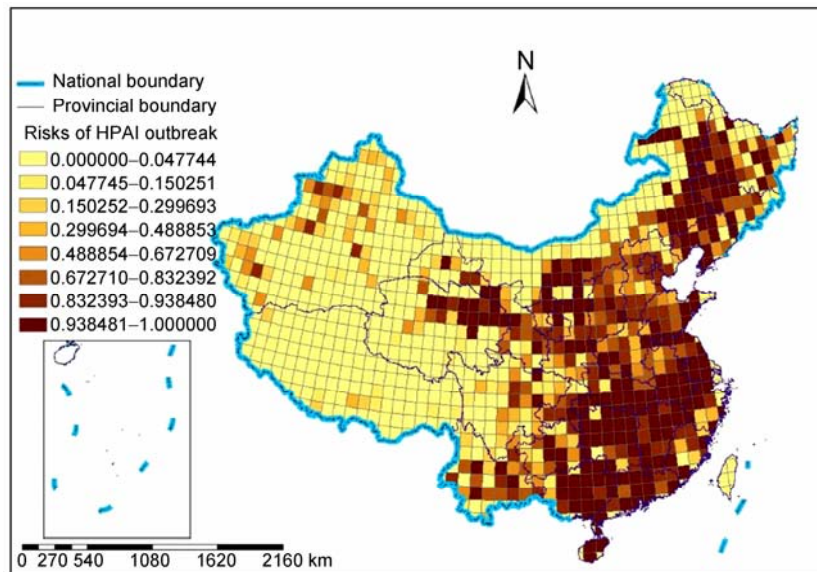| Variable	Coefficient	Odds ratio	Odds ratio (95% confidence interval)	P value
National	−0.113	0.893	0.842–0.948	<0.001
Reservoir	−0.011	0.989	0.976–1.002	<0.1
Lake	−0.039	0.962	0.930–0.994	<0.05
Sigmaw	1.849	6.353	2.502–16.131	<0.001
Constant	1.464	4.325	—	<0.01

Figure 5 Logistic prediction map of Mainland China for January 31, 2004.
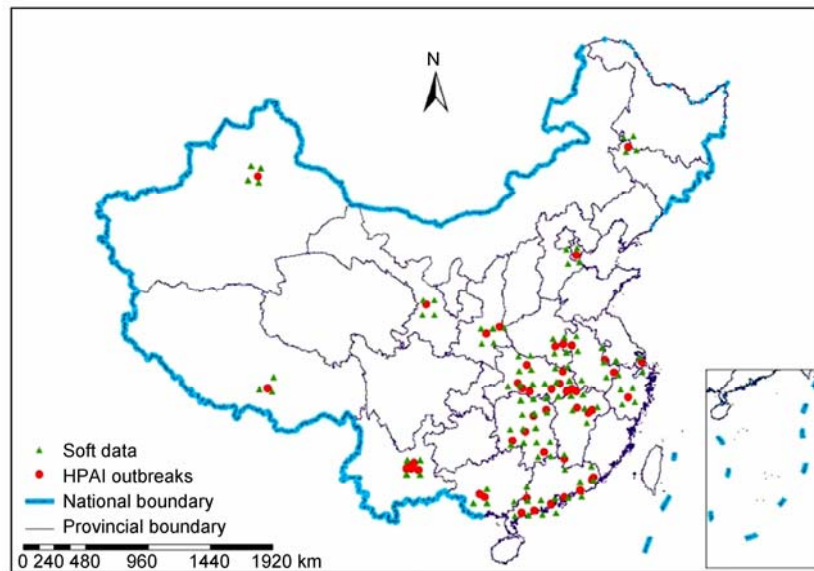

Figure 6 Occurrences of HPAI (hard data) in Mainland China in 2004. Shown are laboratory-confirmed H5N1 cases only. Soft data used in calculations are also labeled on the map.

method to define the error–the difference between the model estimation and observation. While the outbreaks had a value of 1 and non-outbreaks a value of zero, the errors were calculated directly from the difference between the absolute risk probability and observation (see Figure 9). The *X*-axis shows the absolute accuracy of observations. We randomly picked a day for validation (see Figure 10). The results and map were calculated for January 31, 2004. Compared with the result of simply using the binary logistic model, the overall accuracy of meta-modeling was 8.8% better. In addition, Figure 9 shows that the BME method can significantly increase the accuracy stably, and the variance of error decreased by 38.8%.

4 Conclusion and future work

In this paper, we analyzed and predicted the occurrences of HPAI in Mainland China using meta-modeling incorporating binary logistic regression and BME modern geostatistics. Various MODIS products provided fundamental data for the analysis of environmental factors. Through analyzing the relationships of the flyways of migratory birds and the occurrences of HPAI and the characteristics of the spread of HPAI, we defined the spatiotemporal weighting function. Using remote sensing data and GIS technologies, we identified potential environmental variables relevant to the occurrence
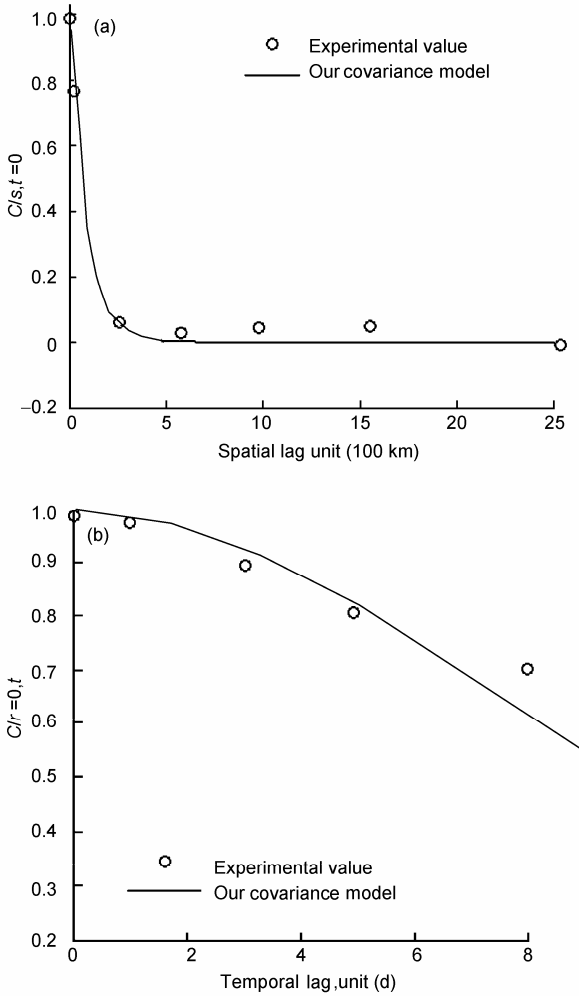
Figure 7 (a) Experimental and model covariance on a spatial scale; (b) experimental and model covariance on a temporal scale.

of HPAI. Finally, a logistic regression model incorporating the defined spatiotemporal weight was built. Using this model, we can determine dominant variables affecting the occurrence of HPAI, and then predict the outbreak of avian influenza A infection, which will help in prevention and control.

Three main factors were determined in our study. Proximity to a national highway, proximity to a reservoir, and proximity to a lake are statistically significant environmental factors relevant to the occurrence of HPAI. A strong connection between the proximity to a water body and HPAI outbreaks is apparent, which might be due to the characteristics of bird migration. Water bodies such as lakes and reservoirs are the main stopover sites of migratory birds. The water bodies constitute the flyways of migratory routes and are essential to some species of poultry. When stopover sites of certain migration routes become hosts for the influenza virus, they are the most probable sites for the severe cross-infection of migratory bird groups. When species of poultry are fed traditionally within an environment of a water body and with exposure to infectious wild birds, there is cross-infection between the migratory birds and poultry. To some extent, this persuades us further that migratory birds, mainly waterfowl, are important infection sources of HPAI, while the role of migratory birds in the spread of HPAI has previously been discussed. This study also demonstrates that proximity to national highways is relevant to the occurrences of HPAI. National highways are important transportation routes for poultry trade. The infected poultry would be contagious to other fowl in transportation, trade, and distribution. Even though the prevalence in poultry is considered to be easier to control, the significance of fowl as an infection source should not be ignored.
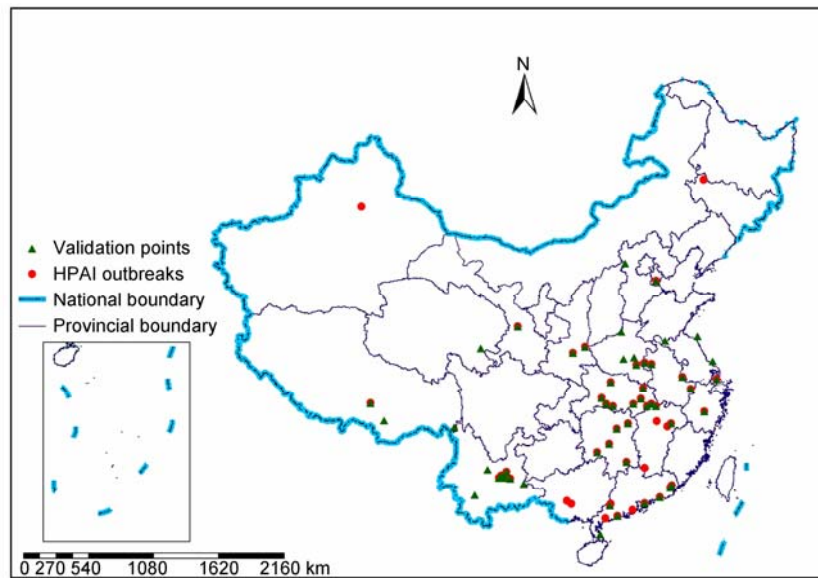
The chosen factors (proximity to a national highway,

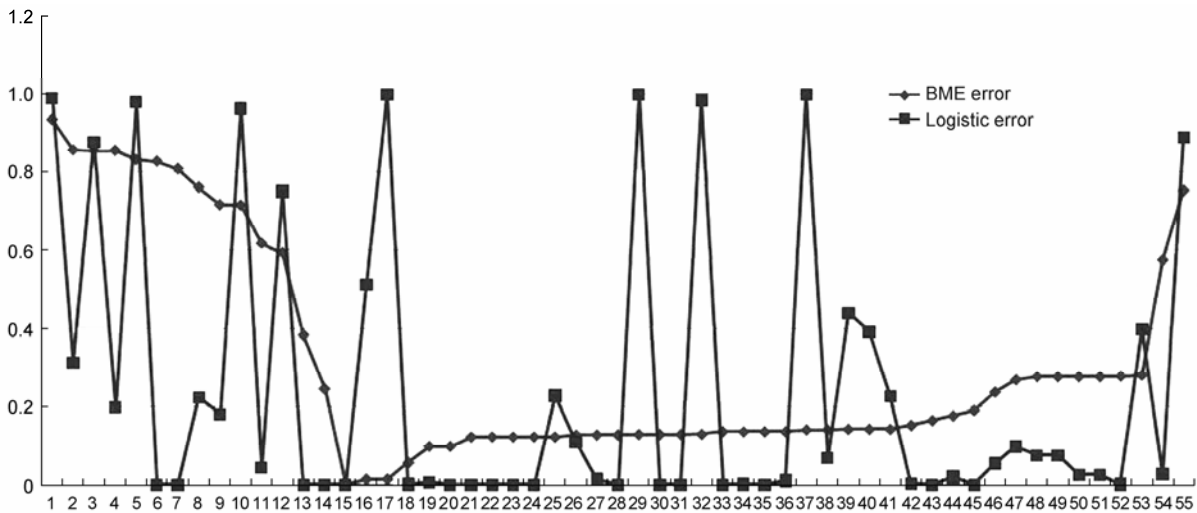**Figure 8** Validation points randomly selected from hard and soft data.

Figure 9 Absolute error of validation points of the logistic model and meta-model showing that the results of the BME estimation are more reliable than those of simple logistic regression.
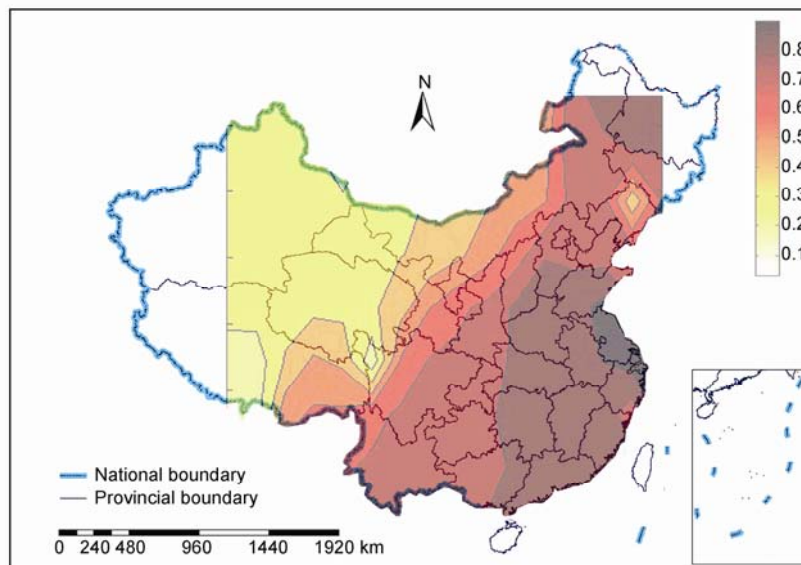

Figure 10 Prediction map of the meta-model for Mainland China on January 31, 2004.

proximity to a reservoir, proximity to a lake, proximity to a railway, proximity to a river, LST, NDVI, and IWV) have relatively high odds ratios to HPAI occurrence. We attribute these phenomena primarily to the strong connections between a water body and migratory birds and nearby poultry and between transportation routes and poultry. Furthermore, the outbreaks were distributed on maps with 1° resolution, and thus, the distance between routes (migratory and transportation) and the overlaid outbreaks would appear to be relatively large, which results in the tolerance of error and higher correlation. Although all of the factors seem useful in the research, autocorrelations between factors should be considered; e.g. locations near reservoirs are always near rivers or lakes. In our study, we chose three factors (prox-

imity to a national highway, proximity to a reservoir, and proximity to a lake) to establish our model because they have relatively low *P*-values. However, there are other, more complicated, factors of the outbreak and spread of HPAI.

Spatiotemporal variability could also be interpreted using three determinants. Two outbreaks within 234.5 km are highly correlated. The distance is possibly the distance between two cities for routine poultry trade or the length of bird migration per day. Within this distance, after an occurrence of HPAI, the transportation of birds has a high probability of spreading the virus to traditional markets (which have open access and confined space) and the flight of birds has a high probability of spreading the virus to poultry

farms near water bodies (where migratory birds stop over for food and rest). This epidemic spatial range has a 45-km fluctuation due to the different distribution patterns of cities and water bodies. In addition, two outbreaks are likely to occur within 22 d. One important reason for the temporal range is the seasonality of HPAI; outbreaks are always in winter and spring. A period of 22 d in these seasons could be a period of stable climate, which increases the odds of virus infection and diffusion. Even though the clarified de-litescence is still controversial, the latent period for HPAI virus survival is longer at lower temperature.

Although this study considered the main environmental factors and the migration of birds in assessing HPAI outbreaks and modeled HPAI for a period in 2004, we could not precisely forecast the evolution of HPAI. As time elapses, the relationship between HPAI and other influential determinants will vary. In future work, we will explore other factors such as the location and scope of a landform, terrain, wind stress and direction, water body temperature, and particularly the socioeconomic situation. We hope to develop a space-time interactive model that has the ability to incorporate various social, economic, and environmental factors to simulate and predict the outbreak and spread of HPAI.

1   Alexander D J. A review of avian influenza in different bird species. Vet Microbiol, 2000, 74: 3–13
2   Alexander D J. An overview of the epidemiology of avian influenza. Vaccine, 2007, 25: 5637–5644
3   Swayne D E, Suarez D L. Highly pathogenic avian influenza. Rev Sci Tech, 2000, 19: 463–482
4   Zhu Q Y, Qin E D, Wang W, et al. Fatal Infection with Influenza A (H5N1) Virus in China. New Engl J Med, 2006, 354: 2731–2732
5   Zhong S B. Application of GIS and remote sensing for study of epidemiology of infectious diseases-case dtudies of hepatitis B and highly pathogenic avian influenza (in Chinese). Dissertation for the Doctoral Degree. Beijing: Graduate University of Chinese Academy of Sciences, 2006. 67–83
6   Chen H, Smith G J, Zhang S Y, et al. Avian flu: H5N1 virus outbreak in migratory waterfowl. Nature, 2005, 436: 191–192
7   Rogers D J, Randolph S E, Snow R W, et al. Satellite imagery in the study and forecast of malaria. Nature, 2002, 415: 710–715
8   Xu B, Gong P, Biging G, et al. Snail density prediction for schistosomiasis control using Ikonos and ASTER images. Photog Eng Rem Sens, 2004, 70: 1285–1294
9   Guo J P, Xue Y, Cao C X, et al. Study on the highly pathogenic avian influenza epidemic using land surface temperature from MODIS data. Int Geosci Rem Sens Symp, 2005, 5: 3599–3602
10 Fang L Q, Cao C X, Chen G S, et al. Studies on the spatial distribution and environmental factors of highly pathogenic avian influenza in China using geographic information system technology (in Chinese). Chin J Epidemiol, 2005, 26: 839–842
11 Gilbert M, Chaitaweesub P, Parakamawongsa T, et al. Free-grazing ducks and highly pathogenic avian influenza, Thailand. Emerg Infectious Dis, 2006, 12: 227–234
12 Brownstein J S, Holford T R, Fish D. Effect of climate change on lyme disease risk in North America. Eco Health, 2005, 2: 38–46
13 Gumpertz M L, Graham J M, Ristaino J B. Autologistic model of spatial pattern of Phytophthora epidemic in bell pepper: Effects of soil variables on disease presence. J Agric Biol Environ Stat, 1997, 2: 131–156
14 Christakos G. Spatiotemporal information systems in soil and environmental sciences. Geoderma, 1998, 85: 141–179
15 Christakos G. Modern Spatiotemporal Geostatistics. New York: Oxford University Press, 2000. 116–215
16 Christakos G, Bogaert P, Serre M L. Temporal GIS. New York: Springer, 2002. 155–166
17 Wang J F, Christakos G, Han W G, et al. Data-driven exploration of "spatial pattern-time process-driving forces" associations of SARS epidemic in Beijing, China. J Public Health, 2008, 30: 234–244
18 Woodward M. Epidemiology: Study Design and Data Analysis. Boca Raton: Chapman & Hall/CRC, 1999. 145–190
19 Hosmer D W, Stanley L. Applied Logistic Regression. 2nd ed. New York: Wiley, 2000
20 Besag J. Spatial interaction and the statistical analysis of lattice systems. J Royal Statist Soc B, 1974, 36: 192–236
21 Cressie N A C. Statist for Spatial Data. New York: Wiley, 1994
22 Kanai K, Ueta M, Germogenov N, et al. Migration routes and important resting areas of Siberian cranes (Grus leucogeranus) between northeastern Siberia and China as revealed by satellite tracking. Biol Conserv, 2002, 106: 339–346