Original article

Identifying spatio-temporal patterns of transboundary disease spread: examples using avian influenza H5N1 outbreaks

Matthew L. Farnsworth[1*], Michael P. Ward[2]

[1] USDA, APHIS, VS, Centers for Epidemiology and Animal Health, 2150 Centre Avenue, Bldg. B, Mail Stop 2W64, Fort Collins, Colorado, 80526-8117, USA
[2] Faculty of Veterinary Science, University of Sydney, Private Mail Bag 3, Camden NSW 2570, Australia

Abstract – Characterizing spatio-temporal patterns among epidemics in which the mechanism of spread is uncertain is important for generating disease spread hypotheses, which may in turn inform disease control and prevention strategies. Using a dataset representing three phases of highly pathogenic avian influenza H5N1 outbreaks in village poultry in Romania, 2005−2006, spatio-temporal patterns were characterized. We first fit a set of hierarchical Bayesian models that quantified changes in the spatio-temporal relative risk for each of the 23 affected counties. We then modeled spatial synchrony in each of the three epidemic phases using non-parametric covariance functions and Thin Plate Spline regression models. We found clear differences in the spatio-temporal patterns among the epidemic phases (local versus regional correlated processes), which may indicate differing spread mechanisms (for example wild bird versus human-mediated). Elucidating these patterns allowed us to postulate that a shift in the primary mechanism of disease spread may have taken place between the second and third phases of this epidemic. Information generated by such analyses could assist affected countries in determining the most appropriate control programs to implement, and to allocate appropriate resources to preventing contact between domestic poultry and wild birds versus enforcing bans on poultry movements and quarantine. The methods used in this study could be applied in many different situations to analyze transboundary disease data in which only location and time of occurrence data are reported.

disease spread / spatio-temporal analysis / epidemic pattern / avian influenza / poultry

1. INTRODUCTION

Knowledge of spread mechanisms informs disease control and prevention strategies. To most effectively control an outbreak of highly pathogenic avian influenza (HPAI) subtype H5N1 in a country, the dominant mechanism of spread needs to be known. Although the potential causes of HPAI spread within national poultry populations are broadly contact with infected wild birds and human-mediated transport of infected poultry and poultry products, simply visualizing the pattern of spread is insufficient for concluding which mechanism might be dominant within a given epidemic [34]. A first step in formulating hypotheses about mechanisms of spread based on outbreak data involves quantifying variations in spread patterns among epidemics. We hypothesize that epidemics mediated by wild bird contact should have a different spatial and temporal structure

* Corresponding author: matt.farnsworth@aphis.usda.gov

than an epidemic in which human spread is dominant. If spatio-temporal patterns can be broadly classified a priori, it might be possible to match them with the spatio-temporal pattern from an epidemic with an unknown mechanism of spread, which will in turn provide information for designing more effective surveillance and response measures.

HPAI virus subtype H5N1 was first identified in Hong Kong in 1996. Prior to 2005, infection was confined to east and south-east Asia. More recently, it has spread across Asia to the Middle East, Eastern and Western Europe, and Africa, causing numerous disease outbreaks in domestic poultry and wild bird populations [1]. HPAI subtype H5N1 is considered a threat to world health because of its potential to infect and cause morbidity and mortality in humans. There is a fear that it could become the next pandemic influenza strain [1].

Avian influenza virus infection is endemic in a range of free-living bird species world-wide [1], particularly species associated with water. Migratory waterfowl might be responsible for the spread of influenza A viruses between regions [22, 27], and HPAI and low pathogenic avian influenza (LPAI) viruses in poultry are often assumed to occur from exposure to wild avian species [17, 20, 24, 25, 31]. Infected migratory waterfowl have been suspected of spreading HPAI virus subtype H5N1 from central Asia to Eastern Europe during the second half of 2005, based on migratory flyways [13]. However, the relationship between infected migratory waterfowl and outbreaks of HPAI and LPAI in poultry remains mostly speculative. The transport of infected poultry and contaminated poultry products have been blamed in some cases for spreading the disease, for example within China [11]. In other cases, the spread of HPAI virus subtype H5N1 has been more consistent with trade routes than with migration flyways [13, 21]. The spread of avian influenza virus via smuggled poultry carcasses and wildlife has been documented [3, 30]. By its very nature, the role played by smuggling and the illegal movement of poultry and poultry products is very difficult to confirm.

In this study, we use three different modeling approaches to quantify spatio-temporal outbreak dynamics among three phases of an epidemic of HPAI subtype H5N1 within a national poultry population. Our goal was to illustrate how such models can be used to generate hypotheses regarding underlying mechanisms of spread. First, we used a set of Bayesian hierarchical models with spatial and spatio-temporal random effects specified by an intrinsic conditional autoregressive structure [4, 5] to identify variations in the patterns of spatio-temporal relative risk among three epidemic phases of HPAI subtype H5N1 outbreaks in village poultry in Romania, 2005–2006. Second, we modeled the space-time correlation, or spatial synchrony [6, 7], of outbreaks within each of the three epidemic phases to identify variations in their spatio-temporal patterns. The patterns of synchrony result from fitting smoothing splines to a data set that reflects the daily outbreak status of each village throughout the course of each phase of the epidemic. This provides an estimate of the correlation between outbreak status for every village on each epidemic day as a function of the spatial separation among all villages within a given outbreak phase. The resulting spatio-temporal patterns were then used to provide insight into the intervening epidemic phase for which the mechanism of spread was most uncertain. Finally, we fit a surface, known as a Thin Plate Spline, to the data to visualize and quantify variations in spread patterns among epidemic phases. These three modeling approaches make clear the distinction in space-time dynamics among outbreak phases and provide an entry point to formulate hypotheses about underlying mechanisms of spread.

2. MATERIALS AND METHODS

2.1. Data source

Between October 2005 and June 2006, 165 outbreaks of HPAI subtype H5N1 were reported in 161 village poultry populations. Outbreaks were reported (Fig. 1) from 23 of the 41 counties in Romania, covering about half (128 000 km^2) the area of the country. Disease control measures used in this epidemic included quarantine of affected

villages, flock depopulation, disinfection, placement of serologically-monitored sentinel flocks prior to village restocking, and national movement controls. Vaccination was not used to control this epidemic[1].

Data available included the outbreak location (X, Y values using the stereographic 60 coordinate system with distances in meters) and the reported date of occurrence. Outbreak locations represented the village in which HPAI subtype H5N1 was identified. Since this is a backyard production system, the entire village flock was assumed to be infected (separate backyards were not differentiated). The coordinate information represents the official geographic location of an affected village.

The epidemic curve of reported outbreaks has been analyzed previously by Ward et al. [34]. Three epidemic phases were apparent: 7 October–20 December 2005 (days 1–75; 23 outbreaks), 7 January–20 March 2006 (days 93–165; 28 outbreaks), and 12 May–6 June 2006 (days 217–243; 110 outbreaks). Half of all outbreaks were reported during an 85-day period, between 27 February and 23 May (epidemic days 144–229).

Previous analysis of this epidemic provided some insight into the pattern of HPAI subtype H5N1 occurrence in space and time [34]. Outbreaks first appeared in eastern and southern Romania, particularly within an area that forms part of the Danube River Delta, and then spread in an east to west direction to central Romania. It was suggested that the evolution of the epidemic consisted of disease introduction, local spread and sporadic outbreaks caused by interaction between wild birds and domestic poultry (phase 1), and long-distance disease spread with rapid epidemic propagation (phase 3) caused by human-related factors [34]. Outbreaks during the intervening period (phase 2) might have been caused by either of these two hypothesized mechanisms of spread, or a mixture of them. Here we extend the above work to characterize the spatio-temporal patterns of the epidemic phases and to develop hypotheses about potential mechanisms of spread.

2.2. Data analysis

We used three approaches to characterize the spatio-temporal patterns. The first analysis used a hierarchical Bayesian model to quantify the spread of disease throughout the entire epidemic, at the county scale of analysis. We considered a generalized

linear mixed model for the number of outbreak villages in each county. For each county, we modelled the probability of observing a given number of outbreak villages as a function of a linear trend in time (fixed effect term in the models) and three random effect terms, which account for any unobserved covariates as well as the spatial, non-spatial, and spatio-temporal patterns in the outbreak data.

This generalized linear mixed model can be described in three stages: the data model, or likelihood linking the data to the model parameters; the process model relating the covariates and random effects to the parameters, and the prior distributions for all model parameters [35]. Our interest focused on the posterior distribution, the distribution of the process and parameters after being informed by the data. For many ecological problems, the high dimensionality of the model can prohibit the use of standard methods. However, Markov Chain Monte Carlo (MCMC) [12, 14] techniques allowed us to estimate the posterior distribution of interest.

2.3. Data model

The data model relates the number of outbreak villages in each county and week to the probability of occurrence, adjusted by the total number of villages in each county. Let Y_{ij} be the number of infected villages in week $i = 1,\ldots,n$ in county $j = 1,\ldots,k$. We assume that the number of villages experiencing an outbreak is Poisson distributed with parameter λ_{ij}, adjusted for the total number of villages per county, n_j:

$$Y_{ij}\,|\,\lambda_{ij}n_j \sim \text{Poisson}\left(\lambda_{ij}n_j\right) \qquad (1)$$

where λ_{ij} is the probability of observing Y_{ij} outbreaks in county j on week i. All observations are assumed to be conditionally independent given this parameter.

2.4. Process model

The process component of the model relates the probability of observing a given number of outbreaks, λ_{ij}, to the linear trend in time and three random effect terms that capture the non-spatial, spatial, and spatio-temporal patterns in the outbreak data. To model the Poisson distributed outbreak probability we use a standard log-linear transform. Thus we model the probability of observing a given number of outbreaks in each county and week as:

$$\log\left(\lambda_{ij}n_j\right) = \mu + v_{ij} + \gamma_{ij} + \beta \cdot t_i + \gamma_{ij} \cdot t_i \quad (2)$$
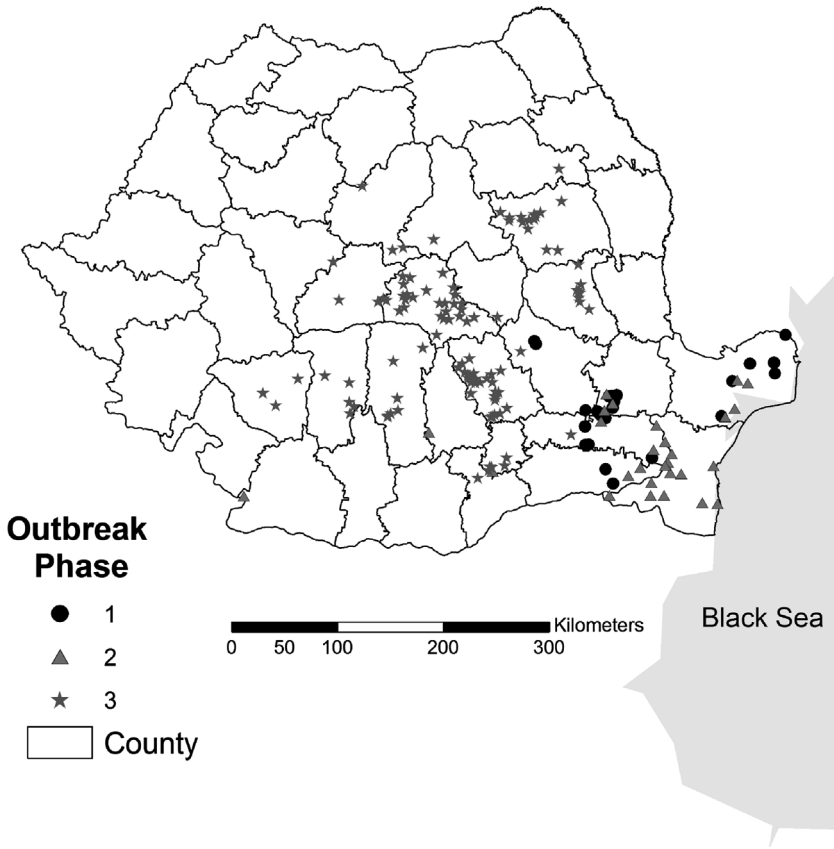
Figure 1. Romanian village poultry outbreaks of HPAI subtype H5N1 between 2005 and 2006.

where μ is the background infection rate common to all counties and times; v_{ij} is an independent random effect term (HET) associated with the ith week in county j; γ_{ij} is the spatial random effect term (SPACE) for ith week in county j; β is a fixed effect modeling the linear trend of the outbreak in time (TIME); and $\gamma_{ij} \cdot t_i$ is an interaction random effect between space and time (SPACE_TIME). All model terms are described below.

2.5. Prior and posterior distributions

Because our analysis is fully Bayesian, we specify prior distributions for all model parameters in the hierarchy. The independent random effect, v_{ij}, corresponds to a latent process operating independently in each county and week. We let $v_{ij} \sim$

$i.i.d.$ $N(0, \sigma_v^2 I)$ for $i = 1, \ldots, n$ and $j = 1, \ldots, k$ where I is an $n_i \times n_j$ indicator matrix. This component models the overall unstructured heterogeneity in the data by assuming no relationship among neighboring villages or weeks, but with a variance that is common to all counties and weeks.

The spatial component, modelled by γ_{ij} describes the latent, or unobserved, transmission process among counties in each week resulting in the spatial structure of the outbreaks each week. We specify the spatially structured variation in outbreak probabilities, γ_{ij}, via an Intrinsic Gaussian Conditional Autoregressive (ICAR) model [4]. For each week i and county j in our problem, the ICAR model states that γ_{ij} is related to the γ terms for the neighboring counties; and, given the γ terms for the neighboring counties, each county is independent of all other counties outside the local neighborhood. Specifically, let the

set of neighbors of county j be denoted by $j+$. Then, for each week i in county j, we assume the conditional relationship

$$\gamma_{ij}|\gamma_{ij+} \sim \text{Normal}\left(\frac{1}{n_{j+}}\sum_{i\,\text{in}\,j+}\gamma_i, \frac{\sigma_\gamma^2}{n_{j+}}\right) \quad (3)$$

where n_{j+} is the number of neighbors of county j and σ_γ^2 is the variance common to all counties. Thus, the conditional mean of γ_{ij} is simply the average value of its neighbors γ_{ij+}, with conditional variance $\frac{\sigma_\gamma^2}{n_{j+}}$ inversely proportional to the number of neighbors. In the conditional mean in Equation (3), the neighboring counties are equally weighted so that all neighbors of county j influence it equally. Spatial variation in our model is limited to counties sharing a border; however there are no a priori restrictions on specifying the neighborhood structure or county weights. We chose to use only the neighboring counties instead of including a larger neighborhood because we wish to maintain a sharp distinction between local dependency and global unstructured heterogeneity, which becomes increasingly blurred as the local neighborhood is extended. The linear trend in time, $\beta \cdot t_i$, is the rate ratio between two consecutive weeks and provides and estimate of the rate at which the number of outbreaks changes between weeks. We let $\beta \sim N(0, 10\,000)$ be the prior distribution for the time trend parameter. Finally, the space-time interaction random effect, $\gamma_{ij} \cdot t_i$, models the interaction between space and time during the course of the epidemic.

All models were fit using WinBUGS [23] software and an MCMC procedure for each model run for 50 000 iterations after a burn-in period of 10 000 iterations to ensure convergence of all model parameters. Convergence diagnostics and autocorrelation statistics were used to assess the mixing of the chains and the MCMC sampling quality for each parameter.

We used a Deviance Information Criteria (DIC), a generalization of the Akaike Information Criteria (AIC), to compare the set of candidate models [29]. These criteria are based on the deviance, $D(\theta) = -2 \ln L$, where L is the likelihood and θ is the vector of model parameters, and a penalty for model complexity. For AIC, the penalty is two times the number of parameters in the model. The complexity of a hierarchical model is measured by the effective number of parameters (p_D), which can be smaller than the total number of parameters. This complexity is defined as $p_D = \overline{D(\theta)} - D(\bar{\theta})$, where $\overline{D(\theta)}$ is the

expected deviance over the posterior distribution of parameter vector θ taken across all MCMC samples, and $D(\bar{\theta})$ is the deviance evaluated at the posterior mean of the parameter vector. Finally, $\text{DIC} = \overline{D(\theta)} + p_D = 2\overline{D(\theta)} - D(\bar{\theta})$. Smaller values of DIC indicate a better-fitting model. As with other penalized likelihood criteria, DIC is a method for comparing a collection of alternative models [9].

Burnham and Anderson [8] derived Akaike weights that, when normalized, can be interpreted as a set of weights that sum to one and estimate the probability that model r is the best Kullback-Leibler model for the data at hand, given the set of models considered. This approach provides a method for assessing model selection uncertainty. An analogous approach based on DIC has been suggested to assess model selection uncertainty within a Bayesian modelling context [29]. We used DIC weights (w_{DIC}) to estimate model selection uncertainty for each model r in the candidate set, calculated using the formula for Akaike weights

$$w_{DIC} = \frac{\exp(-\frac{1}{2}\Delta DIC)}{\sum \exp(-\frac{1}{2}\Delta DIC)} \quad (4)$$

where ΔDIC was the difference between the minimum DIC value in the candidate set and model r, and the denominator was the sum over all models in the set under consideration. The DIC weights are an informal measure and allow easier comparison between models than the DIC value itself.

For the second analysis we used smoothing splines to examine the pattern of spatial synchrony, $\tilde{\rho}(\delta)$, among outbreak villages for each phase of the epidemic. Synchrony among outbreak villages in each phase of the epidemic was quantified using non-parametric spatial covariance functions (package ncf v1.08 within R v2.6; [6, 7]) that allow examination of the decay in spatial correlation for temporally fluctuating spatially-referenced populations, from local to large regional scales, by calculating correlations for all pairs of time series at numerous distances and then fitting a smoothing spline. The non-parametric spatial covariance function is given by;

$$\tilde{\rho}(\delta) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}G(\delta_{ij}/h)(\hat{\rho}_{ij})}{\sum_{i=1}^{n}\sum_{j=1}^{n}G(\delta_{ij}/h)} \quad (5)$$

where G is the kernel function with bandwidth h, δ_{ij} is the distance between villages i and j, and ρ_{ij} is the correlation among outbreaks between locations i and j. The kernel function used in this application is the cubic B-spline [26] as it adapts better to

irregularly spaced data than many regression kernels [19] and is known to provide consistent estimates of the covariance function. The asymptotic kernel function for the cubic B-spline [15] is:

$$K(u) = \frac{1}{2} \exp\left(-\frac{u}{\sqrt{2}}\right) \sin\left(-\frac{u}{\sqrt{2}} + \frac{\pi}{4}\right) \quad (6)$$

where u equals δ_{ij}/h in the current case. This asymptotic curve is nearly bell-shaped, except for the tails which vanish rapidly in a sinusoidal manner. The bandwidth, h (> 0), is the parameter that adjusts the smoothness of the fitted curve. This parameter is analogous to the distance-class width used in defining spatial correlograms. The optimal bandwidth is found by standard means of least squares cross-validation.

Synchrony, which can be decomposed into local and regional components, is the correlation between when and where outbreaks occurred. Local synchrony is defined as the average correlation between time series as the distance approaches zero. Regional synchrony is the average correlation across the study area. We used daily outbreak status of each village as the temporal resolution for fitting the smoothing splines. The covariance functions, and hence estimates of regional synchrony, were estimated across the entire area of each epidemic phase.

Finally, we fit a set of models using a technique known as Thin Plate Spline regression (package Fields v5.01 within R v2.6) [2, 16, 18] for each outbreak phase to produce maps of spatio-temporal spread. This approach differed from hierarchical Bayes mapping by estimating the direction and magnitude of spread for each outbreak village. Unlike Kriging, this interpolation method does not require fitting a variogram model to the data, but instead uses a generalized cross-validation procedure (minimization) to identify the optimal level of smoothing (see [18] for a thorough review of Thin Plate Splines). This approach also differs from Kriging by using non-linear polynomial equations to fit an interpolated surface to the data, resulting in directional derivatives which model the direction of spread for each outbreak location and time.

Each Thin Plate Spline represents a spatial covariance function produced by fitting the outbreak data to a set of radial basis functions whose parameters are found through application of a generalized cross-validation procedure. The objective is to find a suitably smooth function, f, which minimizes:

$$\sum_{i=1}^{n} [y_i - f(x_i)]^2 + \lambda J_m(f) \quad (7)$$

where y_i is the observed outbreak day at location x_i; $J_m(f)$ quantifies the roughness of f which itself is defined in terms of the mth partial derivatives of f; and λ is a positive value known as the smoothing parameter. This parameter controls the amount of data smoothing and its value is determined by generalized cross-validation, which measures the predictive error of the fitted function, f. Generalized cross-validation proceeds by removing each data point in turn and summing the square of the difference of each omitted data value from the Thin Plate Spline fitted to all remaining data points [10]. The solution to Equation (7) can be written as:

$$f(x) = \sum_{j=1}^{M} a_j \varphi_j(x) + \sum_{i=1}^{n} b_i \psi(r_i) \quad (8)$$

where φ_j are a set of M polynomials and ψ is a scalar function of the Euclidean distance r_i between x and x_i. The coefficients a_j and b_i are estimated as part of the generalized cross-validation procedure. Directional derivatives are calculated from this model via finite differencing between an outbreak location and its nearest neighbor in time and space, where the steepness of the gradient based on the differencing estimates the rate of spread.

We formulated multiple candidate models by varying the number of polynomial terms (using first, second, third, and so on, order polynomials) in each Thin Plate Spline regression model and used AIC to select the best approximating model in the candidate set of polynomial equations. This best approximating Thin Plate Spline model was used to visualize the estimated direction and magnitude of spread among epidemic phases with the goal of determining if substantial differences in the Thin Plate Spline patterns existed between outbreak phases. Thus, we used this model to produce a spatially explicit representation of the different hypothesized mechanisms of spread among outbreak phases.

3. RESULTS

The hierarchical Bayes model selection results (Tab. I) highlight the importance of including various spatial and temporal components in models designed to estimate the spatio-temporal patterns of disease spread when multiple mechanisms might drive the dynamics. In particular, the weight of evidence for Model 1, the best model which contained the full suite of spatial and temporal effects, was very high

Table I. Model selection results to identify candidate models best explaining observed spatial and temporal patterns of HPAI subtype H5N1 outbreaks in Romanian villages, 2005−2006. The models incorporated various combinations of temporal (Time), spatial (Space), and unstructured heterogeneous (Het) variation. Time represents the week of an observed outbreak within villages for each of the 23 counties that experienced outbreaks, Space is the spatial random effect, and Het is the unstructured variation in the model, p_D = effective number of parameters, DIC = Deviance Information Criteria, w_{DIC} = DIC weight that informally quantifies model selection uncertainty.

Model	Model	p_D	DIC	w_{DIC}
1	Space + Het + Time + Space × Time	33.0	789.71	0.985
2	Space + Time + Space × Time	36.0	798.13	0.015
3	Space + Het + Time	19.3	949.41	0
4	Het + Time	19.3	949.51	0
5	Space + Time	21.3	954.16	0
6	Time	2.0	1069.46	0
7	Space + Het	18.44	1076.27	0

(0.985). Although the second best model included a spatio-temporal interaction term it received very little support from the data (based on w_{DIC}) because it did not account for overall unstructured spatial heterogeneity (HET) in the epidemic. Although the mean relative spatio-temporal risk (Fig. 2), which quantifies the relative risk of an outbreak occurrence for each county as the epidemic progressed through each week, increased as the epidemic progressed, there were clear differences in the risk among epidemic phases. In particular, the relative spatio-temporal risk for combined epidemic phases 1 and 2 ranged from 0.913 to 1.165 (0.252 units), whereas the relative spatio-temporal risk for phase 3 ranged from 1.184 to 1.352 (0.168 units). These estimates correlate with the visualization of the epidemic (Figs. 1 and 2): epidemic phase 3 evolved much more rapidly than phases 1 and 2.

Within epidemic phase 1, the greatest synchrony (Fig. 3) occurred at the local spatial scale (within approximately 50 km). The function declined smoothly, as expected for a typical spatially structured epidemic process. Clearly the pattern from phase 3 (Fig. 3) of this epidemic (predominantly in central Romania) was very different: local synchrony started close to 0 and gradually increased with increasing distance, reaching a peak at approximately 260 km. Local synchrony (the average correlation between outbreaks as the distance

approaches zero) was 0.04 and 0.001 for phases 1 and 3, respectively, whereas regional synchrony (the average correlation across the study area) was approximately 0.01 and 0.06 for phases 1 and 3, respectively. Thus, the spread dynamics were more of a locally correlated process in the first epidemic phase (80 versus 20%, local versus regional), and a much more regionalized process in the third epidemic phase (2 versus 98%, local versus regional). The pattern of synchrony for phase 2 of this epidemic was more closely aligned with that of phase 1, with local and regional synchronies of 0.06 (75%) and 0.02 (25%), respectively. Because the local synchrony for phase 2 was markedly higher than its regional synchrony, we hypothesize that the mechanism of spread was likely more similar to the mechanism driving phase 1 of the epidemic.

A fourth order polynomial produced the best fitting Thin Plate Spline interpolation surface. Figures 4 and 5 show the clear contrast in outbreak dynamics between phases 1 and 2 combined versus phase 3, respectively, based on both the underlying surface and the arrows representing the estimated direction and relative rate of spread among outbreak village locations. The color scale represents the interpolated outbreak day, with dark blue being the earliest outbreak day and dark red representing the final days of each phase. The length of each arrow is calculated using finite differencing as
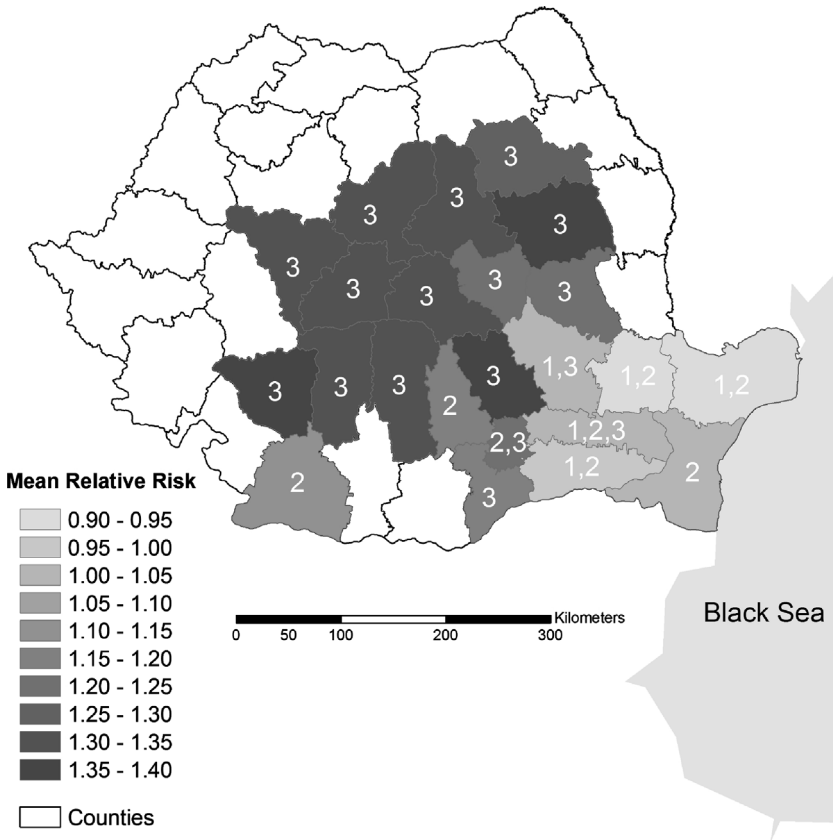
Figure 2. Map of the posterior mean spatio-temporal relative risk of HPAI subtype H5N1 outbreaks from the best (Model 1) hierarchical Bayesian spatial Conditional Autoregressive Poisson model fit at the county scale of analysis. This component models the relative risk of an outbreak as the epidemic unfolds, with darker shades representing greater risk through time. Numbers in the figure indicate which epidemic phases were associated with each county, with some having more than one phase involved. Note the relatively limited and protracted spatio-temporal relative risk for counties associated with phases 1 and 2 of the outbreak, while the spatio-temporal interaction trend is wide spread and rapid across most counties associated with phase 3.

explained above and represents the estimated direction and rate of spread, with longer arrows corresponding to a slower rate of spread.

The combined phases 1 and 2 surface shows an estimated spread pattern clearly following a north to south gradient as the epidemic evolved: the arrows are all of similar length, suggesting a constant rate of spread, and all are oriented from north to south, highlighting the unidirectional spread pattern. In contrast, the phase 3 surface shows no clear directional gradient of spatial

spread over time. Apparently this phase originated in the center of the distribution of outbreak villages (shown in blue) and then radiated out rapidly in multiple directions, visualized here by the discontinuous pattern of patches sharing the same color. In addition, the varying length and orientation of the directional derivative arrows suggests inhomogeneous spread dynamics. For example, clusters of shorter arrows that represent rapid disease spread imply a local source of relatively explosive dynamics.
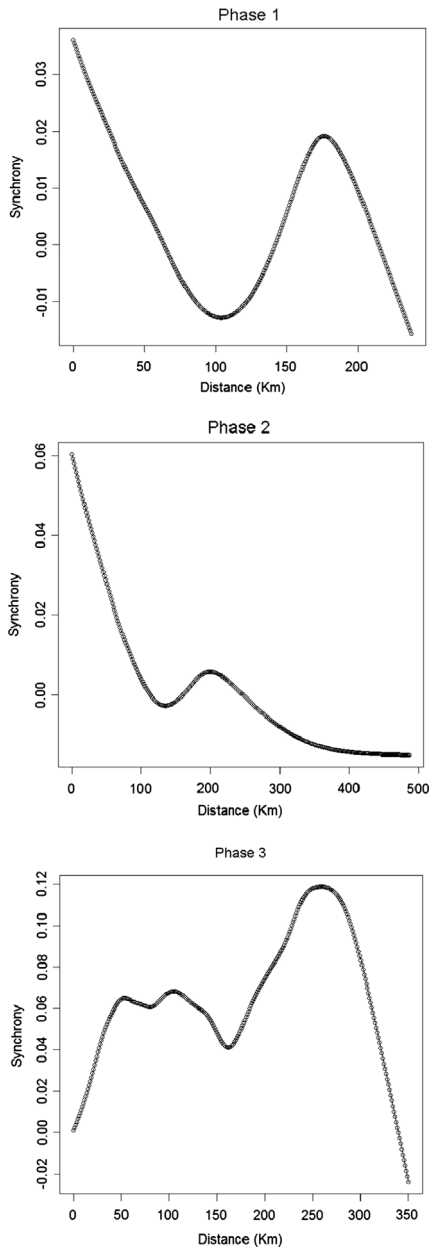
Figure 3. Spatial synchrony, calculated using Equation (5), in each of three epidemic phases of HPAI subtype H5N1 outbreaks in Romanian poultry, 2005−2006. Non-parametric smoothing spline covariance functions were fit to outbreak data in each epidemic phase to compare patterns of synchrony. Synchrony is calculated by Equation (5). Phases 1 and 3 exhibit clear differences in the pattern of spatial synchrony and are hypothesized as having two different mechanisms of spread. The pattern of synchrony in phase 2 is very similar to phase 1.
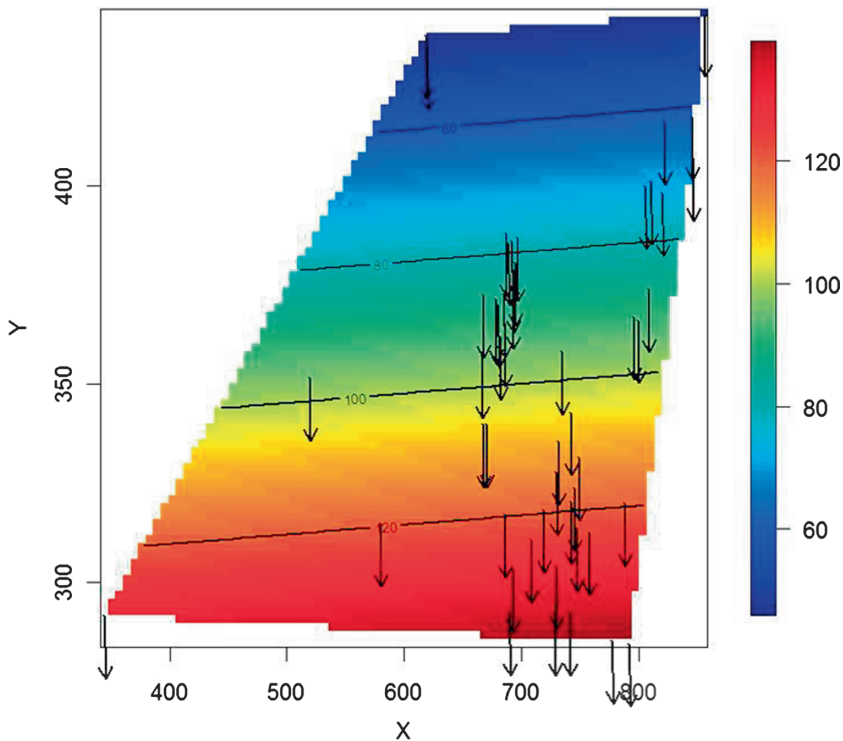
Figure 4. Thin Plate Spline regression model for combined phases 1 and 2 of an epidemic of HPAI subtype H5N1 in Romanian poultry, 2005–2006. The color scale represents the interpolated outbreak day, with dark blue being the earliest outbreak day and dark red representing the final days of each phase. The arrows show the estimated direction and rate of spread, with longer arrows corresponding to a slower rate of spread. Note the unidirectional and relatively homogeneous rate in the pattern (north-south gradient) of spread shown by the arrows and underlying surface as the epidemic evolved across these two protracted epidemic phases.

Locations with relatively few, longer arrows represent slower expansion of the epidemic involving only a few villages. The spatio-temporal pattern of this epidemic phase is not characteristic of an epidemic in which landscape factors drive the evolution of the epidemic; rather, the pattern could be explained by human-mediated spread.

4. DISCUSSION

Information on how infectious diseases spread through susceptible populations is critical for designing control and surveillance programs. For example, the spread of HPAI subtype H5N1 via migratory waterfowl poses very different challenges for disease control (including the prevention of contact between wild birds and domestic poultry) compared to the movement of infected domestic poultry as the disease spread mechanism. Control is much more effective, and efficient, if programs are based on the mechanism of disease spread. The design of surveillance programs (for example, the species that need to be sampled, the spatial and temporal scale of sampling, and sample size) also depends on knowledge of the mechanism of disease spread.

Different modes of spread are expected to produce characteristic spatio-temporal patterns of disease occurrence. Here we have shown how models that capture these patterns can be used to formulate hypotheses regarding
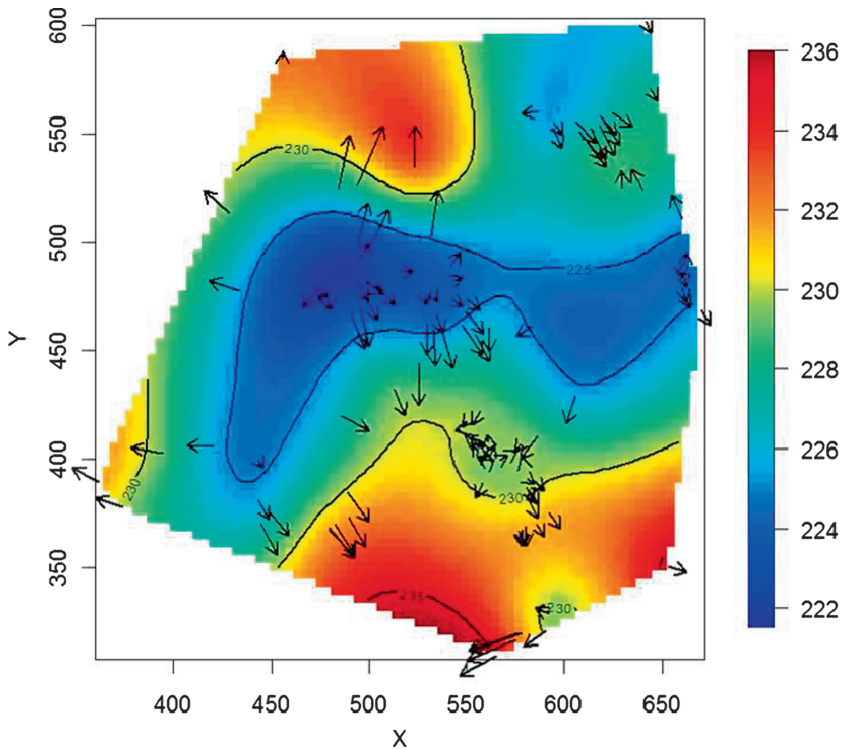
Figure 5. Thin Plate Spline for phase 3 of an epidemic HPAI subtype H5N1 outbreaks in Romanian poultry, 2005–2006. The color scale represents the interpolated outbreak day, with dark blue being the earliest outbreak day and dark red representing the final days of each phase. The arrows show the estimated direction and rate of spread, with longer arrows corresponding to a slower rate of spread. The lack of uniformity in the surface suggests an epidemic originating in the center of the region, followed by rapid expansion in multiple directions occurring over a relatively short period of time. Note that the arrows vary markedly in both length and orientation. Clusters of small arrows represent local rapid spread and long arrows indicate slower disease spread, frequently occurring in disparate locations on similar epidemic dates.

mechanisms of spread. The areal based hierarchical Bayesian models used in this study were able to differentiate between these patterns, even though the data were represented at a coarse level and only included information on date of reported disease occurrence and location. The ability to apply analytical methods to field data that might lack resolution and information on important risk factors is a key attribute of any method selected. The methods used in this study could be applied in many different situations to analyze reported transboundary disease data in which only location and time of occurrence data are reported.

The 2005−2006 epidemic of HPAI subtype H5N1 in Romania consisted of three phases [34]. Based on analysis of the data [32, 33], it was suggested that the first phase of the epidemic was caused by contact between infected migratory wild birds and domestic poultry whereas the third (last) phase was probably caused by the movement of infected domestic poultry. In the present study, the small change in spatio-temporal relative risk over the course of phase 3 of this epidemic (when compared to the combined phases 1 and 2 estimates) indicates a rapidly spreading pattern occurring over a wide geographic scale. These results are

supported by the map of spatio-temporal risk (Fig. 2), in which the combined phases 1 and 2 outbreaks show a relatively limited geographic extent and protracted epidemic duration. Furthermore, a simple spatio-temporal structure was insufficient for modeling the disease spread dynamics across all three phases of this epidemic. For our candidate set of models the combination of multiple spatial, both structured and unstructured random effects, linear trend in time, and spatio-temporal random effects resulted in maximizing the information content in the data. Thus, using only simple models to account for complex epidemic patterns, such as variations in spatial and temporal rates of spread across several phases of an epidemic, can result in poor inference regarding spatio-temporal epidemic processes and hypothesized mechanisms of spread. Prior to attempting to fit such complex models, a thorough understanding of the epidemic (provided by a combination of spatial statistics and geostatistical methods) is needed [34].

The non-parametric smoothing spline analysis further illustrates how quantifying differing spatio-temporal patterns of spread can lead to hypotheses about underlying mechanisms in epidemics in which the cause of disease spread is unclear. Applying this approach, we were able to use the spatio-temporal patterns from all three phases of the epidemic to develop hypotheses about the possible spread mechanism. Because almost all phase 1 outbreaks occurred in association with the Danube River (Fig. 1), and the initial outbreaks in October and November, 2005 occurred in the Danube Delta, a location and season in which large concentrations of waterfowl are known to occur [28], it was suggested that these outbreaks might have been caused by contact between domestic poultry and migratory waterfowl [34]. Further analysis using a spatial regression lag model found that the distance between the nearest migratory waterfowl site was the best predictor ($r^2 = 0.425$) of outbreak timing, but only during phase 1 of the epidemic, supporting the hypothesis that infections of village poultry in Romania during the autumn of 2005 might have occurred via exposure to migratory populations of waterfowl. In contrast to phase 1,

phase 3 (68% of all outbreaks) was of very short (26 days) duration. Although it is clear that these epidemic phases differed temporally and spatially, and it has been hypothesized that phase 1 was wild bird mediated and phase 3 was due to the illegal movement of poultry, until now there has not been a comparison of the spatio-temporal patterns of these phases to relate them to the putative mechanisms of spread.

Phase 1 of the epidemic exhibited the greatest spatial synchrony at the local scale, which remained a declining function of distance until approximately 100 km (the increase in this function at a distance of approximately 175 km can be attributed to an isolated outbreak that occurred during the middle of phase 1). This decaying spatial synchrony indicates wave like spread of HPAI subtype H5N1, as might be expected from a disease transmitted by contact with wildlife concentrated at specific sites. In contrast, phase 3 of the epidemic showed the greatest spatial synchrony at the regional scale, suggesting simultaneous long-distance translocation of HPAI subtype H5N1, reaching its maximum at approximately 260 km. Phase 2 showed the same general pattern as phase 1: a smoothly decaying function of distance, to a distance of approximately 150 km. The similar pattern of spatial synchrony between the first two phases suggests the possibility of a common mechanism of spread, in this case it may be contact between infected wild birds and domestic poultry and potential local spread among villages. In contrast, the increasing spatial synchrony with increasing distance observed in phase 3 suggests simultaneous outbreaks across a large region, which may be indicative of rapid illegal poultry movement from initial outbreaks to distant locations.

Visualization of the epidemic data using Thin Plate Spline regression further reinforced the non-parametric smoothing spline analysis by illustrating the contrast between epidemic dynamics in phases 1 and 2 versus phase 3. It is evident that phases 1 and 2 of the epidemic propagated in a methodical, and spatially uniform, manner, moving from north to south at a relatively constant and slow rate. We suspect that this pattern may be a result of several

introductions of HPAI subtype H5N1 by wild migratory waterfowl within the Danube River delta region. In contrast, the chaotic pattern of spread illustrated by the Thin Plate Spline surface for phase 3 demonstrates rapid epidemic propagation in multiple directions and at rates that vary by location. We hypothesize that this very different pattern is indicative of rapid movement of infected poultry over relatively large distances, consistent with human-mediated spread. Further study is necessary to make definitive statements regarding the potential pathways of disease spread.

The control of epidemic phases with different mechanisms of spread requires the application of very different programs (for example, an emphasis on preventing contact between domestic poultry and wild birds versus quarantine and enforcement of movement controls). Also, different surveillance programs might be appropriate for different regions of Romania (for example, the sampling and testing of wild birds in eastern and southern Romania, versus regular inspection and testing of domestic village poultry in central Romania). In the absence of information from the analysis of epidemic data, such disease control and surveillance decisions are difficult to make. Linking spatio-temporal patterns to epidemic dynamics is a first step in determining the most appropriate course of action in the face of an outbreak. Future work should examine the utility of these, and other, spatio-temporal modeling approaches for hypothesizing about potential mechanisms of spread to determine if these analytical methods can be generalized to locations where even less is known about the mechanism of epidemic spread.

REFERENCES

[1] Alexander D.J., A review of avian influenza in different bird species, Vet. Microbiol. (2000) 74:3–13.

[2] Auchincloss A.H., Roux A.V.D., Brown D.G., Raghunathan T.E., Erdmann C.A., Filling the gaps: spatial interpolation of residential survey data in the estimation of neighborhood characteristics, Epidemiology (2007) 18:469–478.

[3] Beato M.S., Terregino C., Cattoli G., Capua I., Isolation and characterization of an H10N7 avian influenza virus from poultry carcasses smuggled from China into Italy, Avian Pathol. (2006) 35:400–403.

[4] Besag J., York J., Mollie A., Bayesian image-restoration, with two applications in spatial statistics, Ann. Inst. Stat. Math. (1991) 43:1–59.

[5] Besag J., Kooperberg C., On conditional and intrinsic autoregressions, Biometrika (1995) 82:733–746.

[6] Bjornstad O.N., Ims R.A., Lambin X., Spatial population dynamics: analyzing patterns and processes of population synchrony, Trends Ecol. Evol. (1999) 14:427–432.

[7] Bjornstad O.N., Falck W., Nonparametric spatial covariance functions: estimation and testing, Environ. Ecol. Stat. (2001) 8:53–70.

[8] Burnham K.P., Anderson D.R., Model selection and multimodel inference: a practical information-theoretic approach, 2nd ed., Vol. XXVI, Springer, New York, 2002, 488 p.

[9] Carlin B.P., Louis T.A., Bayes and empirical Bayes methods for data analysis, Chapman & Hall/CRC, Boca Raton, 2000, 419 p.

[10] Craven P., Wahba G., Smoothing noisy data with Spline functions – estimating the correct degree of smoothing by the method of generalized cross-validation, Numerische Mathematik (1979) 31:377–403.

[11] Fang L.-Q., de Vlas S.J., Liang S., Looman C.W.N., Gong P., Xu B., et al., Environmental factors contributing to the spread of H5N1 avian influenza in mainland China, PLoS ONE (2008) 3:e2268.

[12] Geman S., Geman D., Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE Trans. Pattern Anal. Mach. Intell. (1984) 6:721–741.

[13] Gilbert M., Xiao X.M., Domenech J., Lubroth J., Martin V., Slingenbergh J., Anatidae migration in the western palearctic and spread of highly pathogenic avian influenza H5N1 virus, Emerg. Infect. Dis. (2006) 12:1650–1656.

[14] Gilks W.R., Richardson S., Spiegelhalter D.J., Markov chain Monte Carlo in practice, Chapman & Hall, Boca Raton, 1998, 486 p.

[15] Green P.J., Silverman B.W., Nonparametric regression and generalized linear models: a roughness penalty approach, Monographs on statistics and applied probability, No. 58, Chapman & Hall, 1994, 182 p.

[16] Handcock M.S., Meier K., Nychka D., Kriging and splines – an empirical comparison of their predictive performance in some applications – Comment, J. Am. Stat. Assoc. (1994) 89:401–403.

[17] Hanson B.A., Swayne D.E., Senne D.A., Lobpries D.S., Hurst J., Stallknecht D.E., Avian influenza viruses and paramyxoviruses in wintering and resident ducks in Texas, J. Wildl. Dis. (2005) 41: 624–628.

[18] Hutchinson M.F., Gessler P.E., Splines – More than just a smooth interpolator, Geoderma (1994) 62:45–67.

[19] Jones M.C., Davies S.J., Park B.U., Versions of kernel-type regression-estimators, J. Am. Stat. Assoc. (1994) 89:825–832.

[20] Khawaja J.Z., Naeem K., Ahmed Z., Ahmad S., Surveillance of avian influenza viruses in wild birds in areas adjacent to the epicenter of an outbreak in federal capital territory of Pakistan, Int. J. Poult. Sci. (2005) 4:39–43.

[21] Kilpatrick A.M., Chmura A.A., Gibbons D.W., Fleischer R.C., Marra P.P., Daszak P., Predicting the global spread of H5N1 avian influenza, Proc. Natl. Acad. Sci. USA (2006) 103:19368–19373.

[22] Krauss S., Walker D., Pryor S.P., Niles L., Li C.H., Hinshaw V.S., Webster R.G., Influenza A viruses of migrating wild aquatic birds in North America, Vector-Borne Zoonotic Dis. (2004) 4:177–189.

[23] Lunn D.J., Thomas A., Best N., Spiegelhalter D., WinBUGS - A Bayesian modelling framework: concepts, structure, and extensibility, Stat. Comput. (2000) 10:325–337.

[24] Morgan I.R., Kelly A.P., Epidemiology of an avian influenza outbreak in Victoria in 1985, Aust. Vet. J. (1990) 67:125–128.

[25] Munster V.J., Wallensten A., Baas C., Rimmelzwaan G.F., Schutten M., Olsen B., et al., Mallards and highly pathogenic avian influenza ancestral viruses, Northern Europe, Emerg. Infect. Dis. (2005) 11:1545–1551.

[26] Nychka D., Splines as local smoothers, Ann. Stat. (1995) 23:1175–1197.

[27] Oyana T.J., Dai D.J., Scott K.E., Spatiotemporal distributions of reported cases of the avian influenza H5N1 (bird flu) in Southern China in early 2004, Avian Dis. (2006) 50:508–515.

[28] Scott D.A., Rose P.M., Atlas of Anatidae populations in Africa and Western Eurasia, Wetlands International, Wageningen, Netherlands, 1996, 336 p.

[29] Spiegelhalter D.J., Best N.G., Carlin B.R., van der Linde A., Bayesian measures of model complexity and fit, J. R. Stat. Soc. Ser. B Stat. Methodol. (2002) 64:583–616.

[30] Van Borm S., Thomas I., Hanquet G., Lambrecht N., Boschmans M., Dupont G., et al., Highly pathogenic H5N1 influenza virus in smuggled Thai eagles, Belgium, Emerg. Infect. Dis. (2005) 11:702–705.

[31] Velkers F.C., Bouma A., Matthijs M.G.R., Koch G., Westendorp S.T., Stegeman J.A., Outbreak of avian influenza H7N3 on a turkey farm in the Netherlands, Vet. Rec. (2006) 159:403–405.

[32] Ward M.P., Maftei D., Apostu C., Suru A., Association between outbreaks of highly pathogenic avian influenza subtype H5N1 and wild bird populations, Zoonoses Public Health (2009) 56:1–9.

[33] Ward M.P., Maftei 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) 32:627–634.

[34] Ward M.P., Maftei D., Apostu C., Suru A., Geostatistical visualisation and spatial statistics for evaluation of the dispersion of epidemic highly pathogenic avian influenza subtype H5N1, Vet. Res. (2008) 39:22.

[35] Wikle C.K., Hierarchical Bayesian models for predicting the spread of ecological processes, Ecology (2003) 84:1382–1394.